The progenitors of short gamma-ray bursts

The progenitors of short gamma-ray bursts

William H Lee[1] and Enrico Ramirez-Ruiz[2, 3]

[1] Instituto de Astronomía, Universidad Nacional Autónoma de México,
Apdo. Postal 70–264, Cd Universitaria, México DF 04510, Mexico
[2] Institute for Advanced Study, School of Natural Sciences,
Princeton, NJ 08540, USA
E-mail: wlee@astroscu.unam.mx and enrico@ias.edu

Abstract. Recent months have witnessed dramatic progress in our understanding of short γ-ray burst (SGRB) sources. There is now general agreement that SGRBs—or at least a substantial subset of them—are capable of producing directed outflows of relativistic matter with a kinetic luminosity exceeding by many millions that of active galactic nuclei. Given the twin requirements of energy and compactness, it is widely believed that SGRB activity is ultimately ascribable to a modest fraction of a solar mass of gas accreting on to a stellar mass black hole (BH) or to a precursor stage whose inevitable end point is a stellar mass BH. Astrophysical scenarios involving the violent birth of a rapidly rotating neutron star, or an accreting BH in a merging compact binary driven by gravitational wave emission are reviewed, along with other possible alternatives (collisions or collapse of compact objects). If a BH lies at the centre of this activity, then the fundamental pathways through which mass, angular momentum and energy can flow around and away from it play a key role in understanding how these prime movers can form collimated, relativistic outflows. Flow patterns near BHs accreting matter in the *hypercritical* regime, where photons are unable to provide cooling, but neutrinos do so efficiently, are discussed in detail, and we believe that they offer the best hope of understanding the *central engine*. On the other hand, statistical investigations of SGRB niches also furnish valuable information on their nature and evolutionary behaviour. The formation of particular kinds of progenitor sources appears to be correlated with environmental effects and cosmic epoch. In addition, there is now compelling

[3] Author to whom any correspondence should be addressed.

evidence for the continuous fuelling of SGRB sources. We suggest here that the observed late flaring activity could be due to a secondary accretion episode induced by the delayed fall back of material dynamically stripped from a compact object during a merger or collision. Some important unresolved questions are identified, along with the types of observation that would discriminate among the various models. Many of the observed properties can be understood as resulting from outflows driven by hyperaccreting BHs and subsequently collimated into a pair of anti-parallel jets. It is likely that most of the radiation we receive is reprocessed by matter quite distant to the BH; SGRB jets, if powered by the hole itself, may therefore be one of the few observable consequences of how flows near nuclear density behave under the influence of strong gravitational fields.

Contents

1. Introduction

1.1. Prologue

In the sections which follow, we shall be concerned predominantly with the theory of short γ-ray bursts(SGRBs)[4]. If the concepts there proposed are indeed relevant to an understanding of the nature of these sources, then their existence becomes inextricably linked to the *metabolic pathways* through which gravity, spin, and energy can combine to form collimated, ultra relativistic outflows. These threads are few and fragile, as we are still wrestling with understanding non-relativistic processes, most notably those associated with the electromagnetic field and gas dynamics. If we are to improve our picture-making we must make more and stronger ties to physical theory. But in reconstructing the creature, we must be guided by our eyes and their extensions. In this introductory chapter, we have therefore attempted to briefly summarize the observed properties of these ultra-energetic phenomena[5]. There are five sections: subsection 1.2 gives a brief account of their history from birth to present-age; subsection 1.3 is devoted to their metabolism – in other words, to their gross energetics, spectra and time variability; subsection 1.4 describes the attributes of the afterglow signals, which, as fading beacons, mark the location of the fiery and brief γ-ray event. These afterglows in turn enable the measurement of redshift distances, the identification of host galaxies at cosmological distances, and provide evidence that many SGRBs are associated with old stellar populations and possibly with no bright supernova. These threads will be woven in subsection 1.5. Finally, subsection 1.6 gives a compendium of the observational *facts*.

1.2. Burst of progress

The manifestations of SGRB activity are extremely diverse. SGRBs are observed throughout the whole electromagnetic spectrum, from GHz radio waves to 10 MeV γ-rays, but until recently, they were known predominantly as bursts of γ-rays, largely devoid of any observable traces at any other wavelengths.

Before 2005, most of what we knew about SGRBs was based on observations from the Burst and Transient Source Experiment (BATSE) on board the *Compton Gamma Ray Observatory*, whose results have been summarized by Fishman and Meegan [2]. BATSE, which measured about 3000 events, detected approximately one burst on a typical day. While they are on, they outshine every other source in the γ-ray sky, including the sun. Although each is unique, the bursts fall into one of two rough categories. Bursts that last less than two seconds are classified as short, and those that last longer—the majority—as long [3]. The two categories differ spectroscopically, with short bursts having relatively more high-energy γ-rays than long bursts do. Figure 1 shows the hardness ratio as a function of the duration of the emission. It is a measure of the slope of the spectrum, where larger values mean that the flux at high energies dominates.

Arguably the most important result from BATSE concerned the spatial distribution of bursts. Both long and short events occur isotropically—that is, evenly over the entire sky with no

---

[4] The literature on this subject has become quite large, and to keep the references to a manageable size, we have in general referred to the most recent comprehensive article in a given topic. In particular, we give only representative references for those topics on which helpful reviews articles already exist. It is thus possible for the reader to trace the early work on which some of these conclusions are based. We apologize to those colleagues whose work is either omitted or not fully represented.

[5] The reader is referred to [1] for an an excellent review of the observations.
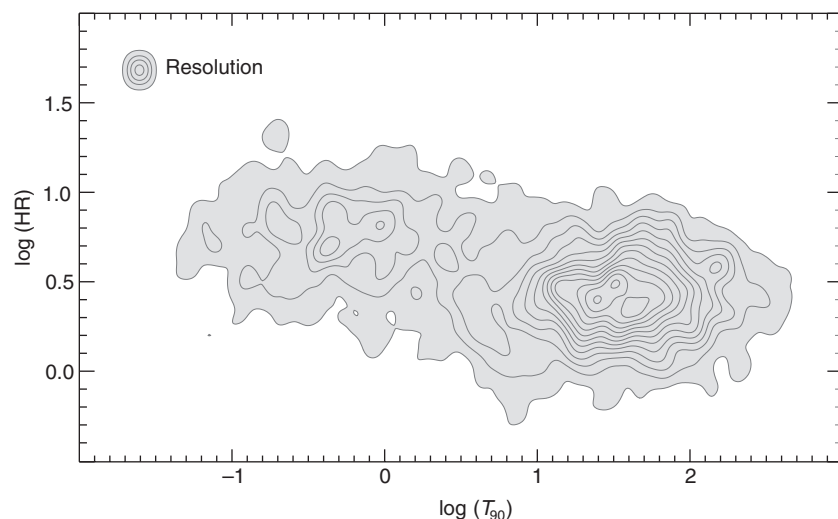
Figure 1. Hardness ratio (i.e., ratio of total counts in the 100–300 keV range, to those in the 50–100 keV range) versus duration for BATSE bursts. The hardness ratio is a measure of the shape of the spectrum: larger values correspond to harder spectra. Different line levels denote number density contours. The bimodality of the distribution of their duration is confirmed by the associated spectral shape.

dipole and quadrupole components, suggesting a cosmological distribution. This finding cast doubt on the prevailing wisdom, which held that bursts came from sources within the Milky Way. The uniform distribution instead led most astronomers to conclude that the instruments were picking up some kind of cosmological event. Unfortunately, γ-rays alone did not provide enough information to settle the question for sure. The detection of radiation from bursts at other wavelengths would turn out to be essential. Visible light, for example, could reveal the galaxies in which the bursts took place, allowing their distances to be measured. Attempts were made to detect these burst counterparts, but they proved fruitless.

Observations of burst counterparts [4] were restricted to the class of long duration bursts[6] until, in 2005, the *Swift* spacecraft succeeded in obtaining high-resolution x-ray images [5, 6] of the fading afterglow of GRB 050509B—so named because it occurred on 9 May 2005. This detection, followed by a number of others at an approximate rate of 10 per year, led to accurate positions, which allowed the detection and follow-up of the afterglows at optical and longer wavelengths [7]–[9]. This paved the way for the measurement of redshift distances, the identification of candidate host galaxies, and the confirmation that they were at cosmological distances [5, 6], [9]–[13]. *Swift* is equipped with γ-ray, x-ray and optical detectors for on-board follow-up, and capable of relaying to the ground arc-second quality burst coordinates within less than a minute from the burst trigger, allowing even mid-size ground-based telescopes to obtain prompt spectra and redshifts.

1.3. Metabolics

SGRBs are brief flashes of radiation at soft and hard γ-ray energies that display a wide variety of time histories. They were first detected at soft γ-ray energies with wide field-of-view instruments,

---

[6] Because *BeppoSAX* is mainly sensitive to bursts longer than about 5–10 s.
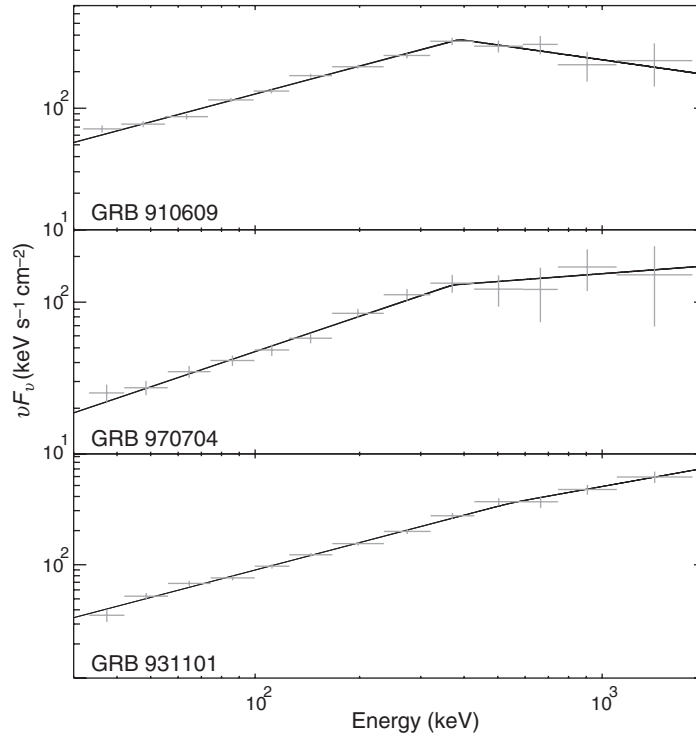
**Figure 2.** Representative spectra $\nu F_\nu \propto \nu^2 N(\nu)$ of various SGRBs. The SGRB spectrum is nonthermal, the number of photons varying typically as $N(\nu) \propto \nu^{-\alpha}$, where $\alpha \sim 1$ at low energies changes to $\alpha \sim 2\text{--}3$ above a photon energy $\sim 1\,\mathrm{MeV}$.

with peak soft γ-ray fluxes reaching hundreds of photons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ in rare cases. The BATSE instrument was sensitive in the 50–300 keV band, and provided the most extensive data base of SGRB observations during the prompt phase.

SGRBs typically show a very hard spectrum in the soft to hard γ-ray regime. The photon index breaks from ≈ -1 at energies $E_{\mathrm{ph}} \leqslant 100$ keV, to a -2 to -3 spectrum at $E_{\mathrm{ph}} \geqslant$ several hundred keV [14]. Consequently, the peak photon energies, E_{pk}, of the time-averaged νF_ν spectra of BATSE SGRBs are typically found in the 500 keV to several MeV range [15]–[17]. The general trend is that the spectrum softens, and E_{pk} decreases, with time. More precise statements must, however, wait for larger area detectors.

In figure 2 representative spectra are plotted, in the conventional coordinates ν and νF_ν, the energy radiated per logarithmic frequency interval. Some obvious points should be emphasized. We measure directly only the specific luminosity $D^2 \nu I_\nu \equiv (1/4\pi)\nu L_\nu$ (the energy radiated in the direction of the earth per second per steradian per logarithmic frequency interval by a source at luminosity distance D), and its dimensionless distance-independent ratios between two frequencies (BATSE triggers, for example, are based on the count rate between 50 and 300 keV). The apparent bolometric luminosity $4\pi \int_{-\infty}^{\infty} D^2 \nu I_\nu \, \mathrm{d}(\mathrm{ln}\,\nu)$ may be quite different from the *true* bolometric luminosity $\int_{4\pi} \int_0^\infty D^2 I_\nu \, \mathrm{d}\nu \, \mathrm{d}\Omega$ if the source is not isotropic. The GRB spectra shown in figure 2 are those of GRB 910609, GRB 970704, and GRB 931101, which were observed by BATSE. The time-integrated spectrum on those detectors ranges from 25 keV to 10 MeV [17].

A *typical* SGRB—if there is such a thing—lasts for a fraction of a second. Observed durations vary, however, by three orders of magnitude, from several milliseconds [2] to
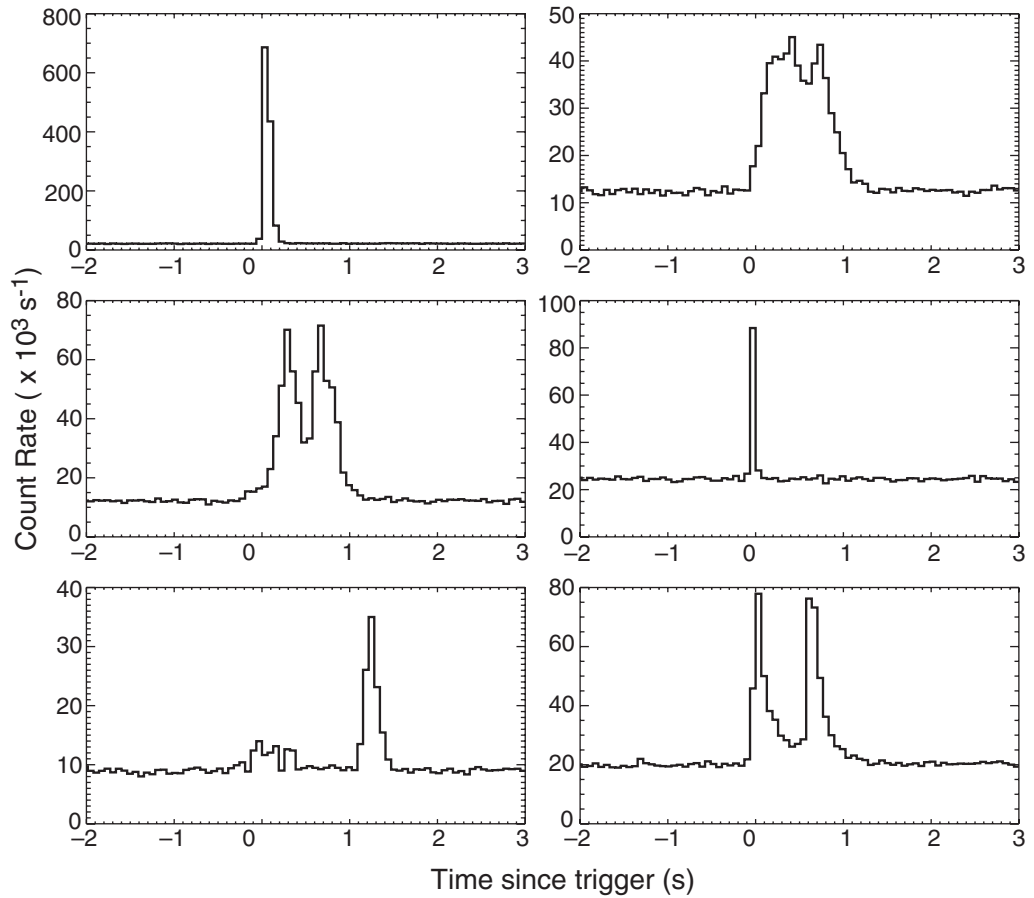
Figure 3. BATSE light curves of various individual bursts. The *y*-axis is the photon count rate in the 0.05–0.5 MeV range; the *x*-axis is the time in seconds since the burst trigger. Both before and after the burst trigger, no γ-rays are detected, above background, from the same direction.

a few seconds [3]. The shortest BATSE burst had a duration of 5 ms with a 0.2 ms structure [18]. Similarly to long bursts, SGRBs have complicated and irregular time profiles which vary drastically from one burst to another [19]. They range from smooth, fast rise and quasi-exponential decay, through curves with several peaks, to variable curves with many peaks. Various profiles, selected from the BATSE catalogue, are shown in figure 3.

The duration of a SGRB is defined by the time during which the middle 50% (t_{50}) or 90% (t_{90}) of the counts above background are measured. A bimodal duration distribution is measured, irrespective of whether the t_{50} or t_{90} durations are considered [3]. About two-thirds of BATSE GRBs are long-duration GRBs with $t_{90} \geqslant 2$ s, with the remainder comprising the SGRBs. The integral size distribution of BATSE SGRBs in terms of peak flux ϕ_p is very flat below ~ 1 ph cm^{-2} s^{-1}, and becomes steeper than the expected power law with index $-3/2$ of a Euclidean distribution of sources, at $\phi_p > 5$ ph cm^{-2} s^{-1} (see figure 12 in [2]). This follows from a cosmological origin of GRB sources, with the decline in the number of faint bursts due to cosmic expansion. Follow-up x-ray observations with *Swift* and *HETE-II* have permitted redshift determinations that firmly establish the distance scale to the sources of SGRBs [5, 6, 9], [11]–[13], [20], [22]–[25].
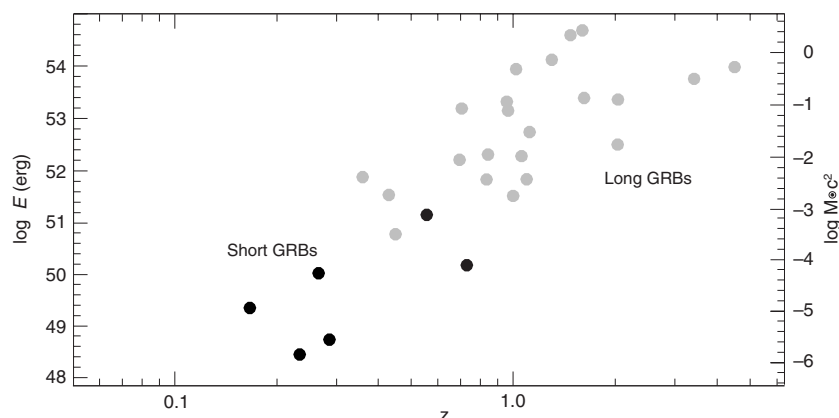
Figure 4. Apparent isotropic γ-ray energy as a function of redshift. The energy is calculated, assuming isotropic emission, for six SGRBs with estimated redshifts. For comparison, the isotropic γ-ray energies for 22 long GRBs are also plotted using the compilation of [21].

The redshifts of about half a dozen SGRBs are now known (late 2006), with the median $\langle z \rangle \sim 0.3$ and the largest *spectroscopically* inferred redshift at $z = 0.71$. It should be noted that the inference of redshift distances relies at present on the statistical connection to a putative host galaxy and spectroscopy of the host, and that none are based on the absorption-line systems seen in the spectra of the afterglows. The corresponding distances imply apparent isotropic γ-ray energy releases $E_{\gamma,\text{iso}} \approx 10^{48}$–$10^{51}(\Omega_\gamma/4\pi)$ ergs, where Ω_γ is the solid angle into which the γ-rays are beamed (figure 4). For a solar-mass object, this implies that an unusually large fraction of the energy is converted into γ-ray photon energy. This spread in the inferred luminosities obtained under the assumption of isotropic emission may be reduced if most GRB outflows are jet-like. A beamed jet would alleviate the energy requirements, and some observational evidence does suggest the presence of a jet (see subsection 1.4).

1.4. A warm afterglow

Among the first SGRBs localized by *Swift* was GRB 050509B. The satellite slewed rapidly in the direction of the burst, enabling its narrow-field x-ray telescope to pinpoint to within 10 arc s a source of faint x-ray emission less than a minute after the trigger. The source decayed with a power-law behaviour, $\phi_X \propto t^\chi$, with $\chi \sim -1.3$ [5, 6]. Despite intense follow-up in the optical and radio, no counterparts were discovered [6, 7, 26]. However, the lack of a detectable afterglow at other wavelengths is not surprising considering, at face value, the existing GRB afterglow theory [27]. It took about two months before the observation of a second short burst, GRB 050709, pinpointed by *HETE-II* [13]. Other detections have followed, at an approximate rate of 10 per year, and permitted the observation and follow-up of afterglows at optical and longer wavelengths. GRB 050709 was the first SGRB from which an optical counterpart was observed [7, 8], and GRB 050724 was the first SGRB for which a radio afterglow was measured [9]. GRB 051221A is probably the best-sampled SGRB to date [22].

Of the few bursts localized by the *Swift* XRT, only four were bright enough to permit detailed study: GRB 050724, GRB 051221A, GRB 051210, and GRB 051227 [28]–[30]. Various
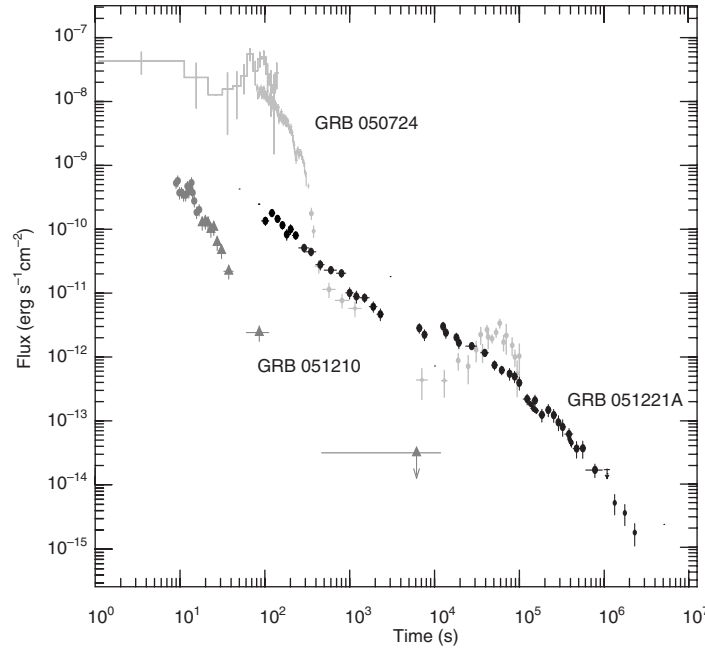
Figure 5. *Swift* lightcurves of various individual bursts: GRB 050724, GRB 051210, and GRB 051221A. The *y*-axis is the flux in the 2–10 keV range; the *x*-axis is the time in seconds since the burst trigger.

x-ray light curves selected from the *Swift* catalogue, are shown in figure 5. *Chandra* follow-up observations obtained for GRB 050724 and GRB 051221A provide the most constraining observations for SGRB jets to date [28, 29]. The lack of an observed early downturn in the x-ray light curves has been interpreted as supporting a mild degree of anisotropy in SGRBs [22, 28, 29, 31, 32]. 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 be masked by the change from ultra relativistic to mildly relativistic flow. It would then be generically difficult to limit the late-time afterglow opening angle in this way if it exceeds 20°. Since some afterglow light curves are unbroken power laws for over 20 days (e.g. GRB 050724), if the energy input were indeed just a simple impulsive shell the opening angle of the late-time afterglow at long wavelengths would probably be greater than 25°, i.e. $\Omega_{opt} \geqslant 0.1$ [28]. However, even this still means that the energy estimates from the afterglow assuming isotropy could be >10 times too high. The beaming angle for the γ-ray emission, Ω_γ, could be smaller, and is much harder to constrain directly.

An observation that attracted much attention was the discovery of long duration ($\sim$100 s) x-ray flares following several SGRBs [11, 13] after a delay of $\sim$30 s (see, e.g. GRB 050724 in figure 5). There is also independent support that x-ray emission on these timescales is detected when lightcurves of many bursts are stacked [33, 34]. These observations may indicate that some sources display continued activity (at a variable level) over a period of minutes [11, 32]. Additional structure in the light curves has also emerged from a continued analysis of some of these objects down to the faintest x-ray flux levels. In some bursts (e.g. GRBs 050724 and 051221A), the x-ray light curve exhibits evidence for a late 'energy refreshment' to the blast

Table 1. Basic properties of the host alaxies of SGRBs (adapted from [39]).

GRB	z	Host classification	In a galaxy cluster	SFR $(M_\odot \, yr^{-1})$	Metallicity $(Z/Z_\odot)$
050509B	0.225	E	Yes	< 0.10	~ 1.00
050709	0.160	Irr/late-type dwarf	No	> 0.30	0.25
050724	0.258	early (E + S0)	No	< 0.05	0.20
050813	0.722?(1.8)	E?	Likely	< 0.20	~ 1.00
051221A	0.5459	late-type dwarf?	?	~ 1.50	~ 1.00
060502B	0.287	E?	?	< 0.40	~ 0.20

wave, on timescales comparable to the afterglow timescale. A reason for the late rise would be, say, if slower moving ejecta catches up with the main shock front injecting a substantial amount of energy and momentum [35, 36]. However, there are other mechanisms that can produce rising afterglow fluxes [37, 38].

1.5. Galactic hosts, supernova family ties and cosmological setting

1.5.1. Demography. Starting with the detection of GRB 050509B, a growing body of evidence has suggested that SGRBs are associated with an older and lower-redshift galactic population than long GRBs and, in a few cases, with large (>10 kpc) projected offsets from the centres of their putative host galaxies (table 1; [39]).

The discovery of GRB 050509B and its fading x-ray afterglow [5] led to the first redshift and host galaxy association [6] for a SGRB. A chance association with such a galaxy was deemed unlikely even under conservative assumptions and stood in stark contrast with the lines-of-sight of long GRBs, for which no association with an early-type host was ever made. Further *Swift* and *HETE-II* detections of SGRBs have continued to support this hypothesis, though SGRBs are not ubiquitously found at large offsets and associated with early-type galaxies. GRB 050724 [9, 12, 23, 40], GRB 050813 [12], and GRB 060502B like 050509B, were found to be in close association with old, red galaxies. GRB 050724 had optical and radio afterglow emission that pinpointed its location to be within its red host, making the association completely unambiguous, though the association of GRB 050813 and GRB 060502B with any single host remains tentative [41]. The absence of observable Hα and [O II] emission constrains the unobscured star formation rates (SFRs) in these galaxies to $< 0.2 M_\odot$ yr^{-1}, and the lack of Balmer absorption lines implies that the last significant star forming event occurred >1 Gyr ago. Based on positions of the afterglows, two bursts (050509B and 050813) are very likely associated with clusters of galaxies [6, 42].

Not all hosts lack active star formation: GRB 050709 [7, 8, 10, 13] and GRB 051221A [22] both had optical afterglows and were associated with galaxies showing evidence for current, albeit low, star formation. The host of GRB 051221A, moreover, exhibits evidence for an evolved stellar population. Despite the availability of both x-ray and optical afterglow locations, no nearby host has successfully been identified for either GRB 060121 or GRB 060313. These observations indicate that these SGRBs occurred during the past ~ 7 Gyr of the universe ($z < 1$) in galaxies with diverse physical characteristics.

In contrast to what is found for SGRBs, all of the confirmed long GRB host galaxies are actively forming stars with integrated, unobscured SFRs $\sim 1 - 10 M_\odot$ yr^{-1} [44]–[46]. These host

galaxies have small stellar masses and bluer colors than present-day spiral galaxies (suggesting a low metallicity; [47]). The ages implied for the long-soft GRBs are estimated to be <0.2 Gyr [44], which is significantly younger than the minimum ages derived for the early-type galaxies found to be associated with SGRBs. The cluster environments of at least two SGRBs contrast strikingly with the observation that no well-localized long-soft GRB has yet been associated with a cluster [48].

On the whole, the hosts of SGRBs, and by extension the progenitors, are not drawn from the same parent population of long GRBs. SGRBs appear to be more diffusely positioned around galaxies, and their associated hosts contain a generally older population of stars.

1.5.2. Soft gamma-ray repeaters in nearby galaxies. It has been noted [49, 50] that the giant flare (GF) observed from the putative galactic magnetar source SGR1806-20 in December 2004 [49, 50] could have looked like a classical SGRB had it occurred much farther away, thus making the tell-tale periodic signal characteristic of the neutron star (NS) rotation in the fading emission undetectable. The two previously recorded GFs of this type, one each from SGR 0520-66 on 5 March 1979 [51] and SGR 1900 + 14 on 27 August 1998 [52], would have been detectable by existing instruments only out to ~ 8 Mpc, and it was therefore not previously thought that they could be the source of SGRBs. The main spike of the 27 December event would have resembled a short, hard GRB if it had occurred within ~ 40 Mpc, a distance scale encompassing the Virgo cluster [50]. However, the paucity of observed giant flares in our own galaxy has so far precluded observationally based determinations of either their luminosity function or their rate.

In the magnetar model, SGRs are isolated neutron stars with teragauss exterior magnetic fields and even stronger fields within [53], making them the most magnetized objects in the Universe. The large-scale reorganization of the magnetic field is thought to produce the observed GFs. The formation rate of magnetars is expected to track that of stars, which is $\sim 0.013\,M_\odot\,\mathrm{Mpc}^{-3}\,\mathrm{yr}^{-1}$ in our galaxy [50]. This suggests that BATSE would have triggered on such events as SGRBs at a rate of $30(\dot{N}_{\mathrm{Gal}}/0.01\,\mathrm{yr}^{-1})\,\mathrm{yr}^{-1}$, compared with the all-sky BATSE rate of about $150\,\mathrm{yr}^{-1}$. Here $\dot{N}_{\mathrm{Gal}}$ is the average rate of GFs in the galaxy similar to the 27 December event. The observed isotropic distribution of short BATSE GRBs on the sky and the lack of excess events from the direction of the Virgo cluster suggests that only a small fraction, $\leqslant 5\%$, of these events can be SGR GFs within 40 Mpc, implying that $\dot{N}_{\mathrm{g}} \leqslant 3 \times 10^{-3}\,\mathrm{yr}^{-1}$ on average for a galaxy like our own [50].

