The Astrophysical Journal, 837:33 (21pp), 2017 March 1 https://doi.org/10.3847/1538-4357/837/1/33

Multi-messenger Light Curves from Gamma-Ray Bursts in the Internal Shock Model

Mauricio Bustamante[1,2], Jonas Heinze[3], Kohta Murase[4,5,6], and Walter Winter[3]
[1] Center for Cosmology and AstroParticle Physics (CCAPP), The Ohio State University, Columbus, OH 43210, USA; bustamanteramirez.1@osu.edu
[2] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[3] Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, D-15738 Zeuthen, Germany; walter.winter@desy.de, jonas.heinze@desy.de
[4] Center for Particle and Gravitational Astrophysics, The Pennsylvania State University, University Park, PA16802, USA; murase@psu.edu
[5] Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA
[6] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA16802, USA
Received 2016 June 21; revised 2016 September 14; accepted 2016 September 30; published 2017 February 28

Abstract

Gamma-ray bursts (GRBs) are promising as sources of neutrinos and cosmic rays. In the internal shock scenario, blobs of plasma emitted from a central engine collide within a relativistic jet and form shocks, leading to particle acceleration and emission. Motivated by present experimental constraints and sensitivities, we improve the predictions of particle emission by investigating time-dependent effects from multiple shocks. We produce synthetic light curves with different variability timescales that stem from properties of the central engine. For individual GRBs, qualitative conclusions about model parameters, neutrino production efficiency, and delays in high-energy gamma-rays can be deduced from inspection of the gamma-ray light curves. GRBs with fast time variability without additional prominent pulse structure tend to be efficient neutrino emitters, whereas GRBs with fast variability modulated by a broad pulse structure can be inefficient neutrino emitters and produce delayed high-energy gamma-ray signals. Our results can be applied to quantitative tests of the GRB origin of ultra-high-energy cosmic rays, and have the potential to impact current and future multi-messenger searches.

Key words: astroparticle physics – cosmic rays – gamma-ray burst: general – methods: numerical – neutrinos

1. Introduction

The most energetic particles discovered—ultra-high-energy cosmic rays (UHECRs), above 10^9 GeV, and high-energy astrophysical neutrinos, above 100 TeV—are detected with regularity, but their sources remain unknown. The extreme physical conditions required to reach these energies restrict the potential source classes. Gamma-ray bursts (GRBs) are one such class. They are the most luminous transient electromagnetic astrophysical phenomena: they emit gamma-rays up to ∼100 GeV, concentrated in only a few tens of seconds. Their observed high luminosities and inferred high particle densities make them prime candidate sources of UHECRs and neutrinos.

One of the leading explanations of GRB emission is the fireball model (Rees & Mészáros 1992). In it, a central engine—likely, a black hole—injects plasma blobs with different relativistic speeds into the jet flow. When they collide with one another, they create mildly relativistic shocks in which their kinetic energy is transformed into internal energy that is radiated away as high-energy particles (Paczynski & Xu 1994; Rees & Mészáros 1994). If the jets contain enough baryons, then protons (Milgrom & Usov 1995; Vietri 1995; Waxman 1995) and nuclei (Murase et al. 2008; Wang et al. 2008) can be accelerated by the shocks into UHECRs which, upon interacting with source photons, create pions and other mesons. Their decays produce PeV gamma-rays and neutrinos (Waxman & Bahcall 1997).

In recent years, IceCube has started testing the hypothesis of GRBs as sources of high-energy neutrinos (Abbasi et al. 2012; Aartsen et al. 2015a, 2016). The simplest versions of the fireball model are now in tension with the data. In these, the neutrino flux prediction for a burst is based on the emission from a single representative plasma blob collision in the jet—this is known as the one-zone approach.

While high-luminosity GRBs cannot account for the diffuse high-energy astrophysical neutrino signal detected by IceCube (Tamborra & Ando 2015; Aartsen et al. 2016), they persist as interesting objects because of the following two reasons. First, they can still be the sources of UHECRs. This can happen if UHECR emission and neutrino emission are not tightly correlated (Baerwald et al. 2013), a possibility that we consider in the present work. Second, they are attractive targets for neutrino telescopes because timing and directional gamma-ray information can be used to reduce neutrino backgrounds. They are ideal sites to look for joint gamma-ray and neutrino signals.

In this paper, we focus on the connection between gamma-rays, UHECRs, and neutrinos in individual GRBs. Unlike most of the existing literature, we consider the multi-messenger emission from a multitude of different plasma collisions along the jet, each one occurring under different physical conditions—this is known as the multi-zone approach.

We construct synthetic GRB light curves—curves of photon rate versus time—that successfully reproduce generic features of real ones. We identify the properties of the central engine that lead to such light curves, and study the consequences for different messengers. For example, light curves dominated by broad pulses overlaid with fast variability imply a "disciplined" engine that ejects shells with little spread in speed at any given time. In such cases, the neutrino production efficiency can be low, and high-energy gamma-ray signals can reach Earth delayed with respect to low-energy signals. Conversely, light curves with no broad pulse structure hint at likely efficient neutrino emitters. We also test the robustness of the assumptions going into the minimal neutrino flux estimate found in Bustamante et al. (2015).

This paper is organized as follows. In Section 2, we comment on GRB emission models. In Section 3, we introduce our multi-zone simulation and a sample of GRBs computed with it. In Section 4, we compute the gamma-ray and neutrino light curves

for each of them. In Section 5, we calculate their associated quasi-diffuse neutrino fluxes and show that different types of particles are emitted from different regions in the jet. In Section 6, we show that, for some GRBs, delays between the light curves in different energy bands are possible, and why. We summarize and conclude in Section 7. Appendix A contains a detailed presentation of the multi-zone collision model used in our simulations. Appendix C contains a discussion of the impact of alternative assumptions for the collision dynamics.

2. One-zone versus Multi-zone Emission

GRBs are conceivably fueled by matter accretion, either in the collapse of a massive star—long-duration bursts, lasting >2 s—or in the merging of two neutron stars or a neutron star and a black hole—short-duration bursts, lasting <2 s. The central engine, likely a newly formed black hole, emits two relativistic matter jets in opposite directions. When one of them points toward Earth, we might detect it as a GRB. We focus on long-duration bursts because they have higher gamma-ray luminosity ($\sim 10^{52}$ erg s^{-1}), represent $\sim$75% of observed bursts, and their emission mechanism has been more deeply studied.

In the internal shock scenario of the fireball model, internal plasma collisions within the jet account for particle emission during the initial, or prompt, phase of the burst, which typically lasts 10–100 s. Ultimately, the jet reaches the circumburst medium, triggering a late emission phase—the afterglow. We will focus exclusively on the prompt phase, where PeV neutrino emission is well-motivated. This is the energy range where existing water-Cherenkov neutrino telescopes, like IceCube and ANTARES, are most sensitive. However, the internal shock scenario, while successful, is not without issues, which we introduce below.

The GRB prompt emission mechanism has been under debate for years (see reviews by Piran 1999; Mészáros 2006; Kumar & Zhang 2014; Mészáros et al. 2015; Pe'er 2015). In the classical model, prompt gamma-ray emission in the MeV range is attributed to optically thin synchrotron emission from non-thermal electrons accelerated at internal shocks (Rees & Mészáros 1994), possibly supplemented by a thermal component from the fireball. Indeed, there are observational indications of such a thermal component (see, e.g., Pe'er 2012; Guiriec et al. 2011). However, this classical internal shock model has difficulty explaining the data, including the low-energy spectral index (Preece et al. 1998), radiation efficiency (Kumar 1999; Zhang et al. 2007), and spectral energy relations (Amati et al. 2002; Yonetoku et al. 2004).

Among alternative theories, photospheric emission models became popular after the *Fermi* satellite was launched (Thompson 1994; Rees & Mészáros 2005; Pe'er et al. 2006; Ioka et al. 2007; Giannios 2008; Beloborodov 2010; Lazzati et al. 2013). In them, the bulk of the prompt emission is attributed to quasi-thermal emission from below the photosphere, i.e., from the region where gamma-rays are unable to escape due to high optical depth to electron-photon scattering. Amid different versions of photospheric emission models, dissipative photosphere models have been discussed both theoretically and observationally, due to their appealing implications (Beloborodov 2013; Hascoet et al. 2013; Vurm et al. 2013). Some authors consider the superposition of pure thermal emission to explain the prompt emission spectrum (Lundman et al. 2013).

Magnetic reconnection models have also been largely of interest (Mészáros & Rees 1997; Lyutikov & Blandford 2003; Zhang & Yan 2011; Bošnjak & Kumar 2012; McKinney & Uzdensky 2012), due to certain advantages (Kumar & Zhang 2014), though detailed microphysics is much more uncertain. For example, they can explain the high polarization observed in some GRBs (Gotz et al. 2009; Yonetoku et al. 2012) and the absence of bright thermal emission (Zhang & Pe'er 2009; Gao & Zhang 2015).

Even though the classical internal shock model has several theoretical issues, internal shocks may still play an important role in the dissipation of the outflowing kinetic energy. For instance, some photospheric emission models require sub-photospheric dissipation that is often attributed to internal shocks (Rees & Mészáros 2005). Magnetic reconnection may also be driven by shocks beyond the photosphere (Zhang & Yan 2011). Even in the optically thin internal shock model, the observed spectra can be explained by synchrotron emission with some modifications such as stochastic acceleration of electrons in the shock downstream (Bykov & Mészáros 1996; Murase et al. 2012; Asano & Terasawa 2015).

In the internal shock model, shell collisions inside the jet are responsible for the shape of the GRB light curves. They typically have a fast time variability t_v, of $\sim$10–100 ms (Bhat 2013; Golkhou et al. 2015), frequently superposed on top of slower pulse structure (Ramirez-Ruiz & Fenimore 1999; Nakar & Piran 2002b). Since the light curve reflects the particularities of the central emitter and jet propagation (Nakar & Piran 2002a; Lopez-Camara et al. 2016; Zhang et al. 2016), no two are equal, though they can be classified on the basis of their morphology.

One-zone collision models assume average shell properties—such as an average speed or Lorentz factor $\langle\Gamma\rangle$—derived from gamma-ray observations. Neutrino emission is computed for a single representative collision (Dermer & Atoyan 2003; Guetta et al. 2004) and is scaled by the total number of collisions—roughly, $T_{90}/t_v \sim 1000$, with $T_{90} \sim 10$ s the burst duration—to yield the flux for the whole burst, implying that all collisions are identical (see, e.g., Asano 2005).

However, it is unrealistic that all collisions occur at the same radius. The main reason is that all shells would have to be emitted with precisely tuned speeds. A different reason comes from the fact that energy deposited in the afterglow cannot be too large; therefore, a sizable part of the energy must be dissipated during the preceding prompt phase. Because dissipation of kinetic energy into radiation scales with the difference of the Lorentz factors of the colliding shells, this implies that the differences must be large. In fact, there is considerable spread in the inferred values of bulk Lorentz factors of GRBs observed by the *Fermi* satellite (Ackermann et al. 2012). Assuming for the Lorentz factors a broad or bi-modal distribution leads to efficient dissipation of kinetic energy in the prompt phase (Kumar 1999; Zhang et al. 2007). As a consequence, collisions occur at many positions throughout the jet, from below the photosphere—where gamma-rays cannot escape—to the circumburst medium, and that particle emission will be different for each collision.

The impact of a distribution[7] of Lorentz factors or emission radii have been addressed qualitatively (Murase & Nagataki 2006) and quantitatively (Guetta et al. 2001a, 2001b; Bustamante et al. 2015; Globus et al. 2015). For example, Bustamante et al. (2015)

[7] Guetta et al. (2004), Becker et al. (2006), Stamatikos (2006), and Baerwald et al. (2012) studied the complementary problem of calculating, in the one-zone approach, the expected neutrino flux from individual GRBs using their observed electromagnetic properties, and the impact of distributions of model parameter values on the quasi-diffuse flux.

predicted a minimal quasi-diffuse prompt neutrino flux from super-photospheric collisions, at the level of $\sim 10^{-11}$ GeV cm^{-2} s^{-1} sr^{-1} per flavor, which is presently below the sensitivity of IceCube. In later sections, we discuss the relationship among different types of messengers in multiple-collision models, focusing on how the properties of the central engine impact the GRB light curves and neutrino emission efficiency.

IceCube recently discovered a diffuse astrophysical flux of neutrinos between about 30 TeV and 2 PeV (Aartsen et al. 2013a, 2013b, 2014a, 2014b, 2015b, 2015c). However, no individual neutrinos are associated to known GRBs (Abbasi et al. 2012; Aartsen et al. 2015a, 2016). The resulting upper limit, extrapolated from a catalog of bursts, is about an order of magnitude below the diffuse flux. Thus, it is unlikely that classical high-luminosity GRBs are the main origin of the observed IceCube neutrinos (Laha et al. 2013; Murase & Ioka 2013; Bustamante et al. 2015). (However, low-luminosity classes of GRBs can describe the diffuse neutrino flux without violating the stacking limits (Murase et al. 2006; Gupta & Zhang 2007), subject to their assumed source density.)

Nevertheless, this does not preclude GRBs from being the dominant sources of UHECRs. So far, results of these observational searches have been interpreted mostly in the context of one-zone models (Waxman & Bahcall 1997; Guetta et al. 2004; He et al. 2012; Hümmer et al. 2012; Li 2012). While the plain one-zone ansatz with commonly assumed parameters with values that are favorable for the UHECR explanation, is starting to be challenged by IceCube data (He et al. 2012; Hümmer et al. 2012), it is not fully ruled out yet.

Another important aspect is that the relationship between neutrino and cosmic-ray emission depends on the UHECR escape mechanism (Baerwald et al. 2013). It has been shown that, depending on the parameter values, a fraction of cosmic rays must directly escape from the sources without neutrino production (Baerwald et al. 2013). Earlier approaches focused on the emission of neutrons only (Ahlers et al. 2011), which leads to a stringent relation between the neutrino and cosmic-ray fluxes. Consequently, bounds on GRB neutrinos translate into bounds on the GRB parameter space; see Baerwald et al. (2015) for the bounds in the one-zone model. Further quantitative tests are necessary to test the possibility that UHECRs mainly come from GRBs, especially taking into account the effects from multiple collisions.

3. Simulating Collisions in a GRB Jet

We study the prompt emission phase of the GRB in the internal shock scenario by simulating collisions of many propagating plasma blobs. Our simulations are based on Kobayashi et al. (1997) and Daigne & Mochkovitch (1998). See Appendix A for details.

Figure 1 illustrates the evolution of the blobs along the jet. To simplify the computations, we take them to be spherical shells. The simulation starts with the central engine emitting $N_{\rm sh} \sim 1000$ shells. Each one propagates at a different relativistic speed; typically, they have Lorentz factors of $\Gamma \sim 100$. After some time, they catch up to one another, collide inelastically, and merge into new shells. At collisionless shocks generated during the collisions, a fraction of the kinetic energy of the colliding shells is used for particle acceleration and creation, and becomes the internal energy of the newly created shells. The new shells cool instantly by emitting high-energy particles, continue

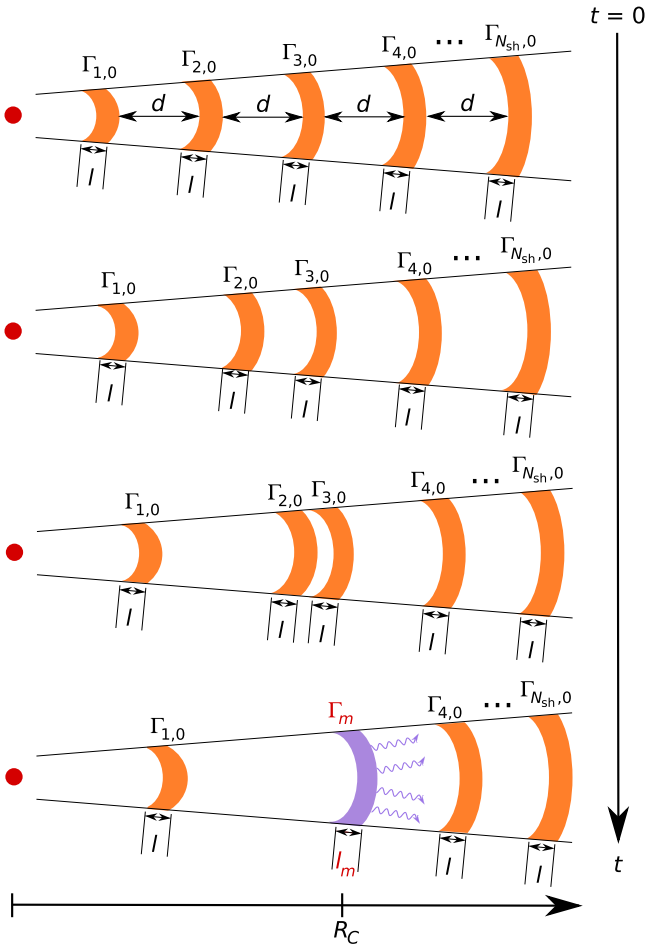

Figure 1. Illustration of the initialization, evolution, and collision of plasma shells in our simulations. See the main text and Appendix A for details.

propagating in the jet, and may collide again. Each simulation in this work comprises about 1000 collisions, as almost all initial shells collide.

Non-thermal electrons receive a fraction ϵ_e of the internal energy of the new shell, protons receive a fraction ϵ_p, and magnetic fields receive a fraction ϵ_B. Because electrons cool quickly via synchrotron emission, ϵ_e can also be regarded as the energy fraction of radiated gamma-rays. We assume that there is energy equipartition between electrons and magnetic fields, and 10 times more energy in protons (Abbasi et al. 2010; Hümmer et al. 2012), i.e., $\epsilon_e = \epsilon_B = 1/12$ and $\epsilon_p/\epsilon_e = 10$; see Appendix A.4. Thanks to the fast cooling, ϵ_p/ϵ_e can be regarded as the non-thermal baryon loading factor, defined as the ratio of cosmic-ray energy to radiation energy (Murase & Nagataki 2006).

The volume of the shell (see Appendix A.3) grows with the distance r to the central emitter $\propto r^2$ until the radial expansion of a shell becomes important at the shell spreading radius. Except in collisions, the number of particles in a shell is conserved. Thus, as a shell propagates and expands, the density of particles in it falls $\propto r^{-2}$. This affects where in the jet different types of particles are produced and emitted: neutrinos come predominantly from close to the photosphere, where densities are high; ultra-high-energy cosmic rays come from intermediate collision radii; and high-energy gamma-rays escape abundantly from large collision radii (Bustamante et al. 2015).

