Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A

B. P. Abbott
California Institute of Technology

K. AultONeal
Embry-Riddle Aeronautical University

S. Gaudio
Embry-Riddle Aeronautical University

K. Gill
Embry-Riddle Aeronautical University

E. M. Gretarsson
Embry-Riddle Aeronautical University

See next page for additional authors

Follow this and additional works at: https://commons.erau.edu/publication

Part of the Cosmology, Relativity, and Gravity Commons

Scholarly Commons Citation

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 license. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
This Article is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Publications by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu, wolfe309@erau.edu.
Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A

LIGO Scientific Collaboration and Virgo Collaboration, *Fermi* Gamma-ray Burst Monitor, and INTEGRAL

(See the end matter for the full list of authors.)

Received 2017 October 6; revised 2017 October 9; accepted 2017 October 9; published 2017 October 16

Abstract

On 2017 August 17, the gravitational-wave event GW170817 was observed by the Advanced LIGO and Virgo detectors, and the gamma-ray burst (GRB) GRB 170817A was observed independently by the *Fermi* Gamma-ray Burst Monitor, and the Anti-Coincidence Shield for the Spectrometer for the *International Gamma-Ray Astrophysics Laboratory*. The probability of the near-simultaneous temporal and spatial observation of GRB 170817A and GW170817 occurring by chance is $5.0 \times 10^{-3}$. We therefore confirm binary neutron star mergers as a progenitor of short GRBs. The association of GW170817 and GRB 170817A provides new insight into fundamental physics and the origin of short GRBs. We use the observed time delay of $(+1.74 \pm 0.05)$ s between GRB 170817A and GW170817 to: (i) constrain the difference between the speed of gravity and the speed of light to be between $-3 \times 10^{-15}$ and $+7 \times 10^{-16}$ times the speed of light, (ii) place new bounds on the violation of Lorentz invariance, (iii) present a new test of the equivalence principle by constraining the Shapiro delay between gravitational and electromagnetic radiation. We also use the time delay to constrain the size and bulk Lorentz factor of the region emitting the gamma-rays. GRB 170817A is the closest short GRB with a known distance, but is between 2 and 6 orders of magnitude less energetic than other bursts with measured redshift. A new generation of gamma-ray detectors, and subthreshold searches in existing detectors, will be essential to detect similar short bursts at greater distances. Finally, we predict a joint detection rate for the *Fermi* Gamma-ray Burst Monitor and the Advanced LIGO and Virgo detectors of 0.1–1.4 per year during the 2018–2019 observing run and 0.3–1.7 per year at design sensitivity.

Key words: binaries; close – gamma-ray burst; general – gravitational waves

1. Introduction and Background

GW170817 and GRB 170817A mark the discovery of a binary neutron star (BNS) merger detected both as a gravitational wave (GW; LIGO Scientific Collaboration & Virgo Collaboration 2017a) and a short-duration gamma-ray burst (SGRB; Goldstein et al. 2017; Savchenko et al. 2017b). Detecting GW radiation from the coalescence of BNS and neutron star (NS)–black hole (BH) binary systems has been a major goal (Abbott et al. 2017a) of the LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) experiments. This was at least partly motivated by their promise of being the most likely sources of simultaneously detectable GW and electromagnetic (EM) radiation from the same source. This is important as joint detections enable a wealth of science unavailable from either messenger alone (Abbott et al. 2017f). BNS mergers are predicted to yield signatures across the EM spectrum (Metzger & Berger 2012; Piran et al. 2013), including SGRBs (Blinnikov et al. 1984; Paczynski 1986; Eichler et al. 1989; Paczynski 1991; Narayan et al. 1992), which produce prompt emission in gamma-rays and longer-lived afterglows.

A major astrophysical implication of a joint detection of an SGRB and of GWs from a BNS merger is the confirmation that these binaries are indeed the progenitors of at least some SGRBs. GRBs are classified as short or long depending on the duration of their prompt gamma-ray emission. This cut is based on spectral differences in gamma-rays and the bimodality of the observed distribution of these durations (Dezalay et al. 1992; Kouveliotou et al. 1993). This empirical division was accompanied by hypotheses that the two classes have different progenitors. Long GRBs have been firmly connected to the collapse of massive stars through the detection of associated Type Ibc core-collapse supernovae (see Galama et al. 1998, as well as Hjorth & Bloom 2012 and references therein). Prior to the reports of GW170817 and GRB 170817A, SGRBs were firmly connected to the collapse of massive stars through the detection of associated Type Ibc core-collapse supernovae (see Galama et al. 1998, as well as Hjorth & Bloom 2012 and references therein). Prior to the results reported here, support for the connection between SGRBs and mergers of BNSs (or NS–BH binaries) came only from indirect observational evidence (Nakar 2007; Berger et al. 2013; Tanvir et al. 2013; Berger 2014), population synthesis studies (Bloom et al. 1999; Fryer et al. 1999; Belczynski et al. 2006), and numerical simulations (e.g., Aloy et al. 2005; Rezzolla et al. 2011; Kiuchi et al. 2015; Baiotti & Rezzolla 2017; Kawamura et al. 2016; Ruiz et al. 2016). The unambiguous joint detection of GW and EM radiation from the same event confirms that BNS mergers are progenitors of (at least some) SGRBs.

In Section 2 we describe the independent observations of GW170817 by the LIGO–Virgo and of GRB 170817A by the *Fermi* Gamma-ray Burst Monitor (GBM) and by the SPECTrometer on board INTEGRAL Anti-Coincidence Shield (SPI-ACS). In Section 3 we establish the firm association between GW170817 and GRB 170817A. In Section 4 we explore the constraints on fundamental physics that can be obtained from the time separation between the GW and EM signals. In Section 5 we explore the implications of the joint detection of GW170817 and GRB 170817A on the SGRB engine and the NS equation of state (EOS). In Section 6 we explore the implications of the comparative dimness of GRB 170817A relative to the known SGRB population and revise the...
expectation rates for joint BNS–SGRB detections in the light of this discovery.

## 2. Observational Results

The observations of GW170817 and of GRB 170817A are described in detail in Abbott et al. (2017e), Goldstein et al. (2017), and Savchenko et al. (2017b). Here we summarize the observations relevant to the results presented in this Letter and report the results of two fully coherent searches for GWs from the sky location of GRB 170817A. For convenience, all measurements of time have been converted to their geocentric equivalent.

### 2.1. LIGO–Virgo Observation of GW170817

GW170817 is a GW signal from the inspiral of two low-mass compact objects and is the first GW observation consistent with a BNS coalescence (Abbott et al. 2017e, 2017f). GW170817 was first observed by a low-latency search (Cannon et al. 2012; Messick et al. 2017) on 2017 August 17 at 12:41:04 UTC as a single-detector trigger in the LIGO-Hanford detector (Abbott et al. 2017e; LIGO Scientific Collaboration & Virgo Collaboration 2017a). The temporal proximity of GRB 170817A was immediately identified by automatic comparison of the Fermi–GBM Gamma-ray Coordinates Network notice to the GW trigger (Urban 2016). Rapid offline re-analysis (Usman et al. 2016; Nitz et al. 2017b) of data from the LIGO/Virgo network confirmed the presence of a significant coincident signal in the LIGO GW detectors with a combined signal-to-noise ratio (S/N) of 32.4. The combination of observations from the LIGO and Virgo detectors allowed a precise sky position localization to an area of 28 deg² at 90% probability shown in green in Figure 1 (Abbott et al. 2017e; LIGO Scientific Collaboration & Virgo Collaboration 2017b). A time-frequency representation of the LIGO data containing GW170817 is shown in the bottom panel of Figure 2. The GPS time of the merger of GW170817 is $T_0^\text{GW} = 1187008882.430 \pm 0.002$ s (Abbott et al. 2017e). At the observed signal strength, the false alarm rate of the all-sky search for compact-object mergers is less than 1 in 80,000 years (Abbott et al. 2017e). The offline searches target binaries with (detector frame) total mass 2–500 $M_\odot$. Signals are required to be coincident in time and mass in the LIGO detectors, but Virgo data are not used in the significance estimates of the all-sky offline search (Abbott et al. 2017e).

We present the results of two offline targeted searches that coherently combine the data from the LIGO and Virgo detectors and restrict the signal offset time and sky-location using information from the EM observation of GRB 170817A. The onset of gamma-ray emission from a BNS merger progenitor is predicted to be within a few seconds after the merger, given that the central engine is expected to form within a few seconds and that the jet propagation delays are at most of the order of the SGRB duration (see, e.g., Finn et al. 1999; Abadie et al. 2012 and references therein). The gravitational and EM waves are expected to travel at the same speed.

The first targeted search (Harry & Fairhurst 2011; Williamson et al. 2014; Abbott et al. 2017b; Nitz et al. 2017a) assumes that the source is a BNS or NS–BH binary merger and is located at the sky-position observed for the optical counterpart to GW170817 and GRB 170817A (Coulter et al. 2017a, 2017b; Abbott et al. 2017b) and that there is a $[-1, +5]$ s time delay in the arrival of gamma-rays (determined by the GBM trigger time) compared to the binary merger time (Abbott et al. 2017b). At the detection statistic value assigned to GW170817, this search has a $p$-value of $<9.4 \times 10^{-6}$ ($>4.2\sigma$), with this significance estimate limited by computational resources used to estimate the noise background. The second coherent search does not assume any particular GW morphology or GRB model (Sutton et al. 2010; Was et al. 2012; Abbott et al. 2017b) and uses the GBM localization of GRB 170817A to constrain the sky location of the source. This search allows for a $[-60, +600]$ s coincidence between the gamma-rays and the GWs in order to include potentially larger delays in collapsar models of long GRBs. At the detection-statistic value observed for GW170817, this search has a $p$-value of $1.3 \times 10^{-5}$ ($4.2\sigma$).
The 90% credible intervals (Veitch et al. 2015; Abbott et al. 2017e) for the component masses (in the $m_1 \geq m_2$ convention) are $m_1 \in (1.36, 2.26) M_\odot$ and $m_2 \in (0.86, 1.36) M_\odot$, with total mass $2.82^{+0.47}_{-0.09} M_\odot$, when considering dimensionless spins with magnitudes up to 0.89 (high-spin prior, hereafter). When the dimensionless spin prior is restricted to $\leq 0.05$ (low-spin prior, hereafter), the measured component masses are $m_1 \in (1.36, 1.60) M_\odot$ and $m_2 \in (1.17, 1.36) M_\odot$, and the total mass is

![Figure 2. Joint, multi-messenger detection of GW170817 and GRB 170817A. Top: the summed GBM lightcurve for sodium iodide (NaI) detectors 1, 2, and 5 for GRB 170817A between 10 and 50 keV, matching the 100 ms time bins of the SPI-ACS data. The background estimate from Goldstein et al. (2016) is overlaid in red. Second: the same as the top panel but in the 50–300 keV energy range. Third: the SPI-ACS lightcurve with the energy range starting approximately at 100 keV and with a high energy limit of least 80 MeV. Bottom: the time-frequency map of GW170817 was obtained by coherently combining LIGO-Hanford and LIGO-Livingston data. All times here are referenced to the GW170817 trigger time $T_{GW}$.](image-url)
2.74^{+0.04}_{-0.01} M_{\odot}. This result is consistent with the masses of all known BNS systems (Ozel & Freire 2016; Tauris et al. 2017).

From the GW signal, the best measured combination of the masses is the chirp mass \( \mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5} \), which in the detector frame is found to be 1.197^{+0.009}_{-0.003} M_{\odot}.

The detection of GW170817 triggered a campaign of EM follow-up observations which led to the identification of NGC 4993 as the host galaxy of GW170817/GRB 170817A (Coulter et al. 2017a, 2017b; Abbott et al. 2017). We evaluate the distance to the host galaxy from the ratio of the Hubble flow velocity of the host 3017 \pm 166 km s^{-1} (Abbott et al. 2017g) and two current measurements of the Hubble constant (Ade et al. 2016; Riess et al. 2016). These two distance measures are within a combined range of (42.9 \pm 3.2) Mpc, which is consistent with the distance of 40^{+5}_{-4} Mpc determined with GW data alone and makes GW170817 the closest GW event ever observed (Abbott et al. 2016a, 2016b, 2017c, 2017d, 2017e).

The GW data constrain the inclination angle \( \theta_{JN} \) between the total angular momentum of the system and the line of sight to be anti-aligned, with \( \cos \theta_{JN} \leq -0.54 \) (Abbott et al. 2017e). As the binary system component masses are comparable, the NS spins have little impact on the total angular momentum which is aligned with the orbital angular momentum within a few degrees. For discussions in this Letter we will assume that the orbital and total angular momenta are aligned. The SGRB jet is expected to be perpendicular to the accretion disk of the central engine if powered by neutrino annihilation or aligned with the magnetic pole of the rotating central object (Shibata et al. 2006), hence we assume the SGRB jet is aligned with the system rotation axis. This yields a jet viewing angle \( \zeta = \min(\theta_{JN}, 180^\circ - \theta_{JN}) \leq 56^\circ \). As the distance measurement is correlated with \( \theta_{JN} \), the known distance to NGC 4993 further constrains the viewing angle to \( \zeta \leq 36^\circ \) or \( \zeta \leq 28^\circ \) depending on the assumed value of the Hubble constant (Abbott et al. 2017g), with smaller values of the Hubble constant giving smaller misalignment angles.

2.2. Fermi-GBM Observation of GRB 170817A

GRB 170817A was autonomously detected in-orbit by the GBM flight software in a process known as “triggering.” Goldstein et al. (2017) showed the signal exceeds 5\( \sigma \) in three (of twelve) GBM NaI detectors. The GBM detection showed two apparently distinct components. The triggering pulse, that lasts about half a second and falls within the usual observer distributions for GBM SGRBs, is shorter and spectrally harder than the subsequent softer, weaker emission that lasts a few seconds (Goldstein et al. 2017). Summed GBM lightcurves from the relevant detectors in two energy ranges, selected to show the two distinct components, are shown in the top two panels in Figure 2. The GBM time-tagged event data is binned to match the SPI-ACS temporal resolution (100 ms) and phase (matching bin edges) to allow for an easier comparison between the gamma-ray instruments.

