Classification of BATSE, Swift, and Fermi Gamma-Ray Bursts from Prompt Emission Alone

Steinhardt, Charles L.; Mann, William J.; Rusakov, Vadim; Jespersen, Christian K.

Published in:
Astrophysical Journal

DOI:
10.3847/1538-4357/acb999

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Classification of BATSE, Swift, and Fermi Gamma-Ray Bursts from Prompt Emission Alone

Charles L. Steinhardt1,2, William J. Mann1,2, Vadim Rusakov1,2, and Christian K. Jespersen3

1 Cosmic Dawn Center (DAWN), Denmark; steinhardt@nbi.ku.dk
2 Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen Ø, Denmark
3 Department of Astrophysical Sciences, Princeton University, Princeton, NJ, 08544, USA

Received 2022 August 1; revised 2022 December 23; accepted 2023 January 1; published 2023 March 8

Abstract

Although it is generally assumed that there are two dominant classes of gamma-ray bursts (GRBs) with different typical durations, it has been difficult to classify GRBs unambiguously as short or long from summary properties such as duration, spectral hardness, and spectral lag. Recent work used t-distributed stochastic neighborhood embedding (t-SNE), a machine-learning algorithm for dimensionality reduction, to classify all Swift GRBs as short or long. Here, the method is expanded, using two algorithms, t-SNE and UMAP, to produce embeddings that are used to provide a classification for 1911 BATSE bursts, 1321 Swift bursts, and 2294 Fermi bursts for which both spectra and metadata are available. Although the embeddings appear to produce a clear separation of each catalog into short and long bursts, a resampling-based approach is used to show that a small fraction of bursts cannot be robustly classified. Further, three of the 304 bursts observed by both Swift and Fermi have robust but conflicting classifications. A likely interpretation is that in addition to the two predominant classes of GRBs, there are additional, uncommon types of bursts which may require multiwavelength observations in order to separate them from more typical short and long GRBs.

Unified Astronomy Thesaurus concepts: Gamma-ray bursters (1878); Gamma-ray bursts (629); Dimensionality reduction (1943); Classification systems (253)

Supporting material: machine-readable table

1. Introduction

The most prominent feature of a $\gamma$-ray burst (GRB) is the short duration of its prompt emission, ranging from $\sim 10^{-2}$ to $\sim 10^3$ s. The distribution of observed durations is predominantly bimodal, leading to a standard division of GRBs into short and long bursts (see Kouveliotou et al. 1993). These two groups have been hypothesized to have distinct astrophysical origins, with short bursts associated with the mergers of neutron stars (Berger et al. 2013; Tanvir et al. 2013; Ghirlanda et al. 2018) and long bursts associated with core collapses of massive stars (Hjorth et al. 2003; Stanek et al. 2003). This would imply that it should be possible to separate the two types of bursts cleanly based on their observed properties.

However, there is considerable overlap between the two distributions in duration. Thus, the standard dividing line at $T_{90} = 2$ s will misclassify the shortest long bursts and the longest short bursts (Tavani et al. 1998; Paciesas et al. 1999). The most prominent additional observed features, spectral hardness (Kouveliotou et al. 1993) or spectral lag (Norris et al. 1986; Norris & Bonnell 2006) do not provide a clean separation between short and long bursts. Environment has also been considered as a possible distinguishing factor (Leśniewska et al. 2022). More complex derived properties, or combinations of the standard summary properties, have also been unable to provide a complete separation (Fruchter et al. 2006; Nakar 2007; Bromberg et al. 2011; Zhang et al. 2012; Bromberg et al. 2013).

Recently, a new approach using the dimensionality reduction algorithm t-distributed stochastic neighborhood embedding (t-SNE) used the full Swift light curves to provide the first clear separation between short and long bursts (Jespersen et al. 2020). Although the same approach should be expected to separate bursts from the BATSE and Fermi catalogs successfully as well, the differences between these data sets require individualized tuning and the data sets cannot be combined into one t-SNE map without significant information loss. Thus, although this work uses very similar methodology to the Jespersen et al. (2020) pilot study, the methodology applied to the other two catalogs is necessarily slightly distinct.

This work expands upon that study in several significant ways:

1. Bursts from BATSE and Fermi are also mapped, providing a complete t-SNE-based classification catalog combining all large data sets.
2. It is shown that the three catalogs yield similar separations, implying that the separation is truly astrophysical rather than due to selection effects, data reduction choices, or other systematics.
3. An additional algorithm, uniform manifold approximation and projection (UMAP; McInnes et al. 2018) is considered as an alternative to t-SNE. UMAP and t-SNE are both dimensionality algorithms, and often produce similar results when well tuned (Becht et al. 2019; Kobak & Berens 2019; Xiang et al. 2021). However, for most large, high-dimensionality data sets, UMAP produces these results in significantly shorter computation times (Hu et al. 2019). \(^4\)

4. A comparison of the classifications of bursts which appear in both the Swift and Fermi data sets is used both as a consistency check and to determine whether additional subclassifications are suggested by the t-SNE and UMAP maps.

