22

RALPH A.M.J. WIJERS*

Gamma-ray bursts

The only feature that all but one (and perhaps all) of the very many proposed models have in common is that they will not be the explanation of γ-ray bursts. Unfortunately, limitations of time prevent me from telling you which model is the exception. (If I did so, I would suggest Black Hole ridden by Accretion as the favorite in the race with Glitch as a dark horse if only because so many different horses and jockeys are riding under that name.) – Mal Ruderman, Texas Conference, 1974

When Mal so ended the first theory review on gamma-ray bursts (GRBs), I was on the other side of the world, practicing my fractions. But had he spoken these same words at the end of the fifth Huntsville GRB meeting in October 1999, his favorite model (though not his reasoning) would have been instantly recognized by all present as a very plausible and popular contender for GRBs. His dark horse is still easily identified with soft gamma repeaters – *Plus ça change, plus que ça reste le même.* What happened roughly in between the two events is the topic of this tale. We shall see how the field emerged out of the mutual distrust of the Cold War adversaries in the 1960s, and was finally resolved by a determined effort of just about the entire membership of the United Nations in the late 1990s. Its progress, of course, was in fits and starts, interwoven with stagnation and frustration.

Since γ rays are absorbed by less than 1% of the atmosphere, they are quintessentially the domain of space science. The initial motivation for launching satellites with γ-ray capability was not, however, scientific. The purpose of the US Air Force's Vela satellite program was the detection of nuclear explosions in the upper atmosphere or in space, which were prohibited by the first ever test ban treaty. Like Jansky (radio), Penzias and Wilson (microwave), and Giacconi *et al*. (X-rays), it was what they were *not* looking for that started a new field: occasional flares of gamma rays lasting for just seconds, from extraterrestrial sources. Soon, other satellites confirmed the discovery.

A number of reviews have appeared that summarize the knowledge up to that point. Ruderman's paper (1975) and its observational counterpart by Strong *et al*. (1975) give an account of the initial explosion of discovery and thought. Higdon and Lingenfelter (1990) give a good overview of the pre-BATSE state of the field. The new discoveries made with the Burst and Transient Source Experiment (BATSE) on board the Compton Gamma Ray Observatory (CGRO), and the bewilderment following its overthrow of the galactic disk neutron star paradigm, are described in the review by Fishman and Meegan (1995). The consequences of the discovery of counterparts to GRBs and the state of the art at the end of the 1990s are described in reviews by Piran (1999; emphasis on theory and early stages of the burst) and by van Paradijs *et al*. (2000; emphasis on observations of afterglows – the later part of this chapter is based on parts of that review). To the historically inclined, review papers are only part of the story because they reflect more the final outcome of research than the struggle to reach that outcome. To get a flavor of the latter throughout the history of the subject, I recommend a number of conference proceedings. Theorists' struggles to come to terms with the phenomenon in the early days are reflected in the contributions to the 1974 Texas Symposium (Bergmann *et al*. 1975). Various stages of progress in the next decade are documented in the proceedings of meetings in La Jolla (Lingenfelter *et al*. 1982) and Stanford (Liang and Petrosian 1986). The first conference of the BATSE/CGRO

* State University of New York at Stony Brook, NY, USA; present address: Universiteit van Amsterdam, The Netherlands

Ralph A.M.J. Wijers, The Century of Space Science, 499–528

era and the accompanying wind of change is documented in the proceedings of the first Huntsville GRB symposium (Paciesas and Fishman 1992). After the discovery of counterparts, a meeting was called on short notice in Elba, Italy (26–27 May 1997). Its atmosphere of excitement mixed with astonishment and confusion is firmly etched in my memory; since the organizers wisely chose not to burden the attendees with writing proceedings papers, this feeling will forever remain the privilege of those who were there. The excitement was still very great during what I think of as the best meeting I ever attended: the fourth Huntsville GRB symposium (Meegan *et al*. 1998).

At this point I should describe my own path into the field, since it may have an impact on the tone and content of this review. I was in primary school when the discovery of GRBs was announced and remained blissfully unaware of them until some time during my PhD training in Amsterdam. I did not start active research in the field until Bohdan Paczyński's unlimited enthusiasm swept me into it in 1991. Therefore my accounts of the pre-BATSE era are entirely derived from the literature, and from conversations with those researchers I interacted with – by no means an unbiased sample of the field. Reviews are written with the sanitizing censorship of 20/20 hindsight and thus do not fully reflect the bouts of inspiration and confusion that characterize ongoing research: here lies a challenge for those who *were* there to see it. Having come into GRBs by way of Princeton also means that my allegiance in the distance scale debate of the early to mid-1990s was firmly extragalactic. While that point of view did eventually prevail, this outcome cannot have been so clearly predictable as I thought it was then, as is nicely illustrated by the reports of the "Great Debate" on the GRB distance scale (Fishman 1995, Lamb 1995, Paczyński 1995). A good perspective on the merits of such debates can be obtained from Trimble's (1995) discussion of the Shapley–Curtis debate. The field of soft gamma repeaters started as part of GRBs but has slowly split off as a separate discipline; its recent development has been similar in rapidity and magnitude to the GRB revolution, but this has been somewhat overshadowed by the latter. I discuss them only briefly in this chapter (Section 4).

