PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Gamma–ray bursts

Martin J. Rees

Phil. Trans. R. Soc. Lond. A 2000 **358**, 853-867
doi: 10.1098/rsta.2000.0563

Gamma-ray bursts

By Martin J. Rees

Institute of Astronomy, University of Cambridge,
Madingley Road, Cambridge CB3 0HA, UK

Gamma-ray bursts, an enigma for more than 25 years, are now coming into focus. They involve extraordinary power outputs, and highly relativistic dynamics. The 'trigger' involves stellar-mass compact objects. The most plausible progenitors, ranging from NS–NS mergers to collapsars (sometimes called 'hypernovae') eventually lead to the formation of a black hole with a debris torus around it, the extractable energy being up to 10^{54} erg. Magnetic fields may exceed 10^{15} G. Details of the afterglow may be easier to understand than the initial trigger. Bursts at very high redshift can be astronomically important as probes of the distant Universe.

Keywords: shocks; gamma-rays; neutron stars

1. Introduction

Astrophysics is an observation-led subject: theorists generally play a subsidiary role—certainly a more modest one than their counterparts in, for instance, particle physics. But in the case of gamma-ray bursts the lag between gathering data and making sense of it has been especially embarrassing, even by astrophysical standards. Until two years ago, there was absolutely no consensus on what, or even where, the bursts are. Owing primarily to the impetus of the Italian/Dutch Beppo-SAX satellite, there is now general agreement that the bursts (or at least a substantial subset of them) are at high redshifts. In this paper I shall review the basic data, and then discuss the physics of the energy production and its conversion into an intense burst and a prolonged afterglow, emphasizing the role of strong magnetic fields.

2. History

The story of gamma-ray bursts (GRBs) started in the late 1960s, when American scientists at Los Alamos developed the Vela satellites, whose purpose was to monitor clandestine nuclear tests in space by detecting the associated gamma-ray emission. Occasional flashes, lasting a few seconds, were indeed recorded. It took several years before these were realized to be genuine; and to be natural, rather than sinister, phenomena. In 1973 a paper was published by Klebesadal, Strong & Olson entitled *Observations of gamma-ray bursts of cosmic origin*. This classic paper reported 16 short bursts of photons in the energy range between 0.2 and 1.5 MeV, which had been observed during a three-year period using widely separated spacecraft. The burst durations ranged from less than 0.1 s up to about 30 s, but complicated time-structure was observed within the longer bursts. It was apparent that the bursts came neither from the Earth nor from the Sun, but little else was clear at that time.

Phil. Trans. R. Soc. Lond. A (2000) **358**, 853–867

It did not take long for the theorists to become enthusiastically engaged. At the 'Texas Conference on Relativistic Astrophysics', held in December 1974, Ruderman (1975) gave a review of models and theories. He presented an exotic menu of alternatives that had already appeared in the literature, involving supernovae, neutron stars, flare stars, antimatter effects, relativistic dust, white holes, and some even more bizarre options. He noted also the tendency (still often apparent; not only among astrophysicists!) for theorists to 'strive strenuously to fit new phenomena into their chosen specialities'.

During the 1970s and 1980s, data on GRBs accumulated, thanks to a number of satellites. Particular mention should be made of the impressive contributions by Mazets and his colleagues in Leningrad. Also important were the extended observations made by the Pioneer Venus Orbiter (PVO). The number of detected bursts rose faster than the number of proposed models—indeed, some of the crazier early conjectures were actually ruled out.

During that period, three classes of models were pursued: those in which the bursts were, respectively, in the galactic disc (at distances of a few hundred parsecs), in the halo (at distances of tens of kiloparsecs) and at cosmological (several gigaparsecs) distances. The most popular idea during the 1980s was that the bursts were relatively local, probably in our galactic disc, and due to magnetospheric phenomena or 'glitches' on old neutron stars (defunct pulsars).

It was already clear that two kinds of statistical information could in principle decide the location of GRBs, as soon as enough data had accumulated, and selection effects were understood. One was the number-versus-intensity of the events, which tells us whether they are uniformly distributed in Euclidean space, or whether we are in some sense probing out to the edge of the distribution. The other was the degree of anisotropy.

The counts of GRBs were already suspected to be flatter than the classic Euclidean slope (-1.54 on a log–log plot), since otherwise it was hard to understand why large-area detectors flown in balloons didn't detect more faint bursts. Flat counts would not of course have been unexpected if the bursts came from within a bounded system such as our galactic disc. However, one would then have expected an anisotropic distribution of sources over the sky: an enhancement towards the galactic plane, and perhaps also towards the galactic centre. It was therefore a real surprise when the Compton Gamma Ray Observatory (GRO) satellite, whose 'burst and transient source experiment' (BATSE) offered systematic all-sky coverage, with good sensitivity over the photon energy range 30 keV–1.9 MeV, revealed that the bursts are highly isotropic over the sky. More than 2500 have now been recorded, and there is still no statistical evidence for any dipole or quadrupole anisotropy, nor for any two-point correlation (Briggs *et al*. 1997). The lack of any enhancement either towards the plane of the galaxy, or towards the galactic centre, was recognized as a very severe constraint on the hypothesis that bursts come from the galaxy. (Note that the bursts cannot be so ultralocal that we were not even probing as far as a scale height in the disc: this would yield isotropy, but is ruled out by the flatness of the number counts.) The 'non-Euclidean' counts imply that the surveys are probing to distances where the sources are, for some reason, thinning out; the problem is to reconcile this with the isotropy.

The BATSE experiment has produced a large body of data on the spectra and time structure of bursts. Despite the large variety, there is little doubt that gamma-ray bursts are a well-defined class of objects, distinguished spectrally from phenomena such as X-ray bursters, and also from the so-called 'soft gamma repeaters' which have substantially softer spectra and are associated with young highly magnetized neutron stars within our galaxy. There are some apparent correlations (though no specially strong ones) between various GRB properties. For instance, shorter bursts tend to be stronger and to have somewhat harder spectra; the histogram of burst durations exhibits two distinct-seeming peaks; and the counts deviate most from the Euclidean slope (i.e. are flatter) for bursts with harder spectra (Pendleton *et al.* 1996).

