10 Progenitor

The progenitor of a GRB is the astrophysical system that precedes the GRB, whose catastrophic self-destruction gives rise to the burst. By definition, when one observes a GRB, the progenitor system has already been destroyed. So it is extremely difficult to directly observe the progenitor system of a GRB.[1] As a result, inference of the progenitor system of a GRB is indirect, and is built upon indirect observational evidence, logical reasoning, and theoretical modeling. In this chapter, various progenitor systems for different types of GRBs are discussed. §10.1 is a general discussion on the observational constraints on the progenitor models. The massive star GRBs and compact star GRBs are discussed in §10.2 and §10.3, respectively. Some other possible progenitor systems are discussed in §10.4.

10.1 General Observational Constraints

Any progenitor system of a GRB has to satisfy the following observational constraints:

- Energetics: Observations show that the collimation-corrected γ-ray emission energy of GRBs is $E_\gamma \sim 10^{51}$ erg (generally in the range of 10^{49}–10^{52} erg). The GRB progenitor therefore must lead to a catastrophic event with an energy of this order.
- Variability time scale: The observed variability time scale δt can be as short as milliseconds. The size of the central engine then has to be smaller than $c\delta t \sim 3 \times 10^7$ cm. This points towards a stellar-size compact object (black hole or neutron/quark star) as the central engine. One requires that the progenitor leaves behind a compact object after the catastrophic event. Therefore, the progenitor system must be of stellar scale.
- Collimation: Various arguments suggest that GRBs are collimated. This requires that the progenitor system has the capability of launching a collimated jet.

Broad-band GRB data collected during past decades, especially in the afterglow era since 1997, have led to the identification of at least two broad types of progenitor systems:

[1] Similar situations also apply to supernovae (SNe). However, since SNe have a much higher event rate density, they can be observed in the nearby universe. After a SN, one may go back to the pre-explosion images to identify the progenitor star of the SN. This was done for some nearby SNe (e.g. SN 1987A). The results greatly enriched our understanding of the stellar explosion physics. Unfortunately, GRBs are too rare and too distant to allow such a direct observation of the progenitor star.

those invoking deaths of massive stars (i.e. *massive star GRBs* or Type II GRBs) and those not invoking massive stars but probably invoking compact stars (i.e. *compact star GRBs* or Type I GRBs) (see §2.7 for a detailed discussion on the physical classification schemes of GRBs). The former typically have durations longer than 2 seconds, whereas the latter typically have durations shorter than 2 seconds. Within each category, one may envisage some sub-categories of the progenitor systems. Below are some examples:

- Massive star GRBs (Type II):
  - Collapse of rapidly spinning single Wolf–Rayet stars;
  - Mergers of binaries containing one massive star (e.g. helium star–black hole mergers or helium star–neutron star mergers);
  - Supernova-induced gravitational collapse (IGS) of a NS in a close CO star–NS binary system;
  - Possible blue supergiants for ultra-long GRBs?
  - Population III stars?
- Compact star GRBs (Type I):
  - NS–NS mergers;
  - BH–NS mergers;
  - NS–NS and NS–BH collisions;
  - BH–WD mergers;
  - Accretion-induced NS collapse (AIC);
  - BH–BH mergers?

10.2 Massive Star GRBs

10.2.1 Observational Evidence

Evidence that most long GRBs are related to deaths of massive stars is overwhelming (Chapter 2 for details). We summarize the key observational evidence in the following:

- A handful of long GRBs are associated with spectroscopically identified SNe of Type Ic;
- Many more long GRBs have a SN red bump in their optical lightcurves about a week after the trigger, which is consistent with the lightcurve template (with stretching) of the Type Ic SN 1998bw associated with GRB 980425;
- Most long GRB host galaxies are star-forming irregular galaxies;
- The location of afterglow tracks the brightest regions in the host galaxy.

All these clues point towards core-collapsing massive stars as the progenitors of this category of GRBs.

10.2.2 Free-Fall Time Scale and Burst Duration

Some theoretical considerations provide clues to a massive star progenitor for long GRBs.

The leading central engine model of GRBs invokes a hyper-accreting black hole. Within this scenario, the duration of the central engine activity should be at least the free-fall time scale of the star. The idea is that, when the massive star core suddenly loses pressure support, the envelope falls to feed the black hole, launch a jet, and power GRB emission.

The typical T_{90} of long GRBs is several tens of seconds (Preece et al., 2000). Including the central engine activity time scale of X-ray flares, the more generally defined burst duration, i.e. $t_{\rm burst}$ (Zhang et al., 2014), peaks around several hundred seconds.

For a massive star with mass M and radius R, a characteristic velocity (free-fall velocity or Keplerian velocity) has the order of magnitude

$$v \sim \left(\frac{GM}{R}\right)^{1/2}. \tag{10.1}$$

The free-fall time scale $t_{\rm ff}$ is therefore roughly

$$t_{\rm ff} \sim \frac{R}{v} \sim \frac{R}{\left(\frac{GM}{R}\right)^{1/2}} \sim \frac{1}{\left(\frac{GM}{R^3}\right)^{1/2}} \sim \frac{1}{(G\rho)^{1/2}}. \tag{10.2}$$

A more exact solution gives

$$t_{\rm ff} = \left(\frac{3\pi}{32G\bar{\rho}}\right)^{1/2} \sim 180\ {\rm s}\ \left(\frac{\bar{\rho}}{100\ {\rm g\ cm^{-3}}}\right)^{-1/2}, \tag{10.3}$$

which is consistent with the typical duration of long-duration GRBs given a typical core density of $\sim$100 g cm^{-3} for massive stars.

In comparison, a compact star has a much higher density ($\rho \sim 10^{14}$ g cm^{-3} for a neutron star and $\rho \sim 10^{6}$ g cm^{-3} for a white dwarf). The corresponding free-fall time scale is

$$t_{\rm ff} \sim 2 \times 10^{-4}\ {\rm s}\ \left(\frac{\bar{\rho}}{10^{14}\ {\rm g\ cm^{-3}}}\right)^{-1/2} \sim 2\ {\rm s}\ \left(\frac{\bar{\rho}}{10^{6}\ {\rm g\ cm^{-3}}}\right)^{-1/2}. \tag{10.4}$$

As a result, compact stars are top candidates for interpreting the short-duration GRBs.

The above estimate of the central engine duration only applies to the black hole central engine, with accretion being the ultimate power of GRB jets. An alternative possibility for both core collapse and NS–NS mergers is that the engine may be a rapidly rotating, highly magnetized neutron star or millisecond magnetar (see Chapter 11 for a more detailed discussion). Within this scenario, the duration of the burst is defined by the spindown or magnetic activity time scales of the magnetar, so that the above estimate is not relevant. For such an engine, NS–NS mergers may also give rise to long-duration events. The extended emission or the internal plateau as observed in some short GRBs might be the manifestation of a magnetar engine in these events.

10.2.3 Wolf–Rayet Stars

The top progenitor star candidate for long GRBs is a Wolf–Rayet (WR) star (Woosley, 1993; MacFadyen and Woosley, 1999; Zhang et al., 2003b).

WR stars (e.g. Crowther, 2007) are evolved, massive stars, which have at least $20M_\odot$ mass at birth but have lost, and are still losing at the time of explosion, mass rapidly by means of a strong stellar wind. The observed WR stars in the Milky Way Galaxy have a mass

$$M_{\rm WR} \geq 10M_\odot \sim 2 \times 10^{34}\ {\rm g}, \tag{10.5}$$

and a radius typically

$$R_{\rm WR} \sim (1\text{–}20)R_\odot \sim (10^{11}\text{–}10^{12})\ {\rm cm}, \tag{10.6}$$

so that the typical density of a WR star is

$$\rho_{\rm WR} \sim 9.5\ {\rm g\ cm^{-3}} \left(\frac{M}{20M_\odot}\right) R_{11}^{-3}, \tag{10.7}$$

and the free-fall time is

$$t_{\rm ff,WR} \sim 680\ {\rm s} \left(\frac{M}{20M_\odot}\right)^{-1/2} R_{11}^{3/2}. \tag{10.8}$$

This time scale is consistent with the typical $t_{\rm burst}$ (duration of GRB internal emission including both γ-rays and X-ray flares) of long GRBs (Zhang et al., 2014).

Type Ic SN Associations

Even though in the original proposal of Woosley (1993) GRBs are not expected to be associated with a bright supernova ("failed" supernova in Woosley's notation), the strongest argument in favor of the long GRB progenitor being WR stars is that several long GRBs are clearly associated with Type Ic SNe. These SNe do not have H and He lines in the spectra, suggesting that both H and He are likely depleted in the progenitor atmosphere. The most straightforward interpretation is that both the H and He envelopes have been stripped by a stellar wind prior to explosion. WR stars are the best candidates to do so. Further modeling of the "collapsar" model of long GRBs (MacFadyen and Woosley, 1999) suggests that associations with SNe are not only possible, but probably also ubiquitous (Woosley and Bloom, 2006).

