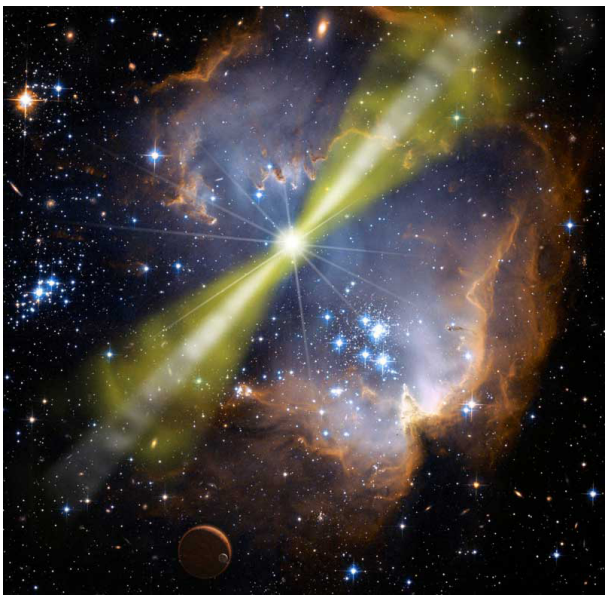
Gamma-Ray Bursts: Theory and Data Analysis

Final Report

Krittika Summer Project 6.0



Guru Jahnavi Madana
Department of Physics
Mentors:
Mehul Goyal and Yashovardhan Rai

"Gamma-ray bursts are the most powerful explosions since the Big Bang and yet they flicker away in the blink of an eye."

Contents

1 A Historical Primer on Gamma-Ray Bursts

1.1 A Lesson in Serendipity

The story of Gamma-Ray Bursts (GRBs) began not in an observatory, but in military secrecy. In the late 1960s, the Vela satellites - designed to monitor compliance with the Nuclear Test Ban Treaty — detected unexpected bursts of gamma rays that clearly did not originate from Earth. After further investigation, these bursts were confirmed to be of extraterrestrial origin.

This accidental discovery launched a completely new field in astrophysics. The bursts were short, unpredictable, and unlike anything observed before. Their origins were a mystery - but their energies hinted at something truly extreme and exotic.

GRB Phenomenology

GRBs are observed as brief, intense flashes of gamma rays, with durations ranging from milliseconds to several minutes. Their temporal profiles vary dramatically , from smooth, single-peaked bursts to chaotic, multi-spiked structures , suggesting a highly complex engine at work.

Spectrally, GRBs are non-thermal in nature, indicating that the emission mechanism is not simply blackbody radiation. Most GRB spectra can be described by a broken power law known as the Band function. The sheer diversity in light curves, durations, and spectral properties pointed toward multiple possible origins or progenitor systems.

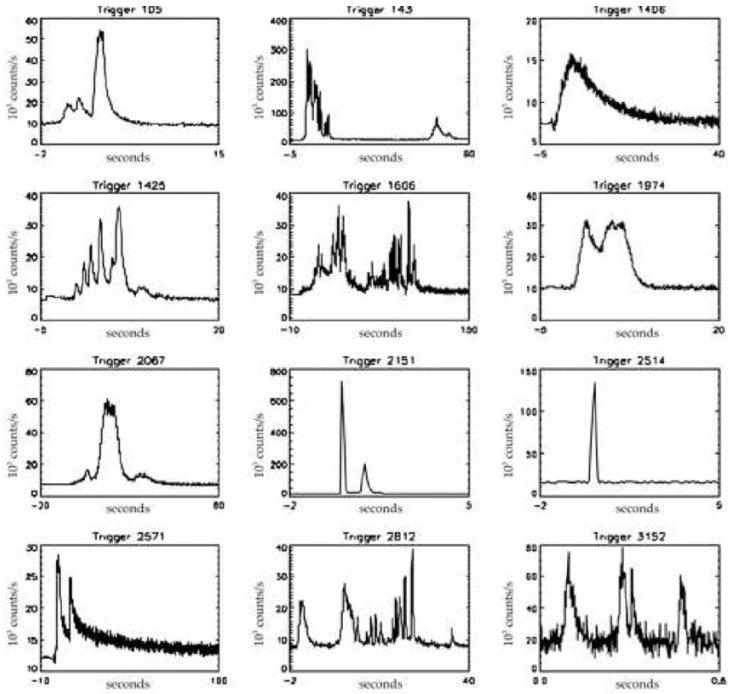


Figure 1: Examples of observed GRB light curves showing variability in shape and duration.

The Early Years

For decades after their discovery, GRBs remained an astrophysical mystery. In the 1970s and 80s, data were sparse and lacked precise localization. Theoretical models ranged from neutron star flares to alien signals, but without firm positional or distance information, none could be confirmed.

During this era, astronomers were flying blind ,observing some of the most powerful events in the Universe without any way to tell how far away they were or what caused them.

1.2 Suggested Models for GRB Creation

A range of models emerged to explain GRBs:

- **Neutron star starquakes:** Sudden crust rearrangements causing gamma flashes.
- **Magnetar flares:** Strong magnetic field-induced energy releases.
- **Black hole accretion events:** Matter falling into black holes releasing energy.
- **Exotic models:** Cosmic string cusps, evaporating primordial black holes, etc.

Each model attempted to match observed properties like variability timescales, energy output, and non-thermal spectra , but none could be definitively proven without knowing distances.

The Great Debate: Galactic or Cosmological?

Perhaps the most intense debate of the early GRB era was over their location. Were GRBs local (within our galaxy) or cosmological (billions of light-years away)?

Galactic advocates pointed to neutron star models (natal kicks), high event rates, and the known population of compact objects in the Milky Way.