Goldstein et al. (2017) quantify the likelihood of GRB 170817A being an SGRB based only on gamma-ray data. This is done by comparing the measured gamma-ray properties of GRB 170817A to the known distributions of short and long GRBs. Both the duration distribution alone and the duration and spectral hardness distributions together show that GRB 170817A is three times more likely to be an SGRB than a long GRB. These analyses are performed in a standard manner, resulting in a longer duration measure than apparent from the hard spike alone because the softer, weaker tail contributes to the calculated duration.

The final GBM localization of GRB 170817A (including systematic error) calculated by the GBM targeted search pipeline is shown in Figure 1. This pipeline performs a coherent search over all GBM detectors (NaI and BGO) and was originally developed to find gamma-ray signals below the onboard triggering threshold around GW triggers (Blackburn et al. 2015; Connaughton et al. 2016; Goldstein et al. 2016). The 50% and 90% credible regions cover \( \sim 350 \text{deg}^2 \) and \( \sim 1100 \text{deg}^2 \), respectively.

Fitting the main pulse in the GBM data with a parameterized function commonly used for GRB pulses indicates a gamma-ray emission onset of 0.310 \pm 0.048 s before \( T_0^{\text{GBM}} \), where \( T_0^{\text{GBM}} \) is defined as the time of the GBM trigger (Goldstein et al. 2017). Based on the position of the optical transient, the signal arrives at Fermi 3.176 ms before it arrives at geocenter. With this correction we find that the start of the gamma-ray emission relative to the GBM trigger is \( (+1.74 \pm 0.05) \text{s} \). In this Letter all derived gamma-ray results use 68\% confidence levels.

The spectral analysis using the standard GBM catalog criteria uses data from the 256 ms time interval between \( T_0^{\text{GBM}} = 0.192 \text{s} \) and \( T_0^{\text{GBM}} + 0.064 \text{s} \). A fit to the “Comptonized” function, a power law with a high-energy exponential cutoff (see Goldstein et al. 2017 for a detailed explanation of this function), is preferred over both a simple power-law fit or models with more parameters. The fit produces values of \( E_{\text{peak}} = (215 \pm 54) \text{keV} \), and a poorly constrained power-law index \( \alpha = 0.14 \pm 0.59 \). The average flux for this interval in the 10–1000 keV range is \( (5.5 \pm 1.2) \times 10^{-7} \text{erg s}^{-1} \text{cm}^{-2} \) with a corresponding fluence of \( (1.4 \pm 0.3) \times 10^{-5} \text{erg cm}^{-2} \). The shorter peak interval selection from \( T_0^{\text{GBM}} - 0.128 \text{s} \) to \( T_0^{\text{GBM}} - 0.064 \text{s} \) fit prefers the Comptonized function, yielding consistent parameters \( E_{\text{peak}} = (229 \pm 78) \text{keV} \), \( \alpha = 0.85 \pm 1.38 \), and peak energy flux in the 10–1000 keV of \( (7.3 \pm 2.5) \times 10^{-7} \text{erg s}^{-1} \text{cm}^{-2} \).

These standard fits are used to compare GRB 170817A to the rest of the SGRBs detected by GBM and to place GRB 170817A in context with the population of SGRBs with known redshift.

More detailed analysis included spectral fits to the two apparently distinct components. The main emission episode, represented by the peak in Figure 2, appears as a typical SGRB best fit by a power law with an exponential cutoff with spectral index \( \alpha = -0.62 \pm 0.40 \) and \( E_{\text{peak}} = (185 \pm 62) \text{keV} \) over a time interval \( T_0^{\text{GBM}} - 0.320 \text{s} \) to \( T_0^{\text{GBM}} + 0.256 \text{s} \). The time-averaged flux is \( (3.1 \pm 0.7) \times 10^{-7} \text{erg s}^{-1} \text{cm}^{-2} \). The tail emission that appears spectrally soft is best fit by a blackbody (BB) spectrum, with temperature of \( k_{\text{BB}} T = (10.3 \pm 1.5) \text{keV} \) and a time-averaged flux of \( (0.53 \pm 0.10) \times 10^{-7} \text{erg s}^{-1} \text{cm}^{-2} \), with selected source interval \( T_0^{\text{GBM}} + 0.832 \text{s} \) to \( T_0^{\text{GBM}} + 1.984 \text{s} \). However, this emission is too weak and near the lower energy detection bound of GBM to completely rule out a non-thermal spectrum.

The temporal analysis yielded a \( T_{90} \), defined as the time interval over which 90\% of the burst fluence between 50–300 keV is accumulated, of \( (2.0 \pm 0.5) \text{s} \) starting at \( T_0^{\text{GBM}} - 0.192 \text{s} \). The duration extends beyond the main emission pulse due to the soft component. This analysis reports a 64 ms peak photon flux of \( (3.7 \pm 0.9) \text{photons s}^{-1} \text{cm}^{-2} \) and occurs from \( T_0^{\text{GBM}} + 0.0 \text{s} \) to \( T_0^{\text{GBM}} + 0.064 \text{s} \). The
minimum variability timescale for GRB 170817A is (0.125 ± 0.064) s.

Using the soft spectral template of the GBM targeted search, a Band function (Band et al. 1993) with a low energy power law index of −1.9, a high energy index of −3.7, and an $E_{\text{peak}}$ of 70 keV, Goldstein et al. (2017) also set 3σ flux upper limits on precursor impulsive gamma-ray emission. The limits on precursor activity out to $T_{0}^{\text{GBM}} > 200$ s are (6.8−7.3) $\times 10^{-7}$ erg s$^{-1}$ cm$^{-2}$ and (2.0−2.1) $\times 10^{-7}$ erg s$^{-1}$ cm$^{-2}$ for signals of 0.1 s and 1.0 s duration, respectively. The tail emission of GRB 170817A is not consistent with the general behavior of SGRBs with extended emission (Kaneko et al. 2015). We set limits on possible extended emission over 10 s intervals out to $\sim 8$ $T_{0}^{\text{GBM}} + 100$ s is (6.4−6.6) $\times 10^{-8}$ erg s$^{-1}$ cm$^{-2}$. Additional upper limits for representative normal and harder spectra are provided in Goldstein et al. (2017) and are up to a factor of a few less constraining.

2.3. INTEGRAL SPI-ACS Observation of GRB 170817A

The orientation of INTErnational Gamma-Ray Astrophysics Laboratory (INTEGRAL) with respect to the LIGO–Virgo localization of GW170817 favored the observation by SPI-ACS and was such that the sensitivity of the Imager on Board the INTEGRAL Satellite (IBIS) was much lower in comparison (Savchenko et al. 2017b). For comparison of relative sensitivities of different INTEGRAL instruments see Savchenko et al. (2017a).

A routine follow-up search for short transients in SPI-ACS identified a single excess at $T_{0}^{\text{ACS}} = T_{0}^{\text{GW}} + 1.88$ s with S/N = 4.6 at the 0.1 s timescale. The correction to the geocentric system assumes the location of the optical transient and results in delay of the signal arrival to INTEGRAL of 148.96 ms. In order to compare the intensity of the event observed by SPI-ACS to the GBM measurement, we compute the range of fluences compatible with the SPI-ACS data in the $[−0.320, +0.256]$ s time interval centered in $T_{0}^{\text{GBM}}$, assuming the GBM best fit spectral model in the same interval. We derive a fluence estimate of $(1.4 ± 0.4) \times 10^{-6}$ erg cm$^{-2}$ (statistical uncertainty only) in the 75−2000 keV energy range, consistent with GBM.

The significance of the association between the GBM observation of GRB 170817A and the event observed by SPI-ACS is 4.2σ. While SPI-ACS would not have alone reported this event as a GRB, it would have reported the event while searching around GW170817, with an independent association significance of 3.2σ (Savchenko et al. 2017b). SGRBs are routinely jointly detected by GBM and SPI-ACS and the association evidence from time coincidence (quoted above) as well as the consistency between the event fluences and temporal properties observed by the two instruments proves that both GBM and SPI-ACS observed the same event. The difference between the time of arrival of the signal in the SPI-ACS and GBM detectors can be exploited to improve the gamma-ray localization of GRB 170817A, which may be beneficial in future joint detections.

The significant interval of the SPI-ACS lightcurve of GRB 170817A is limited to a single pulse with a duration of 100 ms (third panel in Figure 1). GBM and SPI-ACS see the main pulse as appearing to have different durations because they are sensitive in different energy ranges. If the GBM data are shown in an energy range higher than the standard 50-300 keV, the main pulse is consistent with the 100 ms interval seen in SPI-ACS. The lightcurve observed by SPI-ACS reveals a short rise time (<50 ms) and a rapid drop (<50 ms). We therefore constrain the pulse duration in the energy range observed by SPI-ACS (~75−2000 keV) to less than 100 ms.

3. Unambiguous Association

The separation of GRBs into short and long classes was suggested by their duration distributions and reinforced by differences in the prompt gamma-ray emission of the two classes (Dezalay et al. 1992; Kouveliotou et al. 1993). Tying the short class to a different progenitor from the long class was strengthened by redshift measurements of their hosts (Berger 2014). Association of SGRBs with older stellar populations than long GRBs was supported by the types of galaxies that host them (Fong et al. 2013); the connection to BNS mergers was strengthened by the offsets of SGRBs afterglows from their host galaxies (Troja et al. 2008; Church et al. 2011; Fong & Berger 2013) and by the absence of supernovae following nearby, well-observed SGRBs (Fox et al. 2005; Hjorth et al. 2005a, 2005b; Bloom et al. 2006; Soderberg et al. 2006; D’Avanzo et al. 2009; Kocevski et al. 2010; Rowlinson et al. 2010; Berger et al. 2013). We provide conclusive evidence for the BNS-SGRB connection by quantifying the chance temporal and spatial coincidence for GRB 170817A and GW170817 arising from two independent astrophysical events.

To quantify the temporal agreement, we consider the null hypothesis that SGRB and GW detection events are independent Poisson processes and determine how unlikely it is to observe an unassociated SGRB within $\Delta t_{\text{SGRB–GW}} = 1.74$ s of the GW signal. GWs from a BNS merger have been detected once to date, so the $p$-value is $P_{\text{temporal}} = 2 \times 10^{-6}$, which corresponds to a 4.4σ significance in Gaussian statistics.

In order to quantify the spatial agreement of the independent GBM and LIGO–Virgo localizations, we define the statistic $S = \sum_{i=1}^{N_{\text{bins}}} P_{i} P_{2i}$, where $P_{i}$ and $P_{2i}$ are the posterior probabilities from GBM and LIGO–Virgo maps and $i$ is the HEALPix (Gorski et al. 2005) pixel index. $S$ is then compared against a background distribution generated by randomly shifting and rotating GBM posteriors from a representative sample of 164 SGRBs localized by the targeted search. We factor in the estimated localization systematic, and randomly shift and rotate each map 10 times. This background method accounts for the morphology and size distributions of GBM SGRB localizations. We find a $p$-value $P_{\text{spatial}} = 0.01$ that the two independent localizations agree this well by chance.

The temporal and spatial $p$-values are independent quantities, thus the probability that GRB 170817A and GW170817 occurred this close in time and with this level of location agreement by chance is $P_{\text{temporal}} \times P_{\text{spatial}} = (5.0 \times 10^{-6}) \times (0.01) = 5.0 \times 10^{-8}$, corresponding to a Gaussian-equivalent significance.

---

184 https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrsf.html
of 5.3\(\sigma\). This unambiguous association confirms that BNS mergers are progenitors of (at least some) SGRBs.

4. Implications for Fundamental Physics

Little or no arrival delay between photons and GWs over cosmological distances is expected as the intrinsic emission times are similar and the propagation speeds of EM and GWs are thought to be identical. In this Section we discuss the implications on fundamental physics of the temporal offset of \(\pm1.74 \pm 0.05\) s measured between GW170817 and GRB 170817A.

Standard EM theory minimally coupled to general relativity predicts that GWs and light propagate with identical speeds. The refractive index of vacuum is expected to be unity, and both waves are expected to be affected by background gravitational potentials in the same way. The arrival delay of only a few seconds across a distance greater than one hundred million light years places stringent constraints on deviations from fundamental principles. We use the observed temporal offset, the distance to the source, and the expected emission-time difference to place constraints on the deviation of the speed of gravity from the speed of light, and on violations of Lorentz invariance and the equivalence principle.

4.1. Speed of Gravity

Assuming a small difference in travel time \(\Delta t\) between photons and GWs, and the known travel distance \(D\), the fractional speed difference during the trip can be written \(\Delta v/v_{EM} \approx v_{GW} - v_{EM}\), where \(v_{GW} - v_{EM}\) is the difference between the speed of gravity \(v_{GW}\) and the speed of light \(v_{EM}\). This relation is less constraining for small distances, hence we conservatively use here \(D = 26\) Mpc, the lower bound of the 90\% credible interval on luminosity distance derived from the GW signal (Abbott et al. 2017e). If we conservatively assume that the peak of the GW signal and the first photons were emitted simultaneously, attributing the entire \(\pm1.74 \pm 0.05\) s lag to faster travel by the GW signal, this time difference provides an upper bound on \(\Delta v\). To obtain a lower bound on \(\Delta v\), one can assume that the two signals were emitted at times differing by more than \(\pm1.74 \pm 0.05\) s with the faster EM signal making up some of the difference. As a conservative bound relative to the few second delays discussed in Section 2.1, we assume the SGRB signal was emitted 10 s after the GW signal. The resulting constraint on the fractional speed difference is

\[
-3 \times 10^{-15} \leq \frac{\Delta v}{v_{EM}} \leq +7 \times 10^{-16}.
\]

The intergalactic medium dispersion has negligible impact on the gamma-ray photon speed, with an expected propagation delay many orders of magnitude smaller than our errors on \(v_{GW}\).

Lags much longer than 10 s are proposed in alternative models (e.g., Ciolfi & Siegel 2015; Rezzolla & Kumar 2015), and emission of photons before the merger is also possible (Tsang et al. 2012). Hence, certain exotic scenarios can extend this time difference window to \((-100\) s, 1000 s\), yielding a 2 orders of magnitude broadening of the allowed velocity range on either side. While the emission times of the two messengers are inherently model dependent, conservative assumptions yield dramatic improvements over existing indirect (Kostelecky & Russell 2017) and direct (Cornish et al. 2017) constraints, which allow for time differences of more than 1000 years. Future joint GW–GRB detection should allow disentangling the emission time difference from the relative propagation time, as only the latter is expected to depend on distance.