Rapid Spin vs. Low Metallicity

One critical requirement to launch a relativistic jet is high angular momentum at the stellar core. MacFadyen and Woosley (1999) studied a standard collapsar model with a $14M_\odot$ helium core of a $35M_\odot$ main-sequence star. Core collapse would give birth to a $2\text{–}3M_\odot$ black hole. The final outcome critically depends on the specific angular momentum

$$j = \frac{mv_\theta r}{m} = \Omega r^2 \tag{10.9}$$

of the core, where Ω is the angular velocity, and $v_\theta = \Omega r$ is the azimuthal velocity. For reference, a uniformly rotating neutron star with period $P = 2\pi/\Omega \sim 1$ ms has a specific angular momentum at the surface (general relativistic (GR) effects neglected):

$$j_{\rm NS} = 6.3 \times 10^{15}\ {\rm cm^2\ s^{-1}}\ P_{-3}^{-1} R_{\rm NS,6}. \tag{10.10}$$

For an accreting Schwarzschild black hole, the inner radius of the accretion disk is $r_{\rm in} = 6GM/c^2$. Considering Keplerian motion $\Omega = (GM/r^3)^{1/2}$, the specific angular momentum (neglecting GR effects) reads

$$j_{\rm in} = (GMr)^{1/2} = \frac{\sqrt{6}GM}{c} = 3.2 \times 10^{16}\ {\rm cm^2\ s^{-1}} \left(\frac{M_{\rm BH}}{3M_\odot}\right). \tag{10.11}$$

MacFadyen and Woosley (1999) found that if the specific angular momentum is $j < 3 \times 10^{16}$ cm^2 s^{-1}, material would fall into the black hole almost uninhibited, so that no GRB could be launched. If, however, the specific angular momentum falls into the range $j = (3\text{–}20) \times 10^{16}$ cm^2 s^{-1}, a jet would be launched along the rotation axis via neutrino–anti-neutrino annihilations, which would power a GRB after the jet emerges from the star. One can see that, within the collapsar model of long GRBs with a black hole central engine, the required specific angular momentum at the core is very demanding.

A crucial question is how to attain and sustain a rapidly rotating core at the end of a star's life. A key constraint is that dynamo action by differential rotation in a stably stratified star would generate magnetic fields, whose torque would slow down the star (Spruit, 2002). Detailed calculations (Heger et al., 2005) indicated that with the magnetic torque, the final rotation rate of the collapsing iron core would be slower by a factor of 30–50 with respect to the case without the magnetic torque. If the progenitor star is a single star, then one needs to find a mechanism to reduce this magnetic torque. Alternatively, one may introduce an interacting binary progenitor model to spin up the system. Petrovic et al. (2005) showed that this model may provide the required j if the magnetic torque is ignored. However, when the magnetic torque is included, this model is no better than the single star model.

One plausible model is that the progenitor stars are very rapidly rotating at birth. This leads to a mix of the onion layers usually expected in massive stars. Even during the main-sequence phase, hydrogen and helium are mixed, resulting in a quasi-chemically homogeneous evolution (QCHE, Yoon and Langer 2005; Woosley and Heger 2006; Yoon et al. 2006). Such stars burn almost all hydrogen to helium on the main sequence, and go directly to the WR branch without evolving to red giants. This scenario requires a low metallicity (e.g. $Z < 0.004$, Yoon et al. 2006). Fryer et al. (2007) argued that, based on the observational constraints, single stars cannot be the only progenitor for long GRBs, and several binary progenitor scenarios are needed.

In general, low metallicity favors a rapidly rotating core. This is because both line-driven and grain-driven mass losses increase with increasing metallicity. Such mass losses drive a stellar wind, and remove angular momentum from the star. As a result, low metallicity would help to retain the high angular momentum of the star. On the other hand, a WR star needs a strong wind to strip away the hydrogen envelope. This is in apparent contradiction with the high angular momentum requirement. The mixing scenario discussed above helps

to alleviate the contradiction, since there is no hydrogen envelope to begin with. A strong wind is not needed to interpret the lack of hydrogen lines in the SN spectrum.

The theoretical low-metallicity preference seems to be consistent with the long GRB observational data (e.g. Graham and Fruchter, 2013) (§2.4 for details).

The spin and metallicity constraints on the progenitor star are most relevant for a black hole central engine. If the engine is instead a magnetar, the constraints are less demanding, even though a rapidly rotating core (with less required angular momentum) is still needed to make a millisecond rotator.

10.2.4 Jet–Envelope Interaction

One key ingredient of the long GRB massive star progenitor is the stellar envelope. The interaction between the GRB jet and the envelope introduces some interesting physical processes and observational features.

Dynamics

The propagation of a relativistic jet through the stellar envelope of the progenitor star has been studied both numerically (Zhang et al., 2003b, 2004b; Morsony et al., 2007, 2010; Tchekhovskoy et al., 2009; Bromberg and Tchekhovskoy, 2016; López-Cámara et al., 2016; Geng et al., 2016) and analytically (Mészáros and Rees, 2001; Ramirez-Ruiz et al., 2002; Waxman and Mészáros, 2003; Matzner, 2003; Bromberg et al., 2011b). The general features from these investigations include (Figs. 10.1 and 10.2):

- As a fast, supersonic jet propagates into a dense stellar medium, a forward bow shock forms in the envelope. If the jet is matter dominated, a reverse shock propagates into the jet. The region between the FS and RS defines the *jet head*, which propagates substantially slower than the jet itself (Zhang et al., 2003b; Waxman and Mészáros, 2003).
- The strong deceleration of the jet generates significant heat. Matter from the jet head is pushed sideways to form a hot *cocoon* around the jet (Mészáros and Rees, 2001; Ramirez-Ruiz et al., 2002; Zhang et al., 2004b). The cocoon expands sideways towards the envelope material, forming another pair of shocks, one into the envelope and another one into the cocoon itself.
- If the cocoon pressure is high enough, the cocoon reverse shock gives a collimation pressure to the jet to reduce the jet angle (Morsony et al., 2007; Tchekhovskoy et al., 2009).
- If the central engine lasts long enough to allow the jet head to break out of the star, a successful jet is produced. This would power a high-luminosity, highly variable GRB. If, however, the central engine is quenched before the jet breaks out of the star, then the jet would lose pressure from below. The jetted material would in any case keep expanding into the envelope with less confinement. Eventually the forward shock would break out of the star, leading to a less luminous, soft burst with a smooth lightcurve without significant temporal structure (Bromberg et al., 2011b).

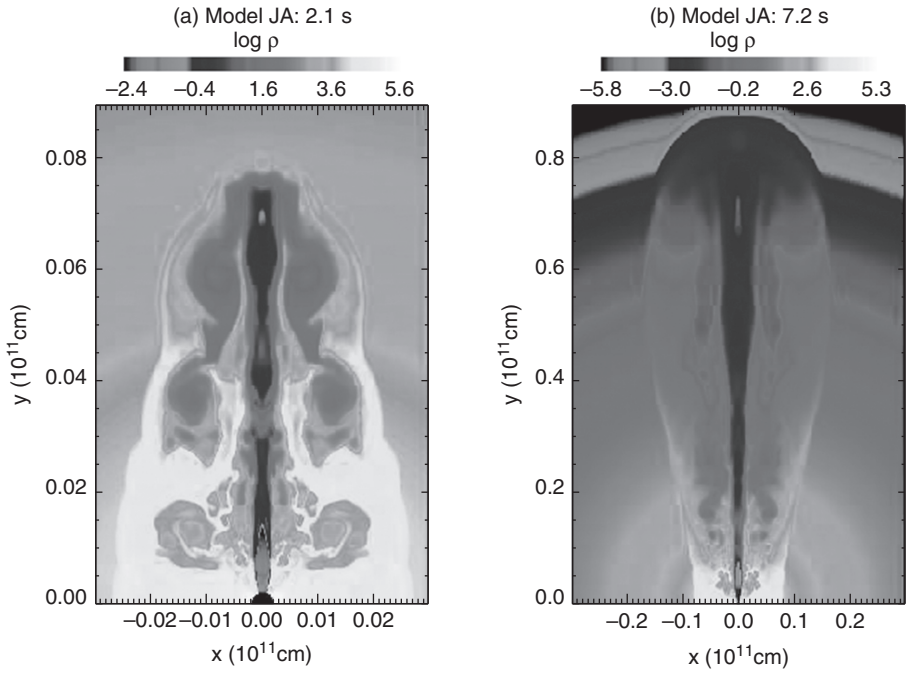

Figure 10.1 The jet–cocoon structure formed during the jet–envelope interaction. Numerical results. Reproduced from Figure 1 in Zhang et al. (2003b) with permission. ©AAS. A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

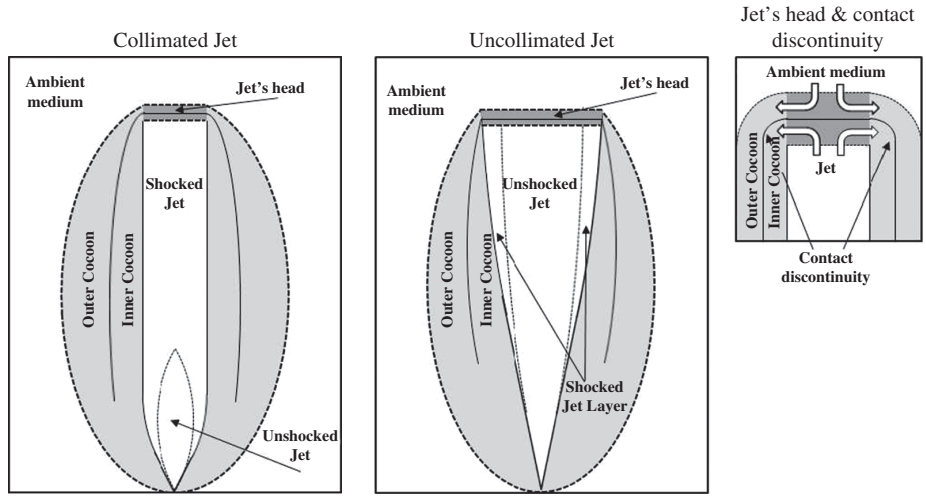

Figure 10.2 A cartoon picture of jet–cocoon structure. Reproduced from Figure 1 in Bromberg et al. (2011b) with permission. ©AAS.

