The origin of gamma-ray bursts (GRBs) is one of the most interesting puzzles in recent astronomy. During the last decade a consensus has formed that long GRBs (LGRBs) arise from the collapse of massive stars, and that short GRBs (SGRBs) have a different origin, most likely neutron star mergers. A key ingredient of the collapsar model that explains how the collapse of massive stars produces a GRB is the emergence of a relativistic jet that penetrates the stellar envelope. The condition that the emerging jet penetrates the envelope imposes strong constraints on the system. Using these constraints we show the following. (i) Low-luminosity GRBs (llGRBs), a subpopulation of GRBs with very low luminosities (and other peculiar properties: single-peaked, smooth and soft), cannot be formed by collapsars. llGRBs must have a different origin (most likely a shock breakout). (ii) On the other hand, regular LGRBs must be formed by collapsars. (iii) While for BATSE the dividing line between collapsars and non-collapsars is indeed at approximately 2 s, the dividing line is different for other GRB detectors. In particular, most Swift bursts longer than 0.8 s are of a collapsar origin. This last result requires a revision of many conclusions concerning the origin of Swift SGRBs, which were based on the commonly used 2 s limit.

1. Introduction

Gamma-ray bursts (GRBs) are among the most amazing transients known. In a few seconds, a GRB emits the energy that a star like our Sun emits in its whole lifetime.
Their origin has puzzled astronomers since their serendipitous discovery in the late 1960s. After two decades in which it was believed that GRBs are Galactic, it was realized, in the early 1990s, that they have a cosmological origin [1–3]. The distance scale immediately set the energy scale to be $\gtrsim 10^{51}$ erg (including beaming corrections that were realized towards the end of the 1990s [4–6]). Together with the short time scale, this has led inevitably to the conclusion that the events involve the formation of a newborn compact object, most likely a black hole. This conclusion has left practically just two progenitor candidates: a collapsing massive star or the merger of two neutron stars (or a neutron star and a black hole).

The observations of a few (long) GRB afterglows in 1997 revealed that those bursts arose in star-forming regions. Paczynski [7], who noticed that, quickly suggested that long GRBs (LGRBs) are related to core collapse events. At roughly the same time MacFadyen & Woosley [8] suggested the collapsar model. According to this model, a GRB is produced by a relativistic jet that emerges from the centre of a massive collapsing star and penetrates the stellar envelope. Now the name collapsar is used with different variations.\footnote{In some cases the term collapsar is used generically for any model that involves a collapsing star, regardless whether there is a jet or not. In other cases, it is used more restrictively to imply a situation in which the collapsing star produces an accreting black hole as a central engine that drives a relativistic jet.}

We stress that here we use this original definition for a collapsar: a jet that penetrates the envelope of a collapsing star. Using numerical simulations, MacFadyen & Woosley [8] demonstrated that a relativistic jet can indeed penetrate a stellar envelope. Again, roughly at the same time, Galama et al. [9] discovered that GRB 980425 was associated with the powerful type Ic supernova SN 1998bw. However, GRB 940425 was a strange GRB. It was very weak, with an energy a few orders of magnitude less than the energy of a typical GRB. Additionally, it was single-peaked and smooth, and it had a very soft spectrum. It was not clear that this was a regular GRB, and hence the association of GRB 980425 with SN 1998bw was not sufficient to demonstrate a GRB–SN association. Shortly after that, Bloom et al. [10] and others discovered red bumps in the afterglows of more distant regular GRBs. These red bumps were interpreted as the signatures of 1998bw-like SN, supporting the GRB–SN association. However, the evidence for a GRB–SN association was inconclusive until SN 2003dh was discovered in association with the regular GRB 030329 [11,12]. Since then a few other GRB–SN associations have been discovered. Even though most of these GRB–SN associations are with weak, smooth and single-peaked GRBs\footnote{Another association of a regular regular GRB and an SN, GRB 101219B and SN 2010ma, was discovered recently.} this is generally considered as a ‘proof’ of the collapsar model for LGRBs.