4.2. Lorentz Invariance Violation Limits

Within a comprehensive effective field theory description of Lorentz violation (Colladay & Kostelecký 1997, 1998; Kostelecký 2004; Tasson 2014), the relative group velocity of GWs and EM waves, is controlled by differences in coefficients for Lorentz violation in the gravitational and EM sectors. We focus here on the non-birefringent, non-dispersive limit at mass dimension \(d = 4\), as it yields by far the most impressive results. In this case, the difference in group velocities for the two sectors takes the form

\[
\Delta v = -\sum_{\ell m} Y_{\ell m}(\hat{n}) \left( \frac{1}{2} (-1)^{\ell + 1} s_{\ell m}^{(4)} - c_{\ell m}^{(4)} \right)
\]

The result is presented in a spherical harmonic, \(Y_{\ell m}\), basis, \(s_{\ell m}^{(4)}\) and \(c_{\ell m}^{(4)}\) being spherical-basis coefficients for Lorentz violation in the gravitational and EM sectors, respectively. The direction \(\hat{n}\) refers to the sky position (provided in Coulter et al. 2017a, 2017b).

For ease of comparison with the many existing sensitivities (Shao 2014a, 2014b; Shao et al. 2017; Kostelecký & Tasson 2015; Bourgoin et al. 2016; Le Poncin-Lafitte et al. 2016; Kostelecký & Russell 2017) to the 4 = 4 gravity-sector coefficients (Bailey & Kostelecký 2006; Hees et al. 2016), an analysis in which the coefficients are constrained one at a time is useful (Flowers et al. 2016), with all other coefficients, including the EM sector ones, set to zero. These results are presented in Table 1 along with the best constraints for each coefficient prior to this work. These results can be compared with the isotropic, \(\alpha_{LV}\) Lorentz violation parametrization (Mirshekari et al. 2012) used by Abbott et al. (2017c) in dispersive GW tests. The \(\alpha_{LV} = 2\) limit of this parametrization is equivalent to the isotropic limit of the framework discussed above, with \(s_{\ell m}^{(4)} \rightarrow \sqrt{4\pi} A\). Constraints on \(A\) for \(\alpha_{LV} = 2\) can be obtained from the first line of Table 1; these cannot be established within the analysis carried out in Abbott et al. (2017c).

4.3. Test of the Equivalence Principle

Probing whether EM radiation and GWs are affected by background gravitational potentials in the same way is a test of the equivalence principle (Will 2014). One way to achieve this is to use the Shapiro effect (Shapiro 1964), which predicts that the propagation time of massless particles in curved spacetime, i.e., through gravitational fields, is slightly increased with respect to the flat spacetime case. We will consider the following simple parametrized form of the Shapiro delay (Krauss & Tremaine 1988; Longo 1988; Gao et al. 2015; Kahya & Desai 2016):

\[
\delta t_S = -\frac{1}{c^3} \int_\epsilon^r U(r(l)) dl,
\]
Table 1

Constraints on the Dimensionless Minimal Gravity Sector Coefficients

<table>
<thead>
<tr>
<th>( f )</th>
<th>Previous Lower</th>
<th>This Work Lower</th>
<th>Coefficient</th>
<th>This Work Upper</th>
<th>Previous Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>(-3 \times 10^{-14})</td>
<td>(-2 \times 10^{-14})</td>
<td>(\sigma_{\text{EM}}^{\text{fi}})</td>
<td>(5 \times 10^{-15})</td>
<td>(8 \times 10^{-15})</td>
</tr>
<tr>
<td>1</td>
<td>(-1 \times 10^{-13})</td>
<td>(-3 \times 10^{-14})</td>
<td>(\Re \sigma_{\text{EM}}^{\text{fi}})</td>
<td>(7 \times 10^{-15})</td>
<td>(7 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-8 \times 10^{-14})</td>
<td>(-1 \times 10^{-14})</td>
<td>(\Im \sigma_{\text{EM}}^{\text{fi}})</td>
<td>(2 \times 10^{-15})</td>
<td>(8 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-7 \times 10^{-14})</td>
<td>(-3 \times 10^{-14})</td>
<td>(\Re \sigma_{\text{EM}}^{\text{gG}})</td>
<td>(7 \times 10^{-15})</td>
<td>(9 \times 10^{-14})</td>
</tr>
<tr>
<td>2</td>
<td>(-1 \times 10^{-13})</td>
<td>(-4 \times 10^{-14})</td>
<td>(\Re \sigma_{\text{EM}}^{\text{fi}})</td>
<td>(8 \times 10^{-15})</td>
<td>(7 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-7 \times 10^{-14})</td>
<td>(-1 \times 10^{-14})</td>
<td>(\Im \sigma_{\text{EM}}^{\text{fi}})</td>
<td>(2 \times 10^{-15})</td>
<td>(7 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-5 \times 10^{-14})</td>
<td>(-4 \times 10^{-14})</td>
<td>(\Re \sigma_{\text{EM}}^{\text{gG}})</td>
<td>(8 \times 10^{-15})</td>
<td>(8 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-6 \times 10^{-14})</td>
<td>(-1 \times 10^{-14})</td>
<td>(\Re \sigma_{\text{EM}}^{\text{gG}})</td>
<td>(3 \times 10^{-15})</td>
<td>(8 \times 10^{-14})</td>
</tr>
<tr>
<td></td>
<td>(-7 \times 10^{-14})</td>
<td>(-2 \times 10^{-14})</td>
<td>(\Im \sigma_{\text{EM}}^{\text{gG}})</td>
<td>(4 \times 10^{-15})</td>
<td>(7 \times 10^{-14})</td>
</tr>
</tbody>
</table>

Note. Constraints on the dimensionless minimal gravity sector coefficients obtained in this work via Equations (1) and (2) appear in columns 3 and 5. The corresponding limits that predate this work are reported in columns 2 and 6; all pre-existing limits are taken from Kostelecký & Tasson (2015), with the exception of the upper limit on \(\sigma_{\text{EM}}^{\text{gG}}\) from Shao (2014a, 2014b). The isotropic upper bound in the first line shows greater than 10 orders of magnitude improvement. The gravity sector coefficients are constrained one at a time, by setting all other coefficients, including those from the EM sector, to zero.

where \(r_e\) and \(r_o\) denote emission and observation positions, respectively, \(U(r)\) is the gravitational potential, and the integral is computed along the wave path. \(\gamma\) parametrizes a deviation from the Einstein–Maxwell theory, which minimally couples classical electromagnetism to general relativity. We allow for different values of \(\gamma\) for the propagation of EM and GWs \((\gamma_{\text{EM}}\) and \(\gamma_{\text{GW}}\), respectively, with \(\gamma_{\text{EM}} = \gamma_{\text{GW}} = 1\) in the Einstein–Maxwell theory).

While obtaining the best bound on the difference between the Shapiro time delays requires modeling the potential \(U(r)\) along the entire line of sight, we determine a conservative bound on \(\gamma_{\text{GW}} - \gamma_{\text{EM}}\) by considering only the effect of the Milky Way outside a sphere of 100 kpc, and by using a Keplerian potential with a mass of \(2.5 \times 10^{11} M_\odot\) (the lowest total mass within a sphere of radius 100 kpc quoted in Bland-Hawthorn & Gerhard 2016, from Gibbons et al. 2014, taking the 95% confidence lower bound) (Krauss & Tremaine 1988; Longo 1988; Gao et al. 2015). Using the same time bounds as Equation (1) we find

\[-2.6 \times 10^{-7} \leq \gamma_{\text{GW}} - \gamma_{\text{EM}} \leq 1.2 \times 10^{-6}.\]

(4)

The best absolute bound on \(\gamma_{\text{EM}}\) is \(\gamma_{\text{EM}} - 1 = (2.1 \pm 2.3) \times 10^{-5}\), from the measurement of the Shapiro delay (at radio wavelengths) with the Cassini spacecraft (Bertotti et al. 2003).

5. Astrophysical Implications

The joint GW–GRB detection provides us with unprecedented information about the central engine of SGRBs. The delay between the GW and the GRB trigger times allows us to examine some basic GRB physics. This delay could be intrinsic to the central engine, reflecting the time elapsed from the moment the binary components come into contact to the formation of a remnant BH and the resulting jet. This interpretation includes the case of a relatively long-lived massive NS remnant, which has been suggested to survive from seconds to minutes after merger (see Faber & Rasio 2012; Baiotti & Rezzolla 2017 and references therein). The delay could also be due to the propagation time of the relativistic jet, including the time it takes for the jet to break out of the dense gaseous environment produced by non-relativistic merger ejecta (Nagakura et al. 2014; Moharan & Piran 2017) and/or the emitting region to become transparent to gamma-rays (Mészáros & Rees 2000).

We first discuss the implications that the time delay between the GW and EM emission has on the physical properties of the emitting region when considering the jet propagation and transparency scenarios. Here we assume that the entire delay is due to the expansion of the emitting region and neglect any intrinsic delays between the moment of binary coalescence and the launching of the resulting jet, thus placing limits on the physical properties of the system. Then we consider the impact of SGRB emission from an NS merger on the EOS of dense matter.

5.1. GRB Physics

The main hard peak observed for GRB 170817A lasted roughly half a second. This peak is consistent with a single intrinsic emission episode as it is well described by a single pulse (Goldstein et al. 2017), showing no evidence for significant substructure (spikes). This interpretation is consistent with the SPI-ACS observation of a single peak (Savchenko et al. 2017b). The GBM detection of GRB 170817A also shows no evidence for photons with energy \(\geq 511\) keV, implying that the outflow does not require a high bulk Lorentz factor \(\Gamma\) to overcome photon–photon absorption at the source.

Explanations for the extreme energetics and short timescales observed in GRBs invoke a near instantaneous release of a large amount of energy in a compact volume of space (Goodman 1986; Paczynski 1986). This is commonly referred to as the fireball model, and it is the framework that we will assume for the remainder of this section. The fireball model is largely independent of the burst progenitor and focuses on the dynamics of such a system after this sudden release of energy. The resulting pair-plasma is optically thick and quickly expands under its own pressure to produce a highly relativistic outflow that coasts asymptotically with a constant Lorentz factor \(\Gamma\). Within the fireball, kinetic energy is imparted to particles...
entained in the outflow, although alternative models exist in which the energy outflow occurs mostly as Poynting flux (Usov 1992; Lyutikov & Blandford 2003). The observed gamma-ray pulses are attributed to shocks internal to this relativistic outflow, which convert some of their kinetic energy into the observed EM radiation (Rees & Meszaros 1994). These shocks could produce the predominantly non-thermal emission observed in most GRBs, although non-shock heating models have also been proposed (e.g., Giannios 2006). The overall multi-pulse duration of a burst is thought to reflect the time that the inner engine was active (e.g., producing inhomogeneities in the outflow represented as shells traveling with different bulk velocities) and the variability of individual pulses reflects the size of the shells producing the emission. For a top-hat jet model, $\Gamma$ is assumed constant over the jet surface and the observer never sees beyond the beaming angle $\theta_b \sim 1/\Gamma$. Therefore, the values inferred from the data are independent of the inclination angle from the total angular momentum axis of the system, as long as the viewer is within the opening angle of the jet.

We can examine the implications of the observed delay between the GW and EM signals in the internal shock scenario if we consider two shells emitted at time $t_{GW} = 0$ and time $t_{GW} + \Delta t_{engine}$. If the Lorentz factor of the second shell, $\gamma_2$, is greater than the Lorentz factor of the first shell, $\gamma_1$, the shells will collide at time

$$t_{delay} = \frac{\Delta t_{engine}}{1 - (\gamma_1/\gamma_2)^2},$$

(5)

which is valid if $\gamma_1, \gamma_2 \gg 1$. If the shells have comparable masses, conservation of energy and momentum leads to a merged shell with Lorentz factor $\gamma_m = (\gamma_1 \gamma_2)^{1/2}$. The resulting pulse profile is determined by two timescales. The rise time (which we equate to the minimum variability timescale) can be attributed to the light-crossing time of the individual emission regions and is expressed as

$$\Delta t_{rise} \approx \frac{\delta R}{2c\gamma_m},$$

(6)

where $\delta R$ is the thickness of the emitting region. The decay time reflects angular effects, where off-axis emission is delayed and affected by a varying Doppler boost due to the curvature of the relativistic shell. This timescale is essentially the difference in light-travel time between photons emitted along the line of sight and photons emitted at an angle $\theta$ along a shell of radius $R$. This timescale may be expressed as

$$\Delta t_{decay} = \frac{R(1 - \cos \Delta \theta)}{c} \approx \frac{R(\Delta \theta)^2}{2c} \approx \frac{R}{2c\gamma_m^2} > \Delta t_{rise},$$

(7)

where we assume that the solid angle accessible to the observer is limited by relativistic beaming and thus given by $\theta \sim 1/\gamma$. At the same time, the distance that the first shell has traveled since ejection is $R_1 \approx 2c\gamma_1^2 t_{delay}$, leading to

$$\Delta t_{decay} \approx t_{delay} (\gamma_1/\gamma_2) = \frac{\Delta t_{engine}}{1 - (\gamma_1/\gamma_2)^2} \frac{\gamma_1}{\gamma_2},$$

(8)

The conclusion is a linear correlation between the delay in the GW and EM signals and the resulting pulse duration, modulo the ratio of the Lorentz factors of the two colliding shells (Fenimore et al. 1996; Kociecki et al. 2007; Krimm et al. 2007).

The relative similarity between the gamma-ray duration $T_{iso}$ and the delay between the GW and the EM emission gives $\Delta t_{decay}/t_{delay} \sim 1$, pointing to an internal shock scenario in which the difference in the Lorentz factors of the colliding shells, $\Delta \gamma$, is much smaller than their typical values, i.e., $\Delta \gamma \ll \gamma$. This would imply that the jet was launched shortly after the time of the merger and points to a relatively short $\Delta t_{engine}$ time in which the central engine was active. Such a scenario would produce a collision that was relatively inefficient at converting the internal energy of the shocks to radiation, resulting in a significant isotropic equivalent kinetic energy remaining in the merged shell (Kobayashi et al. 1997; Krimm et al. 2007). This would lead to a very significant energy injection into the resulting afterglow, producing late time “refreshed shocks” (Rees & Mészáros 1998; Kumar & Piran 2000; Ramirez-Ruiz et al. 2001), which are typically not observed in the X-ray (and optical) lightcurves of SGRBs (e.g., Perley et al. 2009).