Before *Swift* detected SGRBs, searches for nearby galaxies within narrow Inter Planetary network (IPN) error boxes revealed already that only up to $\simeq 15\%$ of them could be accounted for by magnetars capable of producing GFs [54]. Finally, a search for SGRBs with spectral characteristics similar to that of GFs (thermal spectra with $kT \simeq 100\,\mathrm{keV}$) in the BATSE catalogue has concluded that a small fraction (up to a few percent) of them could have originated from magnetars in nearby galaxies [55].

Magnetars are thought to be formed during core-collapse events inside massive stars, and because of their relatively short lifetimes as observable sources, $\sim 10^4$ yr, would naturally be located in predominantly star-forming galaxies, while essentially none should be seen in ellipticals. One possible distinction of these from the classic SGRB population may well come from radio observations, because their radio afterglows should not be detectable beyond ~ 1 Mpc [56, 57]. The fraction of SGR events among what are now classified as short GRBs may not be dominant, but it should be detectable and can be tested with future *Swift* observations.
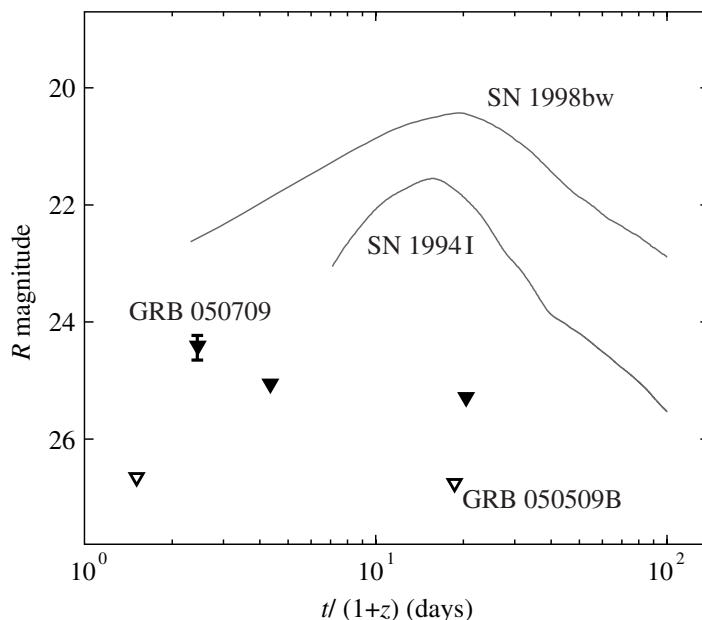
Figure 6. Magnitudes and upper limits of SGRB afterglows, GRB 050709 (filled triangles) and GRB 050509B (empty triangles). For comparison, two Type Ic supernovae are also shown: SN 1998bw was associated with the long GRB 980425.

1.5.3. Supernova partnership. Current observational limits [6, 26, 43] indicate that any supernova(SN)-like event accompanying SGRBs would have to be over 50 times fainter than normal Type Ia SNe or Type Ic hypernovae, five times fainter than the faintest known Ia or Ic SNe, and fainter than the faintest known Type II SNe. These limits strongly constrain progenitor models for SGRBs.

The limits derived for GRB 050709 (filled triangles) and GRB 050509B (empty triangles) are plotted in figure 6 along with two SN light curves as they would appear at $z = 0.225$. The Type Ic SNe plotted are the very energetic Type Ic SN 1998bw associated with the long GRB 980425 [58] and the faint, fast-rise Type Ic SN 1994I [59].

The absence of a SN severely rules out models predicting a normal SN Ia associated with SGRBs. Likewise, current observations disfavour a progenitor uniquely associated with massive stellar collapse. Observations of long GRBs at $z < 0.7$ are consistent with all having SN bumps [60]. The situation is less clear at $z < 0.4$ where all but two long GRBs have had SN features: GRB 980425 [58]; GRB 031203 [61]; GRB 030329 [62]; GRB 011121 [63]; but no SN was detected in GRBs 060505 and 060614 down to a flux limit at least hundreds of times fainter than SN 1998bw [64].

1.5.4. Lifetimes. The distribution of time delays between progenitor formation and explosion is not yet well understood. It can be constrained using, for example, the SGRB rate as a function of redshift [39], [65]–[67]. If SGRB explosions lag star formation by some considerable amount of time, then the intrinsic redshift distribution of SGRBs would be significantly skewed to lower z when compared to the universal SFR. A cursory comparison of the redshift distribution of SGRBs with the SFR reveals what appears to be a significant time delay of a few Gyr. It should

be noted that there are inherent biases in the discovery of a GRB at a given redshift which are often difficult to quantify, such as complex trigger efficiencies and non-detections.

An alternative and perhaps less restrictive approach may be to use the rates of SGRBs in different types of galaxies [68, 69]. On average, early-type galaxies have their stars formed earlier than late-type galaxies, and this difference, together with the time delay between progenitor formation and SGRB outburst, inevitably leads to different burst rates in the two types of galaxies. For instance, the morphological types for the bursts in table 1 reflect a higher incidence of early-type galaxies than type Ia supernovae and suggest progenitor lifetimes significantly exceeding 6 Gyr [69]. One difficulty with this idea is that star formation processes in these two types of galaxies might not be identical. For example, elliptical galaxies can form by the merging of two gas-rich galaxies. Many globular clusters can form in the merging process, which could enhance, for example, the fraction of binary progenitors and also change the lifetime distribution.

Obviously, these estimates are only sketchy and should be taken as an order of magnitude estimate at present. However, it should improve as more hosts are detected. On the other hand, a large progenitor lifetime would help explain the apparent high incidence of galaxy cluster (GC) membership. This can be naturally explained by the fact that galaxies in over dense regions form earlier in hierarchical cosmologies and thus make a substantial contribution to the local stellar mass inventory [69]. Detailed observations of the astrophysics of individual GRB host galaxies may thus be essential before stringent constraints on the lifetime of SGRB progenitors can be placed. If confirmed with further host observations, this tendency of SGRB progenitors to be relatively old can help differentiate between various ways of forming a SGRB.

This concludes our compendium of the *facts*. For ease of reference in the sections that follow, they have been assembled here with a minimum of speculative interpretation.

1.6. Setting the stage

SGRBs have been observed assiduously throughout the electromagnetic spectrum only recently and although we now know much about their collective and individual properties, we are still long way from being sure how they operate.

These SGRB sources involve energies that can exceed 10^{50} ergs, the mass equivalent of 1/10 000 of a sun. Compared with the size of the sun, the seat of this activity is extraordinarily compact, as indicated by rapid variability of the radiation flux on timescales as short as milliseconds. It is unlikely that mass can be converted into energy with better than a few (up to ten) per cent efficiency; therefore, the more powerful SGRB sources must 'process' upwards of $10^{-3} M_\odot$ through a region which is not much larger than the size of a neutron star (NS) or a stellar mass black hole (BH). No other entity can convert mass to energy with such a high efficiency, or within such a small volume.

Well-known arguments connected with opacity, variability timescales and so forth require highly relativistic and variable outflow [70]–[73]. 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^{12}–10^{14} cm [74]–[77]. Because the emitting region must be several powers of ten larger than the compact object that acts as trigger, there is a further physical requirement: the original energy outflowing in a wind would, after expansion, be transformed into bulk kinetic energy. This energy cannot be efficiently radiated as γ-rays unless it is re-randomized. The emitted energy is an observable diagnostic of the microphysical processes of particle acceleration and cooling occurring within the bulk flow [78]–[80]. It is beyond the scope of this article to describe the properties of these

outflows and the physical processes occurring within, and we shall confine our attention to properties directly relevant to their origins. (An excellent account of the structure and energetics of SGRB outflows is given in [1].) This review is therefore not complete and in its emphasis reflects the biases of the authors.

Based on current models, the simplest hypothesis—that the afterglow is due to a relativistic expanding blast wave—seems to agree with the present data [6, 22, 27, 32]. The complex time-structure of some bursts suggests that the central engine may remain active for up to $100\,\mathrm{s}$ [11, 19, 81]. 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 velocity. However we can at present only infer the energy per solid angle; there are reasons to suspect that the afterglow is not too narrowly beamed; on the other hand the constraints on the angle-integrated γ-ray energy are not strong [22, 28, 29, 31, 32].

As regards the trigger, there remain a number of key questions. What is the identity of their progenitors? What is the nature of the triggering mechanism, the transport of the energy and the timescales involved? Does it involve a BH orbited by a dense torus? And, if so, can we decide between the various alternative ways of forming it? The presence of SGRBs in old stellar populations helps rule out a source uniquely associated with recent star formation, while the lack of an accompanying supernova is strong evidence against a core-collapse origin. There is now a stronger motivation to develop models in fuller detail. This article outlines some of these issues.

2. Basic ingredients

In this section, we present a partial summary of some general ways in which gravity, angular momentum and the electromagnetic field can couple to power ultra relativistic outflows, along with a review of the most popular current models for the central source. There are four sections: subsection 2.1 gives a brief account of the arguments in favour of a gravitationally fuelled origin; subsection 2.2 describes the attributes of the most widely favoured and conventional progenitors; the various modes of energy extraction from such systems are then discussed in subsection 2.3; finally, subsection 2.4 gives a compendium of the types of observation that might help discriminate among the various progenitor models.

2.1. General considerations

Shortly after the discovery of quasars in 1963, it was suggested that accretion of gas onto a compact massive body was responsible for their enormous energy output [82, 83], which can be hundreds of times larger than that of entire galaxies. Such a body is essentially a very efficient converter of gravitational binding energy into radiation. The deeper the gas can fall into the potential well before the radiation is converted, the more efficient the process, hence the appealing nature of compact objects. For BHs approximately $\Delta\epsilon \sim GM/R_{\mathrm{ms}} \sim 0.1c^2 \equiv 10^{20}\,\mathrm{erg\,g^{-1}}$ (where R_{ms} is the radius of the marginally stable orbit) can be released, and even more if the hole is endowed with a large angular momentum. This efficiency is over a hundred times that traditionally associated with thermonuclear reactions (hydrogen burning releases $0.007c^2 \sim 6 \times 10^{18}\,\mathrm{erg\,g^{-1}}$). Since the 1960s, we have also learned about stellar–mass BHs, where $M \sim 5 - 10 M_\odot$, in galactic binary systems and ultra-luminous x-ray sources which, with less confidence, we also associate with BHs, primarily on energetic grounds.

In these objects, accretion (and the accompanying radiation) is usually thought to be limited by the self–regulatory balance between Newtonian gravity and radiation pressure. A fiducial luminosity is the *Eddington* limit associated with quasi-spherical accretion, at which radiation pressure balances gravity. If Thomson scattering provides the main opacity and the relevant material is fully ionized hydrogen, then this luminosity is

$$L_{\mathrm{Edd}} = \frac{4\pi G M c m_{\mathrm{p}}}{\sigma_{\mathrm{T}}} = 1.3 \times 10^{38} \left(\frac{M}{M_\odot}\right) \ \mathrm{erg \ s^{-1}}, \tag{1}$$

σ_{T} being the Thomson cross-section. This may be converted to a mass accretion rate if one considers that the accretion luminosity is $L_{\mathrm{Edd}} = L_{\mathrm{acc}} = G M \dot{M}/R$, giving

$$\dot{M} = 10^{18} \left(\frac{R_*}{10 \, \mathrm{km}}\right) \ \mathrm{g \, s^{-1}} = 1.5 \times 10^{-8} \left(\frac{R_*}{10 \, \mathrm{km}}\right) \ M_\odot \, \mathrm{yr^{-1}}, \tag{2}$$

where R_* is the radius of the compact object. Now clearly this applies strictly only in a quasi-spherical configuration and a steady state. Relaxing these assumptions allows for greater luminosities in transient events, or for configurations in which the energy release is somehow collimated. However they will not be greater than the expression given above by more than a factor of a few.

The photon luminosity, for the duration of a typical short burst (a few seconds at most), is thousands of times larger than that of any active galactic nucleus (thought to involve supermassive BHs), and is 12 orders of magnitude above the limit (1). The total energy, however, is not very far off from that of other phenomena encountered in astrophysics, and is in fact reminiscent of that released in the core of a supernova. The Eddington photon limit (1) is circumvented if the main cooling agent is emission of neutrinos rather than electromagnetic waves. The associated interaction cross-section is then many orders of magnitude smaller, and the allowed accretion rates and luminosities are correspondingly higher. For example, using the cross-section for neutrino pair production, the Eddington limit can be rewritten as

$$L_{\mathrm{Edd},\nu} = 8 \times 10^{53} \left(\frac{E_\nu}{50 \, \mathrm{MeV}}\right)^{-2} (M/M_\odot) \ \mathrm{erg \, s^{-1}}, \tag{3}$$

with an associated accretion rate, assuming unit efficiency for conversion of mass into neutrino energy,

$$\dot{M}_{\mathrm{Edd},\nu} = 0.4 (M/M_\odot) \left(\frac{E_\nu}{50 \, \mathrm{MeV}}\right)^{-2} \ M_\odot \, \mathrm{s^{-1}}. \tag{4}$$

The time it would take an object to radiate away its entire rest–mass energy in this way is a mass-independent *Eddington time* given by

$$t_{\mathrm{Edd},\nu} = \frac{M}{\dot{M}_{\mathrm{Edd},\nu}} \sim 2.5 \left(\frac{E_\nu}{50 \, \mathrm{MeV}}\right)^2 \ \mathrm{s}, \tag{5}$$

while the timescale over which an accretion-driven source would double its mass is $\sim(L/L_{\mathrm{Edd},\nu})^{-1} \times (\mathrm{efficiency})^{-1} \times t_{\mathrm{Edd},\nu}$. The dynamical time scales near BHs are modest multiples of R_{g}/c, where R_{g} is the characteristic size of the collapsed object

$$R_{\mathrm{g}} = GM/c^2 \sim 1.5 \times 10^5 (M/M_\odot)\mathrm{cm}, \tag{6}$$

and are therefore much shorter than $t_{\mathrm{Edd},\nu}$. A fiducial Eddington density, characteristic near the horizon when the hole accretes at the Eddington rate, is

$$\rho_{\mathrm{Edd},\nu} = \frac{\dot{M}_{\mathrm{Edd},\nu}}{4\pi R_{\mathrm{g}}^2 c} \sim 10^{11}(M/M_\odot)^{-1}\left(\frac{E_\nu}{50\,\mathrm{MeV}}\right)^{-2} \mathrm{g\,cm^{-3}}. \tag{7}$$

It should be noted that the typical Thomson optical depth under these conditions is

$$\tau_{\mathrm{T}} \sim n_{\mathrm{Edd},\nu}^{1/3} R_{\mathrm{g}} \sim 10^{16} \tag{8}$$

and so, as expected, photons are incapable of escaping and constitute part of the fluid. For completeness, we can also define an Eddington temperature, as the black body temperature if a luminosity $L_{\mathrm{Edd},\nu}$ emerges from a sphere of radius R_{g},

$$T_{\mathrm{Edd},\nu} = \left(\frac{L_{\mathrm{Edd},\nu}}{4\pi R_{\mathrm{g}}^2 \sigma_{\mathrm{SB}}}\right)^{1/4} \sim 5 \times 10^{11}(M/M_\odot)^{-1/4}\left(\frac{E_\nu}{50\,\mathrm{MeV}}\right)^{-1/2} \mathrm{K}, \tag{9}$$

or

$$kT_{\mathrm{Edd},\nu} \sim 45(M/M_\odot)^{-1/4}\left(\frac{E_\nu}{50\,\mathrm{MeV}}\right)^{-1/2} \mathrm{MeV}, \tag{10}$$

and an Eddington magnetic field strength

$$B_{\mathrm{Edd},\nu} = \left(\frac{L_{\mathrm{Edd},\nu}}{R_{\mathrm{g}}^2 c}\right)^{1/2} \sim 3 \times 10^{16}(M/M_\odot)^{11/2}\left(\frac{E_\nu}{50\,\mathrm{MeV}}\right)^{-1} \mathrm{G}. \tag{11}$$

Finally, for comparison, we define T_{th} as the temperature the accreted material would reach if its gravitational potential energy were turned entirely into thermal energy. It is given by

$$T_{\mathrm{th}} = \frac{GMm_{\mathrm{p}}}{3kR_{\mathrm{g}}} \sim 3 \times 10^{12}\,\mathrm{K}. \tag{12}$$

Some authors use the related concept of the virial temperature, $T_{\mathrm{vir}} = T_{\mathrm{th}}/2$, for a system in mechanical and thermal equilibrium. In general, the radiation temperature is expected to be $\leqslant T_{\mathrm{th}}$. In deriving the above estimates we have assumed that the radiating material can be characterized by a single temperature. This may not apply, for example, when a hot corona deforms the neutrino spectrum away from that of a cooler thermal emitter [84].

Although we have considered here the specific case of neutrino pair creation, the estimates vary little when one considers for example coherent scattering of neutrinos by nuclei and/or

free nucleons (except for the energy scaling). Similar overall fiducial numbers also hold for NSs, except that the simple mass scalings obtained here are lost.

It is thus clear from the above estimates that when mass accretes onto a (stellar-mass) BH or NS under these conditions, the densities and temperatures are so large ($\rho \simeq 10^{11}$ g cm^{-3}, $T \simeq 10^{11}$ K) that: (i) photons are completely trapped; and (ii) neutrinos, being copiously emitted, are the main source of cooling since they can mostly escape. This regime, which requires correspondingly large accretion rates, is termed hypercritical accretion, and was considered for SN1987A shortly after the explosion [85, 86]. Such high accretion rates are never reached for BHs in x-ray binaries (XRBs) or Active Galactic Nuclei (AGN), where the luminosity remains well below the photon Eddington rate (1). They can, however, be achieved in the process of forming NSs and stellar–mass BHs during the collapse of massive stellar cores. Note that at sufficiently high accretion rates, the density reaches the threshold for optical thickness even to neutrinos, which cannot then simply stream out (this occurs at $\rho \simeq 10^{11}$ g cm^{-3}).

2.2. Bestiary

The current view is that SGRBs arise in a very small fraction (approximately $\sim 10^{-6}$) of stars which undergo a catastrophic energy release event toward the end of their evolution. One conventional possibility is the coalescence of binary NSs [71, 72], [87]–[94]. Double NS binaries, such as the famous PSR1913 + 16 [95], will eventually coalesce due to angular momentum and energy losses to gravitational radiation. The resulting system could be top–heavy and unable to survive as a single NS. However, a BH would be unable to swallow the large amount of angular momentum present. The expected outcome would then be a spinning hole, orbited by a torus of NS debris.

Other types of progenitor have been suggested—e.g. a NS-BH merger [96]–[98] where the NS is tidally disrupted before being swallowed by the hole; the merger of a White Dwarf (WD) with a BH [99]; the coalescence of binary WDs [71, 87, 100, 101]; or accretion induced collapse (AIC) of a NS [102, 103], where the collapsing NS has too much angular momentum to collapse quietly into a BH. In an alternative class of models, it is supposed that the compact objects are contained within a GC, and that the binary system will evolve mainly through hardening of the binary through three-body interactions [104, 105] or physical star–star collisions [106] rather than by pure gravitational wave emission. Finally, violent reconfigurations of the magnetic field on magnetars [53] may offer another possibility. Table 2 provides a summary of the various rate estimates for some of these possible SGRB progenitors. Aside from the rate of SNe events, the rate of SGRBs and plausible progenitors in table 2 are highly imprecise. The main sources of uncertainty are related to supernova kicks, the mass ratio distribution in binaries, mass limits for BH formation, stellar radii and common envelope evolution [107]–[121].

The rate of gravitational mergers among BH–BH and BH–NS binaries depends on the orbital periods after spiral-in, and hence on the ill-understood details of the process. A binary with total mass $10M_1M_\odot$, mass ratio q, and period P_d (in days) in a nearly circular orbit will merge in

$$\tau_m = 1 \times 10^9 P_d^{8/3} M_1^{-5/3} (1 + q)(1 + 1/q) \text{ yr}. \tag{13}$$

Thus for a coalescence to occur within a Hubble time, spiral–in must have reduced the orbital period to $\leqslant 1$ day. Since a BH can have a mass exceeding that of its companion helium core, spiral-in need not occur, as opposed to the case of a lower mass object like a NS (except in very

Table 2. Estimated progenitors of SGRBs and their plausible rates [66, 101, 107, 111, 116, 122] (in $yr^{-1} Gpc^{-3}$). All of the rates listed scale roughly with the rate of star formation; therefore, at redshift of $z = 1$ they are a factor of ~ 10 higher than locally. Aside from that of SNe events, they are highly uncertain at present and should be taken as an order of magnitude estimate only.

| Progenitor	Rate ($z = 0$)
NS–NS	1–800
BH–NS	0.1–1000
BH–WD	0.01–100
NS AIC	0.1–100
WD–WD	3000
SN Ib/c	60000
SN Ia	150000
SGRBs	$10(4\pi/\Omega)$

non-conservative mass transfer). If the companion has a strong wind, spiral-in might never begin. If it does, envelope ejection might be less efficient than that by a NS, since BH accretion could have a very low efficiency. If spiral-in does occur, however, the subsequent evolution may be simpler than in the double NS case. Helium stars less massive than $3M_\odot$ expand their envelopes dramatically during core carbon burning [123], requiring a second spiral-in to make a NS–NS binary which could merge in a Hubble time. The more massive He cores which could leave BH remnants do not expand much, so a second spiral-in is not needed.

Thus BH–NS binaries form at rates comparable to the NS–NS rate, but the fraction which merge depends on the innards of mass transfer and spiral-in which determine the final period distribution. Gravitational waveforms of any merging systems will allow the masses of the merging bodies to be accurately determined [124, 125], shedding light on the underlying physics, as well as on the minimum BH mass determined by post-collapse infall. The much more certain NS–NS mergers will also allow the mass distribution of NSs to be determined.

How might such a progenitor generate a relativistic outflow or a sudden release of electromagnetic energy? We now address various possible routes.

2.3. Metabolic pathways

It has become increasingly apparent in the last few years that most plausible SGRB progenitors suggested so far (e.g. NS–NS or NS–BH, WD–BH mergers or physical collisions, and NS AIC) are expected to lead to a BH plus debris torus system (figure 7). In this case there is no external agent feeding the accretion disk, and thus the event is over roughly on an accretion time scale (which would be on the order of one second). A possible exception includes the formation from accretion [126] or from a WD–WD [101] or NS–NS merger [127]–[129] of a rapidly rotating (in some cases very massive) NS with an ultrahigh magnetic field (figure 7). If there is an ordered field B, and a characteristic angular velocity ω, for a spinning compact source of radius R_*, then the magnetic dipole moment is $\sim BR_*^3$. General arguments suggest [130, 130] that the non thermal magnetic-dipole-like luminosity will be $\propto B^2 R_*^6 \omega^4/c^3$, and simple scaling from these familiar results of pulsar theory require fields of order 10^{15} G to carry away the rotational or gravitational energy (which is $\sim 10^{53}$ erg) in a time scale of seconds [132, 133].
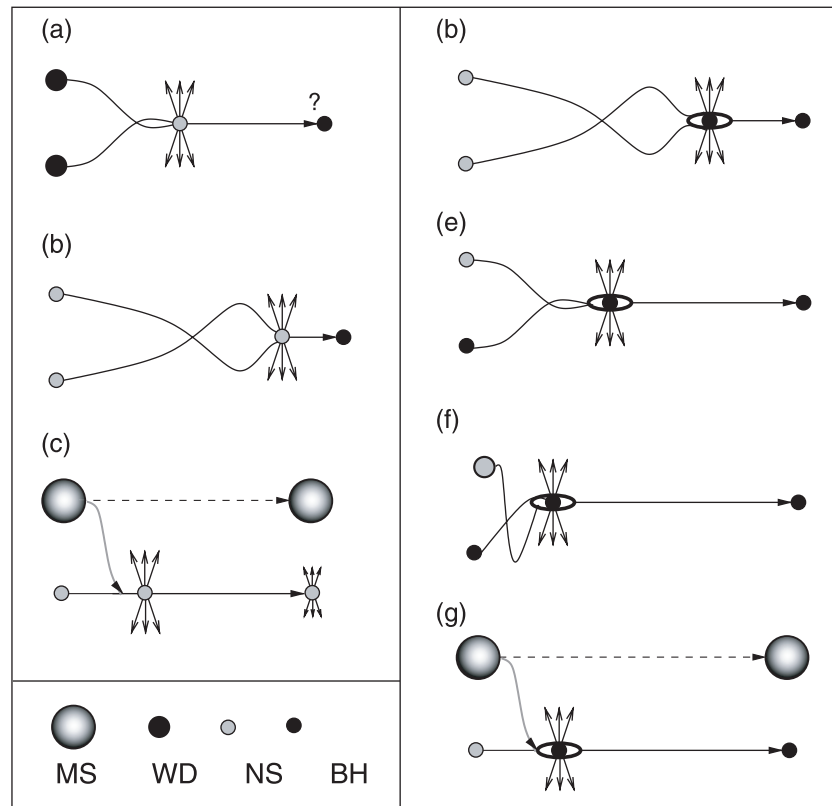
Figure 7. Schematic scenarios for plausible SGRB progenitors: NS mergers (b, d); NS (e) or WD (f) disruption by a BH; WD mergers (a); AIC of NS (g); and a recycled magnetar (c). The dominant production channel for each scenario is depicted, where MS denotes the primary main sequence star. The (rough) relative in-spiral times due to gravitational radiation for compact mergers are shown (not to scale). (d, e, f, g) eventually lead to the formation of a BH with a debris torus around it. Physical star–star collisions or three-body gravitational interactions rather than gravitational mergers could lead to similar conditions. A second way that the progenitor system may liberate its binding and rotational energy is to develop a large magnetic field and act as a magnetar (a, b, c).

One of the most perceptive theoretical discoveries that was made about BHs was that, when they spin, a fraction of their mass can be ascribed to rotational energy and is, in principle, extractable [134]. This is most convincingly demonstrated by observing that there exist orbits of test particles with negative total energy, (including their rest mass), within the ergosphere. If an infalling plasma cloud is attached to magnetic field lines anchored at a large distance (e.g. in an accreting torus), and the field drags the cloud backwards relative to the rotation of the hole placing it on an orbit with negative energy, the work that has been performed can be thought of as energy that has been effectively extracted from the spin of the BH [135].

The binding energy of the orbiting debris, and the spin energy of the BH are thus the two main reservoirs for the case of a BH central engine: up to 42% of the rest mass energy of the torus, and 29% of the rest–mass energy of the BH itself can be extracted for a maximal BH spin. SGRB activity can be powered by the BH only as long as there is interaction with the surrounding

gas, which will probably need to be centrifugally supported (although even for relatively low angular momentum a considerable amount of energy can still be extracted [136]).

The angular momentum is quite generally a crucial parameter, in many ways determining the geometry of the accretion flow. Even a little rotation can make a big difference, breaking the spherical symmetry and producing accretion *disks* instead of *radial* inflow, as envisaged originally by Bondi [137]. If the gas has no angular momentum and the magnetic field is dynamically unimportant, there will be essentially radial inflow. Spherical accretion onto BHs is relatively inefficient despite the deep potential well, because the gas is compressed, but not shocked, and thus cannot easily convert gravitational to thermal energy. The flow pattern changes dramatically if the inflowing gas has a small amount of angular momentum. The quasi-spherical approximation breaks down when the gas reaches a radius $R_{circ} \sim l^2/GM$, where l is the angular momentum per unit mass, and if injection occurs more or less isotropically at large radii, a familiar accretion disk will form (as occurs in x-ray binaries and WD binary systems). The matter will instead dissipate its motion perpendicular to the plane of symmetry and form a differentially rotating disk, the rotational velocity at each point being approximately Keplerian, and then gradually spiral inwards as viscosity transports its angular momentum outwards.

If the emission process is very efficient, the disk is dynamically cold and geometrically thin, in the sense that locally, $kT \ll GMm_p/R$ and the pressure scale height $H(R) \ll R$. If gas passing through a thin disk reaches a radius within which the internal pressure builds up—either because it is unable to cool in an inflow time or because the radiation pressure force is competitive with gravity—the disk will become geometrically thick, with $H \sim R$. In a thick disk or torus the pressure provides substantial support in the radial as well as the vertical direction, and the angular momentum distribution (now as a function of height as well as radius) may be far from Keplerian. Both types of configurations have been studied extensively and the agreement in some cases between observation and theory is extremely good, assuming a fundamentally empirical recipe for angular momentum transport [138]–[141].

As mentioned above in subsection 2.1, in principle flow onto a compact object can liberate gravitational potential energy at a rate approaching a few tenths of $\dot{M}c^2$, where $\dot{M}$ is the mass inflow rate. Even for such high efficiencies the mass requirements of the more luminous SGRB sources are rather high, with

$$\dot{M} \sim 3 \times 10^{-3} \left(\frac{0.1}{\epsilon} \right) \left(\frac{L_{SGRB}}{10^{51} \mathrm{erg\ s^{-1}}} \right) M_\odot\ \mathrm{s^{-1}}, \tag{14}$$

where ϵ is the overall efficiency. The inner regions of disks with mass fluxes in this range are generally able to cool by neutrinos on time scales shorter than the inflow time. If $\dot{m} = \dot{M}/\dot{M}_{Edd,\nu} \leqslant 1$, then the bulk of the neutrino radiation comes from a region only a few gravitational radii in size, and the physical conditions can be scaled in terms of the *Eddington quantities* defined in subsection 2.1. The remaining relevant parameter, related to the angular momentum, is $v_{inflow}/v_{freefall}$, where $v_{freefall} \simeq (2GM/R)^{1/2}$ is the free fall velocity. The inward drift speed v_{inflow} would be of order $v_{freefall}$ for supersonic radial accretion. When angular momentum is important, this ratio depends on the mechanism for its transport through the disk, which is related to the effective shear viscosity. For a thin disk, the factor $(v_{inflow}/v_{freefall})$ is of order $\alpha(H/R)^2$, where H is the scale height at radius R and α is the phenomenological viscosity parameter.