- For a successful jet, as the jet emerges from the star, the hot cocoon will also break out of the star and travel together with the central jet, but with a lower Lorentz factor. This naturally gives rise to a two-component *spine–sheath structure* of the GRB jet (Zhang et al., 2004b).
- For a successful jet, the interaction between the jet and the envelope induces variability, which may leave imprints on the GRB lightcurve (Morsony et al., 2010).
- A Poynting-flux-dominated jet may propagate faster than a matter-dominated jet, due to the non-existence of a reverse shock (Bromberg and Tchekhovskoy, 2016).
- If the central engine injection is intermittent, additional interesting features are produced. The propagation of the unsteady jet is affected by the interaction with the progenitor material (López-Cámara et al., 2016). Some pulses may be quenched, and an early quiescent gap may be produced even if the injection remains periodically intermittent (Geng et al., 2016).

Duration Distribution

The time scale for the jet to penetrate through the star is the *jet breakout time*:

$$t_b = \frac{R_*}{\bar{\beta}_h c} = 33 \text{ s } R_{*,11}\bar{\beta}_{h,-1}. \tag{10.12}$$

Bromberg et al. (2012) argued that, for a successful GRB with an observed duration of $t_{\rm GRB}$, the central engine duration should be

$$t_{\rm eng} = t_{\rm GRB} + t_b. \tag{10.13}$$

For example, if $t_b = 10$ s, a GRB has a duration of only 5 s if the engine lasts for 15 s. The material and energy injected during the first 10 s are released (if the jet is successful). However, most of the energy is given to the cocoon, so that for the first 10 s, the jet emission is too faint to be included in the duration of the GRB. A GRB detector would only register the emission of the central engine in the remaining 5 s.

Under this assumption, Bromberg et al. (2012) made an argument that the observed GRB duration distribution supports the collapsar model. The reasoning is the following: in order to make a successful GRB, one needs to have $t_{\rm eng} > t_b$. The probability of a GRB having a duration $t_{\rm GRB}$ should be equal to the probability of a GRB engine having a duration $t_{\rm eng} = t_{\rm GRB} + t_b$, i.e.

$$p_{\rm GRB}(t_{\rm GRB})dt_{\rm GRB} = p_{\rm eng}(t_{\rm GRB} + t_b)dt_{\rm eng}. \tag{10.14}$$

If the central engine duration $t_{\rm eng}$ is just slightly longer than the breakout time t_b, one has $t_{\rm GRB} = t_{\rm eng} - t_b \ll t_b$. Noting $dt_{\rm GRB} = dt_{\rm eng}$, one would have

$$p_{\rm GRB}(t_{\rm GRB}) = p_{\rm eng}(t_{\rm GRB} + t_b) \simeq p_{\rm eng}(t_b) = \text{const}, \tag{10.15}$$

so that one expects a *plateau* in the GRB duration distribution. By plotting dN/dT_{90} vs. T_{90} (instead of $dN/d\log T_{90}$ as is usually plotted), Bromberg et al. (2012) indeed found a plateau below 20–30 seconds in the data of various GRB missions (BATSE, *Swift*, and *Fermi*/GBM, Fig. 10.3). They suggested that this is consistent with the massive star core

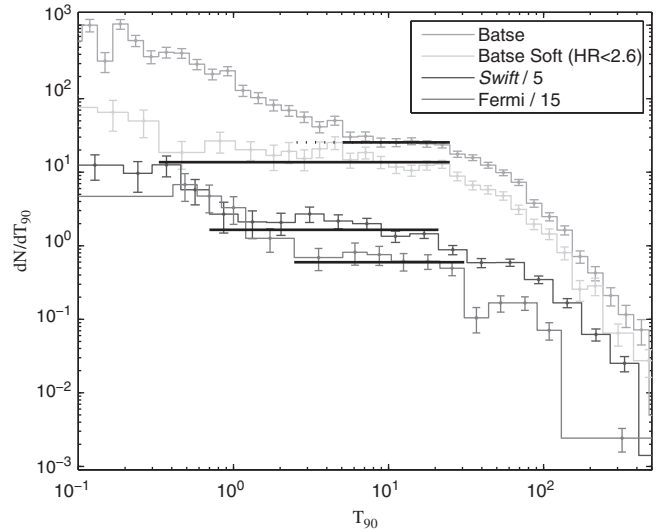

Figure 10.3 A plateau existing in the duration distribution of GRBs, giving direct evidence of jet propagation inside a massive star. Reproduced from Figure 1 in Bromberg et al. (2012) with permission. ©AAS. A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

collapse model, with the upper end of the plateau corresponding to the jet propagation time inside the star, i.e. $t_b \sim 10$ s (after redshift correction).

10.2.5 Shock Breakouts

For the explosion of a massive star that gives rise to a GRB–SN association, shocks traverse the star, break out of the envelope, and leave observational signals. There are two different shock breakout components: one from the collimated jet powered by the central engine, another from the more isotropic, non-relativistic, supernova shock. For the former the jet can be either successful, which gives rise to an ultra-relativistic shock breakout, or choked, which gives rise to a trans-relativistic shock breakout. On the other hand, the supernova shock takes a longer time to break out of the star, and, as it breaks out, the shock is non-relativistic.

These shock breakouts share similar physics, yet show some noticeable differences.

Non-Relativistic Shock Breakouts

When a shock propagates inside the stellar envelope, photons generated in the shock are trapped in the shocked region. These shocks are therefore radiation mediated. If the shock speed is Newtonian (e.g. Colgate, 1974; Matzner and McKee, 1999; Katz et al., 2010; Nakar and Sari, 2010; Katz et al., 2012), which is relevant to SN shocks, thermal equilibrium can be reached. Ignoring the coefficients of order unity, one has $aT^4 \sim \rho v^2$, or

$$T_{\rm BB} \simeq (\rho/a)^{1/4} v^{1/2}. \tag{10.16}$$

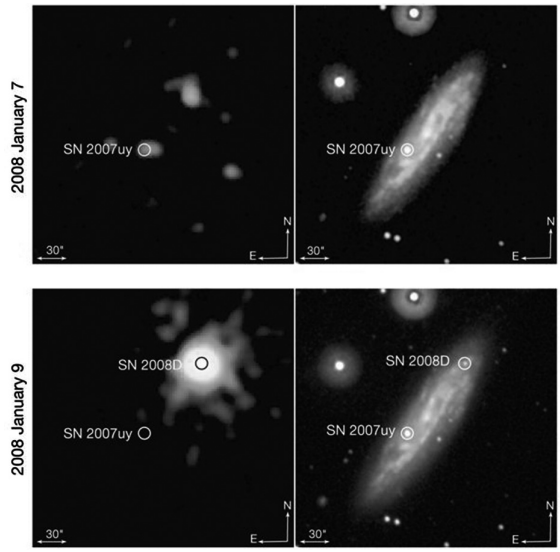

Figure 10.4 The discovery images of XRO 080109 and its associated supernova SN 2008D. From Soderberg et al. (2008). A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

The signal therefore depends on the density of the star and the velocity of the shock, which are directly related to the mass and size of the stars. Larger stars have more extended, stratified envelopes (low ρ) and, hence, lower temperatures. For example, red supergiants have a typical temperature in the 1–10 eV range, and blue supergiants have a typical temperature of several tens of eV. These shock breakouts therefore give rise to UV transients.

WR stars are more compact (hydrogen envelope stripped). Solving detailed shock dynamics and temperature evolution, Nakar and Sari (2010) derived the shock breakout temperature:

$$T_{0,\mathrm{WR}} \simeq 2\ \mathrm{keV} \left(\frac{M_{\mathrm{WR}}}{15 M_\odot}\right)^{-1.7} \left(\frac{R_{\mathrm{WR}}}{5 R_\odot}\right)^{-1.5} E_{51}^{1.8}. \tag{10.17}$$

The main output of the breakout signal is expected in the X-ray band.

On 9 January 2008, Soderberg et al. (2008) made a serendipitous discovery of an X-ray outburst with the *Swift*/XRT, which lasted for several hundred seconds above the XRT sensitivity (Figs. 10.4 and 10.5), with an X-ray peak luminosity $L_{\mathrm{X,p}} \sim 6 \times 10^{43}$ erg s^{-1} and a total X-ray energy $E_{\mathrm{X}} \sim 2 \times 10^{46}$ erg. Follow-up observations of this outburst led to the discovery of He-rich Type Ibc SN 2008D. This established a shock breakout origin of this X-ray outburst, which is named XRO 080109/SN 2008D. This event is consistent with a non-relativistic SN shock breakout.

Trans-Relativistic Shock Breakouts Associated with Failed GRB Jets

If a relativistic jet is launched from a central engine, but the engine is quenched before the jet emerges from the envelope, the jet would spread sideways. In any case, the shock would

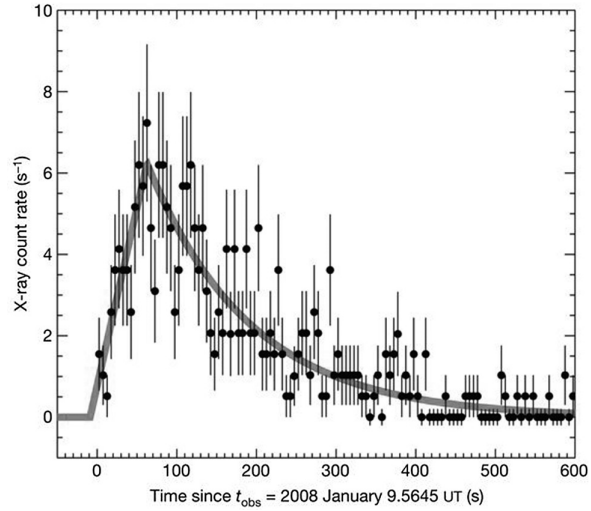

Figure 10.5 The X-ray lightcurve of XRO 080109 as detected by *Swift*/XRT. From Soderberg et al. (2008).

eventually break out of the star, but with a trans-relativistic speed. Shock breakouts in this regime have been studied by many authors (e.g. Tan et al., 2001; Wang et al., 2007c; Li, 2007; Katz et al., 2010; Budnik et al., 2010; Nakar and Sari, 2012).