Some of these energetics constraints can be alleviated if we exclude the soft thermal emission from the gamma-ray duration estimate. In this case, the prompt non-thermal emission of $\Delta t_{decay} \approx 0.5$ s would be due to internal shocks and the soft thermal emission would be attributed to a separate component. In this case we obtain $\Delta t_{delay}/t_{delay} \approx 0.3$, implying $\gamma_2 \approx 3\gamma_1$. These energetics considerations may suggest that the initial hard pulse and the subsequent thermal emission observed by GBM may indeed be distinct components.

Within the context of the internal shock model, if we assume the entire $(+1.74 \pm 0.05)$ s delay between the GW and the EM emission is due to jet propagation time and use a Lorentz factor of $\gamma < 100$ for the first shell, we can estimate an upper limit to the radius of the relativistic outflow to be $R \sim 5 \times 10^{14}$ cm or $\sim 30$ au. The minimum variability timescale $\Delta t_{rise} \approx 0.125$ s (Goldstein et al. 2017) yields an upper limit on the size of the emitting region of $\delta R \sim 4 \times 10^{15}$ cm, or $\sim 3$ au. The ratio of the two is independent of the unknown Lorentz factor and is of order $\delta R/R \sim 10\%$.

The single-pulsed nature of the gamma-ray emission, as well as the observed $\Delta t_{decay}/t_{delay} \sim 1$, also leaves open the possibility that the GBM signal is entirely of an external shock origin. In this scenario, the relativistic outflow converts its internal energy to radiation due to its interaction with an external medium, such as the interstellar matter (Meszaros & Rees 1992). If we associate the duration of the main pulse with the deceleration time, i.e., the timescale over which the jet is significantly decelerated by interstellar matter of constant density $n$, in the external shock scenario (Dermer et al. 1999):

$$t_{dec} = [3E_{k, iso}/(4\pi \gamma^8 n m_p c^5)]^{1/3} = T_{obs},$$

(9)

$E_{k, iso}$ is the kinetic energy of the jet calculated assuming a gamma-ray production efficiency of 20%, $m_p = 1.67 \times 10^{-27}$ kg is the proton mass, $c$ is the speed of light, and $T_{obs}$ is the approximate duration of the main peak. We can thus estimate
the Lorentz factor of the jet in the external shock scenario to be
\[
\gamma \approx 310 \left( \frac{E_{B,\text{iso}}}{2 \times 10^{47} \text{ erg}} \right)^{1/8} \left( \frac{n}{0.1 \text{ cm}^{-3}} \right)^{-1/8} \left( \frac{T_{\text{obs}}}{0.5 \text{ s}} \right)^{-3/8}.
\]
(10)

The deceleration radius represents the upper limit of efficient energy extraction (even for internal shocks) and can be expressed as
\[
R_{\text{dec}} = 2\gamma^2 c T_{\text{obs}} = 3.0 \times 10^{15} \left( \frac{E_{B,\text{iso}}}{2 \times 10^{47} \text{ erg}} \right)^{1/4}
\times \left( \frac{n}{0.1 \text{ cm}^{-3}} \right)^{-1/4} \left( \frac{T_{\text{obs}}}{0.5 \text{ s}} \right)^{1/4} \text{ cm}.
\]
(11)

Therefore, the deceleration radius and associated Lorentz factor also serve as upper limits to the radius and Lorentz factor of the emitting region in the internal shock scenario.

The soft thermal component observed by GBM could also be due to the photosphere of the fireball before it becomes optically thin to gamma-rays. In this interpretation, the delay between the GW and the GRB trigger times may represent the time it takes for the relativistic fireball to expand and become optically thin to gamma-ray radiation. We can examine this scenario by estimating the time it takes for a fireball to become transparent to high-energy radiation in an environment similar to that of a BNS merger.

Following Mészáros & Rees (2000, hereafter MR00), we assume an outflow with an initial radius \( R_0 = 6G M_{\text{BH}}/c^2 \) (the innermost stable circular orbit of a Schwarzschild BH with a mass equal to \( M_{\text{BH}} \)). In our case \( R_0 = 2.5 \times 10^6 \text{ cm} \), and \( M_{\text{BH}} = 2.8 M_\odot \). Given the GBM observations, we estimate an isotropic equivalent energy of the soft thermal BB component to be \( E_{\text{iso, BB}} = 1.3 \times 10^{46} \text{ erg} \) and peak isotropic luminosity of \( L_{\text{iso, p}} = 1.6 \times 10^{47} \text{ erg s}^{-1} \) (see Section 6.1). We take this luminosity as an upper bound of the average luminosity, which may be estimated as \( L_{\text{iso, BB}} = E_{\text{iso, BB}}/\Delta t_{\text{BB}} = 1.1 \times 10^{46} \text{ erg s}^{-1} \), where we have used a duration of the soft BB component of \( \Delta t_{\text{BB}} = 1.15 \text{ s} \).

Using these parameters along with fiducial values (see Appendix B for details), we estimate the photosphere radius to be (MR00)
\[
R_{\text{ph}} = \frac{L_{\text{ph}} Y}{4\pi m_\text{p} c^3 \eta^3} = 2.01 \times 10^{13} \text{ cm}
\times \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right) \left( \frac{\eta}{18} \right)^{-3} Y.
\]
(12)

Where \( L_0 \) is the initial fireball luminosity, \( Y \) is the number of electrons per baryon (in our case \( Y \approx 1 \)), \( \sigma_T = 6.65 \times 10^{-25} \text{ cm}^2 \) is the Thomson cross-section, and \( \eta \) is the dimensionless entropy of the fireball, whose value is much smaller than the canonical one in the standard fireball model (see Appendix B). We note that \( L_0 \) can be much larger than \( L_{\text{iso, BB}} \), since the fireball must expand and convert the remaining internal energy into kinetic energy of the ejecta.

The laboratory frame time needed for the fireball expanding at roughly the speed of light to reach the transparency radius is
\[
t_{\text{ph}} \approx R_{\text{ph}}/c; \quad \text{thus,}
\]
\[
t_{\text{ph}} \approx 672 \text{s} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right) \left( \frac{\eta}{18} \right)^{-3} Y.
\]
(13)

Following Bianco et al. (2001), we can set a upper bound to the conversion from the laboratory to observer frame by assuming the observer is viewing the fireball at most at an angle \( \cos \theta = \nu/c \):
\[
t_a \approx \frac{t_{\text{ph}}}{\gamma^2} \approx 2.1 \text{s} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right) \left( \frac{\eta}{18} \right)^{-5} Y.
\]
(14)

This upper bound can account for the time delay between the GW and the prompt radiation in the soft thermal peak.

Employing MR00’s Equation (8) we can estimate the observer frame temperature of the expanding fireball at the photospheric radius
\[
T_{\text{ph}}^{\text{obs}} = \eta T_{\text{ph}} \approx 2.3 \text{ keV} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right)^{-5/12}
\times \left( \frac{R_0}{2.5 \times 10^6 \text{ cm}} \right)^{-5/6} \left( \frac{\eta}{18} \right)^{-11/3} Y^{-2/3} \gamma^{-5/6}.
\]
(15)

\( T_{\text{ph}}^{\text{obs}} \) can then be compared to the one obtained from observational fits to the GBM data, which provide a BB temperature of \( T_{\text{BB}}^{\text{obs}} \approx 10.3 \text{ keV} \). Our result underestimates the observed BB temperature by a factor \( \approx 4 \), but we are neglecting Comptonization effects, which may slightly raise the estimated temperature. The corresponding BB luminosity is (MR00)
\[
L_{\text{ph}} \approx 1.7 \times 10^{46} \text{ erg s}^{-1} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right)^{7/12}
\times \left( \frac{R_0}{2.5 \times 10^6 \text{ cm}} \right)^{-5/6} \left( \frac{\eta}{18} \right)^{8/3} Y^{-2/3} \gamma^{-5/6}.
\]
(16)

As we have mentioned above, the average luminosity of the BB component is \( L_{\text{iso, BB}} \approx 1.1 \times 10^{46} \text{ erg s}^{-1} \), which is of the order of \( L_{\text{ph}} \) estimated here.

Therefore, based on the observed temperature and luminosity, the delay between the GW signal and the soft BB component can be accounted for as the time it takes the fireball photosphere to radiate. The primary challenge of this interpretation is in explaining the nature of the hard non-thermal emission preceding the BB component. If both components are the result of the same expanding fireball, the photospheric emission is expected to occur earlier than or at the same time as the non-thermal emission. This requirement can be reconciled with the GBM data if the thermal component was subdominant and indistinguishable during the initial hard non-thermal pulse.

Alternatively, energy dissipation below the photospheric radius could also provide an explanation for the timing of the two pulses (Rees & Mészáros 2005). This could be achieved through a range of possible scenarios. Energy dissipation could occur through inelastic collisions between the decoupled neutron and proton populations within the jet (Beloborodov 2010), for example, or through magnetic reconnection processes (Giannios 2006). The emitted radiation would exhibit a modified blackbody spectrum and be released at the
photospheric radius, resulting in similar arrival times to the thermal emission. A non-thermal pulse could also arise from a forward shock if the deceleration radius were located below the photosphere. Such a scenario would be possible if the density of the external medium were sufficiently in excess of the interstellar medium, which is a distinct possibility for such environments (Goriely et al. 2011; Bauswein et al. 2013; Hotokezaka et al. 2013).

The thermal component could also be the result of “cocoon” emission from shocked material surrounding the relativistic jet (Lazzati et al. 2017), which is expected to be softer and fainter than the non-thermal prompt emission (Ramirez-Ruiz et al. 2002; Pe’er et al. 2006). To examine this scenario, we utilize the relation between the radius of the shock breakout duration and observed temperature proposed by Nakar & Sari (2012):

$$R = 1.4 \times 10^9 \left( \frac{t}{1 \text{ s}} \right) \left( \frac{T}{10 \text{ keV}} \right)^2 \text{ cm.}$$

Using the distance measurement, the blackbody spectral fit implies a radius of $R_{\text{BB}} = 3 \times 10^9 \text{ cm}$ that we can use as a proxy for the cocoon radius. This is within a factor of 4 of the relation, in spite of the fact that it applies to spherical geometry that is not guaranteed here, and the $R_{\text{BB}}$ radius derived from the fit assumes thermal equilibrium. We thus consider this as evidence supporting the cocoon scenario.

Finally, the delay between the GW and EM signals may also be due in part to the time it takes for the relativistic jet to break out of the sub-relativistic dense ejecta surrounding the merger (Nagakura et al. 2014; Moharana & Piran 2017). We estimate that the breakout time for typical dynamical ejecta mass values of $\sim 0.1 M_\odot$ in such a merger (Hotokezaka et al. 2013) could not account for the entire observed delay. Lowering the reference isotropic kinetic luminosity of $L_{\text{iso}} = 10^{51} \text{ erg s}^{-1}$ assumed by Moharana & Piran (2017) could be one way to account for a larger delay. However, a luminosity below the one assumed in the breakout scenario substantially increases the likelihood of a “choked” jet that fails to break out of the surrounding medium (Aloy et al. 2005).

5.2. Neutron Star EOS Constraints

The observation of an SGRB associated with the merger of two NSs can be used to derive constraints on the EOS of NS matter (see theoretical studies by Belczynski et al. 2008; Fryer et al. 2015; Lawrence et al. 2015). To do this, we compare the measurement of the binary mass from the GW signal with two possible models of the merger remnant that powered the SGRB: (i) the merger remnant collapsed to a rotating BH with a surrounding disk that powered the SGRB (Shibata et al. 2006), or (ii) the merger formed a rapidly rotating, strongly magnetized NS (millisecond magnetar) with an accretion disk (Metzger et al. 2008).

We consider a representative sample of EOSs: SLy (Douchin & Haensel 2001), LS220 (Lattimer & Swesty 1991), SFHo (Steiner et al. 2013), H4 (Lackey et al. 2006), APR4 (Akmal et al. 1998), SHT (Shen et al. 2011), and MS1 (Müller & Serot 1996). For each EOS, we compute the maximum stable baryonic mass and gravitational mass of a non-rotating (static) NS, denoted $M_{B,\text{static}}$ and $M_{G,\text{static}}$, respectively, and the maximum baryonic mass of a uniformly rotating NS $M_{B,\text{Uniform}}$ (Gourgoulhon et al. 2001). The merger remnant can only collapse to a BH if its baryonic mass is larger than $M_{B,\text{static}}$.

If we neglect rotational corrections, the baryonic masses $m_B, m_{B,2}$ of the initial NSs are functions of their gravitational masses $m_1, m_2$ only. In this approximation, a fixed total initial baryonic mass, $M_{B,\text{Initial}}$, corresponds to a curve in the $(m_1, m_2)$ parameter space. In Figure 3 we show lines of $M_{B,\text{static}}$ and $M_{B,\text{Uniform}}$ that bound the region of the parameter space in which the total mass of the binary is consistent with a stable non-rotating or uniformly rotating remnant, respectively. The figure also contains the 90% credible region of the gravitational masses obtained with a restricted or full spin prior (Abbott et al. 2017e). We note that the latter has a broader distribution of the component masses, such that the heavier NS can exceed $M_{G,\text{static}}$ for various EOS, which would correspond to either a supramassive (or even hypermassive) NS, or to a light BH. The maximum gravitational masses allowed for each EOS, $M_{G,\text{static}}$, are shown in the figure as vertical lines.
Since the total baryonic mass of the system can only be reduced (by mass ejection), the maximum baryonic mass of the merger remnant and accretion disc is bound by $M_{\text{B}}^{\text{initial}}$. From Figure 3, we can see that for the measured NS gravitational masses with the low-spin prior, the MS1 and SHT EOS could not form a BH since $M_{\text{B}}^{\text{initial}} < M_{\text{B}}^{\text{static}}$. Assuming that the magnitude of the spins is small, the MS1 and SHT EOS are incompatible with BH formation. If the dimensionless spins of the NSs are allowed to be larger than 0.05, BH formation is only disfavored: we find that a fraction 83% (MS1) and 84% (SHT) of the posterior distribution satisfies $M_{\text{B}}^{\text{initial}} < M_{\text{B}}^{\text{static}}$. For both spin priors, we find that the H4, LS220, SFHo, and SLy EOS result in $M_{\text{B}}^{\text{initial}} > M_{\text{B}}^{\text{inform}}$. Even when assuming a large ejecta mass of 0.1 $M_\odot$, the remaining mass cannot form a uniformly rotating NS. For those EOS, the merger either results in prompt BH formation or in a short-lived remnant, with a lifetime determined by the dissipation of differential rotation and/or disk accretion.