Suppose that a given accretion rate yields a luminosity L_ν with an efficiency 0.1. Then the characteristic density, at a distance R from the hole, with account of the effects of rotation, is

$$\rho \sim \dot{m}(R/R_g)^{-3/2}(v_{\text{inflow}}/v_{\text{freefall}})\rho_{\text{Edd},\nu}, \tag{15}$$

and the maximum magnetic field, corresponding to equipartition with the bulk kinetic energy, would be

$$B_{\text{eq}} \sim \dot{m}^{1/2}(R/R_g)^{-5/4}(v_{\text{inflow}}/v_{\text{freefall}})^{1/2}B_{\text{Edd},\nu}. \tag{16}$$

Any neutrinos emerging directly from the central *core* would have energies of a few MeV. Note that, as mentioned above, $kT_{\text{Edd},\nu}$ is far below the *virial* temperature $kT_{\text{vir}} \simeq m_p c^2 (R/R_g)$.

The flow pattern when accretion occurs would be then determined by the value of the parameters $L_\nu/L_{\text{Edd},\nu}$, which determine the importance of radiation pressure and gravity, and the ratio $t_{\text{cool}}/t_{\text{dynamical}}$, which fixes the temperature if a stationary flow pattern is set up. The preceding general discussion of neutrino–cooled accretion flows thus provides a basis for general models. So far we have discussed the power output that might be generated by the accretion process, but we have made no attempt to describe in detail the flow of the accreting gas. A hint that its dynamics may not be straightforward is provided by the existence of the Eddington limit for neutrinos, clearly illustrating that for the high accretion rates expected in SGRBs, forces other than gravity may be important. In addition, in many cases (and probably in most) the accreting matter possesses considerable angular momentum per unit mass which it has to lose somehow in order to be accreted at all. The reader is referred to section 4 for fuller details. We now concentrate briefly on the origins of the jets which provide the power for SGRB sources.

Two ingredients are necessary for the production of jets: first, there must be a source of material with sufficient free energy to escape the gravitational field of the compact object; second, there must be a way of imparting some directionality to the escaping flow. Our eventual aim must be to understand the overall flow pattern around a central compact object, involving accretion, rotation, and directional outflow but we are still far from achieving this. Most current works who have discussed outflow and collimation have simply invoked some central supply of energy and material. A self-consistent model incorporating outflow and inflow must explain why some fraction of the matter can acquire a disproportionate share of energy (i.e. a high enthalpy). A brief summary of the various metabolic pathways is presented in figure 8.

The neutrino luminosity emitted when disk material accretes via viscous (or magnetic) torques on a time scale $\Delta t \sim 1$ s is roughly

$$L_\nu \sim 2 \times 10^{52} \left(\frac{M_{\text{disk}}}{0.1 M_\odot}\right) \left(\frac{\Delta t}{1\text{s}}\right)^{-1} \text{ erg s}^{-1} \tag{17}$$

for a canonical radiation efficiency of 0.1. One fairly direct solution is to reconvert some of this energy via collisions outside the disk into electron-positron pairs or photons [92, 143]. If this occurs in a region of low baryon density (e.g. along the rotation axis, away from the equatorial plane of the disk) a relativistic pair-dominated wind can be produced. An obvious requirement for this mechanism to be efficient is that the neutrinos escape (free streaming, or diffusing out if the density is high enough) in a time scale shorter than that for advection into the BH. The efficiency for conversion into pairs (scaling with the square of the neutrino density) is too low if the neutrino production is too gradual, so this can become a delicate balancing act. Typical
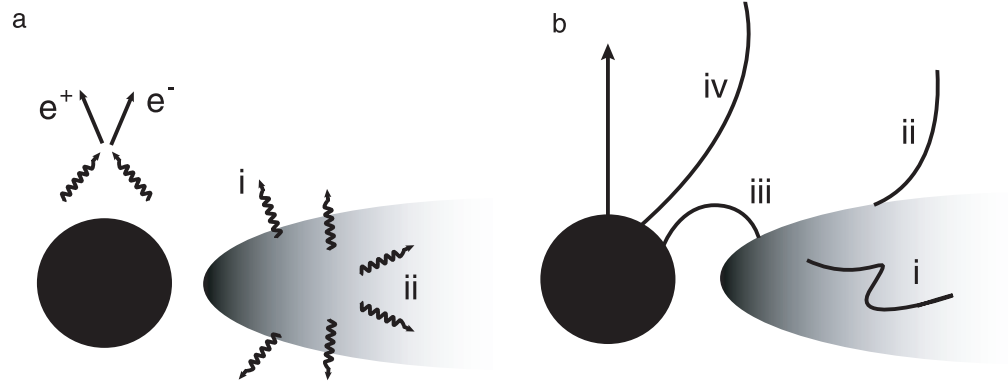
Figure 8. Metabolic pathways for energy extraction. Panel (a): energy released as neutrinos is reconverted via collisions outside the dense core into $e^{\pm}$ pairs or photons (i). The neutrinos that are emitted from the inner regions of the debris deposit part of their energy in the outer parts of the disk (ii), driving a strong baryonic outflow. This wind may be responsible for collimating the jet. Panel (b): strong magnetic fields anchored in the dense matter can convert the binding and/or spin energy into a Poynting outflow. A dynamo process of some kind is widely believed to be able to operate in accretion disks, and simple physical considerations suggest that fields generated in this way would have a canonical length-scale of the order of the disk thickness (i). Open field lines can connect the disk outflow and may drive a hydromagnetic wind (ii). The above mechanism can tap the binding energy of the debris torus, while a rapidly rotating hole could contain an even larger energy reservoir, extractable in principle through magnetohydrodynamic (MHD) coupling to the exterior (iii; iv). Panel (b) adapted from [142].

estimates suggest a lower bound of $L_{\nu\bar{\nu}} \sim 10^{-3} L_{\nu}$ when the entire surface area emits close to a single temperature black-body. The efficiency may be significantly larger if dissipation takes place in a corona-like environment [84].

One attractive energy extraction mechanism that could circumvent the above restriction in efficiency is a relativistic MHD wind [133, 144]. Such a wind carries both bulk kinetic energy and ordered Poynting flux, and it is possible that gamma-ray production occurs mainly at large distances from the source [53, 132, 133], [145]–[147]. A rapidly rotating NS (or accretion disk) releases energy via magnetic torques at a rate

$$L_{\text{em}} \sim 10^{49} \left(\frac{B}{10^{15}\,\text{G}} \right)^2 \left(\frac{P}{10^{-3}\,\text{s}} \right)^{-4} \left(\frac{R}{10^6\,\text{cm}} \right)^6 \text{erg s}^{-1}, \tag{18}$$

where P is the spin period, and B is the strength of the poloidal field at a radius R. The last stable orbit for a Schwarzschild hole lies at a coordinate distance $R = 6R_{\text{g}} = 9(M/M_{\odot})$ km, to be compared with $R_{\text{g}} = 3/2(M/M_{\odot})$ km for an extremal Kerr hole. Thus the massive neutron disk surrounding a Schwarzschild BH of approximately $2\,M_{\odot}$ should emit a spin-down luminosity comparable to that of a millisecond NS. A similar MHD outflow would result if angular momentum were extracted from a central Kerr hole via electromagnetic torques [148]. The

field required to produce $L_{em} \geqslant 10^{51}\,\mathrm{erg\,s^{-1}}$ is enormous, and may be provided by a helical dynamo operating in hot, convective nuclear matter with a millisecond period [53]. A dipole field of the order of 10^{15} G appears weak compared to the strongest field that can in principle be generated by differential rotation ($\sim 10^{17}(P/1\,\mathrm{ms})^{-1}$ G), or by convection ($\sim 10^{16}$ G), although how this may come about in detail is not resolved. Note, however, that it only takes a residual torus (or even a cold disk) of $10^{-3}\,M_\odot$ to confine a field of 10^{15} G. Orbiting debris with such large magnetic seed fields and turbulent fluid motions will give rise to a plethora of electromagnetic activity (figure 8). The topic lies beyond the scope of this paper and we refer the reader to the excellent review by Blandford [142].

A potential death-trap for such relativistic outflows is the amount of entrained baryonic mass from the surrounding medium. For instance, a Poynting flux of 10^{52} erg could not accelerate an outflow to $\Gamma \geqslant 100$ if it had to drag more than $\sim 10^{-5}\,M_\odot$ of baryons with it. A related complication renders the production of relativistic jets even more challenging, because the high neutrino fluxes are capable of ablating baryonic material from the surface of the disk at a rate [149]

$$\dot{M}_\eta \sim 5 \times 10^{-4} \left(\frac{L_\nu}{10^{52}\,\mathrm{erg\,s^{-1}}} \right)^{5/3} M_\odot\,\mathrm{s^{-1}}. \tag{19}$$

Thus a rest mass flux $\dot{M}_\eta$ limits the bulk Lorentz factor of the wind to

$$\Gamma_\eta = \frac{L_{\mathrm{wind}}}{\dot{M}_\eta c^2} = 10 \left(\frac{L_{\mathrm{wind}}}{10^{52}\,\mathrm{erg\,s^{-1}}} \right) \left(\frac{\dot{M}_\eta}{5 \times 10^{-4}\,M_\odot\,\mathrm{s^{-1}}} \right)^{-1}. \tag{20}$$

Assuming that the external poloidal field strength is limited by the vigour of the convective motions, the spin-down luminosity scales with neutrino flux as $L_{\mathrm{wind}} \approx L_{em} \propto B^2 \propto v_{\mathrm{con}}^2 \propto L_\nu^{2/3}$, where v_{con} is the convective velocity. The ablation rate given in equation (19) then indicates that the limiting bulk Lorentz factor Γ_η of the wind decreases as L_ν^{-1}. Thus the burst luminosity emitted by a magnetized neutrino cooled disk may be self-limiting. Mass loss could, however, be suppressed if the relativistic wind were somehow collimated into a jet. This suggests that centrifugally driven mass loss will be heaviest in the outer parts of the disk, and that a detectable burst may be emitted only within a relatively small solid angle centred on the rotation axis.

2.4. Observational tests

In the preceding sections, we have endeavoured to outline some of the basic physical processes that are believed to be of most relevance to interpreting SGRB sources, not so much because we believe that there is strong evidence in favour of them, but instead because it provides a framework in which to discuss the observations and to demonstrate that, as yet, SGRBs pose no threat to conventional physics. Of course, we are conscious that most observations tell us less about the primary source than about the secondary reprocessing of this power in the circumburst and galactic environment—from which we can learn about the central engine only by a chain of uncertain inferences. Unless SGRBs are eventually found to be accompanied by telltale emission features like the supernovae of long-duration GRBs, the only definitive understanding of the progenitors will come from possible associations to direct gravitational or neutrino signals.

Spectral investigations will nonetheless be important as probes of the basic energetics and microphysical parameters of relativistic shocks [27, 32]. We are more sanguine about the

possibility of deriving accurate particle densities from afterglow observations. However, there is the possibility that occurrence in an unusually low-density environment could cause the external shock to occur at much larger radii and over a much longer time scale than in usual afterglows [150, 151]. The x-ray afterglow intensity could be then below the threshold for triggering. This may be the case for SGRBs arising from compact binaries that are ejected from the host galaxy into an external environment that is much less dense that the ISM assumed for usual models. Another possibility for an unusually low density environment, made up only of very high energy but extremely low density electrons, is if the GRB goes off inside a pulsar cavity inflated by one of the NSs in the precursor binary [152, 153]. Such bubbles can be as large as fractions of a parsec or more, giving rise to a deceleration shock months after the SGRB with a consequently much lower brightness that could avoid triggering and detection. The difference between the low-density and high-density environment cases could be tested if future observations of afterglows reveal a correlation with the degree of galaxy clustering or with individual galaxies.

The association of SGRBs with both star-forming galaxies and with ellipticals dominated by old stellar populations, to some researchers, suggested an analogy to type Ia supernovae, as it indicated a class of progenitors with a wide distribution of delay times between formation and explosion. Similarly, just as core-collapse supernovae are discovered almost exclusively in late-time star-forming galaxies, so too are long GRBs. Indeed, a detailed census of the types of host galaxies, burst locations and redshifts should help decide between the various alternatives suggested in subsection 2.2. This is because if the progenitor lifetime is long and the systemic kick is small, then the bursts should correspond spatially to the oldest populations in a given host galaxy. For early-type galaxies, the distribution would presumably follow the light of the host [39]. In contrast, a NS binary could take millions of years to spiral together, and could by then, especially if given a substantial kick velocity on formation, have moved many kiloparsecs from its point of origin. The burst offsets would then presumably be larger for smaller mass hosts. As new redshifts, offsets, and host galaxies of SGRBs are gathered, the theories of the progenitors will undoubtedly be honed.

3. Compact object mergers and accretion disk assembly

SGRB activity manifests itself on many different scales. The simplest hypothesis, however, is that the central *prime mover* or *trigger* is qualitatively similar in all events. The primary energy may be reprocessed in different ways, depending on the details of the circumburst (and galactic) environment. A brief summary of the various evolutionary pathways that may be involved in producing a SGRB is given in subsection 2.2. The important message is that most current alternatives involve either an accreting stellar mass BH (NS) or a precursor stage whose inevitable end point is such an object. Furthermore, as we discussed in subsection 2.1, a BH (NS) embedded in infalling matter offers a more efficient power source than any other conceivable progenitor; it is our firm prejudice that SGRBs are energized by mechanisms that involve either NSs or BH. For this reason, in this section, we turn out attention to the formation of the *trigger*: the actual astrophysical system that is capable of, and presumably powers the observed burst. It is beyond the scope of this article to describe completely all evolutionary pathways, and we shall confine our attention to binary mergers and physical collisions. The study of compact binaries, particularly those involving pulsars, has evolved dramatically over the past thirty years, and carries a great deal of historical baggage. Consequently, most theoretical work has been directed

towards describing these binary encounters. Our discussion will reflect this bias[7]. There are eight sections: subsection 3.1 gives a summary of observed compact object masses; subsection 3.2 gives an overall picture of the merger scenario as we best understand it now; then in subsection 3.3 we describe early considerations of these events; this is followed in subsection 3.4 by a summary of dynamical stability considerations in such systems; the merger itself is considered in subsection 3.5; the effects of general relativity (GR) and the fate of the remnant core are addressed in subsection 3.6 and subsection 3.7 respectively; finally, the outcome of compact object collisions is studied in subsection 3.8.

3.1. The mass of the progenitor

The mass of the compact object dictates a characteristic length and luminosity scale for SGRB activity, so we must consider what we know (and what we do not) from observations of systems containing compact objects in a more leisurely state of affairs.

Measuring stellar masses, although a considerable observational challenge, is not nearly as difficult as measuring their radii (which is why the equation of state at nuclear densities remains one of the primary open questions in compact object and nuclear physics studies). When they occur in binaries, accurate mass determinations are possible (figure 9), and for cases where there are two compact objects in tight orbits, the constraints can be truly spectacular (PSRB1913+16 [157], PSRB1534+12 [158], and PSRJ0737-3039A,B [159, 160]). NS masses are tightly clustered around the Chandrasekhar mass, at 1.4 $M_\odot$. Those that deviate farthest from this value towards high masses are found in low mass x-ray binaries [161], where the slow but steady mass transfer from a companion has produced top–heavy objects (the most dramatic example being J0751+1807 at 2.1 $M_\odot$ [162]). At such an extreme, it is possible that conversion to strange (quark) matter [163]–[166] may have occurred (see Jaikumar *et al* for a review [167]). Incredibly, it is possible that observations of superbursts (unusually long Type I x-ray bursts) may be able to discern if this is actually the case [168, 169].

Regarding BHs, the scatter is considerably greater, probably reflecting the fact that a threshold mass is necessary to create one in a core–collapse event, but no obvious hard limit exists at high masses, and a degree of variation is expected [170]. The constraints are also less stringent in most cases, due to a combination of effects, but nevertheless a boundary can be drawn at roughly 3 $M_\odot$, separating the most massive NS from the least massive BH. This is also close to the maximum mass allowed for a NS in hydrostatic equilibrium in GR, based only on the condition that the equation of state must preserve causality [171]. One would thus expect that encounters between objects of this type would have mass ratios of about unity if two NSs are involved, and 0.2–0.3 if one of each interacts.

Although WDs are not usually considered as direct progenitors of GRBs, one should keep an open mind about their possible role, particularly in the light of unexpected discoveries made by *Swift* concerning late time variability and the inferences about the parent population of SGRBs. They may be responsible for the violent birth of NSs through AIC or the merger of two dwarfs in a binary [172]–[178] and it is possible that this remnant could power a GRB [71, 87, 100, 101]. The mass of a WD is somewhat lower than that of a NS, but they are larger, by a factor of at least $f \sim 10^3$, leading to a natural (dynamical) time scale that is greater than for NSs and BHs by a factor $f^{3/2} \sim 3 \times 10^4$.

---

[7] The reader is referred to the excellent review by Rosswog [154] for an alternative but complementary description of these source models.
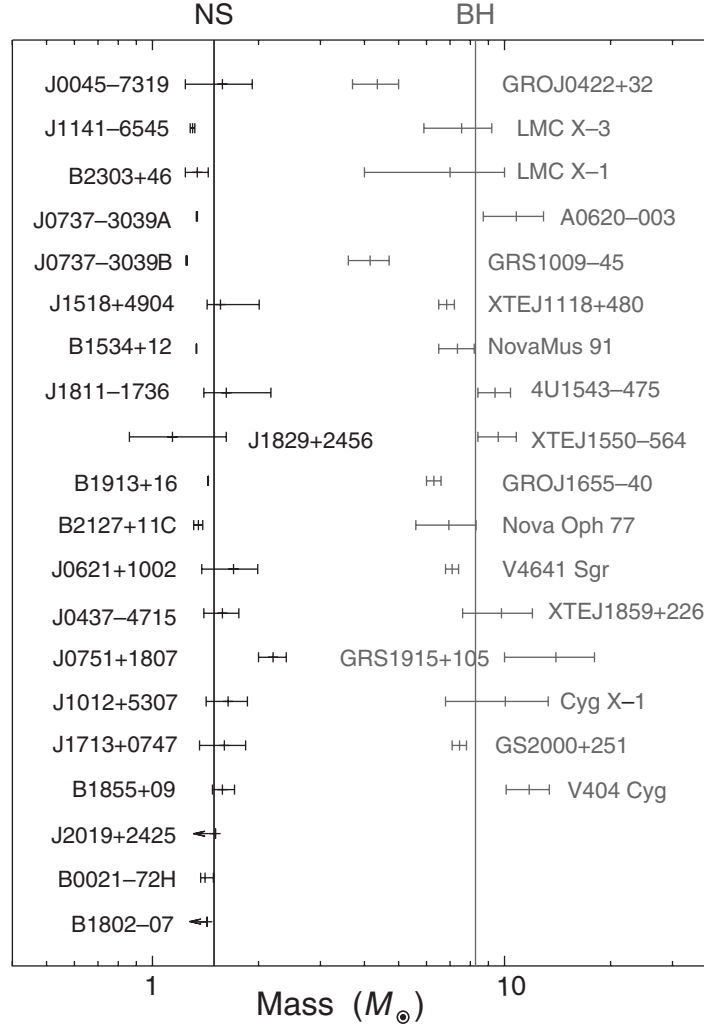
Figure 9. Measured masses for NSs and BHs in binaries. The average values are $M_{NS} = 1.49\,M_\odot$ and $M_{BH} = 8.27\,M_\odot$ (where the scatter is much larger). The data are taken from [155] for NSs, and [156] for BHs.

3.2. Waltzing couples

As a stellar binary revolves, gravitational waves carry away its energy and angular momentum. The loss rates depend on the system mass, semi-major axis, a, and eccentricity, and rapidly increase as orbital decay progresses. In the stage where $a \gg R_*$, where R_* is the typical stellar radius, the evolution can be computed to high accuracy in the weak-field limit, using post-Newtonian expansions for point masses [179]–[188]. The characteristic decay time for circular orbits of period P is given by

$$\frac{t_0}{P} \simeq 10^5 \left(\frac{P}{1\text{s}}\right)^{5/3}, \tag{21}$$

which, when applied to the binary pulsars PSR1913+16 [95] and PSRJ0737-3039 [189], gives $t_0 \simeq 300\,\text{Myr}$ and 85 Myr, respectively. These waveforms, calculated for the last few minutes

of the binary system's life, will be required as templates for detection with interferometric gravitational wave detectors such as the Laser Interferometric Gravitational Wave Observatory (LIGO) [190] and VIRGO [191]. For more common binaries with much longer periods, the decay rate will be negligible and the emission may be observed with the space-based Laser Interferometer Space Antenna (LISA). For systems reaching small separations, the decay rate will increase catastrophically and a collision will ensue, followed by relaxation to a final, nearly steady state in which, in many cases, only a BH will remain. This is the ring-down phase, in which the emission of gravitational waves and the configuration of the fossil BH can be studied through perturbative methods on a Kerr background (see [192] for a review). A composite gravitational radiation signal is shown in figure 10 as an example. The decay at early times ($t < 0$) was computed using the weak-field limit approximation for point masses, while the merger waveform itself ($t > 0$) is the result of a three-dimensional (3D) numerical simulation for the disruption of a NS by a point mass. The initially slow decay accelerates (while the amplitude and frequency of the signal increase) and comes to an abrupt end when the separation is of the order of the stellar radius, giving a complex signal which depends on the details of the NS structure. In principle, observation of such radiation and its frequency power spectrum could help constrain the mass–radius relationship for NSs [195].

The middle ground, during which a violent redistribution of angular momentum and energy takes place, will produce a burst of gravitational, neutrino, and electromagnetic energy, perhaps in the form of a classical GRB [196]. There is no approximate, analytical solution to be found in this regime, and multidimensional numerical simulations are clearly required to address the behaviour of the system. The process is highly dynamical, with thermodynamics and emission processes mattering little, and gravitational dynamics ruling the evolution (with the possible exception of systems containing WDs).

Even before high-resolution numerical simulations were carried out in essentially Newtonian gravity, it was assumed that the dynamical merger would indeed result in the formation of an accretion disk with enough mass and internal energy to account for the energy budget of a typical GRB, either through the tidal disruption of the NS in the former, or post-merger collapse of most of the central core in the latter. Compact binaries were thus seriously suggested as possible progenitors for GRBs at cosmological distances, with possible maximum power in excess of 10^{50} erg s^{-1} [71, 72], [88]–[94].

The characteristics of the problem, and the qualitatively different nature of the questions that can be of interest (or addressed) has dictated a range of approaches when searching for a solution. On the one hand, there is the question of generation of gravitational waves, clearly deserving of substantial effort in its own right, and tackled as such by several researchers. On the other, the generation of an electromagnetic signal (as a GRB, or another observable transient) requires the consideration of a different kind of input physics, mainly in the domain of thermodynamics, the equation of state under a wide array of conditions, nuclear physics, weak interactions and magnetohydrodynamics. Efforts in the field have generally proceeded along these largely orthogonal axes, with little overlap between them until quite recently, and it is only now that numerical simulations with any realistic microphysics are beginning to employ the full machinery of GR. It is thus only natural that results midway along this path have not always been in agreement, and indeed are not in some aspects even now. The recent and spectacular advances in the succesful evolution of BH binaries in numerical relativity give us today a glimpse of what may be possible to study in greater detail in the coming years (see the recent overview by Centrella [197]).
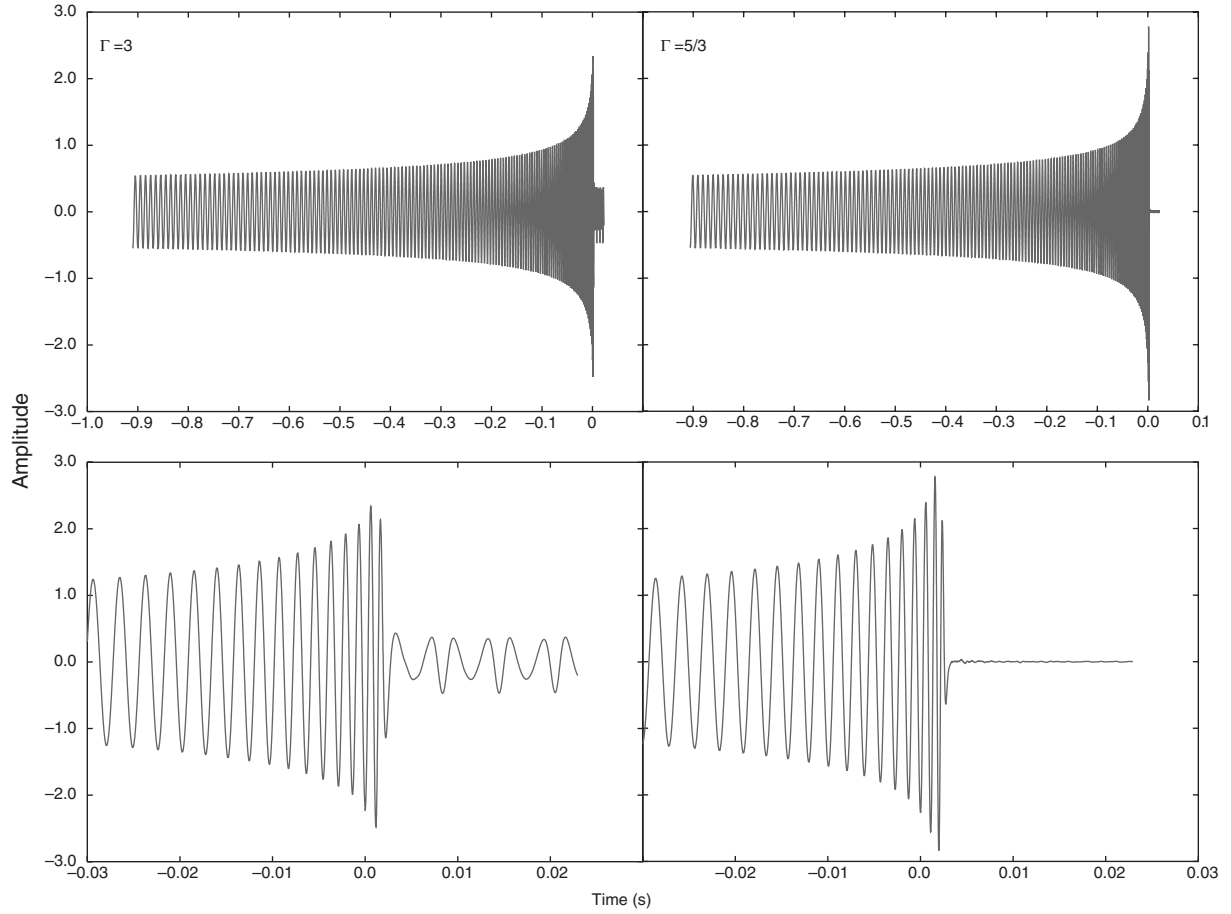
Figure 10. Gravitational wave signals for the final stages of spiral-in of compact binaries consisting of a BH and a NS with initial mass ratio $q = M_{NS}/M_{BH} = 0.3$ and adiabatic index $\Gamma = 3$ (left column [193]) and $\Gamma = 5/3$ (right column [194]). In this example, the observer is directly above the orbital plane of the binary along the rotation axis. The rise in amplitude and frequency shows the characteristic accelerated phase of decay, followed by a rapid transition to a ring–down after the violent tidal disruption of the star at $t \approx 0$. In the bottom panels, details of the merger and tidal disruption waveform are shown, from whose spectrum one could in principle extract valuable information regarding the equation of state at NS densities. Note that the NS with the more compressible equation of state is fully and rapidly disrupted, resulting in the total disappearance of the gravitational wave signal, whereas the one with a stiff equation of state leaves behind a low-mass orbiting remnant, which produces persistent emission at late times (see also figure 11).

3.3. First inroads

To our knowledge, the merger of compact objects (i.e. NSs and/or BHs) was first considered by James Lattimer and David Schramm [198, 199]. They initially studied the dynamics and disruption of a NS by a BH in a bound system, and the discovery of the Hulse–Taylor pulsar

in 1974 [95] soon made it clear that such encounters were an inevitable, if rare in terms of the lifetime of a typical galaxy, consequence of binary stellar evolution pathways. The computations were carried out in full GR, with certain simplifications to make them tractable and their interest originally laid in the possibility of such events ejecting neutron–rich material to the interstellar medium, which might subsequently undergo r-process nucleosynthesis and contribute to the observed abundances of heavy elements [200, 201]. GRBs had only recently been made known to the astrophysical community [202], and the likelihood of them being of Galactic origin made Lattimer and Schramm consider such events to be unrelated. In their words: 'Although the frequency of these events is too small to be important as far as currently observed anti-neutrino, γ-ray, or optical bursts are concerned, enough NS material may be ejected to be of nucleosynthetic importance' [199]. Over thirty years after their pioneering work, coalescing compact binaries stand today as one of the leading runners in models for SGRBs. It is noteworthy that only now are detailed computer simulations of such collisions beginning to fully consider the effects of GR in the dynamical merger phase of tidal disruption (John A Wheeler's 'tube of toothpaste' mechanism [203, 204]).

