GRB Hunters - The Faintest of The Brightest

Ananya Jaywant Nawale

KRITTIKA SUMMER PROJECTS 2023

GRB Hunters - The Faintest of The Brightest

Ananya Jaywant Nawale[1]

[1]Indian Institute of Technology Bombay, Mumbai, 400076, India

Contents

Understanding Gamma-Ray Bursts

1. Introduction to Gamma-Ray Bursts

1.1 Rough Definition

A Gamma-ray Burst (GRB) is a diverse and complex phenomenon characterized by the emission of gamma rays, representing the most energetic and luminous electromagnetic events observed since the inception of the Big Bang. These bursts exhibit a higher peak photon luminosity than any other events in the universe. They are primarily identified through their energy spectrum, known as the νF_ν or EFE spectrum, which reveals their actual flux output. Typically, GRBs display a distinct spectral peak energy in the hundreds of keV range. The emitted energy is concentrated predominantly in the X-ray and γ-ray range, while minimal radiation is observed in the more commonly observed optical window.

1.2 History of First Detections

Gamma-ray bursts (GRBs) were initially discovered as an unintended outcome of the Cold War. To address the escalating arms race and the global increase in radioactive fallout from nuclear weapons testing, Soviet Premier Nikita Khrushchev and U.S. President John F. Kennedy agreed to the Partial Test Ban Treaty. This treaty mandated that all nuclear weapons testing be conducted underground, prohibiting tests in oceans, the atmosphere, or space.

In response, the United States initiated the Vela Satellite Program to monitor compliance with the treaty using space-based surveillance. This initiative was based on our understanding of nuclear explosion physics. While the visible flash of a nuclear detonation could be concealed, the emission of X-rays, gamma rays, and neutrons during the initial second of the explosion are more challenging to hide. These by-products are referred to as the "signature" of a nuclear detonation. Plus conducting surveillance from space was necessary because the Earth's atmosphere blocks X-rays, gamma rays, and neutrons from reaching the surface.

On July 2, 1967, at 14:19 UTC, anomalous gamma radiation was detected by Vela 4 and Vela 3 satellites. Unlike typical nuclear weapon signatures characterized by a brief intense burst of gamma rays followed by decay, the recorded light curve exhibited two distinct peaks without the expected initial flash and decay. Solar flares and new supernovas were ruled out as possible explanations as neither had occurred that day.

Despite limited spatial resolution, subsequent investigations confirmed the extragalactic origin of these flashes and the detection of GRBs was publicly reported in 1973.

1.3   Variations in GRB lightcurves

Edo Berger defines gamma-ray bursts as short, intense, and non-repeating flashes of approximately mega-electron-volt gamma-rays with a wide range of spectral and temporal properties. The observations of gamma-ray bursts present a diverse array of light curves and spectra; however, akin to snowflakes or fingerprints, they also display shared characteristics that unite these distinct events.
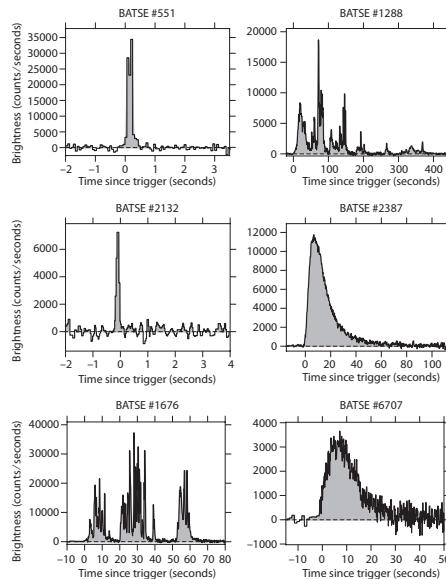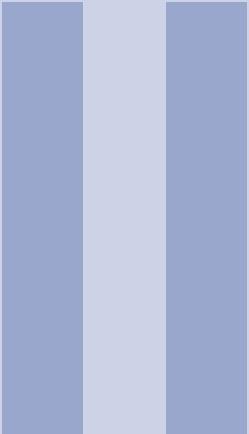
Figure 1.1: A collection of GRB light curves as observed by BATSE. #551 and #2132, are categorized as short-duration bursts, while the rest fall into the long-duration category. Certain events display smooth, single-pulse characteristics (#2387 and #6707), while others exhibit a spiky nature (#1676 and #1288).

Understanding the diverse light curves observed in GRBs presents a non-trivial challenge. Thus far, no cohesive physical model has been identified that can adequately account for all properties associated with GRBs. There exists considerable diversity from one burst to another, with only a limited number of bursts exhibiting similarities.

Notably, GRBs are characterized by their non-repetitive nature, constituting singular occurrences, which suggests that the system responsible for their generation is either destroyed or undergoes a significantly prolonged quiescent period that extends beyond the lifespan of GRB detection missions (the former scenario being more plausible).

They exhibit a diverse range of shapes, ranging from those exhibiting substructure on the scale of submilliseconds to those with smooth time profiles lasting from 10 to 100s of seconds, occasionally accompanied by short spikes. Additionally, certain GRBs demonstrate an "on-off" behavior, temporarily ceasing before producing another burst-like event later, often hundreds or even thousands of seconds afterward.

II

Data Collection through Astrosat CZTI

2. Astrosat CZTI: Introduction and Overview

2.1 Brief Introduction to Astrosat

ASTROSAT, the first multi-wavelength space observatory of India, aims to enhance our comprehension of the universe. Its distinctive characteristic lies in its ability to conduct simultaneous multi-wavelength observations of diverse celestial entities using a singular satellite. Weighing 1513 kg at lift-off, ASTROSAT was launched into a 650 km orbit inclined at a 6-degree angle to the equator on September 28, 2015. This observatory, shaped like a cuboid, carries four co-pointed instruments that provide extensive coverage across the optical, ultraviolet, soft X-ray, and hard X-ray spectrums, alongside a soft X-ray all-sky monitor.
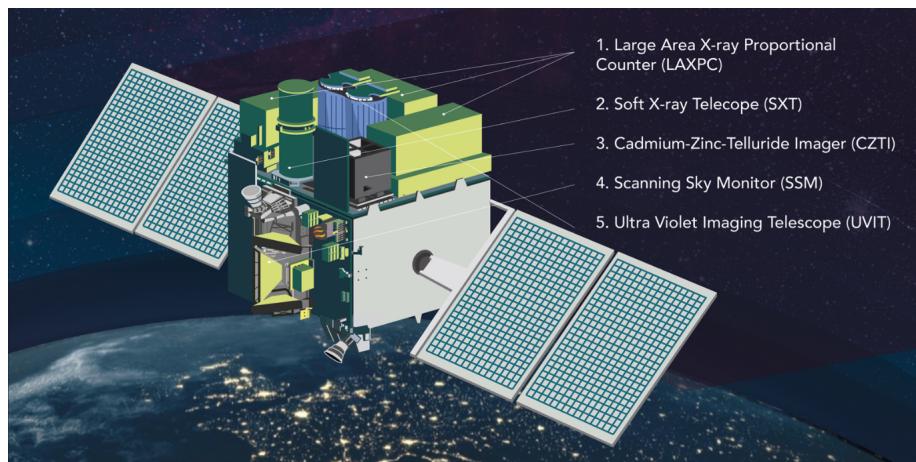

Figure 2.1: Rendering of Astrosat-1 spacecraft in deployed configuration along with its scientific instruments

Its power supply relies on two solar arrays, equipped with Triple Junction solar cells that generate 2100 Watts of electrical power. ASTROSAT's orientation is maintained accurately through its Attitude and Orbit Control System (AOCS), which employs reaction wheels, magnetic torquers, and thrusters. The acquired scientific data from its five payloads is transmitted to the ground station at the Mission Operations Complex (MOX) of ISRO Telemetry, Tracking and Command Network (ISTRAC) in Bangalore. Subsequently, the Indian Space Science Data Centre (ISSDC) in Byalalu, near Bangalore, processes, archives, and disseminates the data.

ASTROSAT carries five astronomy payloads designed for conducting simultaneous observations across multiple wavelength bands:

	UVIT	SXT	LAXPC	CZTI	SSM
Detector	Intensified CMOS, used in photon counting mode or integration mode	X-ray (MOS) CCD (at the focal plane)	Proportional counter	CdZnTe detector array	Position - sensitive proportional counter
Imaging / Non- Imaging	imaging	imaging	non-imaging	imaging	imaging
Optics	Twin Ritchey- Chretian 2 mirror system.	Conical foil (~Wolter-I) mirrors 2-m focal length	Collimator	2- D coded mask	1- D coded mask
Bandwidth	FUV (130-180 nm), NUV (200- 300 nm), VIS (320-550 nm)	0.3 - 8 keV	3-80 keV	10 - 100 keV	2.5 - 10 keV
Geometric Area (cmsq)	~1100	~250	10800	973	~180
Effective Area (cmsq)	10 - 50 (depends on filter)	128@1.5 keV 22@6 keV	8000@5-20 keV	480 (10-100 keV, normal incidence)	~11 @ 2 keV ~53 @ 5 keV for all 3 SSMs
Field of View (FWHM)	28' dia	~40' dia	1° x 1°	6° x 6°	10° X 90°
Energy Resolution	<1000 A (depends on filter)	~5-6%@1.5 keV ~2.5% @6keV	12%@22keV	6% at 100 keV	25% @ 6 keV
Angular Resolution	1.8 arcsec (FUV, NUV) 2.2 arcsec (Vis)	~2 arcmin (HPD)	~(1-5) arcmin (in scan mode only)	8 arcmin	~12 arcmin
Time resolution	1.7 ms	2.4s, 278 ms	10 microsec	20 microsec	1 ms
Typical observation time per target.	30 min	0.5-1 day	1 - 2 days	2 days	10 min
Sensitivity (Obs. Time)	Mag. 20 (5σ) 200 S (for 130-180 nm)	~15 uCrab (5σ) (10000 s)	1 milliCrab (3σ) (100 s)	0.5 milliCrab (3σ) (1000s)	~28 milliCrab (3σ) (600s)
No. of Units	2	1	3	1	3
Total Mass (kg)	230	90	414	50	48
Total Power (W)	85 (pk 117)	80	65	50	30
Sun-avoidance angle	45 deg	> 45 deg	30 deg	30 deg	30 deg from edge of FOV
Prime responsibility	IIA	TIFR	TIFR	TIFR	ISAC

Figure 2.2: Tabulated data of Astrosat-1 Payload Characteristics

1. Twin 38-cm Ultraviolet Imaging Telescopes (UVIT): These telescopes capture images in the Far-UV to optical bands, providing a comprehensive view of celestial objects within this range.

