

Besides being the brightest electromagnetic sources in the universe, GRBs are also *believed* to be an important emitter of non-electromagnetic (or *multi-messenger*) emission in three channels: ultra-high-energy cosmic rays (UHECRs), high-energy neutrinos, and gravitational waves (GWs). This chapter is dedicated to the multi-messenger aspect of GRB physics. Section 12.1 discusses the motivations and issues of the suggestion that GRBs are the dominant sources of UHECRs. Section 12.2 reviews the mechanisms for producing high-energy neutrinos in a very wide energy band (from MeV to EeV) from GRBs, with the focus on the predicted PeV neutrinos and the non-detections from the *IceCube* observatory. Section 12.3 introduces the theoretical expectations of GRB–GW associations and the groundbreaking discovery of the association between the NS–NS merger gravitational wave source GW170817 and the short GRB 170817A.

## 12.1 GRBs as UHECR Sources

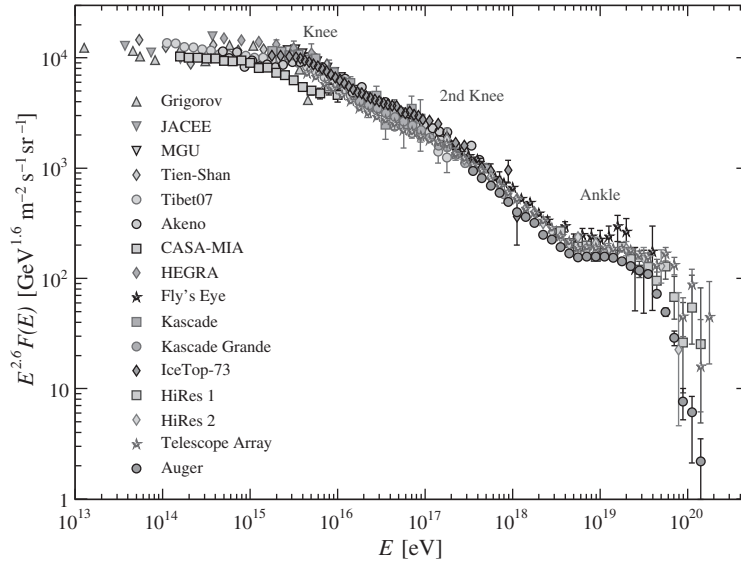
### 12.1.1 Cosmic Rays

Cosmic rays (CRs) may be defined as all relativistic particles originating from cosmic sources that arrive and interact with Earth’s atmosphere. A narrower version of the definition concerns the charged particles only, so that high-energy photons and neutrinos are not included. We apply this latter definition.

Cosmic rays were first discovered by Victor Hess in 1912 using a balloon experiment that measured an increase of the ionization flux at high altitudes. The discovery marked the birth of high-energy astrophysics, which was long before the discoveries of high-energy photons from astrophysical objects (e.g. the first X-ray astronomical source other than the Sun, Sco X-1, was discovered in 1962).

### All-Particle Spectrum

An *all-particle* spectrum of CRs covers more than 11 orders of magnitude in energy, from a few GeV up to a few times  $10^{20}$  eV. Below a few GeV, the CR spectrum is greatly influenced by solar activities, so is not included. Figure 12.1 shows the all-particle spectrum above  $10^{13}$  eV.



**Figure 12.1** The observed all-particle cosmic-ray spectrum as observed from Earth. From Patrignani and Particle Data Group (2016).

The all-particle CR spectrum has the following features:

- It is a nearly featureless power-law spectrum that covers more than 11 orders of magnitude. For an overall fit  $dN/dE \propto E^{-\alpha}$ , one has  $\alpha \sim 2.8$ .
- A closer scrutiny reveals multi-segments in the spectrum, so that the spectrum is better fit with a broken power law. Using the same convention  $dN/dE \propto E^{-\alpha}$ , one has  $\alpha \sim 1.6$ – $1.8$  below several GeV;  $\alpha \sim 2.7$  between several GeV to the *knee*;  $\alpha \sim 3.2$  between the *knee* and the *ankle*, and  $\alpha \sim 2.8$  above the *ankle*.
- The “*knee*” is at  $\sim 4 \times 10^{15}$  eV ( $= 4$  PeV). There is a slight steepening between the *knee* and the *ankle*, with a second *knee* at  $\sim 5 \times 10^{17}$  eV  $= 500$  PeV  $= 0.8$  EeV.
- The “*ankle*” is at  $\sim 5 \times 10^{18}$  eV ( $= 5$  EeV). The spectrum above the *ankle* becomes shallower and lighter (change of composition), suggesting a possible extragalactic origin of these CRs.
- The UHECRs are usually defined as CRs with energy  $> 10^{18}$  eV  $= 1$  EeV.
- Sometimes CRs above  $\sim 5 \times 10^{19}$  eV  $= 50$  EeV are called extreme-energy cosmic rays (EECRs).

## Propagation of Cosmic Rays

Unlike photons, CRs are charged particles and do not travel on null geodesics (straight lines in a flat space-time). They are deflected by magnetic fields between the source and Earth. The arrival directions of the CRs therefore essentially do not carry information of the emission source.

Our Milky Way has a mean magnetic field of  $B \sim 3 \mu\text{G}$ . The scale height of its disk is  $\sim 300$  pc. If the Larmor radius of a CR is shorter than this scale height, the CR is trapped in the Galaxy. Otherwise, the CR is unconfined to the Galaxy.

The Larmor radius of a CR can be expressed as (derived from  $ZeBv/c = \gamma mv^2/R_L$ )

$$R_L = \frac{cp}{ZeB} \simeq \frac{E}{ZeB} \approx (100 \text{ pc}) \frac{3 \mu\text{G}}{B} \frac{E}{Z \times 10^{18} \text{ eV}}, \quad (12.1)$$

where  $p = \gamma mv$  is momentum, and  $E = \sqrt{m^2 c^4 + c^2 p^2} \simeq cp$  is energy. Here  $\gamma \gg 1$  has been assumed. As a result, above  $E \sim$  a few times  $10^{18}$  eV, particles are not confined to the Galaxy, and they can propagate nearly rectilinearly from outside the Galaxy to reach Earth. UHECRs therefore should have an extragalactic origin. Observationally, the arrival directions of UHECRs are consistent with being isotropic (similar to GRBs). This is fully consistent with a cosmological origin of the UHECRs.

### 12.1.2 UHECRs: General Considerations

#### Bottom-Up, Top-Down, and Hybrid Models

There are three general types of models for UHECRs (e.g. Olinto (2000) and references therein). The *bottom-up* models state that UHECRs are accelerated from astrophysical sources through Fermi-like acceleration mechanisms (§4.4). The more exotic *top-down* models, on the other hand, involve the decay of putative very high mass relics from the early universe (e.g. topological defects such as cosmic strings, and super-heavy long-lived relic particles) within the framework of physics beyond the standard model of particle physics. A third group of models are *hybrid*, and combine physics beyond the standard model with known astrophysical processes. One example is to use physics beyond the standard model to generate ZeV neutrinos which propagate to the Galactic halo and local group before annihilating with the neutrino background to form UHECRs through hadronic Z-boson decay.

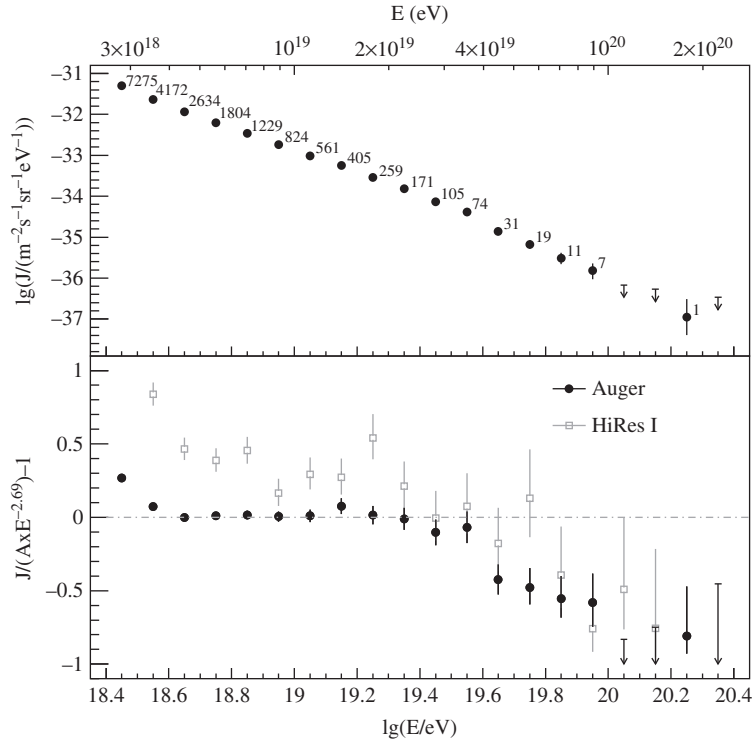
The last two models were motivated by the non-detection of the “GZK-feature” (see below) in the early years of the UHECR observations. Observations with the *HiRes* (Abbasi et al., 2008) and *Pierre Auger Observatory* (Abraham et al., 2008) revealed the existence of the suppression of cosmic-ray flux above  $(5\text{--}6) \times 10^{19}$  eV (50–60 EeV), i.e. the GZK feature (Fig. 12.2). This is consistent with the expectation of the bottom-up models, and the other two groups of exotic models are no longer motivated.

#### The Greisen–Zatsepin–Kuzmin (GZK) cutoff

The Greisen–Zatsepin–Kuzmin limit (Greisen, 1966; Zatsepin and Kuz’min, 1966) states that the source of CRs above  $5 \times 10^{19}$  eV cannot be from a distance much farther outside the “GZK” volume with radius  $l_{\text{GZK}} \sim 30\text{--}50$  Mpc. The reason is that UHECRs above this energy would interact with cosmic microwave background photons through  $p\gamma$  interaction at  $\Delta$ -resonance and hence lose energy at a characteristic distance of  $l_{\text{GZK}}$ . In order to be detected as UHECRs, these CRs must be accelerated within the GZK volume.

