Commentary on the paper by T.-P. Li and Y.-Q. Ma "Analysis methods for results in gamma-ray astronomy", ApJ **272** (1983) 317

Edoardo Milotti

Advanced Statistics for Physics course

A.Y. 2023-24

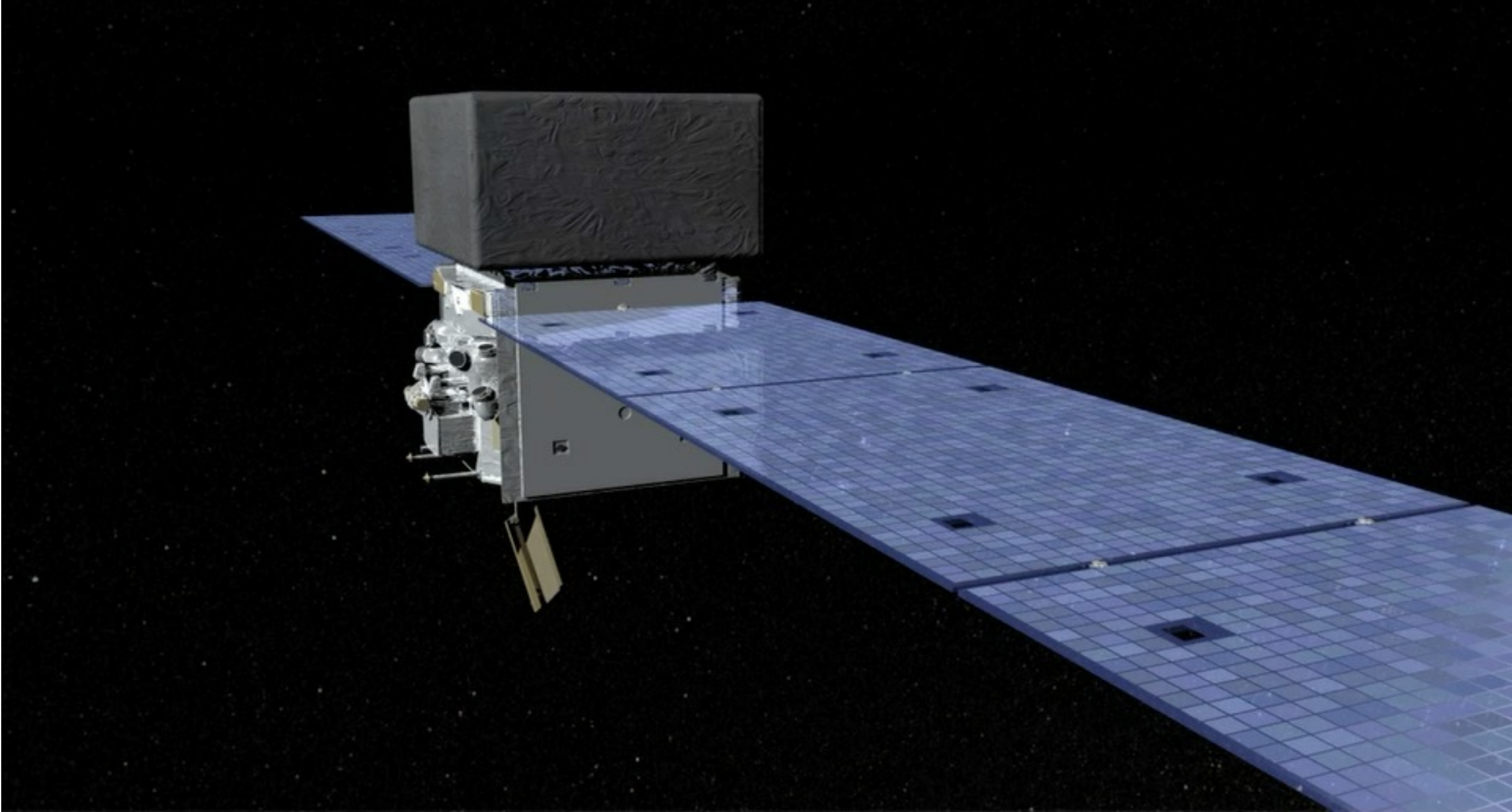
The Fermi Gamma-ray Space Telescope

The Universe is home to numerous exotic and beautiful phenomena, some of which can generate almost inconceivable amounts of energy. Supermassive black holes, merging neutron stars, streams of hot gas moving close to the speed of light … these are but a few of the marvels that generate gamma-ray radiation, the most energetic form of radiation, billions of times more energetic than the type of light visible to our eyes. What is happening to produce this much energy? What happens to the surrounding environment near these phenomena? How will studying these energetic objects add to our understanding of the very nature of the Universe and how it behaves?

