The Astrophysical Journal Letters, 895:L33 (11pp), 2020 June 1 https://doi.org/10.3847/2041-8213/ab9223
© 2020. The American Astronomical Society. All rights reserved.

Ready, Set, Launch: Time Interval between a Binary Neutron Star Merger and Short Gamma-Ray Burst Jet Formation

Paz Beniamini[1], Rodolfo Barniol Duran[2], Maria Petropoulou[3], and Dimitrios Giannios[4]
[1] Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA; paz.beniamini@gmail.com
[2] Department of Physics and Astronomy, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819, USA
[3] Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08544, USA
[4] Department of Physics and Astronomy, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47907, USA
Received 2020 March 20; revised 2020 May 1; accepted 2020 May 1; published 2020 May 29

Abstract

The joint detection of GW170817/GRB 170817 confirmed the long-standing theory that binary neutron star mergers produce short gamma-ray burst (sGRB) jets that can successfully break out of the surrounding ejecta. At the same time, the association with a kilonova provided unprecedented information regarding the physical properties (such as masses and velocities) of the different ejecta constituents. Combining this knowledge with the observed luminosities and durations of cosmological sGRBs detected by the Burst Alert Telescope onboard the Neil Gehrels Swift Observatory, we revisit the breakout conditions of sGRB jets. Assuming self-collimation of sGRB jets does not play a critical role, we find that the time interval between the binary merger and the launch of a typical sGRB jet is $\lesssim$0.1 s. We also show that for a fraction of at least $\sim$30% of sGRBs, the usually adopted assumption of static ejecta is inconsistent with observations, even if the polar ejecta mass is an order of magnitude smaller than that in GRB 170817. Our results disfavor magnetar central engines for powering cosmological sGRBs, limit the amount of energy deposited in the cocoon prior to breakout, and suggest that the observed delay of $\sim$1.7 s in GW170817/GRB 170817 between the gravitational wave and gamma-ray signals is likely dominated by the propagation time of the jet to the gamma-ray production site.

Unified Astronomy Thesaurus concepts: Gamma-ray bursts (629); Relativistic jets (1390); Neutron stars (1108); Gravitational wave sources (677); Astrophysical black holes (98)

1. Introduction

Multi-messenger astronomy has experienced a profound step forward with the observations of the binary neutron star (BNS) merger event, GW170817, in both gravitational waves (GWs) and electromagnetic waves (Abbott et al. 2017). The detection of a short gamma-ray burst (sGRB; Nakar 2007; Berger 2014), GRB 170817, from the BNS merger has renewed the community's interest in these enigmatic explosions (see, e.g., Nakar 2019 for a recent review on sGRBs from BNS mergers). GRB 170817 has forced us to revisit several important properties of GRB jets, such as their angular structure (e.g., Granot et al. 2017; Lamb & Kobayashi 2017; Kathirgamaraju et al. 2018; Beniamini et al. 2020b), as well as possible implications for some as-of-yet mysterious properties of "standard" GRB afterglow observations such as X-ray plateaus (Oganesyan et al. 2020; Beniamini et al. 2020a). More importantly, it has highlighted our need to understand how jets propagate through external media.

As jets propagate out of the central engine of the sGRB, they interact with ejecta made of material launched dynamically during the compact binary merger as well as ejecta driven by the neutrinos released from the neutron star or the accretion disk formed post-merger. The sGRB jet propagation and ejecta interaction (possibly also determining their angular structure) has been studied numerically in numerous works (Aloy et al. 2005; Nagakura et al. 2014; Just et al. 2016; Lazzati et al. 2017; Xie et al. 2018; Geng et al. 2019; Gill et al. 2019a, Kathirgamaraju et al. 2019; Salafia et al. 2020). Such studies are inherently complex, as the relativistic nature of the outflow naturally leads to a large range of temporal and spatial scales. Analytically, the situation may be significantly simplified by considering limiting cases for the dynamics of the ejecta, being either static (Begelman & Cioffi 1989; Marti et al. 1994; Matzner 2003; Bromberg et al. 2011) or homologously expanding (Duffell et al. 2018).

Comparison of model predictions with observed data can help determine the physical properties of breakout, such as the time it takes the jet to break through the ejecta and the time interval between the BNS merger and the launch of the GRB jet. Similar techniques have been employed successfully in the past, mainly for long GRBs breaking out of their surrounding stellar envelopes, for which the static ejecta limit naturally applies (Bromberg et al. 2012; Petropoulou et al. 2017; Sobacchi et al. 2017). Previous studies comparing sGRB data to theory have focused mainly on the static limit by employing either a limited data set of sGRBs with both measured luminosities and durations (Murguia-Berthier et al. 2014, 2017) or a significantly more expanded data set, but with durations only (Moharana & Piran 2017).

Pinning down the timescales involved in the formation and breakout of the jet is at the intersection of several key fields of current study, such as jet formation and propagation, the nature of the central engine (and possibly constraints on the neutron star equation of state, e.g., Lazzati & Perna 2019), and the properties of the radioactive ejecta that may be the dominant source of *r*-process production in the universe (see Hotokezaka et al. 2018 for a recent review).

We show here that the static versus homologous expansion limits for the ejecta propagation can be smoothly combined to form a description that holds also for intermediate situations in terms of the time delay between ejecta and jet launching and intermediate velocities of the ejecta (see also Hamidani et al. 2020; Lyutikov 2020). We then use the current sample of

sGRBs with redshift determination (for which the luminosity and duration can be well determined) to place statistical constraints on the time interval between the moment of the BNS merger and the launching of the GRB jet and on the time it takes the jet to break out of the ejecta.

The paper is organized as follows. In Section 2 we introduce the sample of sGRBs considered in this work. In Section 3 we introduce the two limiting cases (Section 3.1, 3.2) for calculating the properties of jet breakout that have previously been considered in the literature. We then present a treatment that smoothly connects the two regimes (Section 3.3) and show how sGRBs with known durations and luminosities can be used to infer physical properties of the ejecta with respect to the jets. We discuss a variety of implications of these results in Section 4, and finally conclude in Section 5.

Throughout the paper, we adopt a cosmology with $\Omega_{\rm M} = 0.31$, $\Omega_\Lambda = 0.69$, and $H_0 = 69.6$ km s^{-1} Mpc^{-1}.

2. Sample

We use publicly available data from the GRB archive[5] of the Neil Gehrels Swift Observatory (Gehrels et al. 2004). We select sGRBs (i.e., bursts with observed $T_{90} < 2$ s) detected by the Swift Burst Alert Telescope (BAT) from 2005 to 2019 with redshift information (either spectroscopic or photometric). Our sample consists of 27 bursts ($\sim 1/4$ of the Swift–BAT sGRB sample). To estimate the isotropic γ-ray luminosity we use the BAT fluence, Φ, in the 15–150 keV energy range

$$L_{\gamma,\rm iso} = \frac{4\pi d_L^2(z)\Phi}{T_{90}} \frac{\int_{1\ \rm keV}^{10\ \rm MeV} {\rm d}E\ EN(E)}{\int_{(1+z)15\ \rm keV}^{(1+z)150\ \rm keV} {\rm d}E\ EN(E)}, \quad (1)$$

where $d_L(z)$ is the luminosity distance of a burst at redshift z and $N(E)$ is the differential photon spectrum considered in the 1 keV–10 MeV energy range, and described by the so-called Band function (Band et al. 1993) with $\alpha = -0.5$, $\beta = -2.25$, and rest-frame peak energy $E_{\rm p} = 800$ keV (Nava et al. 2011).

We also compare the results we derive from our Swift sGRB sample to the first GRB to be detected in GWs, namely GRB 170817. Since this burst was preceded by a GW trigger, it enabled the detection of a very weak prompt GRB signal with no afterglow signal until days after the event; if there was no GW trigger (i.e., under regular circumstances) there could not have been a redshift determination. Furthermore, for the purposes of this study, we are interested in the luminosities of GRBs along their jet cores. Since GRB 170817 was detected off-axis, its core luminosity is very poorly constrained (Troja et al. 2019). For these reasons, we do not include GRB 170817 in our analysis of deriving upper limits on the time interval between the moment of the BNS merger and the launching of the GRB jet, but return to discuss some specific implications for GRB 170817 in Section 4. We use the duration data from Goldstein et al. (2017) and the constraints on the on-axis luminosity from Troja et al. (2019), accounting for a typical on-axis efficiency, $\eta_\gamma \approx 0.15$ seen in other cosmological sGRBs (see the definition in Section 3.1), when discussing this specific burst.

[5] https://swift.gsfc.nasa.gov/archive/grb_table/

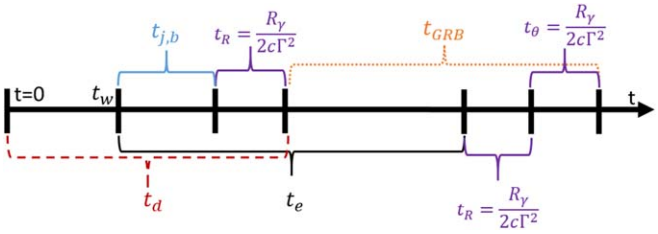

Figure 1. Schematic illustration (not to scale) of the relevant timescales setting the observed duration of the burst, $t_{\rm GRB}$, and the delay time, $t_{\rm d}$, between the burst and a BNS merger occurring at $t = 0$. Propagation to the γ-ray-emitting radius R_γ delays the arrival times of the first and last GRB photons in the same way, thus not contributing to the net GRB duration.

3. Jet Breakout Times

The breakout of the jet through the BNS merger ejecta involves three dynamical timescales, namely the time interval between the BNS merger and the launch of the jet, also referred to here as the "waiting time" ($t_{\rm w}$), the duration of the jet engine operation ($t_{\rm e}$), and the time it takes a GRB jet to break out from the BNS merger ejecta ($t_{\rm j,b}$). Since the ejecta is launched dynamically during the BNS merger, it is launched within several milliseconds from the moment of the merger. As this timescale is much shorter than any of the other timescales of interest considered in this situation, the launching of the ejecta can be considered as concurrent with the BNS merger. The breakout time $t_{\rm j,b}$ has been calculated in the following limits.