## 3.4. Dynamical stability of close binaries

The stiffness of the nuclear equation of state, or equivalently, the compressibility, is a parameter derived from microphysics that has a predominant role in the final outcome of a merger. It is most easily expressed by writing the pressure–density relation as

$$P \propto \rho^{\Gamma}, \tag{22}$$

where Γ is the adiabatic index. Figure 11 shows the effect of its variation on the structure of a star by plotting the density profiles of static, spherically symmetric configurations in hydrostatic equilibrium. For $\Gamma \gg 1$, the result is a nearly-constant density (a stiff equation of state), while for $\Gamma \to 4/3$ (lower values will not produce a stable solution in Newtonian gravity) the star is highly centrally condensed (a soft equation of state). It is immediately obvious that when placed in the vicinity of a massive companion, the star made of more compressible material will feel tidal effects to a much lesser degree. The issue of how matter responds to pressure at extremely high densities is of course far from being solved. For example, while the stiffness of ordinary matter may be quite high at $\rho \simeq \rho_{\mathrm{nuc}}$, with $\Gamma \simeq 2\text{-}3$, the presence of exotic condensates at supra-nuclear densities could soften the equation of state considerably [205], leading to potentially different behaviour.

The equilibrium configurations for an incompressible fluid, and their stability, were considered by Chandrasekhar [206], both for single and binary stars, in combinations that could represent double NSs or NSs in orbit around point masses. This work was extended later by Lai and collaborators [207]–[211] to compressible configurations by the use of polytropic relations as equations of state. In a series of papers, they showed that purely Newtonian hydrodynamic effects can de-stabilize close binaries if the equation of state is sufficiently stiff ($\Gamma \geqslant 2$). This can occur before contact, when the separation is typically $\sim$3 stellar radii. The effective binary potential becomes so steep that a dynamical (i.e. on an orbital timescale) plunge occurs, forcing the two stars to come together. This is independent of any consideration of GR, and clearly shows just how powerful the departures from point-mass behaviour can be when the combination of a finite stellar radius and compressibility are considered. For illustrative purposes, we show
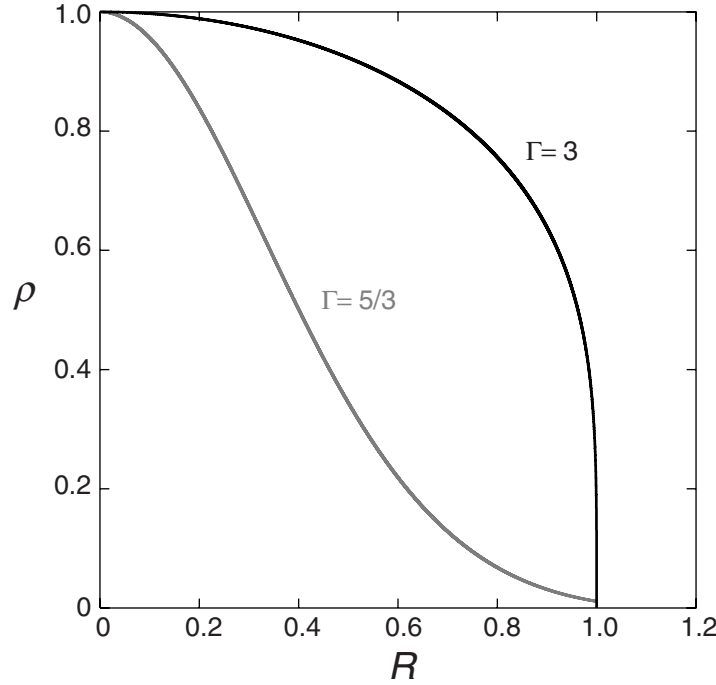
Figure 11. Normalized radial density profiles for spherically symmetric configurations of a given mass in hydrostatic equilibrium, for different compressibilities (parametrized through the adiabatic index Γ). The varying degree of central condensation makes the stiffer stars (with higher values of Γ) more susceptible to tidal effects when placed in a binary system.

in figure 12 the equilibrium curves of total orbital angular momentum for a Newtonian NS-point mass binary for two compressibilities, as well as the point mass result (which is simply $J_{eq}(R) \propto R^{1/2}$). This has important implications for the initiation of mass transfer from one star to the other, the occurrence of tidal disruption and the final configuration of the system once it comes into near contact (see e.g. chapter 4 in [212]).

The compressibility also impacts directly upon the mass radius relation of a star. For polytropes, we have

$$R_* \propto M^{\Gamma-2/3\Gamma-4}, \tag{23}$$

so the star will expand upon mass loss (or contract) for $\Gamma < 2$ (or $\Gamma > 2$). Thus it can further overflow its Roche lobe once mass transfer begins, in which case the process itself is unstable leads to complete disruption, or retreat from the Roche surface and, in principle, shut off mass transfer in the absence of external driving (which will not occur, because gravitational radiation emission will remove orbital angular momentum continuously). In either case, the details of the dynamics will depend on the global adiabatic index at high densities.

3.5. Investigating the merger phase

The numerical studies of coalescing binaries began in the 1980s and 1990s, with computations of the gravitational wave emission and determinations of the stability and dynamics of the system
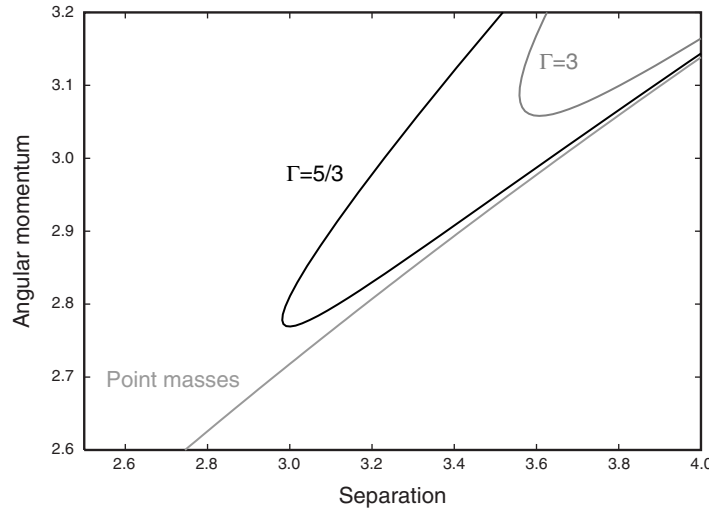
Figure 12. Total dimensionless angular momentum as a function of separation (in units of the NS radius) for equilibrium configurations of BH–NS binaries with mass ratio $q = 0.31$ in a Newtonian approximation. The NS is modelled as a non-spinning tri-axial Roche–Riemann ellipsoid (see [208]) of varying adiabatic index (curves for $\Gamma = 3; 5/3$ are shown). Tidal effects increase the steepness of the effective interaction potential and produce a minimum in $J_{eq}(R)$, indicating the onset of a dynamical instability. The effect of the equation of state is clear, with the less compressible NS encountering instability, which will induce a dynamical plunge, at a greater separation than its more compressible counterpart. For reference, the equilibrium curve for point masses is also shown. As expected, it is a monotonically increasing function of radius and exhibits none of the turning points characteristic of the appearance of an unstable branch.

at separations comparable with the stellar radius under various simplifying approximations [96, 97, 206], [211]–[228]. The thermodynamical evolution of the fluid during merger, with the potential for electromagnetic energy release and GRBs was also considered by various groups [98, 127], [229]–[234]. These studies used Newtonian or post-Newtonian gravity with additional terms included in the equations of motion to mimic gravitational radiation reaction (which is the ultimate agent driving the binary evolution), and in some cases approximations to the equations of GR.

Figure 13 shows the dynamical evolution of a merging NS–BH binary, during which the NS (modelled as a polytrope) is tidally disrupted and an accretion disk promptly forms [98], [235]–[238]. The general conclusion of this body of work (which includes a mock radiation reaction force to account for gravitational wave emission) is that when the orbital separation is of the order of a few stellar radii, the star becomes greatly deformed due to tidal effects, and the system becomes dynamically unstable. For stellar components with relatively stiff equations of state ($\Gamma > 5/2$) the instability is due to a strong steepening of the effective potential because of tidal effects (figure 11). In the opposite limit, the mass–radius relationship for the star is such that mass loss leads to an expansion and Roche lobe overflow, further accelerating the process and leading to a runaway in which disruption occurs. All of the features anticipated in the work of Lai and collaborators concerning the stability of the binary are apparent, and
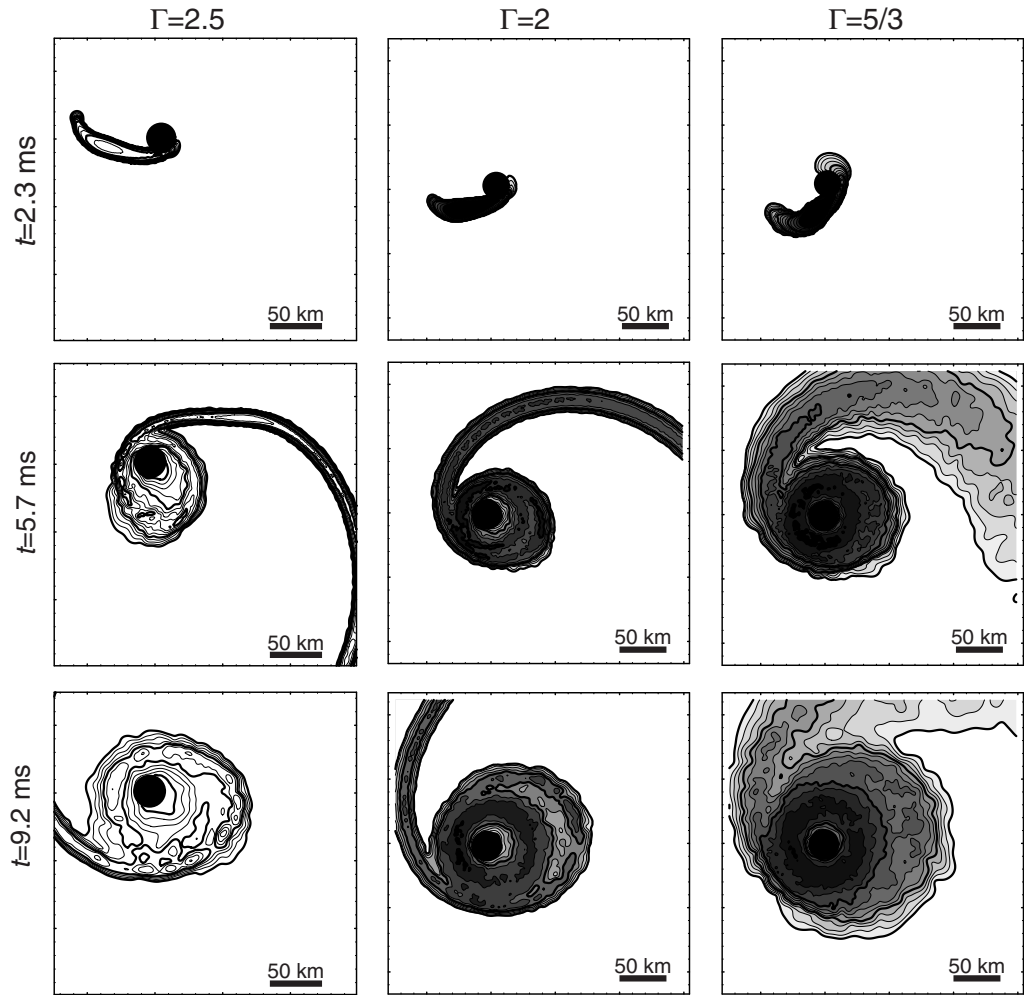
Figure 13. The tidal disruption of a NS by a BH in a binary: each column (top to bottom in time) depicts the interaction (computed in 3D) for a given value of the adiabatic index Γ, going from stiff to soft (left to right). Logarithmic density contours of density (spaced every quarter decade, with the lowest one at $\log \rho(\text{g cm}^{-3}) = 11$), are shown in the orbital plane, with additional shading for $\Gamma = 2$; $5/3$. The distance scale is indicated in each panel, and a black disk of radius $2R_\text{g}$ represents the BH. All three cases initially had identical mass ratios, $q = 0.31$ [193, 194].

are seen as well in the numerical simulations of merging binaries performed by Rasio and collaborators [226, 227, 239]. The main additional result in terms of the dynamics is that configurations that are stable in principle up to Roche lobe overflow are de-stabilized by the mass transfer process itself, and fully merge as well within a few orbital periods. It is also quite clear that in the presence of gravitational radiation reaction (even in its crudest approximation) the system enters a dynamical infall at small separations that no amount of (even conservative) mass transfer can revert. Qualitatively, the result is largely independent of the mass ratio, the assumed initial condition in terms of NS spin (both tidally locked and irrotational configurations were explored in [235]–[238]) and the assumed compressibility. Quantitatively,
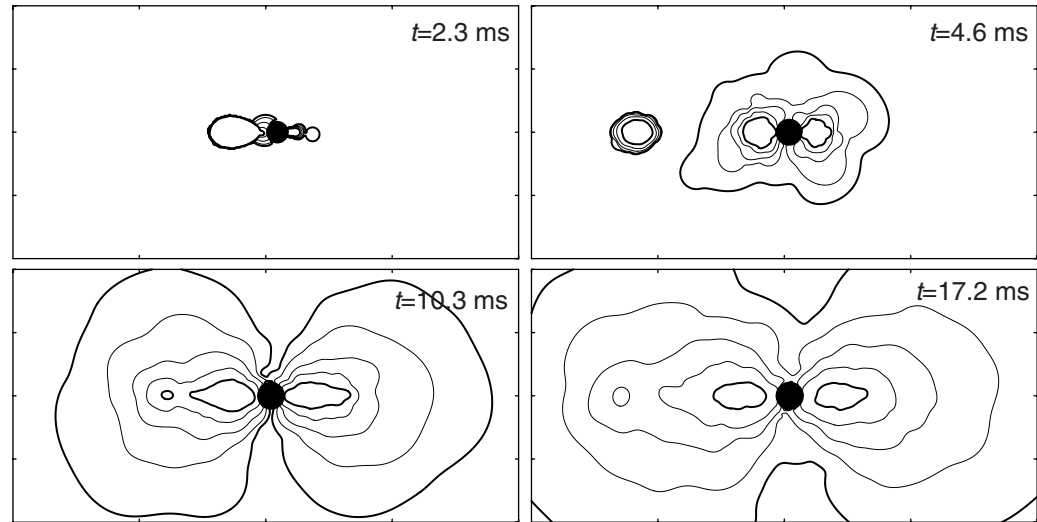
Figure 14. During the tidal disruption of a NS by a BH in a binary, an accretion disk is formed from some of the material stripped from the NS. A sequence of meridional slices through a 3D simulation is shown, plotting logarithmic density contours for the density (contours are spaced every decade, with bold ones at $\log \rho(\mathrm{g\,cm^{-3}}) = 8; 12$). In the first panel, the tidally deformed star is still visible to the left of the BH. By 4 ms,the disk is growing around the BH, and a slice of the tidal tail (see figure 13) is visible to the left of the disk. By 17 ms the disk is practically fully formed, and is draining onto the BH on a much longer time scale, determined by angular momentum transport effects. This particular calculation used $\Gamma = 5/3$, and the box shown covers $400 \times 200\,\mathrm{km}$ [193].

the details are different, and are primarily reflected in the final disk mass and the gravitational wave signal (see, e.g., [240]).

After a few initial orbital periods, a BH of $M \sim 3 - 5\,M_\odot$, surrounded by a thick and hot debris disk ($M_{\mathrm{disk}} \sim 0.01 - 0.1\,M_\odot$) approximately 4×10^7 cm across is all that remains of the initial couple (figure 14).

For double NSs the situation is essentially the same in terms of the orbital dynamics prior to merger. Synchronization during spiral-in is impossible because the viscosity in NS matter is too small [241, 242], and so they merge with spin frequencies close to zero, compared with the orbital frequency (such configurations are termed *irrotational*). Contact occurs at subsonic velocities and a shear layer then develops at the interface. Modelling this numerically is extremely difficult, because the layer is unstable at all wavelengths and vortices develop all along it [243], possibly amplifying the magnetic field to extremely large values very quickly [129]. The mass ratio in double NS binaries is quite close to unity, so a fairly symmetrical remnant would be a natural and expected outcome. However, in general even small departures from unity can have important consequences in the inner structure of the final object: the lighter star is disrupted and spread over the surface of its massive companion, which can remain largely undisturbed [227, 229, 244]. The final configuration now consists of a supra-massive NS (i.e. one with more mass than a cold, non-rotating configuration could support) surrounded again by a thick shock-heated envelope and hot torus similar to that formed in BH–NS mergers. The centre of the remnant is rapidly,

and in many cases, differentially rotating, which can have a profound impact on its subsequent evolution (see below, subsection 3.7).

By definition, in either case the merger entails the formation of an object rotating at the break up limit. Material dynamically stripped from a star is thus violently ejected by tidal torques through the outer Lagrange point, removing energy and angular momentum and forming a large tail. For double NS binaries there are two such structures, and obviously only one in a BH–NS pair. The mass (typically up to a few tenths of a solar mass), energy and composition of these structures can be more easily studied by Lagrangian particle based methods, as they are several thousand kilometres across by the time the merger is over, and have thus moved off the numerical domain in grid based codes. By the inherent asymmetry in BH–NS binaries (in form and mass ratio), the BH can receive a substantial kick (up to 10^2 km s^{-1} [97]) during the process, in addition to that arising from gravitational wave emission. Some of the fluid (as much as a few hundredths of a solar mass) in these flows is often gravitationally unbound, and could, as originally envisaged by Lattimer and Schramm, undergo r-process nucleosynthesis [228, 245]. The rest will eventually return to the vicinity of the compact object, with possible interesting consequences for SGRB late time emission (see section 5).

It is important to note that the debris disks are formed in only a few dynamical time scales (a few milliseconds), which is practically instantaneous considering their later, more leisurely evolution (which will be detailed below).

The possibility that a fraction of the NS in a BH-NS binary would survive the initial mass transfer episode and produce cycles of accretion was explored at long time scales by Davies *et al* [246]. This was motivated by the fact that numerical simulations employing relatively stiff equations of state at high binary mass ratios [97, 194, 247] showed that this might actually occur. However, this is most likely not the case, because: (i) the required equation of state was unrealistically stiff; and (ii) gravity was essentially computed in a Newtonian formalism. More recent calculations with pseudo-Newtonian potentials [27, 248] and the use of GR [249, 250] consistently fail to reproduce this behavior for a range of compressibilities in the equation of state. Instead, the NS is promptly and fully disrupted soon after the onset of mass transfer.

3.6. Effects of GR

Performing calculations with full GR is clearly necessary, but is not an easy task. More importantly, it is not even clear how to estimate its effects with simple analytical considerations. Computations of the location of the *innermost stable circular orbit* (ISCO) in BH systems indicate that tidal disruption may be avoided completely, with the star plunging directly beyond the horizon, essentially being accreted whole in a matter of a millisecond [251]. This would preclude the formation of a GRB lasting 10 to 100 times longer. But we stress that, just as for stability considerations in the Newtonian case, even if the location of the ISCO can be a useful guide in some cases, it cannot accurately describe the dynamical behaviour of the system once mass transfer begins. In pseudo-Newtonian numerical simulations [248] and post-Newtonian orbital evolution estimates [252], even when a near radial plunge is observed, the star is frequently distorted enough by tidal forces that long tidal tails and disk-like structures can form. The outcome is particularly sensitive to the mass ratio $q = M_{NS}/M_{BH}$ and initial GR calculations showed that the spin of the BH is also very important, with rotating BHs favouring the creation of disks [253]. Dynamical calculations of BH–NS systems in pseudo-Newtonian potentials that mimic GR effects typically show that for mass ratios $q \simeq 0.25$ it is possible to

form a disk, although of lower mass than previously thought, $M_{\text{disk}} \approx 10^{-2} M_\odot$ [27, 248]. A long one-armed tidal tail is formed as in the Newtonian calculations, with some material being dynamically ejected from the system. Recently, Faber *et al* [249, 250] have presented their first results for the dynamical merger phase of a BH–NS binary in GR for low mass ratios ($q \simeq 0.1$). While the equation of state is simple (a polytropic relation with $\Gamma = 2$ was assumed), the final outcome is clearly dependent on the complex dynamics of mass transfer. Even though at one point in the evolution most of the stellar material lies within the analytically computed ISCO, tidal torques transfer enough angular momentum to a large fraction of the fluid, producing an accretion disk with $M_{\text{disk}} \simeq 0.1 \, M_\odot$ by the end of the simulation at $t \simeq 70$ ms. The BH horizon is a point of no return, but the ISCO is most definitely *not*.

Extending our own work, and further motivated by the first accurate localizations of SGRB afterglows (starting with GRB050509B, [6]), we computed the binary evolution of BH–NS systems in a pseudo-Newtonian potential in an attempt to better estimate the circumstances under which a disk may form, and its specific configuration if it does [27, 238]. The most important parameters are the disk mass and size, as these set the global energy, density, and time scales for its evolution. The calculations employed the same formalizm and code as before [193], substituting the pseudo-Newtonian potential of (PW) [254], and using irrotational binaries as initial conditions (i.e. ones in which the NS spin is negligible). We explored variations in both the mass ratio ($q \sim 0.1$–0.3) and NS compressibility ($\Gamma \sim 5/3$–2), and found that full tidal disruption takes place as the star approaches the BH, with a torus forming around it. It contains, however, substantially less mass than in the Newtonian case, because a larger fraction is directly accreted by the BH. Additionally, a large portion of the remaining material is on highly eccentric orbits in the long tidal tails, and, while it will return to the vicinity of the BH, it does not constitute immediately part of the accretion disk. Overall the torus mass is at least one order of magnitude lower than previously estimated, and somewhat colder, since it has not been shock-heated to the same degree by self-interaction of the accretion stream (see figure 13). Rosswog [248] has performed high resolution calculations of this type as well, but with the use of a realistic equation of state, and finds similar evolution and outcomes.

In the case of merging NS pairs, general relativistic calculations [255]–[257] have progressed even further, going beyond the adoption of a simple polytropic pressure–density relation [258]–[260]. The outcome is predictably complicated, and depends sensitively on the total mass of the system and the initial mass ratio. Asymmetric binaries tend to produce more massive disks than those with identical components. As mentioned earlier for the case of Newtonian calculations, a small deviation from a mass ratio of unity is significant, since it can lead to the nearly complete tidal disruption of the lighter component, while the more massive star is largely unaffected. Shibata and Taniguchi [259] find that the lower the mass ratio, the more massive the resulting accretion disk, reaching $M_{\text{disk}} = 0.03 \, M_\odot$ for $q = 0.7$. Both PSR1913+16 and PSRJ0737-3039 have $q > 0.9$ [155, 157, 168, 189], so this mass ratio may be unrealistically low, but when $q \simeq 0.9$ a considerable disk containing $\simeq 10^{-3} M_\odot$ still forms. This can easily account for the 10^{49}–10^{50} erg required for a typical SGRB, simply based on its gravitational binding energy.

We give in table 3 a summary of estimates for the disk properties for different progenitors, based on the calculations of various groups described above. The characteristic masses, maximum densities and temperatures are $M_{\text{disk}}/M_\odot \simeq 10^{-4}$–$10^{-1}$, $\rho \simeq 10^{10}$–10^{12} g cm^{-3} and $T \simeq 10^{10} - 10^{11}$ K respectively, and provide a starting point for the more detailed disk evolution calculations we will describe below in section 4.

Table 3. Remnant disk masses in compact mergers for a variety of equations of state and various gravity methods: Newtonian (N); Paczyński and Wiita (PW); and General Relativity (GR).

Programme	$M_{disk}/M_\odot$	Gravity and method	Equation of state	References
BH/NS	0.1–0.3	N, SPH[a]	Polytropes	[235]–[238]
BH/NS	0.03–0.04	PW, SPH	Polytropes	[27]
BH/NS	0.26–0.67	N, Grid	LS [261]	[98]
BH/NS	0.001–0.01	PW, SPH	Shen [262]	[247, 248]
BH/NS	0.001–0.01	GR, SPH	Polytropes	[249, 250]
NS/NS	0.2–0.5	N	SPH, Polytropes	[224, 225]
NS/NS	0.4	N, SPH	Polytropes	[229]
NS/NS	0.01–0.25	N, Grid	LS [261]	[231]–[233]
NS/NS	0.25–0.55	N, SPH	LS [261], Shen [262]	[226]–[229], [234]
NS/NS	0.05–0.26	GR, SPH	Shen [262]	[260]
NS/NS	0.0001–0.01	GR, Grid	APR [263]	[259]

[a] SPH: smooth particle hydrodynamics.

3.7. The fate of the central core

In the case of merging NSs, the ultimate fate of the central object is still unresolved, and depends on the maximum mass that a hot, differentially rotating configuration can support. It has been known for some time that the maximum allowed mass can be increased by the effects of rotation (see, e.g. [264]), and in particular, *differential* rotation [265]. In addition, the post-merger core is certainly not cold, as shock heating (at least for a mass ratio of unity) can raise the internal energy substantially. Shibata and Taniguchi [259] find that this threshold for collapse can be estimated as $M_{thres} \simeq 1.35\, M_{cold}$, where M_{cold} is the corresponding value for a non-rotating and spherical, cold configuration. Based on the recent observation of PSRJ0751+1807 [162], $M_{cold} \geqslant 2.1\, M_\odot$, so a total mass greater than $\simeq 2.83 M_\odot$ is required for prompt collapse to a BH. If $M_{cold} < M_{core} < M_{thres}$, various mechanisms (e.g. emission of gravitational waves, redistribution of angular momentum, magnetic fields) could act to dissipate and/or transport energy and angular momentum, possibly inducing collapse after a delay which could range from seconds to weeks (this has also been suggested as a two-step process which would be responsible for a SN/GRB association in the case of long GRBs [266]).

Differential rotation, besides providing additional support against collapse, could amplify seed magnetic fields in the core to large values [127]–[129] and turn it into a powerful magnetar. Spin-down torques, magnetic field winding and flaring alone could then in principle power a GRB, independently of the presence of a torus-like accretion structure [97, 144]. Note that this would also be one ingredient of the expected outcome of accretion-induced collapse of a WD, or a double WD merger. The surrounding environment would be much different, clearly, and its effect on the generation of a relativistic fireball would certainly be important (and possibly catastrophic [176]). The magneto rotational instability (MRI) may also operate in such an environment, transferring angular momentum to the debris disk fast enough to keep it from being swallowed by the core when (and if) it collapses [267]–[269]. An alternative possibility is that the core may deform into a bar-like structure, provided the equation of state is sufficiently stiff [259]. Gravitational torques

in the presence of this asymmetry may then transfer angular momentum to the orbiting debris and increase the mass of the disk. In either of these last two cases one would then have a system similar to that occurring in a BH–NS merger. In general, the details remain highly uncertain, but powering a GRB is still a possible evolutionary pathway (see figure 21 in [259]).

3.8. Colliding compact objects

The collision of unbound compact objects, rather than their merger in binaries, has not received much attention in the context of GRBs. Janka and Ruffert [232] computed the interaction of two identical NSs on parabolic orbits, and found a post-collision environment so contaminated with baryons that they concluded it was not viable as a GRB central engine. In addition, they estimated the event rate within galaxies, and found it much too low to be of interest. We have reconsidered this problem [106], and find that if interactions in GCs are taken into account, the rates can have an important effect on the production of SGRBs, if the outcome of the collision is a favourable one. Exactly how frequent a collision takes place depends on the velocity dispersion and the stellar density in a given GC, as well as the number density of GCs in a given galaxy [106].

A similar scenario, the tidal disruption of stars by massive BHs (where the mass ratio is typically $q = 10^{-7}$) was considered as a mechanism for feeding AGN, thus accounting for their luminosity [270, 271]. Simple initial estimates and more detailed analytical calculations [272, 273] were subsequently confirmed in their fundamental aspects through numerical simulations by various groups [274]–[276].

The relevant quantity fixing the strength of the interaction is

$$\eta = \left(\frac{M_*}{M_{\mathrm{BH}}} \frac{R_{\mathrm{p}}^3}{R_*^3} \right)^{1/2}, \tag{24}$$

where M_* and M_{BH} are the stellar and BH mass, R_{p} is the pericentre distance and R_* is the stellar radius. When $\eta \approx 1$ the tidal field is sufficiently strong to disrupt the star (the energy for this is ultimately extracted from the orbital motion). For the scenario considered by Rees [273], the star moves essentially in a fixed background metric since $q \ll 1$ (a condition fully exploited in the numerical calculations later carried out), and the typical pericentre distance is up to one hundred times the stellar radius. Thus the star does not directly impact its companion, and the interaction is entirely gravitational. It was found that approximately half the mass of the star is ejected from the system (at greater than escape velocity) and the other half remains bound. Using the Keplerian relation

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}t} = \frac{1}{3}(2\pi G M_{\mathrm{BH}})^{2/3} t^{-5/3}, \tag{25}$$

for the orbiting material, where $\mathrm{d}\epsilon$ is the specific energy, and the fact that the differential mass distribution with energy, $\mathrm{d}M/\mathrm{d}\epsilon$, is practically constant, the rate at which mass returns to the BH is given by $\mathrm{d}M/\mathrm{d}t \propto t^{-5/3}$.

Much can be learned from these calculations, and applied to the case of colliding compact objects of comparable mass. Much is different as well, primarily because the mass ratio is of order $q \approx 1$–0.1, rather than 10^{-7}. This means that the pericentre distance is now comparable to the stellar radius. The initial encounter may thus resemble more a direct impact, with immediate exchange of mass, and a corresponding alteration to the mass ratio on a dynamical time scale. A second effect, as for mergers, is related to the compressibility of the material, and how the star responds in radius once mass transfer begins.

