

GENERALIZED TESTS FOR SELECTION EFFECTS IN GRB HIGH-ENERGY CORRELATIONS

NATHANIEL R. BUTLER^{1,2,3}, DANIEL KOCEVSKI², AND JOSHUA S. BLOOM^{2,4}.

Draft version September 16, 2008

ABSTRACT

Several correlations among parameters derived from modelling the high-energy properties of GRBs have been reported. We show that well-known examples of these have common features indicative of strong contamination by selection effects. We focus here on the impact of detector threshold truncation on the spectral peak versus isotropic equivalent energy release ($E_{\text{pk}}-E_{\text{iso}}$) relation, extended to a large sample of 218 Swift and 56 HETE-2 GRBs with and without measured redshift. The existence of faint Swift events missing from pre-Swift surveys calls into question inferences based on pre-Swift surveys which must be subject to complicated incompleteness effects. We demonstrate a generalized method for treating data truncation in correlation analyses and apply this method to Swift and pre-Swift data. Also, we show that the $E_{\text{pk}}-E_{\gamma}$ (“Ghirlanda”) correlation is effectively independent of the GRB redshifts, which suggests its existence has little to do with intrinsic physics. We suggest that a physically-based correlation, manifest observationally, must show significantly reduced scatter in the rest frame relative to the observer frame and must not persist if the assumed redshifts are scattered. As with the $E_{\text{pk}}-E_{\gamma}$ correlation, we find that the pre-Swift, bright GRB $E_{\text{pk}}-E_{\text{iso}}$ correlation of Amati et al. (2006) does not satisfy these conditions significantly.

Subject headings: gamma rays: bursts — methods: statistical — cosmology: observations

1. INTRODUCTION

Correlations are pervasive in astronomy and generally lead theory in allowing us to discover causal relationships between observed quantities. In the study of Gamma-ray Bursts (GRBs), multiple power-law relations among GRB observables have been used to uncover the intrinsic physics of GRBs themselves (e.g., Eichler & Levinson 2004; Yamazaki et al. 2004; Rees & Mészáros 2005; Levinson & Eichler 2005; Lamb, Donaghy, & Graziani 2005) and to use GRBs as probes to the distant Universe (e.g., Ghirlanda et al. 2005; Schaefer 2007). Several authors have critically examined the limitations of such relations: understanding how well they potentially constrain physics given their form and scatter (e.g., Friedman & Bloom 2005; Schaefer & Collazzi 2007), uncovering possible evolution with cosmic time (e.g., Yonetoku et al. 2004; Li 2007), realizing the commonality of outliers (Nakar & Piran 2005; Band & Preece 2005; Kaneko et al. 2006), and — most fundamentally — determining how the existence and form of the relations vary once spurious correlation imparted by selection effects is treated (e.g., Lloyd, Petrosian, & Mallozzi 2000).

Recently, Butler et al. (2007, hereafter B07) report evidence from Swift satellite (Gehrels et al. 2004) observations of GRBs that several of the correlations exhibit a wide scatter and shift in normalization toward the Swift detection threshold, suggestive of an origin intimately connected to the detection limits of pre-Swift satellites. Because relations intrinsic to the physical processes underlying GRBs should not be instrument-dependent, a broader investigation into the data from pre-Swift satel-

lites is crucial for determining whether and/or how the relations can be trusted to potentially constrain the physics of GRBs or cosmology.

Here, we extend our critique to include the full B07 catalog of Swift GRBs with and without measured redshifts z (for 218 total Swift GRBs) and also to include HETE-2 GRBs with and without measured z (56 GRBs). The uniform B07 catalog is novel for deriving bolometric fluences for all Swift GRBs (between GRBs 041220 and 070509), without requiring tight error bars on the spectroscopic fit parameters. The sample is therefore flux limited by the sensitivity of the Burst Alert Telescope (BAT; Barthelmy et al. 2005); because the X-ray Telescope (XRT; Burrows et al. 2005) localizes to few arcsecond precision nearly all BAT GRB afterglows, additional flux limits associated with afterglow localization and host galaxy detection (important for example when considering the GRBs with measured z) are likely not present or are far less important as compared to stronger flux limits imposed in pre-Swift surveys.

A study similar to this has recently been conducted by Ghirlanda et al. (2008) comparing a smaller sample of bright Swift GRBs to the Beppo-SAX sample (although only the sub-samples with measured z). Ghirlanda et al. (2008) find that faint GRBs detected by Swift (and HETE-2; Section 2) are missing from the Beppo-SAX sample, and there is no clear explanation for the missing data in terms of either the SAX trigger threshold or a (likely higher) threshold resulting from a demand that tight error bars be derived in the GRB spectral modelling. As in Ghirlanda et al. (2008), we also focus on the correlation between the isotropic equivalent energy E_{iso} and the peak in the νF_{ν} spectrum E_{pk} (Lloyd, Petrosian, & Mallozzi 2000; Amati et al. 2002).

Selection effects in high-energy GRBs observables are well-documented (e.g., Lee & Petrosian 1996), if rarely treated. For faint GRBs, an expected departure of the true distributions of flux, duration, etc., from the ob-

¹ Townes Fellow, Space Sciences Laboratory, University of California, Berkeley, CA, 94720-7450, USA

² Astronomy Department, University of California, 445 Campbell Hall, Berkeley, CA 94720-3411, USA

³ GLAST/Einstein Fellow

⁴ Sloan Research Fellow

served distributions is expected given the steep GRB number density versus peak photon flux F relation ($dN/dF \propto F^{-a}$, $a \sim 2$; e.g., Preece et al. 2000), which places most GRBs in a given sample near the detection limit — or near some other imposed sample cutoff at a higher flux level — with a narrow logarithmic dispersion of $\ln(10)/(1-a) \sim 0.4$ dex.

In Section 2 of this paper we study the correlation of $E_{\text{pk,obs}}$ with the energy fluence $S \propto FE_{\text{pk,obs}}$ roughly, for which it is clearly important to compensate for the expected non-intrinsic correlation of fluence with $E_{\text{pk,obs}}$ that arises essentially because GRB detectors are photon counters and not bolometers. Similarly structured correlations are then discussed in Section 3.