An inspection of BATSE’s (Burst and Transient Source Experiment) GRB temporal distribution revealed [13] two groups: short ($T_{90} < 2$ s) and long ($T_{90} > 2$ s). Already in 1995, it was pointed out [14,15] that the two groups have a different spatial distribution. The observed short GRBs (SGRBs) are significantly nearer (and weaker). This suggested the possibility of different physical origin for the two populations. As it takes time (and energy) to cross the relatively large stellar envelope, it was argued that SGRBs cannot be produced by collapsars [16]. In most cases, collapsars produce LGRBs, but by now we know that in some cases collapsars produce SGRBs (see §4 below). However, a variant on this original argument that we discuss here (in §§3 and 4 below) shows that most SGRBs cannot be produced by collapsars. Lack of detection of SGRB afterglows left the situation inconclusive until 2005, when Swift localized the first short bursts and the first SGRB afterglows were detected. It turned out that SGRBs are not associated with star-forming regions (some arise in elliptical galaxies), and as such they are not associated with the deaths of massive stars. The progenitors could be neutron star mergers (as suggested already in 1989 [17]). However, as yet there is no conclusive demonstration of this origin [18].

We describe here new results, derived by Bromberg et al. [19–22], concerning the nature of GRB progenitors. We briefly discuss, in §2, some recent analytical results concerning relativistic jet penetration through the stellar envelope [19]. We then consider their implications on this picture. In §3, we demonstrate that GRBs, those that appear in most GRB–SN associations, cannot be produced by collapsars [20]. Whereas this weakens the case for the association of regular LGRBs with SNe, we show in §4 that, when combined with the GRB temporal distribution, these
considerations demonstrate that the LGRBs originate from collapsars [21], providing a direct observational indication for jets that puncture the stellar envelope. Further inspection of the temporal distribution enables us to estimate (in §5), for the first time, the fraction of collapsars among SGRBs as a function of the observed duration [22]. We show that this fraction depends strongly on the detector (in particular, on its spectral window). In particular, the standard limit of 2 s is invalid for Swift’s observations, for which a limit of 0.8 s is much more appropriate.

2. Jet propagation

A schematic picture of a relativistic jet propagating within a stellar envelope is depicted in figure 1. There are a few critical components. A double shock system appears at the head of the jet [16]. While the jet is highly relativistic, these shocks slow down the head and it typically propagates with a sub- or mildly relativistic velocity. The hot material that streams sideways out of the jet’s head produces a cocoon that engulfs the jet. While expanding sideways into the rest of the stellar envelope (this expansion will eventually blow out a fraction of the stellar envelope), it also squeezes the jet and produces a (radiative) collimation shock within the jet [24].

As long as the jet is within the stellar atmosphere, all its energy is dissipated at the jet’s head. The total dissipated energy equals, therefore, the jet’s luminosity multiplied by the time it takes to cross the envelope. Because the inner engine is much smaller than the envelope, it is decoupled from the jet that crosses the envelope on a much larger scale, and one can expect that the luminosities before and after the jet breaks out are comparable. Using the observed GRB luminosity to estimate the jet power before breakout, we can estimate the duration of the dissipation phase as [19]

$$t_B \simeq 15 \times \left( \frac{L_{iso}}{10^{51} \text{ erg s}^{-1}} \right)^{-1/3} \left( \frac{\theta}{10^\circ} \right)^{2/3} \left( \frac{R_\ast}{5 R_\odot} \right)^{2/3} \left( \frac{M_\ast}{15 M_\odot} \right)^{1/3}.$$ (2.1)
0.01 0.1 1 10
0 5 10 15 20 25 30

Figure 2. The distribution of $T_{90}/t_B$ for LGRBs (blue dotted line), low-luminosity GRBs (llGRBs, red solid line) and SGRBs (green dashed line) (adapted from Bromberg et al. [20]). (Online version in colour.)

Here $L_{\text{iso}}$ is the isotropic equivalent jet luminosity, $\theta$ is the jet half opening angle and we have used typical values for an LGRB; $R_*$ and $M_*$ are the radius and the mass of the progenitor star, where we have normalized their value according to the typical radius and mass inferred from observations of the few SNe associated with LGRBs. For the jet to break out, the central engine must continue operating for a duration longer than $t_B$. If the inner engine stops before the jet’s head crosses the envelope, the jet would not produce a regular GRB.

3. Low-luminosity GRBs

The duration of the prompt emission, approximated by $T_{90}$, is given simply by

$$T_{90} = t_e - t_B,$$  \hspace{1cm} (3.1)

where $t_e$ is the total time that the engine powering the jet is active. Within the collapsar model, without fine tuning, only a small fraction of the bursts should have $T_{90}/t_B \ll 1$ (see §4). Namely, it is unlikely that the engine operates just long enough to let the jet break out of the star and then stops right after breakout. This argument was used by Matzner [16] to argue that collapsars cannot produce SGRBs, for which $T_{90}/t_B \ll 1$. This is indeed confirmed in figure 2, in which the distribution of $T_{90}/t_B$ is shown for both LGRBs and SGRBs. One can clearly see two distinct populations: LGRBs, for the majority of which $T_{90}/t_B > 1$; and SGRBs, all of which satisfy $T_{90}/t_B < 1$.