There are several noticeable differences between trans-relativistic shock breakouts and non-relativistic shock breakouts.

First, $T_{\rm BB} \propto v^{1/2}$ applies when the dimensionless shock velocity $\beta_s < 0.5$. When β_s exceeds 0.5, the temperature behind the shock exceeds 50 keV. A large enough fraction of the photons in the Wien tail of the blackbody emission exceeds $m_e c^2 = 511$ keV, so that $e^{\pm}$ pairs are produced, which become the dominant source of photons. The exponential sensitivity of the number of pairs (in the Wien tail) to the temperature regulates the downstream rest-frame temperature to about 100–200 keV (Katz et al., 2010; Budnik et al., 2010), and also leads to a temperature-dependent opacity. Second, the dynamics of the shock both before and after the shock breakout are modified. Finally, the physical widths of the shocks are also noticeably different between trans-relativistic and non-relativistic shocks.

Nakar and Sari (2012) studied trans-relativistic shock breakouts in great detail. They calculated three important observables: the shock breakout energy $E_{\rm bo}$, the breakout emission temperature $T_{\rm bo}$, and the observed duration of the shock breakout signal $t_{\rm bo}$, in terms of two physical properties: the size of the star R_*, and the Lorentz factor $\gamma_{\rm bo}$ of the breakout shell after the acceleration phase ends. Specifically, they obtained

$$E_{\rm bo} \simeq 2 \times 10^{45}\ {\rm erg} \left(\frac{R_*}{5R_\odot}\right)^2 \gamma_{\rm bo}^{1.36}, \tag{10.18}$$

$$T_{\rm bo} = T'_{\rm bo}\gamma_{\rm bo} \sim 50\ {\rm keV}\ \gamma_{\rm bo}, \tag{10.19}$$

$$t_{\rm ob}^{\rm obs} \simeq \frac{R_*}{c\gamma_{\rm bo}^2} \simeq 10\ {\rm s} \left(\frac{R_*}{5R_\odot}\right) \gamma_{\rm bo}^{-2}, \tag{10.20}$$

where Eq. (10.19) makes use of the effect of constant comoving temperature due to the temperature-mediated pair production rate (as discussed above), and Eq. (10.20) is the standard angular spreading time.

Cancelling out two unknown parameters R_* and $\gamma_{\rm bo}$, Nakar and Sari (2012) obtained a "closure relation" among the three observables:

$$\left(\frac{t_{\rm bo}^{\rm obs}}{20\ {\rm s}}\right) \sim \left(\frac{E_{\rm bo}}{10^{46}\ {\rm erg}}\right)^{1/2} \left(\frac{T_{\rm bo}}{50\ {\rm keV}}\right)^{-2.68}, \tag{10.21}$$

which can be used to check whether a particular transient may be interpreted as a shock breakout event. They argued that the four low-luminosity GRBs they studied are all consistent with being due to shock breakouts: GRB 060218 and GRB 100316D are two weak explosions breaking out at a relatively large radius $R_* \sim 5 \times 10^{13}$ cm with $\gamma_{\rm bo} \sim 1$; GRB 980425 and GRB 031203, on the other hand, are two more energetic explosions breaking out at a smaller radius (still larger than a typical Wolf–Rayet star), but with $\gamma_{\rm bo} \sim$ (3–5). The large breakout radii of these events suggests that the WR star may have a thick wind with a large optical depth, so that the shock breakout is from the wind rather than the stellar surface (Li, 2007). Alternatively, there might be an extended low-mass envelope surrounding low-luminosity GRBs (Nakar, 2015).

Relativistic Shock Breakouts Associated with Successful GRBs

For successful GRBs, the first episode of the jet emission signature would also come from the shock breakout. A precise treatment requires detailed numerical simulations. Wang and Mészáros (2007) estimated that a soft transient signal with temperature $kT \sim 10$ keV and luminosity $L_{\rm X} \sim 2 \times 10^{48}$ erg s^{-1} may signal the breakout signature. This signal can at most lead the main burst by less than 10 s. Using Eq. (10.20) of Nakar and Sari (2012), one can also estimate $E_{\rm bo} \sim 10^{48}$ erg for $\gamma_{\rm bo} \sim 100$. The observed precursor emission of GRBs is typically brighter than this and has a longer gap (e.g. 100 s) with respect to the main burst. So it is likely not due to shock breakout, but rather signals an early, weaker central engine activity episode before the main burst.

10.2.6 Supernova

A GRB-associated SN usually peaks around 1–2 weeks after the GRB. The spectrum around the peak time is characterized by broad emission lines, suggesting a very high (yet still non-relativistic) velocity of the ejecta.

The SN lightcurve is shaped by several physical processes (Arnett, 1982), including decays of ^{56}Ni and ^{56}Co that heat up the ejecta, thermalization of the decay-generated γ-rays, an ejecta opacity that depends on the decay-generated positrons and ionization fraction of heavy elements' atoms, dynamical evolution of the ejecta and photosphere, as well as the possible energy injection from a long-lasting central engine (e.g. a magnetar). Nonetheless, some observational properties may be directly related to the physical properties of the ejecta. In particular, the $^{56}{\rm Ni} \rightarrow {}^{56}{\rm Co}$ decay releases γ-rays at a rate

$$s_{\gamma,\mathrm{Ni}} = (3.9 \times 10^{10}\ \mathrm{erg\ g^{-1}\ s^{-1}})\ \exp\left(-\frac{t}{t_{\mathrm{Ni}}}\right), \tag{10.22}$$

with the nickel decay half-life

$$t_{\mathrm{Ni}} = 8.77\ \mathrm{d}. \tag{10.23}$$

The $^{56}\mathrm{Co} \rightarrow {}^{56}\mathrm{Fe}$ decay, on the other hand, releases γ-rays and positrons at a rate

$$s_{\gamma,\mathrm{Co}} = (6.7 \times 10^{9}\ \mathrm{erg\ g^{-1}\ s^{-1}})\ \exp\left(-\frac{t}{t_{\mathrm{Co}}}\right), \tag{10.24}$$

$$s_{\mathrm{e}^+,\mathrm{Co}} = (6 \times 10^{8}\ \mathrm{erg\ g^{-1}\ s^{-1}})\ \exp\left(-\frac{t}{t_{\mathrm{Co}}}\right), \tag{10.25}$$

respectively, with the cobalt decay half-life

$$t_{\mathrm{Co}} = 111.5\ \mathrm{d}. \tag{10.26}$$

The peak luminosity of the lightcurve depends on the ^{56}Ni mass in the ejecta, as well as the γ-ray energy deposition rate at the peak time:

$$L(t_{\mathrm{max}}) \simeq M(^{56}\mathrm{Ni}) s_{\gamma}(t_{\mathrm{max}}), \tag{10.27}$$

where the peak time t_{max} (and also the width of the peak τ_{LC}) is defined by the optically thin condition, which depends on the optical opacity coefficient κ, total mass of the ejecta M_{ej}, and the total kinetic energy of the ejecta $E_{K,\mathrm{ej}}$, in the form (Arnett, 1982; Nakamura et al., 2001)

$$t_{\mathrm{max}} \sim \tau_{\mathrm{LC}} \sim (\kappa/c)^{1/2} M_{\mathrm{ej}}^{3/4} E_{K,\mathrm{ej}}^{-1/4}. \tag{10.28}$$

There is a degeneracy between M_{ej} and $E_{K,\mathrm{ej}}$ in this expression. However, the width of the emission line depends on the velocity of the ejecta,

$$v_{\mathrm{ej}} = \left(\frac{2E_{K,\mathrm{ej}}}{M_{\mathrm{ej}}}\right)^{1/2}. \tag{10.29}$$

Therefore, a well-observed SN carries information about M_{ej}, E_{ej} (or v_{ej}), and $M(^{56}\mathrm{Ni})$.

This simple picture applies if the SN lightcurve is powered by nickel decay. Recent SN observations suggest that at least some SNe (the superluminous ones) require additional energy injection from the central engine, likely a millisecond magnetar. The dynamics and lightcurve properties of these SNe are modified due to the energy injection process (Kasen and Bildsten, 2010; Woosley, 2010; Wang et al., 2016a).

10.2.7 Other Massive Star Progenitors

Besides the standard Wolf–Rayet star progenitor, several other possible progenitor systems of long GRBs have been discussed in the literature:

- Fryer and Woosley (1998) proposed the merger of a helium star and a black hole as the progenitor of long GRBs. Numerical simulations by Zhang and Fryer (2001) suggested that a successful long GRB with the right duration and energy may be generated via the

merger of a $2M_\odot$ black hole and a $16M_\odot$ helium star. A similar scenario invokes the merger of a helium star and a neutron star (Fryer et al., 2013). Even though these helium star mergers may not produce most of the long GRB population, they have been invoked to explain some special GRBs. For example, Thöne et al. (2011) applied the helium star–neutron star merger scenario to interpret the special "Christmas GRB" 101225A, which showed a blackbody-dominated afterglow. This interpretation was further supported by numerical simulations (Cuesta-Martínez et al., 2015a,b).