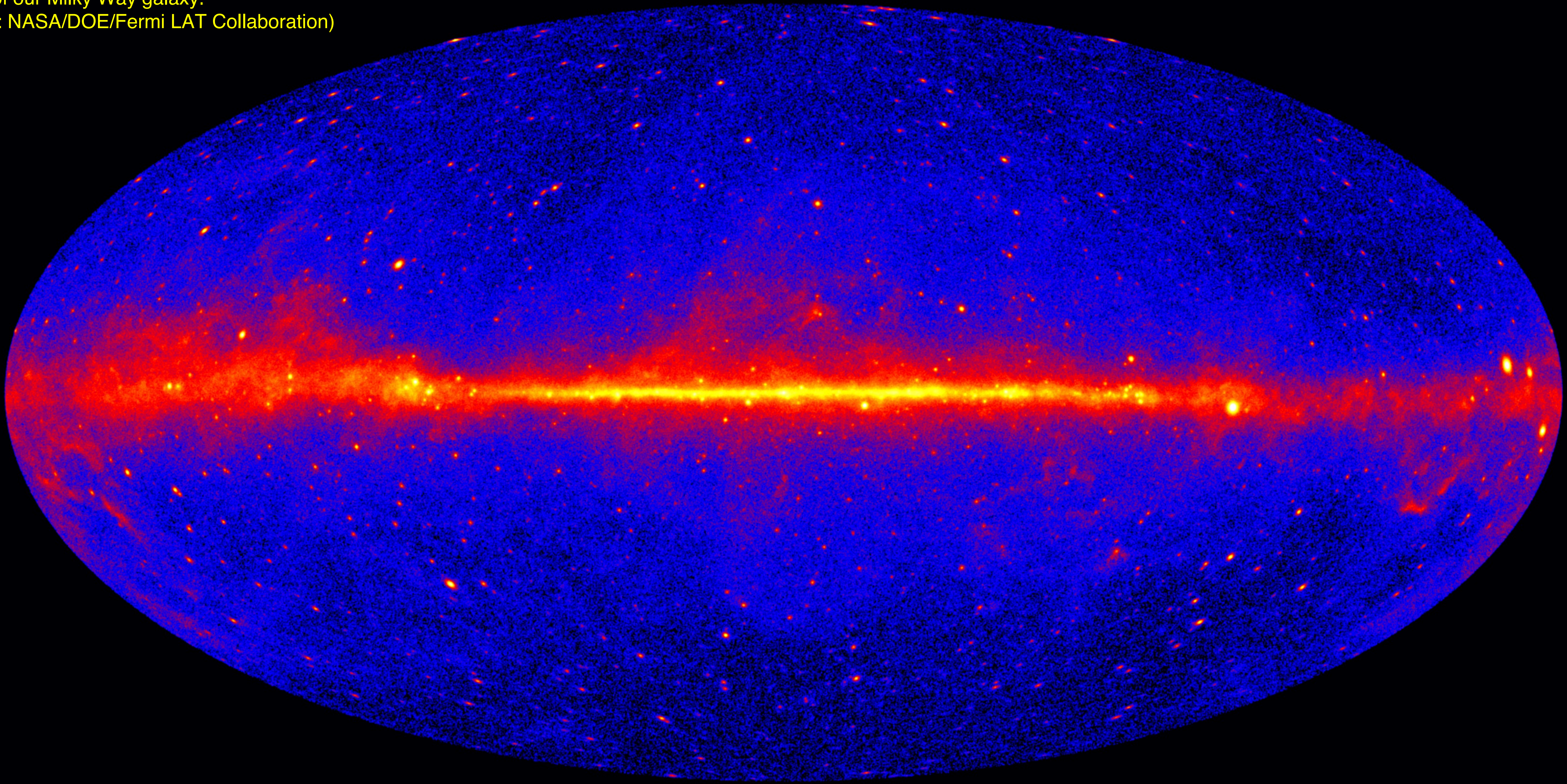
The **Fermi Gamma-ray Space Telescope**, formerly GLAST, is opening this high-energy world to exploration and helping us answer these questions. With Fermi, astronomers have a superior tool to study how black holes, notorious for pulling matter in, can accelerate jets of gas outward at fantastic speeds. Physicists are able to study subatomic particles at energies far greater than those seen in ground-based particle accelerators. And cosmologists are gaining valuable information about the birth and early evolution of the Universe.