5. A new measure of uncertainty is introduced to describe the stability of these classifications.

In Section 2, the methods used in Jespersen et al. (2020) are reviewed, modified as needed, and applied to the BATSE and Fermi catalogs. Corresponding choices for UMAP are also discussed. The resulting classifications are presented in Section 3. In Section 4, the robustness of this classification is evaluated using a combination of multiple catalogs. The results and implications for future surveys are discussed in Section 5. A full catalog including classifications for objects in BATSE, Fermi, and Swift is included, as described in the Appendix.

2. Methodology

The methodology used in this work follows the general approach used in Jespersen et al. (2020)’s analysis of the Swift data set, modified in order to accommodate the differences between various GRB observatories. Dimensionality reduction is applied to the full set of light curves in every observed band. The resulting embedding is examined, and divided into cleanly separated structures. The objects in each structure are then considered to comprise a distinct class of GRBs.

In practice, there are several additional steps. First, the data must be standardized, removing incomplete or missing observations. Then, preprocessing is performed to remove or negate irrelevant information that these algorithms might otherwise interpret as meaningful. Afterwards, dimensionality algorithms can be applied, and that application requires the choice of several hyperparameters which must be selected individually for each data set. Finally, it is necessary to determine which structures in the resulting embedding should be considered distinct. Within the same approach, each of these steps requires different choices for Swift, BATSE, and Fermi. These are described in more detail in the subsections below.

It is important to note that “unsupervised” algorithms such as the ones used here are in practice strongly dependent on the choice of hyperparameters. The proper interpretation of the embeddings presented here should not be that the hyperparameters we have chosen are correct and others are incorrect. Rather, every choice of hyperparameters leads to a valid embedding, which contains potentially useful information about the distribution but is always incomplete. In that sense, it would be more like considering various projections. Unlike projections, however, there is no rigorous formalism such as principal component analysis for determining which will be most useful.

Here, the choices rely on the physical assumption that GRBs can be divided into discrete groups due to distinct progenitors, but avoid asserting any specific number of progenitors. Rather, the hyperparameters which produce the cleanest separations are selected. In that respect, the separation into two primary groups is a property of the data sets. However, because these embeddings focus on groups of a specific size, additional progenitors which are less common and thus have significantly fewer examples will not be revealed in these maps.

Embeddings focusing on small groups were considered in Jespersen et al. (2020), but did not produce easily interpretable, well-separated groups. Further, as described in Section 4.1, the smaller substructures hinted at by the embeddings considered here do not appear to be meaningful. However, given the large space of possible hyperparameters, of which only a small portion has been sampled in this work, it is likely that additional, astrophysically interpretable groups could exist. Identifying how many and which of these groups are meaningful might require additional observational information.

2.1. Burst Selection

One of the restrictions of t-SNE and UMAP is that they cannot be applied to data of different dimensions or labels. This is one of the reasons that Swift, BATSE, and Fermi are examined individually rather than as a group; the different bands and cadences of each set of observations do not allow a direct comparison. Here, bursts for which at least one band is predominantly missing or key metadata do not exist are rejected entirely. More minor flaws are generally adjusted or corrected instead of rejecting the burst. The specific choices for each individual data set are included for reproducibility.

For BATSE (Meegan 1997), the ASCII file is used as the canonical source of each light curve, due to the similar tte_bfits files having various data problems. Out of 2702 total bursts, 527 bursts with missing ASCII files on the HEASARC server were therefore rejected. Summary data including flux, fluence, and $T_{90}$ were taken from https://batse.msfc.nasa.gov/batse/grb/catalog/current/. An additional 264 bursts with missing summary data were rejected. In total, 791 bursts were rejected and the remaining 1911 were included.

For Swift (Lien et al. 2016), the same choices were made as in Jespersen et al. (2020). This work uses a more recent version of the catalog, which includes an additional 67 bursts.

For Fermi (von Kienlin et al. 2020), the bcat file is used as the canonical source of each light curve. All but 34 bursts available at https://heasarc.gsfc.nasa.gov/FTP/fermi/data/gbm/bursts/ had a bcat file available, so 3108 were downloaded. The Fermi GBM Burst Catalog includes multiple models, and the best-fit average flux was used for each burst. These fluxes and $T_{90}$ were taken from https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html. Unfortunately, nearly all of the bursts recorded after mid-2018 had not undergone spectral analysis, therefore lacking best-fit flux and fluence models. These 814 bursts were cut, leaving 2294 remaining for analysis.

A potential concern is that the use of deconvolved flux light curves for Fermi but count rate light curves for the other observatories makes the three no longer directly comparable. However, this was already the case because, e.g., Swift and Fermi provide different bands and thus different information about each burst. Using different data types can be thought of as simply another difference between the data sets and pipelines. As shown in Figure 4, there is still strong agreement between the Swift and Fermi classifications, indicating that all of these differences between observatories, reductions, and choice of light curves do not significantly alter the conclusions presented here about the broad classifications of GRBs.

5 Or, at least, the distance metric selected must make a choice about how to handle the difference between missing and measured dimensions in a consistent way, which is typically very difficult unless the correct answer is already known.