Cosmological proponents argued that GRBs' isotropic distribution on the sky as shown by the BATSE experiment , suggested they were not associated with the Milky Way disk or halo. If GRBs were cosmological, the required energies would be truly enormous orders of magnitude greater than anything previously observed in astrophysics.

Eventually, redshift measurements from optical afterglows (first enabled by *BeppoSAX*) confirmed that GRBs occur at cosmological distances, decisively ending the debate and shifting the field's focus to understanding progenitors and emission mechanisms.

1.3 Intensive Efforts and Large Samples

A true leap forward came in the 1990s with the launch of NASA's *Compton Gamma Ray Observatory* and its onboard BATSE detector (Burst and Transient Source Experiment). Over its mission, BATSE detected more than 2700 GRBs, producing a massive sample that allowed for rigorous statistical analysis.

This large dataset enabled astronomers to study the spatial, temporal, and energetic distributions of GRBs and provided the first strong clues about their nature. It shifted the field from theoretical speculation to data-driven investigation.

The Angular Distribution of GRBs

One of the most compelling pieces of evidence for a cosmological origin came from angular sky maps of GRB positions recorded by BATSE. The distribution was found to be nearly perfectly isotropic — GRBs were spread evenly across the sky, with no apparent preference for the galactic plane or center.

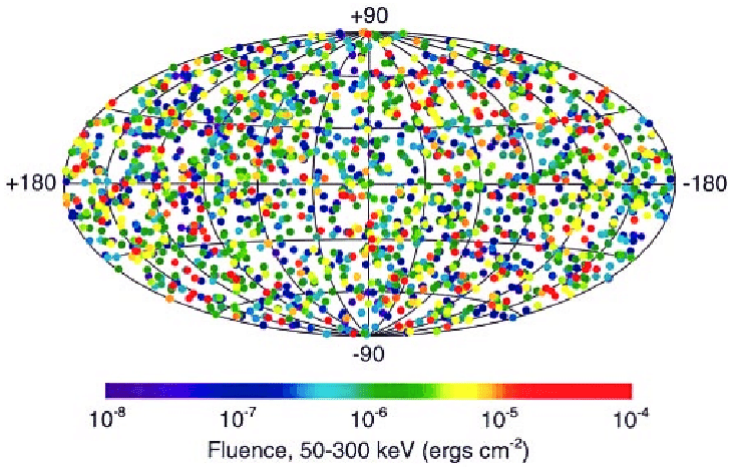


Figure 2: Sky distribution of GRBs detected by BATSE, showing an isotropic pattern.

This ruled out models where GRBs were predominantly located within our galaxy, whose stars are strongly concentrated in a disk. Isotropy was one of the earliest and strongest indicators that GRBs were not just powerful ,they were *distant* and *cosmological*.

Fluence or Luminosity Distributions

The fluence (total received energy) distribution of GRBs also supported a cosmological interpretation. In a uniform Euclidean universe, the number of sources above a given fluence should follow a power-law relationship:

$$N(> F) \propto F^{-3/2}$$

However, the observed distribution deviated significantly from this, with a flattening at low fluences. This effect is naturally explained by the cosmological model - bursts at higher redshifts appear fainter due to cosmic expansion and time dilation.

This deviation provided further evidence that GRBs were not local, but distributed across the observable universe.

Duration Distributions of GRBs

Statistical analysis of GRB durations revealed a clear bimodal structure:

- **Short GRBs:** Duration < 2 seconds
- **Long GRBs:** Duration > 2 seconds

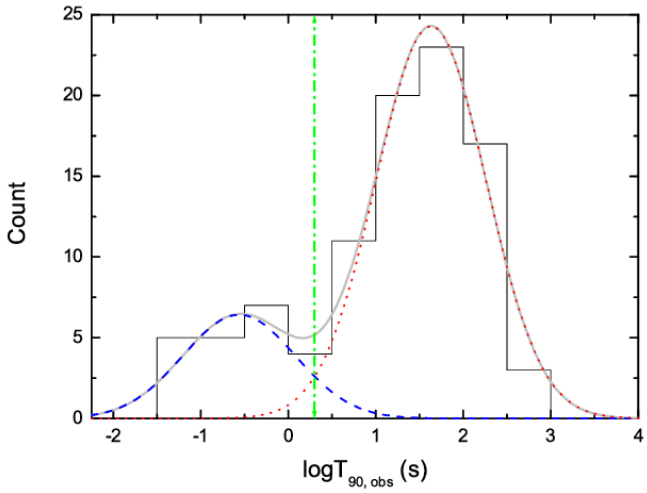


Figure 3: Duration histogram of GRBs showing bimodal distribution .

This discovery suggested that there were at least two distinct classes of GRBs, likely with different progenitor systems. Later observational evidence would support this, linking long GRBs to massive star collapses and short GRBs to compact binary mergers.

1.4 The Fireball Shock Model

To explain the non-thermal spectra and rapid variability of GRBs, the "fireball model" became the prevailing theoretical framework. In this model, a compact central engine (e.g., a newly formed black hole or magnetar) produces an ultra-relativistic outflow with a Lorentz factor $\Gamma 100$.