2. Three units of Large Area Xenon Proportional Counters (LAXPC): These instruments specialize in detecting medium energy X-rays ranging from 3 to 80 keV. With an effective area of 8000 sq.cm. at 10 keV, they are capable of efficiently capturing X-ray emissions.

3. Soft X-ray Telescope (SXT): Equipped with conical foil mirrors and an X-ray CCD detector, the SXT operates in the energy range of 0.3-8 keV. It offers an effective area of approximately 120 sq.cm. at 1 keV, facilitating the observation of soft X-rays.

4. Cadmium-Zinc-Telluride coded-mask imager (CZTI): This imager focuses on detecting hard X-rays ranging from 10 to 150 keV. It possesses a field of view of approximately 6 degrees and an effective area of 480 sq.cm. The CZTI utilizes a coded-mask technique for imaging purposes.

5. Scanning Sky Monitor (SSM): Comprising three one-dimensional position-sensitive proportional counters with coded masks, the SSM is designed to locate transient X-ray sources. The assembly is mounted on a rotating platform, enabling it to scan the available sky every six hours, thereby facilitating the identification of transient X-ray emissions.

2.2 Overview of The Cadmium Zinc Telluride Imager

The CZTI (Cadmium Zinc Telluride Imager) is an imaging instrument deployed on AstroSat, designed to capture high-energy, wide-field images. The CZTI detectors, made of Cadmium Zinc Telluride as the main material, are capable of detecting energy levels ranging from 20 keV to greater than 200 keV. These detectors exhibit an energy resolution of approximately 11% at 60 keV. By utilizing a coded aperture mask, the CZTI achieves an angular resolution of 170 within a field of view measuring $4.6° \times 4.6°$ (FWHM).

Figure 2.3: Fully assembled Cadmium Zinc Telluride Imager prior to its integration with the spacecraft.

The CZTI achieves a total detection area of 976 cm^2 by employing 64 CZT modules, each with an area of 15.25 cm^2. These modules are arranged in four identical and independent quadrants. To measure Veto, the CZTI incorporates a Cesium Iodide (TI) scintillator detector positioned below the CZT modules. A gap of approximately 8 cm exists between the collimator slats' base and the detector plane to accommodate a radioactive calibration source module in each quadrant. This module emits alpha-tagged photons with an energy of 60 keV to facilitate energy response calibration.

The CZTI's scientific goals include analyzing the spectra of Active Galactic Nuclei and X-ray binary systems to determine curvature and reflection components, investigating Quasi-Periodic Oscillations at hard X-ray bands in accreting neutron star

and black hole systems, conducting cyclotron line spectroscopy of high-mass X-ray binaries, characterizing the hard X-ray spectra of magnetars, as well as detecting and studying the early light curves of gamma-ray bursts.

Notably, above 100 keV, the CZTI functions as an unobstructed detector, enabling continuous monitoring of Gamma-Ray Bursts (GRBs) and other transient astronomical events across approximately 30% of the observable sky. Additionally, the CZTI's pixelated detectors possess sensitivity to polarization above the threshold of approximately 100 keV, offering exciting prospects for studying the polarization characteristics of transient phenomena and bright, enduring celestial sources.
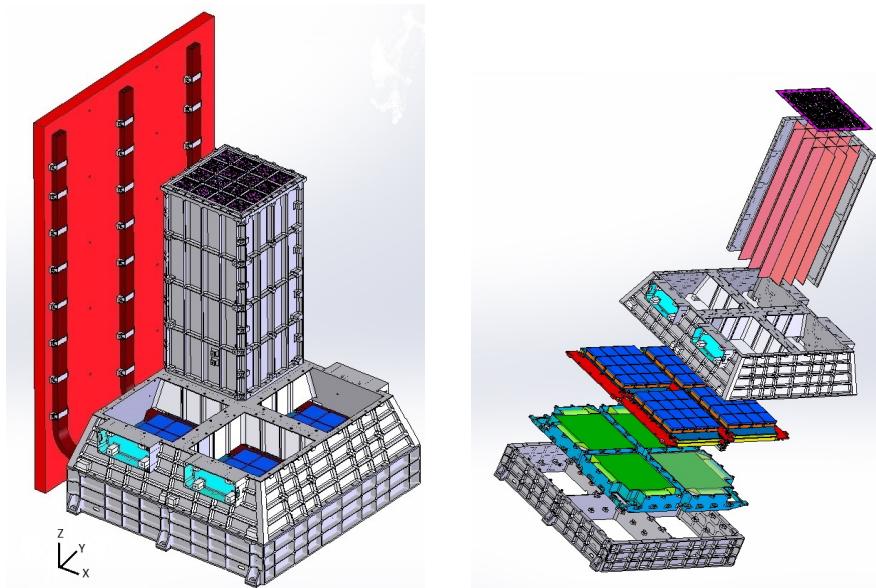

Figure 2.4: Layout: The Cadmium Zinc Telluride Imager (CZTI) consists of four identical quadrants. Each quadrant includes coded aperture masks, CZT detectors, and collimators. The spacecraft co-ordinate system is marked with the Z axis aligned with the instrument bore sight. Radiator plates and side coding masks are positioned along the -X axis. Alpha-tagged 241Am sources are placed near the CZT detectors in ±Y directions. An exploded view of CZTI reveals the components, including the Coded Aperture Mask, collimators, CZT modules, electronics board, CsI(Tl) veto detectors, and 241Am alpha-tagged radioactive sources used for energy response monitoring in orbit.

2.2.1 CZT detector module characteristics

The CZTI utilizes sixty-four detector modules, each composed of 5 mm thick $Cd_{0.9}Zn_{0.1}Te_{1.0}$ crystals, purchased from Orbotech Medical Solutions. The modules feature a continuous anode made of 50 μm aluminised mylar and a cathode divided into a 16x16 pixel grid. Two Application Specific Integrated Circuits (ASICs) facilitate readout from the cathode. These modules exhibit high quantum efficiency in the 20-150 keV range, with reduced efficiency beyond 200 keV. Individual module thresholds can be adjusted, enabling sensitivity from 15 to 200 keV. Ground

calibration demonstrates linear energy response within one percent accuracy.

In space, the modules are temperature-controlled using heaters and cold fin-gers connected to a radiator plate, maintaining a range of $10-15°C$. The CZT modules possess excellent timing capabilities, with a rapid response time of less than 1 μs and negligible dead time. Pileup effects are absent, even during high-count events. In a typical orbit, CZTI is inactive for about 20% of the time due to
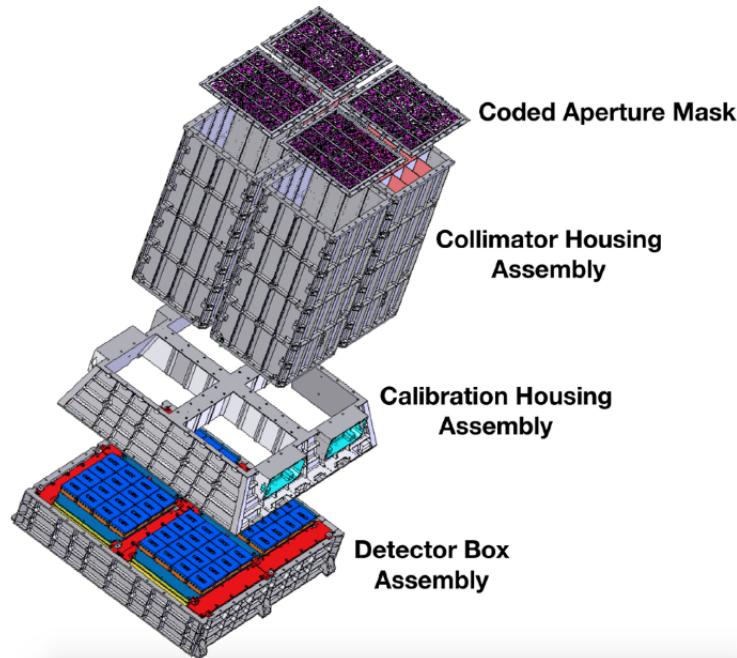

Figure 2.5: Assembly of the CZTI

SAA passage, and about 35% is lost due to occultation of the source by the Earth. including other overheads, the net observing efficiency of CZTI is about 45%.

3. Data Acquisition

3.1 CZTI Data Acquisition Modes

The CZTI can operate on 16 operational modes, consisting of 15 primary modes and one Secondary Spectral Mode which runs in parallel with other primary modes. In the normal course of observations, the modes utilized are:

1. **Normal mode (M0):** This mode is the default operation of CZTI. It records individual events with 10-bit energy information and 20 μs timing accuracy. Photon bunches are handled by the Processing Electronics (PE) and all data is transmitted to the ground. Data frames are generated every second.

2. **SAA mode (M9):** When the spacecraft enters the South Atlantic Anomaly (SAA), CZTI's high voltage is deactivated to ensure device safety. Detection of SAA entry can be automatic using AstroSat's Charge Particle Monitor (CPM) counts or pre-programmed timing. In mode M9, the instrument only records housekeeping, header, and detector temperature information. Data frames are generated once every 100 seconds.

3. **Secondary Spectral mode (SS):** This mode operates in parallel with the normal mode, providing integrated spectra of the CZT, Veto, and Alpha detectors over a duration of 100 seconds.

> (R) The Van Allen radiation belts, encompassing the toroidal regions of energetic charged particles confined by Earth's magnetic field, exhibit a fascinating phenomenon known as the South Atlantic Anomaly (SAA). This SAA affects low-orbit satellites, causing deviations in measurements due to charged particle interference. The impact and intricacies of the SAA will be provided in subsequent discussion.

In addition to these modes, the CZTI has various reduced data modes that can

help reduce data volume while still providing partial spectral or timing information. However, during the instrument's operation, it was observed that these reduced modes were not necessary and are unlikely to be used in the future. Nevertheless, let's briefly explore them:

1. **Fixed No. of Packets (FP):** This mode maintains a frame format similar to the normal mode, but with a fixed maximum number of packets generated from each quadrant.

2. **Veto Spectrum Disabled (VSD):** Packets are generated without including the Veto Spectrum information.