6 https://heasarc.gsfc.nasa.gov/FTP/compton/data/batse/ascii_data/64ms/
2.2. Preprocessing

Several preprocessing steps are required before dimensionality reduction algorithms can be used. Because these algorithms require identical formats and have difficulty handling missing information, light curves are padded with additional zeros to produce data of identical lengths as in Jespersen et al. (2020). The goal of additional preprocessing is to remove extraneous information which might otherwise contribute to the postembedding positions of light curves while retaining as much information as possible. There are two key extraneous parameters: the overall brightness of the burst and the trigger time.

Although the energy carried by a burst is physically meaningful, for most GRBs in the catalog, the redshift is unknown. Under the assumption that bursts with the same underlying astrophysical origins can exist across a wide range of redshifts, the brightness will be a poor indicator of luminosity. Therefore, each burst is normalized by dividing by the total fluence.\(^7\) This retains hardness information because the relative brightness between bands is preserved.

The other issue is handling trigger time offsets, where the time that the burst was detected differs from the actual start of the burst. These offsets are typically due to instrumentation rather than due to the shape of the burst itself, and therefore should be discarded. To accomplish this, the same procedure is used as in Jespersen et al. (2020). A discrete-time Fourier transformation (DTFT) is performed on a concatenation of the light curves in all observed bands. An overall time shift will only change the phase information under this DTFT. However, relative time offsets between different bands, as well as other meaningful information such as duration, hardness, and spectral lag will all contribute to the amplitudes as well. Therefore, the phase information is then discarded, and dimensionality reduction algorithms are run only on the amplitudes.

2.3. Embedding

Dimensionality reduction algorithms are then run on each of the data sets independently. It is necessary to perform different embeddings for Swift, BATSE, and Fermi, since they measure different bands with different cadences and report data in different formats. Two main algorithms are used here: t-SNE and UMAP. Both have hyperparameters which must be tuned for each individual data set. The primary hyperparameter of importance for t-SNE is perplexity. For UMAP, there are two hyperparameters which must be tuned, n_neighbors and set_op_mix_ratio.

In every case, a range of embeddings with different properties will result from various choices of the hyperparameters. Unfortunately, there is no rigorous mathematical formalism for choosing optimal hyperparameters. The multiple embeddings which can result from different choices are in some sense all correct and valid, yet simultaneously are all incomplete descriptions of the full structure. Naturally, some will be more useful for GRB classification. For each data set, both t-SNE and UMAP were run with a variety of hyperparameters, and the ones which produced embeddings with the cleanest separations were chosen. The hyperparameters chosen are summarized in Table 1.

![Figure 1. t-SNE (left) and UMAP (right) embeddings of 1321 Swift lightcurves, colored by classification. The duration distributions (bottom) are consistent with an interpretation as separation into short (orange) and long (purple) GRBs rather than merely a classification purely by duration. The two embeddings agree on the classification of all but one burst (cyan). Three bursts are classified differently in this work than in Jespersen et al. (2020) (black).](image)

Although these embeddings often show clear separations between short and long GRB, that does not guarantee that there are only those two classes of GRBs. The perplexity and n_neighbors hyperparameters for t-SNE and UMAP, respectively, essentially dictate the size of the groups which the embeddings focus on representing properly. Thus, if there are two predominant groups and one or more tiny ones, the tiny groups might be attached to a larger one. An attempt to identify bursts which might not be standard short or longer GRBs is described in Section 4.2.

3. Classification

For each catalog, two embeddings are produced, one with t-SNE and the other with UMAP, for a total of six maps. In each map, a division is made into two large groups, labeled short and long based on the typical duration in each group. A small fraction of bursts are classified either as outliers, distinct from both groups, or as ambiguous, lying in the region between the short and long groups. The BATSE map is more complicated, producing an initial separation which produces a duration distribution qualitatively different than the other two catalogs. A more quantitative comparison is performed in Section 4.

3.1. Swift

Since the technique used in this work is very similar to that in Jespersen et al. (2020), the embeddings and classifications produced using the Swift catalog are nearly identical. There is a clear separation into two groups, one identified as short (orange, Figure 1) and the other as long (purple).

---

\(^7\) For Fermi, we instead normalized by the average flux, which we found to work better on the semiprocessed *bcat* data.
The 1321 bursts, t-SNE classifies 114 as short and 1207 as long. The UMAP embedding classifies just one burst differently: GRB 090813 (cyan, Figure 1) is long in the t-SNE embedding but short in UMAP. In addition, three bursts as classified differently here than in Jespersen et al. (2020). GRB 121226A and GRB 180418A are long in the Jespersen et al. (2020) catalog and short here, while GRB 050724 switches classification from short to long.