As internal variations in the flow collide (internal shocks), prompt emission is generated via synchrotron and inverse Compton mechanisms. Later, as the outflow interacts

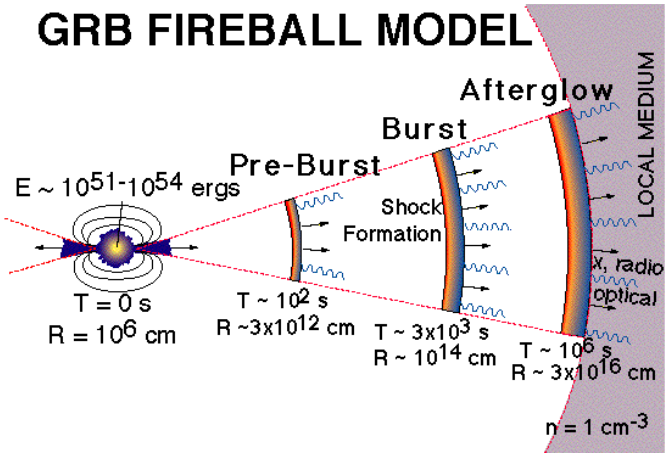


Figure 4: Schematic of the fireball shock model showing prompt (internal) and afterglow (external) shock zones.

with the ambient medium (external shocks), it produces the afterglow observable in X-ray, optical, and radio bands.

This model successfully explained many key GRB features:

- Rapid temporal variability
- Broad-band synchrotron emission
- Delayed onset of afterglows

1.5 The Long GRB Afterglow Revolution

A true turning point came in 1997 with the launch of *BeppoSAX*, which allowed rapid X-ray localizations of GRBs with arcminute precision. This led to the first successful detections of optical and radio afterglows.

The discovery of afterglows revolutionized GRB science by:

- Allowing measurement of GRB redshifts.
- Identifying host galaxies.
- Confirming the cosmological origin of GRBs.

It also provided support for the external shock model and gave insights into the circumburst environment.

Redshifts and Host Galaxies

Thanks to afterglow detections, spectroscopic redshifts of GRBs could be obtained, ranging from $z \sim 0.1$ to beyond $z \sim 8$. Most long GRBs were found in star-forming, irregular galaxies often with high specific star formation rates.

This bolstered the collapsar model, suggesting that long GRBs originate from the core-collapse of massive stars in star-forming regions.

1.6 The Supernova Connection

Perhaps the most compelling evidence for the death of massive stars as the progenitors of long GRBs came from the discovery of supernovae coinciding with GRBs.

Notable examples include:

- **GRB 980425 / SN 1998bw:** The first observed GRB-supernova connection.
- **GRB 030329 / SN 2003dh:** A clear spectroscopic match with a Type Ic supernova.

These events involved highly energetic, asymmetric explosions, sometimes termed 'hypernovae'. The presence of broad-lined Type Ic features suggested a stripped-envelope massive progenitor with a relativistic jet punching through the stellar core.

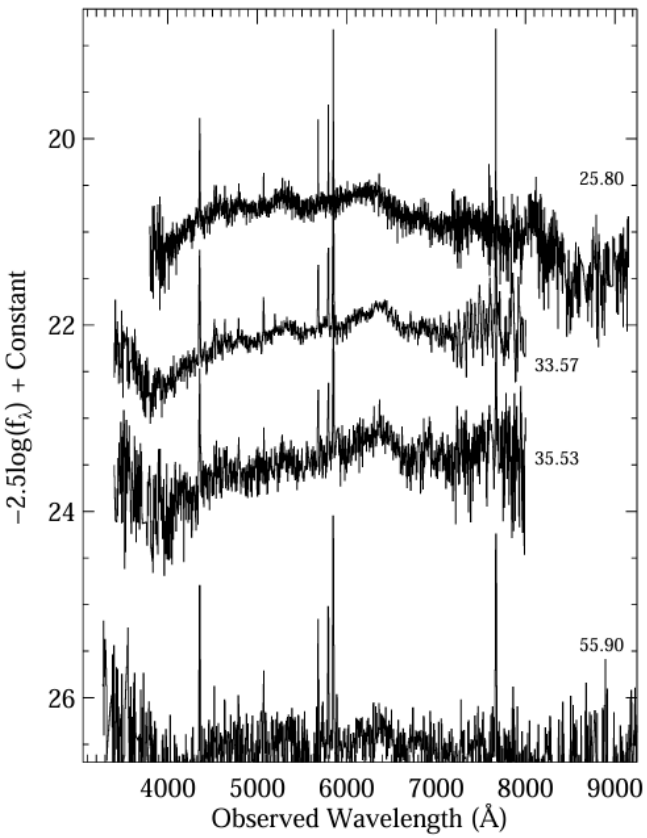


Figure 5: Optical light curves of GRB afterglows showing a supernova bump consistent with Type Ic hypernovae.

GRB Energetics

Assuming isotropic emission, GRBs can emit up to 10^{54} erg of energy in gamma rays ,equivalent to converting the entire mass of the Sun into energy in seconds. However, evidence of beaming implies that true energy outputs are much lower, typically around 10^{50} to 10^{51} erg.

Jet breaks observed in afterglow light curves help constrain the jet opening angle θ_j, allowing for estimates of the "beaming-corrected" energy:

$$E_{\gamma,true} = E_{\gamma,iso} \times (1 - \cos\theta_j)$$

Despite the correction, GRBs remain among the most energetic explosions known in the Universe.

1.7 The Swift Era

Launched in 2004, NASA's *Swift* satellite fundamentally transformed GRB astronomy. With three onboard instruments, namely, BAT (Burst Alert Telescope), XRT (X-ray Telescope), and UVOT (Ultraviolet/Optical Telescope), Swift could detect GRBs, localize them within seconds, and automatically slew to follow up with X-ray and optical imaging.

Swift's rapid response allowed:

- Near-instantaneous afterglow detection.
- Identification of host galaxies.
- Real-time multi-wavelength monitoring.

It enabled redshift measurements for hundreds of GRBs and was critical in discovering afterglows of short GRBs, which had previously been elusive.