The isotropy evidence tilted the balance of opinion strongly towards a cosmological interpretation of GRBs. It still remained conceivable, however, that the bursts were in our galactic halo, but sufficiently far out that the Sun's 8 kpc offset from the galactic centre did not produce a dipole asymmetry. In April 1995 an interesting debate took place on the location of GRBs, in which the two main protagonists were Don Lamb and Bohdan Paczyński (a written version of the argument appears in Lamb (1995) and Paczyński (1995)). It was held in the Washington Museum of Natural History, to commemorate the 75th anniversary of the famous debate that took place there between Shapley and Curtis on whether some of the so-called 'nebulae' were stellar systems (i.e. other galaxies) beyond our Milky Way. I had the privilege of moderating this debate, perhaps because I was one of the few people who had explored both options (cf. Podsiadlowski *et al.* 1995).

There was an agreement among all participants on the kind of new evidence that could settle the issue. Most valuable of all would be firm identification with objects detectable in other wavebands. The stumbling block here is the poor positional accuracy of most gamma-ray detectors. BATSE itself has error circles of 1 or 2° for the brightest bursts, and of more than 5° for the fainter ones. Some bursts had been pinned down with a precision of minutes of arc (or better) by triangulation experiments involving deep-space probes: this latter technique uses the rapid time structure, which, when recorded and timed by detectors separated by 10 light minutes or more, allows accurate positioning. It generally took several days, however, to recover and correlate the required data, and to calculate positions by this technique: nothing was found by looking in any of the resulting error boxes.

The controversies in the Shapley–Curtis debate were settled within a few years; our knowledge of extragalactic astronomy thereby made a forward leap, and astronomers moved on to address more detailed issues. The GRB distances were actually settled even more quickly and decisively: the crucial step was the detection of gradually fading afterglows within some of the arcminute-scale error boxes that the BeppoSax satellite was able to supply within a few hours of the burst. The first such detection occurred in February 1997 (Costa *et al.* 1997; van Paradijs *et al.* 1997); still more crucial was an event, GRB 970528, detected in May of that year, whose optical afterglow (Metzger *et al.* 1997) displayed strong absorption features with $z = 0.835$, indicating that it probably lay in a galaxy with that redshift. Subsequently, further interest was aroused by the report of an afterglow for the burst GRB 971214 at a redshift $z = 3.4$, whose energy output in gamma-rays would amount to 3×10^{53} erg if the emission were isotropic (Kulkarni *et al.* 1998). More recently, 2×10^{54} erg has been claimed for GRB 990123 (Bloom *et al.* 1998). The energies and luminosities would of course be lower if the radiation were beamed rather than isotropic and, indeed, some degree of beaming is a natural consequence of almost all models.

3. What is the trigger?

The photon *luminosity*, for the few-seconds duration of a typical burst, is of course colossal: it exceeds by many thousands the most extreme output from any active galactic nucleus (thought to involve supermassive black holes), and is 14 orders of magnitude above the Eddington limit for a stellar-mass object. The total *energy*, however, is not out of line with some other phenomena encountered in astrophysics—indeed it is reminiscent of the energy released in the core of a supernova, the big difference being that the primary sudden event (with a time-scale of seconds) is not smothered by a stellar envelope, as in a supernova, but manifests itself in hard radiation that escapes more promptly.

Unless they are beamed into less than 1% of the solid angle, the triggers for GRBs are thousands of times rarer than supernovae. A widely discussed possibility is coalescence of binary neutron stars (see, for example, Narayan *et al*. 1992). Systems such as the famous binary pulsar will eventually coalesce, when gravitational radiation drives them together. When a neutron star (NS) binary coalesces, the rapidly spinning merged system would be too massive (for most presumed equations of state) to form a single NS; on the other hand, the total angular momentum is probably too large to be swallowed immediately by a black hole. The expected outcome, after a few milliseconds, would therefore be a spinning black hole (BH), orbited by a torus of neutron-density matter.

Other types of progenitor have been suggested—e.g. an NS–BH merger, where the neutron star is tidally disrupted before being swallowed by the hole; the merger of a white dwarf with a black hole; or a category labelled as hypernovae or collapsars, where the collapsing core is too massive to become a neutron star, but has too much angular momentum to collapse quietly into a black hole (as in a so-called 'failed supernova'). The simple point that I wish to stress, however, is that a BH surrounded by a neutron-density torus is a common feature of all these models; moreover the overall energetics of these various progenitors differ by at most an order of magnitude, the spread reflecting the differing spin energy in the hole and the different masses left behind in an orbiting torus. (There has been some confusion on this point in the literature, through failure to appreciate that the dominant energy from an NS–NS event comes after a black hole forms, rather than during the precursor stage that Narayan *et al*. (1992) discussed.) How might such a system generate relativistic outflow or a release of electromagnetic energy?

4. Energy from a black hole and debris torus?

Two large reservoirs of energy are in principle available: the binding energy of the orbiting debris, and the spin energy of the black hole. The first can provide up to 42% of the rest mass energy of the torus, for a maximally rotating black hole: the second can provide up to 29% (for a maximal spin rate) of the mass of the black hole itself. How can the energy be transformed into outflowing relativistic plasma after such a coalescence event? There seem to be two options. The first is that some of the energy released as thermal neutrinos is reconverted, via collisions outside the dense core, into electron–positron pairs or photons. The second option (which allows higher efficiency) is that strong magnetic fields anchored in the dense matter convert the rotational energy of the system into a Poynting-dominated outflow, rather as in pulsars. Let us consider these two options in turn.