Two basic approaches have been attempted to compensate for GRB flux limits: (1) forward folding of a model GRB rate density, evolution, and luminosity function through a model of the detector response (e.g., Graziani, Donaghy, & Lamb 2005; Lamb, Donaghy, & Graziani 2005), and (2) applying non-parametric statistical methods to the data and the observed flux limit for each data point. In Section 2, we review and apply numerical methods which treat the observed data truncation directly according to (2), while we plan to approach the B07 data via path (1) in a future paper. In order to potentially better understand the origin of pre-Swift correlations not strongly present in the Swift sample, we explore the significance of these correlations after rejecting GRBs from the Swift sample which would not have been detected by HETE-2 (Section 2).

Extending our critique to the Ghirlanda et al. (2008) and similar pre-Swift samples of bright GRBs with redshifts (e.g., Amati et al. 2002, 2006; Schaefer 2007) — where the data truncations are poorly known and the GRBs come from multiple instruments of varying sensitivity — we study (Section 3) how the relations transform from the observer to the source frames. We show that redshift dependence in the correlations is generally weak, which is expected for a correlation that arises due to detection selection effects but also continues to have low scatter in the source frame. We present tests to uncover whether the redshift dependence in other correlations is similarly weak, possibly providing circumstantial evidence that these correlations are non-intrinsic. This general approach, despite its limitations, can potentially find application in other population analyses where thresholds are important.

2. FLUX LIMITS AND THE $E_{\text{PK}}-E_{\text{ISO}}$ CORRELATION

In Figure 1, we reproduce the narrow (< 0.2 dex) scatter in $E_{\text{pk}} \sim E_{\text{iso}}^{0.5}$ observed in the recent review by Amati et al. (2006) for data from multiple missions. Following Nakar & Piran (2005) in realizing that $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ reaches a maximum for $z \approx 3.83$, we assume this redshift value for all GRBs without measured redshift to place the entire B07 sample on the plot (black points). Most (67%) best-fit Swift values are below the lower 1σ red-dotted Amati et al. (2006) line. Nearly half (41%) are below the red dotted line at 90% confidence, indicating a clear preference for a lower flux normalization. We determine 90% confidence error bars directly for $E_{\text{iso}}^{0.5}/E_{\text{pk}}$, given the observed spectrum, rather than attempting to propagate errors on the covariant quantities E_{iso} and E_{pk} .

If we plot the BAT threshold corresponding to the best-fit points, we can see that agreement (right side of the plot) is partly dictated by low-sensitivity, while many sensitive observations are strongly inconsistent. This finding is in excellent agreement with the BATSE studies of faint GRBs (Nakar & Piran 2005; Band & Preece 2005). There is no clear separation in Figure 1 between the Amati et al. (2006)-consistent and Amati et al. (2006)-inconsistent points, precluding obvious (non-circular) cuts which could create the semblance of close consistency.

In order to determine the most accurate threshold for each burst, we utilize the observed GRB spectrum and time profile directly instead of employing a model for the threshold (e.g., Band 2003) which would only utilize results from the spectral fitting (i.e., peak photon flux, $E_{\text{pk,obs}}$, etc.) and possibly the burst duration (e.g., Band 2006). After determining the energy band — taken to be 15-350 keV for Swift but allowed to vary for HETE-2 (see Figure 2) — and temporal region which maximizes the signal-to-noise ratio S/N , we divide the observed fluence by $(S/N)/10$. This assumes a background dominated light curve, which is appropriate for all but the brightest handful of Swift GRBs, for which our threshold calculation takes into account the observed background counts.

Here and below, we choose a cutoff $S/N = 10$, because this is where the observed S/N distributions appear to turn over. The turn-over indicates a drop in detection efficiency, because the number of faint bursts is known from BATSE (e.g., Preece et al. 2000) to increase at low flux levels. Because HETE-2 and Swift have intelligent trigger systems which seek to find the light curve region maximizing the S/N (e.g., Band 2006), the counts that define the S/N also define the fluence at the threshold, approximately. This threshold estimate is conservative in cases where the trigger system fails to find the optimal S/N region.

Many of the BAT points in Figure 1 are likely lower than plotted, because most known GRBs have $z < 3.83$. Also, the lower limit error bars in the figure for most of the points are strongly influenced by a BATSE-based prior on $E_{\text{pk,obs}}$, which inhibits $E_{\text{pk,obs}} > 300$ keV (see, B07).

Studies including some bright BAT GRBs (e.g., Amati et al. 2006; Schaefer 2007) do not report the strong levels of inconsistency because the studies exclude the faint or hard GRBs for which tight $E_{\text{pk,obs}}$ constraints cannot be obtained. However, for the simple question of consistency/inconsistency, tight $E_{\text{pk,obs}}$ error bars are not required.

It is interesting to note that short-duration GRBs tend to appear in the lower left of Figure 1, populating a region where GRBs of longer duration generally cannot be detected. Only one short-duration GRB is detected at a flux level consistent with the pre-Swift $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ level. As for long-duration GRBs, short-duration GRBs tend to be detected near threshold; because of the impulsive nature and low peak flux over fluence ratio for short-duration GRBs, they have more potential than long-duration GRBs to be outliers to an $E_{\text{pk}}-E_{\text{iso}}$ relation. We caution, therefore, against using low $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ values to classify (as opposed to, e.g., Amati 2008) intermediate duration events as short-duration (i.e., potential binary

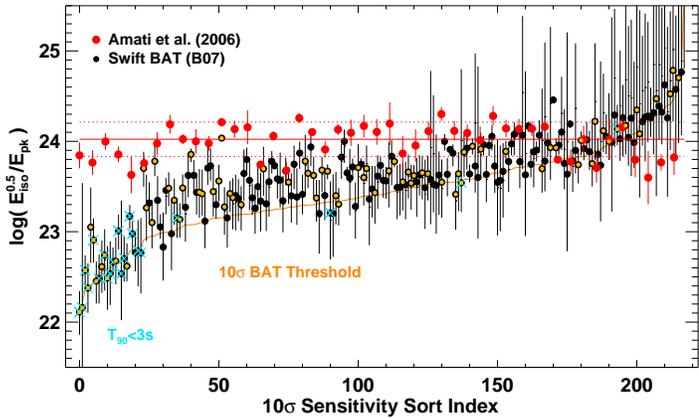


FIG. 1.— The majority ($> 67\%$) of Swift GRBs (black circles) prefer a lower $E_{\text{iso}}-E_{\text{pk}}$ normalization relative to the pre-Swift GRBs in Amati et al. (2006) (red circles). Swift GRBs with measured z are marked with a small yellow circle. For GRBs without measured z , we assume $z = 3.83$, which maximizes $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ (see, also, Nakar & Piran 2005). Swift short GRBs ($T_{90} < 3$ s) are marked with a cyan X . At these z , many (41%) of the Swift sample GRBs exhibit $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ values inconsistent (at 90% confidence) with the (Amati et al. 2006) relation (red horizontal lines). The Amati et al. (2006) consistent GRBs typically lie to the right of the plot, which is sorted by 10σ limiting $E_{\text{iso}}^{0.5}/E_{\text{pk}}$, indicating that these GRBs would be undetectable at significantly lower fluence levels.

merger) events.