(adapted from https://fermi.gsfc.nasa.gov)

https://imagine.gsfc.nasa.gov/observatories/learning/fermi/mission/lat.html

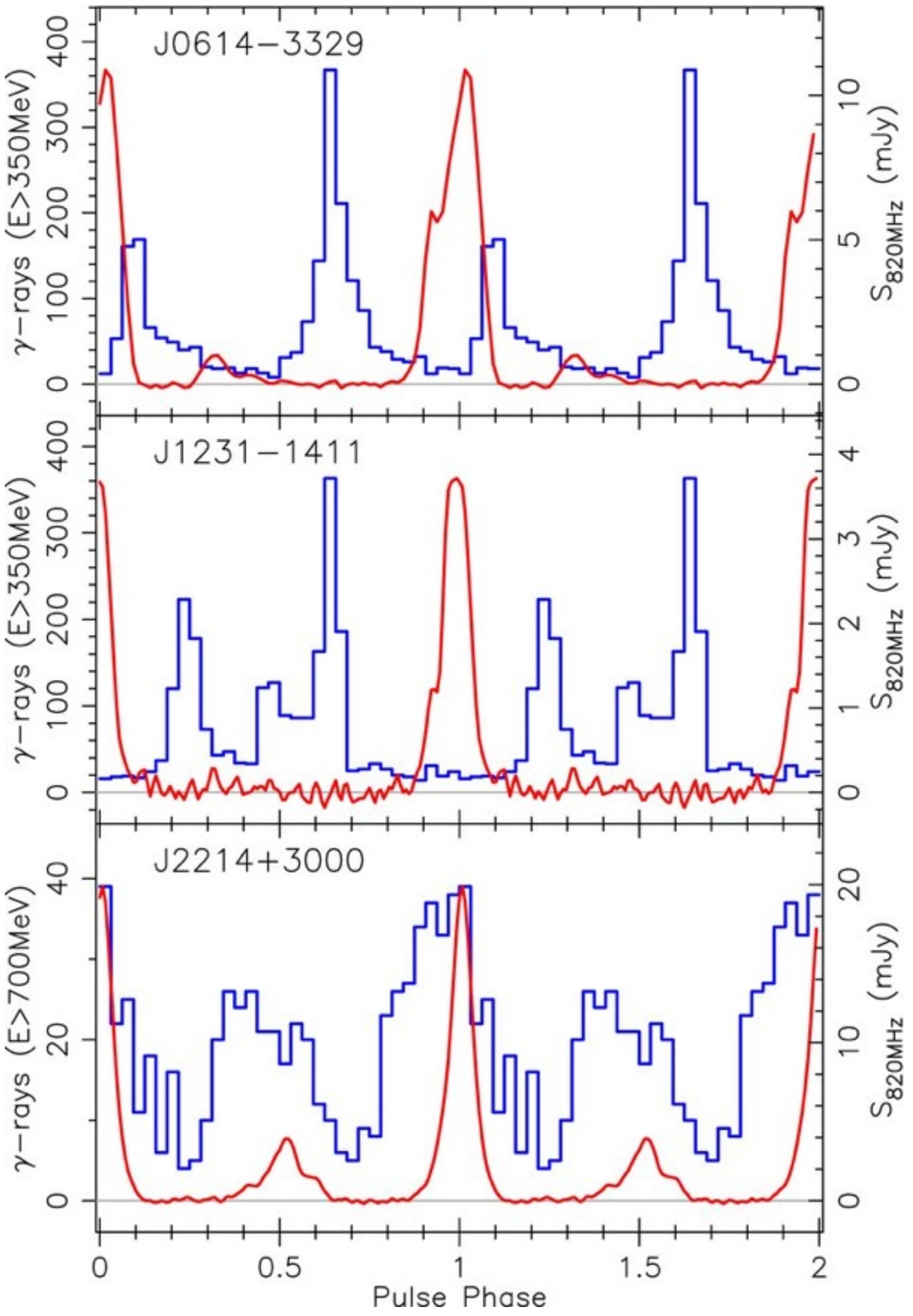


The Fermi LAT 60-month image, constructed from front-converting gamma rays with energies greater than 1 GeV. The most prominent feature is the bright band of diffuse glow along the map's center, which marks the central plane of our Milky Way galaxy.
(Credit: NASA/DOE/Fermi LAT Collaboration)