3. **Two-word event report (2WE):** The number of words representing each event is reduced from 3 (as in the normal mode) to 2, sacrificing some time and energy resolution.

4. **Memory Management (MM):** The behavior of the Memory Managed modes depends on the payload memory occupancy level. Depending on the level, data recording can include Full data, only Secondary Spectral Mode data and Header data, or solely the Header data. Beyond a certain level, no data is recorded.

3.2 Introduction to FITS Files and their Role in Astronomical Data

FITS (Flexible Image Transport System) serves as the dominant data format in astronomy, facilitating the transportation, analysis, and archiving of scientific data files. It distinguishes itself from conventional formats like JPG or GIF by its primary purpose of storing multidimensional arrays (images) and 2-dimensional tables comprising scientific data organized in rows and columns.

3.2.1 HDUs: Header/Data Units

A FITS file comprises segments known as Header/Data Units (HDUs). The initial HDU is referred to as the "Primary HDU" or "Primary Array," capable of accommodating multidimensional arrays of integers or floating-point numbers with various dimensions. Additional HDUs, referred to as FITS extensions, can follow the primary array. Three standard extension types exist:

Image Extensions

Image Extensions represent arrays of pixels, similar to the primary array, with dimensions ranging from 0 to 999. They are denoted by a header starting with `XTENSION = 'IMAGE '`.

ASCII Table Extensions

ASCII Table Extensions store tabular data with numeric values stored in ASCII formats. While ASCII tables may be less efficient than binary tables, they offer human readability and the ability to store numeric data with arbitrary size and precision. The header begins with `XTENSION = 'TABLE '`.

Binary Table Extensions

Binary Table Extensions store tabular data in a binary representation. Each cell in the table can be an array, but the array's dimensionality must remain constant within a column. Although the strict standard supports only one-dimensional arrays, a convention to support multi-dimensional arrays is widely accepted. The header begins with `XTENSION = 'BINTABLE'`.

3.2.2 Header Units

Each HDU comprises an ASCII-formatted "Header Unit" followed by an optional "Data Unit." Both the header and data units have lengths in multiples of 2880 bytes. Padding with ASCII blanks or NULLs, depending on the unit type, is applied when necessary.

The header unit consists of a sequence of fixed-length 80-character keyword records following the format: `KEYNAME = value / comment string`. Keywords have a maximum length of 8 characters, restricted to uppercase letters A to Z, digits 0 to 9, hyphens, and underscores. The keyword is typically followed by an equals sign and a space character in columns 9 and 10 of the record, followed by the value. Values can be integers, floating-point numbers, complex values, character strings enclosed in single quotes, or Boolean values. Certain keywords like `COMMENT` and `HISTORY` do not require an equals sign.

The header unit commences with a series of essential keywords that specify the size and format of the subsequent data unit. Additional optional keywords can be included to describe various data aspects or provide further documentation. The final keyword in the header is always `END`, which contains blank value and comment fields. If needed, the header is padded with additional blank records to achieve a length that is a multiple of 2880 bytes.

3.2.3 Data Units

The data unit, if present, immediately follows the last 2880-byte block in the header unit. However, it is not mandatory, and some HDUs may only contain the header unit.

For primary arrays or image extensions, the image pixels can assume one of five supported data types: 8-bit unsigned integer bytes, signed integers (16-bit or 32-bit), single-precision floating-point numbers (32-bit), or double-precision floating-point numbers (64-bit).

Binary tables and ASCII tables serve as standard extensions for tabular data. Binary tables are faster to read and write, as they utilize a machine-readable binary representation, resulting in more compact storage. Both ASCII and binary tables support various data formats for each column, including integers, floating-point values, character strings, logical values, bit values, and complex data formats. Binary tables additionally allow for N-dimensional arrays within each field, either fixed-length or variable-length arrays.

3.3 Investigating Fast Transients Using CZTI

Although the primary coded field of view of CZTI is $4.6° \times 4.6°$, the collimators and support structure of the instrument become increasingly transparent to radiation above approximately 100 keV, enabling sensitivity to sources across the entire sky. The contribution of off-axis sources is minimal due to the scarcity of bright sources within this energy range, resulting in only a slightly elevated background.

However, special exception to this are extremely bright, short duration transient sources like GRBs. These GRBs, characterized by their high brightness and brief durations of seconds to minutes, lead to increased count rates in CZTI. Prior to conducting data analysis for our investigation, it is essential to perform data pre-processing which is achieved through the CZTI pipeline.

3.3.1 A Brief Summary of CZTI Data and Processing Pipeline

The raw CZTI data undergoes processing through a multi-step pipeline. To facilitate long-term storage, three main data levels have been established:

1. **Level 0:** This represents the raw telemetry data received from the satellite, segregated by instrument, and accompanied by auxiliary data. This raw data is internally archived and not publicly distributed.

2. **Level 1:** The raw data is reorganized and formatted in FITS (Flexible Image Transport System) format for astronomical usage. All relevant auxiliary information required for further processing is consolidated and packed together with the corresponding science data. The Level 1 data is made available through the Astrosat data archive, initially to the Principal Investigator (PI) of the relevant observation proposal, and subsequently, to any interested parties after a specified initial lock-in period.

3. **Level 2:** This level contains standardized science products derived from Level 1 data, also in FITS format. It follows the same release criteria and mechanism as Level 1 data, with specific lock-in conditions.

The software utilized for processing data from Level 1 to Level 2 offers user-configurable elements. While the Payload Operation Centre (POC) runs a default configuration to generate automated Level 2 standard data products, users can customize the software settings to create more customized products for data analysis.

The Level 2 pipeline software is publicly released for general use. In terms of Input/Output, the pipeline software implementation makes use of the CFITSIO and Parameter Interface Library developed and distributed by NASA's High Energy Astrophysics Science Archive Research Centre.

3.3.2 A deeper dive into Level-1 data

The CZTI level-2 data reduction pipeline takes the Level 1 data set as input. The original raw data is received via telemetry at the Indian Space Science Data Centre (ISSDC) during each visible pass over Bengaluru, India. It is then packaged and sent to the Payload Operations Centre (POC) for further processing. At the POC,

data belonging to a specific Observation ID (ObsID) is merged from orbit-wise packages to create a consolidated Level 1 product for distribution and analysis.

The Level 2 pipeline analysis begins with the merged Level 1 package, which includes the following files:

1. **Science Data File**: A FITS file containing sequentially stacked 2048-byte data frames generated by the CZTI payload. This data is mode segregated, i.e., a different FITS file is generated for each distinct data mode.

2. **Time Calibration Table (.tct)**: This file contains a list of time tags expressed in CZTI internal clock, Satellite On Board reference Time (OBT), and Universal Time (UT) derived from the SPS units on ASTROSAT. As all the CZTI science data is tagged with its internal clock, the TCT file is required to correlate them with other on-board events, as well as to obtain absolute timing.

3. **Orbit File (.orb)**: Gives time-tagged orbital position of the satellite in terms of geocentric x, y, z as a function of time, as well as the corresponding velocity components.

4. **Attitude File (.att)**: Gives time-tagged satellite attitude information, in terms of quaternions as well as the RA, Dec values of the pointing directions of the three reference axes of the satellite.

5. **LBT Housekeeping File (.lbt)**: This gives a time-tagged recording of 65 different health parameters monitored by different sensors on the CZTI.

6. **MCAP File (.mcap.xml)**: This XML file has information like source observed, start and end time of observation, etc.

7. **Make Filter File (.mkf)**: This file collects together the time series of a number of selected parameters, including health parameters, Sun angle, Earth angle, charged particle count, etc., based on which the quality of data obtained can be assessed.

These files serve as essential inputs for the subsequent data reduction and analysis processes in the Level 2 pipeline. The workflow of the pipeline is provided below:

The CZTI level-2 data reduction pipeline consists of distinct modules dedicated to specific tasks. The table summarizes the tasks performed by these various software modules. Data reduction for CZTI is carried out in three stages, each encompassing multiple pipeline tasks:

1. **Stage 1:** Event file generation and calibration. This stage involves generating an event file from the level 1 data, applying gain offset corrections, and eliminating multi-hit events.

2. **Stage 2:** Data selection and cleaning. In this stage, events are selected based on the Good Time Interval criteria, and noisy pixel events are removed from the dataset.
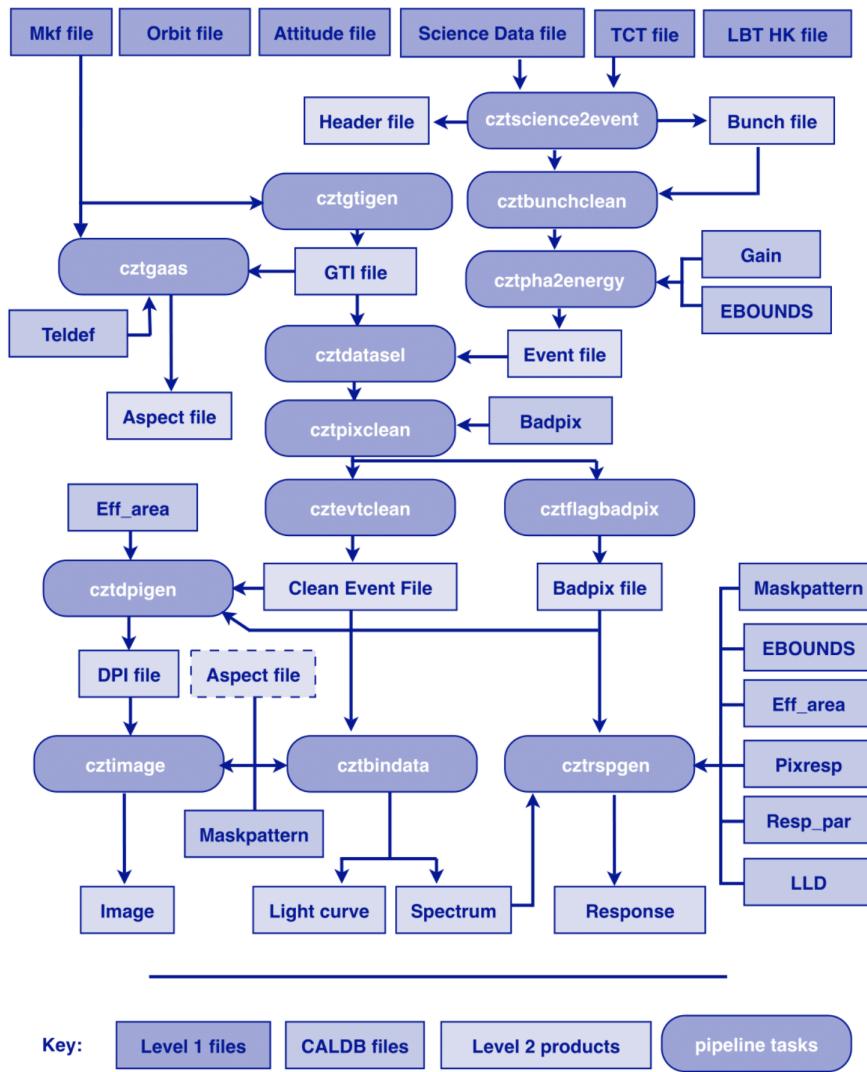
Figure 3.1: Level-2 pipeline work flow

3. **Stage 3:** Science product generation. The final stage entails producing Data Product Interface (DPI), images, spectra, light curves, and response matrices.