1. *Static ejecta.* In this limit, which applies when $t_{\rm j,b}$, $t_{\rm e} \lesssim t_{\rm w}$, the merger ejecta can be considered to be roughly static throughout the breakout (see e.g., Begelman & Cioffi 1989; Marti et al. 1994).
2. *Homologous ejecta expansion*. In this limit, which is relevant when $t_{\rm j,b}$, $t_{\rm e} \gtrsim t_{\rm w}$, the evolution becomes self-similar, namely the jet breakout time is proportional to the engine timescale up to some dimensionless number that is a function of the jet's total energy. In this situation, jets typically break out more easily from the ejecta (Duffell et al. 2018, henceforth denoted as D18).

In addition to the timescales mentioned above ($t_{\rm w}$, $t_{\rm e}$, $t_{\rm j,b}$), there are other important timescales to be considered, which are related to the accompanying observable γ-ray signal. These are the observer's frame[6] duration of the GRB ($t_{\rm GRB}$), the delay time between the GW and γ-ray signals ($t_{\rm d}$), the propagation time of the relativistic ejecta (moving at Lorentz factor Γ) to the γ-ray-emitting radius R_γ, given by $t_{\rm R} \approx R_\gamma/2c\Gamma^2$, and the angular timescale associated with the γ-ray-emitting shell ($t_\theta \approx R_\gamma/2c\Gamma^2$). The latter is the time difference between the arrival of photons emitted on-axis to the observer and those emitted at an angle of $1/\Gamma$ (which due to relativistic beaming is approximately the highest latitude that is visible to the observer). A schematic illustration of the different timescales of the problem is shown in Figure 1.

The delay time between the GW signal and γ-ray signals is the sum of the following three timescales: the time between the BNS merger and jet launch, the time it takes the jet to break out, and the time it takes the jet to reach the γ-ray-emitting

[6] For simplicity, we omit the dependence on cosmological redshift in the expressions throughout this paper. The latter can be trivially included by multiplying all observed timescales by a factor of $1 + z$.

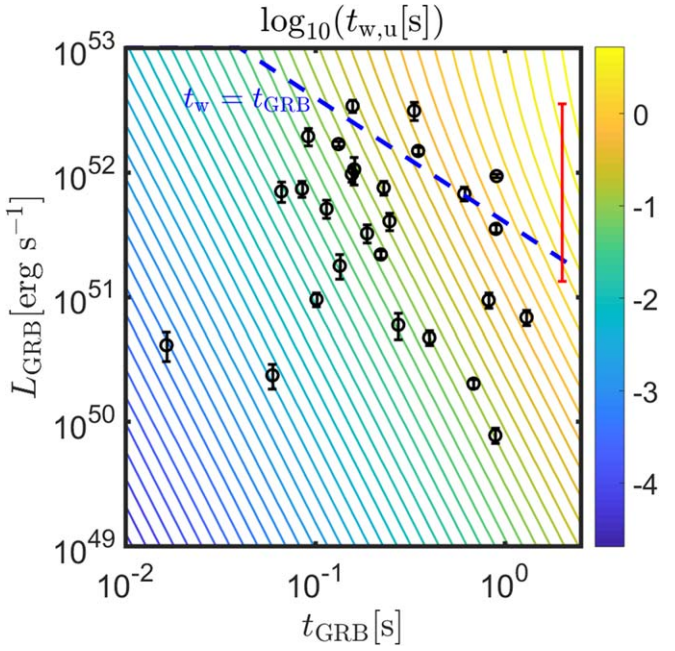

Figure 2. Required (upper limits on the) waiting times (i.e., time intervals between the BNS merger and the launch of the jet—colored solid lines) needed to account for the observed sGRB durations and luminosities within the static ejecta scenario for $t_{\rm GRB}\approx t_{\rm j,b}$. A dashed blue line marks $t_{\rm w}=t_{\rm j,b}$. Below this line, i.e., for $t_{\rm w}<t_{\rm j,b}$, the assumption of a static medium begins to break down. We overplot the data of observed Swift short GRBs with known redshift (corrected for the central engine frame) as black circles and the values for the core of GRB 170817 in red (no circle).

radius,

$$t_{\rm d}=t_{\rm w}+t_{\rm j,b}+t_{\rm R}. \tag{2}$$

The observed duration of the GRB is given by the difference between the engine and jet breakout times (yielding the amount of time during which a successful jet is passing through R_γ) plus the spreading due to the angular timescale from R_γ:

$$t_{\rm GRB}=t_{\rm e}-t_{\rm j,b}+t_\theta. \tag{3}$$

In the following sections, we examine the two limiting regimes (i.e., static ejecta and homologous expanding ejecta). For both regimes, we use the observed distribution of GRB durations and luminosities to set limits on $t_{\rm w}$. This allows us to determine the validity of the approximations corresponding to both regimes and to place overall limits on $t_{\rm w}$, which hold also for any intermediate regime.

3.1. Jet Breakout through Static Ejecta

We begin by considering the static ejecta limit. In this case, the breakout time of a successful jet is given by the time it takes the jet to overpass the merger ejecta

$$t_{\rm j,b}=t_{\rm w}\frac{\beta_{\rm ej}}{\beta_{\rm h}-\beta_{\rm ej}}, \tag{4}$$

where $\beta_{\rm ej}$ is the velocity of the ejecta and $\beta_{\rm h}$ is the velocity of the jet's head. As noted above, self-consistency of this regime requires $t_{\rm j,b}\lesssim t_{\rm w}$ or, equivalently, $\beta_{\rm h}\gtrsim 2\beta_{\rm ej}$ (see Equation (4)). The velocity of the jet's head is related to the ratio between the jet's isotropic-equivalent luminosity, $L_{\rm e}$, and the (isotropic-equivalent) mass outflow rate of the ejecta, $\dot{M}_{\rm ej}$, as follows (Marti et al. 1994; Matzner 2003; Bromberg et al. 2011; Murguia-Berthier et al. 2017):

$$\beta_{\rm h}=\frac{\beta_{\rm j}+\beta_{\rm ej}\tilde{L}^{-1/2}}{1+\tilde{L}^{-1/2}}, \tag{5}$$

where

$$\tilde{L}\equiv\frac{L_{\rm e}\beta_{\rm ej}}{\dot{M}_{\rm ej}c^2}$$
$$=0.14\frac{f_\Omega}{\eta_\gamma}\left(\frac{L_{\rm GRB}}{10^{52}\ {\rm erg\ s^{-1}}}\right)\left(\frac{\beta_{\rm ej}}{0.25}\right)\left(\frac{10^{-2}M_\odot}{M_{\rm ej}}\right)\left(\frac{t_{\rm w}+t_{\rm j,b}}{1{\rm s}}\right). \tag{6}$$

A full derivation of Equations (5) and (6) is given in Appendix A. In particular we note that even in a mildly relativistic regime, with $\beta_{\rm h}=0.25$, Equation (6) holds to better than a 5% accuracy. In addition, this analytical methodology has been shown by Murguia-Berthier et al. (2017) to closely match the results from numerical simulations (with values of $\beta_{\rm h}$ varying by at most 30% between the two). To obtain the numerical value in the last expression, we have inserted values typical for an sGRB jet. The conversion between the isotropic γ-ray luminosity $L_{\rm GRB}$ and the engine power $L_{\rm e}$ can be obtained using the γ-ray efficiency, $\eta_\gamma\equiv L_{\rm GRB}/L_{\rm e}$, which is $\eta_\gamma\approx 0.15$ (Beniamini et al. 2015). This conversion implicitly assumes that the degree of jet collimation within the BNS merger ejecta is the same as after the jet has broken out. We discuss the validity of this assumption and its implications in Section 4.2. We have also assumed that the jet is interacting with the polar component of the merger ejecta. The mass and velocity of the latter can be inferred from the "blue" component of the kilonova (Kasen et al. 2017). It is considered to be associated with the "squeezed" tidal tails, which can be approximated to be roughly isotropically spread up to a polar angle of $\sim\pi/4$ (see, e.g., Kasen et al. 2017). $\dot{M}_{\rm ej}$ can therefore be approximated by $\dot{M}_{\rm ej}=M_{\rm ej}f_\Omega^{-1}/(t_{\rm w}+t_{\rm j,b})$, where $f_\Omega=\int_0^{\pi/4}{\rm d}\theta\sin\theta\approx 0.3$ is the solid angle covered by the blue component (assuming a two-sided jet) and $t_{\rm w}+t_{\rm j,b}$ is the time between the BNS merger and the jet breakout; it is used as a proxy of the ejecta expansion time before the jet breakout.

Equations (4), (5), and (6) allow us to calculate $t_{\rm w}$ as a function of $L_{\rm GRB}$ and $t_{\rm j,b}$. The time interval between the BNS merger and the launch of the jet, $t_{\rm w}$, is found to increase with increasing values of either $L_{\rm GRB}$ or $t_{\rm j,b}$ as we show below. $L_{\rm GRB}$ can be directly constrained from observations for GRBs with redshift determination, while $t_{\rm j,b}$ can be estimated in the following way. We first note that $t_{\rm e}\gtrsim t_{\rm j,b}$ is required in order to avoid most of the jet energy to be deposited in the cocoon instead of the GRB jet (Ramirez-Ruiz et al. 2002). In this case, the GRB duration (in the engine's rest frame) is set by Equation (3) (see also Figure 1). This relation can be better understood in the following limits.

1. $t_\theta\gg t_{\rm j,b}$. In particular, since $t_{\rm GRB}\gtrsim t_\theta$ then $t_{\rm GRB}\gg t_{\rm j,b}$ regardless of $t_{\rm e}$.
2. $t_\theta\ll t_{\rm j,b}$. In this limit, $t_{\rm GRB}\approx t_{\rm e}-t_{\rm j,b}$. For any distribution of $t_{\rm e}$ that has a non-negligible dispersion (i.e., not characterized by a dispersion in $t_{\rm e}$ much smaller than its

average, $\sigma_{t_e} \ll \bar{t}_e$), GRBs with a duration $t_{GRB} \ll t_{j,b}$ would require $t_e \approx t_{j,b}$, which would be fine-tuned and rare (Bromberg et al. 2013). Specifically, there should be approximately one GRB with $t_{GRB} \approx 0.1 t_{j,b}$ ($0.01 t_{j,b}$) for every 10 (100) GRBs with $t_{GRB} \approx t_{j,b}$.