Gamma-ray (blue) and radio (red) light curves of three millisecond pulsars discovered by radio follow-up in Fermi unidentified sources.

(from https://fermi.gsfc.nasa.gov/science/eteu/pulsars/)



THE ASTROPHYSICAL JOURNAL, **272**:317–324, 1983 September 1

ANALYSIS METHODS FOR RESULTS IN GAMMA-RAY ASTRONOMY

TI-PEI LI AND YU-QIAN MA
High Energy Astrophysics Group, Institute of High Energy Physics,
Academia Sinica, Beijing, China
Received 1982 September 20; accepted 1983 February 7

ABSTRACT

The current procedures for analyzing results of γ-ray astronomy experiments are examined critically. We propose two formulae to estimate the significance of positive observations in searching γ-ray sources or lines. The correctness of the formulae are tested by Monte Carlo simulations.

Subject headings: gamma-rays: general — numerical methods

I. INTRODUCTION

Evaluation of the statistical reliability of positive results in searching discrete γ-ray sources or lines is an important problem in γ-ray astronomy. Since both the signal-to-background ratio and detector sensitivity are generally limited in this energy range, one must carefully analyze the observed data to determine the confidence level of a candidate source or line, that is, the probability that the count rate excess is due to a genuine source or line rather than to a spurious background fluctuation, even though all systematic effects are believed to have been removed.

Figure 1 shows a typical observation in γ-ray astronomy. A photon detector points in the direction of a suspected source for a certain time t_{on} and counts N_{on} photons, and then it turns for background measurement for a time interval t_{off} and counts N_{off} photons. The quantity α is the ratio of the on-source time to the off-source time, $\alpha = t_{on}/t_{off}$ (in some cases of searching for lines, N_{on} is the number of counts under a peak in an energy spectrum, and the peak is taken to be n_S channels wide; N_{off} is the number of counts in n_b channels adjacent to the peak; then $\alpha = n_S/n_b$).