In a shell collision, if both protons and electrons are accelerated up to sufficiently high energies, $p\gamma$ interactions create pions via the $\Delta^{+}(1232)$ resonance and other processes. High-energy gamma-rays are produced in $\pi^0 \to \gamma\gamma$ (and by synchrotron and inverse-Compton radiation by electrons), while high-energy neutrinos are produced in $\pi^{+} \to \mu^{+}\nu_{\mu} \to \bar{\nu}_{\mu}e^{+}\nu_{e}\nu_{\mu}$ (and, via additional production channels, by its charge-conjugated version). The pion production efficiency is, on average, 20%; this is the fraction of proton energy received by pions. It is distributed roughly evenly among the pion decay products, so each neutrino carries $\sim$5% of the proton energy. Therefore, production of PeV neutrinos requires 20 PeV protons.

At the source, for the protons, we assume a power-law spectrum $\propto E_p'^{-2}\exp\left(-E_p'/E_{p,\max}'\right)$, with E_p' the proton energy, as expected from Fermi acceleration (primed quantities are in the shock rest frame). The maximum proton energy $E_{p,\max}'$ is computed by balancing the acceleration rate, and the synchrotron, adiabatic, and photohadronic energy loss rates (Baerwald et al. 2013).

For the target photons, we assume a broken power-law spectrum, Equation (14), which resembles the gamma-ray spectrum observed at Earth. This assumption is certainly justified beyond the photosphere, from where the photons can escape. Note that we neither generate the photon spectrum from first principles, nor explain its origin, which typically includes radiation processes such as inverse-Compton scattering, synchrotron emission, and bremsstrahlung. The Band function can be reproduced by invoking various effects, such as the slow heating of electrons (Bykov & Mészáros 1996; Murase et al. 2012; Asano & Terasawa 2015). In this sense, our approach is model-independent; this comes at the expense of insight on the radiation processes at work and their detailed individual implications on secondary production. To calculate the secondary particle production, we use the state-of-the-art NeuCosmA (Hümmer et al. 2012) software (see Appendix A.5).

UHECRs are emitted via two mechanisms: either $p\gamma$ interactions transform protons into neutrons, which escape the merged shell, and later beta-decay back into protons (Mannheim et al. 2001), or protons directly leak out of the shell without interacting. The latter situation occurs if the proton Larmor radius—determined by the physical conditions in the shell—exceeds the shell width at the highest energies; see Baerwald et al. (2013) for a detailed discussion. This direct escape produces a hard proton spectrum (it is dubbed a "high-pass filter" in Globus et al. 2015). Direct proton escape dominates over neutron escape only when the photon densities are low. Therefore, neutrino production associated with neutron escape is higher than that associated with direct proton escape (Mannheim et al. 2001).

Gamma-rays are produced throughout the jet. However, they only escape if they are produced above the "photosphere," i.e., the radius below which Thomson scattering occurs frequently enough to effectively trap them (see Appendix A.5). We adopt the pragmatic point of view that solid predictions about GRB dynamics and secondary production must be grounded in gamma-ray observations, which come exclusively from above the photosphere. The drawback of this assumption is that we cannot accurately calculate secondary production below the photosphere, since the sub-photospheric photons are subject to significant thermalization and the observed spectrum becomes quasi-thermal (Beloborodov 2013; Hascoet et al. 2013; Vurm et al. 2013). For our calculations of the gamma-ray flux, we will use only super-photospheric collisions. For neutrinos, we will additionally estimate possible sub-photospheric contributions. Photomeson production is more efficient below the photosphere, so that neutrino emission around the photosphere can be dominant, especially in photospheric emission models (Murase 2008; Wang & Dai 2009; Bartos et al. 2013; Murase et al. 2013; Bustamante et al. 2015).

For super-photospheric collisions, the gamma-rays are still attenuated in energy and produce particle cascades, via electron–positron pair-production processes inside shells (see below). After gamma-rays leave the source, they scatter off cosmological photon fields—microwave, optical, infrared—and their energies are degraded down to a few hundred GeV or lower. UHECRs also lose energy through photohadronic interactions and inelastic scattering off cosmological photons. Neutrinos are affected only by the adiabatic cosmological expansion, which changes their energies by, at most, a factor of a few.

We normalize the gamma-ray emission of our simulated bursts to typical time-integrated GRB gamma-ray energies, $E_{\gamma,\mathrm{norm}}^{\mathrm{iso}} = 10^{53}$ erg (in the source reference frame). This is equated to the sum energy emitted by all collisions, sub- and super-photospheric (Equation (13)). In our simulated bursts, about 50% of the energy is liberated in super-photospheric collisions.

4. Synthetic Light Curves

GRB light curves are highly irregular. They are extremely variable, and a fraction of GRBs have minimum variability timescales of ms (Golkhou et al. 2015). The time resolution of the detector imposes a lower bound on the variability timescale that can be inferred. Pulses typically have a duration of $\sim$1 s with an asymmetric structure (Ramirez-Ruiz & Fenimore 1999; Nakar & Piran 2002b), and some GRBs show distinct quiescent time between the pulses.

In the internal shock scenario, the shape of the GRB light curve is the result of shell collisions (Kobayashi et al. 1997; Daigne & Mochkovitch 1998; Beloborodov 2000; Spada et al. 2000; Kobayashi & Sari 2001; Daigne & Mochkovitch 2003; Aoi et al. 2010). When shells collide, a gamma-ray pulse and a neutrino pulse are emitted. The superposition of all pulses emitted during the evolution of the burst makes up the light curve. Different features in the light curves are due to differences in the distribution of collision radii, as determined by the initial speeds of the shells, imprinted on them by the central emitter. See Appendix A.4 for an explanation of how we construct the light curves.

The simulation parameters from our benchmark model from Bustamante et al. (2015) are re-branded here as "GRB 1." Note that we slightly adjusted the collision model;[8] therefore, the results are not exactly the same as in Bustamante et al. (2015). Most notably, collisions occur on average at slightly lower radii now. The initial values of shell Lorentz factors, $\Gamma_{k,0}$, are randomly sampled from a log-normal distribution defined by

[8] The main difference compared to the earlier computation in Bustamante et al. (2015) is that, before, we determined which two shells should collide next by computing the absolute value of the times needed for all contiguous shells to collide, and selecting the two shells with the minimum value, whereas now we include causality (so that slower shells cannot catch up with faster ones). We also fixed a problem with the maintenance of the collision data lists (the shell speed was not always updated correctly). As explained in the main text, these modifications do not qualitatively change the results for GRB 1 nor our conclusions.

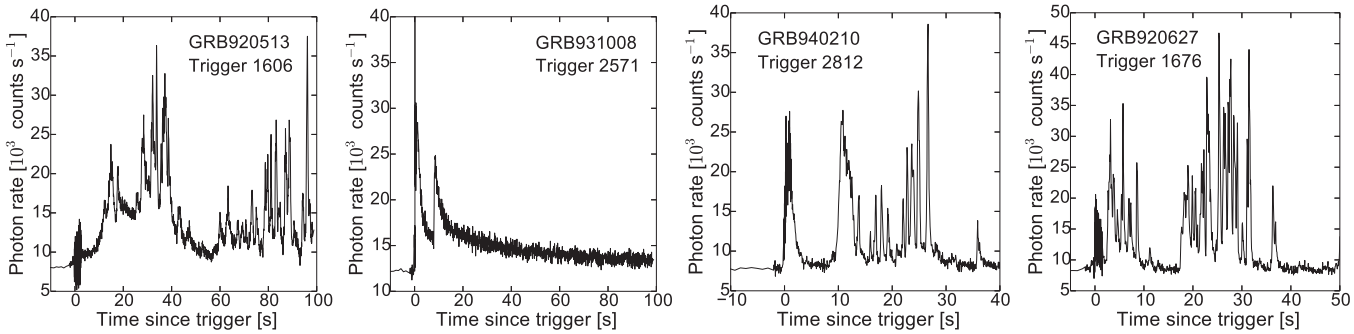

Figure 2. Selected GRB light curves, detected by BATSE (Paciesas et al. 1999) in the >20 keV range (channels 1–4); from left to right: GRB920513, GRB931008, GRB940210, and GRB920627. The implied time resolution is 64 ms.

the characteristic value Γ_0 and the amplitude of fluctuations A_Γ:

$$\ln\left(\frac{\Gamma_{k,0}-1}{\Gamma_0-1}\right)=A_\Gamma\cdot x, \tag{1}$$

where the random variable x follows a Gaussian distribution, $P(x)dx=(2\pi)^{-1}e^{-x^2/2}dx$. When $A_\Gamma<1$, the mean value $\langle\Gamma\rangle\approx\Gamma_0$ and the variance $\Delta\Gamma\approx A_\Gamma\Gamma_0$. When $A_\Gamma>1$, the mean value and variance are significantly affected by fluctuations. Large $A_\Gamma\gtrsim 1$ are typically required for the efficient conversion of kinetic energy to radiated energy (Kobayashi & Sari 2001), since the latter is proportional to the difference in speeds between two colliding shells. In addition, $A_\Gamma\gtrsim 0.1$ is necessary for the number of collisions to be large, i.e., $N_{\rm coll}\sim N_{\rm sh}$. The light curve for GRB 1 is in Figure 4. It is dominated by fast variability, which is determined by the ratio of T_{90} to $N_{\rm coll}$, on the order of tens of ms, which is typical for GRBs.

Figure 2 shows selected real gamma-ray light curves that have slower structure overlaid with fast variability. In the internal shock scenario, the slower structure can be produced by modifying the behavior of the engine. We can change the widths of the initial shells, the separations between them or the spread A_Γ, emit shells intermittently, ramp up or down the Lorentz factors during shell emission, or make them oscillate (see also Daigne & Mochkovitch 1998).

Table 1 lists our sample simulations, GRBs 1–6, where we have implemented these options. We have chosen parameter values such that the associated synthetic light curves reproduce features similar to Figure 2.

Figure 3 shows the randomly sampled initial values of the Lorentz factors $\Gamma_{k,0}$ of the shells in GRBs 1–6, as a function of initial radius $r_{k,0}$. See Table 1 for descriptions of the underlying distributions.

Figure 4 shows that the synthetic gamma-ray and neutrino light curves for GRBs 1–6. GRB 1 has fast time variability without prominent features. GRB 2 has a speedup in Γ during shell emission; a single pulse is overlaid with fast variability. For a slowdown, the pulse occurs earlier and the light curve is time-inverted. However, the efficiency is much lower in this case because the fast shells are emitted first. GRB 3 has three pulses with linear slowdown. The second and third pulses collide with slow shells from the preceding pulse and therefore contribute more strongly to the light curve than the first pulse. GRBs 4 and 6 have more oscillating periods. GRB 5 is oscillating as well, but its lower amplitude increases linearly during emission. Comparison with Figure 2 reveals that GRB 3 was inspired by GRB931008; GRBs 4 and 6, by case GRB920627; and GRB 5, by cases GRB920513 and GRB940210. Appendix B contains four more simulation examples, with different engine assumptions but similar behavior to that of GRBs 1–6, showing that these are representative.

Most GRB light curves detected by *Fermi* (Ajello et al. 2013; von Kienlin et al. 2014) do not have prominent features like the ones in Figure 2. The reason why the latter are featured in the literature more often than simpler single-pulse or fast-variability light curves is possibly a selection effect (i.e., they are more interesting to show and study). From that perspective, it is conceivable that GRBs 3–6 are not "typical" GRBs.

The light curves of GRBs 3 and 5 are qualitatively different from the others in one key aspect: they have a dominant, broad pulse structure overlaid with fast time variability, whereas in the other bursts the fast component is more relevant. This feature can be traced back to the input parameters in Table 1: GRBs 3 and 5 have a "disciplined" central engine that emits shells within a narrow Γ distribution ($A_\Gamma=0.1$), while the average Γ changes slowly. Therefore, most collisions occur at larger radii compared to the cases where the spread in Γ is larger. We will see that this also affects the neutrino production efficiency in the case of GRB 5.

The internal shock model has been invoked as a successful model to explain irregular features of the GRB light curve (Kobayashi et al. 1997; Daigne & Mochkovitch 1998; Beloborodov 2000; Spada et al. 2000; Kobayashi & Sari 2001; Daigne & Mochkovitch 2003; Aoi et al. 2010). However, because the classical internal shock model is being challenged, light-curve predictions of alternative models have also been extensively investigated. In particular, it has been shown that turbulence or magnetic reconnection models can better explain the observed structure of the GRB light curve (Lazar et al. 2009; Narayan & Kumar 2009; Zhang & Zhang 2014; Beniamini & Granot 2016). Some recent studies suggested that GRB light curves may consist of the superposition of slow and fast components, as inferred by a gradual depletion of the fast component at low energies (Vetere et al. 2006). Substructures of the observed GRB pulses can be easily accounted for by assuming relativistic motions in the bulk of a relativistic jet. Although their origins are unclear at present, such relativistic motions could be realized in the ICMART model (Zhang & Zhang 2014). Compared to these explanations, our pulse structure comes from the properties of the engine rather than the jet.

Figure 4 shows that there is no linear correlation between the heights of gamma-ray and neutrino pulses. This is because the height of a neutrino pulse is, via the pion production efficiency,

Table 1
Description of Simulated GRBs 1–6

Model	$\Gamma_{0,1}$	$A_{\Gamma,1}$	$\Gamma_{0,2}$	$A_{\Gamma,2}$	T_p	$N_{\rm up}$	$N_{\rm down}$	$E^{\rm iso}_{\gamma,\rm norm}$ (erg)	Description
1	500	1.0	⋯	⋯	⋯	⋯	⋯	10^{53}	Fixed Γ and A_Γ; benchmark from Bustamante et al. (2015)
2	500	1.0	50	0.1	⋯	⋯	⋯	10^{53}	Linear speedup of Γ
3	50	0.1	500	0.1	0.34	⋯	⋯	10^{53}	Sawtooth Γ (linear slowdown three times) with narrow distribution
4	50	0.1	500	1.0	0.2	⋯	⋯	10^{53}	Oscillating Γ (five periods) with increasing distribution width
5	50	0.1	500	0.1	0.2	⋯	⋯	10^{53}	Oscillating Γ (four periods) with lower amplitude increasing and narrow distribution
6	50	0.1	500	1.0	1/8	250	250	10^{53}	Oscillating Γ (four periods) with distribution widening up; in addition, engine intermittent: $N_{\rm up}$ pulses followed by $N_{\rm down}$ pulses; corresponds to increasing Lorentz factor during uptime

Note. Common values for all models: $N_{\rm sh} = 1000$, $\delta t_{\rm eng} = 10^{-2}$ s, $d = l = c \cdot \delta t_{\rm eng}$, $r_{\rm min} = 10^3$ km, $r_{\rm dec} = 5.5 \cdot 10^{11}$ km, $z = 2$, $\epsilon_e = \epsilon_B = 1/12$, $\epsilon_p = 5/6$, and $\eta = 1.0$ (acceleration efficiency; Baerwald et al. 2013). See Table 3 for an explanation of each parameter. The period T_p for the oscillating cases refers to Γ changing between $\Gamma_{0,1}$ and $\Gamma_{0,2}$, and A_Γ changing between $A_{\Gamma,1}$ and $A_{\Gamma,2}$; T_p is a fraction of the total number of emitted shells. This means that Γ and A_Γ change between first and second value with a factor $\sin^2(k/(N_{\rm sh} \cdot T_p) \cdot \pi)$, where k is the index of the shell ($1 \leqslant k \leqslant N_{\rm sh}$). For GRB 5, the lower amplitude increases from $\Gamma_{0,1}$ to $\Gamma_{0,2}$ linearly with k.

more sensitive to the collision radius than the height of a gamma-ray pulse, i.e., it depends on the proton and photon densities, which drop $\propto R_{\rm C}^2$, where $R_{\rm C}$ is the collision radius.

There are no long time delays between gamma-ray and neutrino pulses: they are within T_{90} (see Equation (23)) of each other. There may be, however, short delays. For example, in GRB 3, the neutrino peak corresponding to the first large gamma-ray peak is suppressed. So the first gamma-ray detection will have occurred ∼20 s, before the neutrino instrument triggers. In GRB 4, the quiescent periods of gamma-ray and neutrino emission coincide, which may be exploited by neutrino telescopes to set further time window cuts.

In GRB 5, the fast time variability gives rise to neutrino spikes, whereas the longer pulses seen in gamma-rays are hardly present in neutrinos. Indeed, the rise time of the spike of particle emission associated with one collision depends strongly on Γ (see Equation (18)) and, consequently, it is $\propto R_{\rm C}$. Therefore, faster rises—sharper structures—are created by collisions at smaller radii, where the neutrino production efficiency is higher.

Table 2 lists the parameters output by GRBs 1–6. GRBs 1, 2, 4, and 5 have time variabilities of ∼50 ms and durations of ∼50 s. In GRB 3, the duration (33 s) and time variability (36 ms) are smaller because mainly two of the three peaks contribute. In GRB 6, pulses are separated by a downtime and, therefore, duration and variability are larger (98 s and 97 ms, respectively).

The dissipation efficiency ε of a burst in Table 2 is the ratio between total energy dissipated by all types of particles in super-photospheric collisions and total kinetic energy available at the start of the simulation. Most simulations[9] have high $\varepsilon \approx 11\%$–27%, which reasonably agrees with previous work (Beloborodov 2000). The efficiency is lower in cases with narrow Γ distribution, as expected.

The gamma-ray emission efficiency $\varepsilon_\gamma \approx \epsilon_e \varepsilon$ is about 10 times smaller since most of the dissipation energy is assumed to be carried by protons rather than electrons. Here, ϵ_e is the fraction of energy in electrons and photons; see Appendix A.4. Such a value for the radiation efficiency may be too small compared to ones preferred by observations. However, if the internal energy is carried by thermal protons or confined cosmic rays, it is natural to expect the reconversion of the internal energy into the kinetic energy; see Appendix C. It has been shown that this effect increases the gamma-ray emission efficiency, represented by the ratio of prompt gamma-ray energy to afterglow kinetic energy, calculated in an approach where shells reflects off each other after colliding, i.e., collisions are not perfectly inelastic (Kobayashi & Sari 2001). Recent results on afterglow modeling also suggest a small value of the gamma-ray emission efficiency (Beniamini et al. 2015).