(i) Neutrinos could give rise to a relativistic pair-dominated wind if they converted into pairs in a region of low baryon density (e.g. along the rotation axis, away from the equatorial plane of the torus). The $\nu\bar{\nu} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ process can tap the thermal energy of the torus produced by viscous dissipation. For this mechanism to be efficient, the neutrinos must escape before being advected into the hole; on the other hand, the efficiency of conversion into pairs (which scales with the square of the neutrino density) is low if the neutrino production is too gradual. Typical estimates suggest a limit of less than *ca.* 10^{51} erg (Ruffert 1997; Ruffert *et al.* 1997; Ruffert & Janka 1999; Woosley *et al.* 1999; Popham *et al.* 1999), except perhaps in the 'collapsar' or failed SN Ib case where Popham *et al.* (1999) estimate 2×10^{53} erg for optimum parameters. If the pair-dominated plasma were collimated into a solid angle Ω_j, then of course the apparent 'isotropized' energy would be larger by a factor $(4\pi/\Omega_j)$, but unless Ω_j is less than *ca.* 10^{-2}–10^{-3} this may fail to reach the apparent isotropized energy of the most luminous bursts.

(ii) An alternative way to tap the torus energy is via magnetic fields threading the torus (Paczyński 1991; Narayan *et al.* 1992; Mészáros & Rees 1997*b*; Katz & Piran 1997). Even before the BH forms, an NS–NS merging system might lead to winding up of the fields and dissipation in the last stages before the merger (Mészáros & Rees 1992; Vietri 1997).

The above mechanisms tap the rotational energy available in the debris torus. However, a hole formed from a coalescing compact binary is guaranteed to be rapidly spinning, and, being more massive, could contain a larger reservoir of energy than the torus; this energy, extractable in principle through MHD coupling to the rotation of the hole by the Blandford & Znajek (1977) (BZ) effect, could be even larger than that contained in the orbiting debris (Mészáros & Rees 1997*b*; Paczyński 1998). Collectively, any such MHD outflows have been referred to as Poynting jets.

Simple scaling from the familiar results of pulsar theory tells us that fields of order 10^{15} G are needed to carry away the rotational or gravitational energy in the time-scales of tens of seconds (Usov 1994; Thompson 1994). If the magnetic fields do not thread the BH, then a Poynting outflow can at most carry the gravitational binding energy of the torus. This is between 0.06 and 0.42 of the rest mass energy of the torus, depending on the spin of the hole. The torus mass in an NS–NS merger is $M_{\mathrm{t}} \sim 0.1 M_{\odot}$ (Ruffert *et al.* 1997), and for an NS–BH or WD–BH merger it may be $M_{\mathrm{t}} \sim 1 M_{\odot}$ (Paczyński 1998; Fryer & Woosley 1998). The extractable energy could amount to several times $10^{53}\epsilon(M_{\mathrm{t}}/M_{\odot})$ erg, where ϵ is the efficiency in converting gravitational into MHD jet energy. Tori masses even higher than *ca.* $1\, M_{\odot}$ may occur in scenarios involving massive supernovae. Conditions for the efficient escape of a high-Γ jet may, however, be less propitious if the 'engine' is surrounded by an extensive envelope.

If magnetic fields of comparable strength thread the BH, its rotational energy offers an extra (and even larger) source of energy that can in principle be extracted via the BZ mechanism (Mészáros & Rees 1997*b*). For a maximally rotating BH, this is $0.29 M_{\mathrm{BH}} c^2$ erg, multiplied, of course, by some efficiency factor. A near-maximally rotating black hole is guaranteed in an NS–NS merger. The central BH will have a mass of *ca.* $2.5 M_{\odot}$; the NS–BH merger and hypernovae models may not produce quite such rapidly spinning holes, but the hole masses are larger, so the expected rotational energy should be comparable. Spinning holes can thus power a jet of up to *ca.* 1.5×10^{54} erg. Even allowing for low total efficiency (say 30%), a system powered by the torus binding energy would only require a modest beaming of the gamma-rays by a factor $(4\pi/\Omega_j) \sim 20$, or no beaming if the jet is powered by the BZ mechanism, to produce the equivalent of an isotropic energy of up to 10^{54} erg. The fields of *ca.* 10^{15} G required for efficient electromagnetic extraction of energy are not significantly higher than those directly inferred in 'magnetors' (cf. Kouveliotou 1999).

Even in the hypernovae model, magnetic extraction of energy seems likely to be more efficient than relativistic pairs generated by neutrons. (It is, however, harder to quantify than the latter, and therefore has not been included in the simulations by Woosley, Janka and their collaborators.)

5. The gamma-ray emission mechanism

Well-known arguments connected with opacity, variability time-scales and so forth (see, for instance, Piran 1997) require highly relativistic outflow. Best-guess numbers are Lorentz factors Γ in the range 10^2–10^3, allowing rapidly variable emission to occur at radii in the range 10^{14}–10^{16} cm. The entrained baryonic mass would need to be below $10^{-4}M_\odot$ to allow these high relativistic expansion speeds.

Because the emitting region must be several powers of ten larger than the compact object that acts as a 'trigger', there is a further physical requirement: the original energy outflowing in a magnetized wind would, after expansion, be transformed into bulk kinetic energy (with associated internal cooling). This energy cannot be efficiently radiated as gamma-rays unless it is rerandomized. This requires relativistic shocks. Impact on an external medium would randomize half of the initial energy merely by reducing the expansion Lorentz factor by a factor of two. Alternatively, there may be internal shocks within the outflow: for instance, if the Lorentz factor in an outflowing wind varied by a factor of more than two, then the shocks that developed when fast material overtakes slower material would be internally relativistic (Rees & Mészáros 1994).

In an unsteady outflow, if Γ were to vary by a factor of more than two on a time-scale δt, internal shocks would develop at a distance $\Gamma^2 c\delta t$, and randomize much of the energy. For instance, if Γ ranged between 500 and 2000, on a time-scale of 1 s, efficient dissipation would occur at 3×10^{16} cm.