The  $\Delta$ -resonance condition is (Eq. (6.7))

$$\epsilon_p \epsilon_\gamma \geq 0.16 \text{ GeV}^2 \sim 1.6 \times 10^{17} \text{ eV}^2. \quad (12.2)$$



**Figure 12.2** The GZK feature as observed by Auger, confirming the *HiRes* results. From Abraham et al. (2008).

The typical CMB photon energy is  $\epsilon_{\gamma,p} \sim 2.8kT_{\text{CMB}} \sim 6.5 \times 10^{-4} \text{ eV}$  ( $T_{\text{CMB}} = 2.7 \text{ K}$ ). Photons at the Wien wing have higher energies and would reach the resonant condition first. In general, one can write  $\epsilon_{\gamma} = (6.5 \times 10^{-4} \text{ eV}) c_{\gamma}$ , so that the GZK energy reads

$$\epsilon_{p,\text{GZK}} \sim (2.5 \times 10^{20} \text{ eV}) c_{\gamma}^{-1} \sim 5 \times 10^{19} \text{ eV} (c_{\gamma}/5)^{-1}, \quad (12.3)$$

where  $c_{\gamma} \sim 5$  is a numerical factor obtained from detailed calculations (e.g. Dermer and Menon, 2009). The energy loss mean free path length can be estimated as

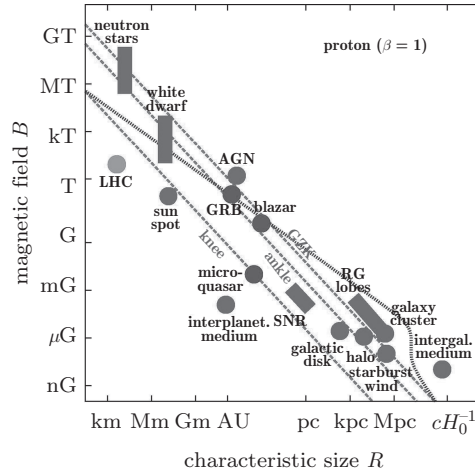
$$l_{\text{GZK}}^{-1} = \frac{1}{E} \frac{dE}{cdt} = \langle K_{p\gamma} \sigma_{p\gamma} n_{\gamma} \rangle, \quad (12.4)$$

where  $K_{p\gamma} = \Delta E/E$  is the energy loss fraction per interaction. Taking  $n_{\gamma} \sim 400 \text{ cm}^{-3}$  for the CMB,  $\sigma_{p\gamma} \sim 10^{-28} \text{ cm}^2$ , and  $K_{p\gamma} \sim 0.5$  for  $E \sim \epsilon_{p,\text{GZK}} \sim 5 \times 10^{19} \text{ eV}$ , we can estimate

$$l_{\text{GZK}}^{-1} = \frac{1}{E} \frac{dE}{dt} \simeq 2 \times 10^{-26} \text{ cm}^{-1}, \quad (12.5)$$

or  $l_{\text{GZK}} \sim 17 \text{ Mpc}$ . A more precise treatment gives

$$l_{\text{GZK}} \sim (30\text{--}50) \text{ Mpc}. \quad (12.6)$$



**Figure 12.3** The Hillas plot showing various sources that potentially accelerate UHECRs. From Aartsen et al. (2017b).

### Hillas Condition

One *necessary condition* to accelerate UHECRs is that the Larmor radius  $R'_L = E\beta/(ZeB')$  of the accelerated particles should fit within the size of the accelerator  $R'_s$ . Equating the two gives the maximum energy of the accelerated particles in the *comoving frame*, so that in the lab frame, where the accelerator is seen to move with a bulk Lorentz factor  $\Gamma$ , the maximum energy is (Hillas, 1984)

$$E_{\max} = \Gamma ZeB'R'_s, \quad (12.7)$$

where  $B'$ ,  $R'_L$ , and  $R'_s$  are the magnetic field strength, Larmor radius, and accelerator size in the comoving frame, respectively. In terms of the lab-frame radius from the central engine,  $r \sim R'_s\Gamma$ , and the lab-frame magnetic field  $B$ , one can write

$$E_{\max} = ZeBr/\Gamma. \quad (12.8)$$

Figure 12.3 shows a “Hillas” plot of different cosmological sources that can potentially accelerate particles to UHECRs. These include GRBs, hypernovae, AGNs, magnetars, and galaxy clusters.

### 12.1.3 High-Luminosity Long GRBs as UHECR Sources

In 1995, three papers proposed that typical (high-luminosity and long-duration) GRBs can be the dominant sources of UHECRs, but with different arguments: Waxman (1995) pointed out that the average UHECR generation rate in the Earth neighborhood is similar to the  $\gamma$ -ray generation rate from GRBs, and that with typical parameters of GRBs, UHECRs can be accelerated from GRB internal shocks. Vietri (1995) proposed that the GRB external forward shock can accelerate protons to energies above  $10^{20}$  eV. Milgrom and Usov (1995), on the other hand, pointed out two potential spatial associations between UHECRs and GRBs with the GRBs leading the UHECRs by several

months, and therefore proposed that GRBs might be the sources of UHECRs. Below we reiterate the arguments of Waxman (1995), which were also re-stated in Waxman (2004).

Observationally, the UHECR energy density at Earth is at the level of

$$\dot{\epsilon}_{\text{UHECR}} \sim 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1}. \quad (12.9)$$

The GRB photon energy density may be estimated as

$$\dot{\epsilon}_{\text{GRB}} \sim \rho_0 E_{\gamma, \text{iso}} = 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1} \left( \frac{\rho_0}{0.5 \text{ Gpc}^{-3} \text{ yr}^{-1}} \right) \left( \frac{E_{\gamma, \text{iso}}}{2 \times 10^{53} \text{ erg}} \right), \quad (12.10)$$

where  $\rho_0$  is the local GRB event rate density, and  $E_{\gamma, \text{iso}}$  is the observed typical isotropic  $\gamma$ -ray energy. Considering the beaming correction factor  $f_b \ll 1$ , the event rate density is  $\rho_0 f_b^{-1}$ , while the true  $\gamma$ -ray energy is  $E_{\gamma, \text{iso}} f_b$ . The beaming factor cancels out in the product, so that Eq. (12.10) remains valid.

This coincidence was adopted by Waxman to make a case for GRB–UHECR associations. One needs to keep in mind some *caveats*. This total  $\gamma$ -ray energy  $E_{\gamma, \text{iso}}$  may not be a good indicator of the proton energy  $E_{p, \text{iso}}$ . Assuming fast cooling, one has  $E_{p, \text{iso}} = E_{\gamma, \text{iso}}(\epsilon_p/\epsilon_e)$ . The energy of the UHECRs, on the other hand, is only a fraction of the total proton energy, i.e.  $E_{\text{UHECR}} = \xi E_p$ , where  $\xi = \xi_{\text{UHE}} \xi_{\text{esc}}$  takes into account the energy fraction in the UHECRs range ( $\xi_{\text{UHE}}$ ), as well as the probability ( $\xi_{\text{esc}}$ ) for the UHECRs to escape the GRB shocks. The coincidence between  $E_{\gamma}$  and  $E_{\text{UHECR}}$  demands  $(\epsilon_e/\epsilon_p)\xi \sim 1$ , which is not guaranteed but is achievable, especially if the proton energy index  $p_p \sim 2$  (so that the energy per decade is a constant) and the escape fraction  $\xi_{\text{esc}}$  is large.

One can show that internal shocks can accelerate protons to UHE.

The maximum proton Lorentz factor  $\gamma_{p, M}$  may be calculated via

$$t'_{\text{acc}} = \min(t'_{\text{dyn}}, t'_c). \quad (12.11)$$

Proton cooling is inefficient. In most cases, one has  $t'_c > t'_{\text{dyn}}$ , so that one can apply

$$t'_{\text{acc}} = t'_{\text{dyn}} \quad (12.12)$$

to define the maximum proton energy. This gives

$$\zeta \frac{\gamma'_{p, M} m_p c}{e B'} = \Gamma \delta t, \quad (12.13)$$

where  $\delta t$  is the observed variability time scale. The comoving magnetic field strength may be estimated via

$$\frac{B'^2}{8\pi} = \epsilon_B \frac{L_w}{4\pi r^2 c \Gamma^2}. \quad (12.14)$$

For fast cooling, the observed GRB luminosity can be written as  $L_{\gamma} = \epsilon_e L_w$ . So the comoving magnetic field is

$$B' = \left( \frac{\epsilon_B \cdot 2L_{\gamma}}{r^2 c \Gamma^2 \epsilon_e} \right)^{1/2} = \left( \frac{2L_{\gamma} \epsilon_B}{\Gamma^6 c^3 \delta t^2 \epsilon_e} \right)^{1/2}, \quad (12.15)$$

where the internal shock radius  $r = R_{\text{IS}} = \Gamma^2 c \delta t$  has been substituted. This finally gives

$$\gamma_{p,M} = \Gamma \gamma'_{p,M} \simeq \left( \frac{2L_\gamma \epsilon_B}{\epsilon_e c} \right)^{1/2} \frac{e}{\zeta \Gamma m_p c^2}, \quad (12.16)$$

or<sup>1</sup>

$$E_{p,\text{max}}^{\text{IS}} = \gamma_{p,M} m_p c^2 \simeq (4 \times 10^{20} \text{ eV}) \zeta^{-1} \left( \frac{\epsilon_{B,-1} L_{\gamma,52}}{\epsilon_{e,-1}} \right)^{1/2} \Gamma_{2.5}^{-1}. \quad (12.17)$$

The condition for the external shock to accelerate protons to UHECRs is more stringent. One may still apply Eq. (12.12) to do the estimation. The dynamical time scale is increased: the variability time scale  $\delta t \sim 10 \text{ ms}$  should be replaced by  $t_{\text{dec}} \sim 100 \text{ s}$ . On the other hand, the acceleration time scale is also increased due to the reduction of  $B'$ . Assuming a constant Lorentz factor from the internal shock radius to the deceleration radius (which is roughly satisfied for the fireball scenario), one has  $t \propto r$ . In order to keep the same condition of particle acceleration, one requires  $B' \propto r^{-1}$ . This is satisfied for a conical jet that has a dominant toroidal magnetic field, or has a random magnetic field with a constant  $\epsilon_B$ .<sup>2</sup> The UHECR condition may then be satisfied in the external reverse shock if the reverse shock region has a high magnetization parameter (say,  $\sigma \gtrsim 0.1$ ). The forward shock, on the other hand, typically has a much lower  $\epsilon_B$  and hence a lower  $B'$ , so that the UHECR condition may not be satisfied. Indeed, a critical condition for the external forward shock to accelerate UHECRs is that the forward shock region has a relatively high  $\epsilon_B$ . According to Vietri et al. (2003), the maximum proton energy accelerated from the external forward shock may reach the UHECR range only if  $\epsilon_B$  is close to unity.