[9] Bursts reach higher efficiencies (∼40%) if they have a square-pulse ("box-like") distribution of Γ. Since this distribution is unrealistic, we do not discuss it further.

5. Multi-messenger Emission

Here we discuss the emission of multiple messengers and their relation.

5.1. Weak versus Strong Neutrino Emitters

The time-integrated neutrino fluence of a simulated burst, for a baryon-rich jet, roughly scales as (Bustamante et al. 2015)

$$\mathcal{F}_\nu \propto \frac{N_{\rm coll}\,(\tau_{p\gamma} \gtrsim 1)}{N_{\rm coll}} \cdot \varepsilon \cdot E^{\rm iso}_{\gamma,\rm tot}, \tag{2}$$

assuming a fixed photon break energy (see Appendix A.4). The first factor gives the fraction of collisions with high optical depth $\tau_{p\gamma} \gtrsim 1$ to $p\gamma$ interactions, ε is the energy dissipation efficiency, and $E^{\rm iso}_{\gamma,\rm tot}$ is the total energy emitted as gamma-rays in super-photospheric collisions. Unlike one-zone predictions (Waxman & Bahcall 1997; Guetta et al. 2004; He et al. 2012; Hümmer et al. 2012; Li 2012), the fluence does not depend on the average Lorentz factor of the shells.

Figure 5 shows $\tau_{p\gamma}$ as a function of $R_{\rm C}$ for collisions in GRBs 1 and 5. In GRB 1, a strong neutrino emitter, about 60 in 1000 collisions occurred above the photosphere—so that they were optically thin to Thomson scattering—*and* were still optically thick to photomeson production. The other simulations with broad Γ distributions have similar results. Therefore, the first factor in Equation (2) is ∼0.05 for strong neutrino emitters. The energy dissipation efficiency in Table 2 lies around $\varepsilon = 0.2$ for these GRBs. As a result, for fixed $E^{\rm iso}_{\gamma,\rm tot}$, the quasi-diffuse neutrino flux that we infer below (Bustamante et al. 2015) is relatively robust.

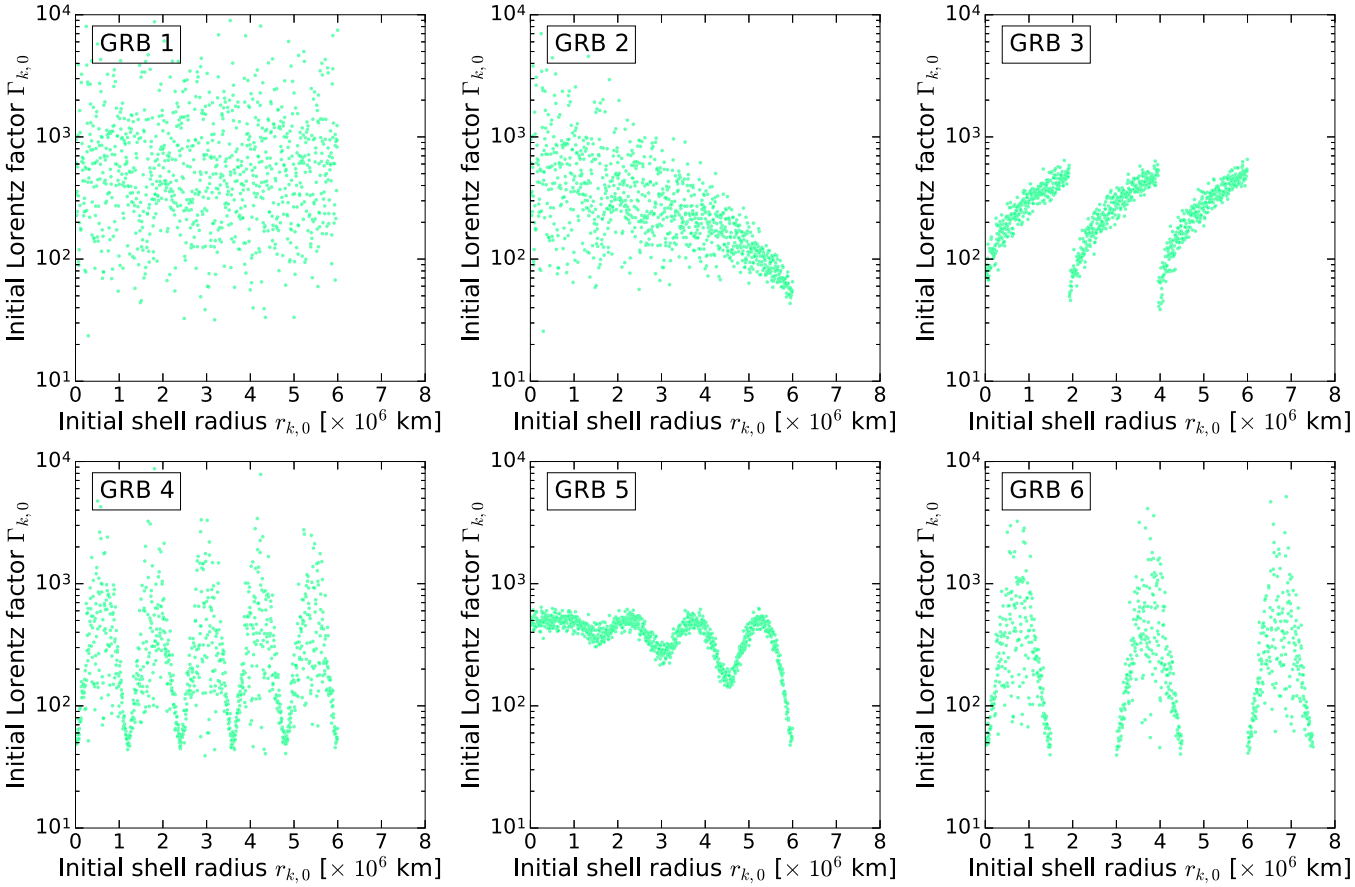

Figure 3. Initial values of the Lorentz factors of the shells in GRBs 1–6, at the start of the simulations. See Table 1 for descriptions of the underlying distributions of initial Lorentz factors in each simulation.

The situation is different for GRBs 3 and 5. They have lower efficiencies $\varepsilon \approx 10\%$. More importantly, they have no optically thick collisions close to the photosphere; see the bottom panel of Figure 5 for GRB 5. The reason is that they emit shells with a variable but narrow Lorentz factor distribution (see Figure 3) that tends to induce collisions at larger radii, where photon densities are low and, therefore, so is neutrino production efficiency. In particular, this makes GRB 5 our weakest neutrino emitter, i.e., it has the lowest ratio of emitted neutrino energy to gamma-ray energy beyond the photosphere. While the same effect should also make GRB 3 a weak neutrino emitter, its neutrino flux is still on a level with our other examples. The reason for this comes from its very specific initial shell setup. It consists of three narrow pulses, each with decreasing Γ. The collisions are therefore dominated by the first, fast shells of a pulse running into the preceding pulse—these shells have largely different Lorentz factors; in particular, the differences are larger than in GRB 5. Most of the neutrino emission comes from these first collisions, which happen below and slightly above the photosphere.

We saw that GRBs with light curves dominated by fast variability are likely to be strong neutrino emitters. However, the reverse conclusion does not hold. While both GRBs 3 and 5 have gamma-ray light curves with broad pulses overlaid with fast variability, only GRB 5 is a weak neutrino emitter. Therefore, in the multi-zone internal shock model, we can tell, by inspection of the gamma-ray light curve alone, whether or not a GRB is likely to be a strong neutrino emitter. Conclusions about weak neutrino emitters require a closer inspection of the specific light-curve morphology.

Figure 6 shows, in the top panel, the neutrino fluence for GRB 2. The neutrino fluence follows the behavior of $\tau_{p\gamma}$ from Figure 5 (which is shown there for different examples). The average fluence per collision drops stronger than approximately $\propto R_{\rm C}^{-2}$. In principle, the fluence from sub-photospheric collisions is high, due to the high extrapolated photon density; however, we do not use those collisions in our flux calculations.

5.2. Quasi-diffuse Neutrino Flux

We derive the all-sky quasi-diffuse $\nu_\mu + \bar{\nu}_\mu$ flux J_{ν_μ} associated with a particular simulated GRB by scaling its fluence by the rate of long-duration GRBs, $\dot{N} = 667$ per year, i.e., $J_{\nu_\mu} = \mathcal{F}_{\nu_\mu} \cdot \dot{N} \cdot (4\pi)^{-1}$. Since this flux does not contain contributions from sub-photospheric collisions, it is effectively a lower limit on the prompt GRB neutrino flux. For our original benchmark, GRB 1, Bustamante et al. (2015) found that the flux is robust against variations in burst parameters like $\delta t_{\rm eng}$ (see Table 3) and $N_{\rm sh}$. We will discuss below how it depends on underlying assumptions, and what the corresponding fluxes are for GRBs 2–6.

Figure 7 shows the fluxes for GRBs 1–6. For all but GRB 5, the flux is $E^2\, J_{\nu_\mu} \approx 2 \cdot 10^{-11}$ GeV cm^{-2} s^{-1} sr^{-1} around 1 PeV—this is close to the value found for GRB 1 in Bustamante et al. (2015). For GRB 5, the flux is somewhat lower because it has fewer optically thick collisions, which is in agreement with

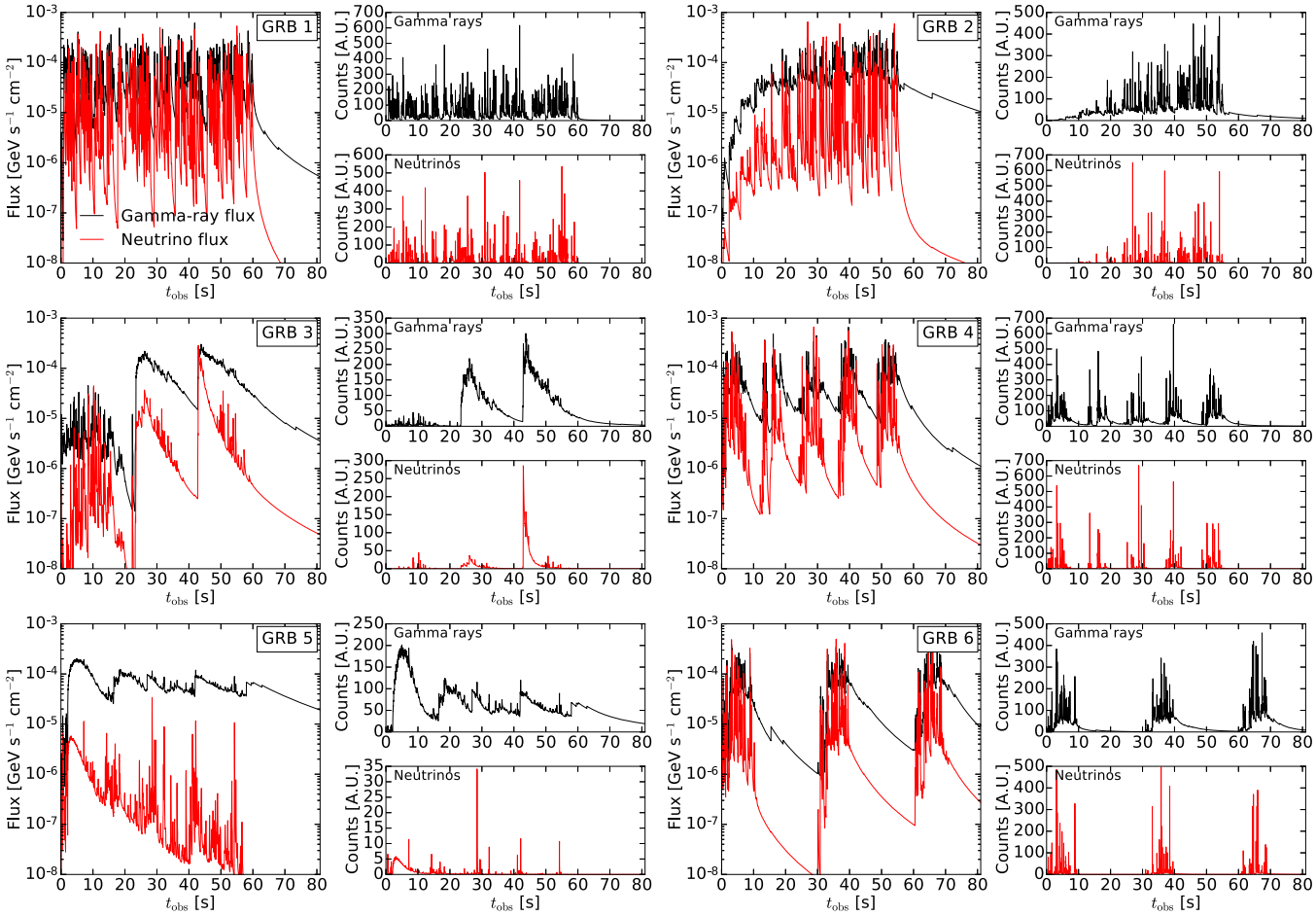

Figure 4. Synthetic gamma-ray and neutrino light curves for the simulated GRBs 1–6, from collisions beyond the photosphere. Photon and neutrino counts are in arbitrary units, obtained by multiplying the flux times a factor of 10^6 GeV^{-1} cm^2 s.

Equation (2). (For GRB 3, the same could be expected, but instead it has a higher neutrino flux, as explained in Section 5.1.) Most real light curves lack the non-trivial features seen in GRB 5 (and GRB 3), and are therefore likely strong neutrino emitters. So it is conceivable that most bursts are instead like GRBs 1, 2, 4, and 6, and that the quasi-diffuse flux lies indeed at the level predicted using those bursts.

IceCube has searched for correlations between neutrino arrival directions and positions of known GRBs (Abbasi et al. 2012; Aartsen et al. 2015a, 2016). No significant signal from GRBs has been found, in consistency with our prediction. Figure 7 includes the differential upper bound from Abbasi et al. (2012).

One-zone and multi-zone models have similar (average) burst parameters and compute the flux of secondaries similarly. However, Figure 7 shows that fluxes calculated with the multi-zone model are typically lower than with the one-zone model (Hümmer et al. 2012; see dashed curves in Figure 7). The reason is that all shells are assumed to have the same collision radius in the one-zone model, which tends to be underestimated: since the neutrino production efficiency decreases nonlinearly with the collision radius, the average value of the collision radius is, in general, not representative of the neutrino production. Nevertheless, we could define a single effective collision radius for neutrino production in the one-zone model that is different from the radius for gamma-rays; this would be done by folding in the production efficiency calculated with the fraction of collisions occurring at that radius. The average or representative jet parameters—such as the typical collision radius—are derived from gamma-ray observations (see Appendix A.4), which implies that parameters representative of gamma-rays may not be representative of the other messengers.

5.3. Cosmic Rays

From Table 2, we can see that all of the GRBs 1–6 are relatively efficient cosmic-ray emitters, though the required energy output per GRB, of at least 10^{53} erg in the discussed energy range, should likely be a factor of a few larger to explain UHECR observations.[10] Within the presented model, it is conceivable that GRBs are the sources of UHECRs.

The connection between cosmic rays and neutrinos depends on how UHECRs escape the shells. Photohadronic interactions will transform protons into neutrons; neutrinos will also be produced. If all cosmic rays escape as neutrons ("neutron escape") the connection is strong (Mannheim et al. 2001; Ahlers et al. 2011): one neutrino of each flavor is expected per observed UHECR proton. This possibility can be clearly ruled out (Abbasi et al. 2012; Baerwald et al. 2015). However, at the

[10] For details, see Section 2 in Baerwald et al. (2015), where the dependence on the source evolution is also discussed. Such an increase can be achieved either by a somewhat larger gamma-ray luminosity or by a somewhat larger baryonic loading.

Table 2
Parameters Output by Simulated GRBs 1–6

Model	$N_{\rm coll}$	$t_{\rm v}$ (ms)	T_{90} (s)	$E^{\rm iso}_{\gamma,\rm tot}$ (erg)	$E^{\rm iso}_{p,\rm tot}$ (erg)	$E^{\rm iso}_{\nu,\rm tot}$ (erg)	$E^{\rm iso}_{\nu,\rm tot}/E^{\rm iso}_{\gamma,\rm tot}$ (%)	ε (%)
1	987	53	54.0	$5.2 \cdot 10^{52}$	$6.2 \cdot 10^{52}$	$1.4 \cdot 10^{52}$	26.9	26.8
2	999	47	47.0	$6.5 \cdot 10^{52}$	$4.4 \cdot 10^{52}$	$9.3 \cdot 10^{51}$	14.3	19.6
3	951	33	35.5	$6.2 \cdot 10^{52}$	$5.0 \cdot 10^{52}$	$7.6 \cdot 10^{51}$	12.3	10.5
4	987	52	52.6	$4.6 \cdot 10^{52}$	$4.0 \cdot 10^{52}$	$9.4 \cdot 10^{51}$	20.4	21.7
5	990	57	57.9	$8.9 \cdot 10^{52}$	$1.7 \cdot 10^{52}$	$6.6 \cdot 10^{50}$	0.7	10.6
6	985	97	98.0	$6.1 \cdot 10^{52}$	$4.2 \cdot 10^{52}$	$1.1 \cdot 10^{52}$	18.0	23.6

Note. The parameters are variability timescale ($t_{\rm v}$), total energy emitted as gamma-rays ($E^{\rm iso}_{\gamma,\rm tot}$), as protons ($E^{\rm iso}_{p,\rm tot}$), and as neutrinos ($E^{\rm iso}_{\nu,\rm tot}$), ratio between neutrino and gamma-ray energies ($E^{\rm iso}_{\nu,\rm tot}/E^{\rm iso}_{\gamma,\rm tot}$), and overall emission efficiency ε. Energies are computed using super-photospheric collisions only. Only protons in the UHECR energy range, above $>10^{10}$ GeV, are counted. The efficiency ε is defined as the ratio of total energy dissipated by all types of particles (gamma-rays, protons, neutrinos) to the total kinetic energy initially available.

highest energies, when the proton Larmor radius exceeds the shell width, protons can directly escape the shells without interacting in them, which leads to a hard spectrum ("direct escape"). In addition, diffusion may lead to escape depending on properties of magnetic fields.