2.1. From Swift BAT to HETE-2

One straight-forward way to diagnose whether a flux limit is truly the origin of inconsistency between the BAT and pre-Swift samples is to compare the BAT data to data from satellites with very different detection sensitivities. If the corresponding surveys both extend to the respective limiting flux limits, then powerlaw fits for each sub-sample should have appropriately different normalizations. (In pre-Swift datasets considered alone as in the next section, the relative detection sensitivities generally do not vary widely, and a different technique must be adopted.) In this sub-section, we compare the Swift and pre-Swift samples using a well-known approach which directly treats the flux limits.

The application and refinement of statistical methods in survival analysis was pioneered in BATSE GRB studies by Petrosian (1993), Lee & Petrosian (1996), and others. By restricting comparison to “associated sets” of data – those GRBs detectable above the estimated thresholds of each other GRB in the set – it is possible to study distributions of and correlations among observables independent of the thresholding, without making assumptions on the nature of the GRBs below threshold.

In the early afterglow-era, these non-parametric methods were applied to correlations between GRB source frame quantities (Lloyd & Petrosian 1999; Lloyd, Petrosian, & Mallozzi 2000; Lloyd-Ronning & Ramirez-Ruiz 2002; Kocevski & Liang 2006). The study of Lloyd, Petrosian, & Mallozzi (2000), in particular, is important for (1) demonstrating that the correlation between isotropic equivalent γ -ray peak flux L_{iso} and E_{pk} likely arises as a result of the detection process while also (2) discovering a possibly intrinsic

correlation between the isotropic equivalent energy release E_{iso} and E_{pk} (also, Amati et al. 2002).

The Kendall’s τ_K statistic employed in these studies reports the fraction of concordant (i.e., correlating) data minus the fraction of discordant data (i.e., anti-correlating). Use of the statistic to rule on a correlation’s strength derives from maximum likelihood principles, and the τ_K -test is non-parametric and has maximum statistical power (Kendall 1938). A simple, elegant, and rigorous extension of the τ_K -test accounts for data truncation (see, Efron & Petrosian 1999): tabulation is restricted to the concordant or discordant data pairs which are detectable above (or below in the case truncations from above) each other’s limits. In this way, both lower and upper limits on either the x or y variable can be treated, without the need to make assumptions regarding the missing data. Assumptions, sometimes difficult to uncover and possibly also hard to justify, regarding the missing data must generally be made with other methods (e.g., Gelman et al. 2004).

A major advantage of the truncated τ_K -test (which employs the limits directly) is that it is in general much easier to calculate a limit for an individual burst than it is to derive a general limit in E_{iso} valid for all E_{pk} . In fact, the general limit can never be calculated precisely, because the flux limit (e.g., a satellite trigger threshold) is not a function of E_{iso} , but of the observed photon flux in some band pass over some shorter time integration.

In the observer frame (Figure 2A), there is a highly significant apparent correlation between S_{bol} and $E_{\text{pk,obs}}$ for both HETE-2 ($\tau_K = 0.47, 5.2\sigma$) and Swift ($\tau_K = 0.45, 9.3\sigma$). Here, S_{bol} is the energy fluence calculated in the source frame $1-10^4$ keV band if z is known (e.g., Amati et al. 2002) or in the observer frame $1-10^4$ keV band if z is unknown. The significance of the correlation becomes modest ($\tau_K = 0.27, 4.4\sigma$) when we account for the Swift threshold.

That some significance remains indicates some amount of true correlation between $E_{\text{pk,obs}}$ and S_{bol} ; although, this residual correlation is strongly affected by the flux limit. Quantifying the slope, scatter, etc., of this possible residual correlation and its origin in local intrinsic or population-wide evolution effects will be the subject of a future study. Here, we are interested in the expected appearance of the correlation in the Swift survey relative to pre-Swift surveys. The significance of the Swift ($\tau_K = 0.21, 2.4\sigma$) or HETE-2 ($\tau_K = 0.16, 1.2\sigma$) correlation becomes marginal if we account for the HETE-2 detection threshold (see, Figure 2A). To apply the HETE-2 threshold to Swift data, we fit a second order polynomial to the HETE-2 threshold and evaluate that curve at the Swift $E_{\text{pk,obs}}$ points.

In the source frame and restricting to the 63 normal, long-duration GRBs with z in B07, we find $\tau_K = 0.59$ (6.8σ) for the $E_{\text{pk}}-E_{\text{iso}}$ relation. However, considering only associated sets of observations above Swift threshold, the correlation strength drops strongly ($\tau_K = 0.28, 2.4\sigma$). The similarly strong decrease in correlation significance for the source and observer frame correlation can be understood by noting that the correlations involve very similar data: observer frame points above and below the HETE-2 threshold transform to source frame points with a very weak dependence on the GRB redshifts. The red dashed curve in Figure 2B — corresponding to the

trajectory one event follows as its z is varied — illustrates how z only weakly affects where points fall on the plot (see also, Section 3). Points in the extreme Bottom-Right of Figure 2A or Top-Left of Figure 2B cannot be made consistent at any z (Figure 1; see also, Nakar & Piran 2005; Band & Preece 2005).

Randomly sampling redshifts from those observed by Swift, we find that $(80 \pm 3)\%$ of Swift events above the HETE-2 threshold are consistent with the pre-Swift relation (solid and dotted lines). Only $(20 \pm 3)\%$ of Swift event below the HETE-2 threshold are consistent. Most (60%) of the events under the HETE-2 threshold are inconsistent at any z . We find similar results — $(62 \pm 4)\%$ consistency above HETE-2 threshold and $(15 \pm 3)\%$ consistency below HETE-2 threshold — if we assume a very different (toy) z distribution and draw $\ln(z)$ from a unit normal distribution. The best-fit $E_{\text{pk}}-E_{\text{iso}}$ curve from B07 is a factor ≈ 2 higher than the pre-Swift curve (solid and dotted lines in Figure 2B).