We have now computed the outcome of parabolic encounters between BHs, NSs and WDs, in various combinations, to determine the structure of the remnant and investigate its viability as a possible GRB central engine. The obvious difference between systems containing a WD and those that do not is that the time scales are typically longer in the former by a factor $(R_{WD}/R_{NS})^{3/2}$. It is therefore natural to expect accretion and possible flares at hundreds, or even thousands of seconds, rather than tens of seconds after the main interaction. Whether these manifest themselves in γ-rays at all is another matter, and one that cannot be easily tackled numerically yet. The reason is that the WD radius is typically a thousand times larger than that of its companion, and hence the dynamical range that needs to be properly resolved is correspondingly increased.

We find that collisions occur in two stages (see figure 15). During the first periastron passage, an outward tidal stream is stripped from the low-mass member and ejected at high velocity. This act binds the surviving core, which rapidly returns for a second passage and is entirely disrupted, forming a second tidal tail, as well as an accretion disk around the massive companion. The duration of the evolution up to this point is comparable to the final disruption in a binary (about 10 ms). The end result is the formation of a system similar to that occurring in a binary merger, in that it contains a massive body (with the same uncertainties in delay and pathway to collapse to a BH as before) surrounded by a debris disk of a few tenths to a few hundredths of a solar mass. We also considered the case of a WD colliding with a massive NS (or low-mass BH). We were unable at this point to follow the entire evolution of the system, because of the longer timescales involved, but a large tidal tail is also formed at first encounter, with the surviving core being bound as a result. Already, however, a disk is apparent, being a thousand times larger than in the NS/BH encounter. A detailed dynamical study of this type of collision will be presented elsewhere [106].

Despite a variety of possible initial configurations, our best guess for the ultimate prime mover thus consists of a similar astrophysical object: a stellar mass BH surrounded by a hot debris torus; moreover the overall energetics of these various progenitors differ by at most a few orders of magnitude, the spread reflecting the differing spin energy in the hole and the different masses left behind in the orbiting debris[8]. On the hypothesis that the central engine involves hyperaccreting BHs, we would obviously expect the hole mass, the rate at which the gas is supplied to the hole, and the angular momentum of the hole to be essential parameters. In this case there is no external agent feeding the accretion disk (excluding tidal tail interactions, which we will consider in detail in section 5), and thus the event is over roughly on an accretion time scale. Any attempt to explain their properties thus requires that we now consider its evolution and associated energy release.

---

[8] A possible exception includes the formation of a rapidly rotating (in some cases very massive) NS with an ultrahigh magnetic field.

NS–NS   BH–NS   WD–BH
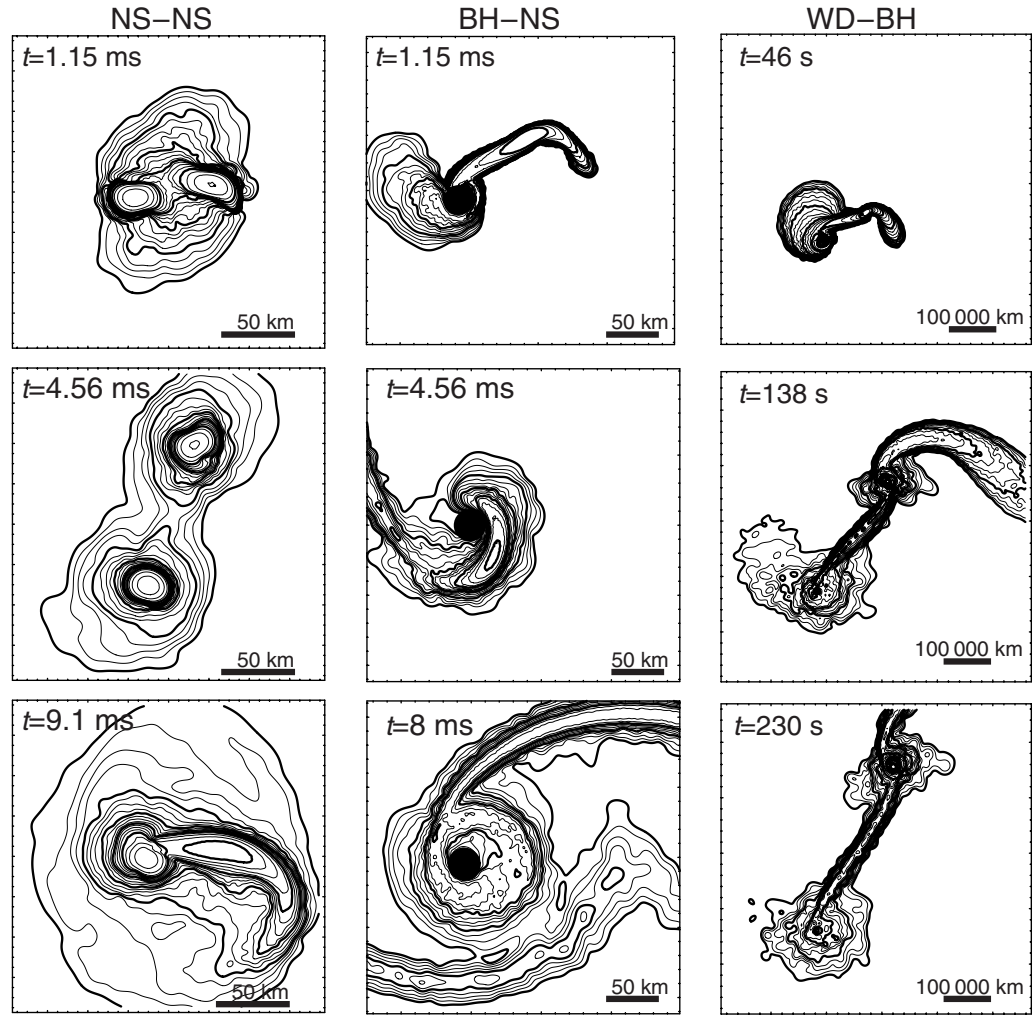
Figure 15. The collision of compact objects, and their remnants: the columns (top to bottom in time) show parabolic encounters between two NSs with masses 1.75 M$_\odot$ and 1.4 $M_\odot$ (left panel), a NS and a BH with masses 1.4$M_\odot$ and 4.5 $M_\odot$ (middle panel), and a WD and a massive NS with masses 0.5 $M_\odot$ and 2.5 $M_\odot$ (right panel). All NSs are modelled with adiabatic indices $\Gamma = 2$, and the WD with index $\Gamma = 5/3$, appropriate for a low-mass configuration. The scale is indicated in each panel, with logarithmic contours of density shown in the orbital plane (spaced every quarter decade, with the lowest one at $\log \rho(\mathrm{g\,cm^{-3}}) = 10; 10; 0$ for each column respectively). The two tidal tails and the accretion disk are clearly visible in the BH–NS collision. Note the large difference in spatial and temporal scales for the case involving the WD. The estimated delay for the WD core to return to the vicinity of the BH is 1000 s. All encounters have a parameter $\eta = 1$, as defined in equation (24).

4. Neutrino cooled accretion flows (NDAFs) and SGRBs

The way in which gas flows on to an accreting object depends largely on the conditions where it is injected. In this section we address the general dynamical evolution of NDAFs under the conditions expected to occur in SGRB central engines. There are five sections: subsection 4.1 is devoted to general considerations and a review of previous work addressing neutrino cooled flows; subsection 4.2 gives a summary of the microphysical and thermodynamical conditions in the fluid and considers the relevant equation of state and cooling processes in the disk; detailed calculations in a pseudo-general relativistic potential are presented in subsection 4.3; the generation of disk-driven winds and their accompanying outflows is discussed in subsection 4.4; finally, stability considerations are addressed in subsection 4.5.

4.1. General considerations

Despite all of the complications inherent in the study of neutrino cooled flows (and accretion flows in general), two main ingredients, easier to qualify, are crucial for understanding the life and death of the accreting source. The first is simply the total mass available in the system at the outset. This fixes the global energy scales, both thermal and gravitational, and allows for an approximate determination of the available energy, with

$$E_{\mathrm{gr}} \simeq E_{\mathrm{th}} \simeq \frac{GM_{\mathrm{BH}}M_{\mathrm{disk}}}{R_{\mathrm{disk}}} \simeq 10^{52} \left(\frac{M_{\mathrm{BH}}}{3M_\odot}\right) \left(\frac{M_{\mathrm{disk}}}{0.1M_\odot}\right) \left(\frac{R_{\mathrm{disk}}}{10^7 \mathrm{cm}}\right)^{-1} \mathrm{erg}. \quad (26)$$

Note that we have estimated the thermal and gravitational energies to be of the same order. This is different than for classical thin disks, because the hypercritical accretion flows we are to consider are born in highly dynamical situations where there is an enormous store of internal energy (this is true for binary mergers as well as for collapsar-type scenarios). The fundamental reason is that in either case, internal energy was the primary source of support prior to the catastrophic gravitationally-induced collapse or tidal disruption.

The second is the nature of the mechanism feeding the inner disk with mass and energy as a function of time. In the standard picture following core collapse, normally envisaged for long GRBs [85, 277, 278], the inflowing mass supplies an accretion rate $\dot{M} \sim 10^{-3}\,M_\odot\,\mathrm{s}^{-1}$ for a time $t_{\mathrm{f}} \simeq 10\,\mathrm{s}$ and is thereafter fed by injection of infalling matter at a rate that drops off as $t^{-5/3}$. For the case of merging compact objects, it is usually assumed that no external agent feeds the previously assembled disk, and thus that the event is over roughly on an accretion timescale. This picture can be altered by the presence of tidal tails formed with material stripped from the stars during the initial merger phase and ejected on eccentric orbits. We will address this later. Regardless of the tidal tails, the physical conditions and instantaneous structure within the flow are similar in both scenarios, and conclusions drawn from one case may be used to infer the situation in the other.

The equivalent Eddington luminosity for neutrinos fixes a rough upper bound for the power output (considering coherent scattering off free nucleons as a source of opacity) of $L_{\mathrm{Edd},\nu} \simeq 10^{54}\,\mathrm{erg\,s}^{-1}$. Combined with the available energy in the disk, one finds a characteristic lifetime at this power level of $t_{\mathrm{Edd},\nu} \sim t_{\mathrm{burst}} \simeq 10\,\mathrm{ms}$. Two important effects alter this estimate: (i) the power is in fact much smaller than $L_{\mathrm{Edd},\nu}$, by at least two orders of magnitude, and (ii) accretion onto the BH drains the disk of mass and energy on a viscous time scale, t_α, which can be shorter if the efficiency of angular momentum transport is high.

Speculation about the structure and evolution of a massive accretion disk surrounding a new-born BH began as soon as they were suggested to be the prime movers of GRB sources [88, 90, 92]. Despite the large theoretical uncertainties, it was clear from the start that indeed dense, rapidly evolving structures with huge accretion rates (on the order of one solar mass per second) were to be expected, with natural time scales (dynamical and viscous) that could in principle account for GRBs. Whether or not the resulting signal turns into a burst or not is still a matter of debate, but as Stan Woosley put it [279] 'If the signature of a 5 $M_\odot$ BH accreting stellar masses of material in a minute is not a gamma-ray burst, what is it?'. The physical regime at the expected densities and temperatures places these systems alongside core-collapse supernovae in the bestiary of astrophysics, and possibly makes them even more energetic.

The detailed structure of NDAFs has been investigated by a number of groups over the past few years. Initially only steady-state, azimuthally symmetric and vertically integrated solutions were considered, with increasing level of detail mainly in what concerns the thermodynamics and the equation of state [280]–[283], but also in the effects of GR [284]. These studies indicated that the instantaneous power output for plausible accretion rates, if maintained, could account for the overall energetics of GRBs. Neutrino cooling, when initiated at the proper temperature, would be an efficient mechanism for the removal of internal energy, and the disk would thin rather abruptly. The 1D solutions of Popham *et al* [280] were matched in the inner regions quite nicely with the initial numerical results for collapsar simulations [278]. Narayan *et al* [281] considered among other things, how the radius of matter injection would affect the flow, and concluded that it needed to be rather small in order for neutrino cooling to be efficient. Otherwise, a large fraction of the accreting mass at large radii would, being unable to get rid of its internal energy, simply be blown away in a large-scale outflow. Kohri and Mineshige [282] then considered microphysical effects in greater detail, both in terms of the equation of state and the cooling processes. Finally, the effects of GR on the disk structure have been studied by Chen and Beloborodov [284] for stationary flows around Schwarzschild BH.

More recently, time-dependent calculations in 1D [285], 2D [286, 287]—assuming azimuthal symmetry—and 3D [288, 289] have tackled the question of time dependence and instabilities. The trade-off between relaxing the assumption of ϕ-symmetry or not is reflected in the time interval that can be reliably modelled at high resolution: $t_{sim} \simeq 50$ ms in 3D versus $t_{sim} \simeq 1$ s in 2D. The former allows for a full exploration of azimuthal modes and instabilities, while the latter may provide reliable physical estimates on timescales comparable to the duration of SGRBs. Both 2D and 3D simulations have relied on initial conditions taken from 3D binary merger calculations [193, 290], and as such are a natural extension of these models. It has become clear from these studies that the neutrino energy release is rather rapid. As long as the accretion (or viscous) time scale $t_{acc} \sim M_{disk}/\dot{M}$ is longer than the cooling time $t_{cool} \sim E_{int}/L_\nu$, the fluid will be able to radiate most of its internal energy. The power is thus maintained at a fairly constant level for up to 100 ms, then drops rapidly. However, the density in the disk remains fairly high for essentially t_{acc}, which can be as long as a few seconds. It is thus in principle able to anchor strong magnetic fields capable of driving MHD flows. Another important result is that for high enough accretion rates (or equivalently, densities), the inner most regions of the disk can become opaque to neutrinos. This has important consequences for the cooling time scale, since the internal energy cannot escape immediately, but must diffuse out. It is also important for the composition, since a negative radial lepton gradient, akin to that occurring above proto-NSs following core-collapse [291], can be established. Despite the stabilizing influence of differential rotation [292], we have found that the disk can become convectively unstable, which can have

important consequences for the generation of magnetic fields and nucleosynthetic products in a possible outflow.

GR, as already pointed out, not only plays a crucial role in the merger dynamics of compact binaries, but may also be important in determining the evolution and energy output from the accretion disks thus formed. The disk is small enough that a substantial amount of material lies in the vicinity of the marginally stable orbit, where the potential departs significantly from its Newtonian form. The BH itself may be rapidly rotating, depending on the previous binary history, and thus produce inertial frame-dragging. Finally, the emitted neutrinos are subject to light bending effects, which can alter the efficiency for annihilation, and thus the corresponding spectrum. Dynamical simulations of post-merger disk evolution have been performed without [286, 287] and with [288, 289] account of these differences, through the use of a pseudo-potential of the form proposed by PW [254] and Artemova *et al* [293] for rotating black holes. These effects have been taken into account in the computation of the energy deposition rates in the vicinity of the disk [93, 94], [294]–[297], taking realistic configurations derived from compact mergers as input conditions in recent computations [298].

4.2. Physical conditions and relevant processes

These disks are typically compact, with the bulk of the mass residing within 4×10^7 cm of the BH (which contains 3–5 $M_\odot$). They are dense ($10^9 \, \text{g cm}^{-3} \leqslant \rho \leqslant 10^{12} \, \text{g cm}^{-3}$) and hot ($10^9 \, \text{K} \leqslant T \leqslant 10^{11} \, \text{K}$). The temperature is in fact high enough that the nuclei are practically fully photodisintegrated in the inner regions. Neutronization due to the high densities can cause the electron fraction to drop substantially below $Y_e = 1/2$, and capture of $e^\pm$ pairs by free neutrons and protons provides the bulk of the cooling for $\rho \geqslant 10^{10} \, \text{g cm}^{-3}$. The gas is composed essentially of α particles and free nucleons in nuclear statistical equilibrium (NSE), $e^\pm$ pairs (the relative importance of positrons is quite sensitive to the degeneracy parameter $\eta_e = \mu_e/kT$, where μ_e is the electron chemical potential), trapped photons and neutrinos of all species ($e^\pm$ captures and annihilation produce only electron neutrinos, but nucleon–nucleon bremsstrahlung and plasmon decays can produce μ and τ neutrinos as well). Studying the thermal evolution and the corresponding energy release thus requires a detailed equation of state and consideration of weak interaction rates and emission processes. In addition, the background gravitational field is intense enough that relativistic effects can come into play, at least in the inner regions of the flow. Finally, the whole situation can hardly be considered to be in a steady state, since it originates, in most cases, either from the collapse of the Fe-core in a massive star, or from the merger or collision of two compact objects.

We have thus considered an equation of state in which the total pressure is given by

$$P = P_{\text{rad}} + P_{\text{gas}} + P_e + P_\nu, \tag{27}$$

where P_ν is the pressure associated with neutrinos, $P_{\text{rad}} = aT^4/3$, $P_{\text{gas}} = (1 + 3X_{\text{nuc}})\rho kT/(4m_p)$ and X_{nuc} is the mass fraction of photodisintegrated (and ideal) nuclei. As an approximate (but quite accurate) solution to the equations of NSE between free nucleons and helium (we do not consider the creation of iron-like nuclei) we use

$$X_{\text{nuc}} = 22.4 \left(\frac{T}{10^{10} \text{K}} \right)^{9/8} \left(\frac{\rho}{10^{10} \, \text{g cm}^{-3}} \right)^{-3/4} \exp\left(-8.2 \frac{10^{10} \text{K}}{T} \right). \tag{28}$$

The degeneracy parameter of Fermions can vary over a large range in the disk, so it is necessary to use an expression (due to [299]) that allows for it (but under the condition of relativity, which translates to $\rho \geqslant 10^6 \text{ g cm}^{-3}$), namely

$$P_\text{e} = \frac{1}{12\pi(\hbar c)^3} \left[\eta_e^4 + 2\pi^2\eta_e^2(kT)^2 + \frac{7}{15}\pi^4(kT)^4 \right], \tag{29}$$

and

$$\frac{\rho Y_\text{e}}{m_\text{p}} = n_- - n_+ = \frac{1}{3\pi^2(\hbar c)^3} \left[\eta_\text{e}^3 + \eta_e\pi^2(kT)^2 \right] \tag{30}$$

for the electron fraction. This formula reduces to the appropriate limits when the temperature is low ($kT \ll \eta_\text{e}$, implying $P \propto \rho^{4/3}$ for a cold Fermi gas) and when it is high ($kT \gg \eta_\text{e}$, giving $P \propto T^4$ for relativistic $e^\pm$ pairs). Note that this automatically takes into account the presence of pairs in the limit of an ultra relativistic gas, making it unnecessary to alter the factor 1/3 appearing in the term for radiation pressure.

There is an additional effect which alters the equilibrium composition of the fluid, related to the optical depth. We approximate this by following Beloborodov [300] and computing the electron fraction as

$$Y_\text{e} = \frac{1}{2} + 0.487 \left(\frac{Q/2 - \eta_\text{e}}{kT} \right), \tag{31}$$

in the optically thin and mildly degenerate case ($Q = [m_\text{n} - m_\text{p}]c^2$), and

$$\frac{1 - Y_\text{e}}{Y_\text{e}} = \exp\left(\frac{\eta_\text{e} - Q}{kT} \right), \tag{32}$$

in the optically thick case (this follows from setting the neutrino chemical potential equal to zero in the reaction $e + p \leftrightarrow n$ and using Maxwell–Boltzmann statistics for the non-degenerate nucleons). We use an interpolated fit weighted with factors involving $\exp(-\tau_\nu)$ to smoothly transition from one regime to the other.

We have included neutrino energy losses from $e^\pm$ pair captures onto free nucleons by the use of tables [301] and $e^\pm$ pair annihilation from fitting functions [302]. Additionally plasmon decays [233] and nucleon–nucleon bremsstrahlung [303] losses are considered, although their contribution to the total luminosity is negligible. The resulting cooling rates are accurate over a wide range of temperature and density, and are vary consistently with the changing composition and conditions within the flow (particularly for the two most important contributions of $e^\pm$ capture and annihilation).

We approximate the optical depth of the material to neutrinos by computing the cross-section for scattering off free nucleons and α particles heavy nucleons. These provide the dominant contribution, and are given by

$$\sigma_\text{N} = \frac{1}{4}\sigma_0 \left(\frac{E_\nu}{m_\text{e}c^2} \right)^2, \tag{33}$$

and

$$\sigma_\alpha = \sigma_0 \left(\frac{E_\nu}{m_e c^2} \right)^2 [4 \sin^2 \theta_W]^2, \tag{34}$$

where $\sigma_0 = 1.76 \times 10^{-44}$ cm^{-2}, θ_W is the Weinberg angle and E_ν is the energy of the neutrinos (roughly equal to the Fermi energy E_F of the mildly degenerate electrons). The optical depth is estimated as $\tau_\nu = H/l_\nu$, where H is the local disk scale height and l_ν is the neutrino mean free path. We find that in the inner regions, $\tau_\nu \sim 10^2$ in high mass disks.

The cooling is then suppressed by a factor $\exp(-\tau_\nu)$ to mimic the effects of diffusion, so the total neutrino luminosity is

$$L_\nu = \int \rho^{-1} (\dot{q}_{ff} + \dot{q}_{plasmon} + \dot{q}_{pair} + \dot{q}_{cap}) \exp(-\tau_\nu) \, dm. \tag{35}$$

The transition from a transparent to an opaque fluid occurs at $\tau_\nu \approx 1$, which corresponds roughly to $\rho \approx 10^{11}$ g cm^{-3} (this happens largely in the high-mass disks, with their low-mass counterparts remaining mostly transparent). Photodisintegration cools the gas at a rate $\dot{q}_{phot} = 6.8 \times 10^{18} (dX_{nuc}/dt)$ erg s^{-1} cm^{-3}, which is included in the energy equation. Finally, angular momentum is transported through the disk by shear stresses, and we use the α prescription to vary the magnitude of the effective viscosity (all terms in the stress tensor are included in the momentum and energy equations [152]).

Clearly, although we have attempted to give an accurate thermodynamic description of the gas, several approximations remain. In the first place, we have assumed that all reactions considered reach equilibrium. This is quite accurate for photodisintegration, but not necessarily true for weak interactions. Secondly, our prescription for neutrino optical depth effects is obviously a far cry from the proper Boltzmann transport used in 1D SN calculations (or even diffusion), but we believe it qualitatively captures the nature of the transparent–opaque transition, at approximately the correct density. Finally, we have *not* considered energy deposition from neutrinos back to the gas through absorption or incoherent scattering. This could be quite important in the outer regions of the disk (see subsections 2.3 and 4.4) and relevant for the baryon loading of a possible GRB-producing outflow.

4.3. Dynamical evolution in a pseudo-GR potential

Here, we present results for our own 2D calculations in azimuthal symmetry for the evolution of hypercritical accretion flows around a BH following compact object mergers. The initial conditions are essentially the same as those in [287], which were taken from 3D simulations of BH–NS mergers, and are evolved in the pseudo-Newtonian potential of PW [254], which reproduces the existence of a marginally stable orbit at $r_{ms} = 6\, GM_{BH}\, c^2$ for non-spinning BHs. The standard disk mass initially is $M_* = 0.3\, M_\odot$. This is probably close to the highest mass that will be found in such disks, and we have also investigated conditions in lighter disks, with $M_{disk} = 0.1, 0.01\, M_*$. The microphysics included is the same as in our previous calculations (and described above in subsection 4.2) thus allowing for a clear identification of the effects of strong field gravity.

The spatial structure of the disk is shown in figure 16, where the cooling is also indicated. The overall structure is similar to that seen in the Newtonian simulations, but the maximum density is
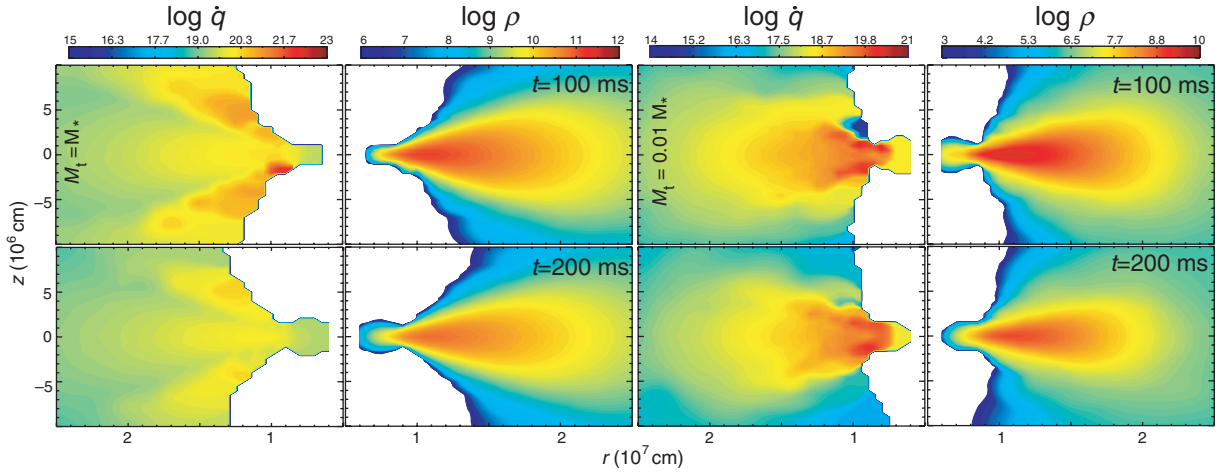
Figure 16. Meridional structure and evolution of hyperaccreting flows around a Schwarszchild BH, computed with the pseudo-Newtonian potential of PW [254]. The density, ρ (g cm^{-3}) and cooling rate through neutrinos $\dot{q}$ (erg g^{-1} s^{-1}) are plotted 100 and 200 ms after the start of the calculation, for disks with an initial mass M_* (left panels) and $0.01\,M_*$ (right panels), where $M_* = 0.3\,M_\odot$. The α viscosity coefficient is 10^{-2}. Note the different bounds for the colour scale between the two cases, in density as well as cooling.

reduced, and only in the high-mass disks and at early times is the presence of an optically thick region (along the equator, and close to the BH) evident. The thermodynamic structure is plotted in figure 17, where the regions in the density–temperature plane occupied by the disk are shown, along with the optical depth. The flow remains almost entirely photo dissociated in the inner regions, thus giving rise to the intense neutrino emission from pair capture onto free nucleons. It is also clear that the electron gas is neither ideal nor fully degenerate, and finite-degeneracy effects must be considered to compute its evolution properly.

In our Newtonian calculations we found that neutronization was important in the inner regions of the flow. This is still the case for the relativistic calculations, although the effect is reduced, particularly for low disk masses. Figure 18 shows one snapshot of height-integrated radial profiles of the electron fraction in the disk for different disk masses. The most massive disk achieves higher densities and neutronization is substantially enhanced. The negative gradient in $Y_e(R)$ at small radii was also observed in Newtonian calculations, where it was a consequence of the transition to the optically thick regime. In the relativistic case it occurs even for low-mass disks, and is related to the appearance of a plunging region, where the radial velocity rapidly increases, once the fluid reaches the marginally stable orbit (see the bottom panel in figure 18). The net effect is also to produce an inversion of the radial lepton gradient, indicating that convection is possible (differential rotation has a stabilizing effect on this, and the full Solberg–Hoiland criterion must be considered, see below).

The accretion rate and neutrino luminosity are plotted as functions of time in figure 19 for $\alpha = 10^{-2}$ and disk masses covering two orders of magnitude. For comparison, the result for high disk mass in the Newtonian case is also shown. Clearly the interval during which a large luminosity is maintained is lower in the relativistic case, simply because the disk drains more rapidly onto the BH. The accretion rates scale roughly linearly with the disk mass, whereas
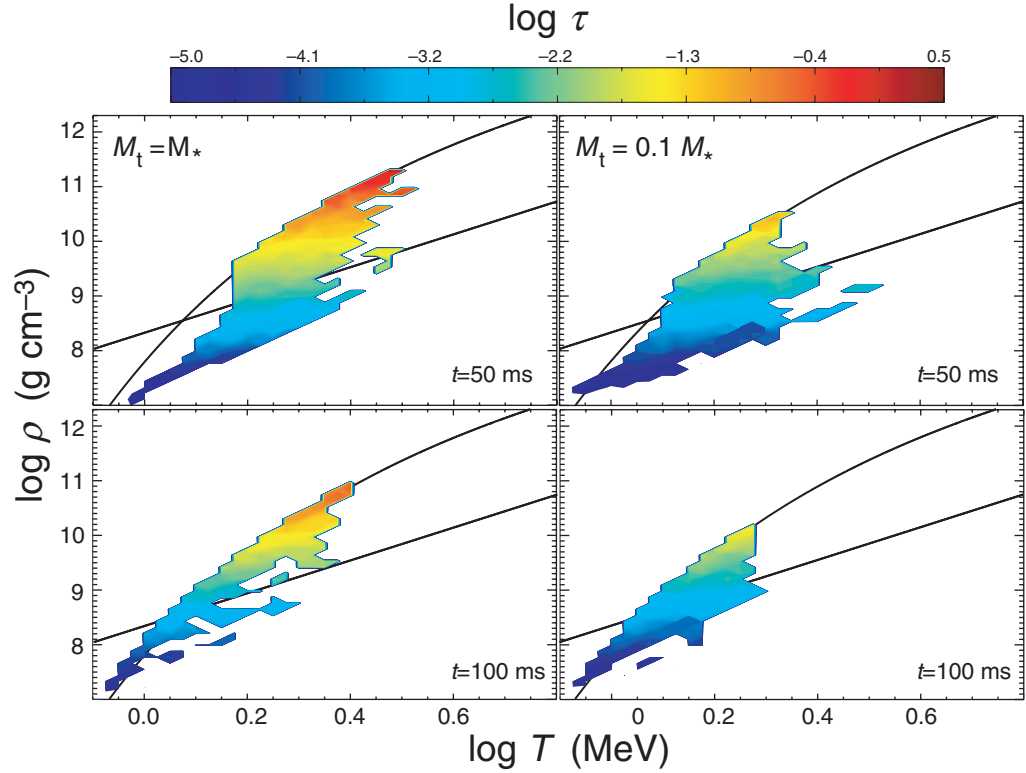
Figure 17. Colour-coded optical depth in the ρ–T plane at 50 and 100 ms for disks with initial mass M_* (left panels) and $0.1\,M_*$ (right panels). The curved solid line marks the photodisintegration threshold between α particles and free nucleons (by mass), and the straight solid line marks the electron degeneracy threshold, given by $kT = 7.7(\rho/10^{11}\,\mathrm{g\,cm^{-3}})^{1/3}$ MeV. As the disk drains into the BH, it becomes more transparent, cooler and less degenerate. The α viscosity coefficient is 10^{-2}.

the power output in neutrinos shows a slightly steeper dependence (it also depends sensitively on the temperature), particularly evident at late times.