In each merged shell, one or another escape component dominates,[11] depending on the properties of the shell (Baerwald et al. 2013). If we consider escape processes other than neutron escape, which is implicitly assumed in most of the previous literature (Waxman & Bahcall 1997; Dermer & Atoyan 2003; Guetta et al. 2004; Murase & Nagataki 2006), the latest IceCube data cannot exclude GRBs as the sources of UHECRs even in a one-zone model, but constraints on the average shell parameters can be obtained from the efficient neutrino production (Baerwald et al. 2015).

Figure 6 shows, in the bottom panel, the maximum proton energy $E_{p,\rm max}$ to which protons are accelerated in collisions in GRB 2, as a function of collision radius. Below the photosphere (black boxes), proton synchrotron losses dominate and $E_{p,\rm max}$ increases with $R_{\rm C}$. Around $R_{\rm C} \simeq 10^{8.5}$–$10^{10}$ km, protons reach 10^{10} GeV and higher. This is where UHECRs are emitted. At large $R_{\rm C}$, falling magnetic fields yield lower acceleration rates and energies. Neutron emission is correlated with efficient neutrino production, since neutrons and charged pions are produced together in $p\gamma$ interactions. However, this occurs only in a few collisions, in a narrow range of low collision radii, where proton and photon densities are high; see red dots. In effect, cosmic-ray emission via neutron escape is limited by Equation (2). Most collisions occur at larger radii, so that the average collision radius for CR emission tends to be higher than for neutrinos (blue circles). There, particle densities are low enough for direct proton escape to dominate, without associated neutrino production. Given that only few collisions are neutron-dominated, the pure neutron escape assumptions for GRBs (Ahlers et al. 2011) cannot be justified. However, a quantitative statement requires further study beyond the scope of this work because it depends on the relative contribution between neutron-dominated and direct escape-dominated shells. For a discussion of UHECR nuclei, see Bustamante et al. (2015) and Globus et al. (2015).

[11] This also holds for UHECR nuclei that have not been photodisintegrated (Globus et al. 2015). The survival of heavy nuclei is shown to be possible and their escaping flux may explain the observed UHECR flux (Murase et al. 2008; Wang et al. 2008).

5.4. Multi-messenger Emission from Different Radii

Figure 8 shows a key feature of the multi-zone GRB model that is not captured by the one-zone model: that neutrinos, gamma-rays, and UHECR protons are emitted from different regions of the jet (Bustamante et al. 2015). This holds regardless of the difference in burst parameters among GRBs 1–6. Neutrinos are produced close to the photosphere, as discussed above. UHECRs tend to be produced at somewhat larger radii. At low radii, UHECRs escape as neutrons; at larger radii, most UHECRs escape directly as protons. Gamma-rays tend to come from even larger radii. While their production is more evenly distributed in collision radius, at low radii, pair production ($\gamma\gamma \to e^+e^-$) drives their energy down, so high-energy gamma-rays mainly come from large radii.

6. Delayed High-energy Gamma-Rays

The detection of time delays between gamma-ray signals in different energy bands can provide insight into the dynamics of the GRB central engine and jet.

The maximum energy with which a photon can escape the shell where it is created is limited by the optical depth $\tau_{\gamma\gamma}$ to pair production via $\gamma\gamma \to e^+e^-$. A photon of a certain energy escapes only if $\tau_{\gamma\gamma} < 1$. Close to the central engine, photon density and, therefore, optical depth, are high. Bustamante et al. (2015) showed that only low-energy gamma-rays escape from that region. Higher-energy gamma-rays escape at larger radii. In this section, we explore whether the different shell opacities lead to time delays between gamma-ray energy bands.

Figure 9—top row, left panel—shows that, in GRB 1, $t_{\rm obs}$ is quite uniformly distributed in $R_{\rm C}$, especially between 10^8 and 10^{10} km, i.e., above the photosphere. The central panel shows that for many collisions in this range the maximum gamma-ray energy is limited by pair production, while, from 10^9 km up, an increasing number of collisions is unaffected by it. However, the right panel shows that no separation exists between the arrival times of gamma-rays limited and not limited by pair production. In other words, at any time during the burst, low- and high-energy gamma-rays arrive indistinctly at Earth from everywhere inside the jet.

For GRB 5 (Figure 9, bottom row), the situation is different. The average $t_{\rm obs}$ increases with $R_{\rm C}$ between 10^9 and 10^{11} km, i.e., some early ($\lesssim$20 s) gamma-ray detections come from mid-range radii, while all late detections tend to come from large radii. As a result, some early gamma-rays have lower energies, in the upper *Fermi*-LAT and lower CTA bands, limited by pair production. Later detections, coming from larger radii, will

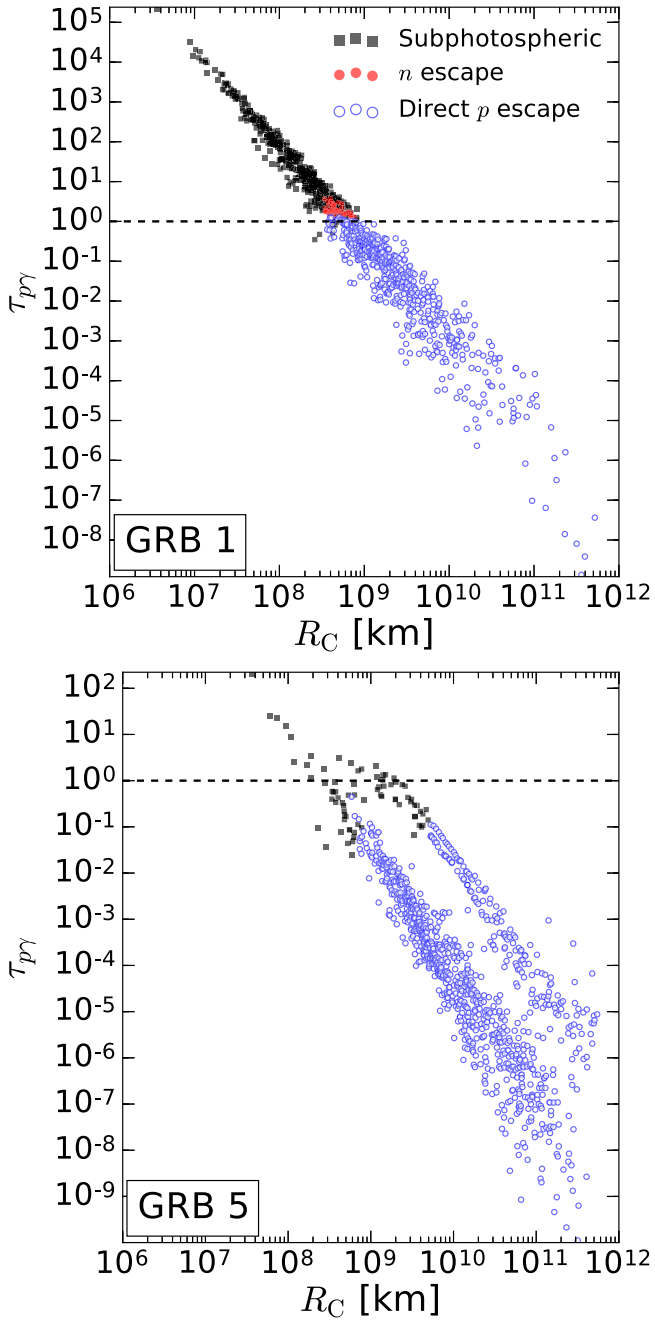

Figure 5. Optical depth $\tau_{p\gamma}$ for all collisions in GRBs 1 (top) and 5 (bottom) as a function of collision radius. The horizontal line corresponds to $\tau_{p\gamma} = 1$. Black filled rectangles are sub-photospheric collisions, red filled dots are super-photospheric collisions, where the dominant UHECR component is neutron escape, and blue unfilled dots are super-photospheric collisions where the dominant component is direct proton escape.

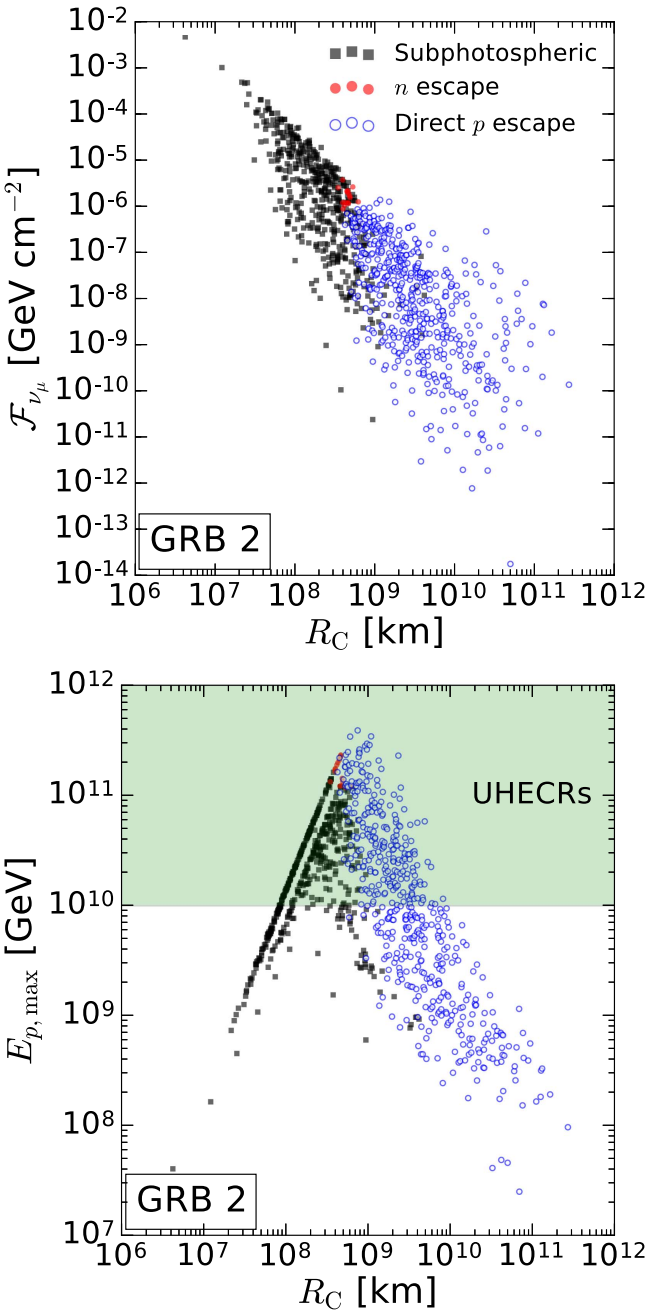

Figure 6. Muon-neutrino ($\nu_\mu + \bar{\nu}_\mu$) fluence (top) and maximum cosmic-ray energy (in source frame, bottom) for collisions in GRB 2. The legend is the same as for Figure 5. In the bottom panel, the UHECR range $E_{p,\rm max} > 10^{10}$ GeV is shaded.

have consistently higher energies, not limited by pair production. There is also an impact on the neutrino light curve: Figure 4 shows that the neutrino flux is much lower for later collisions, which come from larger radii, where neutrino production is inefficient. This behavior is characteristic of bursts with narrow Γ distribution, where collisions tend to occur at large radii and late in the jet evolution.

Figure 10 shows the gamma-ray light curves in different energy bands for GRBs 3 and 5, our two simulations with narrow Γ distributions. To produce them, we have assumed that the power-law photon density in the source extends to high enough energies.[12] GRB 3 has a delayed start of a few seconds in the LAT band compared to the first detected peak in GBM, and of $\sim$10 s in the higher-energy CTA band. These delays depend strongly on the energy threshold of the instrument. In GRB 5, the LAT peak follows the GBM peak after $\sim$2 s, and the signal in CTA grows to comparable levels $\sim$10 s later. For this GRB, the suppression affects mainly the first peak in the light curve (the overall normalization of the light curves is

[12] This might overestimate the relative height of the light curve in the *Fermi*-LAT band.

Table 3
Main Parameters of the Burst Simulation

Parameter	Description	Type	Units	Notes
		Burst initialization		
$N_{\rm sh}$	Initial number of shells	Input	⋯	⋯
$\delta t_{\rm eng}$	Uptime of the engine	Input	s	⋯
$\Delta t_{\rm eng}$	Downtime of the engine	Input	s	⋯
$r_{N_{\rm sh}}$	Distance from innermost shell to central emitter	Input	km	Default: 10^3 km
$r_{\rm dec}$	Deceleration radius (circumburst medium starts)	Input	km	Default: $5.5 \cdot 10^{11}$ km
A_Γ	Fluctuation amplitude of $\Gamma_{k,0}$ distribution	Input	⋯	⋯
$E^{\rm iso}_{\rm kin,0}$	Initial bulk kinetic energy of shells	Input	erg	Common to all initial shells
z	Redshift of the emitter	Input	⋯	Used to calculate $t_{\rm obs}$, Equation (5): $z = 2$ by default
l	Initial shell width	Internal	km	Common to all initial shells: $l = c \cdot \delta t_{\rm eng}$
d	Initial separation between consecutive shells	Internal	km	Common to all initial shell pairs: $d = l$ by default
$r_{k,0}$	Initial radius of the kth shell,	Internal	km	$r_{k,0} = r_{N_{\rm sh}} + (N_{\rm sh} - k)(l + d)$
$\Gamma_{k,0}$	Initial bulk Lorentz factor of the kth shell	Internal	⋯	Sampled from the pre-defined Γ distribution
$m_{k,0}$	Initial mass of the kth shell	Internal	GeV	$m_{k,0} = E^{\rm iso}_{\rm kin,0}/(\Gamma_{k,0}c^2)$
		Burst evolution		
t	Time in the source frame	Internal	s	⋯
r_k	Radius of the kth shell	Internal	km	Grows as $r_k = r_{k,0} + c\beta_k t$
l_k	Width of the kth shell	Internal	km	Changes only in collisions
Γ_k	Bulk Lorentz factor of the kth shell	Internal	⋯	Changes only in collisions
m_k	Mass of the kth shell	Internal	GeV	Changes only in collisions
β_k	Bulk speed of the kth shell	Internal	⋯	$\beta_k = \sqrt{1 - \Gamma_k^{-2}}$
$V_{{\rm iso},k}$	Isotropically equivalent volume of the kth shell	Internal	km^3	$V_{{\rm iso},k} = 4\pi r_k^2 l_k$
$E^{\rm iso}_{{\rm kin},k}$	Bulk kinetic energy of the kth shell	Internal	erg	Changes only in collisions
ρ_k	Mass density of the kth shell	Internal	GeV km^{-3}	$\rho_k = m_k/V_{{\rm iso},k}$
		Shell collisions		
$m_{\rm f(s)}$	Mass of the fast (slow) colliding shell	Internal	GeV	⋯
$\Gamma_{\rm f(s)}$	Bulk Lorentz factor of the fast (slow) shell	Internal	⋯	⋯
$\Gamma_{\rm fs(rs)}$	Lorentz factor of the forward (reverse) shock	Internal	⋯	See Equation (9)
$\beta_{\rm fs(rs)}$	Speed of the forward (reverse) shock	Internal	⋯	$\beta_{\rm fs(rs)} = \sqrt{1 - \Gamma_{\rm fs(rs)}^{-2}}$
$\beta_{\rm m}$	Bulk speed of the merged shell	Internal	⋯	$\beta_{\rm m} = \sqrt{1 - \Gamma_{\rm m}^{-2}}$
$\rho_{\rm m}$	Mass density of the merged shell	Internal	GeV km^{-3}	See Equation (10)
$t_{\rm coll}$	Collision time (source frame)	Internal	s	⋯
$N_{\rm coll}$	Total number of collisions	Output	⋯	⋯
$t_{\rm obs}$	Collision time (observer's frame)	Output	s	See Equation (5)
$\Gamma_{\rm m}$	Bulk Lorentz factor of the merged shell	Output	⋯	See Equation (7)
$E^{\rm iso}_{\rm coll}$	Internal energy liberated in the collision	Output	erg	See Equation (6)
$l_{\rm m}$	Width of the merged shell	Output	km	See Equation (8)
R_C	Collision radius	Output	km	⋯
$\delta t_{\rm em}$	Time for reverse shock to cross fast shell	Output	s	See Equation (17)

Note. All quantities are expressed in the source (engine) frame, except for $t_{\rm obs}$, which is in the observer's frame.

arbitrary, but the relative normalization among the different energy bands is fixed). The early suppression of high-energy emission is consistent with observations; see, e.g., Castignani et al. (2014).

Since bursts with time delays between energy bands have narrow Γ distributions, they are associated with light curves with broad pulses overlaid with fast variability and possibly weak neutrino emitters (see Section 4). It is possible, in principle, to use the observation of delays in population studies to find how common these GRBs actually are, which affects the neutrino flux from the full GRB population.

To summarize, our predictions are as follows.