- Ruffini and collaborators (Ruffini et al., 2016, 2018) envisaged eight different kinds of GRB progenitors that all invoke binary systems, composed of different combinations of carbon–oxygen cores (CO stars), neutron stars, black holes, and white dwarfs. In the scenarios that give rise to long GRBs, the core collapse of a CO star results in a Type Ic supernova. The SN ejecta trigger hyper-accretion onto a NS companion, making the NS collapse into a BH, leading to an *induced gravitational collapse* (IGC) (Ruffini et al., 2008; Rueda and Ruffini, 2012; Fryer et al., 2014). In some other cases, a massive NS instead of a BH is formed after the accretion phase, which may give rise to lower-luminosity events such as XRFs. The advantage of the model is that it naturally accounts for a Type Ic SN associated with the GRB, with a complete absence of (or very little) helium. It is unclear whether the observed GRB event rate density can be matched by the detailed population synthesis studies of these binary systems.
- Mészáros and Rees (2001) proposed that collapsar jets can be launched not only from He stars, but also from blue supergiants. Even though typical long GRBs with Type Ic SN associations are consistent with a hydrogen-depleted progenitor, the discovery of several *ultra-long* GRBs lasting longer than thousands of seconds may call for such a progenitor (e.g. Levan et al., 2014b; Gendre et al., 2013). Whether ultra-long GRBs require a distinct progenitor is still an open question (e.g. Zhang et al., 2014; Virgili et al., 2013; Gao and Mészáros, 2015). A prediction of the model is the existence of a superluminous supernova (e.g. Nakauchi et al., 2013). Greiner et al. (2015) indeed discovered a very luminous supernova (SN 2011kl) associated with the ultra-long GRB 111209A. However, the authors argued that the SN was more likely being powered by a magnetar than being generated from a blue supergiant.

10.2.8 Origin of Low-Luminosity GRBs

A typical low-luminosity GRB has a relatively low luminosity, small energy, low E_p, long duration, smooth lightcurve, and a very low degree (if any) of collimation. Liang et al. (2007a) and Virgili et al. (2009) showed that they form a distinct component in the luminosity function of long GRBs, and therefore may point towards a distinct population from the standard high-luminosity (HL)GRBs. More data later suggested that the distinction between the LL- and HL-GRBs is less clean, and the possibility that they form one single luminosity function component is not ruled out (Sun et al., 2015).

Modeling of the supernova spectra (Mazzali et al., 2006) and observations of the radio afterglow emission (Soderberg et al., 2006) both suggest that the progenitor of the low-luminosity GRB 060218/SN 2006aj has a relatively small mass ($20M_\odot$), so that the central

engine of the GRB/SN association may be a neutron star rather than a black hole. The SN peak flux is also lower than other GRB-associated SNe (Pian et al., 2006). On the other hand, the fact that both LL- and HL-GRBs (e.g. GRB 030329, Stanek et al. 2003; Hjorth et al. 2003 and GRB 130427A, Xu et al. 2013) have Type Ic SN associations suggests that the progenitors of LL-GRBs may not be very different from HL-GRBs. The difference between the two apparent types may lie in the differences at the central engine or in the ability to launch a successful jet.

Wang et al. (2007c), Bromberg et al. (2011a), and Nakar and Sari (2012) proposed that the main difference between the two categories is whether a successful jet can emerge from the stellar envelope: while HL-GRBs are powered by successful jets, LL-GRBs are those events whose central engine shuts off before the jet emerges from the envelope. The observed low-luminosity, long-duration emission is therefore related to shock breakout instead of emission from the successful jet. Alternatively, LL-GRBs may be surrounded by an extended low-mass stellar envelope, which successful GRBs lack (Nakar, 2015). Irwin and Chevalier (2016) suggested that a low-luminosity, mildly relativistic, successful jet may also account for the emission data of GRB060218-like GRBs.

Observationally there is no clear boundary line between successful jet GRBs and shock breakout GRBs. Zhang et al. (2012b) found that the threshold luminosity above which a successful variable jet is observed is about 10^{48} erg s^{-1}. On the other hand, GRB 031203, which was interpreted as a shock breakout event by Nakar and Sari (2012), is above this line.

10.3 Compact Star GRBs

10.3.1 Observational Evidence

Compared with the observational evidence for massive star GRBs, the observational evidence for compact star GRBs was more indirect (Berger, 2014, for a review) before 2017.

- A fraction of short (or short with extended emission) GRBs are found in elliptical or early-type galaxies with little star formation. This suggests that these GRBs can be produced without the existence of a massive star.
- The majority of short GRBs are in star-forming galaxies. However, they usually do not track the bright light (star-forming region), and their local specific star formation rate is low.
- Many short GRBs have a large offset from the center of their host galaxies. Some short GRBs are “hostless”, i.e. no apparent host is found at the afterglow location. In many cases, a bright nearby galaxy is found in the vicinity of the afterglow, suggesting that the short GRB progenitor may have been “kicked” out of the galaxy. Alternatively, the short GRB may be very far away, so that its host galaxy is not detectable.
- For nearby short GRBs, deep upper limits for an associated SN have been placed.

All these observational facts point towards a progenitor source different from massive star GRBs. The leading model is NS–NS mergers or NS–BH mergers. The definite clue for this progenitor was made on 17 August 2017, when the NS–NS merger gravitational wave source GW170817 was found to be associated with the short GRB 170817A (Abbott et al., 2017b; Goldstein et al., 2017) (§12.3.3 for details).

10.3.2 Merger Physics and Jet Launching

Many groups have studied the mergers of two compact objects (NS–NS and NS–BH) both analytically (e.g. Eichler et al., 1989; Narayan et al., 1992; Mészáros and Rees, 1992) and numerically (e.g. Ruffert and Janka, 1999; Rosswog et al., 2003; Aloy et al., 2005; Rezzolla et al., 2011; Rosswog et al., 2013; Hotokezaka et al., 2013; Nagakura et al., 2014; Paschalidis et al., 2015). The general physical picture emerging from these simulations includes the following:

- Both NS–NS mergers and BH–NS mergers can make a system with a central BH surrounded by a dense disk (torus). For NS–NS mergers under certain conditions, a differential-rotation-supported hyper-massive NS is produced, which may survive for $\sim$100 ms before collapsing to a BH (Rosswog et al., 2003). Accretion of the torus material into the BH may launch a jet, which powers a short-duration GRB. There are exceptions, however, for both types of mergers: (1) For NS–NS mergers, if the NS equation of state is stiff and if the masses of the two NSs are small enough, a merger would produce a supra-massive NS or even a stable NS (e.g. Dai et al., 2006; Fan and Xu, 2006; Zhang, 2013; Giacomazzo and Perna, 2013); (2) For NS–BH mergers, if the mass ratio $q \equiv M_{\rm NS}/M_{\rm BH}$ is < 0.1, the tidal disruption radius of the NS is within the BH event horizon and the BH would swallow the NS completely without making a GRB.
- Before a merger event, the NSs in the merger system are tidally distorted. Some material can be dynamically ejected from the system. Another source of ejecta is the neutrino-driven wind from the accretion disk or the post-merger proto-NS. Depending on the merger members, the post-merger remnant, and NS equation of state, the ejected rest mass ranges from 10^{-3} to $10^{-1}M_{\odot}$, with a typical velocity 0.1–0.25c (e.g. Freiburghaus et al., 1999; Rezzolla et al., 2010; Hotokezaka et al., 2013; Rosswog et al., 2013). This ejected material is neutron rich, and would power the so-called "macronova" or "kilonova" (Li and Paczyński, 1998; Kulkarni, 2005; Metzger et al., 2010).
- Both the $\nu\bar{\nu}$ annihilation mechanism and the magnetic mechanism have been considered for jet launching. The magnetic mechanism is the preferred one for launching an energetic jet as observed in some short GRBs (e.g. Rezzolla et al., 2011; Paschalidis et al., 2015). According to Rezzolla et al. (2011), a broad outflow with a half-opening angle $\sim 30^{\rm o}$ may be launched (Fig. 10.6). GRMHD simulations by Paschalidis et al. (2015) showed the launch of a jet from a BH–NS merger system.
- Further collimation by the merger ejected material may be possible (if the ejecta mass $M_{\rm ej} \geq 10^{-2}M_{\odot}$) and can achieve an opening angle of $\theta_j \leq 10^{\rm o}$ (Nagakura et al., 2014).