3. THE ANATOMY OF A CORRELATION

We now turn our attention to a set of correlation tests useful for GRB samples with poorly understood flux limits. As mentioned above, pre-Swift afterglow-era satellites do not have typical sensitivities that vary by more than a factor of a few (see also, Band 2003); we find that diagnosing the reliability of a correlation plotted on a log-log scale for these data cannot effectively be done by looking at variations in the normalizations alone. Moreover, the truncations tests discussed in the preceding section cannot easily be applied because the flux limits for each GRB in most pre-Swift surveys are unpublished. We seek to understand how the regression slope and scatter of the relation fits vary with assumptions on the nature of the relations.

For simplicity, if we assume that all GRB fit parameters have the same error, the errors can be ignored. Consider observer frame quantities $X = \log(S_{\text{bol}})$ and $Y = \log(E_{\text{pk,obs}})$, for example. These translate to source frame quantities $X' = X + bZ = \log(E_{\text{iso}})$ and $Y' = Y + aZ = \log(E_{\text{pk}})$, for example. Here, $Z = \log(1+z)$ and $a = 1$, $b \approx 3$ (accurate to 20% for $0.5 < z < 5$, representing the bulk of known GRB redshifts).

If we assume an origin in the observer frame, which we hope later to rule out, cross terms involving the observables and z can be ignored, and the expected source-frame regression slope is:

$$m_{\text{src}} = \left[m_{\text{obs}} + ab \frac{\langle Z^2 \rangle}{\langle X^2 \rangle} \right] / \left[1 + b^2 \frac{\langle Z^2 \rangle}{\langle X^2 \rangle} \right]. \quad (1)$$

Here, $\langle \cdot \rangle$ denotes an average after subtracting away the mean in X or Z , and m_{obs} is the slope in the observer frame. Equation 1 gives the slope of the *apparent* source frame relation given the observed data, whereas determination of the *true* source frame relation requires detailed knowledge of the GRB rate density and luminosity function to impute the missing data. Although recent progress has been made on measuring these (e.g., Kistler et al. 2008), we do not attempt here to reconstruct $m_{\text{obs}}[m_{\text{src,true}}]$. Moreover, for comparison with most GRB studies which simply assume $m_{\text{src,true}} = m_{\text{src}}$ (e.g., Firmani et al. 2005), we only require Equation 1.

Relatively independent of which survey we use, $\langle Z^2 \rangle / \langle X^2 \rangle \approx 0.1$, and $m_{\text{src}} \approx m_{\text{obs}}$. This is the typi-

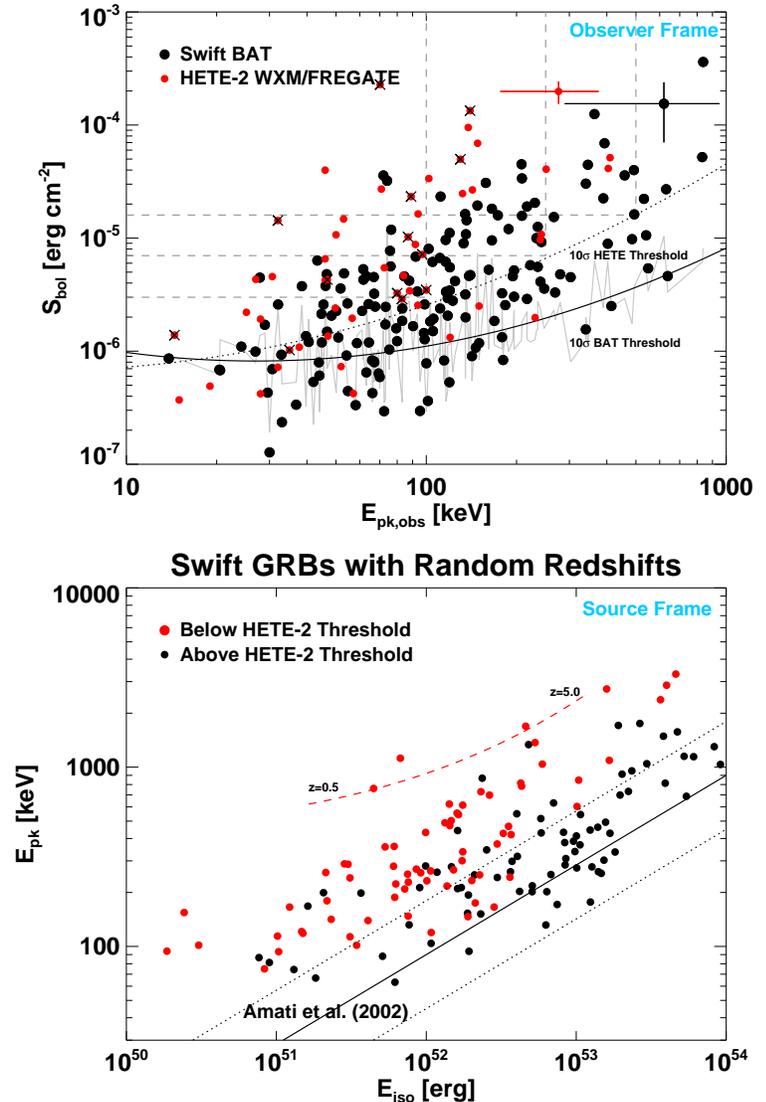


FIG. 2.— From data truncation to spurious amplification of correlation strength in the observer and source frames. Typical error bars are plotted for points in the Top-Right corner. (A) There is a strong ($> 9\sigma$ significant) correlation in the observer frame between $E_{\text{pk,obs}}$ and S_{bol} for Swift (B07) and HETE-2 (Sakamoto et al. 2005). The jagged, gray line shows the observed detection S/N relative to $S/N = 10$ for Swift GRBs, and the solid black line is a fit to those values. Similarly, using the larger S/N value from either FREGATE or WXM from the HETE-2 catalog (Vanderspek et al. 2008), we determine a 10σ threshold curve for HETE-2 bursts. If we consider associated sets (see, e.g., Petrosian 1993) of data points above the HETE-2 threshold (e.g., dashed line regions), the correlation in the Swift data is low (2.4σ). The significance is still only 2.7σ if we lower the estimated HETE-2 threshold by a factor of 2, an appropriate amount to account for error in our threshold estimate (e.g., jagged line). HETE-2 events with measured z 's are marked with an X. (B) Randomly sampling z 's from Swift bursts, we show one iteration placing Swift points from (A) onto the $E_{\text{pk}}-E_{\text{iso}}$ plot. In 10^4 iterations, only $(20 \pm 3)\%$ of Swift event below the HETE-2 threshold are consistent with the relation found by Amati et al. (2002).

cal expected behavior for GRB correlations where X or Y are typically much more broadly distributed than Z . However, artificial correlations in the case m_{obs} may also occur (e.g., consider Willingale et al. 2007), depending on the algebraic arrangement of observables going into X and Y .