1.7.1 Short GRB Afterglows

Prior to Swift, short GRBs had no confirmed afterglows. Swift's quick repointing allowed the detection of faint X-ray afterglows for several short GRBs. These events were typically fainter and shorter-lived than those from long GRBs, and many occurred outside the visible regions of host galaxies.

This suggested a different progenitor population and ejection mechanism compared to long GRBs.

1.7.2 Short GRB Progenitors

The localization of short GRBs led to their association with older stellar populations and elliptical galaxies, supporting the idea that they originate from compact binary mergers (neutron star–neutron star or neutron star–black hole systems).

Their lower energy, shorter duration, and weak or absent supernova signatures all pointed toward a merger scenario. This idea was dramatically confirmed by the multi-messenger detection of GW170817 and GRB 170817A.

1.8 High-Redshift GRBs

Swift also discovered GRBs at extremely high redshifts, such as GRB 090423 ($z \sim 8.2$), making GRBs some of the most distant objects ever observed. These events provide a window into:

- Star formation in the early Universe.
- Reionization-era environments.
- Pop III stellar deaths (theoretical possibility).

High-z GRBs are valuable cosmic probes due to their brightness and detectability at great distances.

1.9 New Classes of GRBs

As observational capabilities improved, new types of GRBs were identified:

- **Ultra-long GRBs:** Events lasting thousands of seconds (e.g., GRB 111209A), possibly linked to blue supergiants or TDEs.
- **Low-luminosity GRBs:** Sub-energetic events closer to Earth, often with clear supernova associations.
- **X-ray flashes (XRFs):** Softer spectral peaks, possibly viewed off-axis or arising from different jet structures.

These discoveries expanded the GRB classification scheme beyond the simple "short vs long" dichotomy.

1.10 New Insights from Fermi

The *Fermi Gamma-ray Space Telescope* (launched 2008) broadened the GRB observational window to GeV energies via its LAT (Large Area Telescope) and GBM (Gamma-ray Burst Monitor). Fermi observed:

- High-energy photons (> 100 MeV) arriving delayed relative to keV–MeV emission.
- Extended GeV afterglows lasting minutes to hours.
- Evidence of multiple emission zones or jet components.

Fermi data challenged simple synchrotron-only models and opened possibilities for hadronic processes in GRB jets.

1.11 Multimessenger Astronomy

Recent years have seen GRBs tied to gravitational wave and neutrino observations, giving rise to multimessenger astrophysics- a field where information is collected from both electromagnetic and non-EM messengers.

1.11.1 Gravitational Waves

The watershed event occurred in 2017 with GW170817, a binary neutron star merger detected by LIGO/Virgo. Just 1.7 seconds later, the Fermi GBM and INTEGRAL satellites detected a weak short GRB (GRB 170817A).

This confirmed that short GRBs can be the EM counterparts of gravitational wave events and demonstrated:

- The role of compact object mergers in GRB production.
- A new method to independently measure cosmological parameters (e.g., Hubble constant).
- A direct connection between kilonovae and GRBs.

1.11.2 High-Energy Neutrinos

GRBs have long been suspected as potential sources of ultra-high energy cosmic rays and neutrinos. Facilities like IceCube aim to detect neutrino counterparts to GRBs, particularly from hadronic interactions in the jet.

While no unambiguous neutrino-GRB association has yet been confirmed, ongoing observations continue to test jet composition models and GRB particle acceleration mechanisms.

2 Prompt Emission

2.1 Observational Properties

The prompt emission phase is what defines a Gamma-Ray Burst- it is the initial flash of high-energy gamma rays detected by satellites like *Swift* and *Fermi*. This phase typically lasts from milliseconds to a few hundred seconds and shows an incredible diversity in shape, intensity, and duration.

The emission occurs in the keV–MeV range, sometimes extending to GeV energies (especially with Fermi-LAT observations). Despite being extremely luminous, prompt emission is short-lived and challenging to catch across multiple wavelengths in real time.

2.2 Temporal Structure

Prompt emission light curves are highly variable and complex. Some GRBs exhibit a single pulse, while others show multiple spikes with substructure and rapid fluctuations on millisecond timescales.

This rich variability implies a compact emission region and an engine capable of producing fast, episodic outflows. The short variability timescale Δt constrains the size of the emitting region to:

$$Rc\,\Delta t\,\Gamma^2$$

where Γ is the Lorentz factor. These constraints point to a highly relativistic outflow from a small, central engine.

2.3 Spectral Structure

Spectrally, prompt emission is typically non-thermal and well described by the empirical Band function—a smoothly broken power law with low-energy and high-energy photon indices and a peak energy E_{peak}.

Key features include:

- A νF_ν spectrum peaking around a few hundred keV.
- Diversity in spectral shapes across GRBs.
- Temporal evolution of E_{peak} within bursts.

While synchrotron emission is a popular explanation, some observed spectra are too hard to be accounted for by standard synchrotron theory, leading to alternative suggestions like photospheric emission or Comptonized components.

2.4 Polarization

Although difficult to measure, several GRBs have shown significant polarization in the prompt phase. Reported polarization degrees range from $\sim 10\%$ to as high as $\sim 80\%$, suggesting an ordered magnetic field structure at the emission site.

Polarization measurements provide unique insights into:

- Jet geometry
- Magnetic field configuration
- Emission mechanisms (e.g., synchrotron vs Comptonized photosphere)

Despite current challenges in statistical significance and sample size, polarization remains a powerful diagnostic tool for future missions.

2.5 Origin of the Prompt Emission

The physical origin of prompt emission is still an open question, but three primary models dominate:

1. **Internal shocks:** Collisions between shells of different velocities within a relativistic jet. Natural explanation for variability but relatively inefficient energetically.
2. **Photospheric emission:** Radiation released as the jet becomes optically thin. Can produce quasi-thermal features but requires additional broadening mechanisms to match observations.
3. **Magnetic reconnection:** Energy release from dissipation of magnetic fields in a Poynting-flux dominated jet. Promising for highly magnetized GRBs.