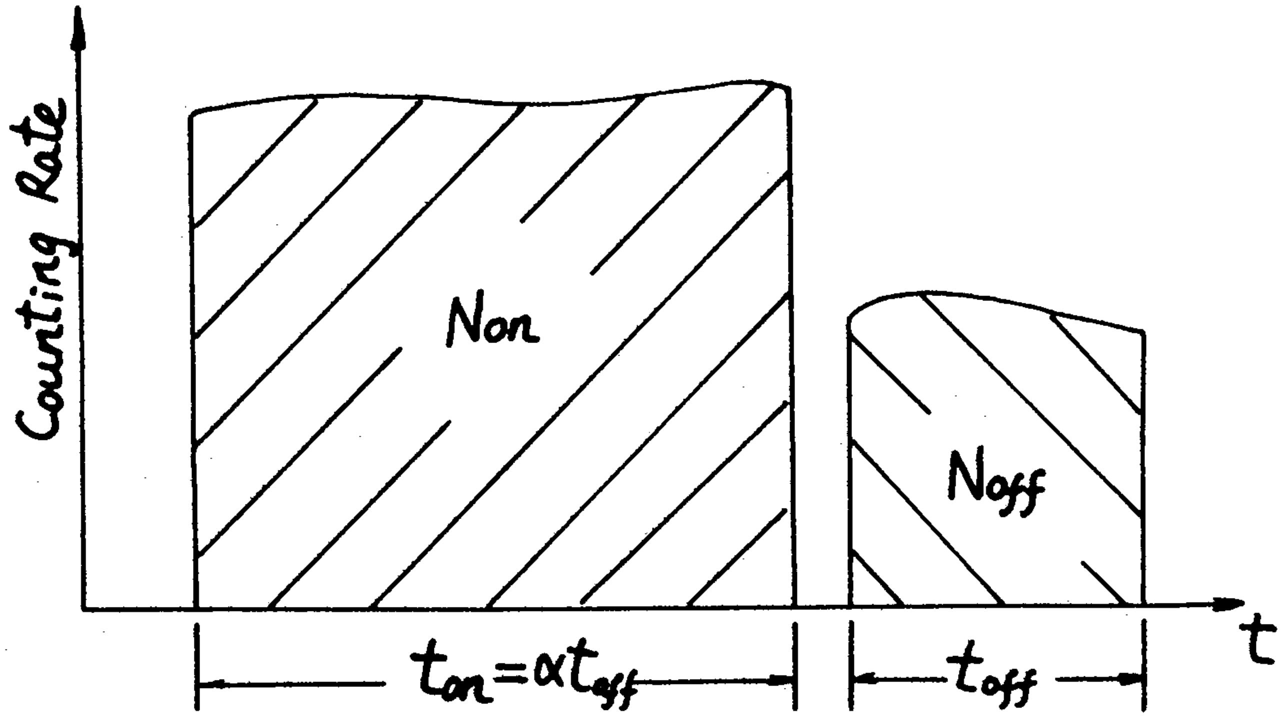


FIG. 1.—A typical observation in γ-ray astronomy

Simple estimate of signal strength and its statistical significance.

- estimate of background photons included in on-source counts $$\hat{N}_B = \alpha N_{\text{off}}$$
- estimate of observed signal $$\hat{N}_S = N_{\text{on}} - \hat{N}_B = N_{\text{on}} - \alpha N_{\text{off}}$$
- standard deviation of signal $$\sigma^2(\hat{N}_S) = \sigma^2(N_{\text{on}}) + \sigma^2(\hat{N}_B) = \sigma^2(N_{\text{on}}) + \sigma^2(\alpha N_{\text{off}}) = \sigma^2(N_{\text{on}}) + \alpha^2\sigma^2(N_{\text{off}})$$
- standard deviation estimate $$\hat{\sigma}_S = \sqrt{N_{\text{on}} + \alpha^2 N_{\text{off}}}$$
- statistical significance $$S = \frac{\hat{N}_S}{\hat{\sigma}_S} = \frac{N_{\text{on}} - \alpha N\text{off}}{\sqrt{N_{\text{on}} + \alpha^2 N_{\text{off}}}}$$

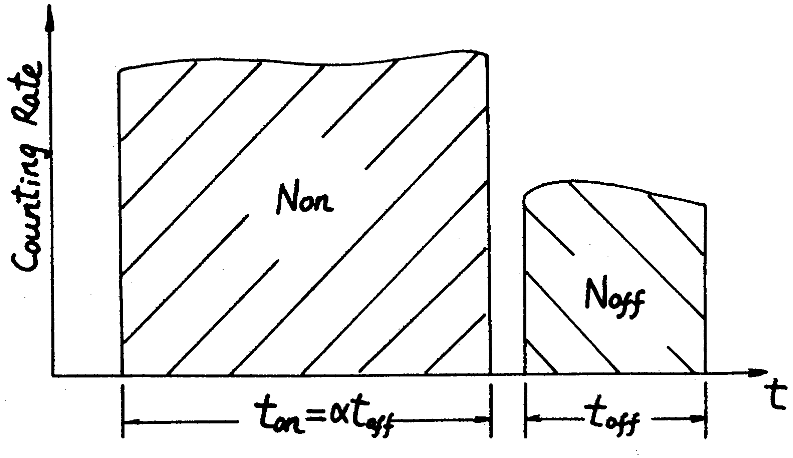


FIG. 1.—A typical observation in γ-ray astronomy

Estimate of result reliability and new estimated significance

Here we calculate the standard deviation under the assumption that there are only background photons.