The scatter in Y about the best-fit regression line is

$$\sigma_{\text{src}}^2 = \langle (Y - m_{\text{src}}X)^2 \rangle + (a - m_{\text{src}}b)^2 \langle Z^2 \rangle. \quad (2)$$

Covariance terms between X and Z or between Y and Z — which generally act to decrease the scatter in the source frame as compared to the observer frame — are dropped because we assume the correlation is non-intrinsic. To be explicit, rearrangement of an equation like $E_{\text{pk}} \propto E_{\text{iso}}^{0.5}$ shows that the observables should vary with z , and this implies cross terms in Equation 2. The cross terms go to zero if the correlation is due to (z -independent) observational effects.

The first term in Equation 2 is bound from the bottom by σ_{obs}^2 ; equality is obtained for $m_{\text{src}} = m_{\text{obs}}$. The second term in Equation 2 implies an increase in σ_{src}^2 over σ_{obs}^2 unless, coincidentally, $a = m_{\text{src}}b$. Interestingly, this “coincidence” is satisfied nearly or exactly for many of the known GRB correlations.

As an example, consider the Ghirlanda et al. (2004) correlation: $E_{\text{pk}} \propto E_{\gamma}^{0.7}$, where $E_{\gamma} \propto E_{\text{iso}}\theta_{\text{jet}}^2$ is the energy release corrected for beaming into a jet of angle θ_{jet} . For the case of a uniform density medium surrounding the GRB (Sari et al. 1999), $\theta_{\text{jet}}^2 \propto E_{\text{iso}}^{-1/4}(1+z)^{-3/4}T_{\text{jet}}^{3/4}$, where T_{jet} is the time in the observer frame where the jetting effects are expected to become apparent. Substituting into equations 1 and 2, it follows that $X = 3/4 \log(S_{\text{bol}}T_{\text{jet}})$, $Y = \log(E_{\text{pk,obs}})$, $a = 1$ and $b \approx 3/2$. Hence, $a \approx m_{\text{src}}b$, as also demonstrated graphically in Figure 3A. The Ghirlanda et al. (2004) correlation is, therefore, effectively independent of the measured spectroscopic redshifts.

For the data in Figure 3A, $\sigma_{\text{obs}} = 0.17$ dex and $\sigma_{\text{src}} = 0.16$ dex (i.e., 0.9σ insignificant decrease from an F-test). The source frame data are correlated strongly with Kendall’s $\tau_K = 0.7$. However, data generated with randomized redshifts (e.g., red points in Figure 3A) exhibit larger τ_K in most (76%) of the simulations. Somewhat more encouraging, we find a marginal 2σ increase in significance in $\sigma_{\text{src,sim}} = 0.18 \pm 0.01$ dex. However, the increase is apparently dominated by 1–2 outlier events in a fraction of the simulations; if we instead employ the median absolute deviation about the median of the fit residuals as a robust measure of scatter, 52% of simulations exhibit lower scatter than for the observed data.

For comparison with the previous section and considering the 47 normal, long-duration bursts in Amati et al. (2006) for the $E_{\text{pk}}-E_{\text{iso}}$ relation, we find a significant change in scatter and slope only if we include two events at $E_{\text{iso}} < 2 \times 10^{51}$ erg. Using the outlier-resistant measures of correlation (τ_K or scatter estimated via the median absolute deviation about the median) we find that 10% of simulations yield a better correlation than the observed correlation, independent of whether the two XRFs are excluded. Hence, there is only very weak (1.5σ) evidence from our tests to favor an intrinsic explanation for the $E_{\text{pk}}-E_{\text{iso}}$ correlation using the bright Amati et al. (2006) data. Exclusion of a small number of outliers is common practice (e.g., Amati et al. 2006); however, it

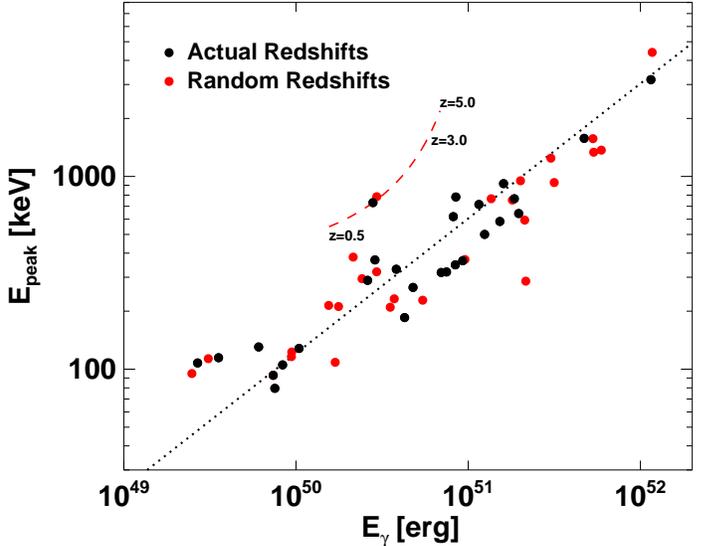


FIG. 3.— Ghirlanda et al. (2004) find a tight correlation between the beaming-corrected GRB energy release E_{γ} and E_{pk} . We also plot recent points from Schaefer (2007). Suggesting a non-intrinsic origin similar to the correlations in Figure 2, it is not necessary to know the z of an event to acceptably place it on the best-fit curve (dotted line). The dashed red curve shows the trajectory with z , holding the other observables fixed. The tightness of the correlation is unaffected whether true or random redshifts are assumed for translating the observer frame to source frame quantities.

may be reasonable to criticize us for allowing the outlier events to be flagged during and not before the simulations.