Therefore, independently of the unknown value of t_θ, for most GRBs the observed t_{GRB} corresponds to an upper limit on $t_{j,b}$. The derived upper limits are most conservative if one assumes $t_\theta \to 0$. This is because t_θ is an extra component in the GRB duration that is completely independent of the breakout. Therefore, a non-zero t_θ only increases the difference between the GRB duration and the jet breakout time; see Equation (3). By assuming $t_{GRB} \approx t_{j,b}$ one typically overestimates the true value of $t_{j,b}$ and, in turn, of t_w, since the latter increases with $t_{j,b}$. For the purpose of placing upper limits on t_w (denoted below as $t_{w,u}$) assuming $t_{GRB} \approx t_{j,b}$ is therefore conservative. As a result, GRBs with short durations and low GRB luminosities place the strongest limits on t_w (see Murguia-Berthier et al. 2017 for a similar approach). We demonstrate this point quantitatively in Appendix B.

Using the assumption that $t_{GRB} \approx t_{j,b}$, as described above, we calculate the value of $t_{w,u}$ for a static medium as a function of t_{GRB} and L_{GRB}. These values are plotted in Figure 2. The distribution of $t_{w,u}$ that is needed to explain the population of the observed 27 GRBs in our sample has a median of $t_{w,u} \approx 0.09$ s and a standard deviation of $\sigma_{\log_{10}(t_{w,u})} = 0.7$. As shown in Figure 2, a large fraction of bursts (those below the diagonal dashed line) have no self-consistent solutions with $t_{GRB} \approx t_{j,b}$ and $t_w \gtrsim t_{j,b}$ under the static medium scenario. Even if we account for the uncertainty in our model parameters and allow M_{ej} to be reduced by a full order of magnitude from our canonically assumed value, we still cannot find consistent solutions for eight out of the 27 GRBs (~30%). Since this inconsistency cannot be easily resolved by changing the ejecta properties within reasonable bounds, it suggests that, at least in some cases, the homologous expansion limit may be a more realistic assumption than the static medium limit. We shall explore the implications of this approach in the next section.

3.2. Jet Breakout through Homologously Expanding Ejecta

This limit has recently been studied analytically and numerically by D18. Assuming that degree of jet collimation does not change during the jet breakout process, these authors have found that jets are successful when $E_j \gtrsim 0.1 E_{ej}$ where $E_{ej} \approx 0.5 M_{ej} \beta_{ej}^2 c^2$ is the kinetic energy of the ejecta and E_j denotes the isotropic-equivalent energy of the jet.[7] D18 have identified two breakout regimes. For energies in the range $0.1 E_{ej} \lesssim E_j \lesssim 3 E_{ej}$, jets barely break out and a significant amount of energy is deposited in a cocoon. This regime is dubbed "late breakout". For higher energies, $E_j \gtrsim 3 E_{ej}$, jets break out easily, and this regime is dubbed "early breakout". Relating the latter condition to observational properties, we find

$$L_{GRB} > 1.7 \times 10^{51} \eta_\gamma \left(\frac{\beta_{ej}}{0.25}\right)^2 \left(\frac{M_{ej}}{10^{-2} M_\odot}\right) \left(\frac{1\,s}{t_e}\right) \text{erg s}^{-1}, \quad (7)$$

[7] D18 used the same notation (i.e., E_j) to denote the beaming-corrected energy, i.e., $E_{j,D18} = E_j \theta_0^2/2$, where $E_{j,D18}$ is the value denoted as E_j in D18 and θ_0 is the jet opening angle, hence the difference in the appearance of the equation.

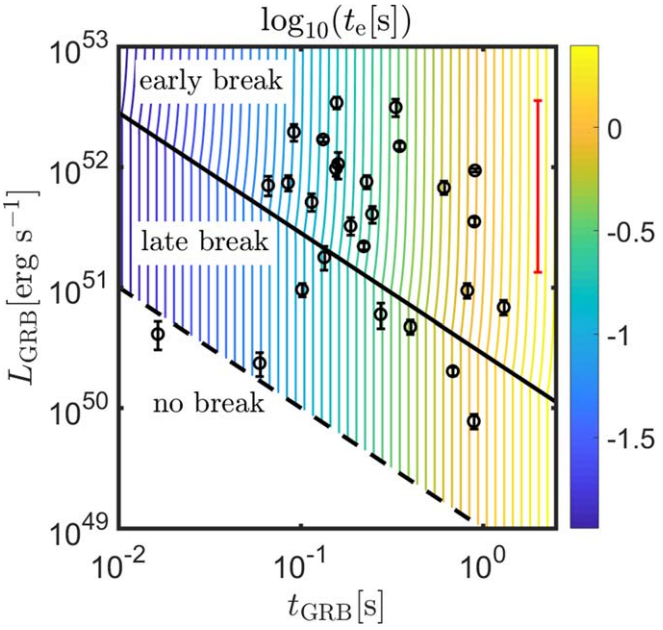

Figure 3. Required engine times (colored solid lines) needed to operate the observed sGRB durations and luminosities for the limit of homologously expanding ejecta. Above the solid line, jets break easily out of the ejecta (i.e., "early breakout"). Between the dashed and solid lines, jets barely break out (i.e., "late breakout"), while below the dashed line, jets no longer break out from the ejecta. Both lines are calculated from Equation (7) using the appropriate expression for $t_{j,b}$ in each regime (see Section 3.2 for details). Symbols have the same meaning as in Figure 2.

where we have taken $E_j \approx t_e L_e \approx t_e L_{GRB}/\eta_\gamma$. The jet breakout time is given by (D18)

$$t_{j,b} = 0.3\, t_e \frac{E_{ej}}{E_j} = 0.15\, \eta_\gamma \frac{M_{ej}(\beta_{ej}c)^2}{L_{GRB}} \quad \text{"early"}$$
$$t_{j,b} = \frac{9 t_e}{\sqrt{\frac{10 E_j}{E_{ej}} - 1}} = \frac{9 t_e}{\sqrt{\frac{20 L_{GRB} t_e}{\eta_\gamma M_{ej}(\beta_{ej}c)^2} - 1}} \quad \text{"late"}. \quad (8)$$

In particular, notice that in the early breakout regime, the jet breakout time becomes independent of the engine duration, and is a function of the jet luminosity only (see also D18 and Lyutikov 2020). The early breakout relation implies that $t_{j,b} < 0.1 t_e$, since $E_j > 3 E_{ej}$ in this case. Thus, the jet breakout time is sub-dominant in determining the GRB duration, i.e., $t_{GRB} \approx t_e - t_{j,b} \approx t_e$ (where we have neglected the potential contribution of t_θ, see Section 3.1 for details).

We can test the validity of the condition given by Equation (7) by directly comparing with sGRB duration and luminosity data. The comparison of both the early breakout and late breakout conditions to the data is shown in Figure 3. The majority of sGRBs (20/27), satisfy the condition given by Equation (7) and reside in the early breakout regime. This is consistent with our previous finding that the majority of sGRBs are successful in breaking out of the BNS merger ejecta (Beniamini et al. 2019). A minority of bursts (5/27) nominally reside in the parameter space for late breakouts. However, these may still be consistent with early breakouts given reasonable changes in the properties of the ejecta (e.g., an ejecta mass lower by a factor of ~4 or with a velocity lower by a factor of ~2). Two bursts (GRB 150101B and GRB 050509B) are close to the limit of jet failure. These are much less likely to have

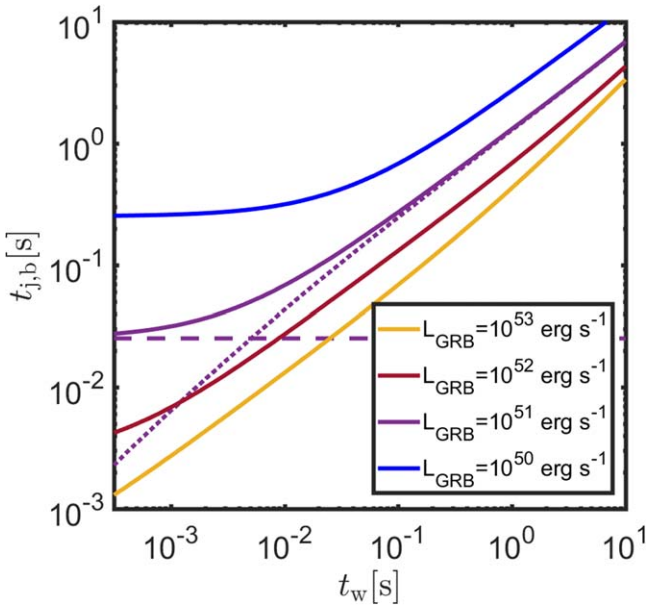

Figure 4. Jet breakout time ($t_{\rm j,b}$) as a function of the time interval between the launch of the BNS merger ejecta and the jet ($t_{\rm w}$) for a generic medium. Results are plotted for $M_{\rm ej} = 0.01 M_\odot$, $\beta_{\rm ej} = 0.25$, $f_\Omega = 0.3$, $\eta_\gamma = 0.15$. Colored curves show the results for different values of the GRB luminosity (see the inset legend). For one case ($L_{\rm GRB} = 10^{51}$ erg s^{-1}), we also show $t_{\rm j,b}(t_{\rm w})$ in the limits of static ejecta (diagonal dotted line) and homologously expanding ejecta (horizontal dashed line). The two curves intersect at $t_{\rm w,c} \approx t_{\rm j,b}/5$.