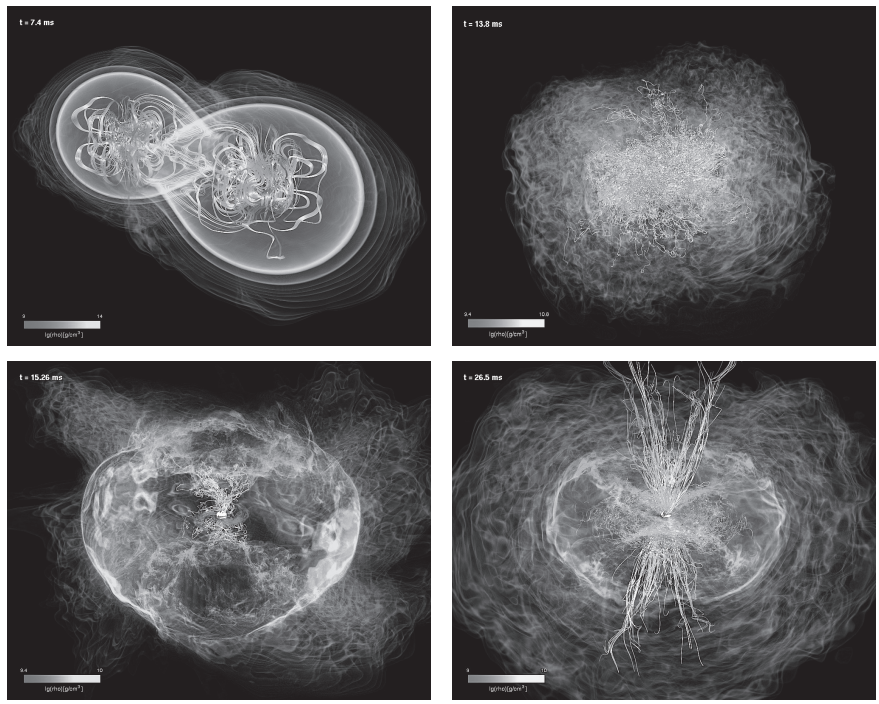

Figure 10.6 Numerical simulations that show launching of a broad outflow from NS–NS mergers. Four snapshot density images with magnetic field lines (green for within the torus and equatorial plane; white for outside the torus and near the BH spin axis) are shown. Reproduced from Figure 1 in Rezzolla et al. (2011) with permission. ©AAS. A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

10.3.3 Macronova/Kilonova

The neutron-rich wind launched during the compact-star-merger process would power a near-isotropic, supernova-like signal in the optical/IR band due to radioactive decay. There are several names to describe the phenomenon in the literature:

- Since Li and Paczyński (1998) first calculated such a signal, the phenomenon is called "Li–Paczyński nova" in some papers;
- Kulkarni (2005) suggested calling the phenomenon "macronova";
- Metzger et al. (2010) calculated the brightness of the phenomenon and found that the peak luminosity is $\sim 10^{41}$ erg s^{-1}, which is about 1000 times that of classical novae. So they suggested calling the phenomenon "kilonova";
- Since the main heating source of the phenomenon is the "r-process" (without energy injection from the central engine), some authors called the phenomenon "r-process nova";
- Yu et al. (2013) and Metzger and Piro (2014) considered the possibility of energy injection from a rapidly rotating magnetar as the post-merger product. Since the magnetar heating can exceed "r-process" heating, and since the peak luminosity can be even

brighter than 10^{41} erg s^{-1}, Yu et al. (2013) suggested calling the phenomenon "merger-nova" to reflect broader possibilities. It was argued that energy injection may also be possible for a BH post-merger product under certain conditions (Song and Liu, 2017; Ma et al., 2017).

The general physical picture of the phenomenon can be summarized as follows (see Metzger 2017 for a detailed review):

Since at least one member of the merger system is a NS, the dynamically launched ejecta (with a mass of 10^{-3}–$10^{-1} M_\odot$) carries a large amount of free neutrons and a small amount of elements lighter than Fe from the NS crust. After escaping from the gravitational potential of the merger remnant, this ejecta expands in space with nearly a constant speed. In a dense neutron-rich environment, the neutron capture time scale is shorter than the β-decay time scale. This allows the rapid neutron capture process, or r-process, to quickly synthesize elements heavier than Fe, including gold and platinum. The r-process proceeds along a nuclear path far on the neutron-rich side of the island of stable isotopes. The neutron-rich nuclei quickly decay to stable nuclei and in the meantime release heat. Photons are trapped in the ejecta, and eventually released from the photosphere, powering the r-process-powered kilonova.

The characteristics of the kilonova may be derived from a simple analytical treatment (Li and Paczyński, 1998; Metzger et al., 2010).

Due to large opacity, photons generated in the ejecta are initially trapped and need to diffuse out. The photon diffusion time scale is

$$t_{\rm dif} \sim \frac{\kappa M_{\rm ej}}{cr}. \tag{10.30}$$

When the dynamical time scale $t_{\rm dyn} \sim r/v$ becomes longer than $t_{\rm dif}$, essentially all the photons trapped in the ejecta are free to escape, so that the lightcurve reaches a peak. The condition $t_{\rm dyn} = t_{\rm dif}$ defines the characteristic radius where the EM emission reaches the peak, i.e.

$$R_{\rm peak} \sim \left(\frac{v\kappa M_{\rm ej}}{c}\right)^{1/2} \simeq (1.2\times 10^{14}\ {\rm cm})\,\kappa_{-1}^{1/2}\beta_{-1}^{1/2}\left(\frac{M_{\rm ej}}{10^{-2}M_\odot}\right)^{1/2}. \tag{10.31}$$

The corresponding peak time is

$$t_{\rm peak} \simeq 0.5\ {\rm day}\ \kappa_{-1}^{1/2}\beta_{-1}^{-1/2}\left(\frac{M_{\rm ej}}{10^{-2}M_\odot}\right)^{1/2}. \tag{10.32}$$

Here $\kappa = 0.1\ {\rm cm^2\ g^{-1}}\,\kappa_{-1}$ is the opacity. The radioactive power can be approximated as a decreasing power-law function with time, i.e. $\dot{Q} \propto t^{-\alpha}$. As long as the decay is not too steep, i.e. $\alpha < 2$, then the total amount of radioactive heating that occurs around the time scale $t_{\rm peak}$ would be

$$Q_{\rm peak} = \int_{t_{\rm peak}}^{\infty} \dot{Q}dt \simeq \dot{Q}(t_{\rm peak})t_{\rm peak} = fM_{\rm ej}c^2, \tag{10.33}$$

where $f \ll 1$ is a fudge dimensionless parameter introduced by Li and Paczyński (1998). The peak bolometric luminosity of the event is then

$$L_{\rm peak} \simeq \frac{Q_{\rm peak}}{t_{\rm peak}} \simeq (5 \times 10^{41}\ {\rm erg\ s^{-1}})\, \kappa_{-1}^{-1/2} f_{-6} \beta_{-1}^{1/2} \left(\frac{M_{\rm ej}}{10^{-2} M_\odot}\right)^{1/2}, \tag{10.34}$$

and the effective temperature is

$$\begin{aligned} T_{\rm peak} &= \left(\frac{L_{\rm peak}}{4\pi R_{\rm peak}^2 \sigma}\right)^{1/4} \\ &\simeq (1.4 \times 10^4\ {\rm K})\, \kappa_{-1}^{-3/8} f_{-6}^{1/4} \beta_{-1}^{-1/8} \left(\frac{M_{\rm ej}}{10^{-2} M_\odot}\right)^{-1/8}. \end{aligned} \tag{10.35}$$

Li and Paczyński (1998) adopted an optimistic value of $f \sim 10^{-3}$, which gives a peak luminosity comparable to that of a supernova. Detailed calculations of radioactive heating by means of a nuclear reaction network by Metzger et al. (2010) gave an effective value of

$$f \sim 3 \times 10^{-6}. \tag{10.36}$$

As a result, the peak luminosity of a $10^{-2} M_\odot$ ejecta gives a V-band luminosity of $\sim 10^{41}$ erg s^{-1}, which is about 10^3 times of that of a typical nova. Metzger et al. (2010) therefore suggested naming these events "kilonovae".

Metzger et al. (2010) adopted a typical opacity $\kappa \sim 0.1$ cm^2 g^{-1}, and predicted a transient peaking in the optical band. Barnes and Kasen (2013) and Tanaka and Hotokezaka (2013) pointed out that the existence of heavy elements, in particular the lanthanides, would greatly increase κ by several orders of magnitude, e.g. tens to hundreds of cm^2 g^{-1}. Noting the $\kappa^{-3/8}$ dependence of $T_{\rm peak}$ (Eq. (10.35)) and $\kappa^{1/2}$ dependence of $t_{\rm peak}$ (Eq. (10.32)), the peak of kilonova emission moves to the infrared band and a later peak time (Fig. 10.7). The putative kilonova discovered in association with GRB 130603B (Tanvir et al., 2013; Berger et al., 2013) peaked in the IR band, which is consistent with such a prediction. Metzger and Fernández (2014) further suggested that there could be two components in a kilonova. Whereas the equatorial directions may be populated with lanthanide-rich ejecta with a large opacity and a "red" emission component, it is possible that near the polar direction lanthanide-free ejecta might be launched from a disk wind with a low opacity, so that a "blue" emission component may emerge. Such a two-component model is consistent with the discovered kilonova transient associated with GW170817 (Shappee et al., 2017; Evans et al., 2017; Nicholl et al., 2017; Chornock et al., 2017)

If a NS–NS merger leaves behind a supra-massive, rapidly rotating magnetar rather than a black hole, the resulting mergernova would be continuously powered by the Poynting flux of the spinning down magnetar. The dynamics of the ejecta would be modified, and the signal enhanced (Yu et al., 2013; Metzger and Piro, 2014).

The following treatment closely follows Yu et al. (2013). To accommodate the possibility of a trans-relativistic ejecta under an extreme condition, the treatment is generalized to the relativistic form, which also properly treats the standard non-relativistic motion of the ejecta.