Although it may seem counterintuitive that a burst can switch classification when an identical technique is run on a superset of the data, this is indeed a property of both t-SNE and UMAP. Both algorithms assign a cost to placing every pair of objects at any specific distance, such that similar objects are less costly at short distances and dissimilar objects less costly at large distances, then seek to minimize the global sum of that cost. The addition of a new point will typically result in a lowest-cost configuration that involves not merely placing that point on an existing map, but shifting their locations as well. For example, an analogous physical system might be one in which every object is attached to every other by a spring, with the stiffness of that spring depending upon their similarity. The addition of a new set of springs will likely result in all of the distances changing in the equilibrium configuration.

One consequence of the choice of perplexity (for t-SNE) and n_neighbors (UMAP) is that both embeddings attempt to place outliers into a cluster where plausible. Thus, bursts which are somewhat dissimilar to both short and long GRBs are often placed on the edges of whichever group is more similar. The addition of a small number of similar objects can therefore change which group they are located close to. Thus, GRB 050724, GRB 121226A, and GRB 180418A are likely neither typical short bursts nor typical long bursts, but instead outliers, either for astrophysical reasons or due to a data processing artifact. An attempt to identify similar bursts in other data sets is described in Section 4.2.

3.2. Fermi

The 2294 bursts in the Fermi catalog are arranged into the two embeddings shown in Figure 2. These maps produce the clearest separation of any of the three data sets using both t-SNE and UMAP. As a result, no bursts were unable to be clearly assigned to either group. A possible interpretation is that the higher-energy bands in the Fermi data set are more useful for distinguishing between types of bursts than the bands available in BATSE and Swift. Had Fermi observed the full set of BATSE and Swift bursts, under this interpretation the Fermi embedding would be expected to look nearly identical on this larger data set.

The t-SNE map classifies 387 bursts as short and 1907 as long. In the UMAP map, there are 385 short bursts and 1907 long bursts. Three bursts (cyan, Fig 2) are classified as short by t-SNE and long by UMAP. GRB 090811696, GRB 110719825, and GRB 110728056. GRB 080828189 (also shown in cyan) is the sole burst classified as long by UMAP but short by t-SNE. The remaining 2290 classifiable bursts agree in both analyses.

Still, given the completeness of the separation on both the t-SNE and UMAP maps, it is perhaps surprising that four bursts change classification between the two embeddings. Several possible causes of these reclassifications are evaluated in Section 4.2. As a result of that analysis, in the catalog presented here, a measure of the uncertainty in classification is developed. For the remainder of this section, bursts will be described as short or long based on their most probable classification and the apparently clear separations in the embeddings in Figures 2–3. The catalog associated with these work includes not only the central values but also estimated likelihoods for these classifications. In that catalog, 2.0% of Fermi bursts have between a 10% and 90% probability of being classified as short (or long). Such objects are therefore labeled as ambiguous rather than as short or long.

3.3. BATSE

The 1911 bursts in the reduced BATSE catalog are arranged into the embeddings shown in Figure 3. Unlike Fermi and Swift, a significant number of BATSE bursts could not be included due to missing data, missing metadata, or high noise. In total, 791 of the 2702 BATSE bursts were discarded, and many of the remaining ones have marginal quality and could not be well constrained.
The t-SNE embedding classifies 491 bursts as short and 1420 as long. The UMAP embedding has similar size groups, with 484 long and 1427 long bursts. However, unlike the Fermi and Swift embeddings, there is a more substantial disagreement in classification. 21 objects are classified as short by UMAP and long by t-SNE, and 28 classified as long by UMAP and short by t-SNE (cyan, Figure 3). The individual objects which switch classification between embeddings are indicated in the catalog associated with this paper.

The disagreement would have been far stronger using the BATSE light curves obtained directly from the HEASARC server. To produce a separation, it was necessary to reprocess each light curve in order to fit and subtract a background. As a rudimentary subtraction, a linear fit was performed to the first (preburst) and last (well after $T_{100}$) data in each of the four bands, then subtracted from the full light curve. The background-subtracted light curves were then used to produce the embeddings and catalog in this work.

A more complete background subtraction would likely require rerunning or modifying the original processing pipeline. It is likely that at the end of such an effort, an improved and more robust separation would be possible. However, since higher-quality GRB data are available from newer observatories, here it is assumed that this would be of limited use. Thus, in this work the decision was made to include this approximate background subtraction for completeness, but to focus on separating the Swift and Fermi data sets, as they will be most suitable for further analysis.

4. Cross-matching and Validation

A significant potential concern when using unsupervised machine-learning methods is that there is no training set, there is an inherent inability to validate the conclusions. Although the GRB light curves can be cleanly separated into two groups, the methodology involved has no knowledge of astronomy or astrophysics. Thus, the statement that there are two distinct classes of GRB light curves does not necessarily mean that there are two astrophysical mechanisms for producing a GRB. It could instead be that the two groups have been separated based on data artifacts or processing pipeline decisions. t-SNE and UMAP are remarkable tools for finding clusters and categories, but not for determining the causes of those categories. Jespersen et al. (2020) demonstrated that the Swift classifications line up well with previous progenitor hypotheses. For example, all Swift bursts with a known, associated supernova afterglow were classified as long, consistent with previous expectations (Hjorth & Bloom 2012; Cano et al. 2017). However, only a few Swift GRBs have observed afterglows, so it is difficult to rule out the possibility of this separation having been caused by data processing rather than astrophysical origin.