Searches of high-energy neutrinos coincident with GRBs both in spatial direction and in time by the *IceCube* collaboration have consistently obtained null results. The current data already pose interesting constraints on the possibility that GRBs are the dominant UHECR sources. In particular, one version of the GRB UHECR theories, i.e. CRs escape from the GRB blastwave region in the form of neutrons (generated from  $p\gamma$  interactions), has been largely ruled out by the *IceCube* data (Abbasi et al., 2012). On the other hand, some other possibilities are still open. For example, if the GRB emission radius (also the proton acceleration radius) is larger than the traditional internal shock radius, the neutrino flux can be greatly reduced (due to the reduced  $p\gamma$  optical depth) without reducing the CR flux. The non-detection of neutrinos from GRBs therefore does not rule out the ability of GRBs to accelerate UHECRs (Zhang and Kumar, 2013).

### 12.1.4 Other GRB-Related Phenomena as UHECR Sources

Besides the traditional high-luminosity long GRBs, the possibility that some other GRB-related phenomena could be the sources of UHECRs has been also discussed in the literature.

<sup>1</sup> Notice that in this chapter  $E_p$  denotes the proton energy, not the peak energy in the GRB spectrum.

<sup>2</sup> As a result, the condition to accelerate UHECRs can also be satisfied for other internal emission models (e.g. the magnetic dissipation models that invoke a larger emission radius) if the GRB emission radius  $R_{\text{GRB}}$  is above the internal shock radius  $R_{\text{IS}}$ . For the models that invoke  $R_{\text{GRB}} < R_{\text{IS}}$ , e.g. the photosphere models, proton cooling may become important, so that  $t_c$  in Eq. (12.11) would define  $E_{p,\text{max}}$ . The UHECR condition may not be satisfied.

## Low-Luminosity GRBs

Equation (12.17) can be re-written as

$$E_{p,\max}^{(1)} \simeq (1.3 \times 10^{20} \text{ eV}) \zeta^{-1} \left( \frac{\epsilon_{B,-1} L_{\gamma,48}}{\epsilon_{e,-1}} \right)^{1/2} \Gamma_1^{-1}. \quad (12.18)$$

This means that if a GRB has a lower luminosity (e.g.  $L_\gamma \sim 10^{48} \text{ erg s}^{-1}$ ) and a lower Lorentz factor (e.g.  $\Gamma \sim 10$ ), it can also accelerate protons to ultra-high energies within the framework of the internal shock model.

Low-luminosity GRBs have the right luminosities and Lorentz factors. One question is whether internal shocks can be developed. If the observed emission all comes from the trans-relativistic shock breakout (Nakar and Sari, 2012), then there is no direct evidence of internal shocks. On the other hand, some low-luminosity GRBs may indeed be successful jets with weak activity, probably due to their relatively low Lorentz factors (e.g. Zhang et al., 2012b; Irwin and Chevalier, 2016). Within this picture the conventional internal shock scenario may operate, so that UHECRs may be accelerated from these events. The beaming-corrected event rate of LL-GRBs may be comparable to that of HL-GRBs (§2.5.3). As a result, LL-GRBs may be regarded as a competent source of UHECRs.

## Mildly Relativistic Type Ib/c Supernovae

Wang et al. (2007c) investigated the ability of the so-called “hypernovae”, those Type Ic supernovae that are associated with GRBs, to accelerate CRs in their mildly relativistic blastwaves. They found that, given typical parameters, protons can be accelerated to an energy as high as  $10^{19} \text{ eV}$ . Using the radio data of SN 2009bb, a transrelativistic radio SN without a GRB trigger (Soderberg et al., 2010), Chakraborti et al. (2011) measured  $B$  and  $R$  of the trans-relativistic SN blastwave, and found that the parameters satisfy the Hillas limit. They then suggested that Type Ib/c SNe with an engine-driven outflow can be a good candidate for generating UHECRs.

# 12.2 High-Energy Neutrinos from GRBs

## 12.2.1 General Considerations

In the energy dissipation sites where electrons are accelerated, protons are also expected to be accelerated. In a GRB outflow, there may be multiple sites where protons/ions are accelerated, including internal shocks, the dissipative photosphere, magnetic reconnection sites, and the external shocks (both forward and reverse shocks). These sites are usually also permeated by copious photons, so that  $p\gamma$  interactions would proceed at the  $\Delta$ -resonance (e.g. Eq. (6.2))

$$p\gamma \rightarrow (\Delta^+ \rightarrow) \begin{cases} n\pi^+ \rightarrow n\mu^+ \nu_\mu \rightarrow ne^+ \nu_e \bar{\nu}_\mu \nu_\mu, \\ p\pi^0 \rightarrow p\gamma\gamma. \end{cases}$$



A  $\pi^+$  meson eventually decays into three neutrinos ( $\nu_e$ ,  $\nu_\mu$ , and  $\bar{\nu}_\mu$ ) and one positron ( $e^+$ ). Accompanying these are photons generated from  $\pi^0$  decay. The branching ratio to the  $\pi^+$  and  $\pi^0$  channels is roughly 1 : 1 after considering all the  $\pi^+$  production channels besides  $\Delta^+$ -resonance (e.g. direct pion production and multiple pion production). The threshold condition for  $\Delta$ -resonance is (Eq. (6.7))

$$E_p E_\gamma \gtrsim \frac{m_\Delta^2 - m_p^2}{4} \left( \frac{\Gamma}{1+z} \right)^2 = 0.16 \text{ GeV}^2 \left( \frac{\Gamma}{1+z} \right)^2, \quad (12.19)$$

and the neutrino energy is

$$E_\nu \sim 0.05 E_p. \quad (12.20)$$

As a result, the characteristic energy of neutrinos generated from a GRB can be estimated as

$$E_\nu \gtrsim 8 \text{ GeV} \left( \frac{\Gamma}{1+z} \right)^2 \left( \frac{E_\gamma}{\text{MeV}} \right)^{-1}. \quad (12.21)$$

For different emission sites, one has different  $E_\gamma$  and  $\Gamma$  values, so that the characteristic neutrino energies would also be different.

Relative to  $p\gamma$  interactions, the  $pp/pn$  interactions are usually much less efficient, since relativistic motion requires low density in the jet. Nonetheless, these interactions can become important at small radii when the compactness of the jet is large, or as the jet is still inside the star so that jet protons can interact with the nucleons inside the stellar envelope.

### 12.2.2 Prompt Emission Region: PeV Neutrinos

#### Theoretical Predictions

A guaranteed target photon source for  $p\gamma$  interactions in a GRB is the GRB prompt emission itself. From Eq. (12.21), one gets  $E_\nu \sim \text{PeV}$  given  $\Gamma^2 \sim 10^5$ ,  $z \sim 1$ , and  $E_\gamma \sim 250 \text{ keV}$ , as long as protons can be accelerated to an energy greater than  $2 \times 10^{16} \text{ eV}$ . This requirement is not a stringent one, since various models predict a proton energy up to  $10^{20} \text{ eV}$  and higher (§12.1 above).

Waxman and Bahcall (1997) first discussed this process within the framework of the internal shock model. Over the years, the PeV neutrino flux from GRBs has been calculated both analytically and numerically by many authors (Waxman and Bahcall, 1997; Razzaque et al., 2003a,b; Dermer and Atoyan, 2003; Guetta et al., 2004; Murase and Nagataki, 2006; Murase et al., 2006; Gupta and Zhang, 2007a; He et al., 2012; Li, 2012; Gao et al., 2012; Zhang and Kumar, 2013). Below we describe the essence of these calculations in a general formalism without specifying the internal shock model, following Zhang and Kumar (2013).

One can approximately delineate the observed Band-function photon flux spectrum of a GRB as a broken power law:

$$F_\gamma(E_\gamma) = \frac{dN(E_\gamma)}{dE_\gamma}$$

$$= f_\gamma \begin{cases} \left(\frac{\epsilon_\gamma}{\text{MeV}}\right)^{\alpha_\gamma} \left(\frac{E_\gamma}{\text{MeV}}\right)^{-\alpha_\gamma}, & E_\gamma < \epsilon_\gamma, \\ \left(\frac{\epsilon_\gamma}{\text{MeV}}\right)^{\beta_\gamma} \left(\frac{E_\gamma}{\text{MeV}}\right)^{-\beta_\gamma}, & E_\gamma \geq \epsilon_\gamma. \end{cases} \quad (12.22)$$

Based on the  $p\gamma$ -threshold matching condition and a  $\pi^+$  cooling constraint, the observed neutrino number spectrum can be expressed as a three-segment broken power law (Waxman and Bahcall, 1997; Abbasi et al., 2010):

$$F_\nu(E_\nu) = \frac{dN(E_\nu)}{dE_\nu} = f_\nu \begin{cases} \left(\frac{\epsilon_{\nu,1}}{\text{GeV}}\right)^{\alpha_\nu} \left(\frac{E_\nu}{\text{GeV}}\right)^{-\alpha_\nu}, & E_\nu < \epsilon_{\nu,1}, \\ \left(\frac{\epsilon_{\nu,1}}{\text{GeV}}\right)^{\beta_\nu} \left(\frac{E_\nu}{\text{GeV}}\right)^{-\beta_\nu}, & \epsilon_{\nu,1} \leq E_\nu < \epsilon_{\nu,2}, \\ \left(\frac{\epsilon_{\nu,1}}{\text{GeV}}\right)^{\beta_\nu} \left(\frac{\epsilon_{\nu,2}}{\text{GeV}}\right)^{\gamma_\nu - \beta_\nu} \left(\frac{E_\nu}{\text{GeV}}\right)^{-\gamma_\nu}, & E_\nu \geq \epsilon_{\nu,2}, \end{cases} \quad (12.23)$$

where

$$\alpha_\nu = p + 1 - \beta_\gamma, \quad \beta_\nu = p + 1 - \alpha_\gamma, \quad \gamma_\nu = \beta_\nu + 2, \quad (12.24)$$

and  $p$  is the proton spectral index defined by  $N(E_p)dE_p \propto E_p^{-p}dE_p$ . The indices  $\alpha_\nu$  and  $\beta_\nu$  are derived by assuming that the neutrino flux is proportional to the proton flux (which has index  $p$ ) and the  $p\gamma$  optical depth  $\tau_{p\gamma} \propto n_\gamma(E_\gamma)$  (which has an index  $-\alpha_\gamma + 1$  or  $-\beta_\gamma + 1$ ) of each proton. The latter is valid when the fraction of proton energy that goes to pion production, i.e.  $f \equiv 1 - (1 - \langle\chi_{p \rightarrow \pi}\rangle)^{\tau_{p\gamma}}$ , is proportional to  $\tau_{p\gamma}$  ( $\langle\chi_{p \rightarrow \pi}\rangle \simeq 0.2$  is the average fraction of energy transferred from protons to pions). This is roughly valid when  $\tau_{p\gamma} < 3$ , which is generally satisfied in most GRB emission models. The break  $\epsilon_{\nu,1}$  is defined by the break in the photon spectrum. The second break  $\epsilon_{\nu,2}$  is defined by the  $\pi^+$  synchrotron cooling effect (Waxman and Bahcall, 1997). In general, one can write (Zhang and Kumar, 2013)

