Krittika Summer Projects 4.0

Faintest of the Brightest: GRB Hunters

Final Report

Soham Mitra
M.Sc Physics, IIT Bombay

Mentors: **Gaurav Waratkar & Meghna Dixit**

Abstract

Gamma Ray Bursts (GRBs) have been at the forefront of astrophysical research in the 21st century. They are caused from various phenomena: imploding stars in supernovas to merging neutron stars in kilonovas. GRBs are also accompanied with X-rays and other lower energy electromagnetic radiation, each of which can/may be generated by different processes. In this project, we study the basics of the formation of GRBs and the detected of GRBs in the X-ray data collected by CZTI instrument onboard India's firsts astronomical space telescope, *AstroSat*. Our goal is to comb through the CZTI data to find transient events on the order of 1 to 10 seconds and make algorithms which may improve the detection of such GRBs.

Contents

1 Introduction

1.1 Gamma Ray Bursts: What and Why

Gamma Ray Bursts (GRBs) are extremely high energy events that are, true to its name, very powerful bursts of gamma rays. They are transient events occurring ranging from a few hundred milliseconds to a few hours. Any burst of astrophysical origin with single photon energies of $> 10,000$eV is considered to be a Gamma Ray Burst. Thus, GRBs contain not only gamma rays but X-rays as well. Moreover, many GRBs have an after glow which produces electromagnetic radiation of lower energies including softer X-rays, UVs and so on.

A Brief History During the cold war, after the USA and the USSR along with other countries signed the Partial Test Ban Treaty for testing nuclear weapons, the US government for ensuring that countries did comply with the treaty started the Vela programme. These network of Vela satellites were supposed to detect any tell tale gamma ray signatures as produced by a non-underground nuclear explosion. In the process, however, they discovered gamma rays of extra-terrestrial origin. In 1969, Klebesadel started probing into these unexplained signals that they had been detecting. After thoroughly analysing the data, in 1973, Klebesadel along with Strong, and Olson published their findings in the Astrophysical Journal: *Observations of Gamma-Ray Bursts of Cosmic Origin.* This was the beginning of these ever expanding field of gamma ray and X-ray research in astronomy.

Over the years, many new missions were specifically designed to observe such events. Some of these *Beppo-SAX*, *CRGO* were game changing in their abilities to localize the events in the sky. Nowadays, almost every major space agency have their own gamma ray and X-ray space telescopes to probe for these phenomena.

Sources and Afterglow Very soon after the discovery, physicists started the hunt to find the source of the bursts. Many candidates were established: supernovas where a star implodes into a neutron star or other type, where a neutron star exceeds its Chandrashekhar mass limit by absorbing too much mass from a bigger companion star. These were explicitly proven to be some of the true sources through using other telescopes like the Hubble and several Earth based telescopes to look at the source in visible light. These discoveries gave the astronomers invaluable information about the nature of the supernovas. It also marked the beginning of multi-messenger astronomy where one event was probed using the entire electromagnetic spectrum.

Afterglows were theorized to form with slowly fading signatures of X-rays. Soon, it was confirmed with the help of corresponding ground and space observations of the same object. These afterglows are formed by the emission of X-rays and other lesser energy radiation from the interaction of the burst ejecta with the interstellar gases. They are extremely useful in characterizing the composition of the gas giving a glimpse into the stellar evolution and formation of galaxies.

However, a few bursts remained unexplained, especially the very short bursts of less than 2 seconds. It was theorized that they are formed due to kilonovas - collision of two neutron stars or one with a black hole. It took many years to definitively proof that such events actually occur. In 2017, the short GRB 170817A was detected by the Fermi Gamma-ray Space Telescope only 1.7 s after the detection of the gravitational wave GW170817 by the LIGO and the Virgo collaborations. This established the definitive proof of neutron star mergers as well as the short GRBs associated with it. The afterglow was also detected by other telescopes including the Swift Gamma-Ray Burst Mission (in UV spectra), Chandra X-ray Observatory and the Hubble Telescope (in visible light). This event expanded the multi messenger astronomy in a whole new direction, opening the skies to very large number of new events being detected using their gravity effects and their radiation.