Current data suggest that multiple emission zones or mechanisms may coexist, and that different GRBs may be dominated by different physical processes.

3 Afterglows

After the brief and intense prompt emission, many GRBs are followed by longer-lasting, multi-wavelength afterglows that decay over time. These afterglows arise from interactions between the GRB outflow and the surrounding medium and can be observed across X-ray, optical, infrared, and radio bands.

3.1 The First Afterglow Searches

Before 1997, no optical or X-ray counterpart to a GRB had ever been observed. That changed with the *BeppoSAX* satellite, which provided arcminute-scale localizations. This allowed ground-based telescopes to detect the first fading afterglow of GRB 970228, confirming that GRBs had observable follow-up emission.

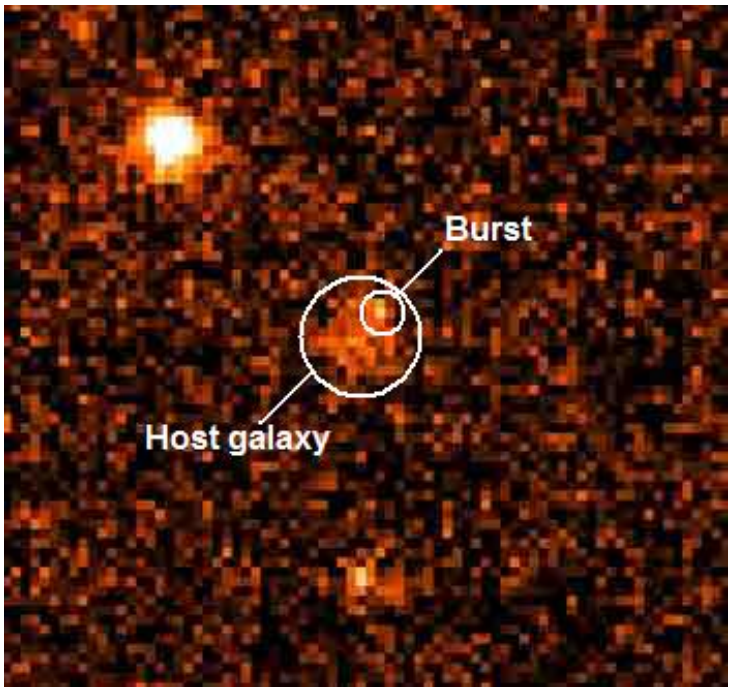


Figure 6: First detected optical afterglow from GRB 970228. This discovery confirmed the afterglow phenomenon.

3.1.1 X-ray Afterglows

X-ray afterglows are the most common and longest-lasting of all bands. They often appear within minutes and can persist for days or even weeks. The *Swift* XRT instrument has observed hundreds of X-ray afterglows, revealing rich structures in their temporal evolution.

These emissions are attributed to synchrotron radiation from forward shocks produced when the GRB ejecta plow into the surrounding medium.

3.2 The Canonical X-ray Afterglow

XRT light curves often follow a "canonical" shape with five distinct segments:

1. Steep decay (internal origin)
2. Plateau phase (energy injection)
3. Normal decay
4. Late-time steepening (jet break)
5. Flares (re-energization or residual activity)

The variety in X-ray behavior provides deep insights into central engine physics, jet collimation, and the density of the external medium.

3.3 Optical Afterglows

Optical afterglows are detected in a significant fraction of GRBs and typically fade over timescales of hours to days. They allow redshift determination via spectroscopy and localization of the host galaxy.

Optical light curves often exhibit power-law decay:

$$F(t) \propto t^{-\alpha}$$

with $\alpha \sim 1$ to 2, depending on viewing angle, jet structure, and environment.

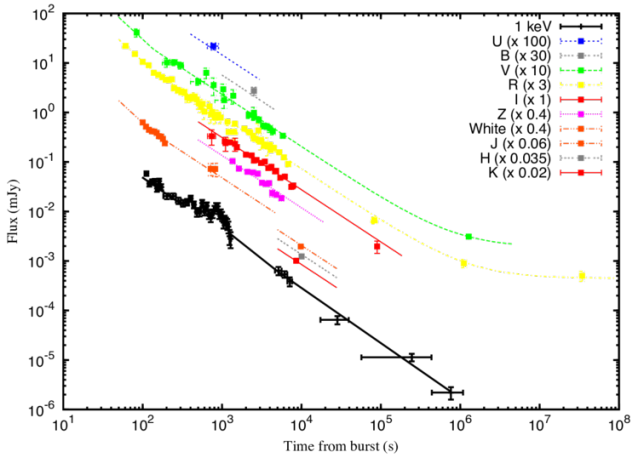


Figure 7: Example of an optical afterglow light curve, enabling redshift determination and host galaxy studies.

3.4 Dark GRBs

Despite sensitive observations, some GRBs show no optical afterglow, these are termed "dark GRBs." Possible causes include:

- High redshift ($z > 6$) causing Lyman-alpha absorption
- Dust extinction in the host galaxy
- Intrinsically faint optical emission

X-ray afterglows can still be detected for these events, helping identify them and analyze their environments.

3.5 Polarisation

Like prompt emission, afterglows can exhibit linear polarization due to synchrotron radiation. Measurements in optical afterglows suggest polarization levels of $\sim$ 1–10%, often varying with time.

Polarization provides clues about magnetic field geometry in the jet and the presence of ordered fields vs turbulence in the emission region.