$$\epsilon_{\nu,1} = \epsilon_{\nu,1}^0 \min(1, (\tau_{p\gamma}^p/3)^{1-\beta_\gamma}), \quad (12.25)$$

where

$$\epsilon_{\nu,1} = \epsilon_{\nu,1}^0 = 7.3 \times 10^5 \text{ GeV} (1+z)^{-2} \Gamma_{2.5}^2 \epsilon_{\gamma, \text{MeV}}^{-1}, \quad (12.26)$$

$$\epsilon_{\nu,2} = 3.4 \times 10^8 \text{ GeV} (1+z)^{-1} \epsilon_B^{-1/2} L_{w,52}^{-1/2} \Gamma_{2.5}^2 R_{\nu,14}, \quad (12.27)$$

and

$$\tau_{p\gamma}^p \equiv \tau_{p\gamma}(E_p^p) \simeq \frac{\Delta R'}{\lambda'_{p\gamma}(E_p^p)} = 0.8 L_{\gamma,52} \Gamma_{2.5}^{-2} R_{\nu,14}^{-1} \epsilon_{\gamma, \text{MeV}}^{-1}, \quad (12.28)$$

where  $\lambda'_{p\gamma}(E_p^p)$  is the comoving proton mean free path for  $p\gamma$  interaction at  $E_p^p$  ( $E_p^p$  is the energy of protons that interact with peak energy photons at  $\Delta$ -resonance),  $\Delta R'$  is the comoving width of the jet,  $R$  denotes the distance of the proton acceleration site (rather than the photon emission site if the two sites are different) from the central engine,  $\epsilon_B$  is the fraction of dissipated jet energy in magnetic fields, and  $L_w$  is the luminosity of the dissipated wind.

Two important parameters define the neutrino luminosity from a GRB: the  $p\gamma$  optical depth  $\tau_{p\gamma}$  (which defines how efficiently each proton generates neutrinos), and the proton luminosity  $L_p$  (which defines how many protons are emitting per unit time). This latter parameter is not directly observed or probed. One may use the  $\gamma$ -ray luminosity as a proxy by defining a dimensionless parameter

$$f_{\gamma/p} \equiv \frac{L_\gamma}{L_p}, \quad (12.29)$$

or its reciprocal

$$f_{p/\gamma} \equiv \frac{L_p}{L_\gamma} = f_{\gamma/p}^{-1} = f_p. \quad (12.30)$$

This latter definition is denoted as  $f_p$  in the *IceCube* collaboration papers (Abbasi et al., 2010, 2011, 2012; Aartsen et al., 2015, 2016, 2017a,b).

One should also consider a correction factor for the energy dependence of the neutrino emission (Li, 2012),

$$\hat{f}_p \equiv \frac{\int_{E_{p,1}}^{E_{p,2}} dE_p E_p^2 dN(E_p)/dE_p}{\int_{E_{p,\min}}^{E_{p,\max}} dE_p E_p^2 dN(E_p)/dE_p} \simeq \frac{\ln(E_{v,2}/E_{v,1})}{\ln(E_{p,\max}/E_{p,\min})} \quad (\text{for } p = 2), \quad (12.31)$$

where  $E_{p,1}$  and  $E_{p,2}$  are proton energies corresponding to  $E_{v,1}$  and  $E_{v,2}$ , respectively (Eq. (12.20)), and  $E_{p,\max}$  and  $E_{p,\min}$  are the maximum and minimum proton energies, respectively. Notice it is different from the commonly defined  $f_p$  (Eq. (12.30)). With  $f_{\gamma/p}$  and  $\hat{f}_p$ , one can then normalize the neutrino spectrum with the total photon fluence (Abbasi et al., 2010):

$$\int_0^\infty dE_\nu E_\nu F_\nu(E_\nu) = \frac{1}{8} \frac{\hat{f}_p}{f_{\gamma/p}} \left[ 1 - (1 - \langle \chi_{p \rightarrow \pi} \rangle)^{\tau_{p\gamma}^p} \right] \int_{1 \text{ keV}}^{10 \text{ MeV}} dE_\gamma E_\gamma F_\gamma(E_\gamma). \quad (12.32)$$

The coefficient  $1/8$  is the product of  $1/4$  (four leptons share the energy of one  $\pi^+$ ) and  $1/2$  (on average roughly half of  $p\gamma$  interactions go to the  $\pi^+$  channel).

In this treatment, both the Lorentz factor  $\Gamma$  and the neutrino emission site  $R_\nu$  are taken as independent parameters. If one adopts the internal shock radius  $R_\nu = R_{\text{IS}} = \Gamma^2 c \delta t$ , the above treatment is reduced to the standard internal shock treatment.

### *IceCube* Upper Limits and Implications

Over the years, the *IceCube* collaboration has been searching for high-energy neutrino signals coincident with GRBs in time and direction, and reached progressively more stringent non-detection upper limits on the neutrino flux from GRBs (Abbasi et al., 2010, 2011, 2012; Aartsen et al., 2015, 2016). In 2012, the *IceCube* upper limit was claimed to be at least a factor of 3.7 lower than the most optimistic theoretical predictions according to the internal shock model (Abbasi et al. 2012, upper left panel of Fig. 12.4). Later more detailed calculations (Li, 2012; Hümmer et al., 2012; He et al., 2012) showed that this limit still allowed the internal shock model in a certain parameter space. The internal shock model was in any case disfavored (He et al., 2012) if the GRB Lorentz factor is correlated to the

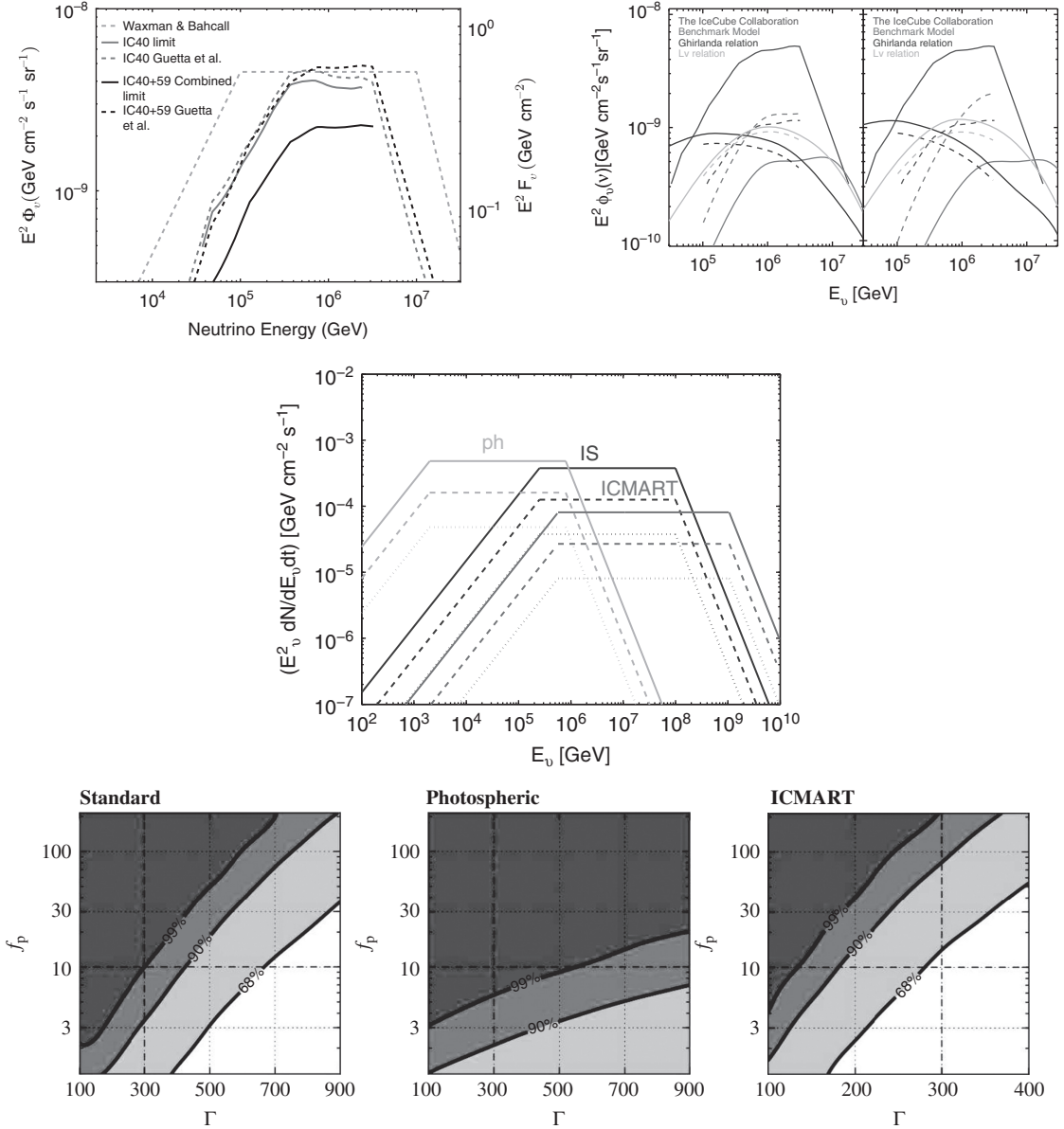
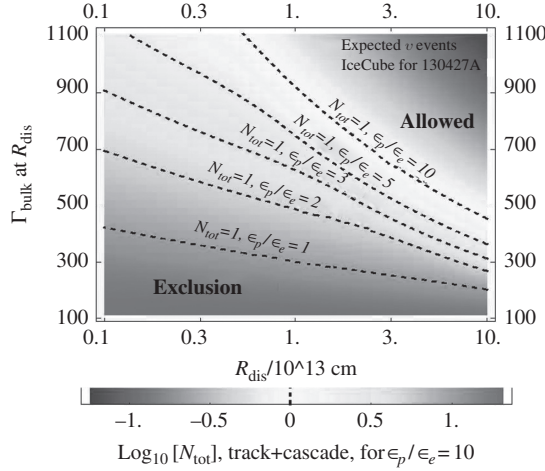


Figure 12.4

The PeV neutrino spectra of GRBs. *Top left*: The 2012 *IceCube* upper limit compared against early model predictions. From Abbasi et al. (2012). *Top right*: The 2012 *IceCube* upper limit confronted with numerical results of the internal shock model. Reproduced from Figure 4 in He et al. (2012) with permission. ©AAS. *Middle*: Model-dependent PeV neutrino flux. From Zhang and Kumar (2013). *Bottom*: The constraints on the internal shock (left), photosphere (middle), and ICMART (right) models using six-year *IceCube* data. Here  $f_p = 1/f_{\gamma/p}$ . Reproduced from Figure 10 in Aartsen et al. (2017a) with permission. ©AAS.