Figure 2 depicts also a third group of GRBs, low-luminosity GRBs (llGRBs). Like SGRBs, the observed duration distribution of llGRBs is inconsistent with the predictions of the collapsar model. In particular, a large fraction of llGRBs have $T_{90}/t_B \ll 1$. The probability that the observed llGRB $T_{90}/t_B$ distribution is consistent with the LGRB distribution is smaller than 5 per cent, implying that llGRBs cannot be generated by collapsars and they must have a different origin.

llGRBs are a group of six GRBs, whose luminosities are around $10^{47}$–$10^{49}$ erg s$^{-1}$, at least two orders of magnitude below the average luminosity of a typical GRB. Remarkably, llGRBs are not characterized just by their low luminosity. They are also single-peaked, smooth and soft. llGRBs include GRB 9890425 (the first GRB detected to accompany a supernova—SN 1998bw) as well as a few other GRB–SN pairs: GRB 031203/SN 2003lw; GRB 060218/SN 2006aj;
GRB 100316D/SN 2010bh. GRB 051109B shows all the common features of llGRBs, but it lacks a reported SN. It is associated with a star-forming region in a spiral galaxy at \( z = 0.08 \) [25]. All llGRBs are at very low redshifts. With such low luminosities, they could not have been detected from further out. While only six llGRBs have been observed so far, given these distances, the llGRB inferred rate per unit volume is much larger than the rate of regular LGRBs [26]. In fact, this rate is so large that llGRBs cannot be significantly beamed, as, even with modest beaming corrections, the rate would exceed the rates of their associated SNe—broad line type Ibc.

An interesting and likely possibility is that llGRB jets are weak and fail to break out from their progenitors. A ‘failed jet’ dissipates all its energy into the surrounding cocoon and drives its expansion. As the cocoon reaches the edge of the star, its forward shock may become mildly or even ultra-relativistic, emitting the \( \gamma \)-rays observed in llGRBs when it breaks out. This idea that llGRBs arise from shock breakouts was suggested shortly following the observations of GRB 980425/SN 1998bw [27–29]. It drew much more attention following the observation of additional llGRBs with similar properties, and especially with the observation of a thermal component in the spectrum of llGRB 060218 [30–32]. Yet, it was hard to explain how shock breakout releases enough energy in the form of \( \gamma \)-rays. Katz et al. [33] realized that the deviation of the breakout radiation from thermal equilibrium provides a natural explanation for the observed \( \gamma \)-rays. More recently, Nakar & Sari [34] calculated the emission from mildly and ultra-relativistic shock breakouts, including the post-breakout dynamics and gas–radiation coupling. They find that the total energy, spectral peak and duration of all llGRBs can be well explained by relativistic shock breakouts. Moreover, they find that such breakouts must satisfy a specific relation between the observed total energy, spectral peak and duration, and that all observed llGRBs satisfy this relation. These results lend strong support to the idea that llGRBs are relativistic shock breakouts. From a historical point of view, this understanding closes the loop with Colgate’s [35] original idea, which preceded the detection of GRBs, that an SN shock breakout will produce a GRB.

As we discuss in the following section, the observed GRB duration distribution indicates the existence of many ‘failed jets’ in which the engine time is shorter than the breakout time. This is consistent with the observations that the rate of llGRBs is much higher than the rate of regular LGRBs.

4. Long GRBs and collapsars

As most of the GRBs associated with SNe are llGRBs, one might think at first that this new understanding rules out the collapsar model for LGRBs. However, on the contrary, these arguments provide a new and unexpected direct observational confirmation of the collapsar origin of LGRBs. Consider again equation (3.1). Under very general conditions, this equation results in a flat duration distribution for durations significantly shorter than the typical breakout time.