To be compatible with scenario (ii), the lifetime of the merger remnant would have to be sufficiently long to power the GRB. We note that prompt BH formation is a dynamic process accessible only to numerical relativity simulations. Although there are parameter studies (Hotokezaka et al. 2011; Bauswein et al. 2013), they only consider equal mass binaries. Considering also the error margins of those studies, we currently cannot exclude prompt collapse for the H4, LS220, SFHo, and SLy EOS. Finally, we note that for the APR4 EOS only the possibility of a stable remnant can be ruled out. More generally, only EOSs with $M_{\text{B}}^{\text{static}} < 3.2 M_\odot$ are consistent with scenario (i) when assuming the low-spin prior, or with $M_{\text{B}}^{\text{static}} < 3.7 M_\odot$ for the wider spin prior. These bounds were derived from the 90% credible intervals of the $M_{\text{B}}^{\text{initial}}$ posteriors (and these, in turn, are determined for each EOS in order to account for binding energy variations). These upper limits are compatible with and complement the lower bounds on $M_{\text{NG}}^{\text{static}}$ from the observation of the most massive known pulsar, which has a mass of $(2.01 \pm 0.04) M_\odot$ (Antoniadis et al. 2013).

In Section 6.5 we will discuss some model-dependent implications of the lack of precursor and temporally extended gamma-ray emission from GRB 170817A on the progenitor NSs.

### 6. Gamma-ray Energetics of GRB 170817A and their Implications

Using the measured gamma-ray energy spectrum and the distance to the host galaxy identified by the associated optical transient, we compare the energetics of GRB 170817A to those of other SGRBs at known redshifts. Finding GRB 170817A to be subluminous, we discuss whether this dimness is an expected observational bias for joint GW–GRB detections, what insight it provides regarding the geometry of the gamma-ray emitting region, what we can learn about the population of SGRBs, update our joint detection estimates, and set limits on gamma-ray precursor and extended emission.

#### 6.1. Isotropic Luminosity and Energetics of GRB 170817A

Using the “standard” spectral information from Goldstein et al. (2017) and the distance to the host galaxy NGC 4993 (42.9 ± 3.2) Mpc, we calculate the energetics of GRB 170817A using the standard formalisms (Bloom et al. 2001; Schaefer 2007). GRBs are believed to be relativistically beamed and their emission collimated (Rhoads 1999). Isotropic energetics are upper bounds on the true total energetics assuming the GRB is observed within the beaming angle of the brightest part of the jet. We estimate that the isotropic energy release in gamma-rays $E_{\text{iso}} = (3.1 \pm 0.7) \times 10^{46}$ erg, and the isotropic peak luminosity, $L_{\text{iso}} = (1.6 \pm 0.6) \times 10^{47}$ erg s$^{-1}$, in the 1 keV–10 MeV energy band. These energetics are from the source interval—i.e., the selected time range the analysis is run over—determined in the standard manner for GBM spectral catalog results, allowing us to compare GRB 170817A to other GRBs throughout this section. The uncertainties on the inferred isotropic energetics values here include the uncertainty on the distance to the host galaxy. As a cross check, the isotropic luminosity is also
estimated using a Bayesian approach proposed by Fan (2017). Assuming a flat prior on isotropic luminosity, we obtain \( L_{\text{iso}} = 1.2 \pm 0.3 \times 10^{47} \) erg s\(^{-1}\), which is consistent with the standard GBM approach. This Bayesian approach can be used to combine future joint GW-GRB observations to provide a redshift-independent estimate of the GRB luminosity function.

The two apparent components of GRB 170817A are sufficiently different that using an average spectrum to estimate the fluence may produce an inaccurate total luminosity. Therefore, we also estimate \( E_{\text{iso}} \) using the “detailed” fits described in Goldstein et al. (2017). Separating the hard peak best fit by a Comptonized function (a power law with an exponential cutoff) and the softer tail best fit by a BB spectrum, we estimate \( E_{\text{iso,comp}} = (4.0 \pm 1.0) \times 10^{46} \) erg, and \( E_{\text{iso,BB}} = (1.3 \pm 0.3) \times 10^{46} \) erg, for a total of \( E_{\text{iso}} = (5.3 \pm 1.0) \times 10^{46} \) erg.

Compared to the distribution of GBM detected GRBs with measured redshift shown in Figure 4, GRB 170817A is 2 orders of magnitude closer and 2 to 6 orders of magnitude less energetic than other SGRBs. In particular, GRB 150101B was previously the weakest SGRB with a firm redshift association (\( \zeta = 0.134; \) Fong et al. 2016), and its energetics (as measured by GBM) \( E_{\text{iso}} = 2.3 \times 10^{49} \) erg, and \( L_{\text{iso}} = 7.5 \times 10^{49} \) erg s\(^{-1}\) are 2–3 orders of magnitude higher. As this was the previous dimmest burst, the minimum luminosity cut of \( 5 \times 10^{49} \) erg s\(^{-1}\) used in Wanderman & Piran (2015) to fit a rate and an \( L_{\text{iso}} \) distribution to existing observations appeared reasonable; however, with GRB 170817A, the lower bound on the isotropic energetic distributions needs to be revised, as discussed in Section 6.4.

### 6.2. Implications of the Dimness on the Central Engine

The broad observed brightness distribution likely arises from a mixture of an intrinsic brightness distribution and geometric effects, which include the inclination angle of the system to Earth, the structure and width of the collimated jet itself, and the relativistic beaming angle \( \theta_b \). We consider several possibilities to explain why GRB 170817A is extremely dim (Figure 5): (i) we viewed it from beyond the half-jet opening angle \( \theta_j \) for a standard top-hat model, (ii) the structure of the jet is more complicated than a simple top-hat model, (iii) the observed emission for GRB 170817A originates from a different mechanism than for most SGRBs, or (iv) it is due solely to the intrinsic luminosity distribution and not the geometry of the system.

**Scenario (i).** Uniform top-hat jets (constant emissivity and Lorentz factor, \( \Gamma \), within the jet aperture) with a sharp edge have been widely used to explain GRB properties, including jet breaks (Rhoads 1999). The top-hat jet is the simplest possible model for calculating off-axis parameters as it captures the basic physics of the system, but it is unable to account for smooth profiles in the Lorentz factor and the emissivity. The observed energetics are significantly lower than they would be if we were within \( \theta_j \).

In the top-hat scenario, off-axis values of physical quantities can be related to the on-axis values through the angle dependence of the relativistic Doppler factor:

\[
\delta_D(\theta) = [\Gamma(1 - \beta \cos \theta)]^{-1} \approx 2\Gamma / (1 + \theta_{\text{obs}}\Gamma^2), \tag{18}
\]

where \( \theta \) is the angle between the velocity vector \( v \) and the line of sight, and \( \beta = v/c \). The relation for duration and peak energy is linear with \( \delta_D \) (see, e.g., Granot et al. 2002):

\[
\frac{T_{\text{iso}}(\text{off - axis})}{T_{\text{iso}}(\text{on - axis})} = \frac{E_{\gamma}(\text{on - axis})}{E_{\gamma}(\text{off - axis})} = \frac{\delta_D(0)}{\delta_D(\theta_j - \zeta)} = \frac{1 - \beta \cos(\theta_j - \zeta)}{1 - \beta} \approx b \approx 1 + \Gamma^2(\zeta - \theta_j)^2, \tag{19}
\]

whereas \( E_{\gamma,\text{iso}}(\text{off-axis}) \) scales approximately \( \propto b^{-2} \) for a viewing angle \( \zeta \) between \( \theta_j \) and 2\( \theta_j \). The duration in the on-axis scenario may be longer than inferred from the above equation, as the variable gamma-ray flux can be discerned above detector noise for a longer fraction of the total activity compared to emission viewed off-axis.

We use the observed quantities for GRB 170817A, \( E_\gamma \approx 200 \) keV, \( E_{\gamma,\text{iso}} = 5.3 \times 10^{46} \) erg, and \( T_{\text{iso}} \approx 2 \) s, as values observed off-axis. If we assume that the on-axis values for GRB 170817A are consistent with typical values observed for SGRBs, we obtain \( E_\gamma = 6(b/30) \) MeV, \( E_{\gamma,\text{iso}} = 5 \times 10^{49}(b/30)^2 \) erg, and \( T_{\text{iso}} = 7 \times 10^{-2}(b/30)^{-1} \) s. In particular using a fiducial range on \( E_{\gamma,\text{iso}}(\text{on - axis}) \) corresponding to the two orders of magnitude spread shown in

---

**Figure 5.** Three potential jet viewing geometries and jet profiles that could explain the observed properties of GRB 170817A, as described by scenarios (i)-(iii) in Section 6.2.
Figure 4 we obtain $b \approx \Gamma^2 (\zeta - \theta_j)^2 \approx 30$ within a factor 3, which is a constraint on the values of $\Gamma$, $\zeta$ and $\theta_j$.

If we assume a viewing angle of $\zeta = 30^\circ$ and $\Gamma = 300$ the uncertainty on $b$ yields $\zeta - \theta_j \approx 1 \pm 0.5$ deg, a solid angle covering only 1% of a full sphere. Hence this configuration would require a fine tuning of the line of sight. However, if we assume $\Gamma = 30$ then the uncertainty on $b$ yields $\zeta - \theta_j \approx 10 \pm 4$ deg, a solid angle that covers 10% of a full sphere, which is plausible without too much fine tuning. This argument only weakly depends on the particular value $\zeta$ and illustrates that for large $\Gamma$ a top-hat jet scenario is disfavored due to the sharp emission fall-off at the edges.

Scenario (ii). A more complex geometry involves a structured jet (Rossi et al. 2002, or Granot 2007 and references therein) which provides a wider range of angles from which the observer could still detect emission, and therefore does not require a fine-tuned viewing angle. Structured jet emission profiles include a uniform ultra-relativistic core surrounded by a power-law decaying wing where the energy and Lorentz factor depend on the distance from the jet axis (Pescalli et al. 2015), a Gaussian with a smooth edge and falloff outside the core (Zhang & Mészáros 2002; Kumar & Granot 2003), and a two-component jet with an ultra-relativistic narrow core and slightly slower outer jet (Frail et al. 2000; Berger et al. 2003; Racusin et al. 2008; Filgas et al. 2011), among other possibilities.

Structured jets can naturally explain the broad observed energetics distribution. Because SGRBs involve relativistic velocities, radiation is strongly beamed into angle $\theta_j$. If the observed brightness depends on viewing geometry, i.e., is not uniform across the angle $\theta_j$, then the part of the beam that we observe may be off-axis to the brightest part of the jet but we may still be within $\theta_b$ of some dimmer part of the emitting region, though in this case we would expect the $\Gamma$ factor to vary as well.

Scenario (iii). Given the closeness of this burst it is possible that the observed emission is due to a different mechanism from other SGRBs, one that is intrinsically dim and thus undetectable at usual SGRB distances. We believe this explanation to be unlikely as the main emission episode of GRB 170817A is a typical SGRB (as measured by the observed gamma-ray properties). It is possible that the soft tail emission arises from a distinct mechanism. One explanation is “cocoon” emission from the relativistic jet shocking its surrounding non-relativistic material (Lazzati et al. 2017). We showed that “cocoon” emission could explain the thermal tail in Section 5.1. A possible full model for GRB 170817A is off-axis emission from a top-hat jet providing the main emission episode, with “cocoon” emission arising from the jet’s interaction with the surrounding torus that powers the main jet. The softer emission is near the detection limits of GBM and would not be detected to much greater distances, suggesting it may be a common property of SGRBs that is otherwise missed.

Scenario (iv). If GRB 170817A is viewed within both the collimated jet and the beaming angle, and the emission is constant across the traditional top-hat jet, then GRB 170817A is intrinsically much dimmer by orders of magnitude compared to other observed SGRBs. This would mean that top-hat jets have an intrinsic distribution covering 6 orders of magnitude, which is difficult to envision given the limited mass ranges in the merger of two NSs (although see Metzger & Berger (2012) and references therein). A broader intrinsic luminosity distribution might be accommodated if we assume that at least some SGRBs arise from the merging of an NS with a BH. It is possible, for example, that the brightest events may arise from NS–BH mergers with optimal mass ratio and spin parameters. Another possibility is that this broad luminosity range could arise from other properties of the system, such as the magnetic field strength of the progenitors or the intrinsic jet-opening angle distribution.

Observations of GW170817/GRB 170817A at other wavelengths (which are not explored in this Letter) will be necessary for a full understanding of this event. For example, evidence for X-ray emission that only arises at late times may provide evidence for this event occurring off-axis (see, e.g., Mészáros et al. 1998; Granot et al. 2002; Yamazaki et al. 2002). However, future joint detections of GW-GRB events can also provide a fuller understanding of the intrinsic energetics distributions and the effect geometry has on our observed brightness. Here the inclination constraint is not particularly informative as the inclination angle constraint, $\zeta \approx 36^\circ$, is comparable to the highest lower limit for a half-jet opening angle, $\theta_j > 25$ deg (Fong et al. 2015). If this is truly off-axis from a top-hat jet then it is unlikely to be a common occurrence. Only joint GW–EM detections will reveal if the intrinsic brightness varies according to the type of progenitor.

The updated expected joint detection rates in Section 6.4 suggest inferences on populations of joint detections may be possible sooner than previously thought.

6.3. Observational Bias Against Low-luminosity GRBs

The fact that GRB 170817A is orders of magnitude dimmer than the population of SGRBs with known redshifts raises the questions: (i) is it unexpectedly dim, and (ii) is there a population of SGRBs with comparable luminosities (and distances) that we are not detecting? We explore here whether the gap in luminosity compared to more luminous SGRBs is a result of the instrumental sensitivity for the detection of either the prompt or the afterglow emission of SGRBs, or whether our problem lies in the association of SGRBs to their host galaxies and thus redshift.