There is a general consensus that the longer complex bursts must involve internal shocks, though simple sharp pulses could arise from an external shock interaction (the latter would in effect be the precursor of the afterglow). An external shock moving into a smooth medium would obviously give a burst with a simple time-profile. A 'blobby' external medium could give features, but only if the covering factor of blobs is low, implying modest efficiency.

Even if the bursts were caused by a completely standardized set of objects, their appearance would be likely to depend drastically on orientation relative to the line of sight. Along any given line of sight, the time-structure would be determined partly by the advance of jet material into the external medium, but probably even more by internal shocks within the jet, which themselves depend on the evolution of the torus, from its formation to its eventual swallowing or dispersal.

The radiation processes for the gamma-rays are probably no more than synchrotron radiation. This would imply the presence of magnetic fields where the shocks occur. If the outflow from the central trigger is Poynting dominated, then a field of 10^{15} G at (say) 10^7 cm would imply a comoving field of $10^7(\Gamma/100)^{-1}$ G out at 10^{13} cm—strong enough to ensure rapid cooling of shocked relativistic electrons. (Note, conversely, that even if magnetic fields were not important near the central trigger, they must be present, with about the same amount of flux that Poynting-dominated models require, at the location of the actual gamma-ray emission.)

We are a long way from modelling what triggers gamma-ray bursts. If we had a precise description of the dynamics, along with the baryon content, magnetic field and Lorentz factor of the outflow, we could maybe predict the gross time-structure. But we could not predict the intensity or spectrum of the emitted radiation—still less answer key questions about the emission in other wavebands—without also having an adequate theory for particle acceleration in relativistic shocks. We need the answers from plasma physicists to the following poorly understood questions.

(i) Do ultrarelativistic shocks yield power laws? The answer probably depends on the ion–positron ratio, and on the relative orientation of the shock front and the magnetic field (see, for example, Gallant *et al*. 1992).

(ii) In ion–electron plasmas, what fraction of the energy goes into the electrons?

(iii) Even if the shocked particles establish a power law, there must be a low-energy break in the spectrum at an energy that is in itself relativistic. But will this energy, for the electrons, be $\varGamma m_{\mathrm{p}} c^2$ (or even, if the positive charges are heavy ions like Fe, $\varGamma m_{\mathrm{Fe}} c^2$)?

(iv) Can ions be accelerated up to the theoretical maximum where the gyroradius becomes the scale of the system? If so, the burst events could be the origin of the highest energy cosmic rays.

(v) Do magnetic fields get amplified in shocks? (This is relevant to the magnetic field in the swept-up external matter outside the contact discontinuity, and determines how sharp the external shock actually is.)

6. Intrinsic time-scales

A question which has remained largely unanswered so far is what determines the characteristic duration of bursts, which can extend to tens, or even hundreds, of seconds. This is of course very long in comparison with the dynamical or orbital time-scale for the 'triggers', which is measured in milliseconds. While bursts lasting hundreds of seconds can easily be derived from a very short impulsive energy input, this is generally unable to account for a large fraction of bursts which show complicated light curves. This hints at the desirability for a 'central engine' lasting much longer than a typical dynamical time-scale.

Observationally (Kouveliotou *et al*. 1993), the short (less than *ca.* 2 s) and long (greater than *ca.* 2 s) bursts appear to represent two distinct subclasses, and one early proposal to explain this was that accretion-induced collapse (AIC) of a white dwarf (WD) into an NS plus debris might be a candidate for the long bursts, while NS–NS mergers could provide the short bursts (Katz & Canel 1996). As indicated by Ruffert *et al*. (1997), $\nu\bar{\nu}$ annihilation will generally tend to produce short bursts of less than *ca.* 1 s in NS–NS systems, requiring collimation by 10^{-1}–10^{-2}, while Popham *et al*. (1999) argued that in WD/He–BH systems longer $\nu\bar{\nu}$ bursts may be possible.

An acceptable model requires that the surrounding torus should not completely drain into the hole, or be otherwise dispersed, on too short a time-scale. There have been some discussions in the literature of possible 'runaway instabilities' in relativistic tori (Nishida *et al*. 1996; Abramowicz *et al*. 1998; Daigne & Mochkovitch 1997): these are analogous to the runaway Roche lobe overflow predicted, under some conditions, in binary systems. These instabilities can be virulent in a torus where the specific angular momentum is uniform throughout, but are inhibited by a spread in angular momentum. In a torus that was massive and/or thin enough to be self-gravitating, bar-mode gravitational instabilities could lead to further redistribution of angular momentum and/or to energy loss by gravitational radiation within only a few orbits. Whether a torus of given mass is dynamically unstable depends on its thickness and stratification, which in turn depends on internal viscous dissipation and neutrino cooling.

The disruption of a neutron star (or any analogous process) is almost certain to lead to a situation where violent instabilities redistribute mass and angular momentum within a few dynamical time-scales (i.e. in much less than a second). A key issue for gamma-ray burst models is the nature of the surviving debris after these violent processes are over: what is the maximum mass of a remnant disc/torus which is immune to very violent instabilities, and which can therefore in principle survive for long enough to power the bursts? If the torus results from the disruption of a compact binary, then the *residual* mass left over after violent instabilities on a dynamical time-scale have done their work is the relevant M_t in the above expressions (in § 3) for the extractable energy of the torus. In the collapsar models discussed by Fryer, Woosley and their collaborators, the torus is not created suddenly, but is replenished by infall from the degenerate stellar core on a time-scale of *ca.* 10 s. In these latter models, the long durations arise naturally, since they do not require a low viscosity (and long residence time) in the relativistic torus.