1.2 CZTI: India's Eyes for Hard X-rays

AstroSat, India's first space telescope was launched in 2015. It is a collaboration between many different institutes in India. It provides a wide coverage of electromagnetic spectrum having instruments for optical, UV, soft and hard X-rays. They provide a wide array of detection capabilities to probe different astrophysical phenomena. Among these, the Cadmium Zinc Telluride Imager (CZTI) is to observe hard X-rays. It has excellent detection capabilities: sensitive to $20 - 200$keV X-rays, coupled with a coded aperture mask to obtain 80 resolution for sources over a $4.6° \times 4.6°$ field of view. CZTI is involved in the measurement of curvature and reflection components in the spectra of Active Galactic Nuclei and Xray binary systems, the study of Quasi-Periodic Oscillations at hard X-ray bands in accreting neutron star and black hole systems, the characterisation of hard X-ray spectra of magnetars as well as the detection of gamma ray bursts and the study of their early light curves.

CZTI over the years have detected hundreds of GRBs many which are completely new, not detected by any other telescope. It has detected all kinds of GRBs from very short transients to very long ones. Due to its comparative low background noise, it has been able to detect many broad band faint GRBs which many other space telescopes have failed to detect.

2 Hunting GRBs

Our task in this project has been to comb through the light curves produced by the CZTI to detect GRBs. We have been exposed to some of the techniques used in the automatic detection of the events from the data. We have to characterize the nature of the signal, its quality and determine how well the detector have been able to observe the event. In practice, most of the GRBs are automatically flagged by the CIFT (CZTI Interface for Fast Transients) system which henceforth are verified by a human. Currently, this system is pretty efficient in flagging the GRBs. However, the number of false positives are also high and thus any confirmation comes after human verification. This makes the process of detection pretty time consuming. The team under Prof Varun Bhalerao at IIT Bombay, has been working to improve the detection capabilities and employ them in future missions.

As an extension, I hope to contribute to this present endeavour in improving the detection capabilities of CZTI, however little.

2.1 Extraction of Data: Using the CZTI Pipeline

So, the first step in detection, is the extraction of the data from the CZTI. It happens in three steps:

- Level 0: Raw data received from satellite telemetry, segregated by instrument along with auxiliary data.
- Level 1: Reorganized raw data, written in FITS format for Astronomical use.
- Level 2: Standard science products derived from Level 1 data.

This level 2 data is also in FITS format. It is generated by the CZTI pipeline: a collection of software modules to process the level 1 data. Some of the level 2 products are already available on the AstroBrowse website. We are provided with the `*_level2.mkf`, `*M0_level2_bc.evt`, `*M0_level2_bc_livetime`.fits where the `*` is the name of the observation file like `AS1A10_073T02_9000004522_31247czt`. We are also provided with a calibration file which provides with the various parameters of the instrument.

We then use the following modules to process the data into light curves of specified time bins:

- `cztgtigen`
- `cztdatasel`
- `cztpixclean`
- `cztevtclean`
- `cztflagbadpix`
- `cztbindata`

In order to process the data through the above modules, I have written a Python script which performs all these steps for a given set of observation to fasten the process. The code of which is given in the appendix A.

2.2 Automating the Process of Detection

CIFT can detect the GRBs automatically. However, instead on relying on the trigger times produced by it for different GRBs, I have created a different way of automatically detecting the GRBs from the data. It has multiple steps. In brief, they are,

Binning	The data is binned into different time bins and energy bands for a comparative study.
Processing	A number of processing is done on the data.
Outliers	From the total stacked data, we find the outliers using a threshold.
Scoring	Each of the outliers are assigned scores based on some weights given to the count rates for the different energy bands.
Grouping	The corresponding outliers for each time bin is grouped together. The suspected GRBs are determined from these groups depending on how many of the time bins has the outlier.
GRB Width	The width (in time domain) of the GRBs is estimated from a given threshold.
SNR	The SNR of the estimated GRBs are found out.

2.3 Binning the Data

In order to detect the GRBs among the noise, we need to characterize the data in different time regimes. CZTI records the rate of X-ray photons falling on it. Dividing the data in the time domain into different time bins allows us to see the data in different resolutions. A comparative study of the signal in each bin size definitively tells us the about the quality and time span of the event.