Burns et al. (2016) examine the observed relationship between redshift and gamma-ray fluence for SGRBs with known redshift and find no strong correlation. SGRBs that appear extremely bright are likely to be nearby because their inferred luminosities would otherwise be unrealistic, but SGRBs near the detection threshold of (current) GRB detectors are as likely to be nearby as far away. The intrinsically dim part of the SGRB luminosity distribution is detectable only at short distances.

GRB 170817A is our only clear case of a subluminous SGRB with known distance, so we investigate the maximum distance at which it could have triggered GBM. Assuming the event occurred at the same time and viewing geometry with respect to Fermi, with comparable detector background rates, we find that GRB 170817A could have been $\sim30\%$ dimmer before falling below the on-board triggering threshold (Goldstein et al. 2017), corresponding to a maximum detection distance of about 50 Mpc. An approximate measure of the detectability distance given optimal detection conditions (e.g., low background, good geometry) suggests the maximum distance we could have detected this burst is about 80 Mpc—closer than any
other SGRB with a firmly determined redshift. While the GW horizon has been considered the limiting factor for joint detections with EM signals, this joint detection shows that we now must also account for an SGRB detection horizon given the sensitivity of the current gamma-ray observatories.

In addition to limited gamma-ray detector sensitivity, determining the redshift from EM observations alone is more difficult for SGRBs than for long GRBs. The fraction of SGRBs with detected X-ray afterglows for Swift BAT detected bursts is $\sim 75\%$ (Fong et al. 2015), compared to over $90\%$ for long GRBs.\[^{185}\] It is possible that SGRBs with subluminous prompt gamma-ray emission also have correspondingly weaker X-ray afterglows, and these could account for a large fraction of the quarter of SGRBs without X-ray detections. Even when the X-ray afterglows are detected, they are fainter and thus fade below detectability threshold faster than the afterglows of long GRBs, making direct measurement of the redshift from the afterglow exceedingly rare (Fong et al. 2015). For SGRBs, the redshift is instead usually determined from the host galaxy. This requires first that the afterglow be tied to a particular host galaxy, which can be difficult because the SGRB progenitors sometimes lie outside their putative hosts, owing to the natal kicks induced by the supernovae that produced the compact objects in the progenitor system (Wong et al. 2010). A well-localized (\sim few arcseconds) SGRB afterglow is associated with a galaxy within a small angular distance on the sky, using probabilistic arguments about chance alignment, and then the redshift of the host galaxy is measured.

Appendix B lists all SGRBs with possible redshifts. Most of the list was compiled by combining three relatively complete and recent literature samples (Fong et al. 2015; Lien et al. 2016; Siellez et al. 2016). Nearly all of these were detected by Swift BAT. It has been suggested that the BAT SGRB distribution is contaminated by the short tail of the long GRB distribution (Bromberg et al. 2013). Burns et al. (2016) find that the BAT sample is not significantly more contaminated than the GBM sample and the redshift distribution based on Swift BAT SGRBs is therefore a valid proxy for the redshift distribution of GBM SGRBs in the following discussion.

Berger (2010) discuss “hostless” SGRBs, which are well-localized SGRBs that have no obvious associated host galaxy despite deep observational limits. They suggested the hosts could be nearby galaxies at larger angular offset to the afterglow than others farther away, but also put forward the possibility of more distant, undetected hosts. Tunnicliffe et al. (2014) show that hostless SGRBs have an excess of nearby galaxies within a few arcminutes, relative to long GRBs or random positions, suggesting that at least some of these hostless SGRBs have nearby hosts. Therefore, the traditional assignment of probability of an SGRB to a host galaxy based solely on angular offset from a well-localized afterglow may exclude real associations with larger offsets, which are more likely to be measured for nearby events.

Tunnicliffe et al. (2014) also includes the closest potential host for an SGRB prior to GRB 170817A, with 81 Mpc for GRB 111020A. If real, this association implies an extremely low $E_{\text{iso}}$ of $\sim 10^{46} \text{erg}$, similar to GRB 170817A. In light of the secure association of GRB 170817A with GW170817, a subenergetic $E_{\text{iso}}$ may no longer be a reason to doubt subluminous nearby SGRB host associations, and may suggest a reconsideration of very nearby host galaxies with large projected angular offsets for hostless SGRBs. We include these putative host associations in Table 2. Also included are SGRBs with extended emission and bursts that have durations exceeding the standard $T_{90}$ cut but are believed to be short based on other evidence such as spectral hardness. Asterisks indicate bursts, where the Swift BAT $T_{90} < 2\text{s}$, that have localizations of a few arcseconds or better (as larger localizations increase the chance of false associations due to chance alignment), and for which the angular offset of the afterglow from the host fulfills standard association criteria. For further analysis this restricted sample is our “gold sample,” and the full sample the “total sample.”

One outstanding question is why we have not detected other SGRBs as close as GRB 170817A. We have established that bursts as dim as GRB 170817A will not be detected by current gamma-ray instruments if they lie much farther away than GRB 170817A. This raises questions about GRBs with luminosities between GRB 170817A and the rest of the GRBs with known luminosities. Some of these are surely being detected, albeit with unassigned redshifts and thus luminosities. While there are only $\sim 40$ SGRBs with possible redshifts, several hundreds have been detected without an assigned redshift. Nearby, subluminous SGRBs surely lie among them. There is a lower priority for following-up weak SGRBs, so if nearby events are systematically detected as weak bursts they may not have the required follow-up observations at lower wavelengths to determine the distance to the burst. It could also be that these weak bursts also have lower brightness at lower wavelengths, making them harder to detect even with follow-up observations. Lastly, we could be detecting these bursts in gamma-rays and X-rays, but failing to properly associate them with their hosts as discussed above.

GRB 170817A is unique in that its distance was first measured by GWs, which are currently detectable out to limited distances (roughly 100 Mpc) compared to other SGRBs with known redshifts (see Table 2). This is analogous to the first association of long GRBs with supernovae. Long GRBs have redshifts systematically higher than SGRBs (Coward et al. 2013). The long burst GRB 980425 is the closest GRB to date with a measured distance (and the only GRB of any class closer than GRB 170817A), and it was the first long GRB associated with a supernovae. GRB 980425 was 4 orders of magnitude less energetic than other GRBs detected at that time (Galama et al. 1998). Because supernovae are less luminous than long GRBs, the long GRBs that are associated with supernovae are systematically closer than the average population. Because of the Malmquist bias, a bias toward detecting intrinsically bright objects (Malmquist 1922), we only see dim long GRBs when they are nearby. This explains the subluminous nature of GRB 980425, and this observational peculiarity has been confirmed by other subluminous long GRB-SN detections, including GRB 031203/SN2003lw (Malesani et al. 2004), GRB 060218/ SN2006aj (Modjaz et al. 2006), and GRB 100316D/SN2010bh (Cano et al. 2011), quantified as a population in Howell & Coward (2012). The history of GRB 980425, the other nearby subluminous long GRBs associated with supernova, and the lack of correlation between SGRB gamma-ray fluence and redshift noted by Burns et al. (2016) motivates the further development of subthreshold searches for counterparts to GW events and for subthreshold SGRBs in general. While GRB 170817A occurred nearby, and its favorable geometry

\[^{185}\] https://swift.gsfc.nasa.gov/archive/grb_table/
to Fermi resulted in an on-board trigger, we anticipate that these untriggered searches of GBM and other gamma-ray data will uncover future counterparts to this GW-selected SGRB population.

Suggestions of nearby subluminous SGRB populations existed prior to this discovery (Tanvir et al. 2005; Siellez et al. 2016). Tanvir et al. (2005) find a statistically significant correlation between a large sample of coarsely localized SGRBs detected by the Burst And Transient Source Experiment (BATSE) and a sample of nearby galaxies. Without associating individual SGRBs with a potential host, they conclude that ∼10% of the SGRB sample could be part of a nearby subluminous population. Siellez et al. (2016) infer the presence of a nearby subluminous SGRB population through a study of SGRBs with known redshift in the context of BNS and NS–BH population evolution. They find an excess of actual nearby low-luminosity SGRBs using the results from their simulations, covering a broad range of assumed lifetimes for the binary system prior to merger.

Giant flares from the highly magnetized NSs known as magnetars can be detected outside our galaxy, with the sole extragalactic example tied to its host coming from SGR 0525-66, in the Large Magellanic Cloud (Evans et al. 1980). A giant flare from the galactic magnetar SGR 1806-20 showed a gamma-ray spectrum measured by the Konus–Wind instrument that was well-fit by a blackbody with temperature ∼175 keV (Hurley et al. 2005), harder than a regular magnetar burst. This hard spectrum led to the idea that giant flares from magnetars in nearby galaxies might be a sub-population hiding among the general SGRB population (Hurley et al. 2005). Tanvir et al. (2005) found a stronger correlation of BATSE SGRBs with early-type than late-type galaxies, which is not expected if nearby SGRBs arise from giant magnetar flares in nearby galaxies, but is consistent with a BNS origin. GRB 170817A is clearly associated with a BNS merger, but even without the connection to GW170817, the spectrum of GRB 170817A in the GBM data strongly disfavored the BB fit expected for a giant magnetar flare. Another possible signature of a giant magnetar flare is the ringing in its tail at the NS period of a few seconds, which could be detected by GBM or by SPI-ACS for flares outside our galaxy providing it was close enough. A search for periodic or quasi-periodic emission in the GBM data for GRB 170817A (Goldstein et al. 2017) found no periodic modulation, providing another discriminant between SGRBs and nearby extragalactic giant magnetar flares that might be masquerading as SGRBs.

6.4. Predicted Detection Rates

The intrinsic specific volumetric SGRB rate is often quoted to be around 10 Gpc−3 yr−1 (see, e.g., Guetta & Piran 2006; Coward et al. 2012; Fong et al. 2015; with the value originating from Nakar et al. 2006, who noted that the true rate could be much higher). However, unlike GW signals, SGRBs do not have a clear relationship between the observed distance and brightness. As discussed in the previous section, this can be due to intrinsic variations in SGRB luminosities, as well as structure in the jet. In this Section, we investigate the former scenario presenting the implications of GW170817/GRB 170817A for future GW and SGRB observations in terms of a simple standard model for the SGRB luminosity distribution. Similar interpretations for other, perhaps more elaborate, models are straightforward.

We model the SGRB luminosity function as a broken power law, with a logarithmic distribution:

$$\phi_{\text{iso}}(L_{\text{iso}}) = \begin{cases} \frac{L_{\text{iso}}}{L_*}^{-\alpha_0} & L_{\text{iso}} < L_* \\ \frac{L_{\text{iso}}}{L_*}^{-\beta} & L_{\text{iso}} > L_* \end{cases}$$

(20)

where \(L_{\text{iso}}\) is the peak isotropic luminosity (in the source frame) between 1 keV and 10 MeV, and \(\alpha_0\) and \(\beta\) give the power law decay below and above the break at \(L_*\). Here, we follow Wanderman & Piran (2015) in using \(L_* \sim 2 \times 10^{52} \text{ erg s}^{-1}\), \(\alpha_0 \approx 1\), and \(\beta \approx 2\). The other important parameter is the minimum SGRB luminosity, which determines the lower cutoff of the luminosity distribution. This is poorly constrained as only nearby low luminosity SGRBs are observable. In Wanderman & Piran (2015) the minimum luminosity is taken to be \(L_{\text{min}} = 5 \times 10^{49} \text{ erg s}^{-1}\), while other studies use values ranging from \(1 \times 10^{49} \text{ erg s}^{-1}\) to \(\sim 10^{50} \text{ erg s}^{-1}\) (Regimbau et al. 2015). We assume a threshold value for detectability in GBM of 2 photons cm−2 s−1 for the 64 ms peak photon flux in the 50–300 keV band, which is higher than the minimum detectability value to account for the sky-dependent sensitivity of GBM. Furthermore, we model the SGRB spectrum using the Band function with parameters taken from Wanderman & Piran (2015) (namely, \(E_{\text{peak}} = 800 \text{ keV}, \nu_{\text{Band}} = -0.5\), and \(\nu_{\text{Band}} = -2.25\)). This spectrum is significantly harder than the one observed for GRB 170817A. The cumulative observed rate predicted for GBM by this base model is shown as a function of redshift in Figure 6 by the purple solid curve.

As discussed in Section 6.1, the inferred \(L_{\text{iso}}\) is \((1.6 \pm 0.6) \times 10^{47} \text{ erg s}^{-1}\), which is significantly lower than any previously detected SGRB, and thus is in tension with this model. In particular, we must extend the lower limit of the luminosity down by a factor of at least 500. At present, there is rather little information available about the low luminosity distribution due to the observational biases discussed in Section 6.3 and, consequently, there is a significant degeneracy between the minimum SGRB luminosity and the rate (Wanderman & Piran 2015). Let us consider the most straightforward extension of the above model and set \(L_{\text{min}} = 1 \times 10^{47} \text{ erg s}^{-1}\) while maintaining \(\alpha_0 = 1\). In order to retain the same prediction for high-luminosity SGRBs, this requires a 500-fold increase in the number of SGRBs, with the majority emitting at low luminosity. The cumulative observed rate predicted for GBM by this simple extension is shown as a function of redshift in Figure 6 by the red solid curve, and is comparable to the measured BNS merger rate shown in black. This simple extension would imply SGRBs are not beamed and that essentially all BNS mergers are accompanied by at least a subluminous SGRB.