As an interesting side-note which may further suggest a paucity of intrinsic physics in Ghirlanda et al. (2004) type relations, the balancing with z is characteristic of similar relations found assuming instead a wind-stratified medium (Nava et al. 2006) or even allowing θ_{jet} to vary arbitrarily with T_{jet} and E_{iso} (Liang & Zhang 2005). The latter relation has the smallest σ_{src} , which leads us to suspect that the only reason that a very large number of correlations do not abound in the literature is that we typically restrict to those that can be interpreted physically, or, short of this, we require the observables to evolve with z in a physically plausible fashion (e.g., $T_{\text{jet,src}} = T_{\text{jet}}/[1+z]$).

Without these constraints, it is possible to invent highly-significant but absurd relations (e.g., redshifted Swift trigger number versus redshifted burst duration) which make a mockery of all correlation studies. This effect alone is likely not sufficient to generate the tightness of the Ghirlanda et al. (2004) correlation, however.

The tightness of the Ghirlanda et al. (2004) correlation, if it is not intrinsic, must stem additionally from the effects outlined in the previous sections for the functionally similar $E_{\text{pk}}-E_{\text{iso}}$ correlation. Excluding GRB 970508 as in Ghirlanda et al. (2004), we find that the $E_{\text{pk}}-E_{\text{iso}}$ relation has only a slight increase in scatter (0.18 dex) relative to the $E_{\text{pk}}-E_{\gamma}$ relation (0.16 dex), considering the same data (Figure 3). In fact, because the inferred T_{jet} values are very narrowly distributed and E_{γ} is proportional to E_{iso} to a power less than unity, a $E_{\text{pk}}-E_{\gamma}$

relation is potentially always tighter than a $E_{\text{pk}}-E_{\text{iso}}$ relation, depending on how the scatter is measured.

4. DISCUSSION

The B07 Swift BAT and Sakamoto et al. (2005) HETE-2 GRB samples both exhibit a statistically significant correlation in the observer frame between S_{bol} and $E_{\text{pk,obs}}$. What creates these correlations? The correlations follow the trigger threshold limits (Figure 2A) in both cases, and the correlation significances drop precipitously when we account for the flux limits (Section 2). Therefore, flux limits must play a strong role shaping the observed correlation, perhaps giving them most of their statistical significance.

What does this tell us about the source frame $E_{\text{pk}}-E_{\text{iso}}$ relation? Because the redshift dependence in the transformation from $E_{\text{pk,obs}}-S_{\text{bol}}$ to $E_{\text{pk}}-E_{\text{iso}}$ is weak, the source frame correlation is likely to have the same origin as the observer frame relation (Section 3). Indeed, we observe a strong decrease in the source frame correlation significance when account for the Swift threshold.

Finally, what have learned about the $E_{\text{pk}}-E_{\text{iso}}$ relation as it appears in Swift data relative to the pre-Swift $E_{\text{pk}}-E_{\text{iso}}$ relation? The normalization appears to be strongly instrument-dependent, and this suggests the pre-Swift normalization is defined largely by pre-Swift satellite flux limits. B07 find that the Swift $E_{\text{pk}}-E_{\text{iso}}$ relation has a lower flux normalization and more scatter relative to pre-Swift $E_{\text{pk}}-E_{\text{iso}}$ relations. We show above that we can raise this normalization (and decrease the scatter about the relation) to a level consistent with the pre-Swift $E_{\text{pk}}-E_{\text{iso}}$ relation by imposing a heightened flux limit corresponding to detection by a satellite of HETE-2-like sensitivity.

Strictly speaking, from these observations we can only rule out the existence of a narrow relation between E_{pk} and E_{iso} , while some physical correlation between the quantities may be present at high flux levels (see, e.g., the paucity of bright events in Figure 2B). In any case, future satellites more sensitive than Swift are expected to further shift and broaden the $E_{\text{pk}}-E_{\text{iso}}$ relation into an inequality.

More strongly, because we can probably attribute the slope (see, Section 3), scatter, and normalization of the correlation to selection effects related to a photon flux F cutoff and the functional correlation between $E_{\text{pk,obs}}$ and $S_{\text{bol}} \propto FE_{\text{pk,obs}}$ (Section 1; also, Massaro et al. 2007), an intrinsic explanation for the correlation in the Swift data may be unnecessary. Our results also call into question any intrinsic explanation for the correlation in pre-Swift data or in surveys only including the brightest Swift events (e.g., Sakamoto et al. 2008), because an $E_{\text{pk}}-E_{\text{iso}}$ relation should be instrument-independent. The existence of faint Swift events (in the observer frame) mostly missing from pre-Swift surveys shows that we do not understand what truncates the data in those surveys (see, also, Ghirlanda et al. 2008).

We could be incorrect in drawing these conclusions if: (a) our spectral fits are systematically incorrect and there are, in fact, no GRBs at flux levels below HETE-2 threshold levels, or (b) we have estimated the HETE-2 threshold incorrectly by a factor > 2 . The quality of the spectral fits is addressed extensively in B07, where direct consistency is established relative to observations from

Konus-WIND or Suzaku of bursts also detected by Swift. We note that Bellm et al. (2008) have independently verified the statistical analysis in B07 for several bursts also observed by RHESSI. Assumptions in B07 regarding the GRB spectra (see, Section 2), useful in compensating for the narrow BAT bandpass when determining bolometric fluences, are conservative in that they err toward large $E_{\text{iso}}^{0.5}/E_{\text{pk}}$ values. It also seems unlikely that the HETE-2 threshold should have any bearing on the censoring of Swift GRBs, considering also that many authors have reported BATSE GRBs detected below HETE-2 fluence limits (Nakar & Piran 2005; Band & Preece 2005; Kaneko et al. 2006).

For (b), we have shown that the $S_{\text{bol}}-E_{\text{pk,obs}}$ correlation significance, after accounting for the HETE-2 threshold, does not increase if we lower our threshold estimates by a factor of two. A larger error than this on our part for many GRBs is extremely unlikely given the straight-forward nature of the threshold calculation (Section 2). Moreover, such a substantial increase in HETE-2 sensitivity would make HETE-2 as sensitive as Swift, which is demonstrably incorrect given the dramatic difference in the GRB localization rate of the two missions ($\sim 20/\text{yr}$ for HETE-2 and $\sim 90/\text{yr}$ for Swift). Given the mean peak photon flux to energy fluence ratio from Sakamoto et al. (2005, $10^{8.0} E_{\text{pk,obs}} \text{ ph s}^{-1} \text{ erg}^{-1} \text{ keV}$), our HETE-2 threshold estimate is within 50% of that estimated in Band (2003).