3.4 Automating the Pipeline

The objective is to streamline automate the entire pipeline process to customize the results according to specific sets of variable values rather than relying on default configurations using a Python-based script. Inputs required for the script include bunch-cleaned event files, mkf thresholds, mkf file, and bunch-cleaned livetime files.

By running the script, the final stage 3 products can be obtained.The Python script defines variables such as bin sizes and handles the automation process, reducing the need for manual intervention. Making changes to the variables is straightforward, as it only requires modifying relevant parameters within the code.

SI No	Module name	Functionality
1	cztscience2event	- Extraction and decoding of level-1 science data into events - Mapping of quadid, detid and pixid to detx, dety - Mapping of 12 bit PHA to 10 bit PHA channels - Time calibration - Recording of temperature history
2	cztbunchclean	clean the event file by identifying and removing bunches
3	cztpha2energy	convert PHA value to PI and write to event file
4	cztgtigen	generate GTI based on current gti, mkf and user input
5	cztgaas	compute time dependent and average aspect of CZTI
6	cztdatasel	select events based on gti
7	cztpixclean	identify noisy pixels and detectors and remove noisy events
8	cztflagbadpix	combine bad pixel list from multiple sources, if required
9	cztevtclean	select events based on veto and alpha tags
10	cztdpigen	generate Detector Plane Image from clean event file
11	cztimage	create image using Fast Fourier Transforms
12	cztbindata	generate light curve/spectrum
13	cztrspgen	generate response matrix file

Table 3.1: Summary of modules in CZTI level-2 pipeline

The automation primarily focuses on the generation of light curves and spectra, streamlining the process efficiently.

The automation script follows a sequential process consisting of several functions to accomplish the automation. Each function performs a specific task, and the output of one function is used as input for the next as depicted in the workflow figure. It utilizes the `subprocess` and `glob` modules to automate the CZTI level-2 data reduction pipeline. These modules enable the execution of command-line commands and facilitate file path matching, respectively.
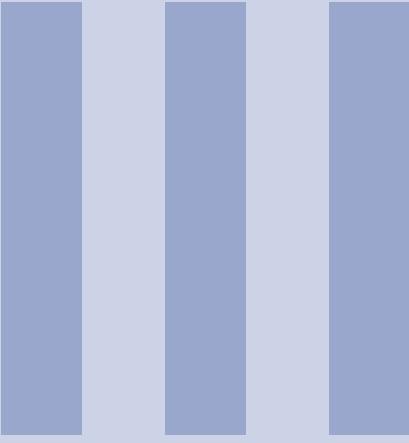
The `subprocess` module in Python provides a way to execute system commands or external processes from within a Python script. In the given code, the `subprocess.run()` function is used to run command-line commands related to the CZTI data reduction pipeline. The command and its arguments are provided as a list, and the function executes the command and waits for it to complete before proceeding with the script. By setting `check=True`, the script raises an exception if the command execution fails.

On the other hand, the `glob` module is employed to search for file pathnames matching specified patterns, using wildcard characters such as `*` and `?` to represent multiple or single characters in filenames. In the provided code, the `glob.glob()` function is used to locate specific files required for the CZTI data reduction pipeline. The function takes a pattern as an argument and searches for files that match the pattern. The pattern is constructed using the `dir_path` and `ob_id` variables, which represent the directory path and observation ID, respectively. The function returns a list of file paths that match the pattern, and `[0]` is used to select the first file path from the list.

By utilizing the `subprocess` module, the script executes command-line commands necessary for each stage of the data reduction pipeline. The `glob` module allows the script to dynamically locate and access the required input files based on the provided observation ID and directory path.

Thus, in the end the script requires two inputs from the user: the directory path where the files are located and the observation ID. These inputs enable the script to identify the specific files needed for analysis.

After providing the inputs, the script executes the `orbit_lightcurves` function, which performs a sequence of tasks within the pipeline and generates the lightcurves and spectra, tailored to specific requirements as defined in the script.

III

Computational Tasks

4. FITS File Handling with astropy.io.fits

The Flexible Image Transport System (FITS) is a widely used digital file format specifically designed for efficient storage of multi-dimensional data arrays and tables, particularly in the field of astronomy. It offers optimized features for handling astronomical data and is highly prevalent in the astronomical community. The `astropy.io.fits` sub-package provides functionalities for reading, manipulating, and writing FITS-formatted data, enabling easy analysis and manipulation of astronomical data in conjunction with Python and Astropy functions.

FITS File Organization:
- FITS files consist of Header Data Units (HDUs), which comprise a header and associated data, such as tables or multi-dimensional data arrays (e.g., images).
- The header contains metadata that describes the corresponding data, providing essential information.
- Multiple HDUs can exist within a FITS file, with the primary HDU designated as extension 0 and subsequent extensions labeled with integers starting from 1.

Reading and Extracting Data:
- The `astropy.io.fits` module allows for determining the structure and contents of FITS files.
- Data and metadata (header information) can be extracted from FITS files using `astropy.io.fits`.
- Headers store keyword/value pairs, functioning akin to dictionaries, and may include optional comments.
- Accessing the metadata is facilitated through the `.header['KEYWORD']` method.

Handling Tables and Images:
- Tables and images within FITS files can be accessed using the `.data` method.

- For tables, the data is assigned to a structured array, allowing for efficient manipulation.
- Image data is assigned to an n-dimensional array, which can be visualized using functions like `matplotlib's imshow`.

To summarize, using `astropy.io.fits` in Python empowers astronomers to explore FITS file structures, extract data and metadata, as well as perform tasks such as writing new FITS files or modifying existing ones. The extracted data can be efficiently handled as structured arrays or n-dimensional arrays, enabling further analysis and visualization.

5. First Elementary Analysis: GRB190928A

5.1 Plotting the clean lightcurve

To begin, the code takes the light curve files as input, which are FITS files containing data. Using the astropy.io.fits module, the code opens the FITS files and extracts the time and rate data. The extracted data is then plotted, with the time values plotted on the x-axis and the corresponding rate values on the y-axis. Each subplot within the figure represents a different quadrant of the CZTI.
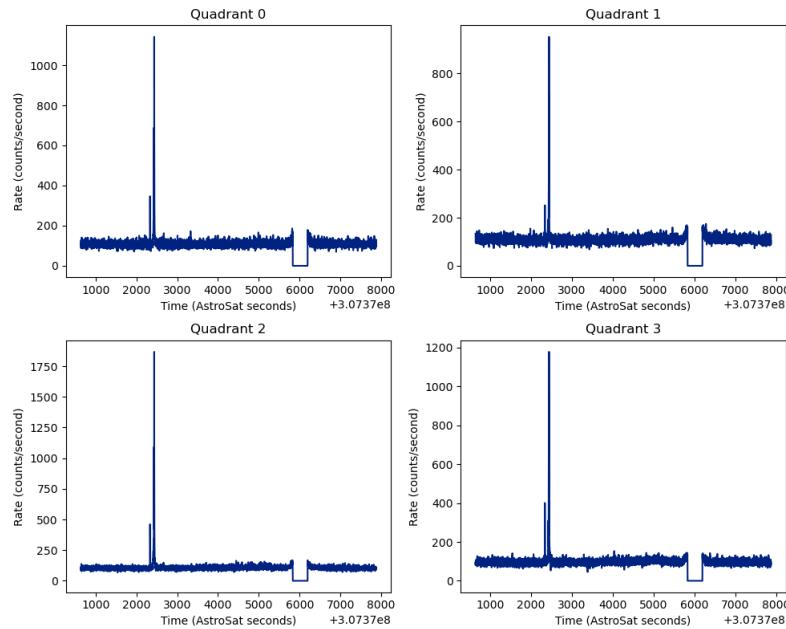

Figure 5.1: Cleaned lightcurves with 1s binning

5.2   Plotting the Focused Lightcurve Centered around the GRB

The peak indices in the rate array are determined by identifying the two highest peaks. A time window is then defined around these peaks. The data within the window, including time and rate values, is extracted from the FITS files. The matplotlib library is used to plot the extracted data, with different colors representing different quadrants. The x-axis represents time in AstroSat seconds, and the y-axis represents the rate in counts per second. Finally, the plot is displayed.
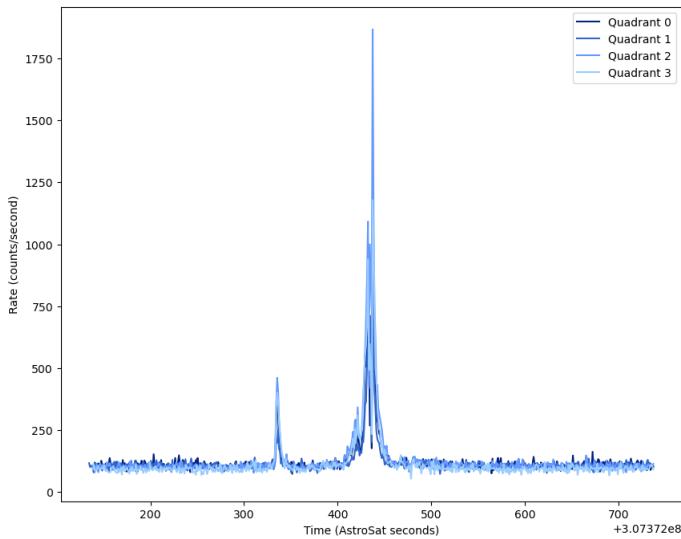

Figure 5.2: Focused Lightcurve Centered around the GRB with 1s binning

5.3   Characterizing Noise, SAA, and GRB Windows

We begin by defining the GRB (Gamma-Ray Burst) window. This window is crucial for capturing the GRB event of interest. Next, the SAA (South Atlantic Anomaly) window is determined. This region corresponds to the time period where the rate of counts drops to zero during a certain period.