Therefore, to reduce this tension and explore other possible extensions, we introduce an additional power law break below

\[186\] To get the linear distribution of luminosities, both \(\alpha_0\) and \(\beta\) must be increased by 1.

\[187\] Other studies use a smaller energy band when defining the luminosity, and this has an impact on the value of \(L_*\), although not on the slopes of the power law components.
Figure 6. Predicted detection rates per year as a function of redshift. The red, blue, and green solid lines refer to the GBM observed SGRB rate assuming a minimum luminosity $L_{\text{min}}$ of $1 \times 10^{47}$ erg s$^{-1}$, and $\gamma_L = 1, \beta_L = 2$ and $\gamma_L = \{1, 0.5, 0\}$ in Equation 21, respectively. The purple solid line refers to the base model with $L_{\text{min}}$ of $5 \times 10^{49}$ erg s$^{-1}$. The four curves are normalized by imposing 40 triggered SGRB per year. As $\gamma_L$ increases, the observed rate is no longer volumetric at lower and redshifts, because a fraction of SGRBs becomes too dim to be detected. For reference, the red, blue and green dot-dashed curves show the local SGRB occurrence rate for broadly compatible with all of the models. The dotted vertical cyan line refers to the redshift of GRB 170817A by Abbott et al. (2017e). For comparison, the measured SGRBs redshift distribution from Table 2 is shown in cyan, and is largely compatible with all of the models. The dotted vertical cyan line refers to the redshift of GRB 170817A host galaxy.

Equation 21: $\phi_{\text{iso}}(L_{\text{iso}}) = \begin{cases} \left( \frac{L_{\text{iso}}}{L_{\text{iso}}} \right)^{-\gamma_L} & L_{\text{iso}} < L_{\text{iso}} < L_{\text{iso}} \leq L_{\text{iso}} < L_{\text{iso}} \\ \left( \frac{L_{\text{iso}}}{L_{\text{iso}}} \right)^{-\beta_L} & L_{\text{iso}} < L_{\text{iso}} \leq L_{\text{iso}} < L_{\text{iso}} \end{cases}$

and consider three values $\gamma_L = \{1, 0.5, 0\}$ for the power law index below this second break. We normalize these three cases to 40 triggered SGRBs per year for GBM, and the $\gamma_L = 1$ case corresponds to the simple extension discussed above. For reference, Figure 6 shows the local SGRB occurrence rate for $L_{\text{min}} = 1 \times 10^{47}$ erg s$^{-1}$ and $\gamma_L = \{1, 0.5, 0\}$ indicated by the red, blue, and green dotted curves, respectively, and the BNS merger rate $1540_{-1200}^{+1200}$ Gpc$^{-3}$yr$^{-1}$ determined with the detection of GW170817 (Abbott et al. 2017e). The $\gamma_L = 1$ case produces the largest number of subluminous SGRBs, and leads to a sharp departure at redshift $\sim 0.005$ from volumetric detection to detection limited by the GBM sensitivity. For $\gamma_L = 0$ the transition is smoother as there is only a small number of subluminous SGRBs, and the observed rate departs gradually from the occurrence rate.

When we include the luminosity of GRB 170817A by setting $L_{\text{min}} = 1 \times 10^{47}$ erg s$^{-1}$ and $\gamma_L = 1$, the expected detection rate at a redshift of $z \approx 0.1$ is around a factor of 2 higher than for the Wanderman & Piran (2015) model. At a redshift of $z \approx 0.01$, which is close to the observed redshift for GRB 170817A, rather than expecting to observe 1 event per 650 years with GBM, this is increased to 1 per year. The expected detection rate at this redshift for the $\gamma_L = 0.5$ and $\gamma_L = 0$ extensions is of roughly 1 observed event per 10 and 65 years, respectively. The expectations we obtain for GBM are consistent with the distribution of SGRBs with known redshifts reported in Table 2, in Appendix B.

Using the BNS merger volumetric rate estimated from GW170817 as a new input to the detection rate calculation presented in Abbott et al. (2017a), the LIGO–Virgo detection rate is narrowed down from 0.04–100 to $\sim 1–50$ BNS coalescences during the 2018–19 observing run, with the remaining uncertainty arising in part from the not-yet-known detector sensitivities during that run. At design sensitivity, the LIGO and Virgo detectors can expect to detect $\sim 6–120$ BNS coalescences per year, as opposed to the previously estimated 0.1–200 BNS coalescences per year. Inclusion of any additional BNS detections in the meantime will allow this prediction to be further sharpened.

Independently, we use the GBM detection rate as a function of redshift to predict joint GW–GBR BNS detection rates (Clark et al. 2015). Both the rates and their relative uncertainties are significantly reduced, compared to the GW-only detection rate estimates above, since the majority of distant mergers will be undetectable by GBM and the GBM...
SGRB detection rate is well measured. The degree to which GBM-selected SGRBs are preferentially on-axis is unclear. When estimating the joint detection rate we include both no selection and a pure on-axis selection hypothesis, the latter implying a larger detection probability by the LIGO and Virgo detectors. During the 2018–19 observing run, we expect 0.1–1.4 joint detections per year for the GW interferometer network and GBM triggered SGRBs, with the high end of the interval corresponding to the $\alpha_L = 1, \beta_{\text{min}} = 1 \times 10^{47}$ erg s$^{-1}$ extension of the luminosity function. At design sensitivity, the expected joint detection rate increases to 0.3–1.7 per year.

Future joint GW-SGRB observations will provide significant new insights into low-luminosity SGRBs. In particular, both joint observations and lower limits on distances to SGRBs not observed in GWs (e.g., GRB 150906B in Abbott et al. 2017b) will constrain the rate of nearby GRBs. Future GW observations of BNS mergers will reduce the uncertainty in the rate of such events, while observation of GW signals with no SGRB counterpart will limit the SGRB beaming angles. Finally, sub-threshold searches in GRB data around the time of GW events could significantly increase the number of joint observations.

### 6.5. Limits on Precursor and Extended Emission

At gamma-ray energies, SGRBs are characterized by a prompt emission episode lasting at most $\sim 2$ s. Observational evidence for precursor flares associated with SGRBs (Troja et al. 2010; Burns 2017; Minev & Pozanenko 2017) and temporally extended emission (Lazzati et al. 2001; Connaughton 2002) is so far inconclusive. Given the small distance to the source, the absence of such emission from GRB 170817A provides an important data point and may constrain models that predict it. The flux upper limits set in Section 2.2 correspond to an intrinsic upper limit of $\sim 2.4 \times 10^{47}$ erg s$^{-1}$ for precursor emission on the 0.1 s timescale, $\sim 7.0 \times 10^{46}$ erg s$^{-1}$ for precursor emission on the 1.0 s timescale, and $\sim 2.2 \times 10^{46}$ erg s$^{-1}$ for extended emission on the 10 s timescale.

Magnetospheric interactions in NS binaries have been proposed as a source of nearly isotropic emission preceding the merger (e.g., Hansen & Lyutikov 2001; Metzger & Zivancev 2016). In the context of such models, the non-detection of precursors associated with GW170817 suggests the absence of strong magnetic fields in the last $\sim 200$ s before merger. Hansen & Lyutikov (2001), for instance, predict a luminosity that depends on the magnetic field $B$ as $L \approx 7.4 \times 10^{45} (B/10^{15} \text{ G})^2 (a/10^6 \text{ cm})^{-7}$ erg s$^{-1}$. We can combine this estimate with the least-constraining GBM intrinsic upper limit above and assume a final separation $a = 3 \times 10^6$ cm before disruption. The resulting limit is $B < 8 \times 10^{13}$ G, which is weaker than the magnetic fields of most known magnetars (Olausen & Kaspi 2014). However, the GBM upper limit still lies within the luminosity range of other similar models (Metzger & Zivancev 2016).

Resonant shattering of the NS crust has also been proposed as a source of emission prior to merger, with a maximum time delay of tens of seconds and nearly isotropic angular distribution (Tsang et al. 2012). The luminosity of such precursor emission depends on the crust breaking strain $\epsilon_b$ and the emission timescale $\Delta t$, as $L \approx 7 \times 10^{38} \epsilon_b^{3/2} \Delta t$ erg, from which we can derive $\epsilon_b^{3/2} \lesssim 10^{-2} (\Delta t/1 \text{ s})$. Assuming this mechanism took place in GW170817, and taking $\epsilon_b = 0.1$ (Horowitz & Kadau 2009), the emission either lasted more than a few seconds or happened below the GBM energy range, i.e., $\sim 10$ keV. There might also be a dependence of the luminosity on the details of the NS EOS, although that is yet to be investigated in detail. Similarly to magnetospheric interaction, however, resonant shattering emission ultimately requires a sufficiently large magnetic field and a simple explanation for the absence of a signal is again the lack of intense magnetic fields prior to merger.

GBM and SPI-ACS observed no temporally extended gamma-ray emission for GRB 170817A. Such emission would be a signature of a long-lived NS remnant powering the SGRB and our flux limits may suggest instead that the remnant is a BH. Metzger et al. (2008) invoke a long-lived millisecond magnetar to explain SGRBs with extended gamma-ray emission (Norris & Bonnell 2006), and millisecond magnetars have also been suggested as possible causes for the plateaus seen in X-ray afterglows of some SGRBs (Rowlinson et al. 2013). The earliest X-ray observation was only performed 50 ks after GRB 170817A (Evans et al. 2017) and hence limits are only set after this time. Future observations may further constrain this scenario, e.g., radio observations on the timescale of a year (Fong et al. 2016).

We encourage the development of quantitative predictions of luminosity as a function of energy, time and physical parameters of the source, as the multiple upcoming joint observations of BNS mergers suggest the possibility of interesting constraints on the pre-merger physics.

### 7. Conclusion

The joint observation of GW170817 and GRB 170817A confirms the association of SGRBs with BNS mergers. With just one joint event, we have set stringent limits on fundamental physics and probed the central engine of SGRBs in ways that have not been possible with EM data alone, demonstrating the importance of multi-messenger astronomy.

The small time offset and independent localizations, though coarse, allowed an unambiguous association of these two events. Because GRB 170817A occurred nearby, an autonomous trigger on-board GBM alerted follow-up observers to the presence of a counterpart to GW170817. At design sensitivity, however, Advanced LIGO and Virgo could in principle detect GW170817 beyond the distance that any active gamma-ray observatory would trigger on a burst like GRB 170817A. Subthreshold searches for SGRBs can extend the gamma-ray horizon and the detection of GRB 170817A provides motivation for further subthreshold search development.

A joint detection at greater distance and for an SGRB with more typical energetics would allow tighter constraints on the temporal offset and the derived inferences. Should NS–BH binaries also be SGRB progenitors, only a joint detection between GW and EM can provide decisive evidence.

In this Letter we propose several explanations for the observed dimness of GRB 170817A. We suggest joint detections should be more common than previously predicted, and future observations of multiple events should enable a study of the populations of mergers and their associated SGRBs, shedding light on the jet geometry and intrinsic brightness distribution. Furthermore, detections with multiple
GW interferometers can provide more stringent constraints on the inclination angles of these systems. The joint detections of SGRBs arising from BNS and NS–BH mergers will constrain the fraction of SGRBs originating from each progenitor class.

The global network of GW detectors and wide-field gamma-ray instruments, such as Fermi–GBM and INTEGRAL/SPI–ACS, are critical to the future of multi-messenger astronomy in the GW era.

We dedicate this Letter to the memory of Neil Gehrels. His pioneering work in gamma-ray astronomy and his vision for multi-messenger astrophysics were instrumental to our discoveries.

The authors thank the referees for their invaluable comments and feedback, especially in a timely manner. The authors gratefully acknowledge the support of the United States National Science Foundation (NSF) for the construction and operation of the LIGO Laboratory and Advanced LIGO as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. The authors gratefully acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French Centre National de la Recherche Scientifique (CNRS) and the Foundation for Fundamental Research on Matter supported by the Netherlands Organisation for Scientific Research, for the construction and operation of the Virgo detector and the creation and support of the EGO consortium. The authors also gratefully acknowledge research support from these agencies as well as by the Council of Scientific and Industrial Research of India, the Department of Science and Technology, India, the Science & Engineering Research Board (SERB), India, the Ministry of Human Resource Development, India, the Spanish Agencia Estatal de Investigación, the Vicepresidência i Conselleria d’Innovació Recerca i Turisme and the Conselleria d’Educació i Universitat del Govern de les Illes Balears, the Conselleria d’Educació Investigació Cultura i Export de la Generalitat Valenciana, the National Science Centre of Poland, the Swiss National Science Foundation (SNSF), the Russian Foundation for Basic Research, the Russian Science Foundation, the European Commission, the European Regional Development Funds (ERDF), the Royal Society, the Scottish Funding Council, the Scottish Universities Physics Alliance, the Hungarian Scientific Research Fund (OTKA), the Lyon Institute of Origins (LIO), the National Research, Development and Innovation Office Hungary (NKFI), the National Research Foundation of Korea, Industry Canada and the Province of Ontario through the Ministry of Economic Development and Innovation, the Natural Science and Engineering Research Council Canada, the Canadian Institute for Advanced Research, the Brazilian Ministry of Science, Technology, Innovations, and Communications, the International Center for Theoretical Physics South American Institute for Fundamental Research (ICTP-SAIFR), the Research Grants Council of Hong Kong, the National Natural Science Foundation of China (NSFC), the Leverhulme Trust, the Research Corporation, the Ministry of Science and Technology (MOST), Taiwan and the Kavli Foundation. The authors gratefully acknowledge the support of the NSF, STFC, MPS, INFN, CNRS and the State of Niedersachsen/Germany for provision of computational resources.

The USRA co-authors gratefully acknowledge NASA funding through contract NNM13AA43C. The UAH co-authors gratefully acknowledge NASA funding from cooperative agreement NNM11AA01A. E.B. and T.D.C. are supported by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center, administered by Universities Space Research Association under contract with NASA. D.K., C.A.W.H., C.M.H., and T.L. gratefully acknowledge NASA funding through the Fermi GBM project. Support for the German contribution to GBM was provided by the Bundesministerium für Bildung und Forschung (BMBF) via the Deutsches Zentrum für Luft und Raumfahrt (DLR) under contract number 50 QV 0301. A.v.K. was supported by the Bundesministeriums für Wirtschaft und Technologie (BMWi) through DLR grant 50 OG 1101. N.C. and J.B. acknowledge support from NSF under grant PHY-1505373. S.M.B. acknowledges support from Science Foundation Ireland under grant 12/1IP/1288.