3.6 Radio / Sub-mm Afterglows

Radio afterglows are rarer but longer-lived, sometimes detectable weeks or even months after the GRB. These emissions come from the slowest ejecta components and probe the late-time evolution of the blast wave.

Sub-millimeter observations (e.g., with ALMA) provide additional data to constrain jet breaks, external medium profiles, and total energetics.

4 Data Analysis Recipes: Using Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is an essential tool for exploring complex, high-dimensional posterior probability distributions in Bayesian inference. It becomes necessary in scenarios where analytic solutions or standard sampling techniques are impractical.

Problem Setting:

Suppose you have a model with parameters $\theta \in R^n$ and observed data D. You are interested in computing or sampling from the posterior distribution:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)}$$

Here,

- $P(\theta \mid D)$ is the posterior probability of the parameters given the data.
- $P(D \mid \theta)$ is the likelihood of the data given the parameters.
- $P(\theta)$ is the prior distribution over the parameters.
- $P(D)$ is the evidence or marginal likelihood.

In high-dimensional spaces or with complicated likelihoods, computing $P(D)$ or directly sampling from the posterior is intractable. This is where MCMC excels.

4.1 When Do You Need MCMC?

You need MCMC when:

1. The posterior distribution $P(\theta \mid D)$ cannot be computed in closed form.
2. The posterior is high-dimensional, multimodal, or has complicated correlations between parameters.
3. You want samples from the posterior to estimate expectations:

$$E[f(\theta) \mid D] = \int f(\theta)\, P(\theta \mid D)\, d\theta$$

Key Insight: MCMC generates a *correlated sequence* of samples $\{\theta_i\}_{i=1}^{N}$ that approximate the target posterior distribution. The beauty of MCMC is that it doesn't require knowledge of the normalization constant $P(D)$.

4.2 What Is a Sampling?

In the context of Bayesian inference, "a sampling" refers to a collection of samples drawn from a target probability distribution — typically the posterior $P(\theta \mid D)$. Rather than describing the distribution analytically, MCMC lets us approximate it using discrete samples.

4.2.1 Mathematical Interpretation

Let $\{\theta_i\}_{i=1}^N$ be a set of samples drawn from the posterior distribution. Then, for any integrable function $f(\theta)$, we can estimate its expectation as:

$$E[f(\theta)] = \int f(\theta)\, P(\theta \mid D)\, d\theta \approx \frac{1}{N} \sum_{i=1}^{N} f(\theta_i)$$

This approximation becomes increasingly accurate as $N \to \infty$ — provided the samples are representative of the underlying distribution.

4.2.2 Properties of a Good Sampling

A good sampling should:

- Cover the full support of the posterior.
- Capture the shape and correlations within the distribution.
- Have minimal autocorrelation between samples.
- Allow accurate approximation of integrals and expectations.

4.2.3 Visual Intuition

Each sample can be thought of as a "probe" into the high-dimensional posterior landscape. A dense clustering of samples in a region implies high posterior density there.

4.2.4 Monte Carlo Estimation

The term "Monte Carlo" comes from estimating quantities via random sampling. If $\theta_i \sim P(\theta \mid D)$, then:

$$\int f(\theta)\, P(\theta \mid D)\, d\theta \approx \frac{1}{N} \sum_{i=1}^{N} f(\theta_i)$$

This bypasses the need for analytical integration, which is often intractable in high dimensions.

Takeaway

Sampling is not merely about generating random points — it's about generating points that represent the underlying distribution so that probabilistic reasoning (like expectation, variance, credible intervals) can be approximated effectively.

4.3 Metropolis–Hastings MCMC

Markov Chain Monte Carlo (MCMC) algorithms are used to draw samples from a probability distribution when direct sampling is difficult. The Metropolis–Hastings (MH) algorithm is one of the most fundamental and widely-used MCMC techniques. It constructs a Markov chain whose stationary distribution is the target posterior $P(\theta \mid D)$.

4.3.1 The Basic Idea

We wish to sample from a posterior distribution $P(\theta \mid D)$ without needing to compute the intractable evidence $P(D)$. The MH algorithm constructs a sequence $\{\theta_i\}$ such that the stationary distribution of the chain is the posterior.

The key idea is to propose a new sample θ' given the current sample θ using a proposal distribution $q(\theta' \mid \theta)$, and then decide probabilistically whether to accept or reject it.

4.3.2 The Algorithm

1. Initialize the chain with some starting point θ_0.
2. For each iteration i:
    - Propose a new point θ' from a proposal distribution $q(\theta' \mid \theta_i)$.
    - Compute the acceptance ratio:

$$r = \frac{P(D \mid \theta')P(\theta') \cdot q(\theta_i \mid \theta')}{P(D \mid \theta_i)P(\theta_i) \cdot q(\theta' \mid \theta_i)}$$

    - Accept θ' with probability $\min(1, r)$. If accepted, set $\theta_{i+1} = \theta'$; else, set $\theta_{i+1} = \theta_i$.

4.3.3 Properties of the Chain

The Markov chain created this way satisfies detailed balance:

$$P(\theta)q(\theta' \mid \theta)A(\theta' \mid \theta) = P(\theta')q(\theta \mid \theta')A(\theta \mid \theta')$$

This ensures that $P(\theta)$ is the stationary distribution of the chain, provided the chain is also irreducible and aperiodic.

4.3.4 Burn-in and Convergence

The early samples of the chain are not representative of the true posterior, especially if the chain is initialized far from high-probability regions. These initial steps are called the *burn-in* phase and are typically discarded.

Convergence diagnostics, such as visual inspection, Gelman–Rubin statistic, or autocorrelation analysis, are used to determine if the chain has sufficiently mixed.