It follows from equation (3.1) that the distribution, \( p_\gamma(T_{90}) \), of the observed GRB durations is a convolution of \( p_e(t_e) \), the distribution of engine operating times, and \( p_B(t_B) \), the distribution of jet breakout times. Under quite general conditions (more specifically, unless \( p_e \) varies very rapidly around \( t_B \), an unlikely situation), the following limits hold:

\[
p_\gamma(T_{90}) \approx \begin{cases} 
p_e(t_B), & T_{90} \lesssim t_B, \\
p_e(T_{90}), & T_{90} \gg t_B. 
\end{cases} \tag{4.1}
\]

Particularly interesting for our purpose here is the appearance at short durations \( T_{90} \lesssim t_B \) of a flat region, a plateau, in which the rate of events is independent of the duration. Remarkably, such plateaus exist in all the observed GRB duration distributions (figure 3). They were not noticed before because the ‘canonical’ distribution plot [13] depicts \( dN/d\log(T) \) instead of \( dN/dT \).
Figure 3. The duration distributions, $dN/dT_{90}$, of BATSE (A, red), Swift (C, blue) and Fermi GBM (D, green) GRBs. Also plotted is the distribution of the soft (hardness ratio < 2.6) BATSE bursts (B, magenta). For clarity, the Swift values are divided by a factor of 5 and the Fermi GBM by 15. Note that the quantity $dN/dT$ is depicted and not $dN/d\log T$ as traditionally shown in such plots [13]. The black lines show the best fitted flat interval in each dataset: 5–25 s (BATSE), 0.7–21 s (Swift) and 2.5–31 s (Fermi). The upper limit of this range indicates a typical breakout time of a few dozen seconds, in agreement with the prediction of the collapsar model. Soft BATSE bursts show a considerably longer plateau (0.4–25 s), indicating that most of the soft short bursts are in fact collapsars (adapted from Bromberg et al. [21]). (Online version in colour.)

The higher end of the plateau enables us to estimate $t_B$ and from this to infer some of the basic properties of the collapsing stars. We find, for example, that a typical progenitor size is $\sim 5 R_\odot$. Another interesting feature seen in figure 3 is the rapid decline at durations longer than $t_B$. In this regime, according to equation (4.1), the distribution is dominated by $p_e(t_e)$, thus $p_e(T_{90}) \approx p_g(T_{90})$. An extrapolation of this distribution to shorter engine operating times suggests that there are numerous cases in which $t_e < t_B$, and the jet fails to break out. This is in a very nice agreement with the very large inferred event rate of LGRBs, if these are interpreted as ‘failed jets’.

The appearance of these plateaus (and their dependence on the observed hardness) is the first direct observational confirmation of the collapsar model. More specifically, these plateaus demonstrate the fact that the duration of an LGRB is the difference between two independent time scales. This confirms a basic prediction of the collapsar model: the overall duration is the difference between the time that the engine operates and the time it takes the jet to penetrate the stellar envelope.

At very short durations, the plateau does not extend all the way to zero. This does not rule out the model. In this regime, non-collapsar SGRBs, which have a different origin, appear and dominate the distribution. As different detectors have different relative sensitivities to long (and soft) versus short (and hard) GRBs, the duration at which short non-collapsars begin to dominate varies from one detector to another. To demonstrate this dependence on the detector’s spectral window, we artificially change BATSE’s effectiveness for detection of hard SGRBs by considering only softer BATSE bursts (hardness ratio < 2.6). As expected, for this softer BATSE sample, the non-collapsar peak shrinks and the plateau extends down to shorter durations.

In principle we should have worked with a redshift-corrected sample. However, such samples of SGRBs are too small and we are forced to use the observed durations. This implies that we have to worry also about time dilation, which in turn depends on the sensitivity of the detector and the corresponding depth of the samples. However, as SGRBs are detected only from relatively small redshifts, this time dilation correction is not significant.
Figure 4. (a) BATSE, (b) Swift and (c) Fermi GBM fractions, $f_{\text{NC}}$ (solid red line), of non-collapsars from the total number of observed GRBs as a function of the observed duration, $T_{90}$. The blue shaded regions represent 67% confidence limits of $f_{\text{NC}}$. Also plotted red as vertical lines are the $T_{90}$ values for which $f_{\text{NC}} = 0.5$ (adapted from Bromberg et al. [22]). (Online version in colour.)

5. Short non-collapsar GRBs

Shortly after Kouveliotou et al. [13] demonstrated that there are two populations of GRBs, long and short ones, it became clear that these two populations have different spatial distributions and a different origin [14,15]. Now we know that LGRBs arise from collapsars. SGRBs have other progenitors, most likely neutron star mergers. As we are still uncertain concerning the origin of SGRBs, we denote them here as non-collapsars. So far it was implicitly assumed that the division line between long and short GRBs is at 2 s regardless of the observing satellite. The existence of a plateau in the observed LGRB (of collapsar origin) duration distribution enables us to determine, for the first time, the fractions of collapsars versus non-collapsars as a function of the observed time for every specific detector. While we cannot determine if a specific burst is a collapsar or not, we can now give a probabilistic estimate for a given duration and hardness.