If the trigger is to liberate its energy over a period of 10–100 s via Poynting flux—either through a relativistic wind 'spun off' the torus or via the BZ mechanism—the required field is a few times 10^{15} G. A weaker field would extract inadequate power; on the other hand, if the large-scale field were even stronger, then the energy would be dumped too fast to account for the longer complex bursts. It is not obvious why the fields cannot become even higher. Note that the virial limit is $B_\mathrm{v} \sim 10^{17}$ G.

Kluźniak & Ruderman (1998) note that, starting with 10^{12} G, it only takes of the order of a second for simple winding to amplify the field to 10^{15} G; amplification in a newly formed torus could well occur more rapidly, for instance via convective instabilities, as in a newly formed neutron star (cf. Thompson & Duncan 1993; Thompson 1994). Kluźniak & Ruderman (1998) suggest, however, that the amplification may be self-limiting because magnetic stresses would then be strong enough for flares to break out. A magnetic-field configuration capable of powering the bursts is likely to have a large-scale structure. Flares and instabilities occurring on the characteristic (millisecond) dynamical time-scale would cause substantial irregularity or intermittency in the overall outflow that would manifest itself in internal shocks. There is thus no problem in principle in accounting for sporadic large-amplitude variability, on all time-scales down to a millisecond, even in the most long-lived bursts. Note also that it only takes a residual torus (or even a cold disc) of $10^{-3} M_\odot$ to confine a field of 10^{15} G, which can extract energy from the black hole via the BZ mechanism.

7. How much beaming?

Computer simulations of compact object mergers and black-hole formation can address the fate of the bulk of the matter, but there are some key questions that they cannot yet tackle. In particular, high resolution of the outer layers is needed because even a tiny mass fraction of baryons loading down the outflow severely limits the attainable Lorentz factor—for instance a Poynting flux of 10^{53} erg could not accelerate an outflow to $\Gamma > 100$ if it had to drag more than *ca.* $10^{-4} M_{\odot}$ of baryons with it. Further two-dimensional numerical simulations of the merger and collapse scenarios are under way largely using Newtonian dynamics, and the numerical difficulties are daunting. There may well be a broad spread of Lorentz factors in the outflow—close to the rotation axis Γ may be very high; at larger angles away from the axis, there may be an increasing degree of entrainment, with a corresponding decrease in Γ. Even if the outflow is not narrowly collimated, some beaming is expected because energy would be channelled preferentially along the rotation axis. Moreover, we would expect baryon contamination to be lowest near the axis, because angular momentum flings material away from the axis, and any gravitationally bound material with low angular momentum falls into the hole. In hypernovae, the envelope is rotating only slowly and thus would not initially have a marked centrifugal funnel; even 10^{53} erg would not suffice to blow out more than a narrow cone of the original envelope with a Lorentz factor of more than 100. So in these models the gamma-rays would be restricted to a narrow beam, even though outflow with a more moderate Lorentz factor (relevant to the afterglow) could be spread over a wider range of angles. A wide variety of burst phenomenology could be attributable to a standard type of event being viewed from different orientations.

Two further effects render the computational task of simulating jets even more challenging. The first stems from the likelihood that any entrained matter would be a mixture of protons and neutrons (neutrons, being unconstrained by magnetic fields, could also drift into a jet from the denser walls at its boundary). If a streaming velocity builds up between ions and neutrons (i.e. if they have different Lorentz factors in the outflow), then interactions can lead to dissipation even in a steady jet where there are no shocks (Derishev *et al.* 1999). A second possibility (Mészáros & Rees 1998*a*, *b*) is that entrained ions in a relativistic jet could become concentrated in dense filaments confined by the magnetic field. As already mentioned, the comoving field strength, even out at 10^{13} cm, is of order 10^6 G. Trapped filaments of iron-rich thermal, with density up to 10^{19} cm^{-3} and with kT of the order of a keV, could be confined by such fields. Such filaments must of course have a small volume-filling factor: otherwise they would load down the jet too much. However, in these strong fields the gyroradii would be so small that filaments could survive against thermal conduction and other diffusion processes even if their dimensions (transverse to the field) were less than 100 cm. Such thin filaments can provide a large covering factor even while filling a tiny fraction of the volume. If they were moving relativistically outwards, they could contribute ultra-blue-shift spectral features—for instance, K-edges of Fe could be shifted up to hundreds of keV.

8. Brief comments on the afterglows

The discovery of afterglows not only has extended observations to longer time-scales and other wavebands, making the identification of counterparts possible, but also provided confirmation for much of the earlier work on the fireball shock model of GRB, in which the gamma-ray emission arises at radii of 10^{13}–10^{15} cm (Rees & Mészáros 1992, 1994; Mészáros & Rees 1993; Paczyński & Xu 1994; Katz 1994; Sari & Piran 1995). In particular, this model led to the prediction of the quantitative nature of the signatures of afterglows, in substantial agreement with subsequent observations (Mészáros & Rees 1997*a*; Costa *et al*. 1997; Vietri 1997; Tavani 1997; Waxman 1997; Reichart 1997; Wijers *et al*. 1997).

Astrophysicists understand supernova *remnants* reasonably well, despite continuing uncertainty about the initiating explosion; likewise, we may hope to understand the afterglows of gamma-ray bursts, despite the uncertainties about the 'trigger' that I have already emphasized. The simplest hypothesis is that the afterglow is due to a relativistic expanding blast wave. The complex time-structure of some bursts suggests that the central trigger may continue for up to 100 s. However, at much later times all memory of the initial time-structure would be lost: essentially all that matters is how much energy and momentum has been injected, its distribution in angle and the mass fractions in shells with different Lorentz factors.

The simplest spherical afterglow model—where a relativistic blast wave decelerates as it runs into ambient matter, leading to a radiative output with a calculable spectrum, and a characteristic power law decay—has been remarkably successful at explaining the gross features of the GRB 970228, GRB 970508 and other afterglows (see, for example, Wijers *et al*. 1997). The gamma-rays we receive come only from material whose motion is directed within one degree of our line of sight. They therefore provide no information about the ejecta in other directions: the outflow could be isotropic, or concentrated in a cone of any angle substantially larger than 1° (provided that the line of sight lay inside the cone). At observer times of more than a week, the blast wave would, however, be decelerated to a moderate Lorentz factor, irrespective of the initial value. The beaming and aberration effects are thereafter less extreme, so we observe afterglow emission not just from material moving almost directly towards us, but from a wider range of angles.