The code identifies the start and end times of the SAA window based on the data. To ensure comprehensive coverage, a small buffer is added around the SAA window. This buffer allows for a margin of error and ensures that any potential effects or variations surrounding the SAA region are included.

Finally, the remaining time intervals, excluding the GRB window and the buffered SAA region, are classified as the noise region. This distinction is crucial for future analysis, as it provides a clear separation between the signal of interest (GRB) and any background noise or unrelated fluctuations in the data.

Thus we partition the data into three distinct windows - the GRB window with a buffer, the SAA window with a buffer, and the noise region.
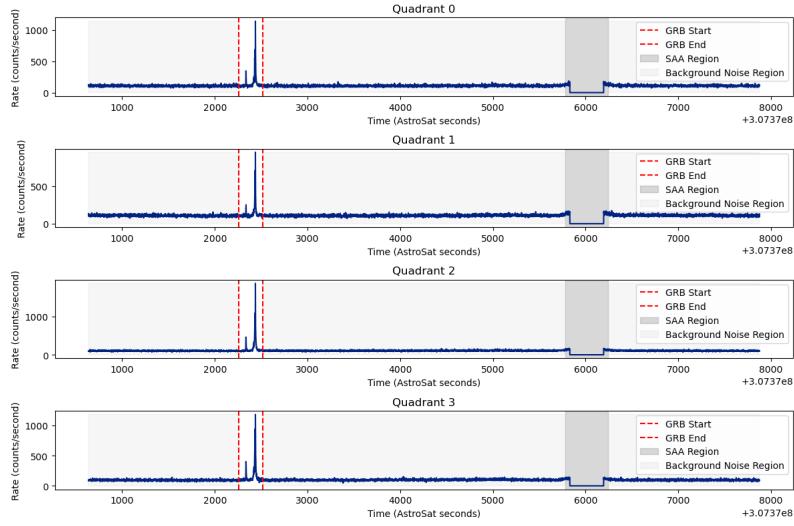
Figure 5.3: Shaded Regions Depicting Noise, Buffer, and GRB
Windows for all quadrants

5.4 Detrending

The median filter, with a window size of 195s, was employed to smoothen the light curve by replacing each data point with the median value within its window. This technique helps to suppress short-term variations and extract the underlying trend in the data. The detrended light curve was then plotted alongside the original data and the variation. Calculations on further properties can be done using this.
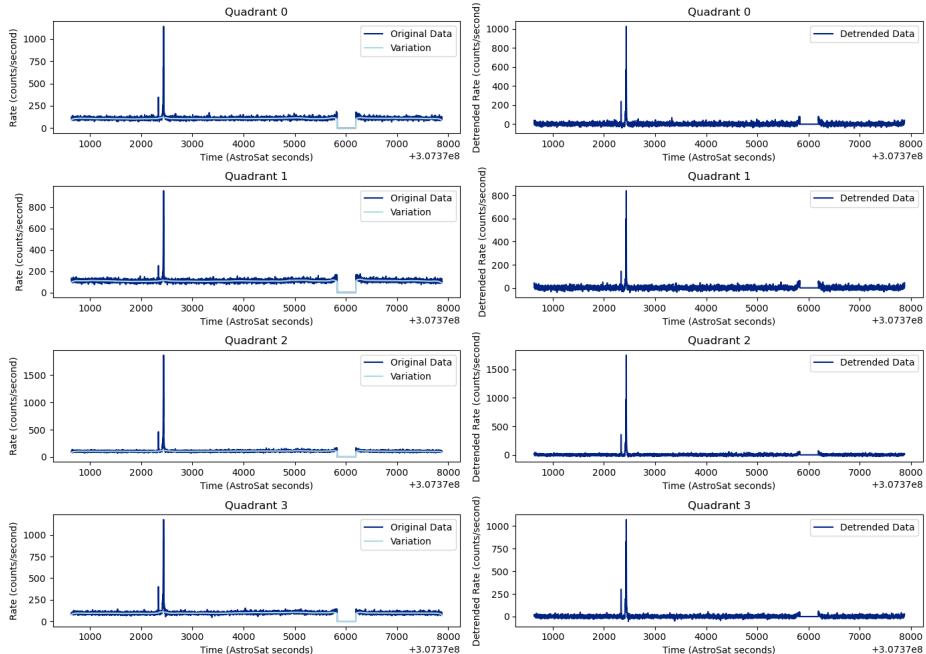

Figure 5.4: Detrended light curve using median filter

5.5 Plotting the Focused Lightcurve Centered around the GRB for differnt Bin sizes
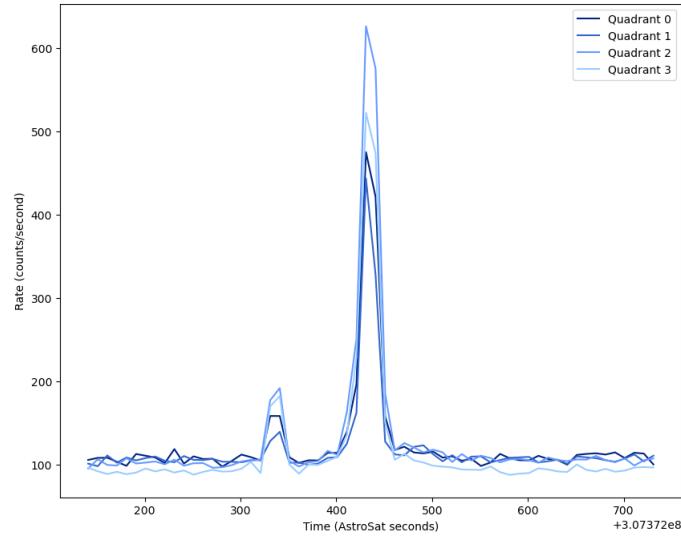

Figure 5.5: Focused Lightcurve Centered around the GRB with 10s binning
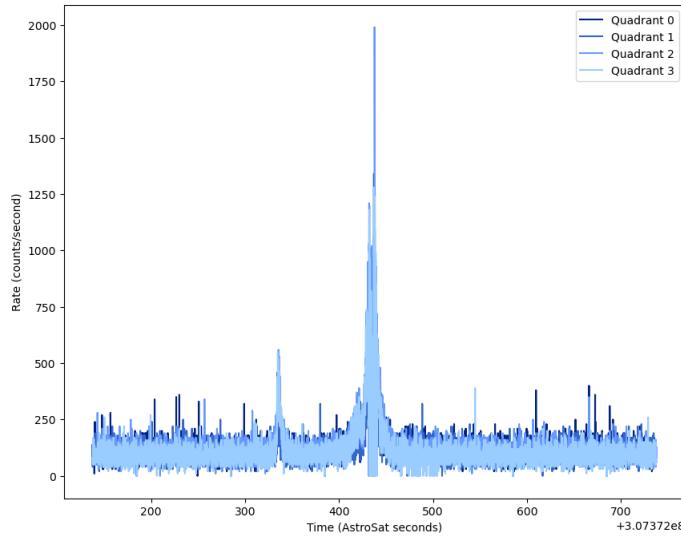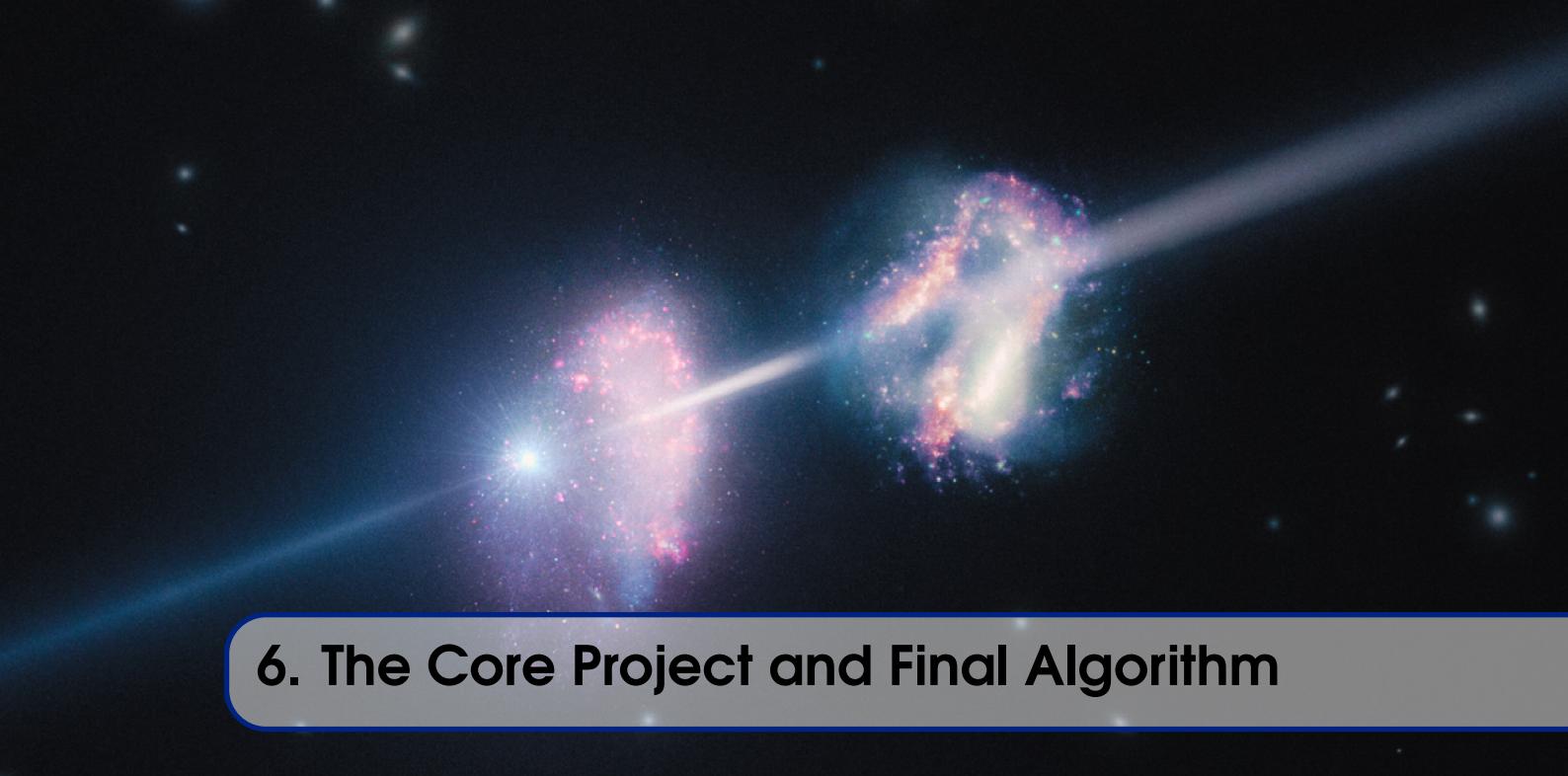
Figure 5.6: Focused Lightcurve Centered around the GRB with 0.1s binning

6. The Core Project and Final Algorithm

6.1 Introduction

A key challenge in the investigation of GRBs lies in the need for efficient and auto-mated methods to analyze observational data.