1. In GRBs with light curves that have broader pulses overlaid with fast variability, time delays in different wavelength bands are possible.
2. Compared to detection in the GBM energy band (10^{-6}–10^{-2} GeV), typical delays are a few seconds long in the LAT band (10^{-1}–10^{2} GeV) and ∼10 s in the CTA band (10^{2}–10^{6} GeV).
3. If such delays are observed, the GRB can be a weak neutrino emitter.
4. The fundamental reason for these apparent delays is an early suppression—rather than an actual delay—of high-energy gamma-rays coming from smaller R_C, where the

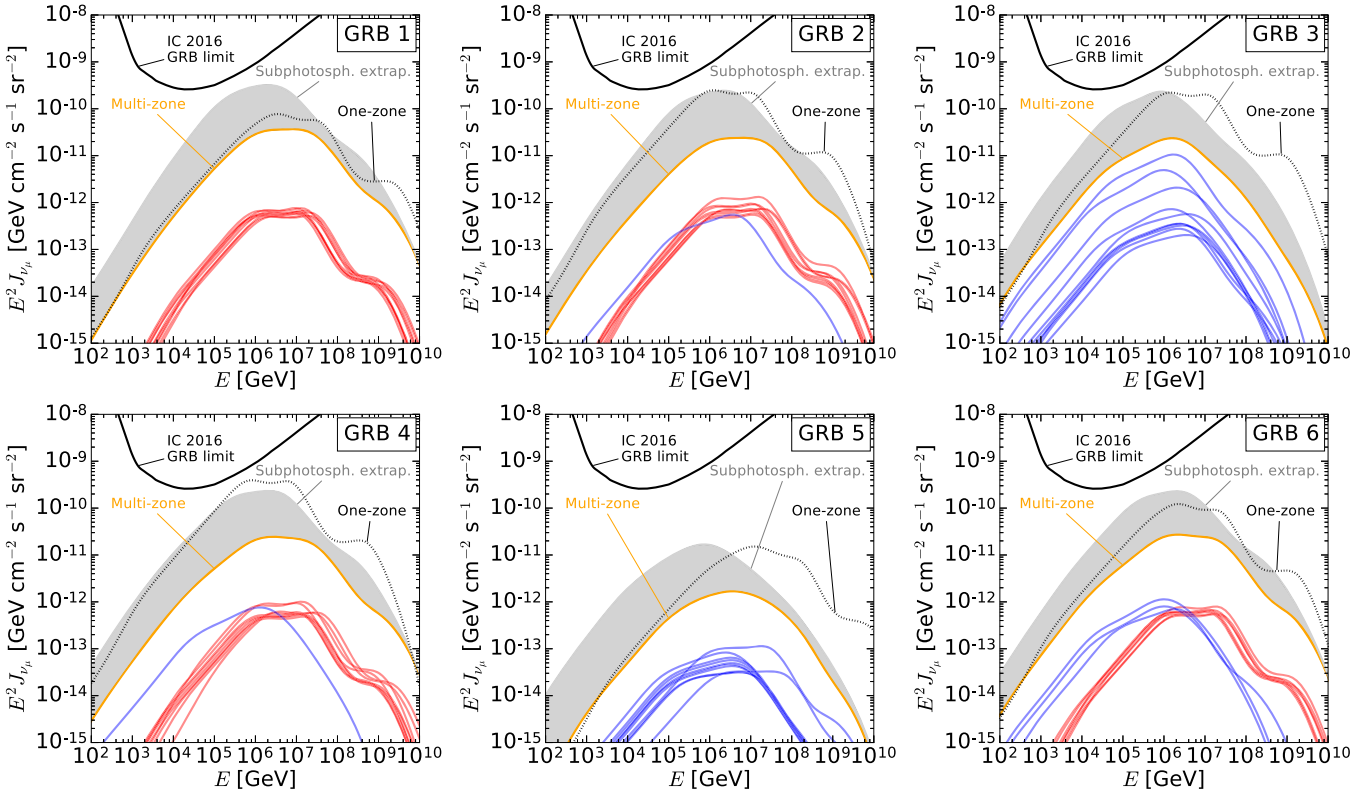

Figure 7. All-sky quasi-diffuse $\nu_\mu + \bar{\nu}_\mu$ fluxes in our simulated multi-zone GRBs 1–6. Numerical one-zone predictions (Hümmer et al. 2012) are included for comparison; they are calculated using the average burst parameters computed as in Appendix A.4. The shaded regions give the potential contribution from sub-photospheric collisions. The dominant contributions from individual collisions are shown as thin curves, corresponding to cases where the optical depth to photohadronic interactions $\tau_{p\gamma}$ is larger (red/light) or smaller (blue/dark) than unity. The IceCube 2016 upper limit was calculated using their latest reported detector effective area and exposure in a stacked GRB search using tracks coming from the Northern Hemisphere (Aartsen et al. 2016).

radiation densities are higher and the gamma-rays cannot escape. This effect is only observable if observation time and R_C in the collisions are correlated.

An example of a GRB that could match these predictions is GRB 080916C (Abdo et al. 2009).

These predictions are a qualitative summary based on examples GRB 1–6, which we believe to be representative of a larger set of examples that we have produced. Quantitative results depend on the chosen parameter values, which reflects the observation that no two light curves are identical. Figure 9 captures the central feature we observe in all examples that exhibit a lag: for these, there is a correlation between $t_{\rm obs}$ and R_C (within the range of R_C values in which the maximal photon energy depends on R_C). This correlation can be traced back to the engine emitting shells in a relatively narrow range of values of Γ. The lags in Figure 10 are clearly visible, though one can see some differences concerning their interpretation. For example, for GRB 3, the first tall peak in the CTA band is clearly delayed by several seconds with respect to the *Fermi*-GBM band, whereas, for GRB 5, the precise size of the lag depends on the instrument threshold of CTA.

7. Summary and Conclusions

We have simulated the evolution of a baryon-rich GRB jet in the internal shock scenario of the fireball model, by keeping track of all relativistic plasma shells that propagate in it, of the collisions between shells, and of the gamma-rays, protons, and neutrinos emitted at the shocks produced during the collisions. Unlike traditional one-zone models that extrapolate the behavior of the whole burst from a single representative collision, our multi-zone simulations consider many such collisions, each happening under different physical conditions.

We have demonstrated that it is possible to generate the various features observed in GRB light curves by varying the behavior of the central emitter, in particular, the initial speeds with which it puts out shells (see Section 4). The initial speeds determine where collisions between shells will occur during the jet evolution. In this approach, one can relate the properties of the central emitter to observables.

If the initial distribution of shell speeds is "disciplined" or narrow—even if the average speed changes with time—the associated gamma-ray light curve will have one or more broad pulses overlaid with fast time variability. The associated neutrino flux can be low, because collisions tend to occur at large collision radii. In addition, there is a correlation between observation time and collision radius, which implies that early-time collisions occur at low radii, where radiation densities are high, and high-energy gamma-ray signals are suppressed. As a consequence, we expect delays in the light curves in the *Fermi*-LAT energy band with respect to the ones in the *Fermi*-GBM band of a few seconds, and in the CTA band with respect to the *Fermi*-GBM band of the order of 10 seconds.

If the distribution of speeds is broader, collisions occur over a wide range of collision radii and the light curve is dominated by fast time variability. In this case, neutrino production is always efficient, because, typically, several collisions occur where the radiation densities are high. In this case, we do not expect delays between gamma-ray wavelength bands, because there is not a strong correlation between observation time and

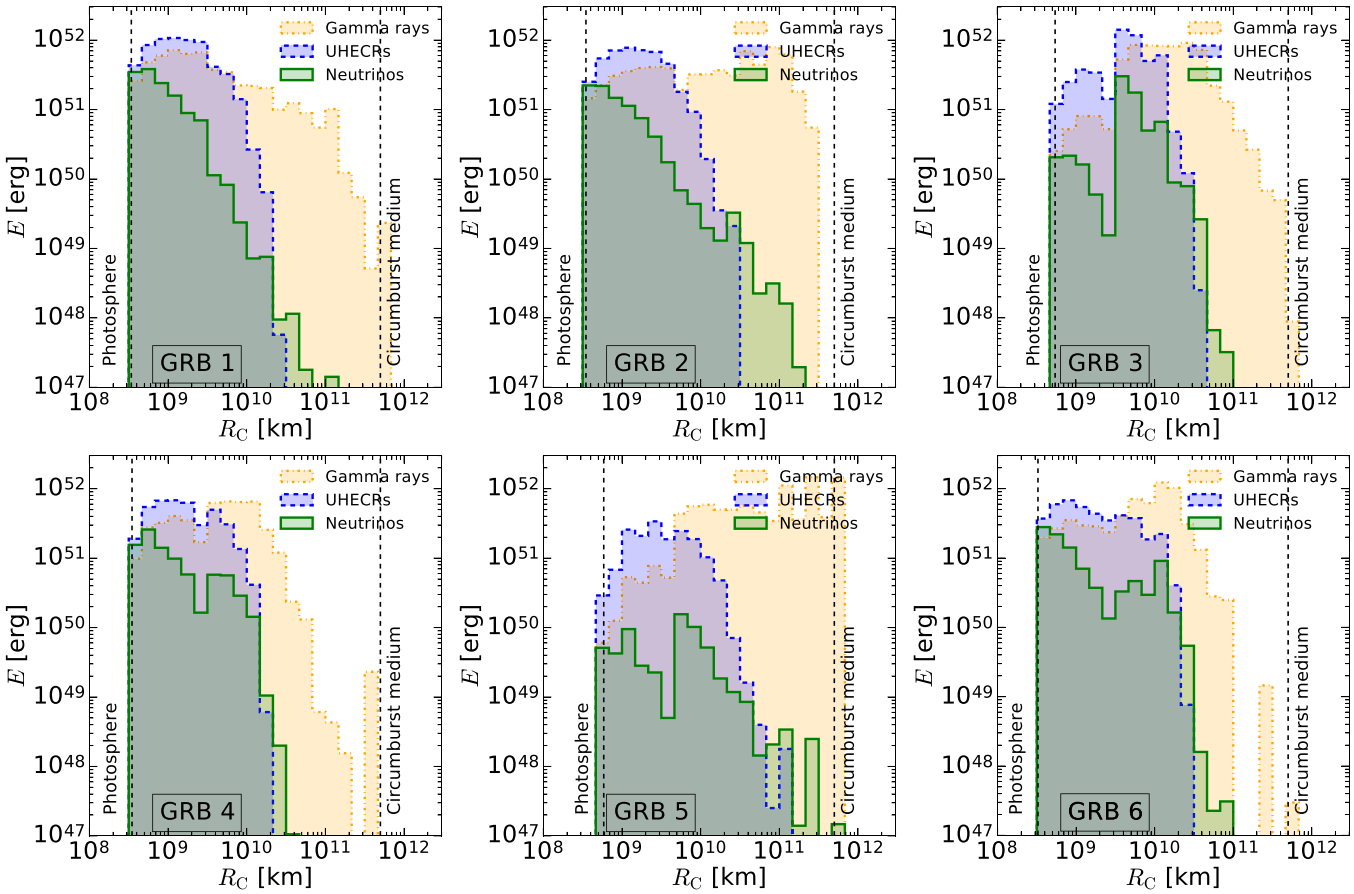

Figure 8. Energy output as a function of collision radius in neutrinos, UHECR protons ($E_p > 10^{10}$ GeV), and gamma-rays. The approximate photospheric and (assumed) circumburst radius are marked, as well as the UHECR escape regions where either neutron escape or direct escape dominates.

collision radius. Inspection of many GRB light curves reveals that most are actually simple, while the ones typically presented in the literature tend to be the more interesting cases with non-trivial features. This means the class of bursts with broad speed distributions may be more representative of the "typical" GRB.

We have also qualitatively confirmed the findings from Bustamante et al. (2015) for very different parameter sets. Notably, we have shown that, regardless of the initial speed distribution of the shells, different messengers come from different regions of the same GRB: neutrinos predominantly come from regions close to the center, UHECR protons come from intermediate regions, and high-energy gamma-rays tend to come from regions further out from the center, where photon densities are low enough that their energies are not limited by pair production on source photons. We have also confirmed the minimal predicted neutrino flux around $\sim 2 \cdot 10^{-11}$ GeV cm^{-2} s^{-1} sr^{-1} per flavor around 1 PeV, as long as a significant fraction of the GRBs has broad initial shell speed distributions, which explain observations better; see, e.g., our examples GRB 1, 2, 4, and 6.

Our main results and conclusions are based exclusively on collisions that occur above the photosphere, where photons are not trapped by Thomson scattering. This allows us to assume that the photon spectra in the shells have the same form as the observed spectra at Earth.

Our results could be exploited in targeted GRB neutrino searches, such as the ones performed by IceCube, to cull a smaller catalog of GRBs that are specially likely to be strong neutrino sources. The non-observation of neutrinos from bursts in such a catalog—where associated backgrounds are smaller—could result in stronger upper bounds on prompt high-energy GRB neutrino emission. Our results could also be tested in different gamma-ray wavelength bands: we do not expect significant delays in GRBs with fast time variability without underlying pulse structure.

Due to their high luminosity, short duration, and the high angular precision with which they are detected, GRBs are arguably our best chance at finding a high-energy neutrino counterpart to electromagnetic emission. The study presented here is a step toward a necessary, realistic multi-messenger understanding of GRBs. The observation of these neutrinos would be a smoking-gun signature of high baryonic loading, and thus of the paradigm that GRBs could be the sources of the UHECRs.

The authors thank John Beacom, Denise Boncioli, Valerie Connaughton, Željka Bošnjak, and Eli Waxman for useful discussion and comments on the manuscript. M.B. acknowledges the hospitality of DESY Zeuthen during the completion of this work. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 646623). M.B. acknowledges support from NSF Grant PHY-1404311. The work of K.M. is supported by NSF Grant PHY-1620777.

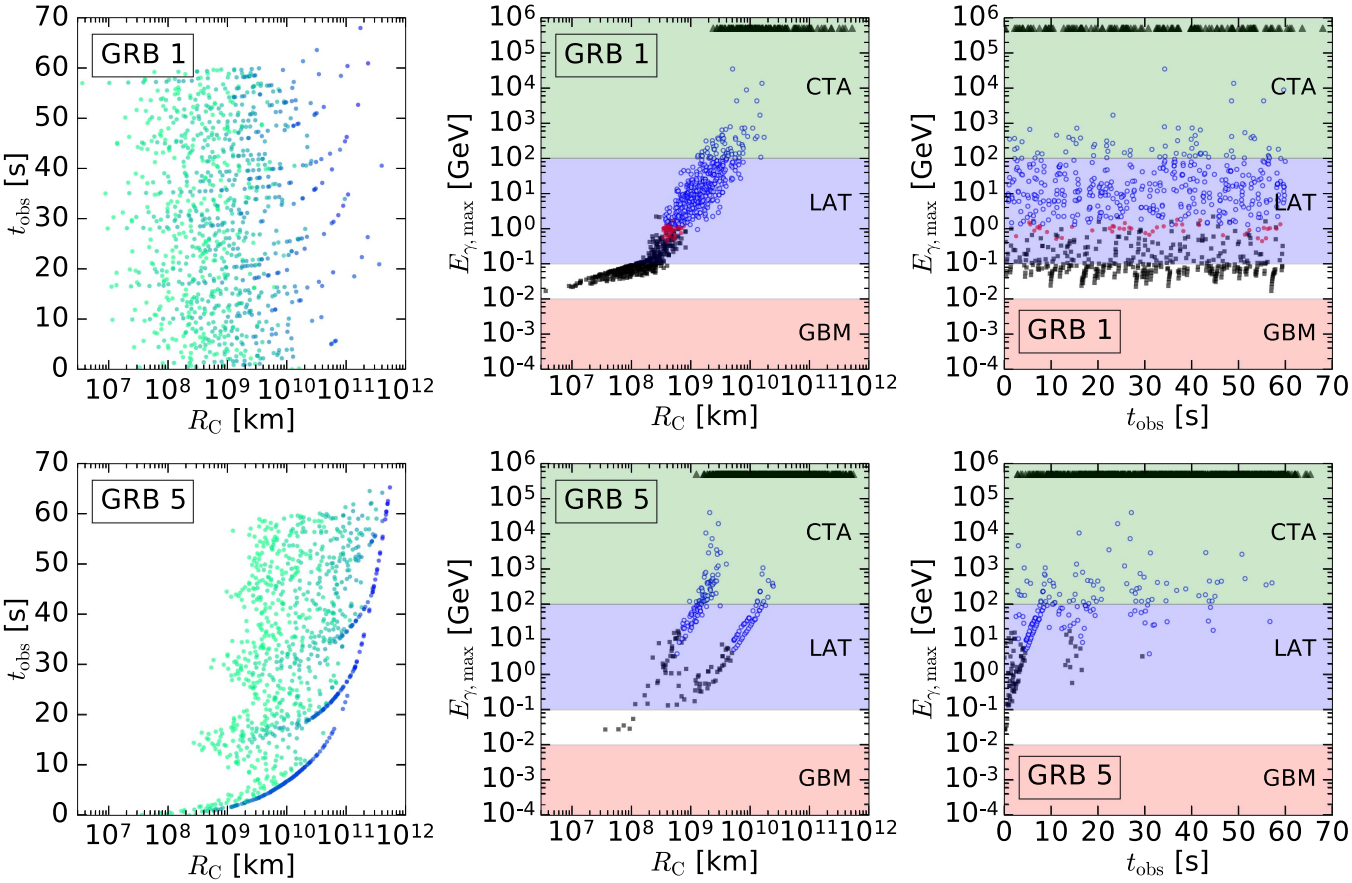

Figure 9. Collision time in the observer's frame as a function of collision radius (left column), and maximum gamma-ray energy in the source frame as a function of collision radius (central column) and time (right column), for collisions in GRBs 1 (top row) and 5 (bottom row). In the left column, collisions between older shells—that have undergone multiple mergers—are darker, while collisions between younger collisions are lighter; the solid black lines are the average trends. In the central and right columns, collisions are labeled as in Figure 6. Energy ranges accessible by *Fermi*-GBM, *Fermi*-LAT, and CTA are shaded. Arrows (▲) represent collisions where the maximum gamma-ray energy is not limited by pair production.

Appendix A
Model Description

Here we describe our GRB jet simulation, based on Kobayashi et al. (1997) and Daigne & Mochkovitch (1998).

A.1. Overview

The central object in a GRB (e.g., the black hole created by a collapsing massive star) emits collimated, relativistic jets that are rich in baryons, which we assume to be protons. When one of the jets points toward Earth, the gamma-ray emission from it may be detected as a GRB. We simulate the particle emission from this jet.

Since we are located inside the opening angle of the jet, we cannot distinguish between the emission geometry of a collimated jet and that of spherical, isotropic emission. For simplicity, we develop our formalism under such an isotropically equivalent scenario. In it, the central engine emits spherical plasma shells that propagate outward along the jet at relativistic speeds. We simulate their propagation and collisions between them, which produce high-energy gamma-rays, protons, and neutrinos.

The simulation covers only the coasting phase of the GRB, during which shells propagate at constant speed, except when they collide. In the preceding, acceleration phase, they reached their maximum individual speeds, limited by the available kinetic energy and their masses. The coasting phase ends when the shells reach the circumburst medium; there, they decelerate, and might produce an afterglow. The acceleration and deceleration phases are not part of the simulation.

In our simulation, shells propagate in one dimension. At any time during the simulation, the kth shell is characterized by four basic parameters: r_k, the shell radius, as measured from the emitter (i.e., the position of the shell inside the jet); l_k, the shell width; Γ_k, the shell bulk Lorentz factor; and m_k, the shell mass.