**Figure 12.5** Constraints on the parameter space with the non-detection of neutrinos from GRB 130427A. Reproduced from Figure 1 in Gao et al. (2013d) 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.

isotropic luminosities, as suggested by the GRB data (Liang et al., 2010; Ghirlanda et al., 2011; Lü et al., 2012), see upper right panel of Fig. 12.4.

The non-detection of neutrinos from GRBs requires an increase in either the neutrino emission site radius  $R_v$  or  $f_{\gamma/p}$  (i.e. a decrease in  $f_p$ ) (Zhang and Kumar, 2013). A larger  $R_v > R_{IS}$  is expected if GRBs are high- $\sigma$  outflows with the magnetic dissipation site at large radii, e.g. in the ICMART model (middle panel of Fig. 12.4). Alternatively, one may increase  $f_{\gamma/p}$  to be above the conventional value of 0.1 (or lower  $f_p$  to be below 10). This is unconventional for shock models, since  $f_{\gamma/p}$  is essentially  $\epsilon_e/\epsilon_p$  for fast cooling, and GRB prompt emission and afterglow data do not support an  $\epsilon_e$  value much greater than 0.1 in the prompt emission region. The photosphere models may predict a large  $f_{\gamma/p}$  value due to the large photon flux advected with the jet below the photosphere. Another possibility for suppressing neutrino flux in the dissipative photosphere models may be to suppress proton acceleration above  $2 \times 10^{16}$  eV. This may not be easy in view that internal shocks or magnetic reconnections likely develop below the photosphere so that Fermi acceleration inevitably accelerates protons to large enough energy to emit neutrinos.

The nearby, very bright GRB 130427A did not show a positive PeV neutrino signal. This non-detection makes even tighter constraints on the internal shock model and the photosphere model of GRBs (Gao et al. 2013d, Fig. 12.5).

The *IceCube* collaboration (Aartsen et al., 2015, 2016) applied progressively more stringent non-detection upper limits to constrain the parameter space of the three models discussed in Zhang and Kumar (2013). The six-year limit by Aartsen et al. (2016) rules out both the photosphere model and the internal shock model for the nominal values  $(\Gamma, f_p) = (300, 10)$ , and the allowed parameter space has greatly shrunk. The ICMART model, on the other hand, still comfortably satisfies the non-detection limit with the nominal parameters. This result is consistent with the prompt emission models that invoke magnetic energy dissipation at large radii, as discussed in Chapter 9.

### 12.2.3 External Shocks: EeV Neutrinos

Protons should also be accelerated in both external forward and reverse shocks. These protons interact with photons generated in the external shock regions to produce neutrinos. The typical neutrino energy can again be estimated from Eq. (12.21).

At the deceleration time, synchrotron emission flux density ( $F_\nu$ ) peaks in the X-ray band for the forward shock and in the optical/IR band for the reverse shock. Plugging in  $E_\gamma \sim \text{keV}$  and  $\text{eV}$ , respectively, and taking  $\Gamma^2 \sim 10^4$  (in the early deceleration phase), one can estimate  $E_\nu \sim 10^{16} - 10^{19} \text{ eV}$ , which is in the EeV range. The large emission radius of the external shock is compensated by a larger photon number density for a lower  $E_\gamma$ , so that the  $\tau_{p\gamma}$  optical depth is not small. The proton luminosity may also be comparable to the case of PeV neutrinos, if  $p \sim 2$  is valid (i.e. protons have equal power per decade in energy). Observationally, however, it is challenging to detect these EeV neutrinos. Given the similar neutrino emission power, the number of neutrinos is much smaller, making them more challenging to detect. Detailed studies of GRB neutrinos from external shocks can be found in, e.g. Waxman and Bahcall (2000), Dai and Lu (2001), Dermer (2002), Li et al. (2002), and Fan et al. (2005b).

### 12.2.4 Jets in Collapsars: TeV Neutrinos

Before a GRB jet breaks out of the progenitor star, or if a jet never breaks out of the star, high-energy neutrinos can be generated inside the star. One may envisage “internal” shocks being developed near the jet head, which is moving slowly. Protons are accelerated in these shocks and interact with photons to produce neutrinos. The seed photons are the thermal photons trapped in the jet and the cocoon, which is likely in the multi-keV energy range. One may estimate the temperature of the cocoon by balancing the ram pressure and radiation pressure, i.e.  $aT^4 \sim \rho v^2 \sim \rho c^2$  (the jet advances with at most a trans-relativistic speed). Estimating  $\rho c^2 \sim (E/\Gamma_0)/(0.01\pi R_*^3) \sim 10^{17} \text{ g cm}^{-1} \text{ s}^{-2}$  (jet approximated as a cylinder of length  $R_*$  and radius  $0.1R_*$ ) for  $\Gamma_0 \sim 300$ ,  $R_* \sim 10^{11} \text{ cm}$ , one therefore estimates  $E_\gamma \sim kT \sim 5 \text{ keV}$ .

Applying Eq. (12.21), with  $E_\gamma \sim 5 \text{ keV}$  and  $\Gamma/(1+z) \sim 1$  (inside the star the jet is at most trans-relativistic), one gets the typical neutrino energies  $E_\nu \gtrsim 1.6 \text{ TeV}$ . Considering the distributions of target photon energies and proton energy as well as the broad, energy-dependent  $p\gamma$  interaction cross sections, one gets the neutrino energy extending to tens of TeV, which is within the detection reach of *IceCube*. One interesting aspect of these neutrinos is that they can be produced in both successful and failed GRBs. As a result, a nearby neutrino burst may probe a failed jet inside a core-collapsing star.

The TeV neutrino emission from successful and choked jets was first proposed by Mészáros and Waxman (2001) and verified by numerical calculations (Razzaque et al., 2003a,b, 2004b). One important physical ingredient addressed later is that shocks inside the stars are radiation mediated. Copious pair production at the shock region prevents a strong shock jump condition. This would smear up the velocity difference between the upstream and the downstream, and therefore suppress cosmic-ray acceleration (e.g. Levinson and Bromberg, 2008). Murase and Ioka (2013) showed that the TeV neutrino component is not

significant for successful GRBs – which is consistent with the non-detection of neutrinos from GRBs – but may be important for low-power (low-luminosity) GRBs or ultra-long GRBs presumably formed from blue supergiants.

The  $pp$  interactions were also considered by Razzaque et al. (2003b), who found that their contribution to the neutrino fluxes becomes important only when the GRB progenitor has an extended hydrogen envelope. These interactions may be relevant to ultra-long GRBs if they have a blue supergiant progenitor.

### 12.2.5 Proton–Proton Interactions in Dissipative Photospheres: TeV Neutrinos

When studying neutrino emission from a baryon-dominated dissipative photosphere, Murase (2008) and Wang and Dai (2009) independently discovered that, besides the conventional  $p\gamma$  interactions,  $pp$  interactions could also be important due to the high compactness of the photosphere zone. This would produce a TeV neutrino component, and in the meantime suppress the PeV component due to the rapid cooling of protons in the dissipative photosphere region. For these baryon-dominated models, internal shocks would in any case develop at larger radii. Protons accelerated from the internal shocks would also interact with these bright photosphere emission photons (along with the photons generated in the internal shocks) and produce PeV neutrinos. The predicted PeV neutrino flux could be comparable to the flux level of the internal shock model that assumes both  $\gamma$ -ray photons and protons are generated from internal shocks (Zhang and Kumar, 2013). Such a high flux is greatly constrained by the *IceCube* upper limit data (Aartsen et al., 2015, 2016, 2017a).

### 12.2.6 Proton–Neutron Decoupling and Collisions: GeV Neutrinos

For neutron-rich ejecta, protons and neutrons can decouple at different radii, coasting with different Lorentz factors (§7.3.5). If the relative speed between the two components is larger than the pion threshold 145 MeV (readily satisfied in an unsteady relativistic ejecta, since the threshold energy is smaller than the rest mass of protons and neutrons), inelastic collisions occur, and pions, muons, and neutrinos are produced. The neutrino signal is in the multi-GeV range (Bahcall and Mészáros, 2000). In this energy range, the atmospheric neutrino background is very strong, which makes it difficult to detect these neutrinos. Kashiyama et al. (2013) found that neutron–proton conversion during inelastic nuclear collisions can serve as a particle acceleration mechanism. This would give a high-energy tail of particles which extend this neutrino emission component to the 10–100 GeV range, facilitating detection with *IceCube* and other high-energy neutrino detectors (Murase et al., 2013).

### 12.2.7 Central Engine: MeV Neutrinos

At the GRB central engine, it is expected that copious neutrinos and anti-neutrinos are generated in the hyper-accreting neutrino-dominated accretion flow (NDAF). These “thermal” neutrinos can reach a luminosity of  $10^{50}$ – $10^{51}$  erg s<sup>−1</sup> peaking at  $\sim 10$  MeV (Liu

et al., 2016b). Only extremely nearby GRBs can have such neutrinos detected by the next generation MeV neutrino detectors such as Hyper-Kamiokande (Hyper-K). The estimated detection rate is  $\sim 0.10\text{--}0.25$  per century by Hyper-K (Liu et al., 2016b).

The MeV neutrinos are generally associated with all core-collapse supernovae. On 23 February 1987, three neutrino detectors (Super Kamiokande, IMB, and Baksan) registered altogether 24 neutrinos within less than 13 seconds from SN 1987A, a nearby supernova in the Large Magellanic Cloud, approximately 168 000 light years away. The SN-associated MeV neutrinos are brighter than NDAF-associated neutrinos by  $\sim 2$  orders of magnitude (Liu et al., 2016b). Detection of a bright MeV neutrino burst followed by a low-level  $\sim 10$  MeV neutrino emission plateau may point towards the existence of a NDAF at the central engine (even if the GRB does not beam towards Earth). Alternatively, a magnetar GRB central engine would show an extended bright MeV neutrino emission phase due to the cooling of the newborn magnetar. For rare, bright nearby events, thermal neutrino signatures can be used to diagnose the central engine of GRBs.