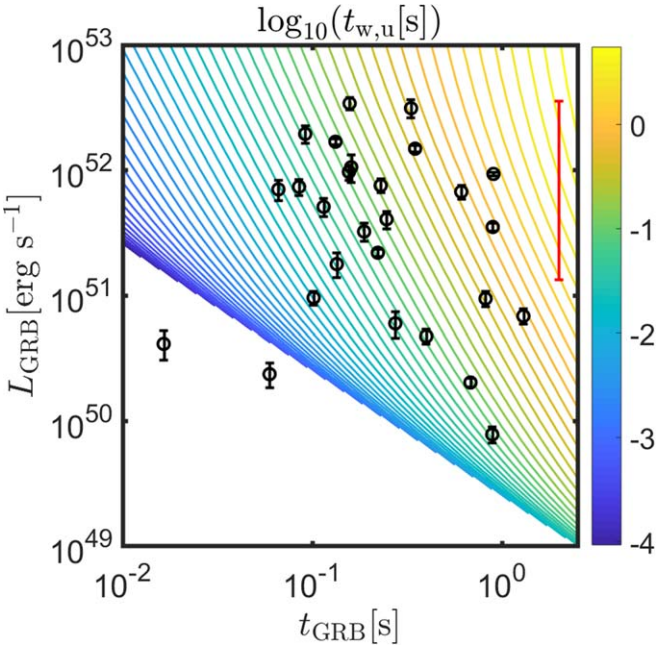

Figure 5. Same as Figure 2 but for a generic medium and for $t_{\rm GRB} \approx t_{\rm j,b}$. Using a generic medium introduces a lower limit on the jet breakout time (corresponding to the homologously expanding medium regime; see Figure 4) and restricts the allowed parameter space compared to Figure 2.

undergone early breakouts, even taking into account variations in the ejecta parameters. We discuss those GRBs in more detail in Section 4.1.

By construction, in the limit of homologously expanding ejecta, the waiting time $t_{\rm w}$ must be sufficiently short that it can be neglected. Thus, $t_{\rm w}$ cannot be directly constrained. Nonetheless, we know that the waiting times must be shorter than the engine times which, for early breakouts, are comparable to the GRB durations. This condition translates to $t_{\rm w} < 0.2$ s, where 0.2 s is the median of the duration distribution of sGRBs in our sample. This upper limit on the waiting time is consistent with our results in Section 3.1, but slightly less constraining. In the next section, we derive limits on the waiting time that are applicable to a generic medium.

3.3. Jet Breakout through a Generic Medium

Combining the results for the jet breakout time obtained in the two regimes of static ejecta and homologously expanding ejecta, we can derive limits on the waiting time for a generic medium. For this, let us recall first that the static approximation is formally valid for $t_{\rm w} \gtrsim t_{\rm j,b}$, while the homologous expansion limit is valid for $t_{\rm w} \lesssim t_{\rm j,b}$. Thus, for a generic medium, the jet breakout time must vary continuously between the solutions obtained in these two limits. Using Equations (4)–(6) and (8), the two limits yield identical jet breakout times at a critical waiting time $t_{\rm w,c} \approx t_{\rm j,b}/5$. Remarkably $t_{\rm w,c}$ is a function of $t_{\rm j,b}$ only, and it is independent of the other physical parameters. The results are shown in Figure 4 where we plot $t_{\rm j,b}(t_{\rm w})$ under both approximations. For a generic medium, we can smoothly connect the two regimes by taking

$$t_{\rm j,b} = t_{\rm j,b,hom} + t_{\rm j,b,stat}, \quad (9)$$

where $t_{\rm j,b,hom}$, $t_{\rm j,b,stat}$ are the jet breakout times in the homologous expansion and static ejecta limits, respectively.

Using the expression for $t_{\rm j,b}$ above we can now obtain upper limits on $t_{\rm w}$ (i.e., $t_{\rm w,u}$) for a generic medium in a similar way to the one outlined in Section 3.1. For a given GRB luminosity, there is a lower limit on the jet breakout time (corresponding to breakout from a homologously expanding medium, see Figure 4). This limit decreases with increasing luminosity. The result, shown in Figure 5, is that for sufficiently short durations and/or low luminosities no self-consistent solutions exist (without changing the ejecta properties and/or the γ-ray efficiency). Within this generic medium scenario, we find again that for the same two out of 27 GRBs in our sample, GRB 050509 and GRB 150101B, no solutions ae available (consistent with our findings in Section 3.2). We return to discuss those bursts in more detail in Section 4.1. Since breakout is more difficult (i.e., $t_{\rm j,b}$ becomes longer) when $t_{\rm w}$ becomes longer, these two bursts likely correspond to shorter waiting times than found for the rest of the sGRB population. Nonetheless, to be more conservative, we ignore these bursts when calculating the upper limits on $t_{\rm w}$ below. For the majority (25/27) of GRBs however, we can self-consistently treat the jet breakout. This leads to upper limits on the waiting times. The median upper limit obtained for those bursts is $t_{\rm w,u} \lesssim 0.1$ s (for the dependence of this result on kilonova ejecta properties, see Section 4.2).

4. Discussion

We have considered the breakout of GRB jets from the BNS merger ejecta. The observed properties of cosmological GRBs, as well as constraints on the merger ejecta from the kilonova counterpart to GW170817, allow us to put limits on the time intervals between the launching of the BNS merger ejecta components and the launching of the relativistic jet. For a generic description of the BNS merger ejecta (which smoothly connects the regimes of static and homologously expanding ejecta), we derive a rough upper limit of $t_{\rm w} \lesssim 0.1$ s.

Furthermore, we argue that for a fraction of sGRBs (at least ∼30%) the assumption of static ejecta, through which the jet punches, is inconsistent with their observed luminosities and durations, even if the polar component of the ejecta mass is a factor of 10 lower than that estimated for the kilonova accompanying GW170817. At the same time, we find that the assumption of homologously expanding ejecta (corresponding to the limit $t_w \to 0$) is consistent with the observed properties (t_{GRB}, L_{GRB}) of our sGRB sample. The derived upper limit on the waiting time has several interesting implications which we discuss below.

4.1. Exceptional GRBs

As mentioned in Sections 3.2 and 3.3 and two bursts (GRB 150101B and GRB 050509B) are close to jet failure, even when considering the limit of negligible waiting time ($t_w \to 0$), for which breakout becomes easiest. These are much less likely to have undergone early breakouts, even taking into account variations in the ejecta parameters.

One possibility is that the prompt GRB in those cases represents a jet that failed to break through the merger ejecta. In such a scenario a γ-ray signal may still result due to the shock breakout from the cocoon created by the jet–merger ejecta interaction. In the case of GRB 150101B, Burns et al. (2018) have found evidence for a short hard spike followed by a soft tail, similar to the GRB counterpart of GW170817, suggesting shock breakout as a possible explanation for the observed γ-rays. Furthermore, GRB 150101B exhibited a bright optical counterpart consistent with a blue kilonova (Troja et al. 2018) with $M_{ej} > 0.02 M_\odot$ and $\beta_{ej} > 0.15$. These values are consistent with a late shock breakout as an explanation for this burst. Nevertheless, it is important to note that GRB 150101B does not appear to satisfy the closure relationship between the energy, duration, and temperature of shock breakout flares (Nakar & Sari 2010). Assuming this relation, we would expect the peak energy of GRB 150101B to be ≈2 MeV, whereas the observed peak energy is of the order of 550 keV for the initial spike and much lower, ∼23 keV, for the soft tail (Burns et al. 2018). In the case of GRB 050509, the breakout estimate for the temperature yields $k_B T \approx 1.5$ MeV. Since νF_ν is seen to be rising within the observed Swift $15 - 150$ keV band (Bloom et al. 2006), only a lower limit on the peak energy can be obtained and the possibility of a shock breakout association satisfying the closure relation cannot be ruled out. A major shortcoming of the shock breakout interpretation for both these bursts, however, is that their γ-ray luminosities, $\sim 3 \times 10^{50}$ erg s^{-1}, are large compared to expectations from shock breakout (Nakar & Sari 2010). Since the shock breakout mechanism releases only a very small fraction of the total energy (see Section 4.5), this interpretation would require much larger engine luminosities to work. This requirement, in turn, would make it much less likely for the jets of those GRBs to have failed to break through the ejecta in the first place (more powerful jets can more easily break out).

The above discussion makes us consider an alternative (and easier to accommodate) scenario for both GRBs. According to this scenario, γ-rays are still produced within a successful relativistic jet, as in regular cosmological sGRBs, but the γ-ray efficiency is much lower than typically used (i.e., $\eta_\gamma \ll 0.1$). One natural way for this to happen is if these GRBs are viewed slightly off-axis from the cores of their jets (see e.g., Bartos et al. 2019; Beniamini et al. 2019; Mandhai et al. 2019; Dichiara et al. 2020).

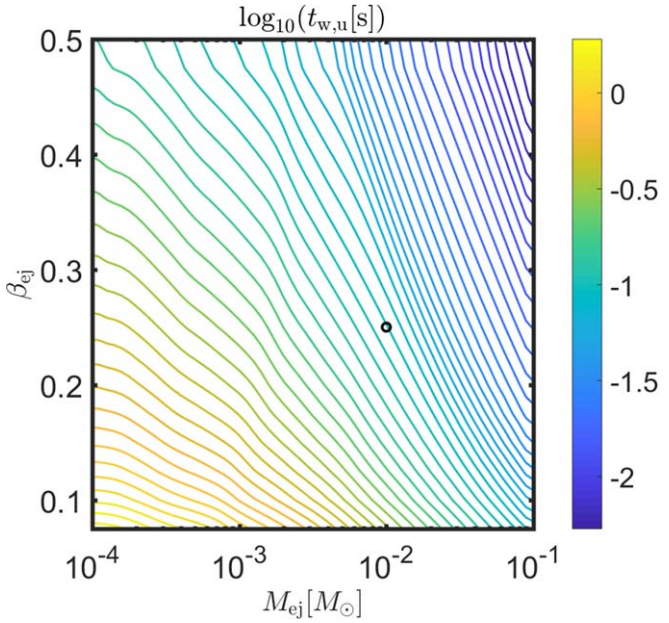

Figure 6. Dependence of the median upper limit on the waiting time on the ejecta mass and velocity corresponding to the blue component of the kilonova. The estimates of those values for GW170817 that are used elsewhere in this paper are marked by a circle.