When two shells collide, they merge into a new shell, with width, speed, and mass calculated from the properties of the shells that collided. The new shell continues propagating in the jet flow and may collide again. Collisions are inelastic. The new shell cools instantly by radiating away its internal energy via particle emission. We compute collisions numerically, following Kobayashi et al. (1997) and Daigne & Mochkovitch (1998), as detailed below.

Table 3 describes all the relevant simulation parameters. Unless otherwise noted, quantities therein and in the text are expressed in the source reference frame.

A.2. Burst Initialization

In the simulation, before shell propagation starts, the central engine has already emitted $N_{\rm sh}$ shells. Each one is described by the initial tuple $(r_{k,0}, l_{k,0}, \Gamma_{k,0}, E^{\rm iso}_{{\rm kin},0})$. Shells closer to the engine are labeled with higher indices.

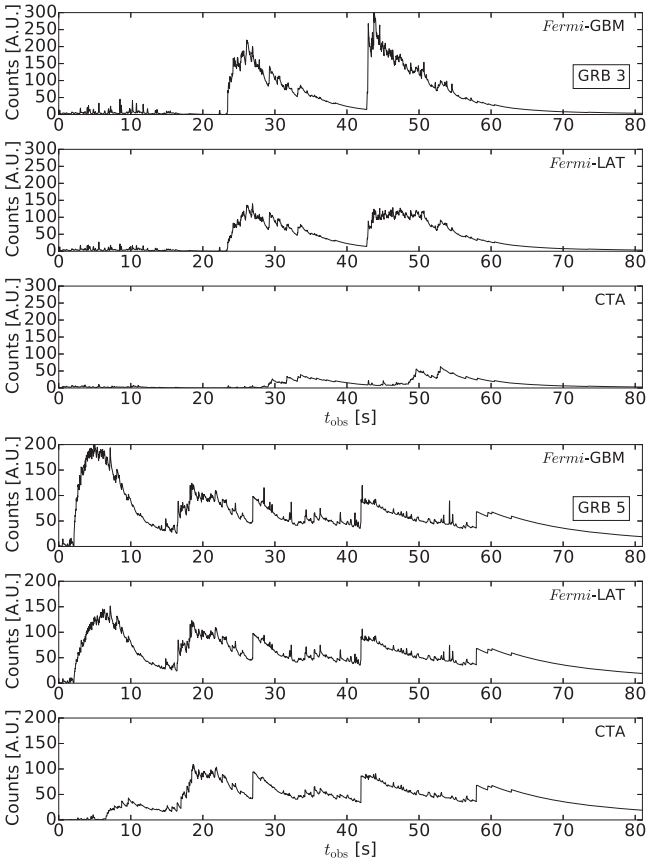

Figure 10. Gamma-ray light curves—using super-photospheric collisions only—for GRBs 3 and 5, in different energy bands: *Fermi*-GBM: 10^{-6}–10^{-2} GeV; *Fermi*-LAT: 10^{-1}–10^{2} GeV; and CTA: 10^{2}–10^{6} GeV.

The behavior of the engine is described by two timescales[13]: an "uptime," $\delta t_{\rm eng}$, during which it emits one shell, followed by a "downtime," $\Delta t_{\rm eng}$, during which it is inactive. The former determines the initial shell width, $l = c \cdot \delta t_{\rm eng}$, which we assume to be common for all shells, and the latter determines the initial separation between consecutive shells, $d = c \cdot \Delta t_{\rm eng}$. Thus, each of the initial shells is located at position $r_{k,0} = r_{N_{\rm sh}} + (N_{\rm sh} - k)(l + d)$, where $r_{N_{\rm sh}}$ is the distance from the innermost shell to the emitter, which is an input parameter of the simulation. Results do not depend on $r_{N_{\rm sh}}$ strongly, unless its value is too large; see Table 3.

We choose the values of l and d to reproduce the timescale of pulses in observed light curves (Nakar & Piran 2002a). If $t_{\rm v}$ is the GRB variability timescale, i.e., the characteristic duration of peaks in the light curve, and $t_{\rm q}$ is the characteristic quiescent time between consecutive peaks, we expect that, roughly, $l \approx c \cdot t_{\rm v}$ and $d \approx c \cdot t_{\rm q}$ (Kobayashi et al. 1997; Aoi et al. 2010). In the internal shock model, d and l should be comparable. The simulations in Kobayashi et al. (1997) set $d = l$, while Aoi et al. (2010) set $d = 5l$. In our simulations, we chose $\delta t_{\rm eng} = \Delta t_{\rm eng}$, such that $d = l = c \cdot \delta t_{\rm eng}$.

The variability timescale of a simulated burst is not an input parameter of the simulation, but a result of it. For our choices of simulation parameter values, we find that the variability timescale, obtained from the post-simulation synthetic light curve (see Appendix A.4), is close to the input value of $\delta t_{\rm eng}$.

The initial values of the shell Lorentz factors follow a predefined distribution. In the benchmark scenario GRB 1, it is a log-normal distribution; see Section 4. Table 1 describes the distributions used in GRBs 2–6.

There are two typical schemes to assign initial masses $m_{k,0}$ to the shells: the equal-mass assumption, i.e., $m_{k,0} = m$ for all k; and the equal-energy assumption, i.e., $m_{k,0} = E^{\rm iso}_{\rm kin,0}/(\Gamma_{k,0}c^2)$, with $E^{\rm iso}_{\rm kin,0}$ the initial bulk kinetic energy, assumed common to all shells. Our simulation uses the latter, since it appears to match observations more closely (Nakar & Piran 2002a).

A.3. Burst Evolution

We simulate the coasting phase of the jet, during which shell speeds do not change while they propagate and expand. In our simplified treatment, the shell width and mass also stay constant.[14] Therefore, the shell volume $V_{{\rm iso},k} = 4\pi r_k^2 l_k$ grows $\propto r_k^2$ and mass density $\rho_k = m_k/V_{{\rm iso},k}$ decreases $\propto r_k^{-2}$. Since the shell mass and Lorentz factor are constant during propagation, its bulk kinetic energy $E^{\rm iso}_{{\rm kin},k} = \Gamma_k m_k c^2$ is constant as well. Speed, width, and mass change only in collisions.

At the start of the simulation, we calculate the collision time for all pairs of neighboring shells, i.e.,

$$\Delta t_{k,k+1} = \frac{d_{k,k+1}}{c(\beta_{k+1} - \beta_k)}\,, \tag{3}$$

where $d_{k,k+1} \equiv r_k - r_{k+1} - l_{k+1}$ is the separation between shells k and $k+1$. The time interval until the next collision occurs is the minimum of these times, i.e.,

$$\Delta t_{\rm next} = \min[\Delta t_{k,k+1}]. \tag{4}$$

We increase the simulation time to $t \to t + \Delta t_{\rm next}$. The collision radius $R_{\rm C}$ is set to the radius of the innermost colliding shell. Light emitted from this collision will be detected by a distant observer at time

$$t_{\rm obs} = \left(\frac{D(z) - R_{\rm C}}{c} + t\right)(1+z), \tag{5}$$

where $D(z)$ is the light-travel distance to the emitter with redshift z. These equations satisfy well-known relations in the internal shock scenario, $\Delta t_{\rm next} \approx 2\Gamma^2(d/c)$ and $R_{\rm C} \approx 2\Gamma^2 d$. All shells propagate to their new positions $r_k \to r_k + c\beta_k \Delta t_{\rm next}$, the time interval for the next collision is calculated, and the process is repeated. (The term $D(z)/c$ is just an offset: it will disappear when, in the simulation output, the first emission is set to start at $t_{\rm obs} = 0$.)

We distinguish between the shell bulk kinetic energy, $E^{\rm iso}_{{\rm kin},k}$, and its internal energy, $E^{\rm iso}_{{\rm int},k}$. The former is related to the motion of the compact shell, measured in the source rest frame, while the latter is the aggregated kinetic energy of particles moving randomly inside the shell, measured in the shock rest frame.

In a collision, the kinetic energy of the two colliding shells is used partly as bulk kinetic energy for the new shell and partly

[13] This is strictly true for the simulated GRBs 1–5. GRB 6 has an overlaid time structure: the engine has an overall active period, where it emits $N_{\rm up}$ shells, followed by a quiescent period that lasts for $N_{\rm down}$ pulses. See Table 1.

[14] Depending on the internal energy, shell spreading is important especially after collisions. Recent dedicated simulations take into account this effect, but it is neglected in the simplest versions, like the one we have adopted (Aoi et al. 2010).

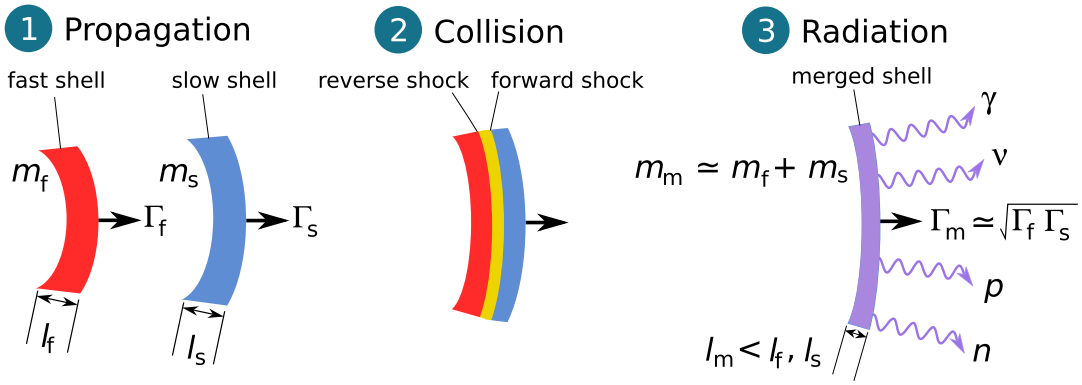

Figure 11. Schematic illustration of the collision process between two plasma shells, and the ensuing emission of high-energy particles.

as its internal energy. For simplicity, we assume that the new shell immediately cools by prompt particle emission; see Kobayashi & Sari (2001) for alternative treatments. While collision details depend on modeling of hydrodynamical properties (Daigne & Mochkovitch 2003), here we adopt the simple collision prescription for the relativistic limit introduced in Kobayashi et al. (1997), which we outline below. Figure 11 is a schematic illustration of the collision process.

In the collision of a slow (s) and a fast (f) shell, the internal energy of the merged (m) shell is the difference of kinetic energy before and after the collision, i.e.,

$$E_{\rm coll}^{\rm iso} = (\Gamma_{\rm f} m_{\rm f} + \Gamma_{\rm s} m_{\rm s}) c^2 - \Gamma_{\rm m}(m_{\rm s} + m_{\rm f}) c^2 \,. \tag{6}$$

We assume that this amount of internal energy of the merged shell is radiated away as secondary particles. From momentum and energy conservation, and assuming $\Gamma_{\rm f}, \Gamma_{\rm s} \gg 1$, the Lorentz factor of the merged shell is

$$\Gamma_{\rm m} \simeq \sqrt{\frac{\Gamma_{\rm f} m_{\rm f} + \Gamma_{\rm s} m_{\rm s}}{m_{\rm f}/\Gamma_{\rm f} + m_{\rm s}/\Gamma_{\rm s}}}, \tag{7}$$

which reduces to $\Gamma_{\rm m} \simeq \sqrt{\Gamma_{\rm f}\Gamma_{\rm s}}$ if $m_{\rm f} \simeq m_{\rm s}$. Its width is given by (Kobayashi et al. 1997; Aoi et al. 2010)

$$l_{\rm m} \simeq l_{\rm s} \frac{\beta_{\rm fs} - \beta_{\rm m}}{\beta_{\rm fs} - \beta_{\rm s}} + l_{\rm f} \frac{\beta_{\rm m} - \beta_{\rm rs}}{\beta_{\rm f} - \beta_{\rm rs}}, \tag{8}$$

where $\beta_{\rm fs(rs)} = \sqrt{1 - \Gamma_{\rm fs(rs)}^{-2}}$ is the speed of the forward (reverse) shock, whose Lorentz factor is

$$\Gamma_{\rm fs(rs)} = \Gamma_{\rm m} \sqrt{\frac{1 + 2\Gamma_{\rm m}/\Gamma_{\rm s(f)}}{2 + \Gamma_{\rm m}/\Gamma_{\rm s(f)}}} \,. \tag{9}$$

The volume of the new shell is $V_{\rm iso,m} = 4\pi R_{\rm C}^2 l_{\rm m}$. The density is different between the shocked faster shell and shocked slower shell. The new shell has an average density obtained from the assumption of an inelastic collision $m_m \simeq m_{\rm f} + m_{\rm s}$ already implied in Equation (6), i.e.,

$$\rho_{\rm m} \simeq \frac{l_{\rm f} \cdot \rho_{\rm f} + l_{\rm s} \cdot \rho_{\rm s}}{l_{\rm m}}. \tag{10}$$

Therefore, its mass is $m_{\rm m} = V_{\rm iso,m}\rho_{\rm m}$, and its kinetic energy is $E_{\rm kin,m}^{\rm iso} = \Gamma_{\rm m} m_{\rm m} c^2$. After the collision, the original fast shell is removed from the simulation and the new shell replaces the former slow shell. It is then propagated with the remaining shells in the jet.

If a shell reaches the circumburst medium, where it decelerates, it is removed from the simulation. Following Rees & Mészáros (1992), we assume $r_{\rm dec} = 5.5 \cdot 10^{11}$ km for the radius at which this happens (see, e.g., Equation (15) in Mészáros 2006).

The simulation finishes when all shells have reached the circumburst medium, all shells have merged into a single remaining shell, or all remaining shells are ordered outward with increasing Lorentz factor, so that no more collisions are possible. The output lists $N_{\rm coll}$ collisions,

$$(t_{{\rm obs},k},\, R_{{\rm C},k},\, l_{{\rm m},k},\, \Gamma_{{\rm m},k},\, E_{{\rm coll},k}^{\rm iso}), \tag{11}$$

where $1 \leqslant k \leqslant N_{\rm coll}$. The minimum $t_{{\rm obs},k}$ is taken to be the start of the observation time of the burst and is set to zero. Collisions are arranged so that $t_{\rm obs,1} = 0 \leqslant t_{\rm obs,2} \leqslant \ldots \leqslant t_{{\rm obs},N_{\rm coll}}$.

Figure 12 shows the time evolution (in the source frame) of macroscopic burst parameters in one of our simulations: average shell mass $\langle m \rangle / \langle m_0 \rangle$ (subscripts of zero indicates values at simulation start), standard deviation of the Lorentz factor[15] $\sigma(\Gamma)/\sigma(\Gamma_0)$, and total available internal energy of the burst $E_{\rm int,tot}^{\rm iso}/E_{\rm int,tot,0}^{\rm iso}$. The latter is calculated directly as $E_{\rm int,tot}^{\rm iso} = \Gamma_{\rm CM} \sum m_i (\tilde{\beta}_i^2/2)$, where $\tilde{\beta}$ is the speed in the CM-frame.[16] The numerical results of our simulation match the analytical power-law estimates from Beloborodov (2000), which assume that fluctuations in the initial Lorentz factors are small, i.e., $A_\Gamma \ll 1$. Deviations occur at late times, when the number of remaining shells is low and the analytical predictions are no longer applicable. This late deviation depends strongly on the random initial setup, so we show ranges obtained after running 1000 different simulations.

Changing the collision dynamics can affect the burst efficiency—i.e., the fraction of kinetic energy dissipated as secondaries—and the radii where most of the gamma-ray energy is dissipated. Changes could include re-converting a fraction of collision energy into kinetic energy, partially inelastic collisions, or even fully penetrating shells. We explore simple modifications of our collision model in Appendix B.

[15] The reference Beloborodov (2000) refers to this as $\Gamma_{\rm rms}$. However, we repeated the derivations of their analytical estimates, which are consistent only when interpreting this as the standard deviation.

[16] The general relation between the total internal, or free, energy of a gas, $E_{\rm int,tot}^{\rm iso}$, and its volume is given by $E_{\rm int,tot}^{\rm iso} \propto V_{\rm iso}^{-(\gamma-1)} \propto r^{-2(\gamma-1)}$, where γ is the adiabatic index. For the relativistic gas in a shell, $\gamma = 4/3$ and $E_{\rm int,tot}^{\rm iso} \propto r^{-2/3} \propto t^{-2/3}$.

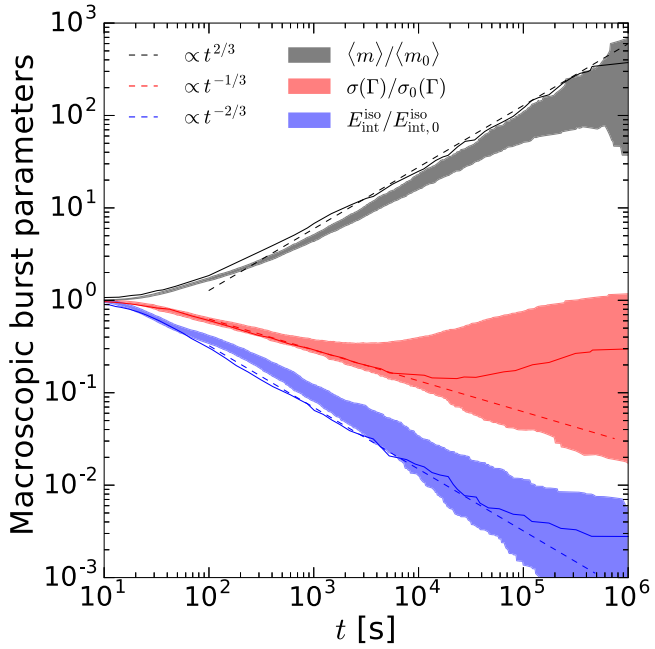

Figure 12. Time evolution, in the source frame, of average shell mass $\langle m\rangle/\langle m_0\rangle$, standard deviation of the Lorentz factor $\sigma(\Gamma)/\sigma(\Gamma_0)$, and total internal energy of a burst, $E^{\rm iso}_{\rm int,tot}/E^{\rm iso}_{\rm int,tot,0}$. The ranges are from our numerical results from a 1000 simulations run with random setups for the parameter values $N_{\rm sh} = 10000$, $\Gamma_0 = 100$, $A_\Gamma = 0.2$, $\delta t_{\rm eng} = 10^{-3}$ s, $d = l$, $z = 2$, and $E^{\rm iso}_{\rm kin,0} = 10^{52}$ erg. Solid and dashed lines come from Figure 1 in Beloborodov (2000) for the same parameter set and refer to numerical calculations and analytical estimates respectively.