The basic idea is very simple. For a given detector, we determine the rate of detection of bursts within the plateau. This provides an estimate for the detection rate of short-duration collapsars by this detector. Now, we can compare this rate with the rate of SGRBs at any given duration, and obtain the collapsar and non-collapsar fractions as a function of duration. This estimate is performed for different detectors or even for different detection windows (hardness) for a specific detector.

We have fitted the different duration distributions with a plateau (representing collapsars) and a lognormal distribution (for non-collapsars). The fit is remarkably good and it enables us to estimate the fraction, $f_{\text{NC}}$, of non-collapsars from the total number of observed GRBs as a function of the observed duration, $T_{90}$ (figure 4). For BATSE, $T_{90} < 2$ s is a reasonable threshold to identify non-collapsars. This limit results in a probability > 70% for a correct classification for BATSE.
bursts. However, this condition is misleading for Swift bursts. At $T_{90} = 2\text{ s}$, a Swift burst has $84 \pm 14\%$ probability to be a collapsar! Clearly, for Swift, a 2 s division line results in a large number of misidentified collapsars as non-collapsars. We propose to draw the division line between collapsars and non-collapsars at the duration where the probability that a GRB is a non-collapsar is 50 per cent. With this condition, a BATSE GRB can be classified as a non-collapsar if its $T_{90} < 3.1 \pm 0.5\text{ s}$. Swift bursts can be identified as non-collapsars only if their duration $T_{90} < 0.8 \pm 0.3\text{ s}$, while the corresponding limit for the Fermi GBM (Gamma-ray Burst Monitor) is $T_{90} < 1.7 \pm 0.5\text{ s}$. The results shown here can be expanded and improved when we consider the hardness of the bursts.

Using these results, one can go back and examine various studies of short bursts that attempted to compare various features of short (standing for non-collapsars) with long (collapsars) and check the samples used. A preliminary inspection of such studies reveals that, while in some cases the short sample has only a small number of potential collapsars, in other cases the short sample used was heavily contaminated by high-probability potential collapsars (with observed duration shorter than 2 s), and it is possible and even likely that these have dominated the results. Note in particular that our results are not inconsistent with those of Berger [36], who finds a larger dispersion in properties of Swift SGRB (with $T_{20} < 2\text{ s}$) host galaxies and the positions of the bursts within the hosts, as compared with the more homogeneous properties of LGRB hosts and their position within the hosts. A mixed sample that contains both collapsars and non-collapsars is expected to show such larger dispersion (see [22] for further details).

6. Conclusions

To conclude, we summarize our basic findings. We define a collapsar as a collapsing massive star that produces in it centre a relativistic jet. The jet penetrates the stellar envelope and produces the GRB once it has left the star. The duration of the envelope penetration phase depends on the jet’s luminosity, its opening angle and the size and density profile of the stellar envelope.

One can expect that the duration of a burst is typically comparable with or longer than the jet breakout time. Indeed, a comparison of the estimated jet breakout time of a typical LGRB shows that it is shorter than the observed duration of the burst. On the other hand, the jet breakout time is much longer than the duration of a short burst. This provides the first indication that SGRBs are not produced by collapsars. We have shown that a third group of low-luminosity GRBs also do not satisfy this condition. This implies that I/GRBs do not arise from collapsars.

Within the collapsar model, the observed duration of a GRB is the difference between the time its central engine operates, producing the jet, and the jet’s breakout time. This directly implies that at short durations the rate of bursts produced by collapsars should be independent of their duration. We have shown that such a behaviour is observed in the duration distributions of all GRB satellites/detectors: BATSE, Swift and GBM. This provides direct observational confirmation of the basic prediction of the collapsar model, and as such demonstrates the collapsar origin of LGRBs.

This last feature also enables us to determine the fraction of collapsars within the observed SGRBs. This fraction depends on the characteristics of the detector. For BATSE the standard division between collapsar and non-collapsar is indeed at $\sim 2\text{ s}$. However, for the softer Swift, many bursts shorter than 2 s are of collapsar origin. This might have led to some confusion in the past in interpreting observations of these short bursts as indications for properties of non-collapsar GRBs.

This research was supported by an Advanced ERC grant and by the Israeli Center for Excellence for High Energy Astrophysics (to T.P.), by an ERC starting grant and an ISF grant (to E.N.) and by Packard, Guggenheim and Radcliffe fellowships (to R.S.).

Note that the term collapsars has also been used in both wider and narrower contexts.
References