One possibility for the driving of a relativistic outflow and the powering of a GRB is $\nu\bar{\nu}$ annihilation, and another is the often-quoted magnetic outflow [148]. If we assume a 1% efficiency for the first case at $L_\nu = 10^{53}\,\mathrm{erg\,s^{-1}}$ (which scales with the square of the neutrino luminosity), and using the present set of calculations, we infer a scaling with disk mass for the total energy release as

$$E_{\nu\bar{\nu}} = 2 \times 10^{48} \left(\frac{M_{\mathrm{disk}}}{0.03 M_\odot}\right)^2 \mathrm{erg}. \qquad (36)$$

This has the same dependence as in the Newtonian case, but with a reduced intensity, by about a factor of five. For the magnetic scenario, the total energy release (assuming the magnetic field energy density is in equipartition with the internal energy density ρc_s^2) scales as

$$E_{\mathrm{BZ}} = 2 \times 10^{49} \left(\frac{M_{\mathrm{disk}}}{0.03 M_\odot}\right) \left(\frac{\alpha}{10^{-1}}\right)^{-0.55} \mathrm{erg}, \qquad (37)$$
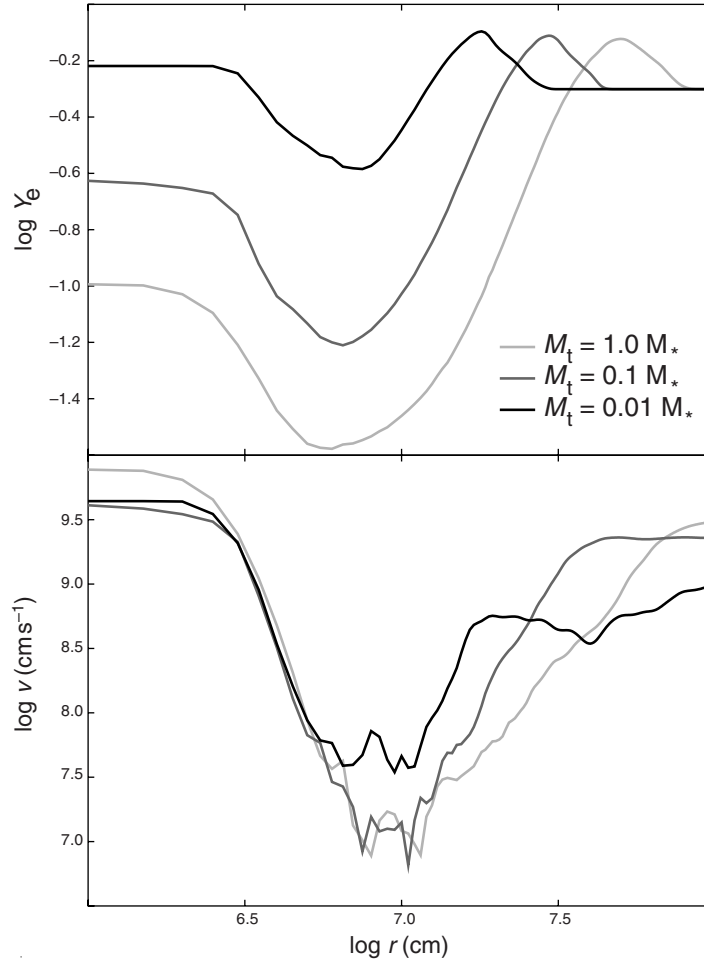
Figure 18. Top: height-integrated radial profiles of the electron fraction Y_e for disks with varying mass in a pseudo-Newtonian relativistic potential. Bottom: meridional velocity magnitude, $(v_r^2 + v_z^2)^{1/2}$, for the same calculation. The marginally stable orbit is located at $\log(R(\mathrm{cm})) \approx 6.6$. In the outer disk, the gas consists mostly of α particles, and $Y_e = 1/2$. As the fluid nears the BH it rises slightly, then becomes smaller as neutronization takes place. The trend is then reversed and Y_e increases again. The minimum is a sensitive function of the disk mass (or equivalently, the density).

again in the same way as in the Newtonian case, but also reduced (this time by about a factor twenty).

A comparison of these scalings and the actual data for the prompt emission of four SGRBs detected in 2005 is shown in figure 20. Clearly neutrinos are probably not the best way to produce a burst lasting more than a few tens of milliseconds, unless the disk mass is quite high. The energy release is somewhat insensitive to the burst duration in this case. On the other hand, magnetic energy extraction seems like a plausible mechanism, and would, at a given disk mass (if this is a standard quantity in any sense), reproduce the trend that longer bursts seem to have a greater fluence. It is also apparent that a reduced disk mass is less of a problem for magnetic mechanisms than for neutrinos as regards the energetics.
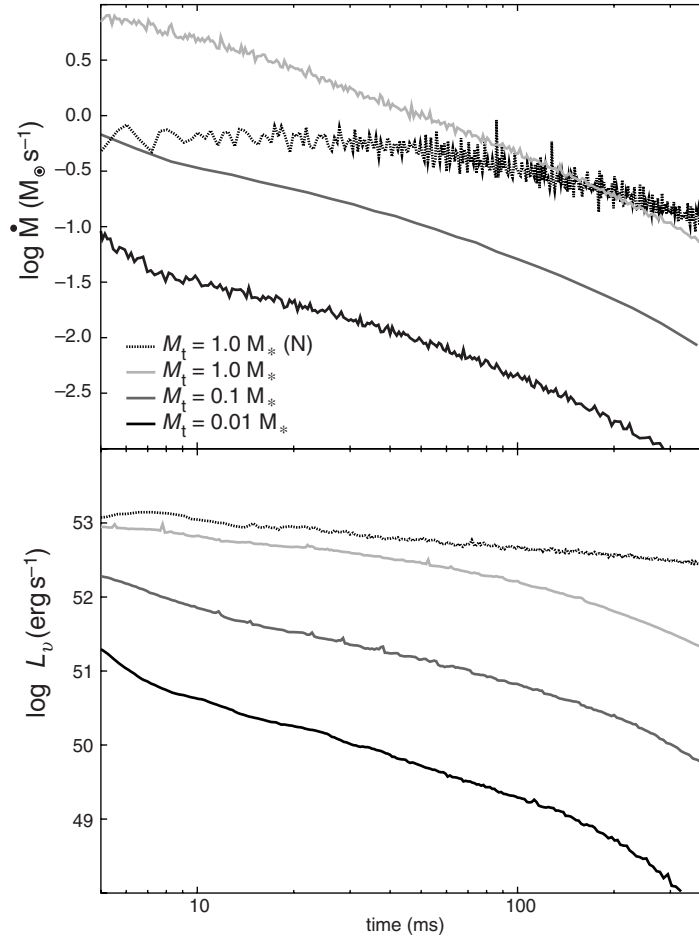
Figure 19. The neutrino luminosity decreases as the disk drains into the BH on a viscous timescale. The top (bottom) panel shows $\dot{M}$ (L_ν) as a function of time for calculations with three different initial disk masses, covering two orders of magnitude ($M_* = 0.3 M_\odot$) in the pseudo-Newtonian potential of PW. For reference, the curve labelled (N) is the result of a calculation evolved in a Newtonian $1/R$ potential. The viscosity parameter is set to $\alpha = 10^{-2}$ in all cases.

4.4. Disk winds and outflows

The outer regions and surface of the accretion flow are subject to various effects that can unbind substantial quantities of mass and drive powerful winds, with important implications for nucleosynthesis [149 307, 308], GRBs and supernovae [309] and the possible collimation of relativistic outflows [143, 306, 310, 311]. One of these is energy deposition by neutrino heating, another is thermonuclear burning and the corresponding energy release. The former has been studied extensively by a number of authors, in the supernova as well as GRB contexts. Dynamical disk evolution calculations do not always include this explicitly (our own simulations fall in this category), but reasonable estimates can nevertheless be derived in some cases. We include the latter in its crudest approximation, considering only the transition from α particles to free nucleons, whose mass fraction is given in equation (28).
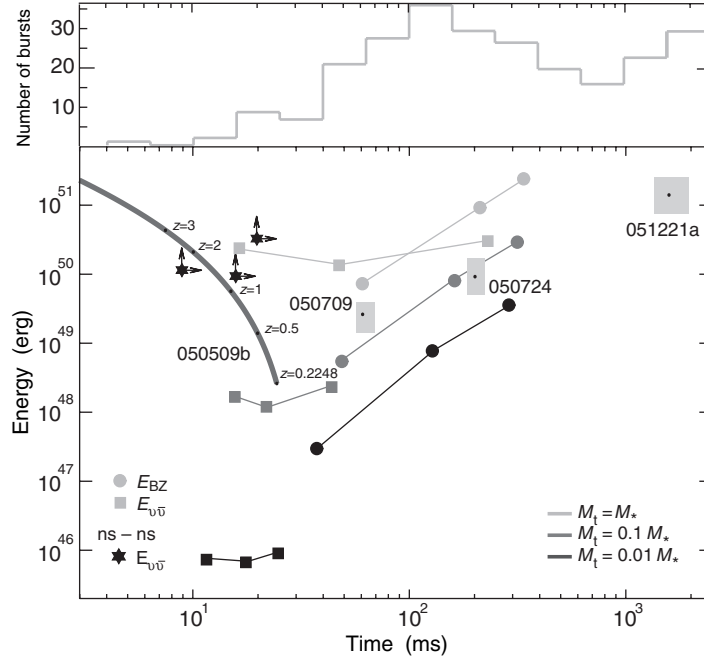
Figure 20. Top panel: histogram of observed burst durations, taken from [304]. Bottom panel: comparison of energy versus duration for GRB959509B (for which the dependence on redshift is indicated by the grey line), GRB050709, GRB050724, and GRB051221A with estimates from compact binary mergers. The connected squares and circles show the total isotropic energy release (assuming collimation into a solid angle $\Omega = 4\pi/10$) and duration (t_{90}) for $\nu\bar{\nu}$ annihilation and Blandford–Znajek powered events, respectively, as computed from our 2D disk evolution calculations in a pseudo-Newtonian potential [254]. The range in initial disk masses covers two orders of magnitude, and the effective viscosities are (from left to right), $\alpha = 10^{-1}, 10^{-2}$ and 10^{-3}. The stars are estimates from $\nu\bar{\nu}$–driven outflows in double NS mergers [305, 306].

Now the gravitational binding energy per nucleon is $E_{gr} \simeq GM_{BH}m_p/R \simeq 5(10^8\,\text{cm}/R)\,\text{MeV}$, and the production of one α particle releases 7.7 MeV/nucleon into the flow, so enough energy is available to produce a large-scale wind where the binding energy becomes sufficiently small. This is in fact seen in our calculations (figure 21), where the outward (mostly vertical) velocity is $v_w \simeq c/10$ and $\dot{M}_{wind} \simeq 2\text{–}5 \times 10^{-2} M_\odot\,\text{s}^{-1}$. The corresponding power is $L_{wind} = \dot{M}_{wind}v_{wind}^2/2 \simeq 1 - 3 \times 10^{50}\,\text{erg s}^{-1}$ for a disk with $M_{disk} \simeq 0.3\,M_\odot$ initially. The mass outflow from a neutrino-driven wind in spherical symmetry was estimated by Qian and Woosley [149] as

$$\dot{M}_{wind} \simeq 5 \times 10^{-4} \left(\frac{L_\nu}{10^{52}\,\text{erg s}^{-1}} \right)^{5/3} M_\odot\,\text{s}^{-1}, \tag{38}$$

so at the disk neutrino luminosity of $L_\nu \simeq 10^{53}\,\text{erg s}^{-1}$, it is interesting to note that the power of each mechanism is comparable (table 4).

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT
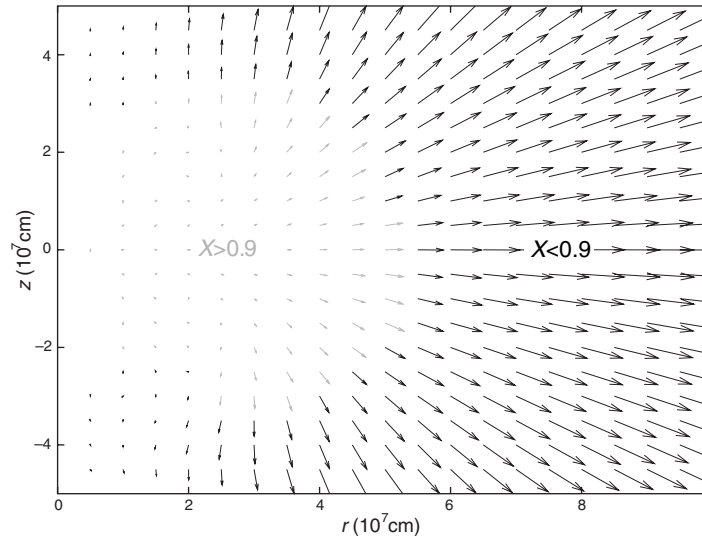
Figure 21. The energy released from He synthesis in the outer regions of the hypercritical accretion flow surrounding the BH is sufficient to unbind the gas and produce a strong wind. The meridional velocity field shows an accelerated outflow in the regions where the mass fraction X of free nucleons drops appreciably (a threshold of 0.9 has been set in the plot), indicating the creation of α particles.

Table 4. Energetics of winds.

α	$\dot{M}_{wind}$ $(10^{-2} M_\odot \, s^{-1})$	ρ $(g \, cm^{-3})$	v_w/c	L_{wind} $(10^{50} \, erg \, s^{-1})$	M_{disk} $(M_\odot)$	Potential[a]
0.100	2.5	10^6	0.08	1.5	0.3	N
0.010	2.5	4×10^5	0.10	2.25	0.3	N
0.010	1.5	4×10^5	0.10	1.30	0.3	PW[a]
0.001	1 0	2×10^5	0.10	0.90	0.3	N
0.100	0.015	4×10^4	0.08	0.009	0.003	N
0.010	0.040	3×10^4	0.05	0.0094	0.003	N
0.010	0.025	3×10^4	0.03	0.009	0.003	PW
0.001	0.020	2×10^4	0.03	0.002	0.003	N

[a] (N): Newtonian; (PW): Paćznski and Wiita [254].

4.5. Stability and convection

An important question which cannot be addressed fully through steady-state calculations is that of stability. Several factors need to be considered; we address each in turn. The 'runaway radial' instability discovered by Abramowicz *et al* [312] applies to configurations with constant specific angular momentum as a function of radius. This is fully an effect of GR, dependent on the existence of an inner Roche lobe for disk accretion, akin to the L_1 Lagrange point in binary stellar evolution. Essentially, if the fluid overflows this lobe, two things occur: (i) the BH mass increases, pushing this critical point outward and (ii) material of the same angular momentum value is now in a position to be accreted. The effect is a runaway, in which the disk is

accreted in a matter of a few dynamical timescales. It turn out that even a small positive gradient of the specific angular momentum, $\ell(R)$, is enough to stabilize this condition, because matter with greater angular momentum will not be easily accreted, and this damps the instability quite rapidly. However, none of the post-merger accretion disks obtained through merger calculations have angular momentum distributions that are even remotely close to being constant, and so this turns out to be most likely irrelevant for the evolution of GRB central engines. Likewise, the self-gravity of the disk is probably unimportant, particularly if one considers the results for disk masses from GR calculations to be more realistic. This has been raised as a possibility concerning the late-time x-ray flares in a few of the SGRBs discovered recently, and will be addressed in the following section.

As for the standard photon-cooled thin disk solution, neutrino cooled flows were shown to be thermally unstable if radiation pressure dominates [281, 282]. Our numerical solutions which fully resolve the vertical as well as radial structure of the disk in time-dependent fashion show this to be the case as well. The difference is that the disks' response to this condition is a substantial thickening in the opaque regions, because of (i) the rise in pressure and (ii) the suppression of cooling. Rather than having a thin disk, the scale height $H = |P/(\mathrm{d}P/\mathrm{d}z)| \simeq R$.

We have also recently found that a convective instability is present in NDAFs at high densities (or equivalently, accretion rates). This is analogous to what occurs above proto–NSs following core collapse [291], and is directly related to the fact that at $\rho \simeq 10^{11}\mathrm{g\,cm^{-3}}$ the fluid becomes opaque to neutrinos (i.e. the equivalent optical depth, τ_ν, is of order unity). A further rise in density is thus accompanied by an increase in entropy per baryon, s_b, and in equilibrium electron fraction Y_e. The classical Solberg–Hoiland requirement for convective stability can be written as two simultaneous conditions (e.g. [292]):

$$\frac{1}{R^3}\frac{\mathrm{d}\ell^2}{\mathrm{d}R} - \left(\frac{\partial T}{\partial P}\right)_\mathrm{s} \nabla P \cdot \nabla s > 0, \tag{39}$$

and

$$-\frac{1}{\rho}\frac{\mathrm{d}P}{\mathrm{d}z}\left[\frac{\mathrm{d}\ell^2}{\mathrm{d}R}\frac{\mathrm{d}s}{\mathrm{d}z} - \frac{\mathrm{d}\ell^2}{\mathrm{d}z}\frac{\mathrm{d}s}{\mathrm{d}R}\right] > 0, \tag{40}$$

where P is the pressure, T is the temperature and s is the specific entropy.

We find that the first of these is satisfied over most of the disk volume in dynamical calculations, and marginally so in the inner, opaque regions. Recall that established convection will erase the conditions which led to its occurrence in a characteristic turnover time $t_\mathrm{con} \simeq l_\mathrm{con}/v_\mathrm{con}$, unless some external condition tries to maintain the instability, in which case (for efficient convection) marginal stability will ensue (see, e.g. [313]). The second condition is clearly not satisfied in the optically thick regime (see figure 22) and drives vigorous motions continuously, as can be seen from inspection of the corresponding velocity field (figure 4 in [287]). Although some circulations are also apparent in the optically thin region, these are of much smaller strength, and are essentially driven by the overshooting of fluid elements out of the convective region and into the outer disk, and their subsequent damping.

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT
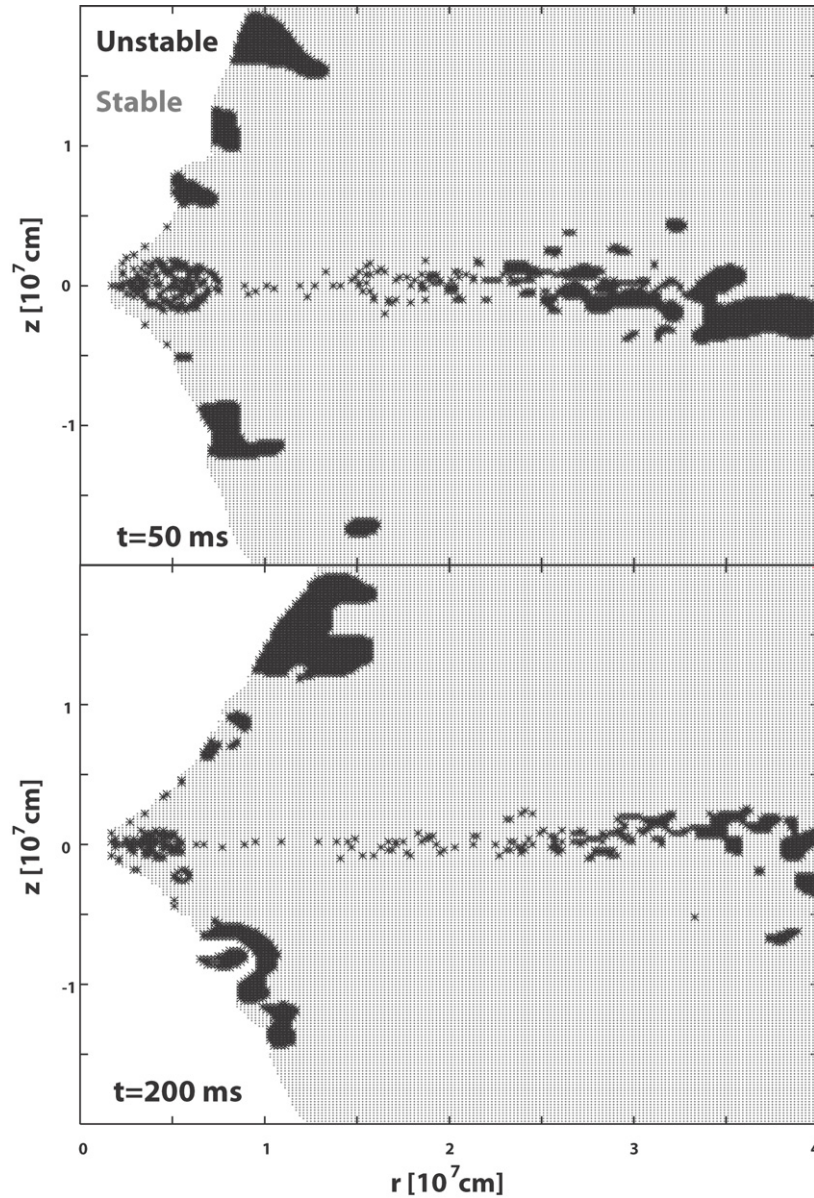
Figure 22. Convective stability of a neutrino-cooled accretion flow around a black hole with $M_{BH} \simeq 4M_\odot$: the dark (light) areas represent regions where the Solberg–Hoiland criterion implies instability (stability). The largest part of the unstable region clearly coincides with the optically thick region of the flow. As the disk drains into the BH, the disk becomes more transparent and this region becomes smaller. The filament-like regions of instability close to the equator are due to the irregularities in the flow. These particular snapshots are taken from run a2M in [287], with a Shakura–Sunyaev viscosity coefficient $\alpha = 0.01$ and $0.3M_\odot$ in the disk initially.

5. Prolonged engine activity

5.1. Motivation

It is possible that the mass and accretion rate shape many of the engine activity's visible manifestations. If we are to distinguish those properties of the activity which can be regulated by the engine itself and those which may varied independently, we must look for guidance from observations of the emitted radiation. A crucial observational development which must be noted is the discovery of late time x-ray flaring in a large number of bursts, both long and short [314, 315]. This has proven difficult to interpret in terms of refreshed shocks, because of their rapid rise and decay. A likely possibility is that it reflects renewed activity (although see [319] for an alternative view) in the central engine itself [316]–[318].

For short bursts, with durations of the order of half a second, these flares occur tens, or even hundreds of seconds after the main burst. There is also independent support that x-ray emission on these time scales is detected when light curves of many bursts are stacked [33, 34]. If the GRB itself is powered by a hypercritical accretion flow around a stellar mass BH, the dynamical time scale is only milliseconds, while the viscous time scale, which is usually related to the burst duration, is at most a few seconds if the effective viscosity is equivalent to $\alpha \sim 10^{-3}$. It is thus in principle a problem to account for a resurgence of activity ten to one hundred viscous timescales (or equivalently, up to ten thousand dynamical timescales) later.

Over the past year, a number of specific suggestions have been made concerning the production of flares, both for long and short events—in some case, suggesting also that their occurrence in both types of events should indicate a common origin: interaction with a binary companion [103, 320]; fragmentation of a rapidly rotating core through non-axisymmetric instabilities [321]; magnetic regulation of the accretion flow [322]; fragmentation of the accretion disk through gravitational instabilities and subsequent accretion [323]; differential rotation in post-merger millisecond pulsars [38, 324] and magnetar-like activity.

Despite all these suggestions, it remains unclear how exactly the regulation (in the case of magnetic mechanisms) or the fragmentation (in the case of self-gravity) would come about. Magnetic regulation was actually addressed by van Putten and Ostriker [325] as a possible way to produce both long and short bursts from the same central engine, depending only on the spin of the BH. Slow-spinning holes would be unable to halt accretion and thus a short burst would ensue. In rapidly spinning holes, the transfer of angular momentum through electromagnetic torques would stop or delay accretion, giving a long event. There are tantalizing clues in numerical simulations indicating that this might actually occur. Krolik *et al* [326] find a large difference in the mass accretion rate on to the BH in MHD simulations depending on the BH spin. They report a calculation with a dimensionless Kerr parameter of $a = 0.998$, in which the accretion rate is strongly suppressed due to electromagnetic stresses, and no stationary state is achieved through the end of the computation (although they note that the magnetic field geometry is fundamentally different from that suggested by van Putten and Ostriker). This effect is absent for lower values of the Kerr parameter. It is not clear how to extrapolate these results to the timescales involved in SGRB flares, however, since they occur at much later times (the calculations by Krolik *et al* span about half a second for a BH of four solar masses).

5.2. *Regulating the accretion flow*

One of the basic unknowns of accretion disk theory is the physical mechanism ultimately responsible for angular momentum transport and energy dissipation in the disk. It is well known that classical hydrodynamical viscosity cannot drive accretion at the rates inferred from observations in almost every astrophysical context where accretion disks are thought to play a crucial role. The usual way to overcome this difficulty is to assume that transport is dominated by some *anomalous* viscous phenomenon, possibly related to collective instabilities in the disk, and to give some adhoc parameterizations for its magnitude.

It has been recently recognized that accretion disks threaded by a weak magnetic field are subject to MHD instabilities (see Balbus and Hawley [327] and references therein), which can induce turbulence in the disk and thereby transport angular momentum, promoting the accretion process. A possible alternative source of transport in cold disks is provided by gravitational instabilities [328]—although the outcome strongly depends on the thermodynamics of the disk. In particular, the fragmentation of a gravitationally unstable disk requires that the disk be able to cool very efficiently [329], with a cooling time

$$t_{\rm cool} < 3\Omega^{-1}, \tag{41}$$

where Ω is the local angular velocity in the disk. Gravitational instabilities would then lead to a self-regulating process: if the disk is initially cold, in the sense that $kT \ll GM/R$, then gravitational instabilities would heat it up on the short dynamical timescale, bringing it toward stability; on the other hand, if the disk is hot enough to begin with, radiative cooling will drive it toward an unstable configuration. As a result of these competing mechanisms, the *switch* associated with the onset of gravitational instabilities will act as a thermostat. For self-regulated disks, a simple relationship holds between the disk aspect ratio H/R and the mass ratio $M_{\rm disk}/M_{\rm BH}$. The disk is in fact marginally stable when its temperature is small enough that

$$M_{\rm disk} > \left(\frac{H}{R}\right) M_{\rm BH}. \tag{42}$$

Of course, this relationship can only hold in the limit where $M_{\rm disk}/M_{\rm BH} \ll 1$.

Disk evolution calculations of neutrino-cooled flows have so far failed to show that the disk is close to instability [282],[286]–[289], but this eventually may be the case when enough time has elapsed to allow the disk to significantly cool. It is not clear how to extrapolate these results to the timescales involved in SGRB flares, since current calculations span at most a few seconds. It remains to be discussed under which conditions the development of a gravitational spiral structure would be indeed able to transport angular momentum in the disk efficiently, hence favouring accretion, and whether this might be able to account for flares. There is an additional concern with the use of a viscous formalizm in self-gravitating accretion disks: Balbus and Papaloizou [330] have shown that in general the energy transport provided by gravitational instabilities contains global terms, associated with wave energy transport, that cannot be directly associated with an effective viscosity. Their relative importance, and the conditions under which they may play a significant role are as yet poorly understood.
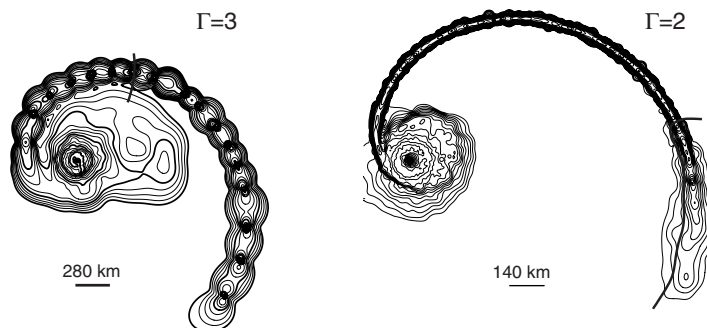
Figure 23. During the merger of a BH and a NS, long one-armed tidal tails of material stripped from the star are formed. Their structure is plotted here for two values of the adiabatic index Γ of the NS fluid, with the scale indicated in each panel. The thick black line across the tail for $\Gamma = 2$ divides material that is bound to the central BH from that which has enough energy to escape the system. Note the condensations at regular intervals in the tail for the stiff equation of state, produced by gravitational instability.

5.3. Flares from tidal tails

In this section, we explore the possibility that the long tidal tails formed during compact object mergers and/or collisions may provide the prolonged mass inflow necessary for the production of late flares in SGRBs. The general idea of extended emission from compact object mergers has been advanced before [11], relying on the gradual disruption of the NS core over a timescale that is much longer than the dynamical one [246]. As we have argued above, we do not believe this will actually occur, but find that the possibility of injecting matter into the central engine at late times remains, in a modified form.