## 12.3 Gravitational Waves from GRBs

### 12.3.1 Gravitational Waves

Gravitational waves (GWs) are predicted from the General Theory of Relativity. They were indirectly “detected” by the observations of double neutron star systems in our Galaxy, such as PSR 1913+16 (e.g. Taylor and Weisberg, 1989). The first GW source due to BH–BH merger, GW150914, was directly detected by the *Advanced LIGO* team in 2015 (Abbott et al., 2016c). This was followed by several more detections of BH–BH merger systems (e.g. Abbott et al., 2016b) and the first detection of the NS–NS merger system GW170817 with the associations of the short GRB 170817A and its multi-wavelength counterpart (Abbott et al., 2017d,b).

Some important properties of GWs include the following:

- GWs are “ripples” in space-time. In contrast to *dipole radiation* for electromagnetic waves, which requires variation of the Coulomb field, GWs are generated via *quadrupole radiation*, which requires the variation of gravitational *tidal* fields. Whereas the static tidal fields (gradient of gravity) fall off as  $\propto r^{-3}$ , the transverse fields fall off as  $\propto r^{-1}$ , which can propagate to infinity.
- GWs travel with the speed of light. This has been confirmed by the joint detection of GW170817 and GRB 170817A, with the  $\gamma$ -ray emission delayed by  $\sim 1.7$  s with respect to the GW signal (Abbott et al., 2017b).
- Unlike electromagnetic waves, for which people measure the “power” of the received waves (flux), the GW detectors directly measure the dimensionless amplitude (or strain) of the wave itself. Since the strain  $h \propto \Delta L/L \propto r^{-1}$  ( $L$  is the length of the detector, and  $\Delta L$  is the change of the length), the capability of detecting distant GWs does not fall as steeply as that of detecting EM waves (flux  $\propto r^{-2}$ ).



- There are two polarization modes: “+” and “×”.

The quadrupole radiation formula can be written as

$$-\dot{E} = L_{\text{GW}} = \frac{G}{5c^5} \langle \ddot{I}_{ij} \ddot{I}^{ij} \rangle, \quad (12.33)$$

where

$$I_{ij} = \int \rho (x_i x_j - r^2 \delta_{ij}/3) d^3x \quad (12.34)$$

is the quadrupole-moment tensor,  $\ddot{I}_{ij}$  and  $\ddot{I}^{ij}$  are the third time derivatives for the covariant and contravariant components of the tensor, and the average is over a period of oscillation.

The simplest system that can emit GWs is a rotating rod. For a rod of length  $L$  and mass  $M$  rotating around its mid-point with angular velocity  $\Omega$ , the rate of GW radiation energy is

$$L_{\text{GW}} = -\dot{E} = \frac{2G}{45c^5} M^2 L^4 \Omega^6 \simeq (1.2 \times 10^{-61} \text{ erg s}^{-1}) M^2 L^4 \Omega^6 \quad (12.35)$$

in c.g.s. units.

### 12.3.2 Compact Object Mergers as Gravitational Wave Emitters

An astronomical binary system has a non-zero  $\ddot{I}_{ij}$ , and therefore is a natural GW emitter. For a binary system with masses  $m_1$  and  $m_2$  with a separation  $a$ , the first-order post-Newtonian GW luminosity can be written as

$$L_{\text{GW}} = \frac{32}{5} \frac{c^5}{G} \left( \frac{G^5 \mu^2 M^3}{c^{10} a^5} \right) f(e), \quad (12.36)$$

where  $M = m_1 + m_2$  is the total mass,  $\mu = m_1 m_2 / (m_1 + m_2)$  is the *reduced mass*, and

$$f(e) = \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1 - e^2)^{7/2}} \quad (12.37)$$

is a correction factor for eccentricity  $e$ , which equals unity when  $e = 0$  (circular orbit).

In terms of orbital angular frequency  $\Omega = (GM/a^3)^{1/2}$ , the GW luminosity can be written as (Maggiore, 2008)

$$L_{\text{GW}} = \frac{32}{5} \frac{c^5}{G} \left( \frac{GM_c \Omega}{c^3} \right)^{10/3} f(e) = \frac{32}{5} \frac{c^5}{G} \left( \frac{GM_c \omega_{\text{GW}}}{2c^3} \right)^{10/3} f(e), \quad (12.38)$$

where

$$M_c = \mu^{3/5} M^{2/5} = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}} \quad (12.39)$$

is the *chirp mass*, and

$$\omega_{\text{GW}} = 2\Omega = 2 \left( \frac{GM}{a^3} \right)^{1/2} \quad (12.40)$$

is the angular frequency of the emitted GWs, which is twice the orbital angular frequency. Noting that  $GM_c\Omega/c^3$  is dimensionless, one can see that the GW luminosity scales with a characteristic luminosity defined by fundamental constants:

$$L_{\text{GW},c} = \frac{c^5}{G} \simeq 3.6 \times 10^{59} \text{ erg s}^{-1} \simeq (2.0 \times 10^5) M_\odot \text{ s}^{-1}. \quad (12.41)$$

For two objects with equal masses, i.e.  $m_1 = m_2 = m$ , one has  $\mu = m/2$  and  $M_c = m/2^{1/5} \simeq 0.87m$ . Equation (12.38) can be reduced to

$$L_{\text{GW}} = \frac{2}{5} \frac{c^5}{G} \left( \frac{r_s}{a} \right)^5 f(e), \quad (12.42)$$

where  $r_s \equiv 2Gm/c^2$  is the Schwarzschild radius of each mass. One immediate inference is that the GW luminosity increases rapidly as  $a$  decreases, i.e.  $(r_s/a)$  approaches the maximum from below before the merger. A strong GW emitter requires that  $a$  is not much greater than  $r_s$ , which explains why BH–BH, NS–BH, and NS–NS mergers are the brightest GW sources. For two Schwarzschild BHs,  $(r_s/a)$  may be as large as  $1/2$ , so that  $L_{\text{GW}}$  would approach  $\sim (1/80)L_{\text{GW},c}$  in the first-order post-Newtonian approximation.<sup>3</sup> For two spinning BHs,  $(r_s/a)$  may even approach unity, resulting in an even larger  $L_{\text{GW}}$ . The radius of a  $1.4M_\odot$  NS is about  $10/(3 \times 1.4) \sim 2.4$  times  $r_s$ , so  $(r_s/a)$  can approach a value smaller than the BH–BH merger systems but still large enough for detection. As a result, NS–NS mergers have weaker GW emission, and hence can be detected only from a smaller distance than BH–BH mergers.

The frequency of GWs increases rapidly towards the final coalescence, which, for the first-order post-Newtonian approximation, can be written as (for  $e = 0$ )

$$f_{\text{GW}}(\tau) = \frac{\omega_{\text{GW}}(\tau)}{2\pi} = \frac{1}{\pi} \left( \frac{5}{256} \frac{1}{\tau} \right)^{3/8} \left( \frac{GM_c}{c^3} \right)^{-5/8} \quad (12.43)$$

$$\simeq 134 \text{ Hz} \left( \frac{1.21M_\odot}{M_c} \right)^{5/8} \left( \frac{1 \text{ s}}{\tau} \right)^{3/8}, \quad (12.44)$$

where  $\tau = t_{\text{coa}} - t$  is the time before coalescence. In the second equation,  $M_c = 1.21M_\odot$  (chirp mass of two NSs with  $m = 1.4M_\odot$ ) has been adopted. Equivalently, one can also write

$$\tau \simeq 2.18 \text{ s} \left( \frac{1.21M_\odot}{M_c} \right)^{5/3} \left( \frac{100 \text{ Hz}}{f_{\text{GW}}} \right)^{8/3}. \quad (12.45)$$

Based on  $L_{\text{GW}}$ , one can solve the orbital period of a binary system  $P_{\text{orb}}$  at any time  $t$  given the initial period  $P_0$  at  $t = 0$ , i.e.

$$P_{\text{orb}}(t) = \left( P_0^{8/3} - \frac{8}{3} kt \right)^{3/8}, \quad (12.46)$$

where

$$k \equiv \frac{96}{5c^5} (2\pi)^{8/3} (GM_c)^{5/3}. \quad (12.47)$$

<sup>3</sup> In reality, near the coalescence of the two BHs, a higher order post-Newtonian approximation is needed to match the full GR results, and the maximum GW luminosity is smaller than this value.

The orbital decay rate is

$$\dot{P}_{\text{orb}} = -\frac{192\pi}{5c^5} \left( \frac{2\pi G}{P_{\text{orb}}} \right)^{5/3} M_c^{5/3} f(e). \quad (12.48)$$

The (average) evolutions of  $a$  and  $e$  read

$$\left\langle \frac{da}{dt} \right\rangle = -\frac{64}{5} \frac{G^3 \mu M^2}{c^5 a^3} f(e), \quad (12.49)$$

$$\left\langle \frac{de}{dt} \right\rangle = -\frac{304}{15} e \frac{G^3 \mu M^2}{c^5 a^4 (1-e^2)^{5/2}} \left( 1 + \frac{121}{304} e^2 \right), \quad (12.50)$$

respectively.

The *gravitational strain* of GWs, defined as the fraction of distortion in the length of detectors induced by the fluctuating gravitational acceleration, can be written generally as

$$h \equiv \sqrt{h_+^2 + h_\times^2} = \left( \frac{32\pi G T_{01}}{c^3 \Omega^2} \right)^{1/2}, \quad (12.51)$$

where  $h_+$  and  $h_\times$  are two polarization modes of the GWs, and  $T_{01}$  is the (0,1) component of the energy–momentum tensor at the detector. For a maximally emitting source, one has

$$T_{01} = \left( \frac{c^5}{4\pi G r^2} \right), \quad (12.52)$$

so that

$$h = \frac{\sqrt{8}c}{\Omega r} \simeq 2.7 \times 10^{-19} \left( \frac{\Omega}{\text{kHz}} \right)^{-1} \left( \frac{r}{100 \text{ Mpc}} \right)^{-1}, \quad (12.53)$$

where  $r$  is the distance to the source. More generally, for compact star mergers, it can be written as (Thorne, 1987)

$$\begin{aligned} h &= 0.237 \frac{\mu^{1/2} M^{1/3}}{r f_{\text{GW}}^{1/6}} \\ &= 4.1 \times 10^{-22} \left( \frac{\mu}{M_\odot} \right)^{1/2} \left( \frac{M}{M_\odot} \right)^{1/3} \left( \frac{r}{100 \text{ Mpc}} \right)^{-1} \left( \frac{f_{\text{gw}}}{100 \text{ Hz}} \right)^{-1/6}. \end{aligned} \quad (12.54)$$

This is within reach with the sensitivity of *Advanced LIGO* and *Advanced Virgo*, suggesting that NS–NS mergers, and certainly NS–BH and BH–BH mergers, are detectable from cosmological distances.