We use the `astropy` module in Python to access the light curves in the FITS file and use plotting modules to display it. The binning is itself done by the `cztbindata` module in the previous step. It bins the data in two time bins, 0.1 s and 1 s and in three different energy intervals of 200 keV, 50 keV and 20 keV between $20 - 220$ keV. Thus, in total, $2 \times (1 + 4 + 10) = 30$ LC files are generated for each quadrant.

From this data, the remaining time bins, which are multiples of 0.1 s and 1 s are generated. 0.2 s to 0.9 s are generated from the 0.1 s data sets while 2 s to 10 s are generated from the 1 s data sets. This is done because there is some processing difference in the way `cztbindata` bins the 0.1 s and 1 s: the 1 s data produced is not an exact multiple that of 0.1 s.

The binning is done for so many energy bands and time bins so as to get a comprehensive insight into the nature of the GRB and determine it definitively, as is shown later on.

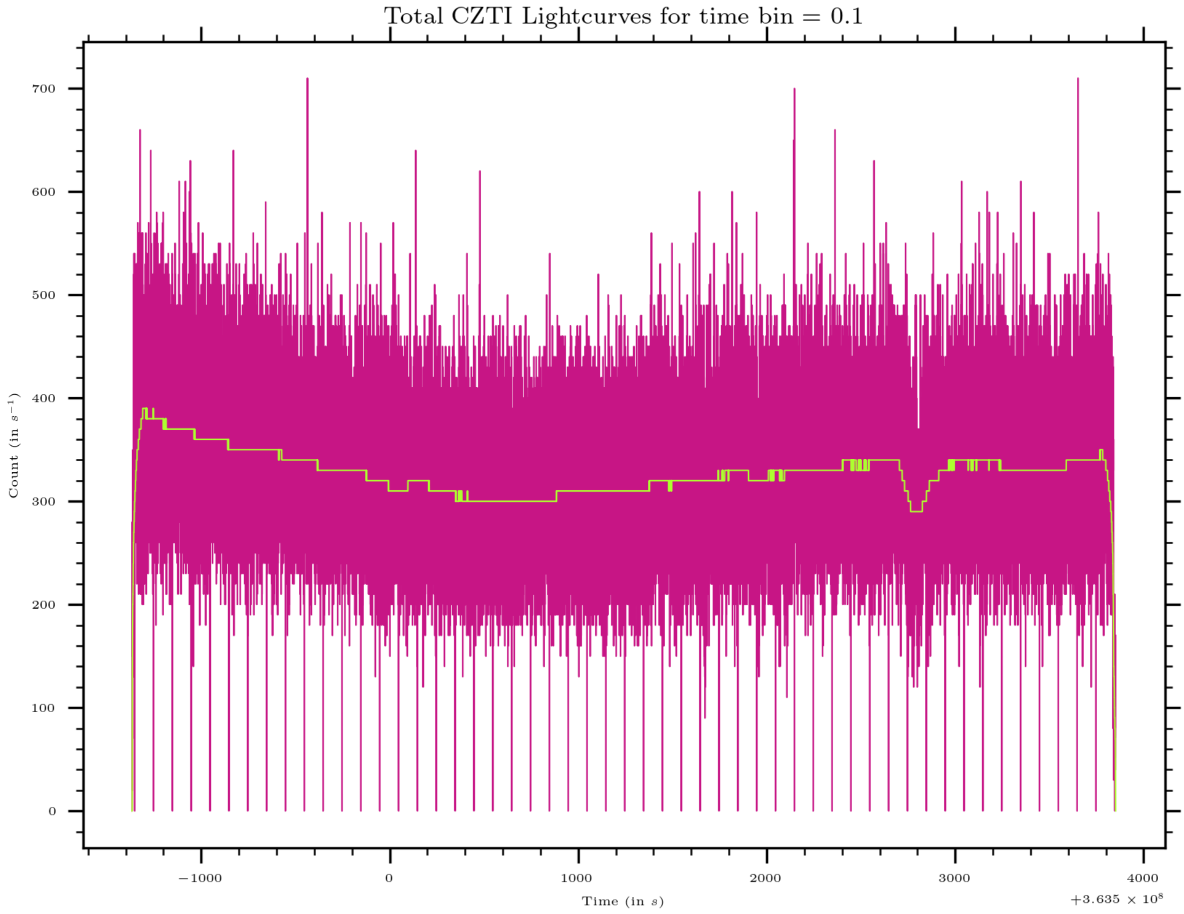

Figure 1: The Total stacked light curve for GRB 210709A. The yellow line is the background.