The aim of this project is to conduct a comprehensive statistical analysis of GRB detections using CZTI Astrosat data, with a particular focus on studying faint GRBs. Traditionally, the search for GRBs has been primarily a manual and labor-intensive process, mainly due to the diverse profiles of these events. However, by developing an algorithm for identifying these bursts and quantifying the strength of their detec-tion, we seek to streamline and automate the search, enabling a more efficient and productive approach to GRB discovery.

Automating the search for GRBs could potentially lead to an increase in the rate of discoveries, allowing us to uncover a more extensive sample of these events. By focusing on the detection of faint GRBs, which often go unnoticed but carry invaluable information, we aim to unlock deeper insights into the fundamental physics driving these explosions.

6.2 Data Preparation for Statistical Analysis

De-Trending

To capture the relevant information surrounding the detected GRB event, we extract a window of data centered at the `trigger_index`, which represents the time point when the GRB event is detected. This time window spans $\pm$ 500 seconds around the `trigger_index`, ensuring that we incorporate an adequate number of data points both before and after the GRB event.

Next we proceed to identify and mask data points that deviate significantly from

the mean (particularly 1σ deviations) of the light curve using method `sigma_clipped_stats` and then apply a Savgol filter on the masked light curve.

After applying the Savgol filter, a quadratic polynomial is fitted to the filtered background data. This quadratic polynomial represents the underlying trend of the background. By subtracting this quadratic fit from the entire lightcurve, the background is effectively removed, leaving us with a de-trended lightcurve for further analysis.
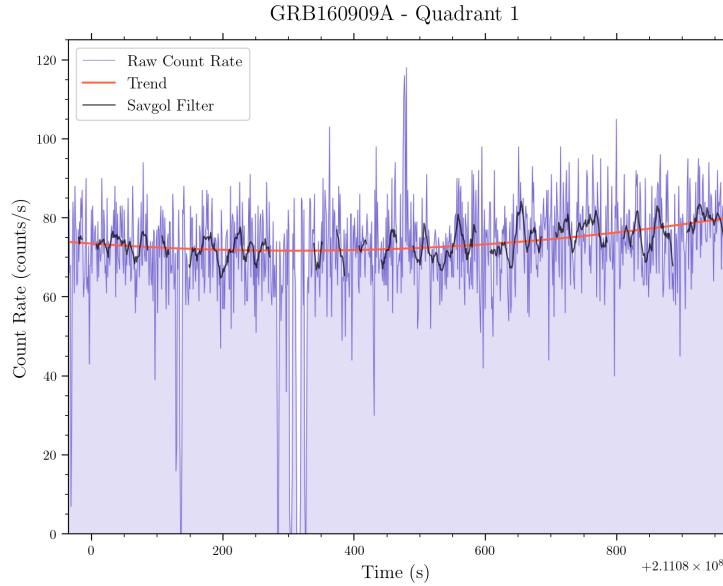

Figure 6.1: GRB 1606909A: Quadratic Fit to Savgol Filter
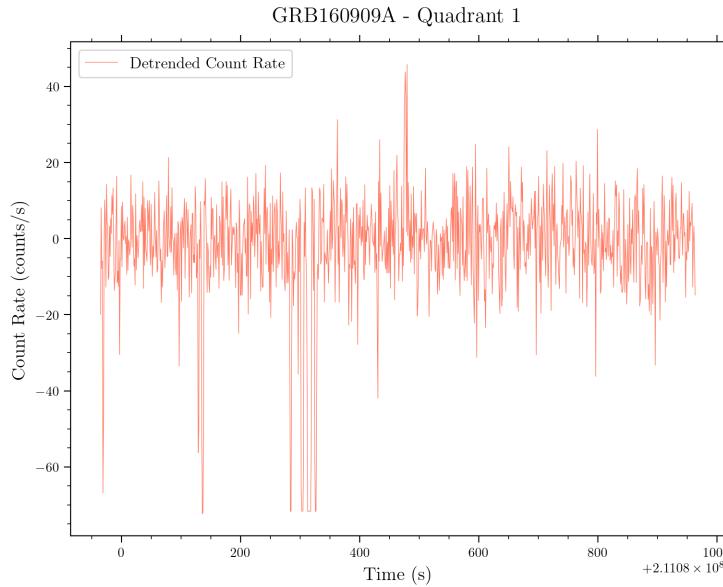
Figure 6.2: GRB 1606909A: Detrended lightcurve

Finding Potential GRB Events

The detection of potential GRB events is done by identifying any spikes in the light curve data for the 20-60keV range that exceed a certain threshold of standard deviation. These detected events serve as essential candidates for further examination. The threshold is set to 1σ by default although the algorithm allows it to be changed as per needs be. Subsequent sections will delve into the detailed methodology of this process.
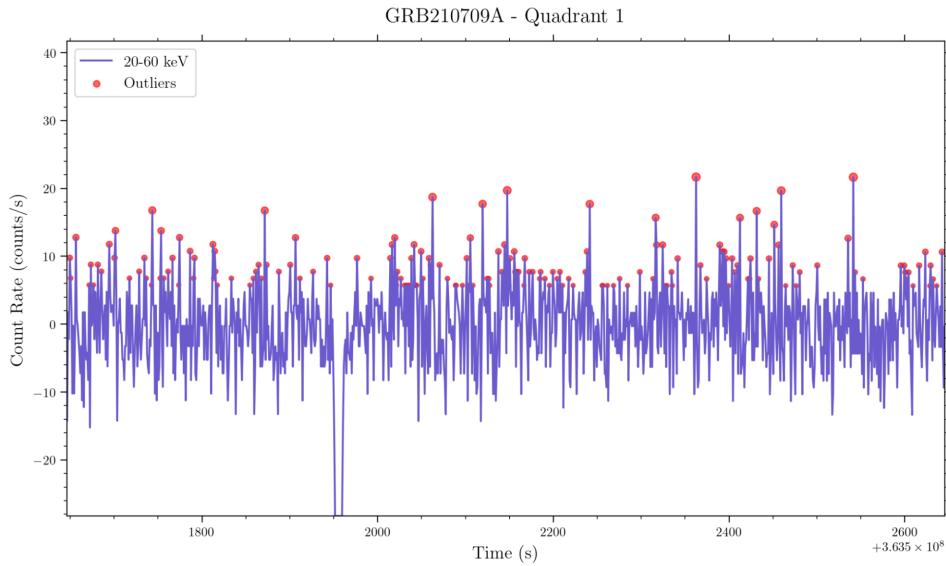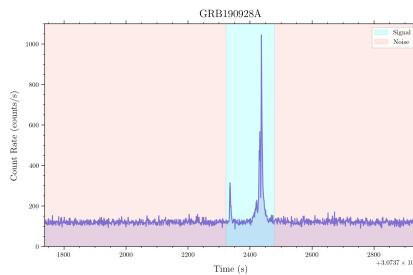

Figure 6.3: GRB210709A: Red dots indicate outliers, with their size representing the Signal-to-Noise Ratios(SNRs)

6.3 SNR Calculations

6.3.1 Definition

The signal-to-noise ratio (SNR or S/N) is a scientific and engineering metric used to compare the strength of a desired signal to the level of background noise. It is calculated by dividing the power of the signal by the power of the noise and is often expressed in decibels. An SNR greater than 1:1 (above 0 dB) indicates that the signal is stronger than the noise. There are different formulas to compute SNR based on how the signal and noise are measured and defined.

Considering the diverse nature of SNR definitions, an exploration of various

methods to quantify both the signal and noise has been done. The signal power

has been equated to the Peak Count Rate, and the noise has been characterized by various methods. The summary of these methods is presented below:

Gaussian Fit to Noise
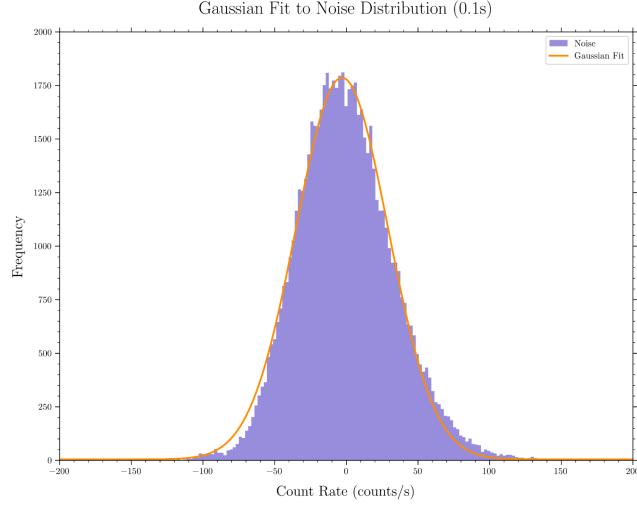
Figure 6.4: Gaussian Fit To Noise

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$SNR = \frac{M_s + \mu}{\mu + \sigma}$$