There are three phases of GW radiation for a compact-star-merger system: *inspiral*, *merger*, and *ring down*. The waveforms of GW emission for BH–BH mergers are well studied, but they are much more complicated for NS–NS mergers, since the merger phase depends on complicated physics such as the equation of state of nuclear-density matter.

The first gravitational wave source GW150914 (Abbott et al., 2016c) was detected by the two detectors of *Advanced LIGO* on 14 September 2015, at 09:50:45 UTC. The signal sweeps upwards in frequency from 35 to 250 Hz with a peak strain  $h = 1.0 \times 10^{-21}$ . The signal is consistent with the coalescence of two stellar-mass BHs with masses  $36_{-4}^{+5} M_\odot$  and  $29_{-4}^{+4} M_\odot$ , at a distance  $410_{-180}^{+160}$  Mpc, or a redshift  $z = 0.09_{-0.04}^{+0.03}$ . The merger gave

birth to a BH of mass  $62^{+4}_{-4}M_{\odot}$ , with a total mass  $3.0^{+0.5}_{-0.5}M_{\odot}$  radiated in the form of GWs.

The first NS–NS merger gravitational wave source GW170817 (Abbott et al., 2017d) was detected on 17 August 2017, at 12:41:04 UTC by both *Advanced LIGO* and *Advanced Virgo* with a combined signal-to-noise ratio of 32.4. The inferred component masses of the binary fall into the range of NSs. Restricting the component spins to the range of known binary NSs, the component masses are constrained in the range  $1.17\text{--}1.60M_{\odot}$ , with the total mass of the system  $2.74^{+0.04}_{-0.01}M_{\odot}$ . There was a short GRB 170817A and an optical transient associated with the GW source, and a host galaxy (NGC 4993) at  $D_L \sim 40$  Mpc was discovered (Abbott et al., 2017e).

From these detections, the estimated BH–BH merger event rate density is (Abbott et al., 2016a)  $\sim 9\text{--}240 \text{ Gpc}^{-3} \text{ yr}^{-1}$ , and the estimated NS–NS merger event rate density is (Abbott et al., 2017d)  $\sim 1540^{+3200}_{-1220} \text{ Gpc}^{-3} \text{ yr}^{-1}$ .

### 12.3.3 Gravitational Waves from GRBs

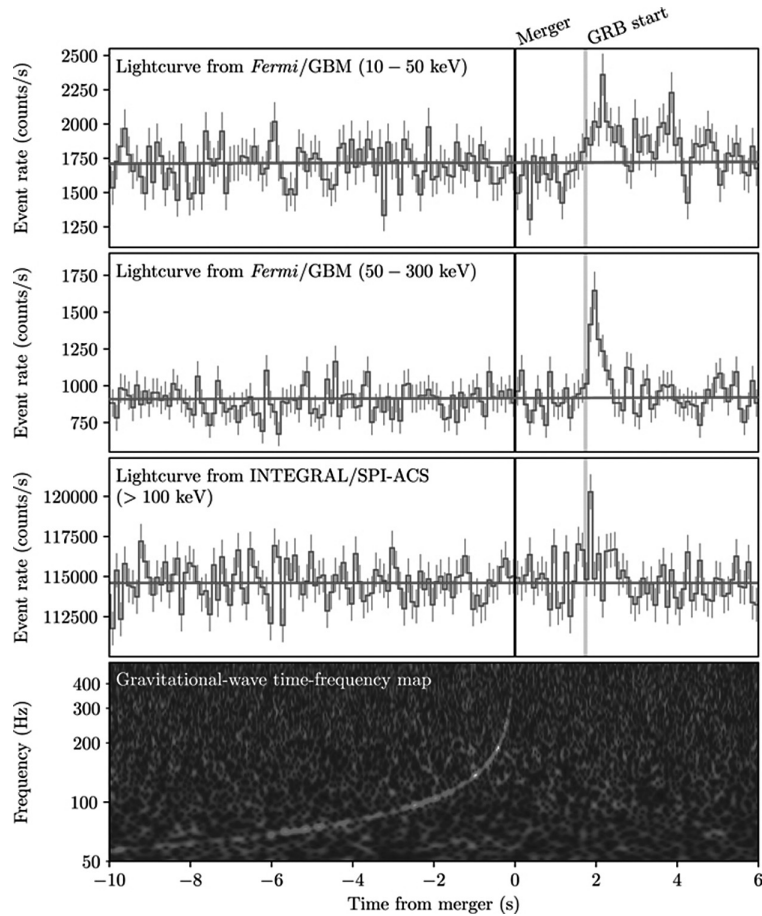
The leading progenitor systems of short GRBs are NS–NS and NS–BH mergers (see §10.3 for detailed discussion). If these models are correct, a GW chirp signal due to inspiral and merger should proceed a short GRB. The joint detection of GW170817 and the short GRB 170817A (Abbott et al., 2017b) has firmly verified this progenitor model at least for some (low-luminosity) short GRBs.

The reason that NS–NS and NS–BH mergers are expected to give rise to short GRBs is that the disrupted NSs can provide ample material to be accreted to the newborn BH formed after the merger (see §11.2 for details). Based on such reasoning, BH–BH mergers are not expected to produce short GRBs. As a consequence, the putative  $\gamma$ -ray signal GW150914-GBM reported by the *Fermi*/GBM team (Connaughton et al., 2016) indeed posted great challenges to theorists. Arguments suggesting that the signal may be spurious have been raised (e.g. Greiner et al., 2016). Some ideas producing GRBs associated with BH–BH mergers have been proposed, which have to invoke either matter (e.g. Loeb, 2016; Perna et al., 2016; Janiuk et al., 2017) or electromagnetic fields (e.g. Zhang, 2016; Liebling and Palenzuela, 2016) around the merging BHs.

Besides compact star mergers, massive star core collapse may also be associated with GW emission (e.g. Kobayashi and Mészáros, 2003; Bartos et al., 2013). For a collapse, since the core is rapidly rotating, GWs can be emitted if the collapse proceeds non-axisymmetrically (e.g. Ott, 2009). At the GRB central engine, bar or fragmentation instabilities in the disk can also excite GWs (Liu et al., 2017a). For both massive star core collapse and NS–NS mergers, if the central engine is a supra-massive, rapidly rotating magnetar, a secular bar-mode instability may develop, which would give rise to strong GW emission lasting from minutes to an hour (Corsi and Mészáros, 2009; Fan et al., 2013c; Gao et al., 2016; Lasky and Glampedakis, 2016; Gao et al., 2017a).

#### GW170817/GRB 170817A/SSS17a (AT 2017gfo)/NGC 4993 Association

GW170817 was detected by the *Advanced LIGO* and *Advanced Virgo* detectors on 17 August 2017, with the merger time at 12:41:04 UTC (Abbott et al., 2017d). This was

**Figure 12.6**

Joint, multi-messenger detection of GW170817 and GRB 170817A. Reproduced from Figure 2 in Abbott et al. (2017b) 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.

followed by the independent detection of GRB 170817A  $\sim 1.7$  seconds later by the *Fermi* GBM detector (Abbott et al., 2017b; Goldstein et al., 2017) (Fig. 12.6). From the GW signal only, the source was initially localized to a sky region of  $31 \text{ deg}^2$  at a luminosity distance  $40^{+8}_{-8} \text{ Mpc}$ , and the component masses were later measured to be in the range  $0.86\text{--}2.26 M_{\odot}$ . The trigger initiated a multi-wavelength, multi-messenger observational campaign (Abbott et al., 2017e), which led to the discovery of a bright optical transient (initially named SSS17a, later with the IAU identification AT 2017gfo) associated with the galaxy NGC 4993 at  $\sim 40 \text{ Mpc}$ . The optical transient was independently discovered by multiple teams, and showed an early blue component (e.g. Coulter et al., 2017; Evans et al., 2017; Nicholl et al., 2017) that faded within 48 hours and later a red component that evolved over  $\sim 10$  days (e.g. Shappee et al., 2017; Chornock et al., 2017). The later component appeared lanthanide rich as revealed from the spectroscopic observations (e.g. Smartt et al., 2017; Pian et al., 2017; Tanvir et al., 2017; Shappee et al., 2017). The entire

optical phenomenology is consistent with the prediction of a kilonova/macronova powered by the r-process (Li and Paczyński, 1998; Metzger et al., 2010; Metzger, 2017; Kasen et al., 2017). X-ray and radio emission were discovered at the transient’s position at  $\sim 9$  days (Troja et al., 2017; Margutti et al., 2017) and  $\sim 16$  days (Hallinan et al., 2017), respectively.

The short GRB 170817A associated with GW170817 is an otherwise normal (relatively weak and soft) short GRB within the *Fermi* short GRB population. The association of it with GW170817 and NGC 4993 at  $D_L \sim 40$  Mpc shows it having a much lower isotropic luminosity and energy compared with other short GRBs. Based on the short GRB observations before this event (typically with isotropic luminosity above  $\sim 10^{50}$  erg s $^{-1}$ ), the detection of a short GRB at such a small distance was a surprise. It is likely the observer’s viewing angle is off the short GRB jet axis. This is consistent with the gravitational wave data that suggested a viewing angle  $\leq 28$  deg (Abbott et al., 2017d). The X-ray (Troja et al., 2017) and radio (Hallinan et al., 2017) emission was detected later, which may be related to a mildly relativistic outflow.

Compared with the estimated NS–NS merger event rate density (Abbott et al., 2017d) of  $1540^{+3200}_{-1220}$  Gpc $^{-3}$  yr $^{-1}$ , the short GRB event rate density above  $\sim 10^{47}$  erg s $^{-1}$  derived from the detection of GRB 170817A is at least  $190^{+440}_{-160}$  Gpc $^{-3}$  yr $^{-1}$ , which is consistent with but can be up to a factor of a few smaller. Assuming that all NS–NS mergers produce short GRBs and vice versa, this suggests that there might be a few more low-luminosity short GRBs in the *Fermi*/GBM archives that are intrinsically similar to GRB 170817A (Zhang et al., 2018a).

### 12.3.4 EM Counterparts of Compact-Star-Merger GW Sources

The triumph of the joint detection of the GW170817/GRB 170817A/SSS17a (AT 2017gfo)/NGC 4993 system opened the era of multi-messenger astrophysics.