We noted already that large scale tidal tails are a common feature formed during mergers and collisions between compact objects. These are typically a few thousand kilometres in size by the end of the disruption event in the case of NSs and BHs, and a thousand times larger for events involving WDs (see figures 23 and 24). The amount of mass contained in these structures is significant, and typically $M_{\text{tail}} \approx 0.01 - 0.05 M_\odot$. Depending on the details of the equation of state, a small fraction of this ($10^{-3} - 10^{-4} M_\odot$) is actually unbound, and will escape to the surrounding medium. The rest will eventually fall back onto the leftover disk, and probably be accreted onto the central mass.

Once the initial dynamical interaction is over, the fluid in the tails is practically on ballistic trajectories, moving in the potential dominated by the central mass. It is thus possible to compute the rate at which this matter will fall back and accrete onto the existing disk or BH. To compare with the simple case of a supermassive BH mentioned in subsection 3.8, we have also computed the differential distribution of mass with energy for this fluid: $\mathrm{d}M/\mathrm{d}\epsilon$. Although it appears at first glance to be roughly constant, there are in fact significant deviations from this. The corresponding accretion rate thus differs from the $t^{-5/3}$ law derived by Rees [273], and is somewhat shallower, with $\dot{M}_{\text{fb}} \propto t^{-4/3}$. The fall back accretion rate illustrated in figure 26 indicates that the bulk of the matter in the tail returns to the BH in roughly 1s.

Now, clearly one second is much too short to account for flares, which typically occur after 30–100 s. However, the fluid does not directly fall on to the BH, since it has considerable
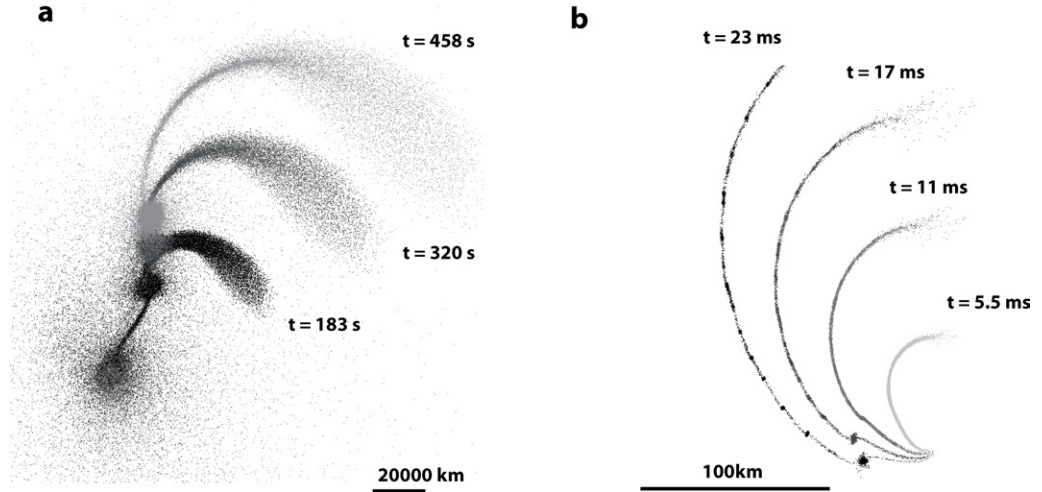
Figure 24. During parabolic collisions of compact objects, similar structures as those found in mergers occur. (a) For a WD–BH collision, a large disk and tail form in a few minutes. The core of the WD will return to the vicinity of the BH in 10^3 s. (b) The collision of a NS with a BH (the initial mass ratio is $q = 0.31$), using the pseudo-potential of PW and [254], reveals the formation of a single large tail, with condensations due to self gravity appearing at late times (see figure 23). Note the different spatial and temporal scales between the two cases.

angular momentum. In fact, its circularization radius is comparable to the size of the disk which was formed around the BH in the first place (typically 200–500 km). This is the point at which the gas will settle if enough of its energy is dissipated, circularizing its orbit. We thus have a situation in which in a short time scale, $t_{fb} \approx 1$ s, a mass $M_{tail} \approx 0.01 M_\odot$ returns to a radius $R_{circ} \approx 5 \times 10^7$ cm. If dissipation occurs (e.g. through shocks with what is left of the initial disk), a new ring of matter may form at $R \approx R_{circ}$. The evolution of this newly injected mass will depend on the previous history of the original accretion disk, and on how much of it is left after the fall back time t_{fb}. Evidently the component with the most energy will dominate the behaviour of the system, and if $M_{disk} \gg M_{tail}$ it is hard to see how the infalling tail could produce a substantial alteration of the overall flow properties and an accompanying observable signal. If the reverse is true and $M_{tail} \gg M_{disk}$, however, perhaps the inner accretion disk and the true accretion rate onto the BH can be modified and a secondary episode of energy release be provoked. If we naively estimate the viscous time scale for the newly formed ring at R_{circ} as

$$t_{visc} \approx R_{circ}^2/10\nu, \qquad (43)$$

where

$$\nu = \alpha c_s^2/\Omega_{Kep} \qquad (44)$$
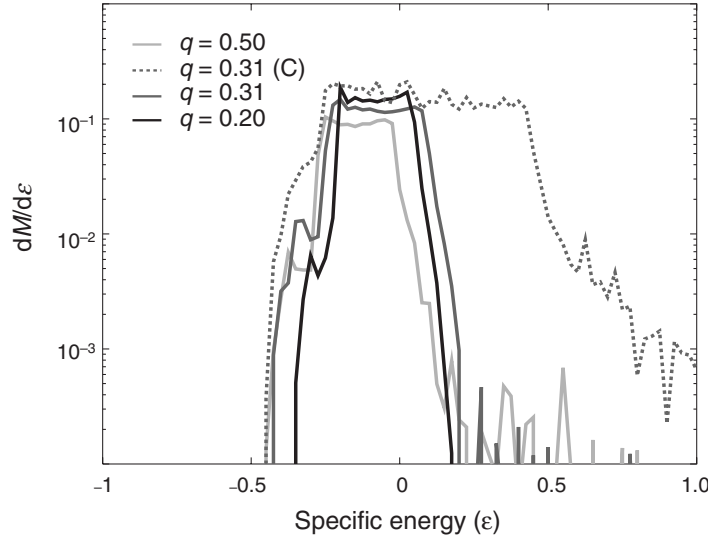
Figure 25. The differential mass distribution, $dM/d\epsilon$ with energy, computed for BH–NS mergers and one parabolic collision (labelled 'C'). Each curve is marked with the initial mass ratio. The material with negative energies will fall back onto the central mass after following essentially ballistic trajectories.

is the viscosity coefficient, c_s is the gas sound speed and Ω_{Kep} is the Keplerian angular frequency, then typical values from the calculations indicate that

$$t_{visc} = 100 \left(\frac{M_{BH}}{4M_\odot}\right)^{1/2} \left(\frac{R_{circ}}{2 \times 10^7 cm}\right)^{1/2} \left(\frac{\alpha}{10^{-2}}\right)^{-1} \left(\frac{c_s}{10^8 cm\,s^{-1}}\right)^{-2} s. \qquad (45)$$

Viscous timescales of $10 - 10^3$ s are thus in principle plausible for $10^{-3} \leqslant \alpha \leqslant 10^{-1}$, and are much longer than the injection interval itself. This is akin to the instantaneous injection of a discrete amount of matter at a given radius, and will produce an accretion episode with a delay proportional to its duration, $t_{delay} \propto t_{acc}$ (e.g. [212]), naturally accounting for this observational fact in SGRB flares.

An additional constraint comes from consideration of the total energy release in the observed flares. Taking neutrino emission as an example, the accretion efficiency is

$$\epsilon_{acc} = L_\nu / \dot{M}_{BH} c^2 \approx 0.1, \qquad (46)$$

based on our own calculations [286, 287] and those reported by other groups [288, 289], where $\dot{M}_{BH}$ is the actual accretion rate that feeds the central BH, and is reduced from the rate at which the disk itself is fed by a factor $\zeta \approx 1/2$, the remainder going into outflows (see also [281]). The accretion efficiency is predominantly dependent upon the flow being optically thin to its own emission, and can thus be extrapolated down to the average accretion rate produced by the fall back material,

$$\dot{M}_{fb} \approx M_{tail}/t_{visc} = 2 \times 10^{-4} M_\odot\,s^{-1}. \qquad (47)$$
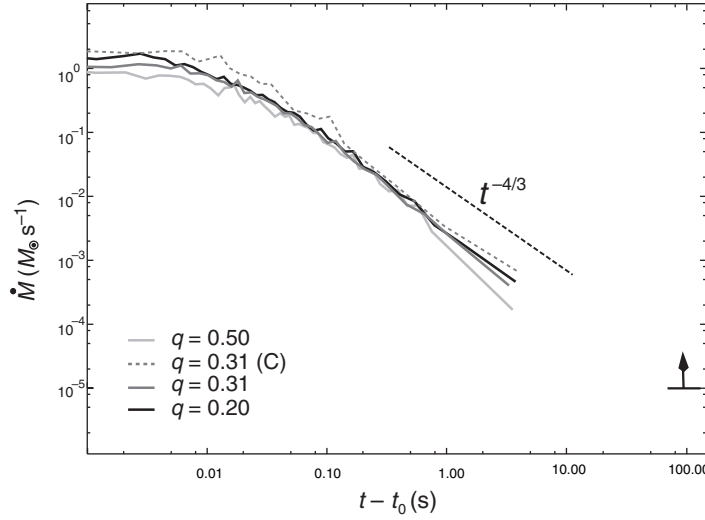
Figure 26. The accretion rate on to the central region from the material in the tails is plotted for the same calculations as shown in figure 25. Initially roughly constant, it rapidly assumes a power law decay, with index $\approx 4/3$. This is roughly independent of the type of interaction (although we note that at late times numerical noise makes an accurate determination of the slope more difficult). The lower limit indicated at 100 s marks the accretion rate required to account for the typical energy seen in a late x-ray flare in a SGRB.

Thus the neutrino luminosity would be

$$L_\nu \approx \epsilon_{\text{acc}} \dot{M}_{\text{BH}} c^2 = 3 \times 10^{49} \text{erg s}^{-1}. \tag{48}$$

Accounting for a total fluence $L_{\text{flare}} \approx 3 \times 10^{46} \text{ erg s}^{-1}$ and a total energy release $E_{\text{flare}} \approx 3 \times 10^{48}$ erg over $t_{\text{flare}} \approx 100$ s (typical numbers for observed events) would require a $\nu\bar{\nu}$ annihilation efficiency of 10^{-3}. This is close to being unrealistically high at this luminosity level, and would thus argue against neutrinos as an ultimate source for the flaring behaviour. On the other hand, magnetic energy extraction could presumably also operate, since the injection can deposit enough mass in the disk to anchor sufficiently strong magnetic fields. It is hard to see, however, how either option could account for extremely delayed flares occurring after 10^3 s, but perhaps those can be due to refreshed shocks [32].

5.4. Instabilities and fragmentation in fluid tails

The hydrodynamic stability of the ejected tail is a question that deserves special comment. The fluid moves globally and to a good approximation in the potential of the central object. It is, nevertheless, prone to an instability due to self-gravity, which may affect its structure on small scales. This is the *varicose* or *sausage* instability, and it can be shown (see chapter XII in [331]) that for a cylinder of *incompressible* fluid, gravitational instability sets in for perturbations with wavelength

$$\lambda \geqslant \lambda^* = 2\pi R_{\text{cyl}}/x^*, \tag{49}$$

where R_{cyl} is the radius of the cylinder and $x^* \approx 1$. The fastest growing mode, which determines the size of the fragments upon manifestation of this effect has $x = 0.58$, corresponding to a wavelength $\lambda = 10.8 R_{cyl}$, and a growth time

$$\tau = \frac{4}{\sqrt{4\pi G\rho}} \tag{50}$$

Binary merger calculations do not employ incompressible fluids, but for high enough adiabatic indices one can see that this behaviour does indeed occur.

The left panel in figure 23 plots the large scale structure formed during a BH-NS merger in a simulation with $\Gamma = 3$. The knots in the tail are spaced at regular intervals, with wavelength $\lambda \approx 2.5 \times 10^7$ cm $\approx 10 R_{tail}$, as predicted by the incompressible analysis. The estimated time scale for their formation is $\tau \approx 20$ ms, in good agreement with what is seen in the calculations. Binary NS mergers with stiff polytropic pressure–density relations performed by Rasio *et al* [227] show exactly the same structures. Once fragmentation occurs, the individual clumps continue to move in the overall gravitational potential, and those that are bound to the central object will return to its vicinity at discrete intervals to inject matter. For lower values of the adiabatic index this behaviour is no longer seen. Instead, the fluid expands smoothly over large volumes, and the tails are less well defined. At the densities encountered in these structures, one would not expect a compressibility low enough to allow this instability to operate, and indeed it is not seen in merger simulations using realistic equations of state [248]. Nevertheless, it is worthwhile to consider their possible formation and subsequent evolution as a general feature. Note that if one substitutes the NS for a quark (strange) star, such droplets are a generic feature upon tidal disruption [332].

6. Summary and future prospects

In this final section, we present a short summary of what we have learned so far about the physics of SGRB sources. Though the field is far from being mature, sufficient progress has been made in identifying some of the essential ingredients. We also describe the observational and theoretical prospects for the near future.

6.1. Summary

The progress in our understanding of SGRBs has been a story of consolidation and integration, and there is every indication that this progression will continue. Until recently, SGRBs were known predominantly as bursts of γ-rays, largely devoid of any observable traces at any other wavelengths. However, a striking development in the last several months has been the measurement and localization of fading x-ray signals from several SGRBs, making possible the optical and radio detection of afterglows, which in turn enabled the identification of host galaxies at cosmological distances. The presence in old stellar populations e.g., of an elliptical galaxy for GRB 050724, rules out a source uniquely associated with recent star formation. In addition, no bright supernova is observed to accompany SGRBs, in distinction from most nearby long-duration GRBs. It is now clear that short and long events are not drawn from the same parent stellar population. Even with a handful of SGRBs detected to date, it has become apparent that they are far from standard, both in their energetics and cosmological niche. This hints at the

underlying possibility that the progenitor itself may be quite different from burst to burst, and not entirely restricted to the most discussed scenario involving the merger of compact binaries such as the Hulse–Taylor pulsar (although see [306]) .

The most fundamental problem posed by SGRB sources is how to generate over 10^{50} erg in the burst nucleus and channel it into collimated plasma jets. The cumulative evidence insistently suggests that the more powerful SGRBs must have 'processed' upwards of $10^{-3} M_\odot$ through a compact entity—only this hypothesis accounts for the high luminosity and compactness inferred from γ-ray variability. We believe that accretion onto a compact object, be it a NS or a stellar mass BH, offer the best hope of understanding the 'prime mover' in all types of SGRB sources although a possible attractive exemption includes a rapidly spinning NS with a powerful magnetic field. Consequently, most theoretical work has been directed towards describing the possible formation channels for these systems, and evaluate those which are likely to produce a viable central engine. The only way to cool the resulting hypercritical accretion flow (other than through direct advective transport of the energy into a gravitational sink hole on a dynamical time scale) is by neutrino emission, circumventing the classical Eddington limit for photons and allowing for the conversion of gravitational binding energy into outflowing radiation. Consideration of the associated energy release thus leads us down the path of detailed thermodynamical and microphysical processes unlike those encountered in most areas of astrophysics, except supernovae, where the physical conditions are quite similar in terms of density, entropy and internal energy.

As of this writing, it is fairly clear from the concerted efforts of many groups working on different aspects of these problems that there is in principle no problem in accounting for the global energy budget of a typical SGRB from the class of systems here considered. The devil is in the details, of course, and the actual modes of energy extraction have yet to be worked out carefully. It would appear, however, that neutrino emission is more confined to be a competitive energy source in the early stages of the dynamical evolution, while magnetically powered events may be able to offer longer staying power.

Still, various alternative ways of triggering the explosions responsible for SGRBs remain: NS–NS, NS–BH, BH–WD or WD–WD binary mergers, recycled magnetars, spun-down supra-massive NS and AIC of a NS. Can we decide between the various alternatives? The progenitors of SGRBs are essentially masked by afterglow emission, largely featureless synchrotron light, which reveals little more than the basic energetics and micro physical parameters of relativistic shocks. In the absence of a supernova-like feature, the interaction of burst ejecta with a stellar binary companion [103] or with its emitted radiation [333] may be the only observable signature in the foreseeable future shedding light on the identity of the progenitor system. A definitive understanding will, however, come with the observations of concurrent gravitational radiation or neutrino signals arising from the dense, opaque central engine.

Our understanding of SGRBs has come a long way since their discovery almost forty years ago, but these enigmatic sources continue to offer major puzzles and challenges. SGRBs provide us with an exciting opportunity to study new regimes of physics. As we have described, our rationalization of the principal physical considerations combines some generally accepted features with some more speculative and controversial ingredients. When confronted with observations, it seems to accommodate their gross features but fails to provide us with a fully predictive theory. What is more valuable, though considerably harder to achieve, is to refine models like the ones advocated here to the point of making quantitative predictions, and to assemble, assess and interpret observations so as to constrain and refute these theories. What we can hope of our present understanding is that it will assist us in this endeavour.

6.2. Observational prospects

High energy astrophysics is a young field. It owns much of the remaining unexplored 'discovery space' in contemporary astronomy. Two examples of this discovery space are at the extremes of observation of the electromagnetic spectrum. At the high end, there are already a few tens of TeV sources, while at the low end of $\leqslant 50$ MHz radio astronomy there are essentially no sources. Neutrino astronomy claims only two cosmic sources so far, the sun and SN1987a. Finally, as many of the most interesting high energy sources are ultimately BHs and NSs, the exciting field of gravitational wave astronomy—perhaps a decade away from birth—is inextricably linked to high energy astrophysics. These are the next frontiers and we consider the role of SGRBs within them in turn.

Atmospheric Cerenkov techniques are being used to detect γ-rays in the GeV–TeV range. These are important as both sources and as probes. Persistent sources have been identified with pulsars, blazars and supernova remnants and in each case are likely to provide the best approach we have to understanding the fundamental nature of these sources. A tentative $\geqslant 0.1$ TeV detection of a long GRB has been reported with the water Cherenkov detector Milagrito [334]. Here the big question is to determine whether the SGRB jets comprise ultrarelativistic protons, that interact with either the radiation field or the background plasma, or if they are $e^{\pm}$ pairs. TeV sources should be far more plentiful in the latter case. The combination of the Gamma Ray Large Area Space Telescope (GLAST) and observatories such as the High Energy Stereoscopic System (HESS) and the Very Energetic Radiation Imaging Telescope Array System (VERITAS) ought to be able to sort this out.

The same shocks which are thought to accelerate the electrons responsible for the non-thermal γ-rays in SGRBs should also accelerate protons [335]–[337]. Both the internal and the external reverse shocks are mildly relativistic, and are expected to lead to relativistic protons [338]. The maximum proton energies achievable in SGRB shocks are $E_{\mathrm{p}} \sim 10^{20}$ eV, comparable to the highest energies measured with large cosmic ray ground arrays [339]. For this, the acceleration time must be shorter than both the radiation or adiabatic loss time and the escape time from the acceleration region [340]. The accelerated protons can interact with the fireball photons, leading to charged pions, muons and neutrinos. For internal shocks producing observed 1 MeV photons this implies $\geqslant 10^{16}$ eV protons, and neutrinos with $\sim 5\%$ of that energy, $\epsilon_{\nu} \geqslant 10^{14}$ eV [341]. Another copious source of target photons in the UV is the afterglow reverse shock, for which the resonance condition requires higher energy protons leading to neutrinos of 10^{17}–10^{19} eV [342]. Whereas photon–pion interactions lead to higher energy neutrinos and provide a direct probe of the shock proton acceleration as well as of the photon density, inelastic proton–neutron collisions may occur even in the absence of shocks, leading to charged pions and neutrinos [343] with lower energies than those from photon–pion interactions. The typical neutrino energies are in the ~ 1–10 GeV range, which could be detectable in coincidence with observed SGRBs. This is the province of projects like AMANDA, IceCube and ANTARES. Success in the former will suggest that ultrarelativistic outflows comprise mainly protons. Neutrino astronomy has the advantage that we can see the universe up to $\sim$EeV energies. By contrast, the universe becomes opaque to γ-rays above $\sim$TeV energies through absorption by the infrared background.

Finally, the last and most challenging frontier is that of gravitational radiation, which is largely unknown territory. A time-integrated luminosity of the order of a solar rest mass ($\sim 10^{54}$ erg) is predicted from merging NS–NS and NS–BH models, while that from collapsar models is less certain, but estimated to be lower. Ground-based facilities, like LIGO, TAMA

and VIRGO, will be seeking such stellar sources. The observation the associated gravitational waves would be facilitated if the mergers involve observed SGRB sources; and conversely, it may be possible to strengthen the case for (or against) NS–NS or NS–BH progenitors of SGRBs if gravitational waves were detected (or not) in coincidence with some bursts. The technical challenge of achieving the sensitivities necessary to measure waves from assured sources should not be understated. It may well take more than another decade to reach them. Neither, however, should the potential rewards. Gravitational waves offer the possibility of observing in an entirely *new* spectrum, not merely another window in the same (electromagnetic) spectrum. Furthermore, they will by their very nature tell us about events where large quantities of mass move in such small regions that they are utterly opaque and forever hidden from direct electromagnetic probing, and are the only way (except perhaps for neutrinos) through which we can learn about them. There have been few regions of the electromagnetic spectrum where observation conformed to previous expectation. So it would be indeed remarkable if gravitational wave astronomy, or any of the other frontiers, turned out to be as we have described. Indeed, if past experience is any guide, they will surely provide us with new surprises, challenges and potential loose threads through which we can unravel another piece of the fabric of the universe.

6.3. Theoretical prospects

Although some of the features now observed in GRB sources (especially afterglows) were anticipated by theoretical discussions, the recent burst of observational discovery has left theory lagging behind. There are, however, some topics on which we do believe that there will be steady work of direct relevance to interpreting observations.

One of the most important is the development and use of hydrodynamical codes for numerical simulation of SGRB sources with detailed physics input. Existing 2D and 3D codes have already uncovered some gas-dynamical properties of relativistic flows unanticipated by analytical models (e.g. [344]), but there are some key questions that they cannot yet address. In particular, higher resolution is needed because even a tiny mass fraction of baryons loading down the outflow severely limits the maximum attainable Lorentz factor. We must wait for useful and affordable 3D simulations before we can understand the nonlinear development of instabilities. Well-resolved 3D simulations are becoming increasingly common and they rarely fail to surprise us. The symmetry-breaking involved in transitioning from 2D to 3D is crucial and can lead to qualitatively new phenomena. A particularly important aspect of this would be to link in a self-consistent manner the flow within the accretion disk to that in the outflowing gas, allowing for feedback between the two components. The key to using simulations productively is to isolate questions that can realistically be addressed and where we do not know what the outcome will be, and then to analyse the simulations so that we can learn what is the correct way to think about the problem and to describe it in terms of elementary principles. Simulations in which the input physics is so circumscribed that they merely illustrate existing prejudice are of limited value!

A second subject ready for a more sophisticated treatment is related to the intensity and shape of the intrinsic spectrum of the emitted radiation. Few would dispute the statement that the photons which bring us all our information about the nature of GRBs are the result of particle acceleration in relativistic shocks (e.g. [345]). Since charged particles radiate only when accelerated, one must attempt to deduce from the spectrum *how* the particles are being accelerated, *why* they are being accelerated, and to identify the macroscopic source driving the microphysical acceleration process.

Collisionless shocks are among the main agents for accelerating ions as well as electrons to high energies whenever sufficient time is available (e.g. [346, 347]). Particles reflected from the shock and from scattering centres behind it in the turbulent compressed region have a good chance of experiencing multiple scattering and acceleration by first-order Fermi acceleration when coming back across the shock into the turbulent upstream region. Second-order or stochastic Fermi acceleration in the broadband turbulence downstream of collisionless shocks will also contribute to acceleration. In addition, ions may be trapped at perpendicular shocks. The trapping is a consequence of the shock and the Lorentz force exerted on the particle by the magnetic and electric fields in the upstream region. With each reflection at the shock the particles gyrate parallel to the motional electric field, picking up energy and surfing along the shock surface. All these mechanisms are still under investigation, but there is evidence that shocks play a most important role in the acceleration of cosmic rays and other particles to very high energies.

There is no *in situ* information available from astrophysical plasmas. So one is forced to refer to indirect methods and analogies with accessible plasmas, found only in near-Earth space. Actually, most of the ideas about and models of the behaviour of astrophysical plasmas have been borrowed from space physics and have been refitted to astrophysical scales. However, the large spatial and long temporal scales in astrophysics and astrophysical observations do not allow for the resolution of the collisionless state of the plasmas. For instance, in the solar wind the collisional mean free path is of the order of a few AU. Looked at from the outside, the heliosphere, the region which is affected by the solar wind, will thus be considered collision dominated over time scales longer than a typical propagation time from the Sun to Jupiter. On any smaller and shorter scales this is wrong, because collisionless processes govern the solar wind here. Similar arguments apply to stellar winds, molecular clouds, pulsar magnetospheres and the hot gas in clusters of galaxies. One should thus be aware of the mere fact of a lack of small-scale observations. Collisionless processes generate anomalous transport coefficients. This helps in deriving a more macroscopic description. However, small-scale genuinely collisionless processes are thereby hidden. This implies that it will be difficult, if not impossible, to infer anything about the real structure, for instance, of collisionless astrophysical shock waves. The reader is refered to [348, 349] for an excellent presentation of the basic kinetic collisionless (space) plasma theory.

The most interesting problem remains, however, in the nature of the central engine and the means of extracting power in a useful collimated form. In all observed cases of relativistic jets, the central object is compact, either a NS or BH, and is accreting matter and angular momentum. In addition, in most systems there is direct or indirect evidence that magnetic fields are present—detected in the synchrotron radiation in galactic and extragalactic radio sources or inferred in collapsing supernova cores from the association of remnants with radio pulsars. This combination of magnetic field and rotation may be very relevant to the production of relativistic jets (e.g. [142]). Much of what we have summarized in this respect is conjecture and revolves largely around different prejudices as to how 3D flows behave in strong gravitational fields. There are serious issues of theory that need to be settled independently of the guidance we obtain from observations of astrophysical BHs.