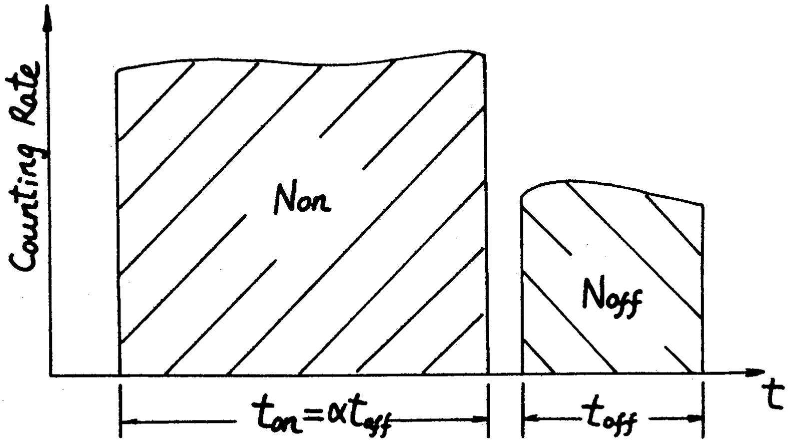


FIG. 1.—A typical observation in γ-ray astronomy

- estimate of photon arrival rate
$$\frac{N_{\text{on}} + N_{\text{off}}}{t_{\text{on}} + t_{\text{off}}}$$

- estimate of background on-source photons
$$\hat{N}_B = \frac{N_{\text{on}} + N_{\text{off}}}{t_{\text{on}} + t_{\text{off}}} t_{\text{on}} = \frac{\alpha}{\alpha + 1}(N_{\text{on}} + N_{\text{off}})$$

- estimate of background off-source photons
$$\frac{N_{\text{on}} + N_{\text{off}}}{t_{\text{on}} + t_{\text{off}}} t_{\text{off}} = \frac{\hat{N}_B}{\alpha}$$

- estimate of on-source standard deviation
$$\sigma^2(\hat{N}_S) = \sigma^2(N_{\text{on}}) + \alpha^2 \sigma^2(N_{\text{on}}) \approx \hat{N}_B + \alpha^2(\hat{N}_B/\alpha)$$
$$= (1 + \alpha)\hat{N}_B = \alpha(N_{\text{on}} + N_{\text{on}})$$

- estimated significance
$$S = \frac{\hat{N}_S}{\hat{\sigma}_S} = \frac{N_{\text{on}} - \alpha N_{\text{off}}}{(\sqrt{\alpha(N_{\text{on}} + N_{\text{off}})}}$$

Short recap of the Likelihood Ratio Method (Wilks' theorem) – 1

- Taylor expansion

$$\frac{\partial \ln L(D|\theta)}{\partial \theta} \approx - \left.\frac{\partial^2 \ln L(D|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} (\hat{\theta} - \theta) \approx -E\left[\left.\frac{\partial^2 \ln L(D|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}}\right] (\hat{\theta} - \theta)$$

- Integration

$$L(D|\theta) \propto \exp\left\{\frac{1}{2}E\left[\left.\frac{\partial^2 \ln L(D|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}}\right] (\hat{\theta} - \theta)^2\right\}$$

- Extension to more than one parameters (split into two subsets)

$$L(D|\boldsymbol{\theta}) = L(D|\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \propto \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T I (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]$$

where Fisher's information matrix is split into submatrices $I = \begin{pmatrix} I_{rr} & \vdots & I_{rs} \\ \cdots & & \cdots \\ I_{sr} & \vdots & I_{ss} \end{pmatrix}$

Short recap of the Likelihood Ratio Method (Wilks' theorem) – 2

- Then, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_r \\ \boldsymbol{\theta}_s \end{pmatrix}$ and therefore

$$L(D|\boldsymbol{\theta}_r, \boldsymbol{\theta}_s) \propto \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)^T I_{rr}(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r) - (\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)^T I_{rs}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s) - \frac{1}{2}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)^T I_{ss}(\hat{\boldsymbol{\theta}}_s - \boldsymbol{\theta}_s)\right]$$

- We know that asymptotically, the estimator $\hat{\boldsymbol{\theta}}$ has a Gaussian distribution with covariance matrix I^{-1}, therefore, asymptotically, the likelihood approaches the distribution of the estimator.

- When we maximize the likelihood with respect to the whole parameter vector, we find that the estimators for the subvectors are

$$\theta_r' = \hat{\boldsymbol{\theta}}_r; \quad \theta_s' = \hat{\boldsymbol{\theta}}_s$$

  and the corresponding maximum likelihood has a fixed value that depends only on data.