A key goal of this work is to combine observations from all three available GRB observatories, with different bands, sensitivities, selections, and processing pipelines. If all produce a similar classification, this would validate the separation as being due to astrophysics. Here, two tests are performed using multiple catalogs in an effort to determine whether this classification is robust.

4.1. Cross-matching Swift and Fermi GRBs

307 bursts were observed by both Swift and Fermi, allowing a comparison between the two catalogs. If the classifications are robust and due to astrophysical origins, then bursts common to both catalogs must have the same label in both data sets. Of the 307, 298 have the same classification and only nine disagree (Figure 4), which strongly suggests that the separation is indeed based on the emission itself rather than artifacts induced by the data reduction.

The same approach also allows an investigation of the several smaller clusters which appear in the t-SNE and UMAP maps. The objects common to both telescopes which appear in a compact substructure of the long GRB groups (e.g., toward the top-left of the Swift map in Figure 1 and shown as the green points on Figure 4) do not comprise a distinct substructure in the other data set, but rather are merely part of the long GRB group. Therefore, these substructures are not interpreted as a distinct type of GRB or as having a distinct astrophysical origin, but rather as merely lying at one end of the parameter space of long GRBs.

4.2. Classification Stability

The nine bursts classified differently in Swift and Fermi suggest that even the seemingly unambiguous separations in these embeddings might not be entirely robust. Here, three sources of potential instability in classification are considered.

4.2.1. Data Ordering

First, the exact embeddings produced by t-SNE and UMAP depend upon the order in which the bursts are fed into the algorithm. Although embeddings produced by different orders have almost identical structures, the actual locations of individual objects will vary (Figure 5). In order to investigate this, 1000 maps were generated for each of the three data sets using burst lists sorted randomly into different orders. For each map, the bursts were divided into groups using spectral clustering (Fiedler 1973; Ng et al. 2001), directed to split into exactly two groups. In all 1000 trials, an identical set of short and long bursts was produced by both t-SNE and UMAP. Thus,
it can be concluded that the classifications are resilient to changes in ordering.

4.2.2. Outliers and Rare Bursts

Perhaps a greater concern comes from the choice of hyperparameters and resulting handling of outlier bursts. In this paper, the issue is described in terms of t-SNE hyperparameters, but is common to both algorithms. Dimensionality reduction algorithms must choose between preserving more local and more global structure, as the data are not truly two-dimensional and some information must be lost in the mapping. For t-SNE, this is controlled by perplexity. The embeddings here have been tuned to focus on separating the two major groups of GRBs.

During the gradient descent as t-SNE iteratively optimizes its embedding, each burst is attracted by similar bursts and repelled by dissimilar ones. For a perplexity of $N$, t-SNE imposes a probability density function which can be thought of as optimizing for a typical object having $N$ attractive neighbors. This will produce a clean separation between two groups each larger than $N$. However, tiny groups or individual objects with unique properties can be attached to the most similar group. If a burst is, e.g., far more similar to short bursts than to long bursts, it will be classified as short. However, if it has properties in common with both groups, then it will be attracted to both groups, and classified based on whichever set of attractors are stronger.

In order to search for these groups, a resampling-based approach is adopted. Subsets of (600, 900, 1000) bursts are drawn from the full (Swift, BATSE, Fermi) catalog, corresponding to \(~50\%\) of the total bursts, and an embedding produced for each subset. As before, spectral clustering is applied to separate the embedding into two groups. With the smaller samples, there are not always enough bursts to produce a clean separation without manual hyperparameter tuning. Thus, only subsets which produce a clean separation similar to the original grouping are included. Embeddings for which a Kolmogorov–Smirnov (KS) test indicates a $p < 0.10$ probability of being drawn from the same distribution as the separation in the full catalog are rejected.\(^8\) A KS test is chosen rather than a test such as Anderson–Darling in order to emphasize the bulk of the distribution rather than the tails, so that embeddings which move outliers will not be excluded.

The hope is that if a burst is attracted by both groups, then it might change location. In some of the random trials, many of its closest neighbors from one group or the other will be excluded. However, a prototypical short burst will always be most similar to short bursts, even if some of its closest analogs are excluded.

For the most part, this procedure indicates that the classifications are stable (Figure 6). In the Swift catalog, the median resampling trial has 0.5% of bursts change location. 6.6% of bursts change location in at least one of the \(~750\) trials, and 2.0% change location in more than 10% of trials. This latter group is labeled as ambiguous in the catalog from this work. In the Swift catalog, 0.5% change location in the median trial, 13.0% change location at least once, and 2.6% are ambiguous. In the BATSE catalog, 2.8% of bursts change location in the median trial, 20.6% change location at least once, and 9.2% are ambiguous.

4.2.3. Insufficient Information

Of the nine bursts classified differently by Swift and Fermi, seven change location in at least one resampled map in at least one catalog and six change location more than 10% of the time (Table 2).