The afterglow is thus a probe for the geometry of the ejecta—at late stages, if the outflow is beamed, we expect a spherically symmetric assumption to be inadequate; the deviations from the predictions of such a model would then tell us about the ejection in directions away from our line of sight. It is quite possible, for instance, that there is relativistic outflow with lower Γ (heavier loading of baryons) in other directions (see, for example, Wijers *et al*. 1997); this slower matter could even carry most of the energy (Paczyński 1998). Rhoads (1997) noted that if the energy were channelled into a solid angle Ω_j, one expects a faster decay of Γ after it drops below $\Omega_j^{-1/2}$. A simple calculation using the usual scaling laws leads then to a steepening of the flux power law in time. Anisotropy in the burst outflow and emission affects the light curve at the time when the inverse of the bulk Lorentz factor equals the opening angle of the outflow. If the critical Lorentz factor is less than three or so (i.e. the opening angle exceeds 20°), such a transition might, however, be masked by the transition from ultrarelativistic to mildly relativistic flow, so quite generically it would be difficult to limit the late-time afterglow opening angle in this way if it exceeds 20°.

The beaming angle for the gamma-ray emission, requiring Γ to be greater than about 100, could be far smaller than for the overall relativistic outflow, and is much harder to constrain directly. The ratio of Ω_γ/Ω_x has been considered by Grindlay (1999) using data from Ariel V and HEAO-A1/A2 surveys, who did not find evidence for a significant difference between the deduced gamma-ray and X-ray rates, and concluded that higher sensitivity surveys would be needed to provide significant constraints. More promising for the immediate future, the ratio $\Omega_\gamma/\Omega_{\rm opt}$ can also be investigated observationally (see also Rhoads 1997). The rate of GRB with peak fluxes above $1\,{\rm ph\,cm^{-2}\,s^{-1}}$ as determined by BATSE is *ca.* $300\,{\rm yr}^{-1}$, i.e. $0.01\,{\rm sq}^\circ\,{\rm yr}^{-1}$. According to Wijers *et al.* (1998) this flux corresponds to a redshift of 3.

If the gamma-rays were much more narrowly beamed than the optical afterglow, there should be many 'homeless' afterglows, i.e. ones without a GRB preceding them. The transient sky at faint magnitudes is poorly known, but there are two major efforts underway to find supernovae down to about $R = 23$ (Garnavich *et al.* 1998; Perlmutter *et al.* 1998). These searches have by now covered a few tens of 'square degree years' of exposure and would be sensitive to afterglows of the brightness levels thus far observed. It therefore appears that the afterglow rate is not more than a few times $0.1\,{\rm sq}^\circ\,{\rm yr}^{-1}$. Since the magnitude limit of these searches allows detection of optical counterparts of GRB brighter than $1\,{\rm ph\,cm^{-2}\,s^{-1}}$, it is fair to conclude that the ratio of homeless afterglows to GRB is unlikely to exceed about 20. It then follows that $\Omega_\gamma > 0.05\Omega_{\rm opt}$, which combined with our limit to $\Omega_{\rm opt}$ yields $\Omega_\gamma > 0.02$. The true rate of events that give rise to GRB is therefore at most 600 times the observed GRB rate, and the opening angle of the ultrarelativistic gamma-ray-emitting material is no less than 5°. Combined with the most energetic bursts, this begins to pose a problem for the neutrino annihilation type of the GRB energy source.

Obviously, the above calculation is only sketchy and should be taken as an order of magnitude estimate at present. However, it should improve as more afterglows are detected and the modelling gets more precise.

9. Conclusions and prospects

There are two key questions regarding the 'trigger'. First, does it involve a black hole orbited by a dense torus (which I have advocated as a 'best buy')? Second, if so, can we decide between the various alternative ways of forming it: NS–NS, NS–BH or collapsar/hypernova?

The locations should help to settle the second question. This is because a collapsar/hypernova would be expected to lie in a region of recent star formation; on the other hand, a neutron-star binary could take hundreds of millions of years to spiral together, and could by then (especially if given a 'kick velocity' on formation) have moved many kiloparsecs from its point of origin (Bloom *et al.* 1998). There is also already tentative evidence that some detected afterglows arise in relatively dense gaseous environments—e.g. by evidence for dust in GRB 970508 (Reichart 1998) and the absence of an optical afterglow and strong soft-X-ray absorption in GRB 970828 (Groot *et al.* 1997; Murakami *et al.* 1997). On the other hand, fits to the observational data on GRB 970508 and GRB 971214 suggest external densities in the range $0.04\text{–}0.4\,{\rm cm}^{-1}$, which would be more typical of a tenuous interstellar medium (Wijers & Galama 2000).

We must nonetheless remain open-minded about other possibilities. For instance, we may be wrong in supposing that the central object becomes dormant after the gamma-ray burst itself. It could be that the accretion-induced collapse of a white dwarf, or (for some equations of state) the merger of two neutron stars, could give rise to a rapidly spinning pulsar, temporarily stabilized by rapid rotation. The afterglow could then, at least in part, be due to a pulsar's continuing power output (cf. Usov 1994). It could also be that mergers of unequal-mass neutron stars, or neutron stars with other compact companions, lead to the delayed formation of a black hole. Such events might also lead to repeating episodes of accretion and orbit separation, or to the eventual explosion of a neutron star which has dropped below the critical mass, all of which would provide a longer time-scale episodic energy output.