A.4. Gamma-Ray Observables

The internal energy of a merged shell is split among electrons, magnetic field, and protons. They receive a fraction ϵ_e, ϵ_B, and ϵ_p, respectively. We assume energy equipartition between electrons and photons and fix $\epsilon_e = \epsilon_B = 1/12$ and $\epsilon_p = 5/6$, since this yields the frequently used value of baryonic loading $1/f_e = \epsilon_p/\epsilon_e = 10$ (Waxman & Bahcall 1999; Abbasi et al. 2012). Thus, the kth collision dissipates an energy $E_{\gamma,k} = \epsilon_e E_{{\rm coll},k}$ as gamma-rays, and energy $E_{\gamma,p} = \epsilon_p E_{{\rm coll},k}$ as protons, and supports a magnetic energy density of $U_B = \epsilon_B E_{{\rm coll},k}/V_{\rm iso}$. The latter translates into a magnetic field intensity, in the shock rest frame, of

$$B' \simeq 44.7\left(\frac{\Gamma_m}{10^{2.5}}\right)^{-1}\left(\frac{\epsilon_B}{\epsilon_e}\right)^{\frac{1}{2}}\left(\frac{E_{\gamma,k}}{10^{50}\ {\rm erg}}\right)^{\frac{1}{2}} \times \left(\frac{R_{\rm C}}{10^9\ {\rm km}}\right)^{-1}\left(\frac{l_{\rm m}}{10^3\ {\rm km}}\right)^{-\frac{1}{2}}{\rm kG}\,. \quad (12)$$

We normalize these individual energies by requiring that the total dissipated energy as gamma-rays, an experimentally accessible quantity,

$$E^{\rm iso}_{\gamma,{\rm tot}} \equiv \sum_{k=1}^{N_{\rm coll}} E^{\rm iso}_{\gamma,k}, \quad (13)$$

matches a given value $E^{\rm iso}_{\gamma,{\rm norm}} = 10^{53}$ erg. This also fixes the energy in protons and magnetic field.

Our simulation does not generate the photon spectrum in the shell. Instead, as in Aoi et al. (2010), Abbasi et al. (2012), we assume that its shape is that of observed GRB spectra. We parametrize the spectrum as a broken power law, i.e.,

$$n'_\gamma(\varepsilon') \equiv \frac{dn'_\gamma}{d\varepsilon'} = C'_\gamma \begin{cases} (\varepsilon'/\varepsilon'_{\gamma,{\rm break}})^{-\alpha_\gamma}, & \varepsilon' < \varepsilon'_{\gamma,{\rm break}} \\ (\varepsilon'/\varepsilon'_{\gamma,{\rm break}})^{-\beta_\gamma}, & \varepsilon' \geqslant \varepsilon'_{\gamma,{\rm break}} \end{cases}. \quad (14)$$

Primed quantities are in the shock rest frame. We fix[17] $\alpha_\gamma = 1$, $\beta_\gamma = 2$, and $\varepsilon'_{\gamma,{\rm break}} = 1$ keV.

The photon spectrum in each shell is normalized via

$$V'_{{\rm iso},k}\int_{\varepsilon'_{\gamma,\min}}^{\varepsilon'_{\gamma,\max}} d\varepsilon'\varepsilon' n'_\gamma(\varepsilon') = \frac{E^{\rm iso}_{\gamma,k}}{\Gamma_{{\rm m},k}}, \quad (15)$$

where the minimum and maximum energies are $\varepsilon'_{\gamma,\min} = 0.2$ eV and $\varepsilon'_{\gamma,\max} = 1$ PeV, respectively (Aoi et al. 2010). Pair production via $\gamma\gamma \to e^+e^-$ may limit the maximum energy of escaping photons; see Figure 9, right column.

Each collision emits a gamma-ray pulse. The superposition of all pulses, propagated to Earth, is the light curve of the burst; see Section 4. Following Kobayashi et al. (1997), we parametrize the luminosity of the pulse from the kth collision (in the observer's frame) as a peaked profile, with a fast rise and exponential decay ("FRED"), i.e.,

$$L_{\gamma,k}(t_{\rm obs}) = \begin{cases} 0, & t_{\rm obs} < 0 \\ h_k\left[1 - \frac{1}{(1+ct/R_{{\rm C},k})^2}\right], & 0 \leqslant t_{\rm obs} < t_{{\rm rise},k} \\ h_k\left\{\frac{1}{[1+(t-\delta t_{{\rm em},k})c/R_{{\rm C},k}]^2}\right. & \\ \left. \quad - \frac{1}{(1+ct/R_{{\rm C},k})^2}\right\}, & t_{\rm obs} \geqslant t_{{\rm rise},k} \end{cases} \quad (16)$$

where the emission timescale, i.e., the time at which the reverse shock crosses the fast shell, is

$$\delta t_{{\rm em},k} \equiv \frac{l_{{\rm m},k}}{c(\beta_{\rm f} - \beta_{\rm rs})} \quad (17)$$

and the "rise time,"

$$t_{{\rm rise},k} \equiv \frac{\delta t_{{\rm em},k}}{2\Gamma^2_{{\rm m},k}}(1+z), \quad (18)$$

is the time elapsed since the start of the emission until the peak luminosity is reached. For an illustration, see Figure 1 in Kobayashi et al. (1997). The peak luminosity is

$$h_k = \frac{E^{\rm iso}_{\gamma,k}}{1+z}\frac{1}{t_{{\rm rise},k}}. \quad (19)$$

[17] It is uncertain how $\varepsilon'_{\gamma,{\rm break}}$ changes with $R_{\rm C}$, since the scaling expected in the internal shock model has a problem (Daigne & Mochkovitch 2003), which is why we prefer to set the photon spectra to observations. However, depending on the origin of the prompt gamma-ray emission—e.g., synchrotron radiation, inverse-Compton scattering—one can implement model-specific assumptions, as in Guetta et al. (2001a, 2001b; though the models therein are already in tension with GRB neutrino searches by IceCube (Abbasi et al. 2012; Aartsen et al. 2015a)). To correctly calculate the photon break energy for each collision, one needs a time-dependent radiative code, as implemented in Bošnjak et al. (2008). Detailed information on mildly relativistic collisionless physics, such as the injection Lorentz factor (which depends on the number fraction of accelerated particles (Eichler & Waxman 2005)), is also necessary. However, such an improvement will not change our conclusions in Section 7.

The time t in the source frame is related to $t_{\rm obs}$ through

$$t = \frac{2\Gamma_{{\rm m},k}^2 t_{\rm obs}}{1+z}\,. \tag{20}$$

Hence, the synthetic light curve L_γ is

$$L_\gamma(t_{\rm obs}) = \sum_{k=1}^{N_{\rm coll}} L_{\gamma,k}(t_{\rm obs}). \tag{21}$$

Unless noted otherwise, we only show the light curves for collisions beyond the photosphere (see Section 4), while we use all collisions for the normalization in Equation (13). In our simulations, the fraction of energy dissipated below the photosphere is around 50%, so that the total super-photospheric energy output in gamma-rays is around half of $E_{\gamma,{\rm norm}}^{\rm iso}$.

In Section 6, we showed light curves in different energy bands. Each band spans the range $[E_{\gamma,{\rm low}}^{\rm band}, E_{\gamma,{\rm high}}^{\rm band}]$, with $E_{\gamma,{\rm low}}^{\rm band} = 10^{-6}$, 10^{-1}, 10^{2} GeV, and $E_{\gamma,{\rm high}}^{\rm band} = 10^{-2}$, 10^{2}, 10^{6} GeV, for *Fermi*-GBM, *Fermi*-LAT, and CTA, respectively. To compute the gamma-ray contribution of the kth collision to each band, we calculate the fraction

$$f_k^{\rm band} = \begin{cases} 0, & E_{\gamma,{\rm max},k} < E_{\gamma,{\rm low}}^{\rm band} \\ 1, & E_{\gamma,{\rm max},k} > E_{\gamma,{\rm high}}^{\rm band} \\ \dfrac{\log(E_{\gamma,{\rm max},k}) - \log(E_{\gamma,{\rm low}}^{\rm band})}{\log(E_{\gamma,{\rm high}}^{\rm band}) - \log(E_{\gamma,{\rm low}}^{\rm band})} & E_{\gamma,{\rm low}}^{\rm band} \leqslant E_{\gamma,{\rm max},k} \leqslant E_{\gamma,{\rm high}}^{\rm band} \end{cases}, \tag{22}$$

where $E_{\gamma,{\rm max},k}$ is the maximum gamma-ray energy emitted by the collision. The light curve for each band is then simply calculated using Equation (21), replacing $L_{\gamma,k} \to f_k^{\rm band} L_{\gamma,k}$. Note that this simplified treatment assumes that the product of instrument response times flux is approximately constant within the anticipated energy band—which is typically a good estimate within the energy bands the instruments are designed for. A more detailed model for the instrument response or a different shape of the target photon spectrum at TeV energies—which we assumed to be a power law beyond the break—will not affect the qualitative shape of the light curves, but may slightly change the relative power in different energy bands or light-curve peaks.

The burst duration and variability timescale $t_{\rm v}$ are derived from the light curve. For the duration, we use T_{90}, the time elapsed between the detection of 5% and 95% of the total gamma-ray energy, i.e.,

$$T_{90} \equiv t_{95} - t_5, \tag{23}$$

where

$$\int_0^{t_f} L_\gamma(t_{\rm obs})\,dt_{\rm obs} = f\,\frac{E_{\gamma,{\rm tot}}^{\rm iso}}{1+z} \tag{24}$$

and $f = 0.05$ or 0.95. The variability timescale is estimated as

$$t_{\rm v} = T_{90}/N_{\rm coll}. \tag{25}$$

This procedure yields values of $t_{\rm v}$ close to $\delta t_{\rm eng}$.

In some cases (e.g., Figure 7), we have compared simulation results to "standard" estimators from the one-zone model:

$$t_{\rm v}^{\rm std} = \frac{d+l}{c}(1+z), \tag{26}$$

$$T^{\rm std} = N_{\rm coll}\, t_{\rm v}^{\rm std}, \tag{27}$$

$$\Gamma_{\rm m}^{\rm std} = \langle\Gamma_{\rm m}\rangle, \tag{28}$$

$$R_{\rm C}^{\rm std} = 2(\Gamma_{\rm m}^{\rm std})^2 \frac{c t_{\rm v}^{\rm std}}{1+z}. \tag{29}$$

A.5. Neutrinos and Cosmic-rays

In analogy to gamma-rays, the proton and neutrino spectra of the complete burst are obtained by summing over the spectra emitted by all the individual collisions.

We compute secondary particle production using the NeuCosmA software (Baerwald et al. 2012; Hümmer et al. 2012; Baerwald et al. 2013). This assumes a proton density $n_p' \propto E_p'^{-2}\exp(-E_p'/E_{p,{\rm max}}')$, with the maximum proton energy $E_{p,{\rm max}}'$ obtained from balancing the acceleration rate (with perfect efficiency) with synchrotron, adiabatic, and photohadronic energy loss rates (Baerwald et al. 2013). The proton density is normalized like the photon density, Equation (15), but replacing $E_{\gamma,k}^{\rm iso} \to E_{\gamma,k}^{\rm iso}/f_e$. Secondary pion, muon, and kaon spectra, and, consequently, neutrino spectra are computed as in Hümmer et al. (2012), which includes magnetic field effects on the secondaries, state-of-the-art normalization of the spectra, helicity-dependent muon decays, and flavor mixing.

Following Baerwald et al. (2013), UHECRs escape from each shell as neutrons, produced in photohadronic interactions ("neutron escape"), and as protons that leak out of the shell when their Larmor radius exceeds the shell thickness ("direct escape"). At the highest energies, protons can always leak out—provided their maximum energy is limited by adiabatic cooling. However, which escape component dominates in each shell depends on the optical depth to photohadronic interactions of the shell in question. Figure 13 is a schematic illustration of the components contributing to UHECR emission depending on the optical depth of the merged shell. (If magnetic fields decay fast enough in the bulk of the shell, the direct escape component may be larger. However, this possibility is not contemplated in our framework.)

Our main results, e.g., our light curves and neutrino spectra, are based only on shell collisions that occurred above the photosphere. Below it, Thomson scattering off electrons keeps the photons trapped in the shell. Since our adopted photon spectra are chosen to reproduce observed gamma-ray spectra, we cannot accurately calculate secondary production below the photosphere, where the photon spectra might be different. We mark "sub-photospheric" collisions clearly as such (see, e.g., Figures 5 and 6) and exclude them from our flux calculations. Excluding sub-photospheric collisions does not qualitatively change the shape of the light curves. However, in cases with broader pulse structures, the onset of each pulse is usually dominated by sub-photospheric collisions. Excluding these collisions slightly delays the onset of each peak.

The optical depth to Thomson scattering is calculate from shell properties. Since shells are, on average, electrically

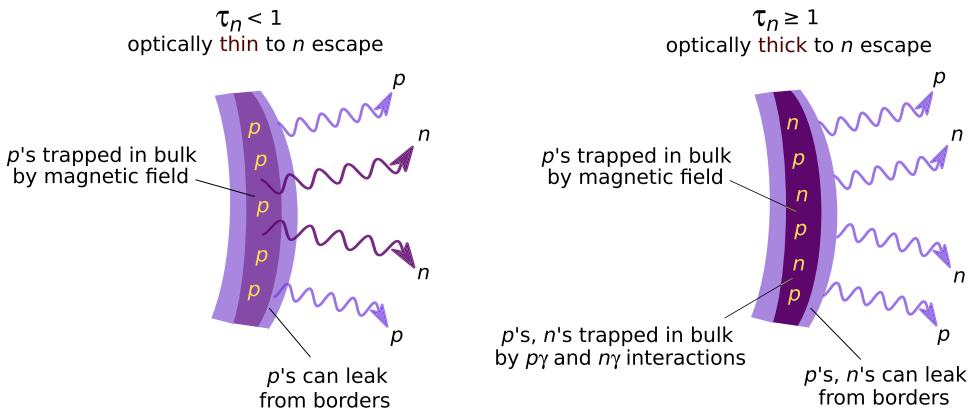

Figure 13. Schematic illustration of the UHECR emission from a merged shell produced in a collision, when the shell is optically thin (left) and thick (right) to $p\gamma$ and $n\gamma$ interactions.

Table 4
Description of Simulated GRBs 7–10

Model	$\Gamma_{0,1}$	$A_{\Gamma,1}$	$\Gamma_{0,2}$	$A_{\Gamma,2}$	T_p	$N_{\rm up}$	$N_{\rm down}$	$E^{\rm iso}_{\gamma,\rm norm}$ (erg)	Description
7	5	⋯	1500	⋯	⋯	⋯	⋯	10^{53}	Box-like Γ distribution
8	50	0.1	500	1.0	⋯	⋯	⋯	10^{53}	Linear speedup of Γ
9	500	1.0	⋯	⋯	⋯	100	100	10^{52}	100 pulses separated by 100 pulse-times
10	500	1.0	⋯	⋯	⋯	300	350	10^{52}	300 pulses separated by 350 pulse-times

Note. Common values for all models: $N_{\rm sh} = 1000$, $\delta t_{\rm eng} = 10^{-2}$ s, $d = l = c \cdot \delta t_{\rm eng}$, $r_{\rm min} = 10^3$ km, $r_{\rm dec} = 5.5 \cdot 10^{11}$ km, $z = 2$, $\epsilon_e = \epsilon_B = 1/12$, $\epsilon_p = 5/6$, $\eta = 1.0$ (acceleration efficiency, see Baerwald et al. 2013). See Table 3 for an explanation of each parameter.

neutral, the electron density is equal to the proton density, i.e.,

$$n'_{e,k} \simeq \frac{m_k}{m_p\, V'_{{\rm iso},k}}, \tag{30}$$

where m_p is the proton mass. This assumes that electron–positron pair production does not increase the electron and positron densities significantly. The optical depth to Thomson scattering is then

$$\tau'_{{\rm Th},k} \simeq \frac{1}{n'_{e,k}\, \sigma_{\rm Th}\, l'_k}, \tag{31}$$

with $\sigma_{\rm Th} \approx 66.52$ fm^2 the Thomson cross-section. A collision is sub-photospheric if $\tau'_{\rm Th} > 1$.

For a burst with $E^{\rm iso}_{\gamma,\rm tot} \simeq 10^{53}$ erg, $\epsilon_e = 1/12$, and a dissipation efficiency of $\varepsilon = 25\%$ (such as GRB 1), the initial kinetic energy per shell is about $10^{51.6}$ erg if 1000 collisions occur. This yields $m_{k,0} = E^{\rm iso}_{{\rm kin},0}/\Gamma_{k,0}c^2 \simeq 10^{49}$ erg for $\Gamma_{k,0} \sim 500$. From Equations (30) and (31), the photospheric radius of the kth shell is

$$R_{{\rm ph},k} \simeq 1.8 \cdot 10^8\ {\rm km} \left(\frac{m_k}{10^{49}\ {\rm erg}}\right)^{1/2}. \tag{32}$$

Bustamante et al. (2015) showed that, since the photohadronic optical depth scales similarly to Equation (31) (replacing $\sigma_{\rm Th} \to \sigma_{p\gamma}$), the pion production efficiency at the photosphere is independent of the isotropic volume and only weakly dependent on Γ_k. The total neutrino flux of the burst is dominated by a few collisions that occur just above the photosphere, where the pion production efficiency is highest.

Appendix B
Additional Simulations

Here we include four additional examples of GRB simulations, GRBs 7–10, that complement the ones showed in the main text. Table 4 describes the simulations. Figure 14 shows the corresponding light curves.

The light curves of GRB 7 and GRB 8 are similar to that of GRB 1; see Figure 4. The light curves of GRB 9 and GRB 10 are similar to that of GRB 4 and GRB 6.

The similarities in the light curves exist in spite of fundamental differences between the behavior of the engine. This illustrates our point that the qualitative behavior of the examples shown in this work are representative of a larger class of models.