4.3.5 Symmetric Proposals

If the proposal distribution is symmetric, i.e.,

$$q(\theta' \mid \theta) = q(\theta \mid \theta'),$$

then the acceptance ratio simplifies to:

$$r = \frac{P(D \mid \theta')P(\theta')}{P(D \mid \theta)P(\theta)}$$

This is the classical Metropolis algorithm, a special case of Metropolis–Hastings.

4.3.6 Choosing a Proposal Distribution

Choosing a good proposal distribution q is crucial. A narrow q leads to slow exploration (low acceptance rate), while a very wide q results in most proposals being rejected. Ideally, q should balance local exploration and global movement across modes.

Common choices include:

- Gaussian centered at the current point.
- Adaptive proposals based on past covariance.
- Mixture models or heavy-tailed proposals for multimodal posteriors.

4.3.7 Acceptance Fraction

The acceptance fraction ,the ratio of accepted proposals to total proposals, is a key indicator of sampler performance. Typical target values are between 0.2 and 0.5, but the optimal value depends on dimensionality.

If the acceptance rate is too low, the sampler is stuck; if too high, the steps are too small. Tuning the scale of the proposal distribution is essential.

4.3.8 Example: A Simple Gaussian Posterior

Consider a one-dimensional Gaussian likelihood:

$$P(D \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(D-\theta)^2}{2\sigma^2}\right)$$

Assuming a flat prior on θ, the posterior is also Gaussian. The MH algorithm can be used to sample from this posterior, validating the correctness of the implementation.

Key Insight: The MH algorithm builds a Markov chain that mimics the target posterior using only local moves and accept/reject decisions. It avoids the need for direct normalization or integration.

4.4 Likelihoods and Priors

Bayesian inference is built on two key ingredients: the likelihood and the prior. Their combination, via Bayes' theorem, defines the posterior probability distribution that we wish to explore using MCMC.

4.4.1 Likelihood Function

The likelihood function $P(D \mid \theta)$ quantifies how probable the observed data D are given the model parameters θ. It embodies the statistical model of the measurement process.

$$P(D \mid \theta) = \mathcal{L}(\theta)$$

It is often defined based on assumptions like:

- Gaussian noise:
$$\mathcal{L}(\theta) \propto \exp\left(-\frac{1}{2}\sum_i \left(\frac{y_i - f(x_i;\theta)}{\sigma_i}\right)^2\right)$$
- Poisson likelihoods for count data:
$$\mathcal{L}(\theta) = \prod_i \frac{\lambda_i^{k_i} e^{-\lambda_i}}{k_i!}, \quad \lambda_i = f(x_i;\theta)$$
- Bernoulli or Binomial likelihoods for binary data.

The form of the likelihood is dictated by the measurement process and noise model , not by convenience.

4.4.2 Prior Distribution

The prior $P(\theta)$ encodes our knowledge or belief about the parameters before seeing the data. Priors can be:

- **Uninformative**: Flat or weakly informative; e.g., uniform over a large domain.
- **Informative**: Based on previous experiments, theory, or strong expectations.
- **Improper**: Not normalized, but still usable if the posterior is normalizable.

$$P(\theta) = belief about parameters before data$$

Choosing appropriate priors is subtle and problem-dependent. Poor choices can bias the inference, while good priors can regularize underconstrained models.

4.4.3 Posterior Distribution

Combining likelihood and prior via Bayes' theorem gives:

$$P(\theta \mid D) = \frac{P(D \mid \theta)\, P(\theta)}{P(D)} \propto P(D \mid \theta)\, P(\theta)$$

The denominator $P(D)$ (called the evidence) serves as a normalization constant and is often ignored in MCMC, since MCMC only requires the unnormalized posterior.

4.4.4 Log-likelihood and Log-prior

For numerical stability and efficiency, the log-posterior is often used:

$$\log P(\theta \mid D) = \log P(D \mid \theta) + \log P(\theta) + constant$$

This converts products into sums and is especially helpful when working with very small probability values.

4.4.5 Practical Tips

- Always visualize the likelihood and prior separately to understand their influence.
- Use priors to encode physical constraints (e.g., $\theta > 0$).
- In hierarchical models, priors play a larger role .
- Flat priors are not always "non-informative"; reparameterization can change that.

Key Insight: The likelihood maps models to data, and the prior maps belief to parameters. Together, they define the landscape that MCMC explores.

4.5 Autocorrelation

In MCMC sampling, successive samples are generally not independent. The degree to which a sample depends on previous samples is measured by the **autocorrelation**.

High autocorrelation implies that samples are strongly dependent on each other and contain less independent information, reducing the effective sample size.

Low autocorrelation is desirable , it means the chain explores the posterior efficiently, producing nearly independent samples.

5 The Multi-Mission Maximum Likelihood Framework

5.1 Introduction

Multi-wavelength observations are essential to understanding astrophysical sources and their emission mechanisms. However, combining data from instruments across different energy bands is challenging due to variations in technology, software, and noise characteristics.

The **3ML** (Multi-Mission Maximum Likelihood) framework aims to solve this by providing a common structure to perform joint likelihood analyses using multiple instruments, each with its own response characteristics.

In a typical observation, the sky emits a true flux $S(E, P)$ from energy E and position P. Instruments detect this signal with some response $R(e, p \mid E, P)$, which includes effects like limited energy resolution (energy dispersion- D), imperfect angular resolution (PSF), and effective area.

This response is often factorized as:

$$R(e, p \mid E, P) = A(E, P) \cdot PSF(E, P, p) \cdot D(e \mid E)$$

The observed count rate in a detector is then modeled by integrating this response over the source flux. Instruments bin the data spatially and spectrally, and the actual counts in each bin are modeled as random variables (e.g., Poisson-distributed due to photon noise).