And there could be more subclasses of classical GRB than just short ones and long ones. There is for instance the apparent coincidence of GRB 980425 with the SN Ib/Ic 1998bw (Galama *et al*. 1998). Much progress has been made in understanding how gamma-rays can arise in fireballs produced by brief events depositing a large amount of energy in a small volume, and in deriving the generic properties of the long-wavelength afterglows that follow from this. There still remain a number of mysteries, especially concerning the identity of their progenitors, the nature of the triggering mechanism, the transport of the energy and the time-scales involved.

Gamma-ray bursts, even if we do not understand them, may still be useful as powerful beacons for probing the high-redshift ($z > 5$) Universe. Even if their total energy is reduced by beaming to a 'modest' *ca.* 10^{52} erg in photons, they are the most extreme phenomena that we know about in high-energy astrophysics. The modelling of the burst itself—the trigger, the formation of the ultrarelativistic outflow and the radiation processes—is a formidable challenge to theorists and to computational techniques. It is, also, a formidable challenge for observers in their quest for detecting minute details in extremely faint and distant sources. And if the class of models that we have advocated here turns out to be irrelevant, the explanation of gamma-ray bursts will surely turn out to be even more remarkable and fascinating, perhaps implicating magnetic fields even stronger than 10^{15} G.

I am especially grateful to Peter Mészáros and Ralph Wijers for extended collaboration on this subject, and to Josh Bloom and Stan Woosley for discussions. This research has been supported by The Royal Society.

References

Abramowicz, M. A., Karas, V. & Lanza, A. 1998 *Astron. Astrophys.* **331**, 1143.

Blandford, R. D. & Znajek, R. L. 1977 *Mon. Not. R. Astron. Soc.* **179**, 433.

Bloom, J. S., Sigurdsson, S. & Pols, O. 1998 *Mon. Not. R. Astron. Soc.* **305**, 763.

Bloom, J. S. *et al*. 1998 *Astrophys. J. Lett.* **506**, L105.

Briggs, M. S. *et al*. 1997 *Astrophys. J.* **469**, 40.

Costa, E. *et al*. 1997 *Nature* **387**, 783.

Daigne, F. & Mochkovitch, R. 1997 *Mon. Not. R. Astron. Soc.* **285**, L15.

Derishev, E. V., Kocharovsky, V. V. & Kocharovsky, Vl. V. 1999 Preprint.

Fryer, C. & Woosley, S. 1998 *Astrophys. J. Lett.* **502**, L9.

Galama, T. *et al*. 1998 *Nature* **395**, 670.

Gallant, Y. A., Hoshino, M., Langdon, A. B., Arons, J. & Max, C. E. 1992 *Astrophys. J.* **391**, 73.

Garnavich, P. *et al*. 1998 *Astrophys. J.* **493**, L53.

Grindlay, J. 1999 *Astrophys. J.* **510**, 710.

Groot, P. *et al*. 1997 In *Gamma-ray bursts* (ed. C. Meegan, R. Preece & T. Koshut), p. 557. New York: AIP.

Katz, J. 1994 *Astrophys. J.* **422**, 248.

Katz, J. & Canel, L. M. 1996 *Astrophys. J.* **471**, 915.

Katz, J. & Piran, T. 1997 *Astrophys. J.* **490**, 772.

Klebesadal, R. W., Strong, I. B. & Olsen, R. A 1973 *Astrophys. J.* **182**, L85.

Kluźniak , W. & Ruderman, M. 1998 *Astrophys. J.* **508**, L113.

Kouveliotou, C. 1999 *P. Natn. Acad. Sci.* **96**, 5351.

Kouveliotou, C. *et al*. 1993 *Astrophys. J.* **413**, L101.

Kulkarni, S. *et al*. 1998 *Nature* **393**, 35.

Lamb, D. 1995 *Publ. Astron. Soc. Pacific* **107**, 1152.

Mészáros, P. & Rees, M. J. 1992 *Astrophys. J.* **397**, 570.

Mészáros, P. & Rees, M. J. 1993 *Astrophys. J.* **405**, 278.

Mészáros, P. & Rees, M. J. 1997*a* *Astrophys. J.* **476**, 232.

Mészáros, P. & Rees, M. J. 1997*b* *Astrophys. J.* **482**, L29.

Mészáros, P. & Rees, M. J. 1998*a* *Astrophys. J. Lett.* **502**, L105.

Mészáros, P. & Rees, M. J. 1998*b* *Mon. Not. R. Astron. Soc.* **299**, L10.

Metzger, M. R. *et al*. 1997 *Nature* **387**, 878.

Murakami, T. *et al*. 1997 In *Gamma-ray bursts* (ed. C. Meegan, R. Preece & T. Koshut). New York: AIP.

Narayan, R., Paczyński, B. & Piran, T. 1992 *Astrophys. J. Lett.* **395**, L83.

Nishida, S., Lanza, A., Eriguchi, Y. & Abramowicz, M. A. 1996 *Mon. Not. R. Astron. Soc.* **278**, L41.

Paczyński, B. 1991 *Acta. Astron.* **41**, 257.

Paczyński, B. 1995 *Publ. Astron. Soc. Pacific* **107**, 1167.

Paczyński, B. 1998 *Astrophys. J.* **494**, L45.

Paczyński, B. & Xu, G. 1994 *Astrophys. J.* **427**, 708.

Pendleton, G. *et al*. 1996 *Astrophys. J.* **464**, 606.

Perlmutter, S. *et al*. 1998 *Nature* **391**, 51.

Piran, T. 1997 In *Unsolved problems in astrophysics* (ed. J. N. Bahcall & J. P. Ostriker). Princeton University Press.

Podsiadlowski, P., Rees, M. J. & Ruderman, M. 1995 *Mon. Not. R. Astron. Soc.* **273**, 755.

Popham, R., Woosley, S. & Fryer, C. 1999 *Astrophys. J.* **518**, 356.

Rees, M. J. & Mészáros, P. 1992 *Mon. Not. R. Astron. Soc.* **258**, 41P.