This work is based on observations with INTEGRAL, an ESA project with instruments and science data center funded by ESA member states (especially the PI countries: Denmark, France, Germany, Italy, Switzerland, Spain), and with the participation of Russia and the USA. The INTEGRAL SPI project has been completed under the responsibility and leadership of CNES. The SPI-ACS detector system has been provided by MPE Garching/Germany. The SPI team is grateful to ASI, CEA, CNES, DLR, ESA, INTA, NASA and OSTC for their support. The Italian INTEGRAL team acknowledges the support of ASI/INAF agreement n. 2013-025-R.1. R.D. and A.v.K. acknowledge the German INTEGRAL support through DLR grant 50 OG 1101. A.L. and R.S. acknowledge the support from the Russian Science Foundation (grant 14-22-00271). A.D. is funded by Spanish MINECO/FEDER grant ESP2015-65712-C5-1-R. Some of the results in this paper have been derived using the HEALPix (Gorski et al. 2005) package. We are grateful VirtualData from LABEX P2IO for enabling access to the StratusLab academic cloud. We acknowledge the continuous support by the INTEGRAL Users Group and the exceptionally efficient support by the teams at ESAC and ESOC for the scheduling of the targeted follow-up observations.

Appendix A
Full Derivation of Photospheric Radius

In order to explain the observed soft BB component, we present a model that simultaneously yields predictions on (1) the time difference between the GW emission and the beginning of the gamma-ray radiation, (2) the estimated temperature of the BB component ($T_{BB} \sim 0.3 ± 1.5$ keV) and (3) its average isotropic luminosity, $L_{iso, BB} = E_{iso, BB}/\Delta t_{BB} = 1.1 \times 10^{46}$ erg s$^{-1}$, where we have used the source interval width of the soft tail spectral fit of $\Delta t_{BB} = 1.15$ s.

Following MR00, our model depends upon three main parameters to provide predictions of the aforementioned three observable quantities. First, the radius from which the fireball
is initiated, which is assumed to be $R_0 = 6GM_{BH}/c^2$ (the innermost stable circular orbit of a Schwarzschild BH with a mass equal to $M_{BH}$). In our case $R_0 = 2.5 \times 10^6$ cm, and $M_{BH} = 2.8 M_\odot$. From numerical models of GRB jets produced in BNS merger remnants (e.g., Aloy et al. 2005), the value of $R_0$ can be associated to the stagnation point of a relativistic outflow, and it is fairly well constrained to be a few gravitational radii of the BH. Second, the initial luminosity of the fireball, $L_0$. This is a free parameter of the model and we note it can be much larger than $L_{iso, BB}$, since the fireball must expand from its initial volume ($\sim R_0^3$) to the size where the photosphere appears (see below). This is also needed, since the observed gamma-ray luminosity, will be a fraction $0 < \epsilon_\gamma \lesssim 1$ of the total (kinetic) luminosity. Finally, the third parameter is the dimensionless entropy of the fireball, $\eta$. For a fireball baryon load $L_0/\eta = L/(M c^2)$. Typical values of $\eta$ are larger than 100 to prevent the compactness problem (Goodman 1986), which is not an issue for GRB 170817A owing to the lack of emission detected above 511 keV. Therefore, our model may allow for values of $\eta$ substantially smaller than $\sim 100$. As we shall see, a combination of $L_0 \approx 10^{50}$ erg and $\eta \approx 18$, results in a viable model to account for the delay of GRB with respect to the GW signal and the average luminosity of the soft BB component.

The initial BB temperature in units of the electron rest mass is given by (MR00, Equation (5))

$$\Theta_0 = \left(\frac{k_B}{m_e c^2}\right) \left(\frac{L_0}{4\pi R_0^2 c a_r}\right)^{1/4} \approx 1.5 \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right)^{1/4} \left(\frac{R_0}{2.5 \times 10^6 \text{ cm}}\right)^{-1/2},$$

(22)

where $a_r = 7.57 \times 10^{-18}$ kg cm$^{-1}$ s$^{-2}$ K$^{-4}$ is the radiation constant. The value of $\Theta_0$ corresponds to a comoving temperature $k_B T_0 \approx 750$ keV. The radius at which the internal energy of the fireball is converted into kinetic energy, i.e., the saturation radius is

$$R_s = \eta R_0 \approx 4.5 \times 10^7 \text{ cm} \left(\frac{\eta}{18}\right) \left(\frac{R_0}{2.5 \times 10^6 \text{ cm}}\right).$$

(23)

The critical baryon load $\eta_b$ below which the photosphere of an expanding fireball happens after the fireball coasts at constant Lorentz factor $\gamma \approx \eta$, i.e., at radii larger than $R_s$, is given by (MR00)

$$\eta_b = \left(\frac{L_0 \sigma_T Y}{4\pi m_p c^3 R_0}\right)^{1/4},$$

(24)

where $Y$ is the number of electrons per baryon, $\sigma_T = 6.65 \times 10^{-25}$ cm$^2$ is the Thomson cross-section. Using fiducial values, we obtain

$$\eta_b \approx 470 \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right)^{1/4} \left(\frac{R_0}{2.5 \times 10^6 \text{ cm}}\right)^{-1/4} \gamma^{1/4}. $$

(25)

In the previous expression, we have taken $Y \approx 1$, which is appropriate once pairs are not present in the system. This is the case for radii larger than $R_p$ (MR00)

$$R_p = R_0 \frac{\Theta_0}{\Theta_p} \approx 1.1 \times 10^8 \text{ cm} \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right)^{1/4} \times \left(\frac{R_0}{2.5 \times 10^6 \text{ cm}}\right)^{1/2} \left(\frac{\Theta_p}{0.03}\right)^{-1},$$

(26)

where the comoving dimensionless temperature below which $e^\pm$ pairs drop out of equilibrium is $\Theta_p \approx 0.03 \approx 17$ keV. Since we have set $\eta \approx \eta_b$, the photosphere will happen at a radius (MR00)

$$R_{ph} = \frac{L \sigma_T Y}{4\pi m_p c^3 \eta^3} \approx 2 \times 10^{13} \text{ cm} \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right) \left(\frac{\eta}{18}\right)^{-3} Y.$$  

(27)

Note that $R_p \lesssim R_{ph}$ for our choice of tunable parameters.

The laboratory frame time needed for the fireball expanding at roughly the speed of light to reach the transparency radius is $t_{ph} \approx R_{ph}/c$, thus,

$$t_{ph} \approx 672 \text{ s} \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right) \left(\frac{\eta}{18}\right)^{-3} Y.$$  

(28)

To compute the time delay between a photon emitted at $R_0$ at $t = 0$ (namely, signaling the GW detection) and another one at $R_{ph}$, we must consider that the fireball begins its expansion from rest, in which case we shall apply the following relation between the arrival time and the time at which the photosphere appears

$$t_a \approx t_{ph} \left(\frac{1}{2\gamma^2} + 1 - \cos \vartheta\right),$$

(29)

where $\vartheta$ is the angle between the radial direction and the line of sight and $\cos \vartheta$ takes values only in the interval $[\gamma/c, 1]$ (Bianco et al. 2001). Using fiducial values for $\eta$, we obtain

$$t_a^{(1)} \approx \frac{t_{ph}}{2\gamma^2} \approx 1.0 \text{ s} \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right) \left(\frac{\eta}{18}\right)^{-5} Y, \text{ for } \cos \vartheta = 1$$

(30)

$$t_a^{(2)} \approx \frac{t_{ph}}{\gamma^2} \approx 2.1 \text{ s} \left(\frac{L_0}{10^{50} \text{ erg s}^{-1}}\right) \left(\frac{\eta}{18}\right)^{-5} Y, \text{ for } \cos \vartheta = v/c.$$  

(31)

These values account for a significant fraction of the time delay between the GW and the prompt radiation in the soft thermal peak.
We can now estimate the comoving temperature of the expanding radius (MR00)

\[ k_B T_{\text{ph}} = k_B T_0 \left( \frac{R_{\text{ph}}}{R_s} \right)^{2/3} \]

\[ \simeq 0.13 \text{ keV} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right)^{-5/12} \left( \frac{R_0}{2.5 \times 10^6 \text{ cm}} \right)^{-5/6} \times \left( \frac{\eta}{18} \right)^{8/3} Y^{-2/3}. \]

This corresponds to an observed temperature, \( T_{\text{ph}}^{\text{obs}} = \eta T_{\text{ph}} \)

\[ k_B T_{\text{ph}}^{\text{obs}} = \eta k_B T_{\text{ph}} \simeq 2.3 \text{ keV} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right)^{-5/12} \times \left( \frac{R_0}{2.5 \times 10^6 \text{ cm}} \right)^{-5/6} \left( \frac{\eta}{18} \right)^{11/3} Y^{-2/3}. \]

The corresponding BB luminosity is (MR00),

\[ L_{\text{ph}} = L_0 \left( \frac{R_{\text{ph}}}{R_s} \right)^{-2/3} \]

\[ \simeq 1.7 \times 10^{46} \text{ erg s}^{-1} \left( \frac{L_0}{10^{50} \text{ erg s}^{-1}} \right)^{7/12} \times \left( \frac{R_0}{2.5 \times 10^6 \text{ cm}} \right)^{-5/6} \left( \frac{\eta}{18} \right)^{8/3} Y^{-2/3}. \]

Thus, with the reference values of our model for \( L_0 \) and \( \eta \), we obtain \( L_{\text{ph}} \simeq L_{\text{iso,BB}} \). On the other hand, \( T_{\text{ph}}^{\text{obs}} \) underpredicts \( T_{\text{BB}}^{\text{obs}} \) by a factor \( \sim 4 \). Nonetheless, we are neglecting Comptonization effects, which may slightly raise the estimated temperature.

**Appendix B**

**List of SGRBs with Associated Redshift**

Table 2 is a list of possible redshifts for GRBs that have been argued to belong to the short class. The asterisks show the “gold sample” selection with standard cuts on duration (\( T_{\text{ph}} < 2 \) s) and localization uncertainty (~arcsecond or better). The others include SGRBs with extended gamma-ray emission, those slightly longer than 2 s that are spectrally hard or show negligible spectral lag, and bursts that are best localized by *Swift* BAT (so a chance association is possible). Most of these redshifts come from Lien et al. (2016); Siellez et al. (2016), and Fong et al. (2015); for the original citations see references therein. They also include bursts from Tunnicliffe et al. (2014). For these bursts, and those best localized by *Swift* BAT, an individual nearby galaxy may be a chance alignment, but it is statistically unlikely that most of them are false associations.

<table>
<thead>
<tr>
<th>GRB</th>
<th>Any Claimed Redshift</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>161104A*</td>
<td>0.788</td>
<td></td>
</tr>
<tr>
<td>160821B+</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>160624A+</td>
<td>0.483</td>
<td></td>
</tr>
<tr>
<td>150423A+</td>
<td>1.394</td>
<td></td>
</tr>
<tr>
<td>150120A+</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>150101B+</td>
<td>0.134</td>
<td></td>
</tr>
<tr>
<td>141212A+</td>
<td>0.596</td>
<td></td>
</tr>
<tr>
<td>140903A+</td>
<td>0.351</td>
<td></td>
</tr>
<tr>
<td>140622A+</td>
<td>0.959</td>
<td></td>
</tr>
<tr>
<td>131004A+</td>
<td>0.717</td>
<td></td>
</tr>
<tr>
<td>130603B+</td>
<td>0.356</td>
<td></td>
</tr>
<tr>
<td>120804A+</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>111117A+</td>
<td>2.211</td>
<td>Updated value from Selsing et al. (2017)</td>
</tr>
<tr>
<td>111020A+</td>
<td>0.019</td>
<td>Tunnicliffe et al. (2014)</td>
</tr>
<tr>
<td>100721A+</td>
<td>0.718</td>
<td></td>
</tr>
<tr>
<td>100724A+</td>
<td>1.288</td>
<td></td>
</tr>
<tr>
<td>100628A+</td>
<td>0.102</td>
<td></td>
</tr>
<tr>
<td>100625A+</td>
<td>0.452</td>
<td></td>
</tr>
<tr>
<td>100206A+</td>
<td>0.407</td>
<td></td>
</tr>
<tr>
<td>100117A+</td>
<td>0.915</td>
<td></td>
</tr>
<tr>
<td>090927</td>
<td>1.37</td>
<td>Arcminute Localization</td>
</tr>
<tr>
<td>090515+</td>
<td>0.403</td>
<td></td>
</tr>
<tr>
<td>090510</td>
<td>0.903</td>
<td><em>Swift</em> BAT ( T_{\text{ph}} &gt; 2 ) s</td>
</tr>
<tr>
<td>090426+</td>
<td>2.609</td>
<td></td>
</tr>
<tr>
<td>090417A</td>
<td>0.088</td>
<td>Arcminute Localization</td>
</tr>
<tr>
<td>080905A+</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>071227</td>
<td>0.381</td>
<td>SGRB w/Extended Emission</td>
</tr>
<tr>
<td>070923</td>
<td>0.076</td>
<td>Arcminute Localization</td>
</tr>
<tr>
<td>070809*</td>
<td>0.473</td>
<td>Tunnicliffe et al. (2014) mention of possible host at ( z = 0.047 )</td>
</tr>
<tr>
<td>070729*</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>070724A+</td>
<td>0.457</td>
<td>SGRB w/Extended Emission</td>
</tr>
<tr>
<td>070714B</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>070429B+</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>061217+</td>
<td>0.827</td>
<td></td>
</tr>
<tr>
<td>061210</td>
<td>0.41</td>
<td>SGRB w/Extended Emission</td>
</tr>
<tr>
<td>061201+</td>
<td>0.111</td>
<td></td>
</tr>
<tr>
<td>061006</td>
<td>0.438</td>
<td>SGRB w/Extended Emission</td>
</tr>
<tr>
<td>060801+</td>
<td>1.13</td>
<td></td>
</tr>
<tr>
<td>060502B+</td>
<td>0.287</td>
<td></td>
</tr>
<tr>
<td>051221A+</td>
<td>0.546</td>
<td></td>
</tr>
<tr>
<td>050906</td>
<td>0.032</td>
<td>Tunnicliffe et al. (2014)</td>
</tr>
<tr>
<td>050813+</td>
<td>0.72/1.8</td>
<td>It is unclear which redshift is correct</td>
</tr>
<tr>
<td>050724A</td>
<td>0.257</td>
<td>SGRB w/Extended Emission</td>
</tr>
<tr>
<td>050709</td>
<td>0.161</td>
<td>Not detected by <em>Swift</em> BAT</td>
</tr>
<tr>
<td>050509B+</td>
<td>0.225</td>
<td></td>
</tr>
</tbody>
</table>

**References**


Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, *CQG* *e*, 32, 024001