The remaining bursts, GRB 090531B and GRB 130716A, are consistently classified differently, in this case as long bursts by Swift and short bursts by Fermi. That is, in the Swift data set alone, they are similar to typical long burst and in the Fermi data set alone, they are similar to typical short bursts. A reasonable interpretation is that these are extended-emission bursts (Norris & Bonnell 2006; Kaneko et al. 2015), which are known to be shorter in the harder emission observed by Fermi and longer in Swift.

The broader implication is that for some uncommon types of bursts, the information provided by one data set alone is insufficient to classify them. Extended-emission bursts might only be detectable with a combination of harder and softer emission, which currently no single telescope can provide. These bursts do appear distinct in an analysis which combines both Swift and Fermi data. However, such a combination only exists for around 15% of these catalogs. Thus, if GRB 090531B and GRB 130716A are indeed extended-emission bursts, there are likely an additional \(~15\) undetected extended-emission bursts classified as long in Swift with no Fermi data, and a similar number of undetected extended-emission bursts classified as short in Fermi with no Swift data.

\(^8\) Note that by looking for similar distributions, this procedure estimates the stability of the specific bimodal classification being evaluated. In principle, given computing resources well beyond those available to the authors, one could systematically search for the most stable bimodal classification.
4.3. Bulk Properties of Short and Long GRBs

Another way to evaluate whether the separations in these three catalogs are identical is to compare the bulk properties of the short and long populations in all three data sets. If GRBs truly are correctly separated into classes by astrophysical origin, then the short and long GRBs observed by each telescope should be drawn from identical distributions, and thus have identical distributions of properties. Further, with a clean separation between short and long GRBs, it should now be possible to determine which other properties correlate with GRB type, something that would not have been possible for complete samples without this method.

Such a comparison is significantly complicated by different selection functions for each telescope. In particular, the duration is dilated by a factor of $(1 + z)$, and therefore telescopes sensitive to GRBs across a wider range of redshifts should also have a broader distribution of durations. Because redshifts are only known for a small fraction of GRBs, it is not possible to compare rest-frame durations for the full samples.

Still, a comparison of the observed durations indicates a similar separation in each data set (Figure 7). The most similar, given their similar frequency ranges, are BATSE and Fermi. The short GRB durations in the Swift data set include a longer-duration tail than in BATSE or Fermi. Thus, it is possible that all three telescopes are selecting a similar set of short bursts down to redshift distributions and detection thresholds. However, the softer bands in Swift produce a dissimilar duration distribution of long bursts when compared with BATSE and Fermi.

The distributions of short and long GRBs in a hardness–duration plot further indicates that the t-SNE and UMAP classifications are separating objects with similar bulk properties (Figure 8). Only BATSE and Swift are shown here, as

---

Table 2

<table>
<thead>
<tr>
<th>Burst</th>
<th>Swift</th>
<th>Fermi</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRB 090531B</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>GRB 090927</td>
<td>0.070</td>
<td>1.000</td>
</tr>
<tr>
<td>GRB 130716A</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>GRB 131044A</td>
<td>0.194</td>
<td>0.754</td>
</tr>
<tr>
<td>GRB 140209A</td>
<td>0.682</td>
<td>0.081</td>
</tr>
<tr>
<td>GRB 140320A</td>
<td>1.000</td>
<td>0.346</td>
</tr>
<tr>
<td>GRB 141205A</td>
<td>0.257</td>
<td>1.000</td>
</tr>
<tr>
<td>GRB 151020A</td>
<td>0.146</td>
<td>0.860</td>
</tr>
<tr>
<td>GRB 170318B</td>
<td>0.926</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note. Six of the bursts have unstable classifications in at least one of the two catalogs, which implies that these are neither typical short nor long bursts. One additional burst, GRB 090927, changes location in 7% of the resampled Swift maps. However, two, GRB 090531B and GRB 130716A, have entirely stable but conflicting classifications. These are likely extended-emission bursts, and indicate that there is not enough information in either data set alone to determine that they are atypical.
5. Discussion

Following a pilot study using t-SNE to classify Swift GRBs from the full observed light curves (Jespersen et al. 2020), here dimensionality reduction is shown to be able to classify observations from all three major GRB observatories, BATSE, Swift and Fermi. As with Swift, all three data sets produce a separation into two substantial groups. This fits with previous work proposing two distinct classes of progenitors; that short bursts are likely associated with mergers (Berger et al. 2013; Tanvir et al. 2013; Ghirlanda et al. 2018) and long bursts with the core collapse of massive stars (Hjorth et al. 2003; Stanek et al. 2003).

In previous studies, the definition of a short or long burst has generally been based on duration only. Since the two distributions overlap in duration, some bursts will therefore be misclassified. In this work, the separation into short and long is based on the entire observed light curve, and produces a clean separation into two groups for each data set without overlap. The hope is that this clean separation will correspond to physical properties, and that the resulting short and long bursts will indeed have distinct astrophysical origins. Jespersen et al. (2020) found that a comparison of their classifications with both known and proposed supernova afterglows supports this interpretation.