2.4 Processing the Data

The data, thus obtained after binning, has to be first processed before doing anything. Three steps are there:

1. Slicing the data away from SAA zones. The South Atlantic Anomaly is a zone over the South Atlantic ocean where the earth's magnetosphere is comparatively weak and as a result, there are elevated levels of X-ray activity from the interaction of cosmic rays with the atmosphere, as here they can reach much lower than otherwise possible. As a result of these extra noisy signals from this zone, the CZTI detector is switched off when it passes across this zone. The algorithm that is employed cannot detect any GRBs near the SAA. To prevent false positives, the data is manually sliced slightly away from the SAA zones, so that there is no effect of the increased particle events that will occur here.

2. Finding out the background signal. The background signal is found from the sliced data using median or a Savgol filter. The median filter is found to be more robust and one size fits all method and is thus, employed here in the processing. The filter window is chosen as 100 s as in CIFT.

3. Subtracting the background to get a signal centred across zero. We then subtract this background from the data that we find, to get the zero centred data. This enables us in making much better estimations of the statistical parameters of noise/signal.

All these steps are done for all the data bands. In case of the 50 keV and 20 keV intervals, the background is calculated for each of the energy bands. They are also sigma normalized for each of the bands.

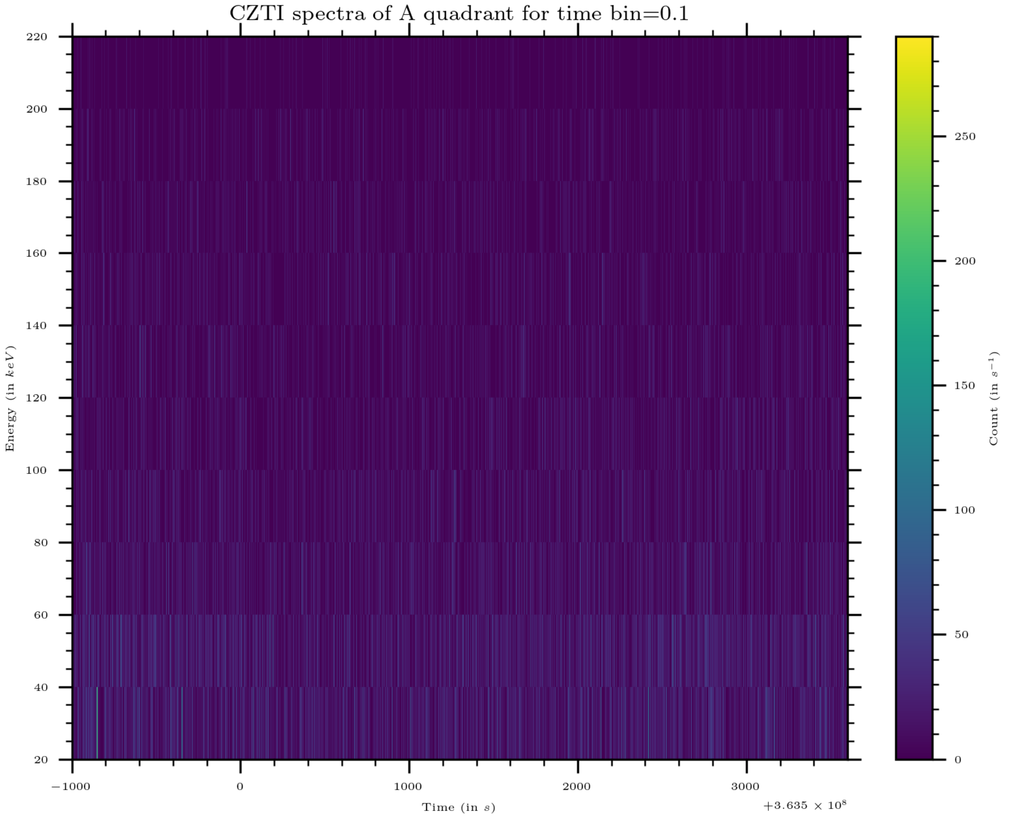

Figure 2: The raw spectra for quadrant A.