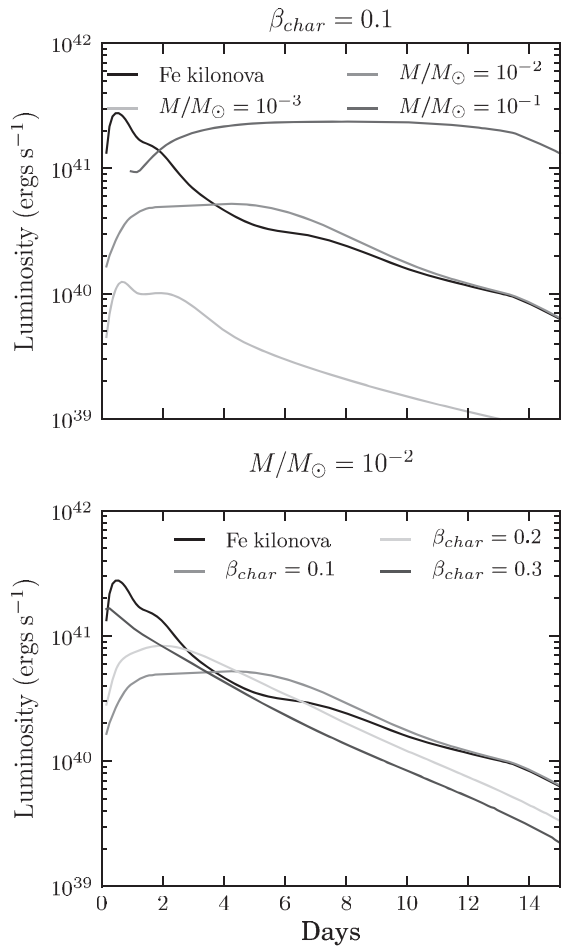

Figure 10.7 The kilonova bolometric lightcurves comparing the low-κ Fe ejecta and high-κ lanthanide ejecta. Reproduced from Figure 2 in Barnes and Kasen (2013) with permission. ©AAS. A black and white version of this figure will appear in some formats. For the color version, please refer to the plate section.

The total energy of the ejecta excluding the rest mass energy may be expressed as

$$E_{\rm ej} = (\Gamma - 1)M_{\rm ej}c^2 + \Gamma E'_{\rm int}, \tag{10.37}$$

where Γ is the Lorentz factor and $E'_{\rm int}$ is the internal energy in the comoving frame.[2] The two terms in Eq. (10.37) represent the kinetic energy and the thermal energy of the ejecta, respectively. For each time step dt in the observer frame, the change of ejecta energy is

$$dE_{\rm ej} = (\zeta L_{\rm sd} + L_{\rm ra} - L_e)dt, \tag{10.38}$$

[2] A more rigorous treatment should replace Γ in the second term of the right hand side by $\Gamma_{\rm eff}$ defined in Eq. (8.54). For the sub- or trans-relativistic ejecta, the current treatment gives a good approximation.

where $L_{\rm sd}$ is the spindown luminosity (which will be discussed in Chapter 11, Eq. (11.50)) and ζ is the fraction of $L_{\rm sd}$ that is injected into the ejecta, $L_{\rm ra}$ is the radioactive power, and L_e is the bolometric radiation luminosity of the heated electrons. Noting Eq. (10.37) and $dt' = \mathcal{D}dt$ (t' is the comoving time, where $\mathcal{D} = 1/[\Gamma(1-\beta\cos\theta)]$ is the Doppler factor), for an on-beam observer ($\theta = 0$), one has

$$\frac{d\Gamma}{dt} = \frac{\zeta L_{\rm sd} + L_{\rm ra} - L_e - \Gamma\mathcal{D}(dE'_{\rm int}/dt')}{M_{\rm ej}c^2 + E'_{\rm int}}. \tag{10.39}$$

The change of the internal energy includes heating due to the magnetar and radioactivity, and cooling due to radiation and pdV work (Kasen and Bildsten, 2010), i.e.

$$\frac{dE'_{\rm int}}{dt'} = \xi L'_{\rm sd} + L'_{\rm ra} - L'_e - P'\frac{dV'}{dt'}, \tag{10.40}$$

where ξ is an efficiency parameter defining the fraction of the magnetar spindown luminosity that is used to heat the ejecta.[3] The comoving luminosities are defined as $L'_{\rm sd} = L_{\rm sd}/\mathcal{D}^2$, $L'_{\rm ra} = L_{\rm ra}/\mathcal{D}^2$, and $L'_e = L_e/\mathcal{D}^2$. The comoving radiative heating luminosity as a function of time depends on the details of r-process nuclear reactions, which require nuclear-chain numerical simulations to quantify. For a rough treatment, one may use a simplified quasi-analytical formula (Korobkin et al., 2012):

$$L'_{\rm ra} = 4\times10^{49}M_{\rm ej,-2}\left[\frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{t'-t'_0}{t'_\sigma}\right)\right]^{1.3}\ {\rm erg\ s^{-1}}, \tag{10.41}$$

with $t'_0 \sim 1.3$ s and $t'_\sigma \sim 0.11$ s. The comoving-frame bolometric emission luminosity of the heated electrons can be estimated as

$$L'_e = \begin{cases} E'_{\rm int}c/(\tau R/\Gamma), & \text{for } t < t_\tau, \\ E'_{\rm int}c/(R/\Gamma), & \text{for } t \geq t_\tau, \end{cases} \tag{10.42}$$

where the first expression takes into account the skin-depth effect of an optically thick emitter. Finally, noting $P' = E'_{\rm int}/(3V')$ for a relativistic gas, one gets

$$\frac{dV'}{dt'} = 4\pi R^2\beta c \tag{10.43}$$

and

$$\frac{dR}{dt} = \frac{\beta c}{1-\beta}. \tag{10.44}$$

The dynamics of the mergernova with energy injection can be properly solved.

The observed spectrum is nearly a blackbody with a typical temperature

$$\varepsilon_{\gamma,p} \approx 4\mathcal{D}kT' = \begin{cases} 4\mathcal{D}k\left(\frac{E'_{\rm int}}{aV'\tau}\right)^{1/4}, & \text{for } \tau > 1, \\ 4\mathcal{D}k\left(\frac{E'_{\rm int}}{aV'}\right)^{1/4}, & \text{for } \tau \leq 1, \end{cases} \tag{10.45}$$

[3] The parameters ζ and ξ have different physical meanings, which were assumed to be the same in Yu et al. (2013). Here we use different symbols to allow a more general discussion.

where k is the Boltzmann constant and a is the blackbody radiation constant. For a blackbody spectrum with comoving temperature T', the luminosity at a particular frequency ν is given by

$$(\nu L_\nu)_{\rm bb} = \frac{8\pi^2 \mathcal{D}^2 R^2}{h^3 c^2} \frac{(h\nu/\mathcal{D})^4}{\exp(h\nu/\mathcal{D}kT') - 1}. \tag{10.46}$$

Coupling this with the dynamical evolution of the ejecta, the lightcurves of a mergernova in different energy bands can be obtained. Figure 10.8 (upper panel) presents the lightcurves of some magnetar-powered mergernovae in different energy bands for different parameters. A comparison of the optical lightcurves between mergernovae, a normal kilonova, and GRB-associated supernovae are presented in Fig. 10.8 (lower panel).

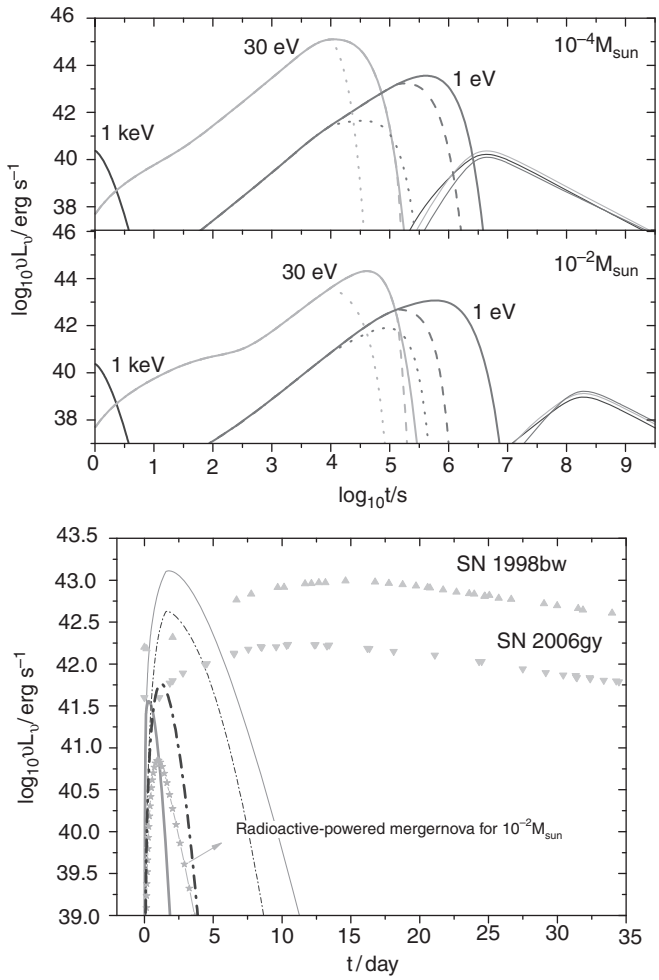

Figure 10.8 *Upper:* Mergernova lightcurves as observed in different bands. *Lower:* A comparison of magnetar-powered mergernova lightcurve and the traditional kilonova lightcurve (with small κ). Adapted from Yu et al. (2013). Figure courtesy Yun-Wei Yu.

Observationally, several kilonova candidates have been reported to be associated with short GRBs, i.e. those associated with GRB 130603B (Tanvir et al., 2013; Berger et al., 2013), GRB 060614 (Yang et al., 2015), and GRB 050709 (Jin et al., 2016). A systematic search for magnetar-powered mergernovae revealed three more candidates, i.e. those associated with GRBs 050724, 070714B, and 061006 (Gao et al., 2017b). The peak luminosities of these events were estimated to be above 10^{42} erg s^{-1}, more than 1 order of magnitude brighter than that of a standard kilonova. It seems that the mergernova phenomenology may have a wide range of peak luminosity.

The NS–NS merger GW event GW170817 (Abbott et al., 2017d) was associated with an optical/IR transient, which has both a blue (Evans et al., 2017; Nicholl et al., 2017) and a red component (Shappee et al., 2017; Chornock et al., 2017). Its lightcurve and spectrum are generally consistent with a macronova/kilonova origin (Kasen et al., 2017; Pian et al., 2017; Tanvir et al., 2017; Shappee et al., 2017; Villar et al., 2017), even though energy injection from a long-lived low-B millisecond pulsar may be helpful to interpret the phenomenology (Yu et al., 2018; Li et al., 2018).