Poisson Fit to Noise
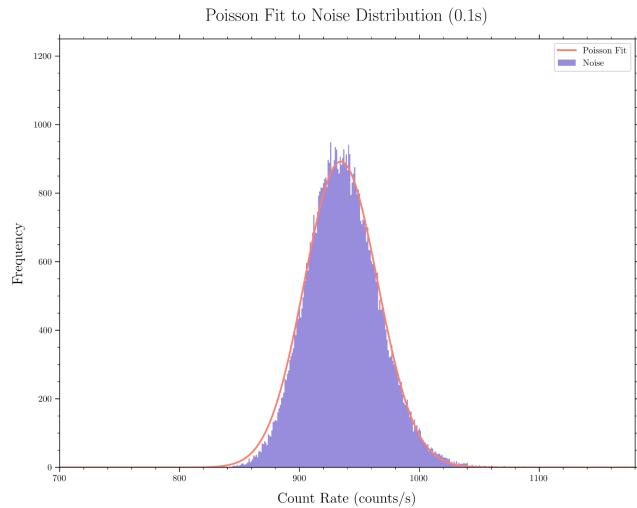
Figure 6.5: Poisson Fit To Noise

$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$SNR = \frac{M_s + \lambda}{\lambda + \sqrt{\lambda}}$$

Gamma Fit to Noise
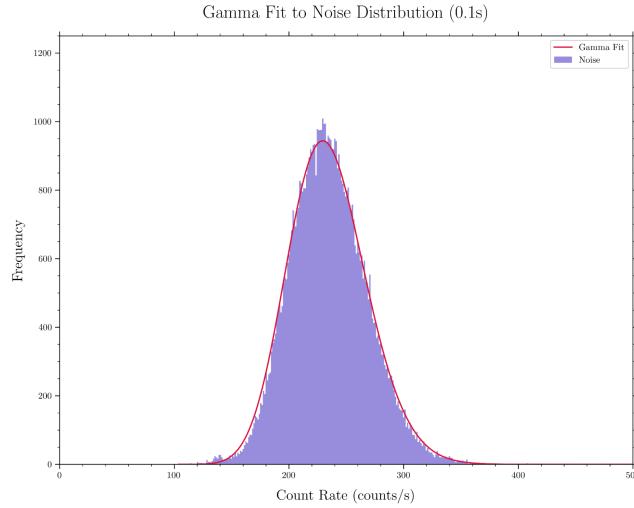
Figure 6.6: Gamma Fit To Noise

$$f(x;k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$$

$$SNR = \frac{M_s + k\theta}{k\theta + \sqrt{k\theta^2}}$$

Skew-norm Fit to Noise
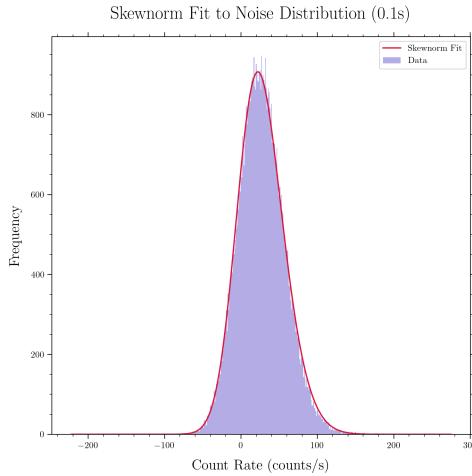
Figure 6.7: Skew-norm Fit to Noise

No Fit

$$SNR = \frac{M_s}{\mu + \sigma}$$

By computing the Root Mean Square Error (RMSE), Kurtosis, Skewness, and Variance parameters and examining their values, we can gain insights into the characteristics of each method. This analysis will help us determine which method is most suitable for quantifying the SNR.

6.3.2 Overall Comparison

|	RMSE	Kurtosis	Skewness	Variance
Gaussian	14.35	0	0	855.78
Possion	16.51	0.001	0.032	934.95
Gamma	13.12	0.145	0.311	937.23
Skewnorm	13.10	0.175	0.299	932.53
Noise	-	2.856	0.632	1201.03

Indeed, the results demonstrate that none of the fittings adequately represent the noise. Therefore, for our current purposes, we will proceed with the simple SNR calculation method without any fitting.

6.3.3 Time Dependence of SNR

The value of the signal-to-noise ratio (SNR) is indeed influenced by the time interval used for binning the light curve. When the time interval is too small, the SNR tends to approach zero due to limited counts in each bin. Conversely, when the time interval is too large, the SNR also tends to approach zero, as the binning averages out the signal, making it indistinguishable from the noise.

To determine the optimal time interval that yields the maximum SNR, the algorithm generates a plot of SNR against different time intervals. The plot illustrates the trend of the SNR as the time interval varies.
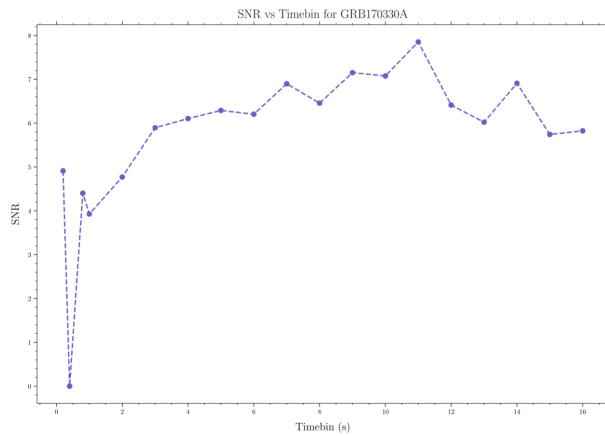
Figure 6.8: Variation of SNR values according to chosen timebins during `cztbindata` for GRB170330A

6.4 Splitting Energy Bands

Our main objective is to analyse the light curve data and identify signal characteristics and we will do so by binning the count rate into different energy bands. So the main question here is, what energy bands do we consider? This choice would depend on our algorithm, but we can still begin by making initial estimations.

If we consider too few energy bands then we cannot extrapolate any meaningful trends.

Dividing the data into several energy bands gives us a more specific view of how the count rate is spread across them. However, our data is not an infinite distribution and dividing it too many times will lead to statis tical loss, rendering the division meaningless and obscures from actual trends that might be present.

What we want to do is find the optimal sweet spot corresponding to our algorithm.

6.4.1 Determining the Number of Energy Bands:

Considering that most modules in CZTI are sensitive to hard X-rays in the range of 20-200 keV, this range is chosen for our master light curve file. For our analysis, we focus on the energy range of 20-60 keV, 60-100 keV, and 100-200 keV. The selection of these energy ranges is based on specific characteristics observed at different energy levels:

- **20-60keV:** In the energy range of 20-50 keV, light curves from a single quadrant occasionally show noise spikes similar to astrophysical transients. These events are primarily caused by charged particles or electronic noise and are more prevalent at lower energies (<50 keV).
  It is important to note that the CIFT paper utilizes segregations of 20-50 keV, 50-100 keV, and 100-200 keV. While we also considered these segregations, we found that they yielded significantly worse results when applied to our algorithm. Thus, we tried to understand the reasoning behind this.
  The 20-50 keV region contains a substantial portion of electronic noise and charged particle events, which can obscure or hide genuine signals. To overcome this issue and facilitate proper signal detection, we incorporated an additional 50 to 60 keV region. This extension allows us to leverage some of the GRB counts present in the 50 to 60 keV range and and contributes to a more reliable statistical analysis of the data.

- **100-200keV:** In the energy range of 100-200 keV, the CZTI instrument's collimators exhibit increased transparency, enabling it to be sensitive to radiation from the entire sky. Notably, the typical spectrum of Gamma Ray Bursts (GRBs) often peaks within the 100 keV to 1 MeV energy range. This characteristic makes the CZTI well-suited for detecting transient events, such as GRBs, with optimal efficiency in this particular energy range.

- **60-100keV:** In the energy range of 60-100 keV, there is a notable aspect worth discussing based on the findings from a study conducted by Sujay Mate et al. titled "The AstroSat mass model: Imaging and flux studies of off-axis sources with CZTI."
  The researchers performed simulations and observed data analysis, which revealed the presence of significant flux in tantalum fluorescence lines within the 55-65 keV region (with K_α at 57.52 keV and K_β at 65.21 keV). As a result of this observation, they chose to exclude the 50-70 keV band from their analysis.
  However, this presents us with two viable options for our analysis. One option is to discard the 50–70 keV range altogether, which would remove any potential GRB counts that may occur within this region, leading to a loss of relevant

statistical data. The second option, which we have chosen to implement, is to keep the 60-100 keV range separate from the 100-200 keV range and analyze them independently.
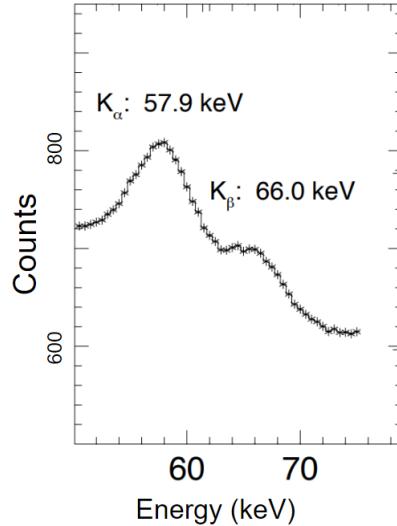

Figure 6.9: Background spectra with Ta fluorescence lines at 57.52 and 65.21 keV, recovered at $\sim$ 57.9 keV and $\sim$ 66.0 keV, respectively.

By adopting this approach, we acknowledge the potential presence of tantalum fluorescence lines in the 60-100 keV region and avoid any undue influence on the analysis in the 100-200 keV range. So, the final energy bands we have implemented for our analysis are 20-60 keV, 60-100 keV, and 100-200 keV. This segmentation enables us to carefully consider the unique characteristics of each energy band and facilitates a more accurate and meaningful analysis of our data.