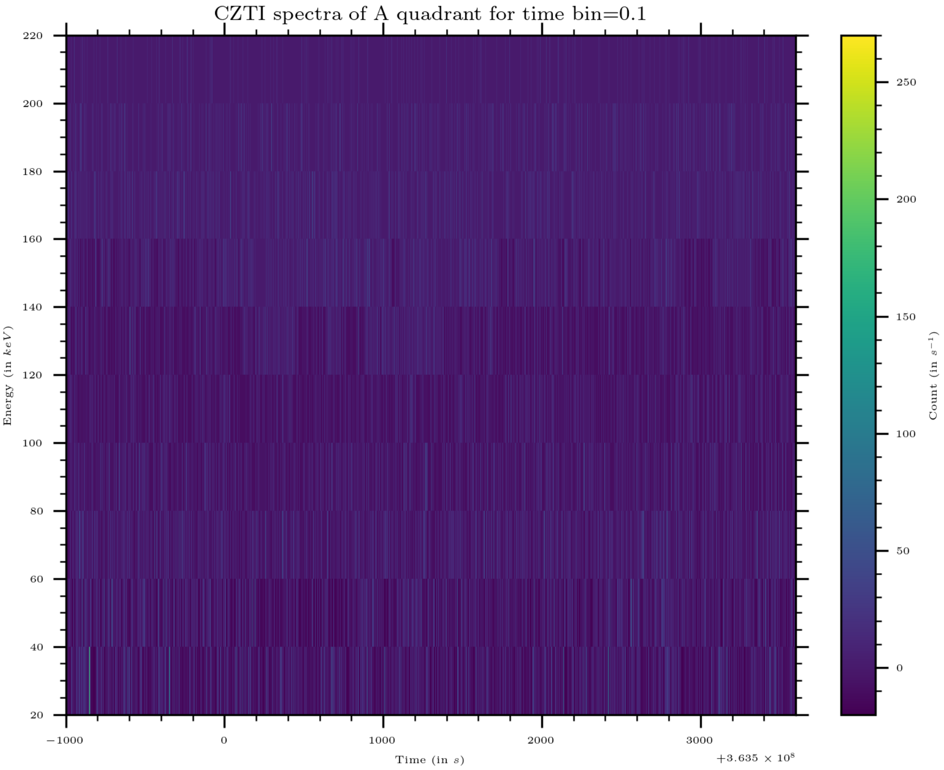

Figure 3: The background subtracted spectra for quadrant A.