A significant issue with many machine-learning methods is that classifications can be based on extraneous information, confounding variables, or metadata. Thus, the clean separation in the Swift data set reported by Jespersen et al. (2020) might occur due to different astrophysical origins, as hoped, but could also be produced by differences in data processing. A comparison of all three catalogs shows a similar separation with similar properties. Further, nearly all of the objects observed by both Swift and Fermi are classified in the same way. Therefore, it can be concluded that the separations reported here are truly astrophysical in origin.

5.1. Additional Classes

Although 97% of the bursts common to Swift and Fermi are classified identically, nine are not. Several of these bursts are part of the small fraction with unstable classifications (Section 2). However, at least two bursts, GRB 090531B and GRB 130716A, have entirely stable but conflicting classifications. In the Swift data alone, they appear to be typical long bursts, but in the Fermi data alone, they appear to be typical short bursts. A strong possibility is that these are part of the previously reported group of extended-emission bursts (Norris & Bonnell 2006; Kaneko et al. 2015). If they have a distinct astrophysical origin, then GRBs should be properly divided into at least three groups, not two.

However, their presence should also suggest that additional classes of bursts might exist which are difficult to classify from any single one of these three observatories. It might have been hoped that the use of additional information could separate them from more typical short or long bursts. However, dimensionality reduction algorithms which use all of the available information have been unable to do so. The authors of this work attempted to tune the preprocessing and hyperparameters in order to separate these bursts from the others, but were unable to do so for any single catalog.

Of course, given the combination of the Swift and Fermi observations, it is easy to identify these objects as the

---

Figure 8. Distribution of objects in hardness and duration in the Swift (left) and BATSE (right) data sets. It should be noted that due to the different bands used in the calculation, the Swift and BATSE hardness measurements are on different scales and cannot be compared directly. The Fermi catalog does not include a direct measure of hardness. The two catalogs exhibit a qualitatively similar, overlapping distribution of short and long GRBs. This is consistent with a selection that depends upon intrinsic GRB properties rather than details of the observatory or processing pipeline.

Fermi does not produce hardness as part of its data release. It should be noted that due to the different available bands, the hardness measured by BATSE cannot be directly compared with that measured by Swift. For both BATSE and Swift, short GRBs are generally harder and shorter than long GRBs, but with some overlap. This is again consistent with the hypothesis that short and long GRBs are robust categories intrinsic to GRB emission rather than to details of the observatories or processing pipelines used to measure their properties.
only ones which are classified differently. Perhaps this is not so surprising. Multiwavelength astronomy has proven to be far more powerful than any single observatory, and multiinstrument astronomy is poised to be similarly powerful. Thus, a key conclusion here is that the next generation of gamma-ray observatories should be constructed with multiwavelength observations in mind, and that the combination of Swift and Fermi has already proven to be more powerful than an improved version of either observatory would be alone.

It is important to note that “unsupervised” algorithms such as the ones used here are in practice strongly dependent on the choice of hyperparameters. The proper interpretation of the embeddings presented here should not be that the hyperparameters we have chosen are correct and others are incorrect. Rather, every choice of hyperparameters leads to a valid embedding, which contains potentially useful information about the distribution but is always incomplete. In that sense, it would be more like considering various projections. Unlike projections, however, there is no rigorous formalism such as principal component analysis for determining which will be most useful. Thus, the inability of this work to find hyperparameters which separate these possible extended-emission bursts from a single catalog does not guarantee that insufficient information exists to do so.

There is also considerable literature on the possibility of a third class consisting of intermediate-duration bursts found in the BATSE catalog (Mukherjee et al. 1998; Hakkila et al. 2003; Horváth et al. 2004; Chattopadhyay et al. 2007; Ípás et al. 2009; Zhang et al. 2022). Thus, a search for hyperparameters that separate such an intermediate group is well motivated. Using the same techniques presented here, it was indeed possible to produce an embedding that separates a group of intermediate duration using the BATSE 4B catalog. However, this group does not appear in a catalog restricted to post-4B BATSE bursts, and no similar group was found when tuning the hyperparameters to search for one in the Fermi and Swift catalogs. As a result, a reasonable conclusion is that this group is not of astronomical origin (see Hakkila et al. 2000), but rather is related to instrumentation or data reduction techniques applied only to earlier part of the BATSE data set.

5.2. Physical Interpretations

Norris et al. (2010) suggest that approximately 1/4 of a sample of Swift short GRBs have signatures of extended emission, and estimate that the true fraction may be as high as 50%. This is not observed in our classifications, which could be due to detector effects, but possibly also due to the “choking” mechanism suggested by Bucciantini et al. (2012), which would reduce the rate of observed extended-emission bursts. All the bursts classified here as short but with $T_{90} > 2$ s are within or close to the approximate theoretical boundary of $\approx 100$ s suggested by Metzger et al. (2008) and Bucciantini et al. (2012).