The core problem: Given observed data and known instrument responses, infer the true source flux $S(E, P)$. This can be done by either:

1. Forward-folding a model through the response and comparing it to data,
2. Attempting deconvolution (often unstable) to reconstruct the signal.

3ML enables the first strategy in a unified, modular way -critical for modern multi-instrument astrophysics.

5.2 Multi-wavelength Analysis of a Source

Multi-wavelength observations provide rich information about astrophysical sources, but each instrument records data in its own format and resolution. To extract meaningful physical insight, we must relate these diverse data sets to a common representation of the source , typically its spectral energy distribution (SED).

5.2.1 Bringing the Signal Back to the Reality Domain: Spectral Energy Distribution

One traditional approach involves unfolding the observed data to estimate the true source flux - i.e., reconstructing the SED in the "reality domain." This process involves correcting for the instrument's response by attempting to invert it.

While appealing in principle, this deconvolution is often unstable or biased:

- It amplifies noise in low-count regimes,
- It relies on regularization, which can suppress real features,
- The uncertainties are hard to propagate rigorously.

As a result, unfolded SEDs are best treated as visual summaries, not inputs to quantitative modeling.

5.2.2 Bringing the Model to the Data Domain: Forward Folding

A more robust approach is to work entirely in the data domain. In this method, a physical model for the source is proposed in the reality domain (e.g., a synchrotron spectrum), then "folded" through the instrument response R to predict expected counts in each detector bin.

This predicted data is then compared to actual observations using an appropriate statistical method (e.g., likelihood analysis based on Poisson noise). This strategy:

- Preserves full statistical rigor,
- Avoids unstable deconvolutions,
- Allows simultaneous fitting across instruments.

3ML adopts this forward-folding approach as its core philosophy, enabling coherent multi-instrument analyses by modeling the full system: source $\rightarrow$ instrument $\rightarrow$ data.

5.3 The Multi-Mission Maximum Likelihood Framework (3ML)

The 3ML framework is designed to make forward-folded, multi-instrument analyses modular and reproducible. It does this by abstracting each instrument's specifics behind a common interface, allowing different backends to communicate with a unified statistical model.

At the heart of 3ML is the concept of a **plugin**: a lightweight wrapper that connects the shared model to the instrument's native analysis software. Each plugin knows how to:

- Read the instrument's data and response files,

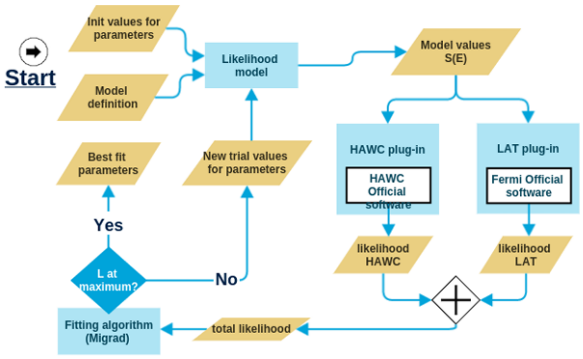


Figure 8: 3ml framework

- Translate model parameters into expected counts,
- Compute the instrument-specific likelihood contribution.

The total likelihood is then formed as the product (or sum of log-likelihoods) of all plugin contributions:

$$\mathcal{L}_{total} = \prod_i \mathcal{L}_i^{plugin}$$

This design makes it possible to:

- Perform joint fits across instruments with differing characteristics,
- Leverage existing mission-specific software without modification,
- Incorporate complex models and priors through external samplers (e.g., MCMC).

3ML thus acts as a flexible, scalable bridge between theoretical modeling and real multi-wavelength data — enabling consistent, high-quality scientific inference.

Installation and setup of threeml is done - going through the tutorials from the threeml package examples in jupyter notebooks .

6 Tasks Completed

6.1 MCMC Tutorial: Bayesian Straight-Line Fit

To build a foundational understanding of Bayesian inference and MCMC, I implemented the standard example of linear regression with uncertainties using the `emcee` package. The task involved fitting a straight line to synthetic data with Gaussian noise, treating the slope and intercept as parameters to sample. I defined the log-likelihood, prior, and posterior, and used a small ensemble of walkers to explore the parameter space. The results were visualized through the marginalized posterior distributions and the resulting best-fit line, consistent with standard Bayesian linear regression.

Notebook: tutorial_for_st_line.ipynb

6.2 Metropolis–Hastings Sampler Implementation

To deepen my understanding of the Metropolis–Hastings (MH) algorithm, I implemented a simple M–H sampler in Python from scratch and applied it to several progressively complex sampling problems, based on the exercises from the *Data Analysis Recipes: Using MCMC* paper.

Problems Implemented:

- **Univariate Gaussian:** Sampled from a Gaussian distribution with mean 2 and variance 2 using a symmetric proposal distribution.
- **Uniform Distribution:** Modified the input density to be uniform on $3 < x < 7$, requiring adjustments to initialization and acceptance rules.
- **Bivariate Gaussian:** Sampled from a correlated 2D Gaussian with a nontrivial covariance matrix, visualized using scatter plots and marginal histograms.
- **Top-hat Distribution:** Sampled from a 2D uniform distribution over a rectangular region, comparing its behavior to the Gaussian case.
- **Proposal Variance Study:** Explored the effects of different proposal variances (too small and too large) on convergence and efficiency.
- **Shifted Proposal Mean:** Tested what happens if the proposal mean is not centered on the current state, highlighting the importance of symmetric or well-chosen proposals.

Throughout, I visualized the histograms of sampled parameters and analyzed acceptance rates and chain behavior, building strong intuition about tuning and sampler performance.

Notebook: MH_MCMC_Sampler.ipynb