4.2. Variation of Jet and Kilonova Ejecta Properties

We have argued that, taking the ejecta properties to be similar to those inferred from the kilonova accompanying GW170817, the majority of observed short GRBs would not have been able to break out through a static ejecta. An alternative option is that there is a very wide variation in the BNS ejecta properties of different BNS mergers. We caution the reader that if indeed the ejecta mass varied very widely between different events, this would tend to increase our upper limits on the waiting times discussed above. The required level of ejecta mass variation under the static ejecta interpretation, however, would be very large. For example, lowering the ejecta mass by roughly two orders of magnitude as compared with the inferred values for GRB 170817 is required to enable a successful breakout of all GRBs in our sample under this interpretation. The general dependence of our limit on the waiting time on the ejecta mass and velocity is depicted in Figure 6. Future observations of kilonovae accompanying GW events from nearby BNS mergers would enable us to directly test the validity of this possibility.

Another potential caveat regards the possibility of self-collimation. For long GRB jets, which are propagating through the envelope of the collapsed star, self-collimation is expected to make the effective opening angle of the jet, as it is passing through the stellar envelope, narrower than its final opening angle after breakout (Bromberg et al. 2011). For short GRBs, the smaller amount of ejecta mass and the expanding nature of the ejecta suggest that self-collimation plays a lesser role, especially in the homologous expansion limit (D18; Gill et al. 2019b; Hamidani et al. 2020). If, however, the jets were indeed significantly narrower during their propagation through the BNS merger ejecta, this would result in an effective increase of the isotropic-equivalent engine luminosity and, in turn, an

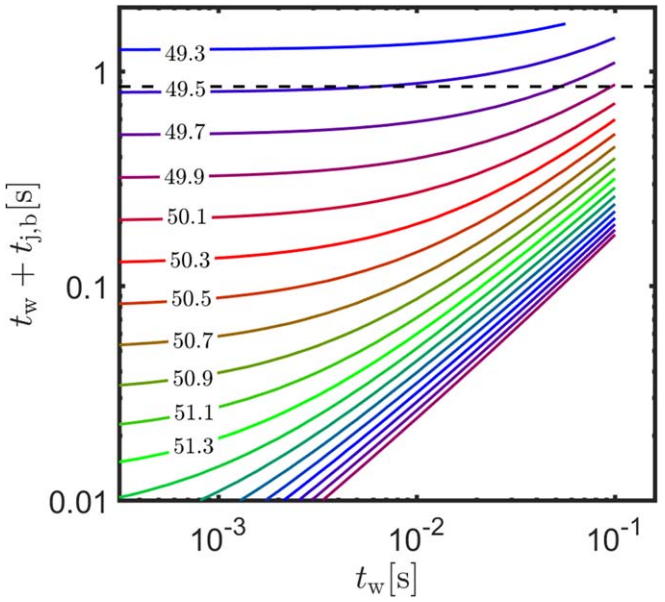

Figure 7. GRB luminosity along the core of the jet and waiting times that correspond to delay times between GW and γ-rays that are consistent with observations in GRB 170817. The curves correspond to combinations of these properties that ensure $t_{\rm w} < 0.1$ s (as found in this paper) and $t_{\rm w} + t_{\rm j,b} < 1.7$ s (as required by the observed delay). From top to bottom contours depict jet core isotropic-equivalent GRB luminosities with equal logarithmic intervals (labels indicate $\log_{10}(L_{\rm GRB})$). A dashed horizontal line marks $t_{\rm w} + t_{\rm j,b} = t_{\rm d}/2$. Afterglow fitting of GRB 170817 strongly favors $L_{\rm GRB} \gtrsim 1.3 \times 10^{51}$ erg s^{-1}, which is the region below the solid green line, implying that $t_{\rm w} + t_{\rm j,b} < t_{\rm d}/2$.

increase in $\tilde{L}$ and an increase in our upper limit on $t_{\rm w}$. To test the importance of self-collimation, we have considered a situation in which the effective area of the jet decreases by a full order of magnitude within the merger ejecta (i.e., increasing $\tilde{L}$ by a similar amount). Our limit on the waiting time changes in that case from $t_{\rm w} < 0.1$ s to $t_{\rm w} < 0.25$ s.

Expanding further on the previous point, one may question how well the analytical treatment adopted in this work represents the true physical situation. Comparison of analytical and numerical studies (Mizuta & Ioka 2013; Harrison et al. 2018; Hamidani et al. 2020) suggests that, while the analytical results match rather well the numerical ones, they tend to slightly overestimate $\tilde{L}$. This will correspond to a slight decrease in our upper limit on $t_{\rm w}$, making our analytical treatment conservative from this perspective.

4.3. Engine Duration Distribution

In either the static or the homologous expansion case, very luminous short GRBs have jet breakout times that are much shorter than the GRB duration. The result is that the GRB duration is almost the same as the engine duration and implies that the duration distribution of luminous bursts directly maps the distribution of engine durations. At the moment, the numbers of such events are rather low, due to the sparsity of sGRBs with redshift determination and the intrinsic rarity of the most luminous bursts. Increasing this sample in the future would enable us to glean critical information regarding the nature of the central engine.

Since the difference between the observed and source frame durations of short GRBs is typically less than a factor of two, the duration distribution of short GRBs can be studied even for GRBs with no redshifts. This has the advantage of increasing the data set significantly, but at the cost of removing information about the intrinsic luminosity. Moharana & Piran (2017), using a large sGRB sample (with and without z measurements) within the framework of Bromberg et al. (2012), found possible evidence for a plateau in the duration distribution ($dN/dt_{\rm GRB}$) at $t_{\rm j,b} \approx 0.4$ s, followed by a power-law-like distribution at longer durations, $dN/dt_{\rm GRB} \propto t^{-1.4}$. Such a plateau is expected if the jet breakout time of the most commonly observed sGRBs is also $\approx$0.4 s. This interpretation is consistent with $t_{\rm w} \lesssim 0.1$ s if the most commonly observed sGRBs have characteristic luminosities of $L_{\rm ch} \sim 3 \times 10^{50}$ erg s^{-1} (see Figure 4). This is consistent with the results of Wanderman & Piran (2015), who showed that the number of sGRBs is dominated by the low end of the observed luminosity function (i.e., $L_{\rm min} \approx 5 \times 10^{49}$ erg s^{-1}). The proximity of $L_{\rm ch}$ to $L_{\rm min}$ and, in particular, the fact that $L_{\rm ch}$ is orders of magnitude smaller than the characteristic break of the luminosity function, $L_* \approx 2 \times 10^{52}$ erg s^{-1}, suggest that it is unlikely that $L_{\rm ch}$ plays any significant role in shaping the sGRB luminosity function (see Beniamini et al. 2019 for details). This is consistent with the conclusions of Beniamini et al., namely that the fraction of failed jets cannot explain the broken-power-law nature of the sGRB luminosity function, contrary to the case of long GRB jets (Petropoulou et al. 2017).

4.4. Extension of the Analysis to GRBs with no Redshift Determination

As pointed out in Section 4.3, considering GRBs with undetermined redshift has the advantage of significantly increasing the sample size, but at the cost of leaving the luminosity highly uncertain. For this reason, we consider now the 14 yr Swift sGRB sample that consists of 119 sGRBs without redshift determination only as a consistency check to the main results presented in Section 3.3. For this purpose, we make the simplifying assumption that all Swift GRBs without redshift determination originate from the same redshift z_0. As test values we consider $z_0 = 0.55$, which is the median redshift of GRBs in our sample, and $z_0 = 0.9$, which is roughly the peak of the sGRB redshift distribution found by Wanderman & Piran (2015). The limits on $t_{\rm w}$ can then be obtained in a similar way to that described in Section 3.3. The results are $t_{\rm w,u} \lesssim 0.09$ s ($t_{\rm w,u} \lesssim 0.12$ s) for $z_0 = 0.55$ ($z_0 = 0.9$). The proximity of these values to the upper limits derived from the sample of bursts with redshift suggests that our results can be reasonably extended to the general sGRB population.

4.5. Shock Breakout Energy

The energy released during the breakout phase is limited by the thermal energy stored in the cocoon, $E_{\rm Th}$. The latter is limited by the (collimation-corrected) energy deposited in the cocoon before the moment of breakout, $E_{\rm Th} \lesssim \frac{\theta_0^2}{2} L_{\rm e} t_{\rm j,b}$. Using the relation $t_{\rm w}(t_{\rm j,b}, L_{\rm GRB})$ derived in Equations (4)–(8), we find that $E_{\rm Th} \lesssim 4 \times 10^{49}$ erg for $t_{\rm w} \lesssim 0.1$ s and $L_{\rm e} = 10^{53}$ erg s$^{-1} \approx L_*/\eta_\gamma$. If the engine power is reduced or the waiting time is shorter, the upper limits on $E_{\rm Th}$ would become more constraining. As a comparison, in the homologous case, the thermal energy of the cocoon is (for both the early and late breakout scenarios) $E_{\rm Th} \lesssim 5 \times 10^{47}(\theta_0/0.1)^2$erg (see Beniamini et al. 2019 for details). Furthermore, since initially the cocoon is highly optically thick (Nakar & Sari 2010), only a small fraction of this energy is expected to

be released as prompt γ-rays during the breakout phase. Overall, we expect the quasi-isotropic shock breakout signal accompanying sGRBs to be typically rather weak.

4.6. The Delay Time of GRB 170817

The observed delay of $t_{\rm d} \sim 1.7$ s between the GW and the γ-ray signal from GRB 170817 (Abbott et al. 2017) can be expressed as the sum of $t_{\rm w}$, $t_{\rm j,b}$, and $t_{\rm R}$ (see Equation (2) and Figure 1). A natural question then arises: which of the three timescales dominates the observed $t_{\rm d}$? Since the first two timescales are a function of the jet luminosity, the answer depends critically on the luminosity of GRB 170817 along its core.