- When we maximize the likelihood with respect to the s parameters only, we find $\theta_s'' = \hat{\boldsymbol{\theta}}_s$ and

$$L(D|\boldsymbol{\theta}_r, \boldsymbol{\theta}_s'') \propto \exp\left[-\frac{1}{2}(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)^T I_{rr}(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)\right]$$

Short recap of the Likelihood Ratio Method (Wilks' theorem) – 3

- This means that when we define the likelihood ratio $\lambda = \dfrac{L(D|\boldsymbol{\theta}_r, \boldsymbol{\theta}_s'')}{L(D|\boldsymbol{\theta}_r', \boldsymbol{\theta}_s')}$, and recall that the estimators are

  asymptotically Gaussian, we find that

$$-2 \ln \lambda = (\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)^T I_{rr} (\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta}_r)$$

  has a chi-square distribution with *r* degrees of freedom (Wilks' theorem).

Application of the Likelihood Ratio Method to estimating N_S and N_B

- The problem at hand is defined by

data: $(N_{\rm on}, N_{\rm off})$

unknown parameters: $\boldsymbol{\theta} = (\langle N_B \rangle, \langle N_S \rangle)$

null hypothesis: $\langle N_S \rangle = 0$

alternative hypothesis: $\langle N_S \rangle \neq 0$

- maximum of a Poisson likelihood with just one count

$$L(N|\theta) = \frac{\theta^N}{N!} e^{-\theta} \quad \Rightarrow \quad \ln L(N|\theta) \sim N \ln\theta - \theta \quad \Rightarrow \quad \frac{\partial L}{\partial \theta} = \frac{N}{\theta} - 1 = 0 \quad \Rightarrow \quad \hat{\theta} = N$$

(the count is the MaxL estimate).
This means that the previous estimates ARE MaxL estimates, and we can use them to calculate the likelihood ratio.

Application of the Likelihood Ratio Method to estimating N_S and N_B (ctd.)

- **MaxL estimates**

alternative hypothesis: $$\langle \hat{N}_B \rangle = \alpha N_{\mathrm{off}}, \quad \langle \hat{N}_S \rangle = N_{\mathrm{on}} - \alpha N_{\mathrm{off}}$$

null hypothesis: $$\langle \hat{N}_B \rangle = \frac{\alpha}{\alpha+1}(N_{\mathrm{on}} + N_{\mathrm{off}}), \quad \langle \hat{N}_S \rangle = 0$$

- **Likelihoods**

alternative hypothesis: $$L(D|H_1)|_{\max} = \frac{N_{\mathrm{on}}^{N_{\mathrm{on}}}}{N_{\mathrm{on}}!} e^{-N_{\mathrm{on}}} \frac{N_{\mathrm{off}}^{N_{\mathrm{off}}}}{N_{\mathrm{off}}!} e^{-N_{\mathrm{off}}}$$

null hypothesis: $$L(D|H_0)|_{\max} = \frac{1}{N_{\mathrm{on}}!} \left(\frac{\alpha}{\alpha+1}(N_{\mathrm{on}} + N_{\mathrm{off}}) \right)^{N_{\mathrm{on}}} \exp\left(-\frac{\alpha}{\alpha+1}(N_{\mathrm{on}} + N_{\mathrm{off}}) \right)$$
$$\times \frac{1}{N_{\mathrm{off}}!} \left(\frac{1}{\alpha+1}(N_{\mathrm{on}} + N_{\mathrm{off}}) \right)^{N_{\mathrm{off}}} \exp\left(-\frac{1}{\alpha+1}(N_{\mathrm{on}} + N_{\mathrm{off}}) \right)$$

Application of the Likelihood Ratio Method to estimating N_S and N_B (ctd.)

- **MaxL ratio**

$$\lambda_{\max} = \frac{L(D|H_0)|_{\max}}{L(D|H_1)|_{\max}} = \left(\frac{\alpha}{\alpha+1}\frac{N_{\mathrm{on}}+N_{\mathrm{off}}}{N_{\mathrm{on}}}\right)^{N_{\mathrm{on}}}\left(\frac{1}{\alpha+1}\frac{N_{\mathrm{on}}+N_{\mathrm{off}}}{N_{\mathrm{off}}}\right)^{N_{\mathrm{off}}}$$

therefore the significance can be obtained from $-2\ln\lambda_{\max}$ because $-2\ln\lambda$ has a chi-square distribution with 1 degree of freedom.