Although the classifications presented here do not immediately allow for distinctions between different progenitor scenarios for the extended emission, there are several signatures that will be exciting to follow with the next generation of GRB observatories. Currently, the only viable way to distinguish between the two primary proposed progenitor scenarios, neutron star—neutron star (NS-NS) mergers and accretion-induced collapse (AIC), is by the signature of the elements produced during the event (Metzger et al. 2008). This was done for a NS-NS merger by Watson et al. (2019), who identified strontium in the spectra of the afterglow. However, these spectra have low signal-to-noise ratios, and need to be taken within a few days, making them hard both to obtain and analyse.

Another possible way of distinguishing between different short GRB extended-emission progenitor mechanisms would be to rely on the extra polarization that would be produced during the prompt emission in the AIC scenario. The proposed Daksha mission will carry X-ray polarimeters which will be able to measure the polarization of the prompt emission shortly after a trigger (Bhalerao et al. 2022). Including the polarization would then allow t-SNE/UMAP to subdivide the short group further, corresponding to either NS-NS mergers (without extended emission) or AICs (with extended emission), if both progenitors classes do exist.

This approach would also lend itself to distinguishing between different progenitor mechanisms for long GRBs (Toma et al. 2009), but will require having a large statistical sample due to the large uncertainties in the observed prompt emission polarizations (Kole et al. 2020). For the planned Daksha mission, the lowest energy that will be detectable is currently 1 keV, but based on the models of Metzger et al. (2008) and Bucciantini et al. (2012), this should ideally be even lower, preferably past the Swift 0.3 keV limit, in order to distinguish between the different progenitor classes.

5.3. Robustness of Machine Learning for Astronomical Problems

It is surprising that even though the embeddings produce clear separations, the classifications are not entirely stable. Even though the results of Jespersen et al. (2020) appeared unambiguous, and have been reproduced by independent groups from the same codebase, a few objects may still have been misclassified. Resampling shows that in a study measuring 1000 similar bursts, a small fraction of bursts could end up being classified as either short or long. Even for the Fermi catalog, with the most robust classifications, 2.0% of the bursts change classification in at least 10% of trials. However, in any individual embedding, there is a clear separation providing an unambiguous assignment of each GRB as either short or long.

This is one example of a more generic issue when using machine-learning methods in astronomy. Statistical methods typically are associated with theorems proven from a set of axioms. Thus, it is known with certainty that if those axioms apply to a data set, a particular method will produce, e.g., the minimum variance unbiased estimator. Such a method needs a proven theorem in order to be considered valid.

However, machine-learning algorithms are often instead validated via benchmark problems and data sets. An algorithm which outperforms previous attempts at those problems is considered successful. However, there is rarely a rigorous formalism proving optimality. One of the advantages of UMAP over t-SNE is that there is a stronger mathematical argument for its embedding.

Further, often these benchmark problems are on idealized, noiseless data sets. One of the standard dimensionality reduction problems is to classify handwritten digits in the MNIST data set (LeCun et al. 1998). These digits are noiseless and have no missing pixels. Indeed, a change as simple as inverting the colors, which leaves the digits equally legible for
A data table (Table 3) containing the classification of each burst is given here and available online in a machine-readable format. For each burst, a classification is given for each telescope which observed it. The possible classifications are the following types:

1. S (short): a burst identified as short by both t-SNE and UMAP and in at least 90% of the resampled t-SNE maps.
2. L (long): a burst identified as long by both t-SNE and UMAP in at least 90% of the resampled t-SNE maps.
3. A (ambiguous): a burst which was classified as both short and long in at least 10% of the resampled t-SNE maps.
4. D (disagreement): a burst which is classified differently by t-SNE and UMAP, but which is classified consistently in at least 90% of the resampled t-SNE maps.

In addition, there is an overall classification given for each burst. If the burst is observed by only one telescope, the overall classification is same as for that telescope. For bursts observed by both Swift and Fermi, if both give the same classification, this also the overall classification. A burst which is S or L in one catalog and A or D in the other is classified as S or L based on the unambiguous classification. The three bursts which are classified as L by Swift but S by Fermi, GRB 090531B, GRB 090927, and GRB 130716A, are classified as D (disagreement) in the overall classification. These may be extended-emission bursts, as discussed in the main text.

The primary name used for each burst, i.e., the name given each individual catalog, with the Fermi name used for objects observed by both Fermi and Swift. Probabilities given are the probability that any given burst is short or long in the resampled data.

**ORCID iDs**

Charles L. Steinhardt @ https://orcid.org/0000-0003-3780-6801
William J. Mann @ https://orcid.org/0000-0001-7691-0762
Vadim Rusakov @ https://orcid.org/0000-0001-7633-3985
Christian K. Jespersen @ https://orcid.org/0000-0002-8896-6496
References

Becht, E., McInnes, L., Healy, J., et al. 2019, NatBi, 37, 38
Fiedler, M. 1973, CzMJ, 23, 298
Kobak, D., & Berens, F. 2019, NatCo, 10, 5416
Nakar, E. 2007, PRL, 442, 166
Sun, Y., Zeng, Y., & Zhang, T. 2021, iScience, 24, 102880

The Astrophysical Journal, 945:67 (11pp), 2023 March 1

Steinhardt et al.