In Figure 7 we plot the allowed parameter space given by the requirements $t_{\rm j,b} + t_{\rm w} < 1.7$ s and $t_{\rm w} < 0.1$ s. We caution the reader that the latter constraint is found from a statistical analysis of the entire sGRB sample. It is of course possible that for any specific event, the waiting time may be longer. Indeed several groups have demonstrated that they can reproduce the observed signatures of GRB 170817 with numerical simulations involving longer ($\sim$1 s) delays (Mooley et al. 2018; Xie et al. 2018). In addition, Gill et al. (2019b) have shown that such longer delays could be favourable for explaining the observed properties of the associated kilonova emission in that event. Nonetheless, we expect the analysis outlined here to be representative of future GW-detected sGRBs. Mooley et al. (2018) have demonstrated that GRB 170817 involved a powerful jet that broke out of the merger ejecta. Given the energy at the core of GRB 170817 inferred from afterglow fitting (Troja et al. 2019), its GRB isotropic-equivalent luminosity (along its core) is estimated to be $L_{\rm GRB} \gtrsim 1.3 \times 10^{51}\,{\rm erg\,s^{-1}}$ (region below the solid green line in Figure 7), clearly in contradiction with $t_{\rm j,b} + t_{\rm w} > t_{\rm d}/2$ (region above the horizontal dashed line in Figure 7). This implies that $t_{\rm j,b} + t_{\rm w} < t_{\rm d}/2$ for GRB 170817, which translates to $t_{\rm d} \approx t_{\rm R} \approx R_\gamma/2c\Gamma^2$. It is worth noting that, due to angular spreading, this situation can naturally lead to a pulse duration which is also of the order of $t_\theta \approx R_\gamma/2c\Gamma^2$. This is realized in prompt emission models for which other timescales involved in the prompt GRB phase, such as the cooling time and engine variation timescale, are shorter than or equal to t_θ. Models of this kind include the internal shocks model (Sari & Piran 1997; Daigne & Mochkovitch 1998) and several magnetic reconnection-based models where dissipation takes place far from the central engine (e.g., Kumar & Narayan 2009; Zhang & Yan 2011; Beniamini & Granot 2016; Beniamini & Giannios 2017; Beniamini et al. 2018). Interestingly, the duration of GRB 170817 was of the same order as the time delay. If future events continue to show a similar trend, this will be a strong indication in favor of $R_\gamma/2c\Gamma^2$ dominating the observed time delay (see Zhang 2019 for a detailed discussion on this point).

As opposed to regular cosmological GRBs, which are observed on-axis, GRB 170817 was observed off-axis (Mooley et al. 2018). As a result, the propagation and angular timescales ($\propto R_\gamma/\Gamma^2$) are likely to be larger than for on-axis GRBs. The extent of this effect depends on the nature of the mechanism producing the prompt emission. For example, in photospheric models, $R_\gamma \propto L_{\rm e}\Gamma^{-3}$ leading to $R_\gamma/\Gamma^2 \propto L_{\rm e}\Gamma^{-5}$. For typical expected profiles of power and Lorentz factor in the jet (e.g., Kathirgamaraju et al. 2018; Beniamini & Nakar 2019), the latter is an increasing function of polar angle, suggesting longer propagation and angular timescales for GRBs seen off-axis.

Regardless of the specific prompt emission model, a measurement of the delay, $t_{\rm d}$, leads to a lower limit on the Lorentz factor of the prompt producing material due to the following argument. Radiation is decoupled from the jet material when the Thomson optical depth $\tau_{\rm T}$ becomes of order unity. This happens at the so-called photospheric radius, which can be determined by setting $\tau_{\rm T} = 1$ as: $R_{\rm ph} \simeq 10^{14} L_{\rm e,47}\Gamma^{-3}$ cm, where $L_{\rm e} = 10^{47}\,L_{\rm e,47}$ erg s^{-1} is the jet's isotropic-equivalent power of the material that dominates the emission toward the observer (which in the case of GRB 170817 was outside of the jet core; e.g., Finstad et al. 2018) and Γ is the Lorentz factor of the same material (see, e.g., Giannios 2012). Using the observed time delay $t_{\rm d}$ and noting that $R_\gamma \gtrsim R_{\rm ph}$, we can derive a lower limit on the bulk Lorentz factor of the material that is independent of the prompt emission model: $\Gamma \gtrsim 4\,L_{\rm e,47}^{1/5}(t_{\rm d}/1.7{\rm s})^{-1/5}$. Note that the exact numeric value is not so sensitive to the jet power of the material dominating the observed emission. Furthermore, this material need not necessarily lie directly along the line of sight. The argument outlined here would hold for material from an intermediate angle $\theta_0 < \theta < \theta_{\rm obs}$, such that it is Doppler boosted toward the observer, i.e., $(\theta_{\rm obs} - \theta)\Gamma < 1$.

4.7. Nature of the Central Engine

In any GRB central engine, the jet breakout time $t_{\rm j,b}(t_{\rm w}, L_{\rm e})$ must be large enough that, by the time the jet breaks out and starts emitting, its bulk Lorentz factor is sufficiently high to avoid the compactness problem (i.e., to avoid a very large optical depth of the emitting material; see the definition of $R_{\rm ph}$ in Section 4.6).

For magnetar central engines, where the flow is initially heavily baryon loaded, this is not easily achieved. Since the power released by the magnetar, $\dot{E}$, typically decreases more slowly in time as compared to the mass outflow rate, $\dot{M}$ (Metzger et al. 2011), there is generally a minimum time, $t_{\rm d,mag}$, before the energy per baryon at the base of the jet ($\eta \propto \dot{E}/\dot{M}$) becomes sufficiently high (i.e., $\eta \gtrsim 100$) to power an ultra-relativistic GRB (Beniamini et al. 2017). For typical parameters of the magnetar model, $t_{\rm d,mag} \gtrsim 3$ s. The main parameter affecting this timescale is the magnetar's magnetic field, B. Only for an extreme value of $B \sim 3 \times 10^{16}$ G does one find $t_{\rm d,mag} \approx 0.2$ s. As shown in Figure 4, this is still too high given the expected values of $t_{\rm j,b}$ with a waiting time of $t_{\rm w} \lesssim 0.1$ s and a (rather common) $L_{\rm GRB} \gtrsim 3 \times 10^{51}\,{\rm erg\,s^{-1}}$. Fall-back accretion onto the magnetar may alter the timescale $t_{\rm d,mag}$, but given the high accretion rates expected for BNS mergers, this effect tends to reduce the initial energy per baryon and therefore increase $t_{\rm d,mag}$ even more (Metzger et al. 2018). As a result we conclude that magnetar central engines are severely challenged as possible engines of short GRBs.

In the context of black hole central engines, a waiting time of $t_{\rm w} \lesssim 0.1$ s suggests a relatively prompt collapse of a neutron star to a black hole. This suggests that the remnant mass from the BNS merger should be massive enough to form at least a hypermassive neutron star (a short-lived neutron star supported by differential rotation); see also Murguia-Berthier et al. (2014, 2017). The collapse time to a black hole depends on the tidal deformability parameter and on the equation of state of the neutron star (Flanagan & Hinderer 2008; Favata 2014). For example, Radice et al. (2018) have shown that for GW170817 this time is indeed expected to be 1–10 ms, consistent with our limit on $t_{\rm w}$.

Physically, the waiting time can be composed of the sum of time it takes to form the central black hole and the time it takes it to accrete a significant amount of mass. Our limit on the waiting time therefore limits also the accretion time, $t_{\rm acc} < t_{\rm w,u}$. Recent simulations of magnetically launched jets from neutron star merger accretion discs (Christie et al. 2019) find that the jet power peaks within $\lesssim$0.05 s. This is consistent with the results found in this paper. Useful intuition on the accretion timescale can be obtained from an α-viscosity thick disk model for the accretion. This limit has been applied to different neutron star merger simulations by Fryer et al. (2015) to constrain the outcome of the merger as a function of, e.g., the individual neutron star masses and the equation of state. Using $t_{\rm acc} = 2\pi r^{3/2}/(\alpha\sqrt{GM_{\rm enc}})$ (where we assume that immediately after the merger there is a thick disk with radius r, enclosing a mass $M_{\rm enc}$) and requiring $t_{\rm acc} < t_{\rm w,u}$ we can constrain the α viscosity parameter of the disk,

$$\alpha \gtrsim 0.01\left(\frac{r}{2\times10^{6}\,{\rm cm}}\right)^{3/2}\left(\frac{0.1{\rm s}}{t_{\rm w,u}}\right)\left(\frac{2.6M_{\odot}}{M_{\rm enc}}\right)^{1/2}. \quad (10)$$

In the last expression we have considered a radius for the disk immediately after merger which is of the order of two neutron star radii (in general this can be considered as a lower limit). We also took the enclosed mass to be approximately the minimum total mass required for producing a hypermassive neutron star (Baumgarte et al. 2000) that would quickly collapse to a black hole (a significantly smaller mass would correspond to a long-lived neutron star in contrast with our limits on the waiting time, as discussed above, while a significantly higher mass would lead to a prompt collapse and would result in little amount of mass in the disk that would be available to power the following GRB). This mass is also consistent with the total mass of known Galactic BNSs (e.g., Beniamini & Piran 2016). Our limits on α from Equation (10) are consistent with the considerations of Fryer et al. (2015).

5. Conclusions

We have revisited the conditions for breakout of an sGRB jet from the BNS merger ejecta. Using published results from analytical and numerical works on this topic, which apply either to the case of static merger ejecta or homologously expanding ejecta (e.g., Begelman & Cioffi 1989; Marti et al. 1994; D18), we derive the conditions of successful jet breakout for a generic medium that smoothly connects these limiting cases. Using the Swift–BAT sample of sGRBs with measured redshift, we derive limits on the waiting time, i.e., the time interval between the BNS merger and the launching of the jet (assuming that the ejecta is launched immediately after the BNS merger).

For all the cases we examined (i.e., static ejecta, homologously expanding ejecta, and generic medium), we set an upper limit of $\sim$0.1 s on the waiting time. Decreasing the ejecta mass (velocity) by a factor of 10 (2) increases the upper limit on $t_{\rm w}$ by a factor of $\sim$4 (2). Our results on the waiting time can be also extended to the complete Swift sample of sGRBs with no redshift determination. Our upper limit on the waiting time is consistent with previous results (e.g., Murguia-Berthier et al. 2014, 2017) obtained with a smaller sample and in the limit of static merger ejecta. We also show that this typically adopted limit for the BNS ejecta is inconsistent with at least $\sim$30% of our sGRB sample.