A large part of this chapter is devoted to the events following 28 February 1997, when the first X-ray and optical counterpart to a GRB was discovered. The developments since then have been fast and furious: the first two counterparts settled the distance debate over GRBs firmly in favor of the extragalactic scale, making them the most powerful explosions since the Big Bang. Many issues are still not quite settled, and thus this part of the story has a decidedly less finished feel. Still, the discovery that they are somehow associated with young stars and can be observed out to very high redshifts opens up many new avenues of development. At the beginning of a new century, this field is likely to explode into a major branch of astrophysics.

The organization of this chapter is mostly historical, though where dictated by logic I complete the discussion of some topics when they first arise, to prevent too much subdivision of the narrative. It begins in the days when every gamma-ray photon had a name, and it ends with the demise of the CGRO, halfway between the popular and the logical end of the twentieth century. I dedicate this account to Jan van Paradijs, who taught me so much. Despite his untimely death on 2 November 1999, his science will last well into the twenty-first century.

1 THE EARLY DAYS

1.1 Coming in from the cold

As described by others in this volume, high-energy astronomy started in the 1950s, because it required high-altitude balloon and satellite technology to get equipment above the atmosphere for long times. Even 1% of the atmosphere absorbs most gamma rays, so being in space is essential (though some observations and new technology tests can be done with balloon experiments). In the history of GRBs, the other important point to note is that the early days of satellite astronomy coincided with the peak of the Cold War. A practical way was needed to verify compliance with the first test ban treaty, which prohibited the detonation of nuclear weapons in space. Since it was possible in principle to hide the initial explosion (by detonation behind the Moon), one type of detector was designed to catch the MeV gamma rays from radioactive decays in the debris cloud that would drift into view some time after the explosion. For this purpose, the US Air Force in collaboration with the Los Alamos and Sandia Laboratories operated the Vela satellites (Figure 1; the name is derived from the Spanish *velar*, which means "to watch over" or "to hold a vigil"). They flew in pairs on opposite sides of a 250 000 km diameter orbit.

Clandestine nuclear explosions were not found, but something else was – short, intense bursts of gamma rays from random directions in the sky. The first one on record dates from 2 July 1967 (Figure 1), but the discovery was not published until 1973 (Klebesadel *et al*. 1973). It is sometimes stated that this was due to the classified nature of the mission, but this is false: since the objective of the satellites was to deter violations of the treaty, their existence was well advertised (just as shops conspicuously advertise their closed-circuit cameras). However, since only a few people worked on the data, and did so by hand, it took time for them to convince themselves that spikes in the highly variable γ-ray background were real signals and had a cosmic origin. The basic method for doing this is still