10.3.4 Global Properties

Confronting the NS–NS and NS–BH merger models with the global properties of short GRBs suggests a general agreement between theoretical expectations and data.

Redshift Distribution

Most short GRBs have a relatively low z as compared to long GRBs. This is consistent with the expectation of the compact-star-merger models, which predict a delay time τ_m with respect to star formation, defined by the inspiral time scale of the two compact stars due to gravitational wave radiation. This time scale depends on the initial orbital period $P_{\rm orb}$, the masses of the two compact objects M_1 and M_2, as well as the initial eccentricity e of the orbit. It can be estimated as

$$\tau_m \simeq \frac{P_{\rm orb}}{\dot{P}_{\rm orb}}. \tag{10.47}$$

The orbital decay rate of the binary system due to gravitational wave radiation reads (Taylor and Weisberg, 1989)

$$\dot{P}_{\rm orb} = -\frac{192\pi}{5c^5}\left(\frac{2\pi G}{P_{\rm orb}}\right)^{5/3} f(e)\frac{M_1 M_2}{(M_1+M_2)^{1/3}}, \tag{10.48}$$

where

$$f(e) = \left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right)(1-e^2)^{-7/2}. \tag{10.49}$$

See §12.3.2 for a more detailed discussion.

The distribution of τ_m is unknown. A power-law distribution model, $f(\tau_m) \propto \tau_m^{\eta}$ (Piran, 1992; Nakar et al., 2006), predicts a dominant population in the local universe, which is found inconsistent with the data. Insisting on this model requires a significant contamination (30% and higher) of massive star GRBs in the short GRB population (Virgili et al.,

2011a; Wanderman and Piran, 2015). Alternatively, the delay time scale may have a narrow distribution around 2–3 Gyr (Virgili et al., 2011a; Wanderman and Piran, 2015). However, such a narrow distribution may pose a strong constraint on the birth parameters of the binary merger systems.

Supernova Kicks and Offset Distribution

For NS–NS or NS–BH merger systems, the formation of each compact object is associated with one supernova. Since supernova explosions are typically asymmetric, the binary system will receive a "kick" after each explosion. Under certain conditions, the binary system can survive both SN explosions. After the formation of the second compact object, five "initial" parameters of the compact star binary system define the final fate of the system: M_1, M_2, the initial semi-major axis a_0 (which defines the initial orbital period $P_{\rm orb}$), the initial eccentricity e, and the initial kick velocity of the system $\mathbf{v_k}$. The first four parameters define the merger time scale τ_m (Eq. (10.47)), whereas $\mathbf{v_k}$ together with τ_m defines the location of the merger with respect to the birthplace of the binary system, and with respect to the center of the host galaxy. This requires a proper treatment of the evolution of the binary system in the gravitational potential field of the host galaxy.

Monte Carlo simulations taking into account all these effects were carried out by Bloom et al. (1999). A general consistency between the model and the data was reached (Bloom et al. 2002; Fong and Berger 2013; Berger 2014, see Fig. 2.41).

Luminosity Function and Flux Distribution

The luminosity function and redshift distribution of compact star GRBs may be constrained making use of the observed two-dimensional $L - z$ distribution of the z-known sample and the $\log N - \log P$ distribution of the general sample of short GRBs (Virgili et al., 2011a; Wanderman and Piran, 2015; Sun et al., 2015). The results suggest that there is a non-negligible contamination of massive star GRBs in the short GRB sample, and that there likely exists a characteristic merger delay time scale. The constrained luminosity function is typically a power law (Sun et al., 2015). It turns out that the low-luminosity short GRB 170817A associated with GW170817 falls naturally on the extension of this luminosity function to lower luminosities (Zhang et al., 2018a).

10.3.5 Other Compact Star Progenitors

Besides NS–NS and NS–BH mergers, several other compact star progenitor models have been discussed in the literature.

- Grindlay et al. (2006) argued that a large fraction of short GRBs may be generated from NS–NS interactions in globular clusters. They argued that this model naturally accounts for the large offsets of short GRB locations with respect to their host galaxies, and the extreme density of compact stars in the globular clusters naturally accounts for a high event rate. In fact, the interactions between compact objects in globular clusters are more

likely through direct *collisions* rather than mergers (coalescence due to gravitational wave energy loss). Rosswog et al. (2013) found that dynamical collisions are at least as promising as mergers to produce short GRBs. On the other hand, collisions eject more masses, which may cause a baryon contamination problem to launch a relativistic GRB jet, but would enhance the kilonova signal.

- Black hole–white dwarf mergers were considered as one possible progenitor of GRBs (e.g. Fryer et al., 1999). Narayan et al. (2001) argued that these systems may not be favorable for launching a GRB jet, since the accretion proceeds as a convection-dominated accretion flow (CDAF). As a result, most materials are released as a disk wind, and only a small fraction of mass is accreted. The accretion rate may not be enough to power a GRB.
- Accretion-induced collapse of a neutron star for powering a GRB was discussed by several authors (Qin et al., 1998; MacFadyen et al., 2005; Dermer and Atoyan, 2006). This model can produce GRBs in non-star-forming galaxies, but may not give the large offset of GRB location with respect to the host galaxy center, as well as "hostless" short GRBs.
- The putative short GRB event, GW150914-GBM, was detected 0.4 s after the merger time of the BH–BH merger gravitational wave event GW150914 and lasted for 1 second (Connaughton et al., 2016). Even though whether the signal was genuine is subject to debate (cf. Greiner et al., 2016), the event nonetheless triggered a wave of investigations of the possibility of producing GRBs associated with BH–BH mergers. Most models invoking accretion require the existence of a massive star or accretion disk (Loeb, 2016; Perna et al., 2016; Janiuk et al., 2017), which most likely produce a long-duration GRB due to the free-fall time scale argument presented in §10.2.2. If at least one BH is charged, the merger would give rise to a short-duration electromagnetic counterpart associated with the merger (Zhang, 2016). GW150914-GBM may be generated if the dimensionless charge is as high as $\hat{q} \sim (10^{-5}$–$10^{-4})$ (Zhang, 2016; Liebling and Palenzuela, 2016).

10.4 Other Progenitors that Can Give Rise to Bursts of γ-rays

Besides the two main categories of cosmological GRBs, many other objects may also give rise to bursts of γ-rays. Some of these (e.g. SGRs and TDE jets) were historically confused as GRBs, but are now separated from GRBs. Some others are theoretical speculations that have not been confirmed observationally.

- Soft gamma-ray repeaters (SGRs) are repeating bursts from Galactic magnetars. They emit repeating bursts through magnetic energy dissipation, and occasionally (once a century or so) release *giant flares* with a total isotropic energy up to 2×10^{46} erg within a short duration less than a second followed by a periodic, oscillating X-ray tail (e.g. Palmer et al., 2005). These events are now considered to be a completely different phenomenon from GRBs.

- The existence of giant flares from Galactic SGRs suggests that some SGR giant flares in nearby galaxies could make short GRBs detectable from Earth (the long, X-ray oscillating tail is undetectable from the distance). Indeed, the giant flare of SGR 1806-20 would be detectable up to 80 Mpc (Hurley et al., 2005). A systematic search for the associated nearby galaxies with short GRBs suggested that such associations should make up less than 5% of the short GRB population (Tanvir et al., 2005). The short GRB 051103 may be associated with the M81/M82 system, which is a candidate for a short GRB due to a SGR giant flare (Frederiks et al., 2007).
- "GRB 110328" triggered *Swift* multiple times, and had a peculiar afterglow lightcurve (Burrows et al., 2011). It was soon realized that it is not a traditional GRB, but is powered by a tidal disruption event (TDE) of a super-massive black hole swallowing a nearby star (Bloom et al., 2011; Burrows et al., 2011). The event was later renamed Sw J1644+57. Similar events, e.g. Sw J2058+05, were discovered later (Cenko et al., 2012). The super-Eddington luminosity of these events suggests that we are observing the collimated jet emission from these systems. The condition for launching a relativistic jet from some TDEs is unknown. One possibility is that a BH with a relatively rapid spin tends to launch a jet (Lei and Zhang, 2011). Quasi-periodic dip variations in the X-ray lightcurve of Sw 1644+57 may be caused by the precession of the accretion disk, and hence the jet axis (Lei et al., 2013). The jet activity may be related to the presence of strong magnetic flux threading the BH (Tchekhovskoy et al., 2014). For jetted TDEs, one would expect to observe misaligned TDEs, which may display bright radio emission due to the interaction between the jet and the medium. Such a system might have been observed in another event, IGR J12580+0134 (Irwin et al., 2015; Lei et al., 2016; Yuan et al., 2016).
- It was suggested by Hawking (1975) that black holes evaporate due to a quantum effect near the BH event horizon. If the initial density fluctuation in the early universe makes primordial BHs with a certain mass distribution that covers a wide mass range, those BHs with mass $\sim 5.1 \times 10^{14}$ g are evaporating at the present time (e.g. Rice and Zhang, 2017). At the final stage of evaporation, temperature increases in a runaway manner, which would lead to a very short GRB (Cline and Hong, 1992). Since the black hole temperature increases with time during the final phase of evaporation, high-energy photons lag behind low-energy photons, giving rise to a clear negative spectral lag feature (Ukwatta et al., 2016b). Searches for evaporating primordial BHs in the short GRB data have been carried out, but no robust evidence for the existence of such GRBs has been collected (Ukwatta et al., 2016a).