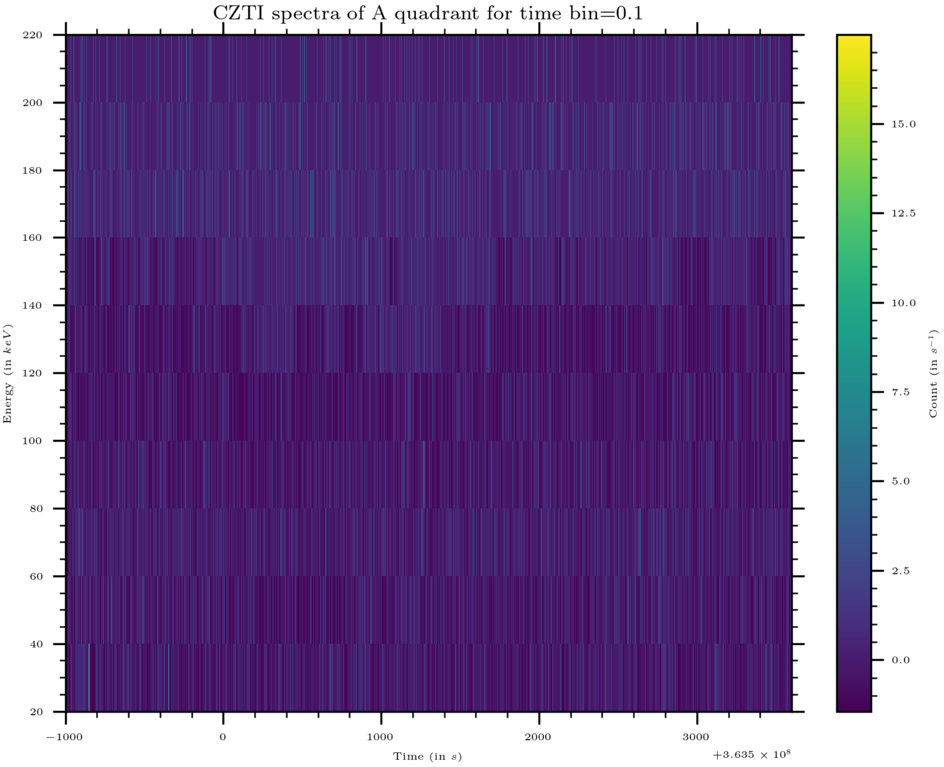

Figure 4: The sigma normalized spectra for quadrant A. The colour bar label is misleading to call it counts, it's merely indicating the relative counts, and hence should be unit-less. Rest of the quadrants and time bins are given in the note book in appendix A.

Moreover, all these processing is also done for the `total` stacked data from all four quadrants for each of the time bins and intervals individually. This enables us to find the outliers.

2.5 Finding out the Outliers

From the stacked data for each of the time bins of interval 200 keV, we find the outliers. Here, the outliers are such data points which are 4σ away from the mean, where σ is the standard deviation of the `total` stacked data. The choice of the 4σ is empirical and observation based. It can be improved on after running the programme on many different kinds of data of GRBs along with false positives. However, it has been found to work pretty well in the GRBs that are sharp or are faint, with short to mid time lengths.

2.6 Scoring the Energy Bands

Here, we come to another empirical values for the assignment of weights to the different energy bands. The justification of these different weights is also based on the fact that the lower energy bands will always contain noise while the upper ones can be filled only in case of particle event or a GRB event. The weights for the 50 keV energy bands are as 1.75 for the lowest and the rest are given weights of 2.75. In case of the 20 keV energy bands, the lowest two and the highest bands are given weights of 0.79 while all the rest in-between are given as 1.03. The highest band getting the lower weight is due to the fact that the detector is near its sensitivity limit in that band and as a result, may produce extra noise. The scores are calculated as,

$$\text{Score} = \sum_{i=1} \text{weight}_i \cdot \text{count}_i$$

2.7 Grouping the Suspected GRBs

Upon finding the scores in both the interval data sets, the outliers having a score more than 6 and 7 respectively for 50 keV and 20 keV intervals are found. These are treated as suspected GRB events. The data from each time bin are grouped together for the same event. It has then found out which groups has three or more than three time bins in which the score criteria is being fulfilled. This narrows down the suspected GRBs even more.

2.8 Finding the suspected GRB widths

We now turn to the `total` stacked data of the 200 keV energy band to find out the widths of the suspected GRBs. This is done by finding out how many data points on either side of the outlier is more than 2σ (another empirical threshold) from the mean of the stacked data. This region is considered to be the time when the suspected GRB is active.

2.9 Finding the SNRs

Upon finding the widths, we characterize rest of the data as noise. The standard deviation of the noise is found. Then the signal-to-noise ratio (SNR) is calculated as,

$$\text{SNR} = \frac{\text{Peak of the suspected GRB}}{\text{Standard deviation of noise}}$$

The SNR calculation is done for all the suspected GRBs. Since initially, not all the time bins are centred about the event i.e. the events can divided one or more time bins with the peak being divided into multiple bins. This reduces their standard deviation and as a result are not flagged as an outlier in all the different sized time bins. Once the SNR for the detected outliers are known, the time bin having the highest SNR is found out. This is treated as the time stamp for the event peak. All the rest of the time bins are re-centred in such a way that the event completely falls into one bin for that time interval. This maximizes the SNRs. The maximized SNRs vs the time bins gives us a definite way of telling whether the suspected GRB is indeed a gamma ray burst event or just a strong particle event.