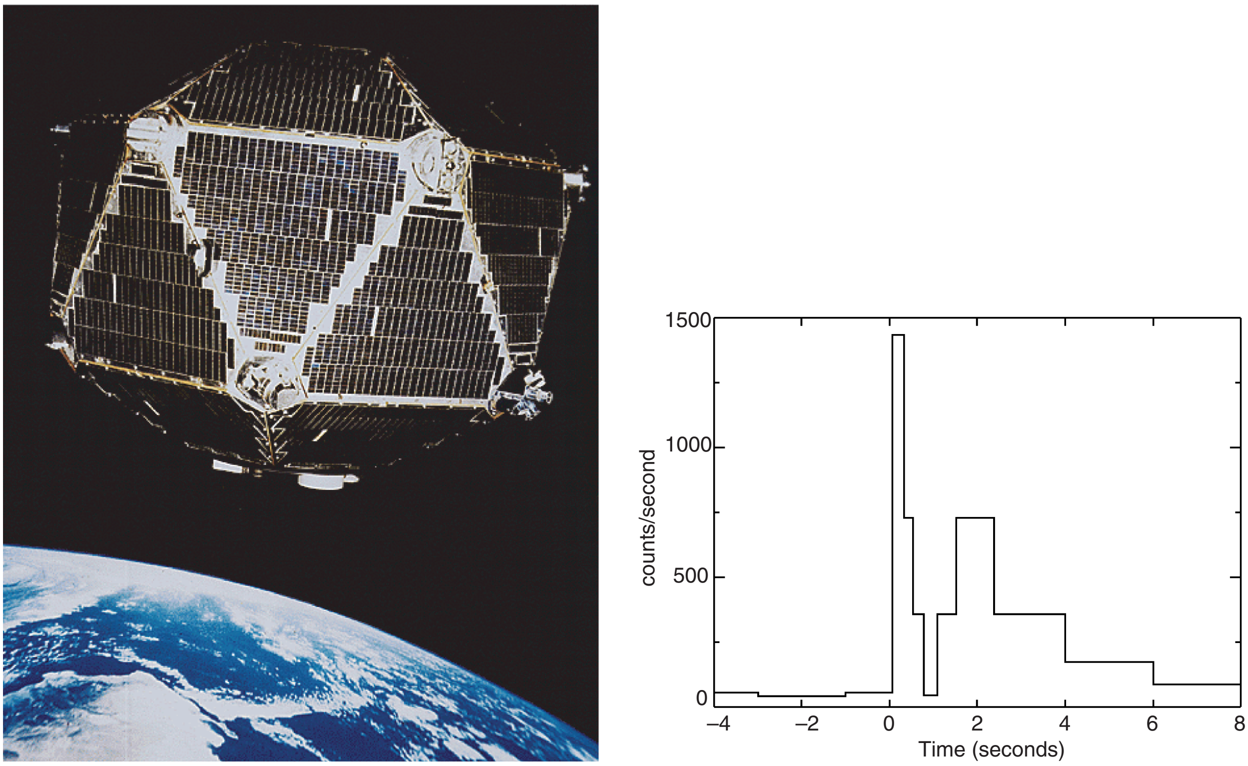

Figure 1 The Vela satellite (here Vela 5b), used by the USA to verify compliance with a ban on nuclear explosions in space (left). The light-curve of the first GRB it ever saw, on 2 July 1967, is also shown (right). (Images courtesy of "Astronomy Picture of the Day," at http://antwrp.gsfc.nasa.gov/apod/astropix.html.)

used in modern instruments: First, one requires more than one detector to see the signal; and second, one uses the difference in arrival time of the burst signal between two widely separated satellites to constrain the arrival direction (Figure 2). In this way, the early researchers eventually convinced themselves that the transient gamma-ray events they saw did not come from the Earth or the Sun.

It is interesting to note that the first observational paper already had a model to discuss. A few years earlier, Colgate (1968) had published his idea that a flash of gamma rays may be produced when the shock powering a Type II supernova breaks out of the surface of a red giant. Because the shock conserves energy, it becomes ever hotter as it moves down the density gradient, and at the point where the (Thomson) optical depth becomes unity, it emits photons with energy comparable to the electron rest mass. Since the medium is moving towards us with a Lorentz factor of about 1500, we see these blueshifted to GeV energies, and the flash time shortened to tens of microseconds. In later papers (Colgate 1973, 1974), he considers cooling before shock break-out and revises his estimates to more nearly the observed burst durations and photon energies. The scanned version of the 1974 paper held in NASA's ADS database (http://adsabs.harvard.edu/abstract_service.html) contains an interesting scribbled note to the conclusion section: the reader comments that such intense flashes in TeV gamma rays should be visible with the TeV telescope at Mount Hopkins. With the kind help of the people at ADS, I discovered that the owner of that copy of the *Astrophysical Journal* was Professor J. Grindlay, who confirmed to me that he had indeed considered the issue seriously. The problem, of course, was the same as with all other counterpart searches: one did not have any means of getting the very early and accurate locations required to aim a telescope at the GRB locations quickly enough.

Klebesadel *et al*. (1973) commented that there have been no detected supernovae near the GRBs they detected, and therefore dismissed the model. Very recent developments have put supernovae very much back into the picture (Section 6.2).

The first few facts about GRBs emerged quickly. Cline *et al*. (1973) used a hard X-ray detector on the IMP 6 satellite, designed for solar flare observations, to confirm the discovery and show that the spectra of the bursts really did peak in gamma rays (thus excluding the possibility that they were just high-energy tails of some type of X-ray event). A telescope with some directional capability on board OSO 7 confirmed the extrasolar origin of one burst