Appendix C
Alternative Collision Dynamics

Here we discuss the impact of modifications to our canonical collision model, which is used in the main text and described in Appendix A. We focus on alternative scenarios that can be easily implemented in our framework; that is, we assume that, in each collision, the colliding shells merge and do not consider the case in which they reflect off each other, as in Kobayashi & Sari (2001).

One extreme modification is to remove colliding shells from the system after they collide and radiate, which means that multiple collisions are not allowed. This makes simulation results insensitive to details of how shells are treated after colliding. However, removing the shells will modify the whole system, since, in the canonical collision model, collisions

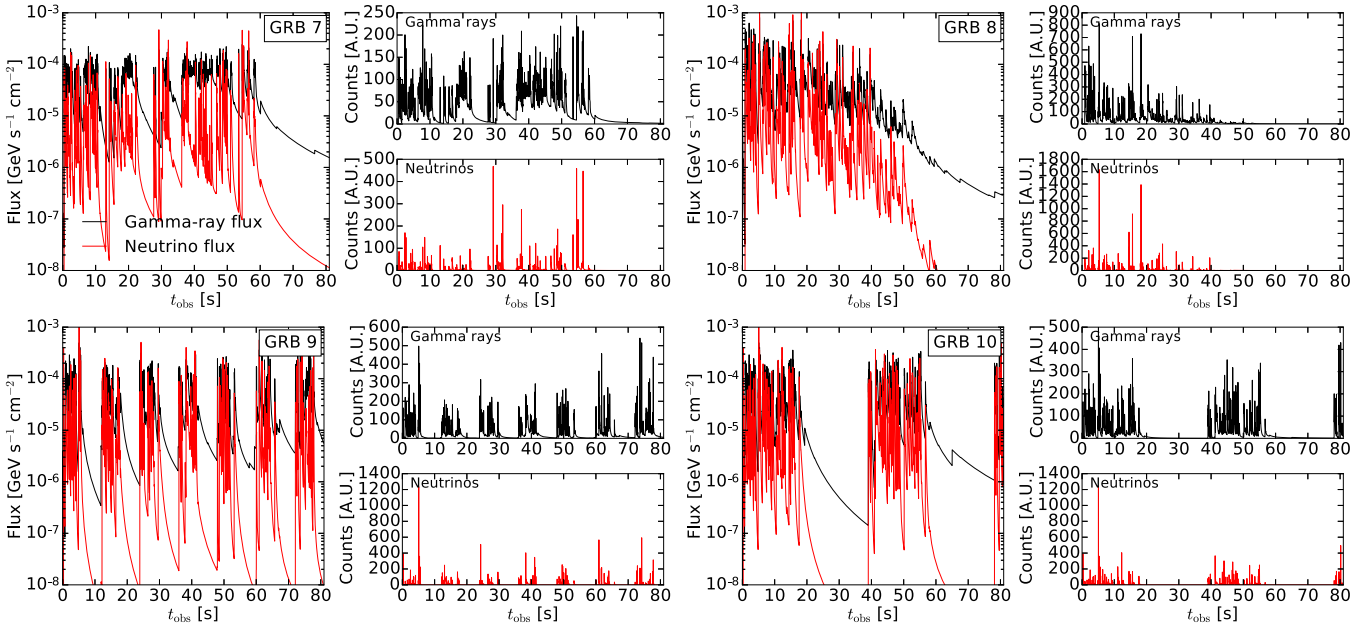

Figure 14. Synthetic gamma-ray and neutrino light curves for the simulated GRBs 7–10, from collisions beyond the photosphere. Photon counts are in arbitrary units, obtained by multiplying the flux times a factor of 10^6 GeV^{-1} cm^2 s.

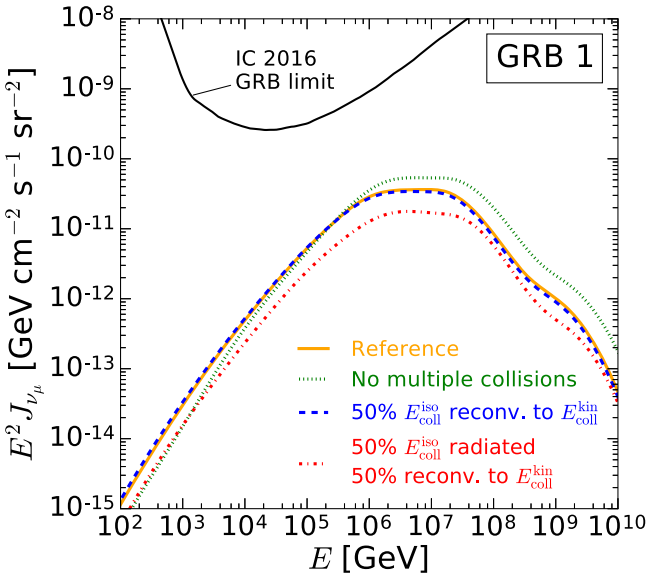

Figure 15. Three modifications to the canonical collision model described in Appendix A and applied to GRB 1. The modified case with no multiplication collisions (green dotted) uses the same per-collision normalization as the reference case (orange, solid). The other two modifications—50% of internal energy reconverted to kinetic energy (blue, dotted–dashed) and 50% radiated with 50% reconverted (red, dashed)—are normalized so in each case the burst yields 10^{53} erg in gamma-rays when adding sub-photospheric and super-photospheric collisions.

among old shells, and between young and old shells, occurred relatively early on.

Figure 15 shows the effect on the quasi-diffuse neutrino flux of applying this modification to a simulation that has the same parameters as our reference case, GRB 1. The modified case is labeled "no multiple collisions." The collision energies were normalized using $E^{\rm iso}_{\gamma,{\rm norm}} = 10^{53}$ erg. The number of collisions is reduced to 499, about half that of the reference GRB 1 model. The modification results in a higher neutrino flux, because, by forbidding multiple-time collisions, most collisions —all of them first-time—occur at low radii, around $10^{8.5}$ km. As we normalize to the same total energy, the per-collision normalization is slightly higher, which further increases the neutrino flux.

Another modification is to assume that only a fraction η of the internal energy in each collision, Equation (6), is attributed to the non-thermal spectra of the secondaries, while $1-\eta$ is instantly reconverted into kinetic energy of the merged shell. This fraction $1-\eta$ can, for instance, describe a fraction of thermal protons not directly participating in the prompt emission. For simplicity, we assume that the extra kinetic energy translates into an instantaneous increase of the Lorentz factor the merged shell after cooling: $\Gamma_{\rm m} = [(1-\eta)E^{\rm iso}_{\rm coll} + E^{\rm iso}_{\rm kin,m}]/m_{\rm m}$.

However, another modification is linked to our assumption that protons can only directly escape the merged shells—and not leave them by diffusion (Baerwald et al. 2013) or other processes—which implies that a substantial fraction of the non-thermal baryonic energy will remain trapped by the magnetic fields and eventually be reconverted into kinetic energy. The typical fraction ξ of electromagnetic energy and non-thermal baryonic energy, which is actually radiated for this escape process, is 40%–50%, estimated from energy partition and from the proportion of baryonic energy in the UHECR energy range compared to the full energy range. To account for this, we consider that the amount of internal energy used for computing the secondary production is still given by Equation (6), but we modify the dynamics so that a fraction of internal energy is reconverted into kinetic energy. This case does not include a fraction of energy going into thermal protons, unlike the previous case.

Figure 15 shows the result of both of these modifications to the kinetic and radiated energy, for the case $\eta = \xi = 0.5$. In both cases, the burst was normalized to yield 10^{53} erg in gamma-rays when adding sub-photospheric and super-photospheric collisions. The number of collisions is similar to that of the reference GRB 1 model. Because the Lorentz factors of the

merged shells are higher due to the increased kinetic energy, collisions occur further out in the jet, where particle densities are lower and neutrino production is less efficient. As a result, the neutrino flux associated with these two modifications is slightly lower than the one associated with the reference case, especially if only a fraction η of the energy is radiated. Therefore, the minimal super-photospheric flux prediction of $\sim 10^{-11}$ GeV cm^{-2} s^{-1} s^{-1} holds.

The neutrino flux scales with the fraction η going into the non-thermal spectra (and magnetic field), which means that, for $\eta = 0.1$, it would be about one order of magnitude lower than our nominal case. On the other hand, the result is rather insensitive to the fraction of reconverted non-thermal energy $1 - \xi$. This means that for the combined case—a fraction η into non-thermal spectra and a fraction $1 - \xi$ of non-thermal energy reconverted—we expect that the result is dominated by the effect of η.

Finally, the three modifications to the collision dynamics that we have explored do not affect our conclusion about the distribution of particle emission with collision radii: neutrinos still come from low radii, UHECR protons come from intermediate radii, and gamma-rays come from large radii.

References

Aartsen, M., Abbasi, R., Abdou, Y., et al. 2013a, Sci, 342, 1242856
Aartsen, M., Abbasi, R., Abdou, Y., et al. 2013b, PhRvL, 111, 021103
Aartsen, M., Abbasi, R., Ackermann, M., et al. 2014a, PhRvD, 89, 062007
Aartsen, M., Ackermann, M., Adams, J., et al. 2014b, PhRvL, 113, 101101
Aartsen, M., Ackermann, M., Adams, J., et al. 2015a, ApJL, 805, L5
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2015b, PhRvL, 115, 081102
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2016, ApJ, 824, 115
Aartsen, M. G., Ackermann, M., Adams, J., et al. 2015c, PhRvD, 91, 122004
Abbasi, R., Abdou, Y., Abu-Zayyad, T., et al. 2010, ApJ, 710, 346
Abbasi, R., Abdou, Y., Abu-Zayyad, T., et al. 2012, Natur, 484, 351
Abdo, A. A., Ackermann, M., Arimoto, M., et al. 2009, Sci, 323, 1688
Ackermann, M., Ajello, M., Baldini, L., et al. 2012, ApJ, 754, 121
Ahlers, M., González-García, M., & Halzen, F. 2011, APh, 35, 87
Ajello, M., Ajello, M., Asano, K., et al. 2013, ApJS, 209, 11
Amati, L., Frontera, F., Tavani, M., et al. 2002, A&A, 390, 81
Aoi, J., Murase, K., Takahashi, K., Ioka, K., & Nagataki, S. 2010, ApJ, 722, 440
Asano, K. 2005, ApJ, 623, 967
Asano, K., & Terasawa, T. 2015, MNRAS, 454, 2242
Baerwald, P., Bustamante, M., & Winter, W. 2013, ApJ, 768, 186
Baerwald, P., Bustamante, M., & Winter, W. 2015, APh, 62, 66
Baerwald, P., Hümmer, S., & Winter, W. 2012, APh, 35, 508
Bartos, I., Beloborodov, A., Hurley, K., & Márka, S. 2013, PhRvL, 110, 241101
Becker, J. K., Stamatikos, M., Halzen, F., & Rhode, W. 2006, APh, 25, 118
Beloborodov, A. M. 2000, ApJL, 539, L25
Beloborodov, A. M. 2010, MNRAS, 407, 1033
Beloborodov, A. M. 2013, ApJ, 764, 157
Beniamini, P., & Granot, J. 2016, MNRAS, 459, 3635
Beniamini, P., Nava, L., Barniol Duran, R., & Piran, T. 2015, MNRAS, 454, 1073
Bhat, P. 2013, arXiv1307.7618B
Bošnjak, Z., Daigne, F., & Dubus, G. 2008, A&A, 498, 677
Bošnjak, Z., & Kumar, P. 2012, MNRAS, 421, 39
Bustamante, M., Baerwald, P., Murase, K., & Winter, W. 2015, NatCo, 6, 6783
Bykov, A. M., & Mészáros, P. 1996, ApJL, 461, L37
Castignani, G., Guetta, D., Pian, E., et al. 2014, A&A, 565, A60
Daigne, F., & Mochkovitch, R. 1998, MNRAS, 296, 275
Daigne, F., & Mochkovitch, R. 2003, MNRAS, 342, 587
Dermer, C. D., & Atoyan, A. 2003, PhRvL, 91, 071102
Eichler, D., & Waxman, E. 2005, ApJ, 627, 861
Gao, H., & Zhang, B. 2015, ApJ, 801, 103
Giannios, D. 2008, A&A, 480, 305
Globus, N., Allard, D., Mochkovitch, R., & Parizot, E. 2015, MNRAS, 451, 5270
Golkhou, V. Z., Butler, N. R., & Littlejohns, O. M. 2015, ApJ, 811, 93
Gotz, D., Laurent, P., Lebrun, F., Daigne, F., & Bošnjak, Z. 2009, ApJL, 695, L208
Guetta, D., Hooper, D., Alvarez-Muniz, J., Halzen, F., & Reuveni, E. 2004, APh, 20, 429
Guetta, D., Spada, M., & Waxman, E. 2001a, ApJ, 557, 399
Guetta, D., Spada, M., & Waxman, E. 2001b, ApJ, 559, 101
Guiriec, S., Connaughton, V., Briggs, M. S., et al. 2011, ApJL, 727, L33
Gupta, N., & Zhang, B. 2007, APh, 27, 386
Hascoet, R., Daigne, F., & Mochkovitch, R. 2013, A&A, 551, A124
He, H.-N., Liu, R.-Y., Wang, X.-Y., et al. 2012, ApJ, 752, 29
Hümmer, S., Baerwald, P., & Winter, W. 2012, PhRvL, 108, 231101
Ioka, K., Murase, K., Toma, K., Nagataki, S., & Nakamura, T. 2007, ApJL, 670, L77
Kobayashi, S., Piran, T., & Sari, R. 1997, ApJ, 490, 92
Kobayashi, S., & Sari, R. 2001, ApJ, 551, 934
Kumar, P. 1999, ApJL, 523, L113
Kumar, P., & Zhang, B. 2014, PhR, 561, 1
Laha, R., Beacom, J. F., Dasgupta, B., Horiuchi, S., & Murase, K. 2013, PhRvD, 88, 043009
Lazar, A., Nakar, E., & Piran, T. 2009, ApJL, 695, L10
Lazzati, D., Morsony, B. J., Margutti, R., & Begelman, M. C. 2013, ApJ, 765, 103
Li, Z. 2012, PhRvD, 85, 027301
Lopez-Camara, D., Lazzati, D., & Morsony, B. J. 2016, ApJ, 826, 180
Lundman, C., Pe'er, A., & Ryde, F. 2013, MNRAS, 428, 2430
Lyutikov, M., & Blandford, R. D. 2003, arXiv:astro-ph/0312347
Mannheim, K., Protheroe, R. J., & Rachen, J. P. 2001, PhRvD, 63, 023003
McKinney, J. C., & Uzdensky, D. A. 2012, MNRAS, 419, 573
Mészáros, P. 2006, RPPh, 69, 2259
Mészáros, P., Asano, K., Murase, K., et al. 2015, arXiv:1506.02707
Mészáros, P., & Rees, M. J. 1997, ApJL, 482, L29
Milgrom, M., & Usov, V. 1995, ApJL, 449, L37
Murase, K. 2008, PhRvD, 78, 101302
Murase, K., Asano, K., Terasawa, T., & Mészáros, P. 2012, ApJ, 746, 164
Murase, K., & Ioka, K. 2013, PhRvL, 111, 121102
Murase, K., Ioka, K., Nagataki, S., & Nakamura, T. 2006, ApJL, 651, L5
Murase, K., Ioka, K., Nagataki, S., & Nakamura, T. 2008, PhRvD, 78, 023005
Murase, K., Kashiyama, K., & Mészáros, P. 2013, PhRvL, 111, 131102
Murase, K., & Nagataki, S. 2006, PhRvD, 73, 063002
Nakar, E., & Piran, T. 2002a, ApJL, 572, L139
Nakar, E., & Piran, T. 2002b, MNRAS, 331, 40
Narayan, R., & Kumar, P. 2009, MNRAS, 394, L117
Paciesas, W. S., Meegan, C. A., Pendleton, G. N., et al. 1999, ApJS, 122, 465
Paczynski, B., & Xu, G. 1994, ApJ, 427, 708
Pe'er, A. 2012, ApJL, 757, L31
Pe'er, A. 2015, AdAst, 2015, 907321
Pe'er, A., Mészáros, P., & Rees, M. J. 2006, ApJ, 642, 995
Piran, T. 1999, PhR, 314, 575
Preece, R. D., Briggs, M. S., Mallozzi, R. S., et al. 1998, ApJL, 506, L23
Ramirez-Ruiz, E., & Fenimore, E. 1999, ApJ, 539, 712
Rees, M., & Mészáros, P. 1992, MNRAS, 258, 41
Rees, M., & Mészáros, P. 2005, ApJ, 628, 847
Rees, M. J., & Mészáros, P. 1994, ApJL, 430, L93
Spada, M., Panaitescu, A., & Mészáros, P. 2000, ApJ, 537, 824
Stamatikos, M. 2006, PhD thesis, SUNY, Buffalo
Tamborra, I., & Ando, S. 2015, JCAP, 1509, 036
Thompson, C. 1994, MNRAS, 270, 480
Vetere, L., Massaro, E., Costa, E., Soffitta, P., & Ventura, G. 2006, A&A, 447, 499
Vietri, M. 1995, ApJ, 453, 883
von Kienlin, A., Meegan, C. A., Paciesas, W. S., et al. 2014, ApJS, 211, 13
Vurm, I., Lyubarsky, Y., & Piran, T. 2013, ApJ, 764, 143
Wang, X.-Y., & Dai, Z.-G. 2009, ApJL, 691, L67
Wang, X.-Y., Razzaque, S., & Mészáros, P. 2008, ApJ, 677, 432
Waxman, E. 1995, PhRvL, 75, 386
Waxman, E., & Bahcall, J. N. 1997, PhRvL, 78, 2292
Waxman, E., & Bahcall, J. N. 1999, PhRvD, 59, 023002
Yonetoku, D., Murakami, T., Gunji, S., et al. 2012, ApJL, 758, L1
Yonetoku, D., Murakami, T., Nakamura, T., et al. 2004, ApJ, 609, 935
Zhang, B., Liang, E., Page, K. L., et al. 2007, ApJ, 655, 989
Zhang, B., & Pe'er, A. 2009, ApJL, 700, L65
Zhang, B., & Yan, H. 2011, ApJ, 726, 90
Zhang, B., & Zhang, B. 2014, ApJ, 782, 92
Zhang, B.-B., Zhang, B., & Castro-Tirado, A. J. 2016, ApJL, 820, L32