Acknowledgments

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

References

[1] Nakar E 2006 *Phys. Rep.* submitted
[2] Fishman G and Meegan C 1995 *Annu. Rev. Astron. Astrophys.* **33** 415
[3] Kouveliotou C *et al* 1993 *Astrophys. J.* **413** L101
[4] van Paradijs J, Kouveliotou C and Wijers R A M J 2000 *Annu. Rev. Astron. Astrophys.* **38** 379
[5] Gehrels N *et al* 2005 *Nature* **437** 851
[6] Bloom J S *et al* 2006 *Astrophys. J.* **638** 354
[7] Hjorth J *et al* 2005 *Nature* **437** 859
[8] Covino S *et al* 2006 *Astron. Astrophys.* **447** L5
[9] Berger E *et al* 2005 *Nature* **438** 988
[10] Fox D B *et al* 2005 *Nature* **437** 845
[11] Barthelmy S D *et al* 2005 *Nature* **438** 994
[12] Prochaska J X *et al* 2006 *Astrophys. J.* **642** 989
[13] Villasenor J S *et al* 2005 *Nature* **437** 855
[14] Band D *et al* 1993 *Astrophys. J.* **413** 281
[15] Mallozzi R S *et al* 1995 *Astrophys. J.* **454** 597
[16] Ghirlanda G, Ghisellini G and Celotti A 2004 *Astron. Astrophys.* **422** L55
[17] Kaneko Y *et al* 2006 *Astrophys. J. Suppl.* **166** 298
[18] Bhat P N *et al* 1992 *Nature* **359** 217
[19] Nakar E and Piran T 2002 *Mon. Not. R. Astron. Soc.* **330** 920
[20] Donaghy T Q *et al* 2006 *Astrophys. J.* submitted (*Preprint* astro-ph/0605570)
[21] Bloom J S *et al* 2001 *Astron. J.* **121** 2879
[22] Soderberg A *et al* 2006 *Astrophys. J.* **650** 261
[23] Gorosabel J *et al* 2006 *Astron. Astrophys.* **450** 8
[24] Bloom J S *et al* 2007 *Astrophys. J.* **654** 878
[25] Roming P W A *et al* 2006 *Astrophys. J.* **651** 985
[26] Castro-Tirado A J *et al* 2005 *Astron. Astrophys.* **439** L15
[27] Lee W H, Ramirez-Ruiz E and Granot J 2005 *Astrophys. J.* **630** L165
[28] Grupe D *et al* 2006 *Astrophys. J.* **653** 462
[29] Burrows D N *et al* 2006 *Astrophys. J.* **653** 468
[30] La Parola V *et al* 2006 *Astron. Astrophys.* **454** 753
[31] Watson D *et al* 2006 *Astron. Astrophys.* **454** L123
[32] Panaitescu A 2006 *Mon. Not. R. Astron. Soc.* **367** L42
[33] Lazzati D, Ramirez-Ruiz E and Ghisellini G 2001 *Astron. Astrophys.* **379** L39
[34] Montanari E *et al* 2005 *Astrophys. J.* **625** L17

[35] Rees M J and Mészáros P 1998 *Astrophys. J.* **496** L1
[36] Ramirez-Ruiz E, Merloni A and Rees M J 2001 *Mon. Not. R. Astron. Soc.* **324** 1147
[37] Fan Y and Xu D. 2006 *Mon. Not. R. Astron. Soc.* **372** L19
[38] Dai Z G *et al* 2006 *Science* **311** 1127
[39] Bloom J S and Prochaska J X 2006 *AIP Conf. Ser.* **836** 473
[40] Berger E *et al* 2006 *Astrophys. J.* submitted (*Preprint* astro-ph/0608498)
[41] Berger E 2006 *AIP Conf. Ser.* **836** 33
[42] Pedersen K *et al* 2005 *Astrophys. J.* **634** L17
[43] Hjorth J *et al* 2005 *Astrophys. J.* **630** L117
[44] Christensen L, Hjorth J and Gorosabel J 2004 *Astron. Astrophys.* **425** 913
[45] Trentham N *et al* 2002 *Mon. Not. R. Astron. Soc.* **334** 983
[46] Tanvir N R *et al* 2004 *Mon. Not. R. Astron. Soc.* **352** 1073
[47] Le Floc'h E *et al* 2003 *Astron. Astrophys.* **400** 499
[48] Bornancini C G *et al* 2004 *Astrophys. J.* **614** 84
[49] Hurley K *et al* 2005 *Nature* **434** 1098
[50] Palmer D M *et al* 2005 *Nature* **434** 1107
[51] Fenimore E E *et al* 1996 *Astrophys. J.* **460** 964
[52] Hurley K *et al* 1999 *Nature* **397** 41
[53] Duncan R C and Thompson C 1992 *Astrophys. J.* **392** L9
[54] Nakar E, Gal-Yam A, Piran T and Fox D B 2006 *Astrophys. J.* **640** 849
[55] Lazzati D, Ghirlanda G and Ghisellini G 2005 *Mon. Not. R. Astron. Soc.* **362** L8
[56] Gaensler B M *et al* 2005 *Nature* **434** 1104
[57] Granot J *et al* 2006 *Astrophys. J.* **638** 391
[58] Galama T *et al* 1998 *Nature* **387** 479
[59] Richmond M W *et al* 1996 *Astron. J.* **111** 327
[60] Zeh A, Klose S and Hartmann D H 2004 *Astrophys. J.* **609** 952
[61] Malesani D *et al* 2004 *Astrophys. J.* **609** L5
[62] Hjorth J *et al* 2003 *Nature* **423** 847
[63] Garnavich P M *et al* 2003 *Astrophys. J.* **582** 924
[64] Fynbo J P U *et al* 2006 *Nature* **444** 1047
[65] Guetta D and Piran T 2005 *Astron. Astrophys.* **435** 421
[66] Guetta D and Piran T 2006 *Astron. Astrophys.* **453** 823
[67] Nakar E *et al* 2006 *Astrophys. J.* **650** 281
[68] Gal-Yam A *et al* 2006 *Astrophys. J.* submitted (astro-ph/0509891)
[69] Zheng Z and Ramirez-Ruiz E 2006 *Astrophys. J.* submitted (*Preprint* astro-ph/0601622)
[70] Cavallo G and Rees M J 1978 *Mon. Not. R. Astron. Soc.* **183** 359
[71] Paczynski B 1986 *Astrophys. J.* **308** L43
[72] Goodman J 1986 *Astrophys. J.* **308** L47
[73] Shemi A and Piran T 1990 *Astrophys. J.* **365** L55
[74] Rees M J and Mészáros P 1994 *Astrophys. J.* **430** L93
[75] Sari R and Piran T 1997 *Astrophys. J.* **485** 270
[76] Ramirez-Ruiz E and Lloyd-Ronning N M 2002 *New. Astron.* **7** 197
[77] Kobayashi S, Piran T and Sari R 1997 *Astrophys. J.* **490** 92
[78] Granot J and Sari R 2002 *Astrophys. J.* **568** 820
[79] Sari R, Piran T and Narayan R 1998 *Astrophys. J.* **497** L17
[80] Ramirez-Ruiz E 2005 *Mon. Not. R. Astron. Soc.* **363** L61
[81] Ramirez-Ruiz E and Fenimore E E 2000 *Astrophys. J.* **539** 712
[82] Zel'dovich Ya B 1964 *Sov. Phys.—Dokl.* **9** 195
[83] Salpeter E E 1964 *Astrophys. J.* **140** 796

[84] Ramirez-Ruiz E and Socrates A 2005 *Astrophys. J.* submitted (*Preprint* astro-ph/0504257)
[85] Chevalier R A 1989 *Astrophys. J.* **346** 847
[86] Houck J C and Chevalier R A 1991 *Astrophys. J.* **376** 234
[87] Goodman J, Dar A and Nussinov S 1987 *Astrophys. J.* **314** L51
[88] Eichler D, Livio M, Piran T and Schramm D N 1989 *Nature* **340** 126
[89] Paczyński B 1991 *Acta Astron.* **41** 257
[90] Narayan R, Paczyński B and Piran T 1992 *Astrophys. J.* **395** L83
[91] Mészáros P and Rees M J 1992 *Astrophys. J.* **397** 570
[92] Mochkovitch R, Hernanz M, Isern J and Martin X 1993 *Nature* **361** 236
[93] Jaroszynski M 1993 *Acta Astron.* **43** 183
[94] Jaroszynski M 1996 *Astron. Astrophys.* **305** 839
[95] Hulse R A and Taylor J H 1975 *Astrophys. J.* **195** L51
[96] Lee W H and Kluźniak W 1995 *Acta Astron.* **45** 705
[97] Kluźniak W and Lee W H 1998 *Astrophys. J.* **494** L53
[98] Janka H Th, Eberl T, Ruffert M and Fryer C L 1999 *Astrophys. J.* **527** L39
[99] Fryer C L *et al* 1999 *Astrophys. J.* **520** 650
[100] Katz N and Canel L M 1996 *Astrophys. J.* **471** 915
[101] Levan A J *et al* 2006 *Mon. Not. R. Astron. Soc.* **368** L1
[102] Vietri M and Stella L 1999 *Astrophys. J.* **527** L43
[103] MacFadyen A I, Ramirez-Ruiz E and Zhang W 2006 *Astrophys. J.* submitted (*Preprint* astro-ph/0510192)
[104] Hansen B M and Murali C 1998 *Astrophys. J.* **505** L15
[105] Grindlay J, Portegies Zwart S and McMillan S 2006 *Nature Phys.* **2** 116
[106] Lee W H, Ramirez-Ruiz E and van de Ven G 2006 *Astrophys. J.* submitted
[107] Fryer C L, Woosley S E and Hartmann D H 1999 *Astrophys. J.* **526** 152
[108] Phinney E S 1991 *Astrophys. J.* **380** L17
[109] Bethe H and Brown G E 1998 *Astrophys. J.* **506** 780
[110] Izzard R G *et al* 2004 *Mon. Not. R. Astron. Soc.* **348** 1215
[111] Clark J P A, van den Heuvel E P J and Sutantyo W 1979 *Astron. Astrophys.* **72** 120
[112] Lipunov V M, Postnov K A and Prokhorov M E 1987 *Astron. Astrophys.* **176** L1
[113] Hills D, Bender P L and Webbink R F 1991 *Astrophys. J.* **369** 271
[114] Tutukov A V and Yungelson L R 1993 *Mon. Not. R. Astron. Soc.* **260** 675
[115] Lipunov V M *et al* 1995 *Astron. Astrophys.* **298** 677
[116] Portegies-Zwart S F and Spreeuw H N 1996 *Astron. Astrophys.* **312** 670
[117] Fryer C L, Burrows A and Benz W 1998 *Astrophys. J.* **496** 333
[118] Belczyński K and Bulik T 1999 *Astron. Astrophys.* **346** 91
[119] Bloom J S, Sigurdsson S and Pols O R 1999 *Mon. Not. R. Astron. Soc.* **305** 763
[120] Belczyński K and Kalogera V 2001 *Astrophys. J.* **550** L183
[121] Belczyński K *et al* 2002 *Astrophys. J.* **571** 394
[122] Belczyński K *et al* 2006 *Astrophys. J.* **648** 1110
[123] Habets G M H J 1986 *Astron. Astrophys.* **167** 61
[124] Schutz B 1986 *Nature* **323** 310
[125] Cutler C and Flanagan E E 1994 *Phys. Rev.* D **49** 2658
[126] Spruit H C 1999 *Astron. Astrophys.* **341** L1
[127] Rosswog S, Ramirez-Ruiz E and Davies M B 2003 **345** 1077
[128] Rosswog S and Ramirez-Ruiz E *AIP Conf. Ser.* **727** 361
[129] Price D J and Rosswog S 2006 *Science* **312** 719
[130] Pacini F 1967 *Nature* **216** 567
[131] Gold T 1968 *Nature* **218** 731
[132] Usov V V 1994 *Mon. Not. R. Astron. Soc.* **267** 1035

[133] Thompson C 1994 *Mon. Not. R. Astron. Soc.* **270** 480
[134] Penrose R and Floyd G R 1971 *Nature Phys. Sci.* **229** 177
[135] Ruffini R and Wilson J R 1975 *Phys. Rev.* D **12** 2959
[136] Lee W H and Ramirez-Ruiz E 2006 *Astrophys. J.* **641** 961
[137] Bondi H 1952 *Mon. Not. R. Astron. Soc.* **112** 195
[138] Shakura N I and Sunyaev R A 1973 *Astron. Astrophys.* **24** 337
[139] Ichimaru S 1977 *Astrophys. J.* **214** 840
[140] Narayan R and Yi I 1994 *Astrophys. J.* **428** L13
[141] Abramowicz M A, Chen X, Kato Y, Lasota J P and Regev O 1995 *Astrophys. J.* **438** L37
[142] Blandford R D 2002 *Lighthouses of the Universe: The Most Luminous Celestial Objects and Their Use for Cosmology* vol 381 ed M Gilfanov, R Sunyaev and E Churazov (Berlin: Springer)
[143] Rosswog S and Ramirez-Ruiz E 2002 *Mon. Not. R. Astron. Soc.* **336** L7
[144] Usov V V 1992 *Nature* **357** 472
[145] Mészáros P and Rees M J 1997 *Astrophys. J.* **482** L29
[146] Spruit H C, Daigne F and Drenkhahn G 2001 *Astron. Astrophys.* **369** 694
[147] Lyutikov M and Blandford R 2003 *Preprint* astro-ph/0312347
[148] Blandford R D and Znajek R L 1977 *Mon. Not. R. Astron. Soc.* **179** 433
[149] Qian Y Z and Woosley S E 1996 *Astrophys. J.* **471** 331
[150] Panaitescu A, Kumar P and Narayan R 2001 *Astrophys. J.* **561** L171
[151] Perna R and Belczyński K 2002 *Astrophys. J.* **570** 252
[152] Lee W H and Ramirez-Ruiz E 2002 *Astrophys. J.* **577** 893
[153] Königl A and Granot J 2002 *Astrophys. J.* **574** 134
[154] Rosswog S 2006 *Revista Mexicana de Astronomía y Astrofísica* at press
[155] Stairs I H 2004 *Science* **304** 547
[156] McClintock J E and Remillard R A 2006 *Annu. Rev. Astron. Astrophys.* at press
[157] Weisberg J M and Taylor J H 2003 *ASP Conf. Ser. 302: Radio Pulsars* **302** 93
[158] Stairs I H, Thorsett S E, Taylor J H and Wolszczan A 2002 *Astrophys. J.* **581** 501
[159] Lyne A G 2004 *Science* **303** 1153
[160] Kramer M *et al* 2006 *Science* **314** 97
[161] Tauris T M and van den Heuvel E 2006 *Camb. Astrophys. Ser.* **39**
[162] Nice D J *et al* 2005 *Astrophys. J.* **634** 1242
[163] Bodmer A R 1971 *Phys. Rev.* D **4** 1601
[164] Witten E 1984 *Phys. Rev.* D **30** 272
[165] Fahri E and Jaffe R L 1984 *Phys. Rev.* D **30** 2379
[166] Alcock C, Fahri E and Olinto A 1986 *Astrophys. J.* **310** 261
[167] Jaikumar P, Reddy S and Steiner A W 2006 *Mod. Phys. Lett.* A **21** 1965
[168] Page D and Cumming A 2005 *Astrophys. J.* **635** L157
[169] Cumming A, MacBeth J, in't Zand J J M and Page D 2006 *Astrophys. J.* **646** 429
[170] Heger A, Fryer C L, Woosley S E, Langer N and Hartmann D H 2000 *Astrophys. J.* **591** 288
[171] Rhoades C E and Ruffini R 1974 *Astrophys. J.* **32** 324
[172] Mochkovitch R and Livio M 1989 *Astron. Astrophys.* **209** 111
[173] Mochkovitch R and Livio M 1990 *Astron. Astrophys.* **236** 378
[174] Nomoto K and Kondo Y 1991 *Astrophys. J.* **367** L19
[175] Segretain L, Chabrier G and Mochkovitch R 1997 *Astrophys. J.* **481** 355
[176] Fryer C, Benz W, Herant M and Colgate S A 1999 *Astrophys. J.* **516** 892
[177] King A R, Pringle J E and Wickramasinghe D T 2001 *Mon. Not. R. Ast. Soc.* **320** L45
[178] Dessart L, Burrows A, Ott C D, Livne E, Yoon S Y and Langer N 2006 *Astrophys. J.* **644** 1063
[179] Kidder L E, Will C M and Wiseman A G 1992 *Class. Quantum Grav.* **9** L125
[180] Cutler C *et al* 1993 *Phys. Rev. Lett.* **70** 2984

[181] Blanchet L *et al* 1995 *Phys. Rev. Lett.* **74** 3515
[182] Buonanno A, Chen Y and Vallisneri M 2003 *Phys. Rev.* D **67** 4016
[183] Blanchet L, Damour T, Esposito-Farese G and Iyer B R 2004 *Phys. Rev. Lett.* **93** 1101
[184] Blanchet L, Damour T, Esposito-Farese G and Iyer B R 2004 *Phys. Rev.* D **71** 4004
[185] Pati M E and Will C M 2000 *Phys. Rev* D **62** 4015
[186] Pati M E and Will C M 2002 *Phys. Rev.* D **65** 4008
[187] Will C M 2005 *Phys. Rev.* D **71** 4027
[188] Blanchet L 2006 *Living Reviews in Relativity* online at www.livingreviews.org/lrr-2006-4
[189] Burgay M *et al* 2003 *Nature* **426** 531
[190] Abramowici M *et al* 1992 *Science* **256** 325
[191] Bradaschia C *et al* 1990 *Nucl. Instrum. Methods Phys. Res.* A **289** 518
[192] Kokkotas K and Schmidt B G 1999 *Living Reviews in Relativity* oneline at www.livingreviews.org/lrr-1999-2
[193] Lee W H 2001 *Mon. Not. R. Astron. Soc.* **328** 583
[194] Lee W H 2000 *Mon. Not. R. Astron. Soc.* **318** 606
[195] Faber J A, Grandclément P and Rasio F A 2002 *Phys. Rev. Lett.* **89** 231102
[196] Rees M J 1999 *Astron. Astrophys. Suppl.* **138** 491
[197] Centrella J 2006 *6 Int. LISA Symp. Proc.* at press (AIP) (*Preprint* astro-ph/0609172)
[198] Lattimer J M and Schramm D N 1974 *Astrophys. J.* **192** L145
[199] Lattimer J M and Schramm D N 1976 *Astrophys. J.* **210** 549
[200] Burbidge E M, Burbidge G R, Fowler W A and Hoyle F 1975 *Rev. Mod. Phys.* **29** 547
[201] Wallerstein G *et al* 1997 *Rev. Mod. Phys.* **69** 995
[202] Klebesadel R W, Strong I B and Olson R A 1973 *Astrophys. J.* **182** L85
[203] Wheeler J A 1971 *Pontificae Acad. Sci. Scr. Varia* **35** 539
[204] Mashhoon B 1973 *Astrophys. J.* **181** L65
[205] Glendenning N K 2000 *Compact Stars* (New York: A&A Library Springer)
[206] Chandrasekhar S 1969 *Ellipsoidal Figures of Equilibrium* (New Haven, CT: Yale University Press)
[207] Lai D, Rasio F A and Shapiro S L 1993 *Astrophys. J.* **406** L63
[208] Lai D, Rasio F A and Shapiro S L 1993 *Astrophys. J. Suppl.* **88** 205
[209] Lai D, Rasio F A and Shapiro S L 1994 *Astrophys. J.* **420** 811
[210] Lai D, Rasio F A and Shapiro S L 1994 *Astrophys. J.* **423** 344
[211] Lai D, Rasio F A and Shapiro S L 1994 *Astrophys. J.* **437** 752
[212] Frank J, King A and Raine D 1992 *Accretion Power in Astrophysics* (Cambridge: Cambridge University Press)
[213] Oohara K and Nakamura T 1989 *Prog. Theor. Phys.* **82** 535
[214] Oohara K and Nakamura T 1990 *Prog. Theor. Phys.* **83** 906
[215] Oohara K and Nakamura T 1992 *Prog. Theor. Phys.* **88** 307
[216] Nakamura T and Oohara K 1989 *Prog. Theor. Phys.* **82** 1066
[217] Nakamura T and Oohara K 1991 *Prog. Theor. Phys.* **86** 73
[218] Zhuge X, Centrella J M and McMillan S L W 1994 *Phys. Rev.* D **50** 6247
[219] Zhuge X, Centrella J M and McMillan S L W 1996 *Phys. Rev.* D **54** 7261
[220] Shibata M, Nakamura T and Oohara K 1992 *Prog. Theor. Phys.* **88** 1079
[221] Shibata M, Nakamura T and Oohara K 1993 *Prog. Theor. Phys.* **89** 809
[222] Shibata M 1997 *Phys. Rev.* D **55** 6019
[223] Wilson J R, Mathews G J and Marronetti P 1996 *Phys. Rev.* D **54** 1317
[224] Marronetti P, Mathews G J and Wilson J R 1998 *Phys. Rev.* D **58** 107503
[225] Mathews G J and Wilson J R 2000 *Phys. Rev.* D **61** 127304
[226] Rasio F A and Shapiro S L 1992 *Astrophys. J.* **401** 226
[227] Rasio F A and Shapiro S L 1994 *Astrophys. J.* **432** 242
[228] Rosswog S *et al* 1999 *Astron. Astrophys.* **341** 499

Institute *of* **Physics** Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

[229] Rosswog S, Davies M B, Thielemann F K and Piran T 2000 *Astron. Astrophys.* **360** 171
[230] Rosswog S and Davies M B 2002 *Mon. Not. R. Astron. Soc.* **334** 481
[231] Davies M B, Benz W, Piran T and Thielemann F K 1994 *Astrophys. J.* **431** 742
[232] Janka H T and Ruffert M 1996 *Astron. Astrophys.* **307** L33
[233] Ruffert M, Janka H Th and Schafer G 1996 *Astron. Astrophys.* **311** 532
[234] Ruffert M, Janka H Th, Takahashi K and Schafer G 1997 *Astron. Astrophys.* **319** 122
[235] Ruffert M and Janka H Th 2001 *Astron. Astrophys.* **380** 544
[236] Rosswog S and Liebendorfer M 2003 *Mon. Not. R. Astron. Soc.* **342** 673
[237] Lee W H and Kluźniak W 1999 *Astrophys. J.* **526** 178
[238] Lee W H and Kluźniak W 1999 *Mon. Not. R. Astron. Soc.* **308** 780
[239] Rasio F A and Shapiro S L 1995 *Astrophys. J.* **438** 887
[240] Duez M, Baumgarte T W, Shapiro S L, Shibata M and Uryu K 2002 *Phys. Rev.* D **65** 024016
[241] Bildsten L and Cutler C 1992 *Astrophys. J.* **400** 175
[242] Kochanek C 1992 *Astrophys. J.* **398** 234
[243] Faber J A and Rasio F A 2002 *Phys. Rev* D **65** 084042
[244] Benz W, Bowers R L, Cameron A G W and Press W H 1990 *Astrophys. J.* **348** 647
[245] Freiburghaus C, Rosswog S and Thielemann F-K 1999 *Astrophys. J.* **525** L121
[246] Davies M B, Levan A J and King A R 2005 *Mon. Not. R. Astron. Soc.* **356** 54
[247] Rosswog S, Speith R and Wynn G A 2004 *Mon. Not. R. Astron. Soc.* **351** 1121
[248] Rosswog S 2005 *Astrophys. J.* **634** 1202
[249] Faber J A *et al* 2006 *Phys. Rev.* D **73** 4012
[250] Faber J A *et al* 2006 *Astrophys. J.* **641** L93
[251] Miller M C 2005 *Astrophys. J.* **626** L41
[252] Prakash M, Ratkovic S and Lattimer J M 2004 *J. Phys. G: Nucl. Part. Phys.* **30** S1279
[253] Rasio F A et al 2005 *Proc. JGRG14 Yukawa Institute for Theoretical Physics* (*preprint* astro-ph/0503007)
[254] Paczynsky B and Wiita P J 1980 *Astron. Astrophys.* **88** 23
[255] Shibata M 1999 *Phys. Rev.* D **60** 104052
[256] Shibata M and Uryu K 2000 *Phys. Rev.* D **61** 064001
[257] Shibata M, Taniguchi K and Uryu K 2003 *Phys. Rev.* D **68** 084020
[258] Shibata M, Taniguchi K and Uryu K 2005 *Phys. Rev.* D **71** 084021
[259] Shibata M and Taniguchi K 2006 *Phys. Rev.* D **73** 064027
[260] Oechslin R and Janka H Th 2006 *Mon. Not. R. Astron. Soc.* **368** 1489
[261] Lattimer J M and Swesty F D 1991 *Nucl. Phys.* A **535** 331
[262] Shen H *et al* 1998 *Nucl. Phys.* A **637** 435
[263] Akmal A, Pandharipande D R and Ravenhall V G 1998 *Phys. Rev.* C **58** 1804
[264] Cook G B, Shapiro S L and Teukolsky S A 1994 *Astrophys. J.* **424** 823
[265] Baumgarte T W, Shapiro S L and Shibata M 2000 *Astrophys. J.* **528** L28
[266] Stella L and Vietri M 1998 *Astrophys. J.* **507** L45
[267] Duez M D, Liu Y T, Shapiro S L, Shibata M and Stephens B C 2006 *Phys. Rev. Lett.* **96** 03101
[268] Shibata M, Duez M D, Liu Y T, Shapiro S L and Stephens B C 2006 *Astrophys. J.* **96** 031102
[269] Duez M D, Liu Y T, Shapiro S L, Shibata M and Stephens B C 2006 *Phys. Rev.* D **73** 104015
[270] Frank J 1978 *Mon. R. Not. Astron. Soc.* **184** 87
[271] Lacy J H, Townes C H and Hollenbach D J 1982 *Astrophys. J.* **262** 120
[272] Carter B and Luminet J P 1983 *Astron. Astrophys.* **121** 97
[273] Rees M J 1988 *Nature* **333** 523
[274] Evans C R and Kochanek C S 1989 *Astrophys. J.* **346** L13
[275] Laguna P, Miller W A, Zurek W H and Davies M B 1993 *Astrophys. J.* **410** L83
[276] Ayal S, Livio M and Piran T 2000 *Astrophys. J.* **545** 772
[277] Michel F C 1988 *Nature* **333** 644

[278] MacFadyen A I and Woosley S E 1999 *Astrophys. J.* **524** 262
[279] Woosley S E 1993 *Astrophys. J.* **405** 273
[280] Popham R, Woosley S E and Fryer C L 1999 *Astrophys. J.* **518** 356
[281] Narayan R, Piran T and Kumar P 2001 *Astrophys. J.* **557** 949
[282] Kohri K and Mineshige S 2002 *Astrophys. J.* **577** 311
[283] DiMatteo T, Perna R and Narayan R 2002 *Astrophys. J.* **579** 706
[284] Chen W X and Beloborodov A M 2006 *Astrophys. J.* (*Preprint* astro-ph/0607145) at press
[285] Janiuk A, Perna R, DiMatteo T and Czerny B 2004 *Mon. Not. R. Astron. Soc.* **355** 950
[286] Lee W H, Ramirez-Ruiz E and Page D 2004 *Astrophys. J.* **608** L5
[287] Lee W H, Ramirez-Ruiz E and Page D 2005 *Astrophys. J.* **632** 421
[288] Setiawan S, Ruffert M and Janka H Th 2004 *Mon. Not. R. Astron. Soc.* **352** 753
[289] Setiawan S, Ruffert M and Janka H Th 2006 *Astron. Astrophys.* **458** 553
[290] Ruffert M and Janka H Th 1999 *Astron. Astrophys.* **344** 573
[291] Janka H Th and Müller E 1996 *Astron. Astrophys.* **306** 167
[292] Tassoul J L 1978 *Theory of Rotating Stars* (Princeton, NJ: Princeton University Press)
[293] Artemova I V, Björnsson G and Novikov I D 1996 *Astrophys. J.* **461** 565
[294] Salmonson J D and Wilson G R 1999 *Astrophys. J.* **517** 859
[295] Asano K and Fukuyama T 2000 *Astrophys. J.* **531** 949
[296] Asano K and Fukuyama T 2001 *Astrophys. J.* **546** 1019
[297] Miller W A, George N D, Kheyfets A and McGhee J M 2003 *Astrophys. J.* **583** 833
[298] Birkl R, Aloy M A, Janka H Th and Mueller E 2007 *Astron. Astrophys.* in press (*Preprint* astro-ph/0608543)
[299] Blinnikov S I, Dunina–Barkovskaya N V and Nadyozhin D K 1996 *Astrophys. J. Suppl.* **106** 171
[300] Beloborodov A M 2003 *Astrophys. J.* **588** 931
[301] Langanke J and Martínez–Pinedo G 2001 *At. Data Nucl. Data. Tables* **79** 1
[302] Itoh N, Hayashi S, Nishikawa A and Kohyama Y 1996 *Astrophys. J. Suppl.* **102** 411
[303] Hannestad S and Raffelt G 1998 *Astrophys. J.* **507** 339
[304] Paciesas W S 1999 *Astrophys. J. Suppl.* **122** 465
[305] Rosswog S and Liebendorfer M 2003 *Mon. Not. R. Astron. Soc.* **342** 673
[306] Rosswog S and Ramirez-Ruiz E 2003 *Mon. Not. R. Astron. Soc.* **343** L36
[307] Pruet J, Woosley S E and Hoffman R D 2003 *Astrophys. J.* **586** 1254
[308] Pruet J, Thompson T A and Hoffman R D 2004 *Astrophys. J.* **606** 1006
[309] Kohri K, Narayan R and Piran T 2005 *Astrophys. J.* **629** 341
[310] Levinson A and Eichler D 2000 *Phys. Rev. Lett.* **85** 236
[311] Aloy M A, Janka H Th and Müller E 2005 *Astron. Astrophys.* **436** 273
[312] Abramowicz M A, Calvani M and Nobili L 1983 *Nature* **302** 597
[313] Quataert E and Gruzinov A 2000 *Astrophys. J.* **539** 809
[314] Burrows D N *et al* 2005 *Science* **309** 1833
[315] Nousek J A *et al* 2006 *Astrophys. J.* **642** 389
[316] Giannios D 2006 *Astron. Astrophys.* **455** L5
[317] Ramirez-Ruiz E and Merloni A 2001 *Mon. Not. R. Astron. Soc.* **320** L25
[318] Zhang B 2006 *AIP Conf. Ser.* **836** 392
[319] Liang E W *et al* 2006 *Astrophys. J.* **646** 351
[320] Dermer C D and Atoyan A 2006 *Astrophys. J.* **643** L13
[321] King A R *et al* 2005 *Astrophys. J.* **630** L113
[322] Proga D and Zhang B 2006 *Mon. Not. R. Astron. Soc.* **370** L61
[323] Perna R, Armitage P J and Zhang B 2005 *Astrophys. J.* **636** L29
[324] Gao W H and Fan Y Z 2006 *Chin. J. Astron. Astrophys.* **6** 513
[325] van Putten M H P M and Ostriker E C 2001 *Astrophys. J.* **552** L31
[326] Krolik J H, Hawley J F and Hirose S 2005 *Astrophys. J.* **622** 1008

[327] Balbus S A and Hawley J F 1998 *Rev. Mod. Phys.* **70** 1
[328] Lin D N C and Pringle J E 1987 *Mon. Not. R. Astron. Soc.* **225** 607
[329] Gammie C F 2001 *Astrophys. J.* **553** 174
[330] Balbus S A and Papaloizou J C B 1999 *Astrophys. J.* **521** 650
[331] Chandrasekhar S 1961 *Hydrodynamic and Hydromagnetic Stability* (Oxford: Oxford university Press)
[332] Lee W H, Kluźniak W and Nix J 2001 *Acta Astron.* **51** 331
[333] Ramirez-Ruiz E 2004 *Mon. Not. R. Astron. Soc.* **349** L38
[334] Atkins R *et al* 2000 *Astrophys. J.* **533** L119
[335] Waxman E 2004 *New J. Phys.* **6** 140
[336] Dermer C 2005 *Nuovo Cimento* **28** 789
[337] Waxman E 2006 *AIP Conf. Ser.* **836** 589
[338] Waxman E 2004 *Astrophys. J.* **606** 988
[339] Hayashida N *et al* 1999 *Astrophys. J.* **522** 225
[340] Waxman E 1995 *Phys. Rev. Lett.* **75** 386
[341] Waxman E and Bahcall J N 1997 *Phys. Rev. Lett.* **78** 2292
[342] Waxman E and Bahcall J N 1999 *Phys. Rev.* D **59** 023002
[343] Derishev E V, Kocharovsky V V and Kocharovsky Vl V 1999 *Astrophys. J.* **521** 640
[344] McKinney J C 2006 *Mon. Not. R. Astron. Soc.* **368** 1561
[345] Granot J. 2006 *Revista Mexicana de Astronomia y Astrofisica* at press
[346] Blandford R and Eichler D 1987 *Phys. Rep.* **154** 1
[347] Achterberg A *et al* 2001 *Mon. Not. R. Astron. Soc.* **328** 393
[348] Baumjohann W and Treumann R A 1996 *Basic Space Plasma Physics* (London: Imperial College Press)
[349] Treumann R A and Baumjohann W 1997 *Advanced Space Plasma Physics* (London: Imperial College Press)