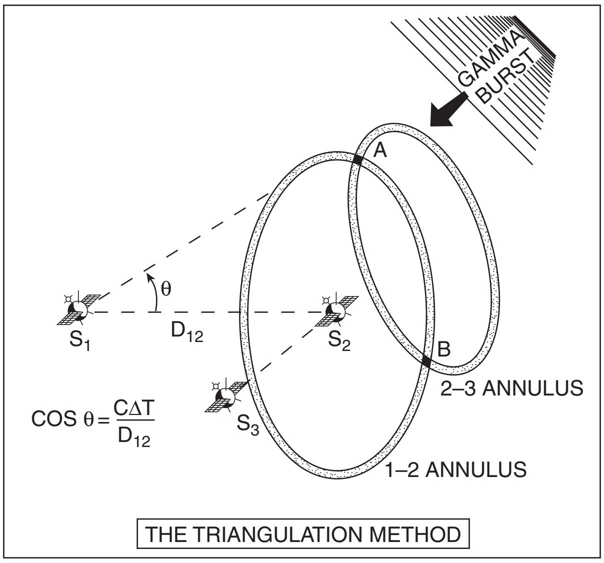

Figure 2 When two or more satellites detect a GRB, the arrival time difference of the signal between each pair of satellites can be used to constrain the location of the burst to a narrow circle on the sky. With three or more satellites, a true error box – sometimes as small as a few arc minutes across – can be constructed. This principle underlies the Interplanetary Network (IPN), still in operation. (Image courtesy of K. Hurley.)

(Wheaton *et al.* 1973). Incidentally, the first GRBs for which a cosmic origin could be established were observed with the Vela 5 and 6 series satellites, which had enough timing accuracy to pin down the directions to GRBs sufficiently well. The burst of 2 July 1967 was detected with the Vela 3 and 4 series, and is classified as the first GRB only on the basis of its similarity to the later ones, not because the direction to it is known to be inconsistent with a solar or terrestrial origin.

The first published Soviet observation of a GRB, using the Kosmos 461 satellite, dates from 1974, and is by the Leningrad group (Mazets *et al.* 1974). Igor Mitrofanov has told me that they too knew earlier that they might have something cosmic on their hands, probably in about 1971.

1.2 The earliest observations and theories

From here on the pace of discovery picked up quickly. In the review by Strong *et al.* (1975), in the proceedings of the seventh Texas Symposium, 34 bursts are listed. Their "sizes" (now called fluence or time-integrated flux) are in the range $10^{-5}-10^{-4}$ erg cm^{-2}. The spectra, where available, show a power law at low energies and then a bending to a steeper decline at a few hundred keV. At first this bending is fitted with a thermal bremsstrahlung function, that is, an exponential cutoff, but there is already one burst observed with Apollo 16 that has enough signal above 1 MeV to show that the high-energy part is also a power law (Metzger *et al.* 1974). The time histories are usually quite spiky, with overall durations of 0.1 to 100 s and sharp fluctuations down to 16 ms, the finest time resolution of the Velas. The locations of the bursts seem to show no significant preference for any part of the sky, except that two events are consistent with the location of the well-known accreting source Cyg X-1 (a fact that contributed considerably to Ruderman's pronouncement on theories). It is also noted that the fluence distribution for bright events is $N(>S) \propto S^{-3/2}$, the expectation for sources whose density does not depend on distance (often referred to as the "Euclidean" distribution). Cline (1975), in the same volume, confirms the high-energy end of the spectra to be a power law, and argues that the claimed flattening of the fluence distribution at low fluence is not significant. With hindsight, his balloon detections must have been something other than GRBs, since the fluence distribution is now known to be flatter than Euclidean at the fluence values he reports.

Enter Ruderman, who, also in the same volume, has the task of deciding which of the models to date survive scrutiny. The overview of models later given by Nemiroff (1994) lists 15 models dated 1974 or earlier; most of them are discussed in Ruderman's (1975) paper. He recognizes the limitations of theorists: "Most theoretical astrophysicists function well in only one or two normal modes. Therefore, we often tend to twist rather strenuously to convince ourselves and others that observations of new phenomena fit into our chosen specialties." He finds the brilliant solution of inviting as many as possible to enlist: "For theorists who may wish to enter this broad and growing field, I should point out that there are a considerable number of combinations, for example, comets of antimatter falling onto white holes, not yet claimed."

Ruderman largely dismisses the extragalactic models, in part because Colgate's supernova model is damaged by the non-detection of supernovae coincident with known GRBs. However, he also derives a constraint on distances from a black-body limit to the luminosity, assuming the observed photon energies are comparable to the thermal energies of the emitting particles, which virtually excludes extragalactic models beyond 30 Mpc. He acknowledges this to be wrong for some radiation mechanisms, but nonetheless the constraint appears to have impressed the audience. One of his acknowledged ways around the constraint is synchrotron radiation, since for this mechanism the emitted photons do have much lower energies than the particles that emit them; of course, we now know synchrotron radiation to be the dominant contributor. Among the galactic models he discusses a variety of "conventional" models, namely magnetic flares on various types of object and accretion events in compact objects. He then presents a number of

Springer
http://www.springer.com/978-0-7923-7196-0

The Century of Space Science
Bleeker, J.A.; Geiss, J.; Huber, M. (Eds.)
2001, XLIX, 1846 p., Hardcover
ISBN: 978-0-7923-7196-0