Finally, there is strong indication that flux limits associated with spectroscopic redshift determination are important to estimate and consider when restricting to GRBs with measured redshift (e.g, as in Ghirlanda et al. 2008). In Figure 2 we mark the HETE-2 events with measured redshift; these are on average a factor two brighter relative to threshold than the events without measured redshifts.

The quality of the HETE-2 GRB localization depends on the brightness of the GRB as does likely also the brightness of the optical transient; both effects contribute to whether the afterglow and host galaxy can be detected. Additional complicated selection effects related to determining the GRB z are discussed in Bloom (2003). These are expected to be less important for Swift due to arcsecond X-ray localizations.

We stress that a correlation is not de facto intrinsic simply because trigger thresholding can be ruled out as influencing the surveys. Other flux limits may dominate (and probably do dominate when the samples are restricted to GRBs with measured redshift). A crucial step in utilizing pre-Swift data to rule on the nature of correlations is to establish completeness in the surveys.

4.1. New Correlations: Handle with Care

To gauge the importance of systematic effects in this area of research, we have isolated from the literature two paths that likely have lead to apparently highly significant correlations: (1) selection effects truncate the data in various ways and the “missing” data are not treated; or (2) partial correlation with a hidden variable or variables is ignored. In this paper, we have studied type (1) errors. However, both type (1) and (2) errors can be unmasked using the tools outlined above.

Type (2) errors potentially arise in studies which employ one correlation (e.g., Lag-Luminosity, Norris et al.

2000) to infer z for another correlation with similar variables (e.g., E_{pk} -Luminosity), without controlling for partial correlation with the variable in common (see, e.g., Lloyd-Ronning & Ramirez-Ruiz 2002; Kocevski & Liang 2006).

For example, from BATSE satellite observations where the detection threshold is very well characterized in terms of peak luminosity, we know that the E_{pk} -Luminosity correlation is largely formed in the detection process (Lloyd, Petrosian, & Mallozzi 2000). However, this potential strong bias is ignored when Yonetoku et al. (2004) consider a purely intrinsic E_{pk} -Luminosity to derive a significant evolution in the rest-frame properties of GRBs (e.g., the correlation between Luminosity and z). The potential instrumental origin of the E_{pk} -Luminosity relation is also ignored by Schaefer (2007).

We recommend testing potential new correlations (or known correlations not explicitly mentioned here) by proving:

1. $\sigma_{\text{obs}} > \sigma_{\text{src}}$, and
2. the correlation quality — scatter or significance using outlier resistant measures — degrades when the z 's are randomized and the correlations are recalculated, and
3. the observables, grouped to one side of the correlation equation, vary with z as predicted.

These tests are to establish basic confidence in an intrinsic nature for the correlations and are not merely to establish that a correlation can be used to estimate z (e.g., Li 2006; Schaefer & Collazzi 2007).

For these tests to be accurate, however, it is also necessary to identify selection effects acting on the data. Flux or fluence values should be present down to the established survey completeness level. Well-established methods can then be applied to compensate for data truncation (Section 2) and to control for partial correlations (e.g., Akritas & Siebert 1996). Covariance between the measured quantities, if present, must also be treated (e.g., Lee & Petrosian 1996; Cabrera et al. 2007, B07).

5. CONCLUSIONS

We show above for GRBs observed with multiple satellites that four example correlations reported in the literature have features (instrument-dependent normalizations, weak z -dependence, etc.) indicative of strong contamination by or even an origin in selection effects. Contrarily, there is no widely-accepted, non-a-posterior explanation for the correlations in the source frame. Also, the a-posteriori theoretical explanations (e.g., Eichler & Levinson 2004; Schaefer 2007) fix the correlation normalizations by requiring GRBs to have one intrinsic spectrum or to be standardizable candles (e.g., a narrow E_{γ} distribution; see, Bloom, Frail, & Kulkarni 2003), a hypothesis no longer well-supported by the data (see, Kocevski & Butler 2008).

The common z -independence of GRB correlations is either very odd or damning. As we discuss, this is one characteristic of a tight, apparent source-frame correlation which arises purely due to selection effects. For correlations between luminosity and some other measured quantity that does not depend on luminosity distance,

redshift balancing is likely not a consequence of intrinsic physics, because the GRB cannot possibly know how the distance to the observer should vary with z . However, redshift balancing might be expected for correlations between luminosities (e.g., the recently reviewed $E_{\text{iso}}-L_{X,10}$ correlation, Nysewander et al. 2008).

We note that redshift balancing trivially explains the relative slopes of the $E_{\text{pk}}-E_{\text{iso}}$ and $E_{\text{pk}}-E_{\gamma}$ relations (Section 3), something theory can apparently do as well but only with reference to complicated biases stemming from GRB beaming (Levinson & Eichler 2005).

As an obvious point, we caution that the balancing of $1+z$ terms implies the correlations cannot be used to infer z . More speculatively — because redshift balancing allows a non-intrinsic observer-frame correlation to appear in the source frame as a low-scatter correlation — this balancing may have played a role in the discovery of potential intrinsic correlations. If the correlations currently known have been selected in place of correlations that do not balance redshift, and if the fitting of GRB properties and the pruning of outliers has been conducted with these correlations in mind, then using these correlations to construct a Hubble diagram and test concordance cosmology (e.g., Ghirlanda et al. 2005; Schaefer 2007) is circular.

We stress that these redshift dependency problems — and our interpretation of them — are unique to and potentially only characteristic of high- z objects like GRBs (as opposed to, e.g., SNe) where there is essentially no low- z calibration, and luminosity distance must generally be calculated using z and assuming a cosmology model.

Small sample sizes may also have played an important role in allowing for tight apparent correlations through over-fitting. The correlation involving Luminosity, E_{pk} , and T_{45} duration for 22 GRBs in Firmani et al. (2006), which does not appear to suffer from redshift balancing (see Figure 7 in Firmani et al. 2006), is an interesting case. When looking at larger datasets, however, there appears to be increased scatter and no statistically significant improvement in scatter relative to the E_{pk} -Luminosity (Collazzi & Schaefer 2008) or $E_{\text{pk}}-E_{\text{iso}}$ (Rossi et al. 2008) correlations. Investigation of the correlations in the largest possible datasets is, therefore, critical.