### 6.4.2 Energy-Dependent Statistics
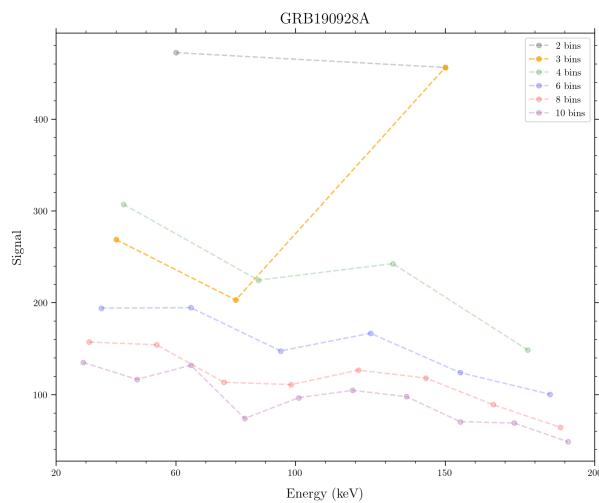
Figure 6.10: Peak Value Variation in Different Energy Bands for GRB190928A: GRB trend
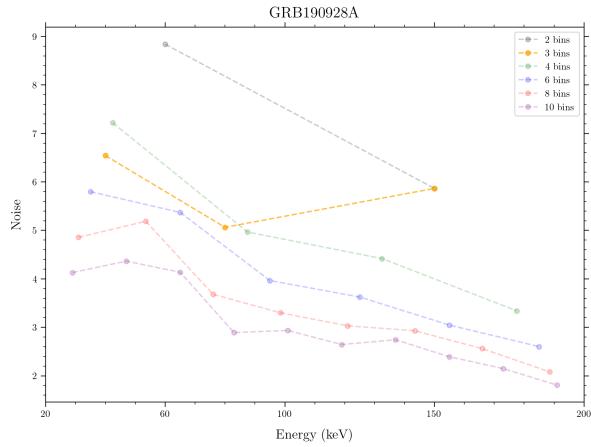
Figure 6.11: Peak Value Variation in Different Energy Bands for
GRB190928A: Background Noise trend
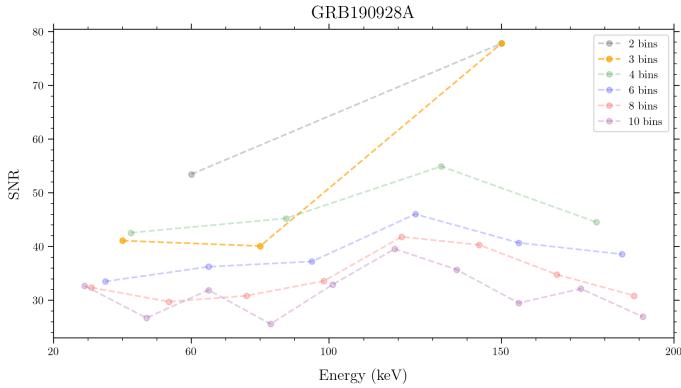
Figure 6.12: SNR Value Variation in Different Energy Bands for
GRB190928A: GRB trend

Again, we see that too few bands limit us and excessive division too showcases nothing of significance. We can see that in contrast to the noise, GRBs exhibit a distinctive and rapid rise in the 100-200 keV range. This energy band proves crucial in capturing the sharp jump characteristic of GRB signals, setting it apart from the noise fluctuations.

The Signal-to-Noise Ratio (SNR) trend roughly follows the Peak Value Variation, reflecting the relationship between the peak value of the signal and the mean plus standard deviation of the noise. A higher SNR indicates a stronger and more distinguishable signal amidst the background noise.

6.5 Algorithm Overview

6.5.1 Outlier Classification

Before providing an overview of the entire algorithm's functioning, it is crucial to outline the steps through which outliers are classified as "potential GRB candidates" and the conditions imposed on them. These filtration processes are essential in selecting the most promising events for further analysis.

Type 1 "Outliers"	Obtained from a 1 SNR clip of 20-60 keV light curve
Type 2 "Filtered Outliers" (subset of type 1)	Obtained by 3 SNR clip of type 1 outliers in 60-100 and 100-200 keV range as well as a 1 SNR clip of 20-200 keV.
Type 3 "Possible GRBs" (subset of type 2)	Obtained by cross matching timestamps of type 2 outliers.

6.6   Directory Structure and Algorithm Flowchart

```
Parent Directory
        ├──── bc.evt
        ├──── bc_livetime.fits
        ├──── level2.mkf
        ├──── mkfThresholds.txt
        ├──── output.pdf
        ├──── 0.2
        │         ├──── 20-60 keV .lc file
        │         ├──── 60-100 keV .lc file
        │         ├──── 100-200 keV .lc file
        │         └──── Master .lc file (20-200 keV)
        ├──── 0.4
        │         ├──── 20-60 keV .lc file
        │         ├──── 60-100 keV .lc file
        │         ├──── 100-200 keV .lc file
        │         └──── Master .lc file (20-200 keV)
        ├──── .   .
              .   .
              .
              .
```
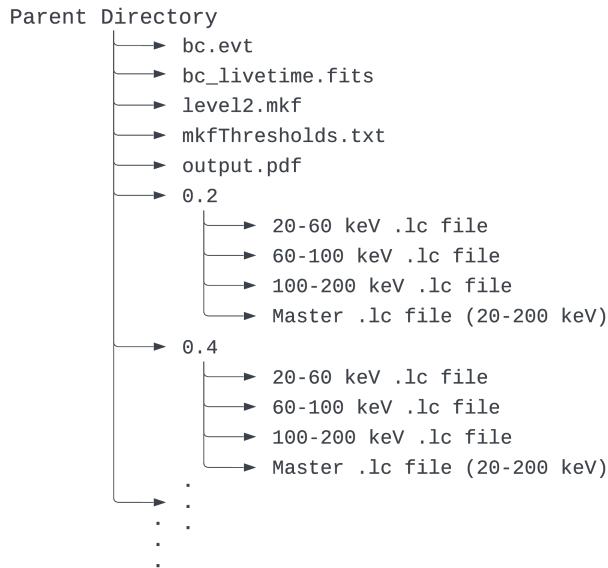
Figure 6.13: Final Directory Structure
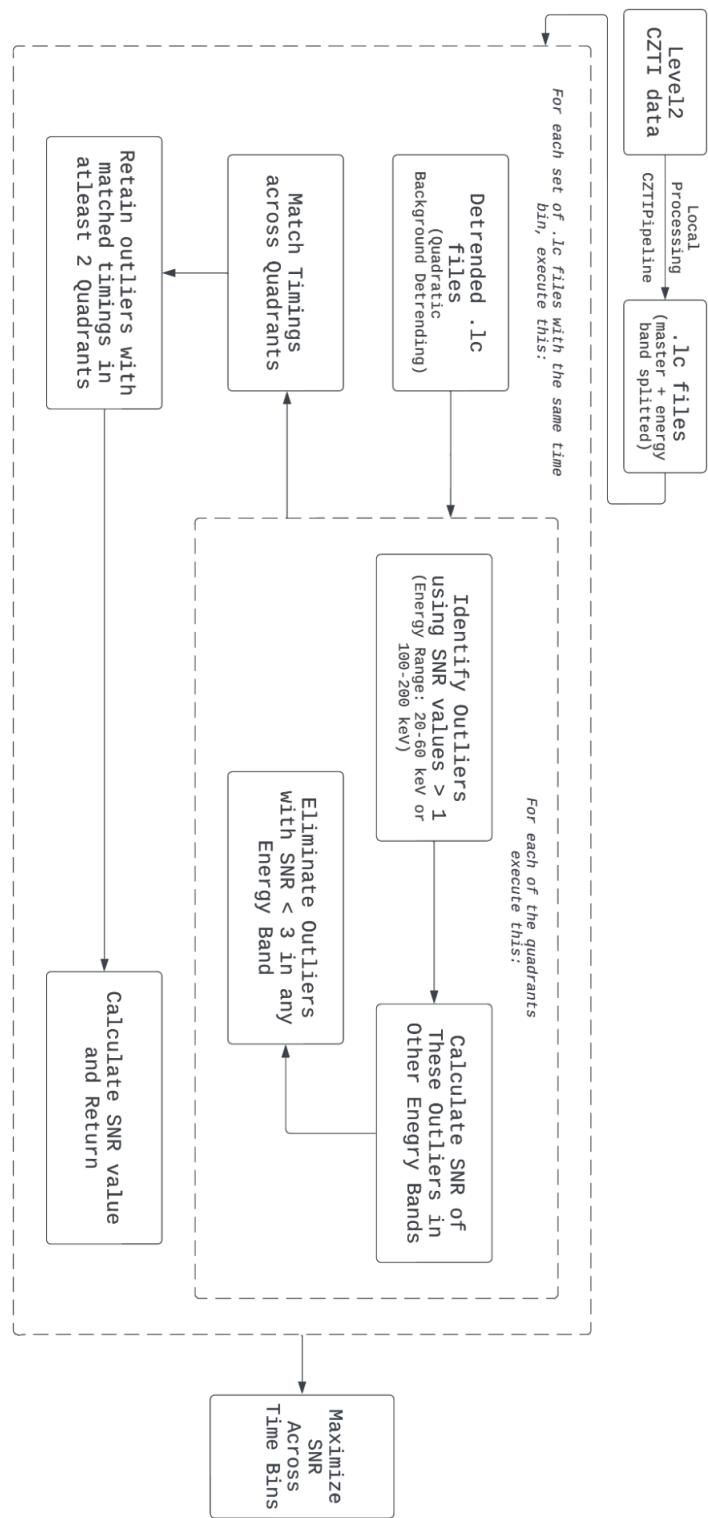
Figure 6.14: Final Algorithm Flowchart

IV

Results

7. Analysis and detection of GRBs

In our analysis using the pipeline, we successfully detected 9 out of the 10 given GRBs. The detected GRBs exhibit varying Signal-to-Noise Ratios and durations, indicating the diversity of these events.

However, one GRB was extremely faint and was not detected by our pipeline. The low SNR and/or brief duration of this event might have made it challenging to discern it from the background noise.

Overall, our pipeline demonstrated effective GRB detection capabilities, successfully identifying the majority of GRBs in the analyzed dataset. Further optimizations and refinements may enhance the pipeline's sensitivity, enabling more accurate detection of fainter GRBs in the future.

Name	Trigger Time (on archive)	Calculated Peak Time	Duration (on archive)	Best Timebin (s)	Q0	Q1	Q2	Q3
GRB160909A	211080464.0	211080477.5	-	7	6.1	7.27	-	-
GRB170330A	228608821.0	228608855.5	10	11	5.55	9.6	8.91	7.25
GRB190928A	307372337.0	307372434.5	119	11	112.78	107.86	147.74	127.56
GRB200803A	334121667.0	334121678.5	21	1	4.12	-	-	4.53
GRB200907B	337200672.0	337200671.3	1	0.2	3.83	-	6.51	7.56
GRB210515B	358786607.0	358786609.5	9	13	-	6.37	6.28	-
GRB210519A	359102334.9	359102336.5	25	9	4.11	6.44	-	-
GRB210709A	363502145.0	363502147.5	14	9	7.01	8.65	-	-
GRB210730A	365317050.0	365317049.6	3	0.8	7.36	11.23	-	-
GRB200920B	338327820.5	-	2		Not Detected			

Figure 7.1: Analysis and detection of 10 GRBs using our pipeline