Although the analytical treatment adopted in this paper is approximated, and does not take into account some of the finer details of jet propagation observed in numerical simulations, such as collimation shocks, it is in good agreement with several numerical studies (Murguia-Berthier et al. 2017; D18; Hamidani et al. 2020). We stress that our overall result is rather intuitive. Given that sGRBs typically last $\sim$0.2 s and that the rate of BNS mergers and successful sGRB jets are similar (Beniamini et al. 2019), it is unlikely that the characteristic breakout and waiting timescales could be much longer than the sGRB durations, as this would require a fine-tuning between the engine activity time and these timescales.

The limit on the interval between the BNS merger and the launch of the jet has profound consequences for the origin of γ-ray emission (i.e., cocoon shock breakout versus jet) and the nature of the sGRB central engine (i.e., magnetar versus black hole). It restricts the amount of thermal energy stored in the cocoon (e.g., $E_{\rm Th} \lesssim 4\times10^{49}$ erg for $L_{\rm e} = 10^{53}$ erg s^{-1}), suggesting that the shock breakout signal accompanying sGRBs is expected to be rather weak. It also suggests that central engines of sGRBs are unlikely to be millisecond magnetars (i.e., with $B \lesssim 3\times10^{16}$ G), since the time interval of $\lesssim$0.1 s is too short to produce a jet with sufficiently high energy per baryon at its base to allow its bulk acceleration to ultra-relativistic speeds. Our results are therefore in favor of a relatively prompt collapse (i.e., within $<$100 ms) of a neutron star to a black hole.

In the context of GRB 170817 our work places strong constraints on the physical origin of the observed γ-ray signals, assuming that the statistical limit on $t_{\rm w}$ found in this work applies also for this specific GRB. We find that the observed delay between the GW and the γ-ray signal is dominated by the time it takes the jet to reach the location at which it will radiate (see also Zhang 2019). The consequence of this interpretation is that the γ-ray duration may naturally (depending on the prompt emission model; see Section 4.6) be of the same order of the observed delay, which is the case for GRB 170817. Future observations would indicate if this is the case for other bursts. This could provide a much needed independent test for comparing between the many prompt emission models that remain viable at this point.

The research of P.B. was funded by the Gordon and Betty Moore Foundation through Grant GBMF5076. R.B.D. and D. G. acknowledge support from the National Science Foundation under Grants 1816694 and 1816136. M.P. acknowledges support from the Lyman Jr. Spitzer Postdoctoral Fellowship and the Fermi Guest Investigator Program Cycle 12, grant 80NSSC18K1745. D.G. acknowledges support from the NASA grant NNX17AG21G and the Fermi Guest Investigator Program Cycle 12, grant 80NSSC19K1506.

Appendix A
Estimating the Jet Head's Velocity

The jet head velocity can be found by the requirement of pressure balance between the shocked jet head material and the shocked ambient medium (see, e.g., Begelman & Cioffi 1989)

$$h_{\rm j}\rho_{\rm j}c^2(\Gamma\beta)^2_{\rm j,h} + P_{\rm j} = h_{\rm ej}\rho_{\rm ej}c^2(\Gamma\beta)^2_{\rm h,ej} + P_{\rm ej} \quad (A1)$$

where h, ρ, P are the dimensionless enthalpy, the density, and the pressure of the fluid materials in their respective rest frames. The quantity $(\Gamma\beta)_{\rm j,h}$ ($(\Gamma\beta)_{\rm h,ej}$) measures the proper velocity of the jet (head) relative to the head (BNS merger ejecta). Assuming both the jet and the ejecta are initially cold, we can neglect the terms $P_{\rm j}$, $P_{\rm ej}$ in Equation (A1). Using the Lorentz transformation we write $(\Gamma\beta)_{\rm x,y} = (\beta_{\rm x} - \beta_{\rm y})\Gamma_{\rm x}\Gamma_{\rm y}$. Plugging back into Equation (A1), we find

$$h_{\rm j}\rho_{\rm j}\Gamma_{\rm j}^2\beta_{\rm j}^2\left(1-\frac{\beta_{\rm h}}{\beta_{\rm j}}\right)^2 = h_{\rm ej}\rho_{\rm ej}\Gamma_{\rm ej}^2(\beta_{\rm h}-\beta_{\rm ej})^2 \quad \text{(A2)}$$

noting that

$$\frac{4\pi r^2 h_{\rm j}\rho_{\rm j}\Gamma_{\rm j}^2 c^3}{4\pi r^2 h_{\rm ej}\rho_{\rm ej}\Gamma_{\rm ej}^2 c^3} = \frac{L_{\rm e}\beta_{\rm ej}}{\dot{M}_{\rm ej}c^2\Gamma_{\rm ej}h_{\rm ej}} = \frac{\tilde{L}}{\Gamma_{\rm ej}h_{\rm ej}} \quad \text{(A3)}$$

where in the last transition we have plugged in the definition of $\tilde{L}$ given in Equation (6). Since for $\beta_{\rm ej} = 0.25$, we have $(\Gamma_{\rm ej}h_{\rm ej})^{-1} \approx 1.05$ (where we have used an approximation for the enthalpy of monoenergetic relativistic gas, introduced by Mathews 1971, according to which $h = 1 + \frac{1}{3}(1-\bar{\gamma}^{-2})$ where $\bar{\gamma}$ is the Lorentz factor of particles in the ejecta and is of the order of $\Gamma_{\rm ej}$), we can assume to a $\sim$5% accuracy that $\tilde{L} \approx \frac{h_{\rm j}\rho_{\rm j}\Gamma_{\rm j}^2}{h_{\rm ej}\rho_{\rm ej}\Gamma_{\rm ej}^2}$. Plugging back into Equation (A2), we find

$$\tilde{L}^{1/2}\beta_{\rm j}\left(1-\frac{\beta_{\rm h}}{\beta_{\rm j}}\right) = (\beta_{\rm h}-\beta_{\rm ej}) \quad \text{(A4)}$$

leading to

$$\beta_{\rm h} = \frac{\beta_{\rm j} + \beta_{\rm ej}\tilde{L}^{-1/2}}{1+\tilde{L}^{-1/2}} \quad \text{(A5)}$$

which is the same as Equation (5).

Appendix B
A Monte Carlo Estimation of the Fraction of Successful Short GRB Jets

As mentioned in Section 3.1, GRBs with lower luminosities and shorter durations place the most stringent limits on the waiting time $t_{\rm w}$. Here, we complement the qualitative discussion in Section 3.1 with a Monte Carlo (MC) estimation of the fraction of events that result in successful sGRBs with a certain observed γ-ray luminosity $L_{\rm GRB}$ and duration $t_{\rm GRB}$. For the purposes of this calculation, we use our general jet breakout calculation that holds for a generic medium (i.e., not necessarily static or homologously expanding; see Section 3.3).

As we are not interested in reproducing the exact distribution of Swift–BAT bursts in the $L_{\rm GRB} - t_{\rm GRB}$ parameter space, but rather aim to highlight the effect of the breakout, we employ the following method. We assume log-uniform priors[8] for the engine time ($t_{\rm e}$) and engine power ($L_{\rm e}$) distributions and using Equations (4)–(8) we calculate the luminosity and duration of each jet that successfully breaks out from the BNS ejecta (i.e., with $t_{\rm e} > t_{\rm j,b}$).

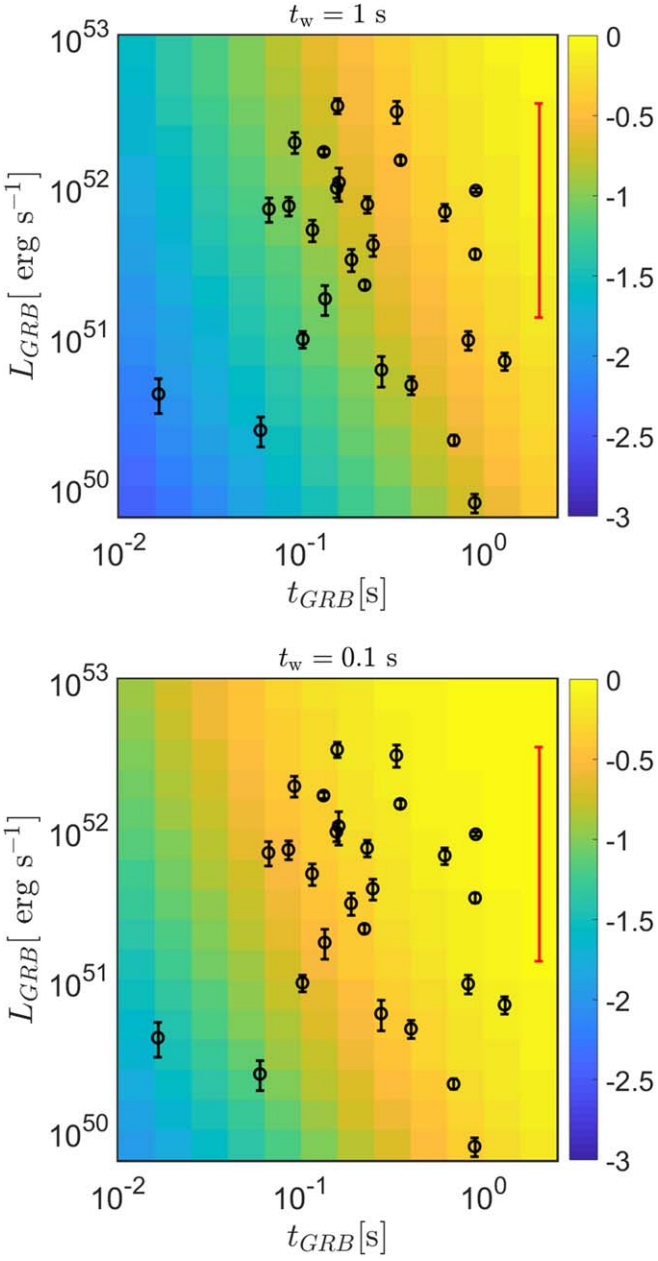

Figure 8. Density map of GRB luminosities and durations, as obtained from 10^8 Monte Carlo realizations, assuming log-uniform priors for the engine time and engine power, and different waiting times (marked on the plot). Colour denotes the probability of breakout (in logarithm). Black and red symbols have the same meaning as in Figure 2. For increasing values of $t_{\rm w}$, the probability of obtaining bursts with short duration and low GRB luminosity decreases.