In general, there are three ultimate energy sources to power EM counterparts of compact-star-merger GW sources: (1) the *gravitational energy* of the post-merger debris may be tapped through accretion to power a relativistic jet, which, when beaming towards Earth would appear as a short GRB; (2) the *nuclear energy* of the radioactive material launched during NS–NS mergers or NS–BH mergers would power a nearly isotropic macronova/kilonova; (3) in the case of a supra-massive or stable neutron star remnant following a NS–NS merger, the *spin energy* of the central object may be injected into the merger environment, making additional radiation signatures.

The properties of the EM counterparts of GW signals due to compact star mergers depend on the merger remnant.

#### The Case of a BH Post-Merger Remnant

For NS–BH mergers or NS–NS mergers that form a BH directly, the post-merger product is a BH. There are three widely discussed EM counterparts (Fig. 12.7; Metzger and Berger, 2012):

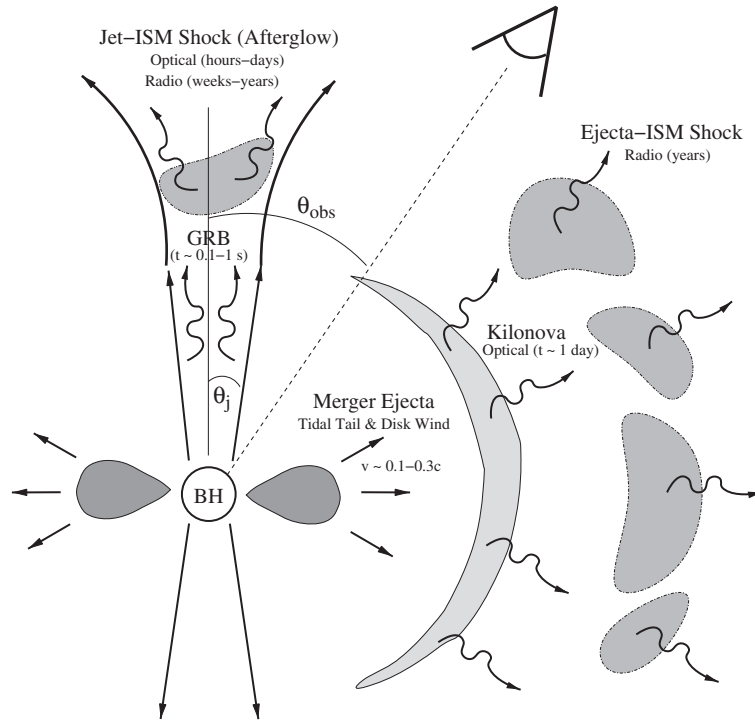


Figure 12.7

A cartoon picture showing various EM counterparts of GW sources due to compact star mergers, if the post-merger remnant is a hyper-accreting black hole. Reproduced from Figure 1 in Metzger and Berger (2012) with permission. ©AAS.

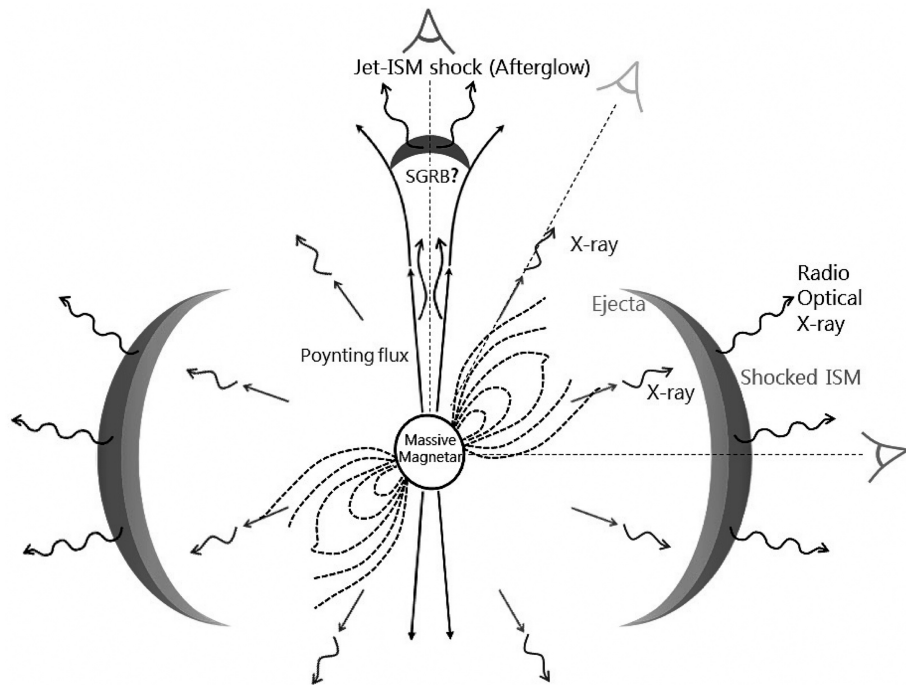
- *Short GRB*: A short GRB is likely generated along the angular momentum axis of the merger system due to rapid accretion of matter from a torus around the BH. This component is believed to be collimated. For a large enough viewing angle, one may not be able to detect this component. For GW170817, however, a low-luminosity short GRB 170817A was detectable at a viewing angle  $\leq 28^\circ$  (Abbott et al., 2017b), suggesting that the solid angle for a detectable short GRB is very wide. It is possible that short GRB jets are structured, so that a large-angle observer would observe the wing of the jet that is less energetic than the core (e.g. Lamb and Kobayashi, 2016, 2017; Lazzati et al., 2018). The model seems to be relevant to the GW170817–GRB 170817A association (e.g. Xiao et al., 2017; Zhang et al., 2018a; Lazzati et al., 2018).
- *Macronova/kilonova*: This signal (see §10.3.3) arises from the r-process nuclear heating of the neutron-rich material dynamically launched during the merger process or ejected from the neutron-rich disk. Since this component is expected to be essentially isotropic, one expects to observe these transients at all viewing angles, so that one might have such transients associated with the GW events but not with a short GRB. The optical transient SSS17a/AT 2017gfo associated with GW170817 was identified as a macronova/kilonova. One expects that a similar event may be discovered to be associated with every NS–NS or NS–BH merger event detected in the future.

- *Radio afterglow of the macronova/kilonova:* As the matter of the macronova/kilonova eventually gets decelerated by an ambient gas, an afterglow (similar to the afterglow of a successful jet) would be produced. Detailed calculations (e.g. Nakar and Piran, 2011; Piran et al., 2013) showed that the emission peaks in radio in the time scale of years. It was also termed a “radio flare”. Due to the non-relativistic nature of the ejecta, this radio emission component is faint.

### The Case of a NS Post-Merger Remnant

In the case of a NS–NS merger giving rise to a supra-massive or stable NS as the merger product, the EM counterpart features are richer. A newborn massive NS due to a NS–NS merger likely spins rapidly, and an  $\alpha$ – $\Omega$  dynamo may operate to amplify the magnetic field into the magnetar range (e.g. Duncan and Thompson, 1992). As a result, the merger product may be a millisecond magnetar. In general, there might be several types of EM counterparts for NS–NS mergers with a millisecond magnetar post-merger product (Fig. 12.8, Gao et al. 2013b):

- *Short GRB:* A short GRB may also be launched after the merger through accretion or differential rotation, as discussed in §11.3.3.



**Figure 12.8**

A cartoon picture showing various EM counterparts of GW sources due to compact star mergers, if the post-merger remnant is a rapidly spinning magnetar. From Gao et al. (2013b).



- *Off-beam X-ray emission:* If the X-ray plateau observed in some short GRBs is indeed powered by a post-merger magnetar (Rowlinson et al., 2010, 2013; Lü et al., 2015), a direct inference is that the magnetar wind is essentially isotropic so that the X-ray emission may be detected in all directions (Zhang, 2013). The lightcurve of the signal depends on the viewing direction (Sun et al., 2017): within the so-called “free zone” where X-ray emission is not blocked by the launched ejecta, the luminosity may peak around  $\sim 10^{49.6}$  erg s $^{-1}$ . In the so-called “trapped zone”, where the X-ray emission is initially blocked, the X-rays may emerge later after the mergernova becomes transparent. The luminosity of these X-ray transients may peak around  $\sim 10^{46.4}$  erg s $^{-1}$ . Also the mergernova itself may produce bright X-rays for extreme magnetar parameters (Siegel and Ciolfi, 2016a).<sup>4</sup>
- *Mergernova:* With a magnetar central engine, the neutron-rich ejecta launched before and during the merger would be continuously energized by the magnetar wind (in the form of a Poynting flux) and the X-rays due to magnetic wind dissipation. The brightness of the mergernova is therefore enhanced (Yu et al., 2013; Metzger and Piro, 2014). A magnetar-powered mergernova may exist in GRBs 080503 (Gao et al., 2015c), 050724, 070714B, 061006 (Gao et al., 2017b), and probably 130603B as well (Fan et al., 2013a).
- *Broad-band afterglow of the mergernova:* Energy injection from the magnetar central engine would also enhance the afterglow emission of the mergernova. Under certain conditions the ejecta dynamics may even reach the trans-relativistic regime, giving rise to bright emission in broad bands (X-rays, optical, and radio) (Gao et al., 2013b).
- *Fast radio burst:* If the merger product is a supra-massive neutron star that collapses hundreds of seconds after the merger, an FRB may be produced at the collapsing time (Zhang, 2014) due to the abrupt ejection of the NS magnetosphere when the newly formed BH expels the closed field lines (Falcke and Rezzolla, 2014). The FRB is detectable only in the jet direction, so one expects a GW/GRB/FRB association (Zhang, 2014).

The optical counterpart of GW170817 (e.g Coulter et al., 2017; Nicholl et al., 2017; Chornock et al., 2017; Smartt et al., 2017; Evans et al., 2017; Shappee et al., 2017; Pian et al., 2017) can be modeled with a macronova/kilonova without additional energy injection. A search for  $\gamma$ -ray emission before and after the short GRB 170817A also did not reveal any signature of an underlying magnetar (Zhang et al., 2018a). A search for gravitational wave signals from a putative hyper-massive or supra-massive neutron star after the merger led to negative results, even though the upper limits are still much higher than the predicted level (Abbott et al., 2017f). In general, the data of GW170817 are consistent with a BH-post-merger product. The case of a long-lived NS is not ruled out, even though the dipole  $B$  field of the putative NS cannot be too high (Ai et al., 2018). The optical transient, AT2018gfo, associated with GW170817 can be also interpreted as being partially powered by a long-lived low- $B$  NS (Yu et al., 2018; Li et al., 2018).

<sup>4</sup> The calculations of Siegel and Ciolfi (2016a,b) did not consider adiabatic cooling of the ejecta due to expansion (D. M. Siegel, 2016, personal communication), so that their X-ray luminosities of the mergernovae are over-predicted.