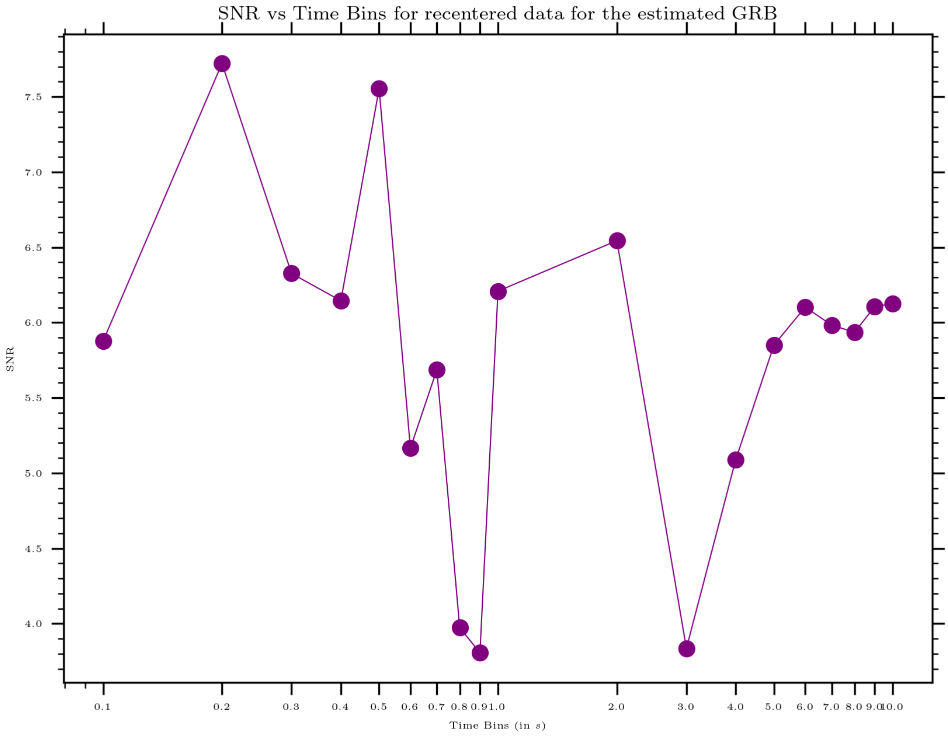

Figure 5: The SNR data for the GRB 210709A. Clearly, the SNR values are high consistently. There are dips because this GRB is actually, longer than 10 s at 14 s. The fact that the SNR value again increases towards the 10 s bins confirms the correctness of the analysis.

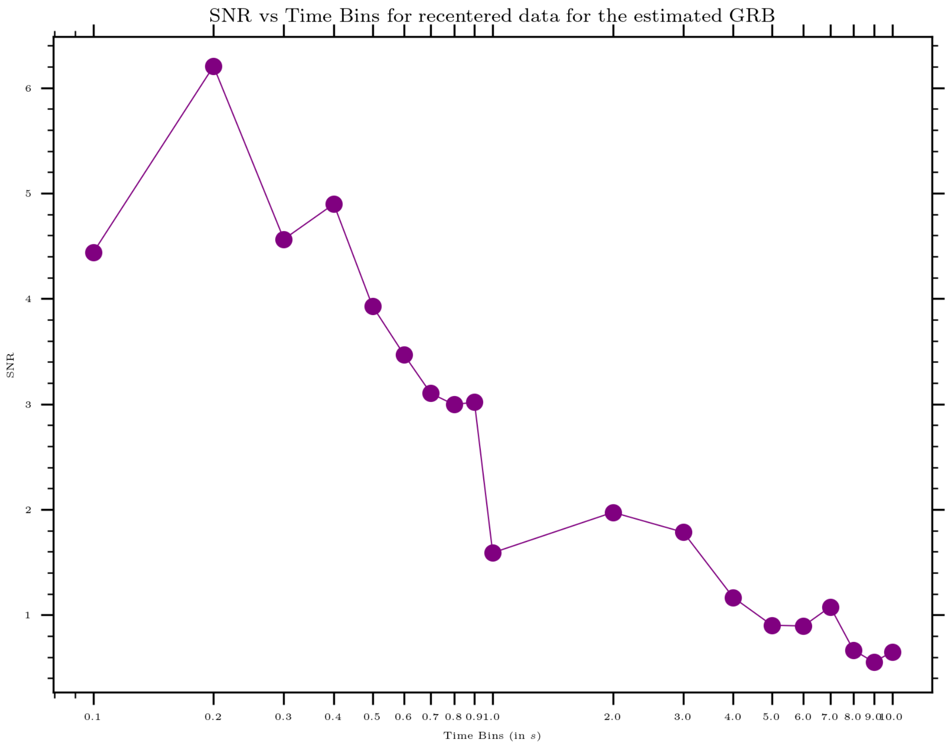

Figure 6: The SNR data for the falsely tagged GRB which is an actually a particle event of about 0.2 s. The values are generally low and falls rapidly away from that particular bin of 0.2 s which should not be the case for a true GRB.