Figure 8 shows the density maps of all simulated bursts that successfully break out from the BNS ejecta, computed for two values of the waiting time ($t_{\rm w} = 1$ s and 0.1 s) and using 10^8 MC realizations for each case. Swift–BAT bursts with measured redshift (black circles) and GRB 170817 (red symbol) are overplotted. For a given value of $t_{\rm w}$, bursts with lower luminosity (or, equivalently, engine power) have longer breakout times $t_{\rm j,b}$ (see, e.g., Figure 4). Thus, they are less likely to successfully break out from the BNS merger ejecta (they require longer engine activity durations corresponding to a smaller fraction of simulated events). This results in a

[8] The limits on the distributions of $t_{\rm e}$, $L_{\rm e}$ do not affect the probability as long as (i) we reach sufficiently large $t_{\rm e}$, $L_{\rm e}$ such that the probability of breakout is essentially 100% and (ii) the adopted ranges are wide enough that all possible combinations of $t_{\rm e}$, $t_{\rm GRB}$ are realized.

deficiency of simulated bursts with low $L_{\rm GRB}$. Furthermore, longer breakout times mean that a short GRB duration requires fine-tuning in terms of $t_{\rm e}/t_{\rm j,b}$ (see Equation (3)). Therefore, the fraction of successful bursts decreases also with diminishing $t_{\rm GRB}$.

For increasing values of $t_{\rm w}$, the probability of obtaining bursts with short duration and low GRB luminosity decreases (see the top panel of Figure 8). This effect is largely insensitive to the assumed priors on the probabilities, as it is due to the low breakout probability. As an example, for $t_{\rm w}=1$ s, there are two out of 27 bursts that have breakout probabilities of $\sim$0.01 and nine out of 27 with breakout probabilities $\lesssim$0.1. These are well below the expected statistical fluctuations from Poisson statistics, corresponding to a $>5\sigma$ deviation. Of course, the exact probabilities depend on the assumed priors, but the overall conclusion, that low values of $t_{\rm w}$ are required to explain the observed sGRBs, is largely insensitive to those priors.

We finally note that the top right corners of both panels in Figure 8 (which for the adopted priors are expected to be the most populated) are in practice empty of Swift–BAT sGRBs. This result should not be surprising, as it simply reflects the fact that the true distributions of the engine properties (i.e., $t_{\rm e}$ and $L_{\rm e}$) are not expected to be as simple as the statistically independent log-uniform priors assumed here.

ORCID iDs

Paz Beniamini https://orcid.org/0000-0001-7833-1043
Maria Petropoulou https://orcid.org/0000-0001-6640-0179

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Aloy, M. A., Janka, H. T., & Müller, E. 2005, A&A, 436, 273
Band, D., Matteson, J., Ford, L., et al. 1993, ApJ, 413, 281
Bartos, I., Lee, K. H., Corsi, A., Márka, Z., & Márka, S. 2019, MNRAS, 485, 4150
Baumgarte, T. W., Shapiro, S. L., & Shibata, M. 2000, ApJL, 528, L29
Begelman, M. C., & Cioffi, D. F. 1989, ApJL, 345, L21
Beniamini, P., Barniol Duran, R., & Giannios, D. 2018, MNRAS, 476, 1785
Beniamini, P., Duque, R., Daigne, F., & Mochkovitch, R. 2020a, MNRAS, 492, 2847
Beniamini, P., & Giannios, D. 2017, MNRAS, 468, 3202
Beniamini, P., Giannios, D., & Metzger, B. D. 2017, MNRAS, 472, 3058
Beniamini, P., & Granot, J. 2016, MNRAS, 459, 3635
Beniamini, P., Granot, J., & Gill, R. 2020b, MNRAS, 493, 3521
Beniamini, P., & Nakar, E. 2019, MNRAS, 482, 5430
Beniamini, P., Nava, L., Duran, R. B., & Piran, T. 2015, MNRAS, 454, 1073
Beniamini, P., Petropoulou, M., Barniol Duran, R., & Giannios, D. 2019, MNRAS, 483, 840
Beniamini, P., & Piran, T. 2016, MNRAS, 456, 4089
Berger, E. 2014, ARA&A, 52, 43
Bloom, J. S., Prochaska, J. X., Pooley, D., et al. 2006, ApJ, 638, 354
Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2011, ApJ, 740, 100
Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2012, ApJ, 749, 110
Bromberg, O., Nakar, E., Piran, T., & Sari, R. 2013, ApJ, 764, 179
Burns, E., Veres, P., Connaughton, V., et al. 2018, ApJL, 863, L34
Christie, I. M., Lalakos, A., Tchekhovskoy, A., et al. 2019, MNRAS, 490, 4811
Daigne, F., & Mochkovitch, R. 1998, MNRAS, 296, 275
Dichiara, S., Troja, E., O'Connor, B., et al. 2020, MNRAS, 492, 5011
Duffell, P. C., Quataert, E., Kasen, D., & Klion, H. 2018, ApJ, 866, 3
Favata, M. 2014, PhRvL, 112, 101101
Finstad, D., De, S., Brown, D. A., Berger, E., & Biwer, C. M. 2018, ApJL, 860, L2
Flanagan, É. É., & Hinderer, T. 2008, PhRvD, 77, 021502
Fryer, C. L., Belczynski, K., Ramirez-Ruiz, E., et al. 2015, ApJ, 812, 24
Gehrels, N., Chincarini, G., Giommi, P., et al. 2004, ApJ, 611, 1005
Geng, J.-J., Zhang, B., Kölligan, A., Kuiper, R., & Huang, Y.-F. 2019, ApJL, 877, L40
Giannios, D. 2012, MNRAS, 422, 3092
Gill, R., Granot, J., De Colle, F., & Urrutia, G. 2019a, ApJ, 883, 15
Gill, R., Nathanail, A., & Rezzolla, L. 2019b, ApJ, 876, 139
Goldstein, A., Veres, P., Burns, E., et al. 2017, ApJL, 848, L14
Granot, J., Guetta, D., & Gill, R. 2017, ApJL, 850, L24
Hamidani, H., Kiuchi, K., & Ioka, K. 2020, MNRAS, 491, 3192
Harrison, R., Gottlieb, O., & Nakar, E. 2018, MNRAS, 477, 2128
Hotokezaka, K., Beniamini, P., & Piran, T. 2018, IJMPD, 27, 1842005
Just, O., Obergaulinger, M., Janka, H. T., Bauswein, A., & Schwarz, N. 2016, ApJL, 816, L30
Kasen, D., Metzger, B., Barnes, J., Quataert, E., & Ramirez-Ruiz, E. 2017, Natur, 551, 80
Kathirgamaraju, A., Barniol Duran, R., & Giannios, D. 2018, MNRAS, 473, L121
Kathirgamaraju, A., Tchekhovskoy, A., Giannios, D., & Barniol Duran, R. 2019, MNRAS, 484, L98
Kumar, P., & Narayan, R. 2009, MNRAS, 395, 472
Lamb, G. P., & Kobayashi, S. 2017, MNRAS, 472, 4953
Lazzati, D., Deich, A., Morsony, B. J., & Workman, J. C. 2017, MNRAS, 471, 1652
Lazzati, D., & Perna, R. 2019, ApJ, 881, 89
Lyutikov, M. 2020, MNRAS, 491, 483
Mandhai, S., Tanvir, N., Lamb, G., Levan, A., & Tsang, D. 2019, arXiv:1908.00100.
Marti, J. M., Mueller, E., & Ibanez, J. M. 1994, A&A, 281, L9
Mathews, W. G. 1971, ApJ, 165, 147
Matzner, C. D. 2003, MNRAS, 345, 575
Metzger, B. D., Beniamini, P., & Giannios, D. 2018, ApJ, 857, 95
Metzger, B. D., Giannios, D., Thompson, T. A., Bucciantini, N., & Quataert, E. 2011, MNRAS, 413, 2031
Mizuta, A., & Ioka, K. 2013, ApJ, 777, 162
Moharana, R., & Piran, T. 2017, MNRAS, 472, L55
Mooley, K. P., Deller, A. T., Gottlieb, O., et al. 2018, Natur, 561, 355
Murguia-Berthier, A., Montes, G., Ramirez-Ruiz, E., De Colle, F., & Lee, W. H. 2014, ApJL, 788, L8
Murguia-Berthier, A., Ramirez-Ruiz, E., Montes, G., et al. 2017, ApJL, 835, L34
Nagakura, H., Hotokezaka, K., Sekiguchi, Y., Shibata, M., & Ioka, K. 2014, ApJL, 784, L28
Nakar, E. 2007, PhR, 442, 166
Nakar, E. 2019, arXiv:1912.05659.
Nakar, E., & Sari, R. 2010, ApJ, 725, 904
Nava, L., Ghirlanda, G., Ghisellini, G., & Celotti, A. 2011, A&A, 530, A21
Oganesyan, G., Ascenzi, S., Branchesi, M., et al. 2020, ApJ, 893, 88
Petropoulou, M., Barniol Duran, R., & Giannios, D. 2017, MNRAS, 472, 2722
Radice, D., Perego, A., Zappa, F., & Bernuzzi, S. 2018, ApJL, 852, L29
Ramirez-Ruiz, E., Celotti, A., & Rees, M. J. 2002, MNRAS, 337, 1349
Salafia, O. S., Barbieri, C., Ascenzi, S., & Toffano, M. 2020, A&A, 636, A105
Sari, R., & Piran, T. 1997, ApJ, 485, 270
Sobacchi, E., Granot, J., Bromberg, O., & Sormani, M. C. 2017, MNRAS, 472, 616
Troja, E., Ryan, G., Piro, L., et al. 2018, NatCo, 9, 4089
Troja, E., van Eerten, H., Ryan, G., et al. 2019, MNRAS, 489, 1919
Wanderman, D., & Piran, T. 2015, MNRAS, 448, 3026
Xie, X., Zrake, J., & MacFadyen, A. 2018, ApJ, 863, 58
Zhang, B. 2019, FrPhy, 14, 64402
Zhang, B., & Yan, H. 2011, ApJ, 726, 90