Rees, M. J. & Mészáros, P. 1994 *Astrophys. J.* **430**, L93.

Reichart, D. 1997 *Astrophys. J.* **485**, L57.

Reichart, D. 1998 *Astrophys. J.* **495**, L99.

Rhoads, J. E. 1997 *Astrophys. J.* **487**, L1.

Ruderman, M. 1975 *Ann. N.Y. Acad. Sci.* **262**, 164.

Ruffert, M. 1997 *Astron. Astrophys.* **317**, 793.

Ruffert, M. & Janka, H.-T. 1999 *Astron. Astrophys.* **344**, 573.

Ruffert, M., Janka, H.-T., Takahashi, K. & Schaefer, G. 1997 *Astron. Astrophys.* **319**, 122.

Sari, R. & Piran, T. 1995 *Astrophys. J.* **455**, L143.

Tavani, M. 1997 *Astrophys. J.* **483**, L87.

Thompson, C. 1994 *Mon. Not. R. Astron. Soc.* **270**, 480.

Thompson, C. & Duncan, R. C. 1993 *Astrophys. J.* **408**, 194.

Usov, V. V. 1994 *Mon. Not. R. Astron. Soc.* **267**, 1035.

van Paradijs, J. *et al*. 1997 *Nature* **386**, 686.

Vietri, M. 1997 *Astrophys. J.* **478**, L9.

Waxman, E. 1997 *Astrophys. J.* **489**, L33.

Wijers, R. A. M. J. & Galama, T. 2000 *Astrophys. J.* (In the press.)

Wijers, R. A. M. J., Rees, M. J. & Mészáros, P. 1997 *Mon. Not. R. Astron. Soc.* **288**, L51.

Wijers, R. A. M. J., Bloom, J. S., Bagla, J. S. & Natarajan, P. 1998 *Mon. Not. R. Astron. Soc.* **294**, L13.

Woosley, S., Eastman, R. & Schmidt, B. 1999 *Astrophys. J.* **516**, 788.

Discussion

T. G. Forbes (*EOS Institute, The University of New Hampshire, USA*). What about the rapid time variations in the gamma-ray emission during the burst. Are those also due to scintillation or is something else going on?

M. J. Rees. Scintillations in the intervening medium can influence the observed afterglow in the radio band, but not the gamma-rays themselves. The time-structure in the complex bursts is probably caused by intrinsic variations in the 'engine'. For instance, if a relativistic outflow persisted for 10–100 s, but its intensity or Lorentz factor were unsteady, then internal shocks within the outflow could generate sharp peaks in the intensity, on any time-scale down to milliseconds (since the dynamical time-scale close to a collapsed stellar-mass object is of that order).

M. Osmaston (*Address?*). From your description and interpretation, gamma-ray bursts offer a hitherto unavailable combination of source signal brevity and immense distance. Have you considered the possibility that the time-spread variability of the received signal, which you have referred to as being problematic if intrinsic to the source, could actually be detecting, for the first time, tiny fluctuations in the refractive index/transmission time along the path? This might valuably supplement the information already to be had about such paths from the absorption line 'forests' in quasar spectra.

M. J. Rees. The ordinary plasma dispersion effects would be negligible in the gamma-ray band. There have, however, been speculations that a different kind of dispersion associated with quantum gravity could yield fractional time-delays of order the ratio of the photon energy to the Planck energy; such effects might be marginally detectable.

I would add, in response to your comment on quasar spectra, that observations of the optical flashes or afterglows from high-redshift bursts could in principle offer even higher signal-to-noise spectra of the Lyman forest than we already have from quasars (and might, if we were lucky, allow optical astronomers to probe still higher redshifts).

L. P. Grischchuk (*Cardiff University, UK*). From everything you have said it appears that beaming is necessary and you even suggested that Ω_γ is only about 10^{-4}.

M. J. Rees. I know what you're going to say—good news for gravitational waves?

L. Grischchuk. No, it's unfortunately bad news for us both because it would raise the event rate up to about 10^{-2} per galaxy per year. Nobody would be happy with this because the estimated coalescence rates, from binary star evolution models, are about 10^{-5}–10^{-6} per year.

M. J. Rees. I agree that these considerations would raise a problem for ultra-narrow beaming in binary-coalescence models. However, I think it's actually physically implausible that the Poynting outflow in such models would yield very narrow beams, even though some beaming along the rotation axis would seem natural. In contrast, supernova-type models could probably only produce a narrow beam, because it's unlikely that there would be enough energy available to expel all the envelope, except in a rather narrow cone.

Perhaps I could mention a recent claim, that a relatively nearby supernova generated a gamma-ray burst (albeit at a much lower luminosity than 'classic' bursts). If correct, this may be encouraging for the Laser Interferometric Gravitational Wave Observatory: if a substantial subset of ordinary supernovae involved a highly asymmetric core-collapse, then it could be that these would yield a higher rate of detectable gravitational wave events than the (much rarer) coalescing compact binaries.

K. Horne (*University of St Andrews, UK*). When one of these goes off in our galaxy, is there any danger for any forms of life on Earth?

M. J. Rees. We might expect, very roughly, one 'classical' burst to be detected in our own Galaxy every 10 million years. This estimate depends on how the burst rate depends on cosmic epoch, but is independent of how narrowly beamed each object is. The gamma-ray flux would, for a few seconds, be comparable with the optical energy flux from the Sun. But the high-energy photons would of course dump their energy in the upper atmosphere. If there were a burst within a kiloparsec, the optical emission during the flash (lasting up to a minute) could be brighter than the Sun.

D. Lynden-Bell (*Queen's University, Belfast, UK*). This has been a marvellous meeting, and so I would like to thank The Royal Society for bringing us all together and Nicholas Boross-Toby for his efficient organization. The meeting was a great success with wonderful talks and lively discussion. I thank everyone who attended.