For an actual GRB, the SNR is typically found to be high above 6 at least for 3-4 time bins which are comparable to the event duration. This starkly differs from that of a particle event where only the small time bins and that too only one or two have similar SNR values and then it falls off quickly for other time bins.

3 Comments on the Methodology

There are many points to be noted here regarding the techniques that have been used in this algorithm to finally find out the SNR.

- A lot of empirical estimates are made in the analysis of the data. Although it might seem random, the values are pretty good in determining a true event from particle events. Another advantage this technique has that the values can be modified from running the analysis on all past data and then adjusting the values to get even more accurate results faster. Further improvements, thus are possible to narrow down the events and the subsequent processing.
- Many statistical tools that have been used in the code are from the `scipy` package of Python. They work satisfactorily for the required use cases.
- The SNR data is not just calculated for the stacked data. It is also parallelly calculated for each of the quadrants individually. They are not used for any comparison

or filtering. But they too can be used for even finer filtering as is done in CIFT. The alternative approach used here, focusing mainly the quadrant-stacked data is something to be noted.

- The SNR data, as hypothesized in the beginning, is indeed very useful in filtering out the true events from the various particle events. It acts as a completely new parameter to characterise the GRBs and the strengths of the instrument.
- The SNR data also enables us to estimate the duration of the GRB. In this algorithm, it is mostly accurate for short GRBs less than 10 s because here it is the maximum time bin. Larger time bins and the corresponding SNR data will enable us to determine the SNR of longer GRBs as well.
- Although most of process from the generation of light curves to detection and calculation of SNRs is automated, the part of separating the data from the SAA is not. Its done manually on inspecting the data. Perhaps, this can as well be automated but currently, it is difficult to establish any such criteria, to eliminate the data from near the SAA. This task is more tough than it seems because of the non-definite boundaries of the SAA coupled with the unpredictability of the space weather.
- The above algorithm works very well when the data is produced from CZTI Pipeline Version 3. Its not determined whether it will be effective with older versions of the pipeline. Although it is suspected there needs to be drastic changes in values due to the worse noise cleaning in the previous versions of the pipeline.

4 Conclusion

Coming to the end of this small endeavour to improve the detecting capabilities, it is very clear that the signal to noise ratio is an excellent parameter in detecting true events from false ones. This can be revolutionary when employed on a large scale and the parameters improved upon from the past data. This will also enable in faster and accurate detections, which in turn helps in quick correlation with the data from other missions. This improves the amount of science that can be done with such data more productively. The SNR data also goes a long way in improving future detections which are aimed for such GRB hunting.

Appendix A

The `https://github.com/CaptSM271/CZTI-Data-SNR-Analysis.git` repository contains the Python script to produce the LC files from the level 1 data. It can be run only after installation of the CZTI Pipeline Version 3 in a Linux environment. Once that is done, the Jupyter notebook can be run to do the analysis. The name of the files has only to be edited and the clipping has to be done manually, then let the entire code run, the result will be the SNR data which definitively find the GRB. The Jupyter notebook's files are for the data for the GRB 210709A.

References

[1] Bloom, Joshua S. *What are Gamma-Ray Bursts?*, Princeton University Press, 2011

[2] `https://en.wikipedia.org/wiki/Gamma-ray_burst`

[3] `https://en.wikipedia.org/wiki/GW170817`

[4] V. Bhalerao, D. Bhattacharya et. al., *The Cadmium Zinc Telluride Imager on AstroSat*, Journal of Astrophysics and Astronomy, Volume 38, Issue 2, article id.31, June 2017

[5] Y. Sharma, V. Bhalerao, D. Bhattacharya et. al., *The search for fast transients with CZTI*, Journal of Astrophysics and Astronomy, Volume 42, Issue 2, article id.73, October 2021