The GRB community is no longer starved for data. The next critical step toward uncovering intrinsic correlations is to combine all available data by establishing sample completeness in pre-Swift surveys (e.g., Ghirlanda et al. 2008) and treating the dominant flux truncations using methods like those outlined above.

In looking for new relations relevant to the physical processes underlying GRBs, and to avoid an inherent difficulty in deciding an origin for the correlations in the source or observer frames (Section 3), it may be important to choose observables less-broadly distributed than the characteristic range in $1+z$ or to abandon observables with strong and complicated truncations (i.e., fluxes or fluences) altogether. Functional correlations in $1+z$ could be minimized by choosing observables (e.g., powerlaw indices) which do not vary explicitly with z . Most directly and circumventing all concerns raised above, we should establish calibration for GRBs at $z \lesssim 0.1$. It is, therefore, also crucial to understand whether GRB prop-

erties evolve with redshift.

N.R.B gratefully acknowledges support from a Townes Fellowship at the U. C. Berkeley Space Sciences Lab-

oratory, as well as partial support from J.S.B. and A. Filippenko, as well as support from a GLAST/Einstein Fellowship. J.S.B. is partially supported by the Hellman Faculty Fund.

REFERENCES

- Akritas, M. G., & Siebert, J. 1996, *A&A*, 278, 919
 Amati, L., et al. 2002, *A&A*, 390, 81
 Amati, L. 2006, *MNRAS*, 372, 233
 Amati, L. 2008, GCN #7612
 Barthelmy, S. D., et al. 2005, *Space Science Reviews*, 120, 143
 Band, D. L. 2003, *ApJ*, 588, 945
 Band, D. L. 2006, *ApJ*, 644, 378
 Band, D. L., & Preece, R. D. 2005, *ApJ*, 627, 319
 Bellm, E. C., et al. 2008, *AIP Conf. Proc. GRB 2007*, 1000, 154
 Bloom, J. S. 2003, *AJ*, 125, 2865
 Bloom, J. S., Frail, D. A., & Kulkarni, S. R. 2003, *ApJ*, 594, 674
 Burrows, D. N., et al. 2005, *Space Science Reviews*, 120, 165
 Butler, N., et al. 2007, *ApJ*, 671, 656
 Cabrera, J. I., et al. 2007, *MNRAS*, 382, 342
 Collazzi, A. C., & Schaefer, B. E. 2008, arXiv:0808.2061, *ApJ*, in press
 Efron, B., & Petrosian, V. 1999, *J. of Am. Stat. Assoc.*, 94, 824
 Eichler, D., & Levinson, A. 2004, *ApJ*, 614, L13
 Firmani, C., et al. 2004, *ApJ*, 611, 1033
 Firmani, C., et al. 2006, *MNRAS*, 370, 185
 Friedman, A. S., & Bloom, J. S. 2005, *ApJ*, 627, 1
 Gehrels, N., et al. 2004, *ApJ*, 611, 1005
 Gelman, A., et al. 2004, *Bayesian Data Analysis* (2nd ed.; Boca Raton:Chapman & Hall/CRC)
 Ghirlanda, G., Ghisellini, G., & Lazzati D. 2004, *ApJ*, 616, 331
 Ghirlanda, G., et al. 2005, *ApJ*, 613, L13
 Ghirlanda, et al. 2008, *MNRAS*, 387, 319
 Graziani, C., Donaghy, T. Q., & Lamb, D. Q. 2005, *Il Nuovo Cimento C*, 28, 681
 Kaneko, Y., et al. 2006, *ApJS*, 166, 298
 Kendall, M. G. 1938, *Biometrika*, 30, 81
 Kistler, M. D., et al. 2008, *ApJ*, 673, L119
 Kocevski, D., & Liang, E. 2006, *ApJ*, 642, 371
 Kocevski, D., & Butler, N. 2008, *ApJ*, 680, 531
 Lee, T. T., & Petrosian, V. 1996, *ApJ*, 470, 479
 Lamb, D. Q., Donaghy, T. Q., & Graziani, C. 2005, *ApJ*, 620, 335
 Levinson, A., & Eichler, D. 2005, *ApJ*, 629, L13
 Li, L. X. 2006, *MNRAS*, 374, L20
 Li, L. X., 2007, *MNRAS*, 79, L55
 Liang, E., & Zhang, B. 2005, *ApJ*, 633, 611
 Lloyd, N. M., & Petrosian, V. 1999, *ApJ*, 511, 550
 Lloyd, N. M., Petrosian, V., & Mallozzi, R. S. 2000, *ApJ*, 534, 227
 Lloyd-Ronning, N. M., & Ramirez-Ruiz, E. 2002, *ApJ*, 576, 101
 Massaro, F., Cutini, S., Conciatore, M. L., & Tramacere, A. 2007, arXiv:0710.2226
 Nakar, E., & Piran, T. 2005, *MNRAS*, 360, 73
 Nava, L., et al. 2006, *A&A*, 450, 471
 Norris, J., et al. 2000, *ApJ*, 534, 248
 Nysewander, M., Fruchter, A. S., & Peér, A. 2008, arXiv:0808.2610
 Petrosian, V. 1993, *ApJ*, 402, L33
 Preece, R. D., et al. 2000, *ApJ*, 126, 19
 Rees, M. J., & Mészáros, P. 2005, *ApJ*, 628, 847
 Rossi, F., et al. 2008, *MNRAS*, 388, 1284
 Schaefer, B. E. 2007, *ApJ*, 660, 16
 Schaefer, B. E., & Collazzi, A. C. 2007, *ApJ*, 656, L53
 Sakamoto, T., et al. 2005, *ApJ*, 629, 311
 Sakamoto, T., et al. 2008, *ApJ*, 679, 570
 Sari, R., Piran, T., & Halpern, J. P. 1999, *ApJ*, 524, L43
 Vanderspek, R., et al. 2008, *HETE-2 catalog*, in prep.
 Willingale, R., et al. 2007, arXiv:0710.3727
 Yamazaki, R., Ioka, K., & Nakamura, T. 2004, *ApJ*, 606, L33
 Yonetoku, D., et al. 2004, *ApJ*, 609, 935