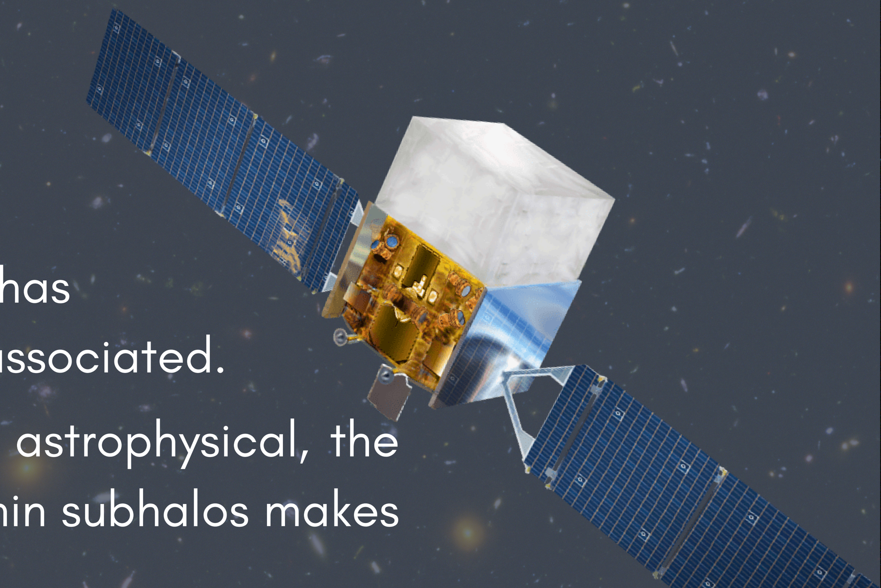
Search for dark matter subhalos among Fermi-LAT sources in presence of dataset shift

A. Amerio[a], D. Malyshev[b], B. Zaldivar[a], V. Gammaldi[c, d, e]
M.A. Sánchez-Conde[d, e]



Abstract

Over 16 years, the Fermi gamma-ray space telescope has identified over 7000 sources, with 2577 remaining unassociated.

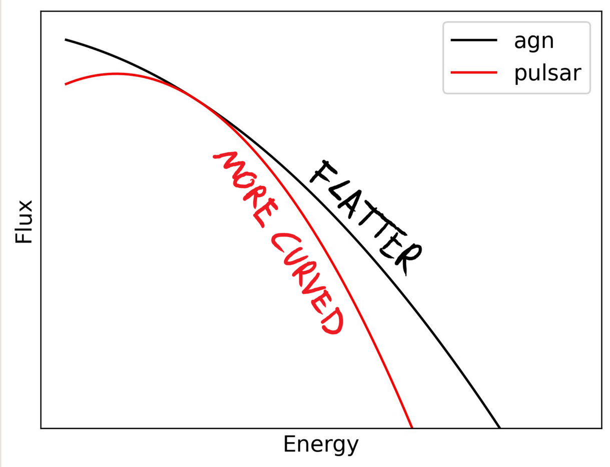
While many of these unassociated sources are likely astrophysical, the potential for signals from dark matter annihilation within subhalos makes the analysis of this population extremely compelling.

In this work, we introduce a novel generative statistical model to characterize the likelihood of unassociated gamma-ray sources, incorporating both astrophysical and dark matter contributions. Using this model, we conduct a statistical classification and establish new bounds on the dark matter annihilation cross-section, yielding competitive results.

Methodology

The origin of gamma-ray sources can be studied through several properties, including an analysis of the energy spectrum, variability index, or multi-wavelength approaches.
We focus on the **energy spectrum**, which we parametrize with the so-called log parabola fitting formula:

$$\Phi(E) = \Phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta \log(E/E_0)}$$

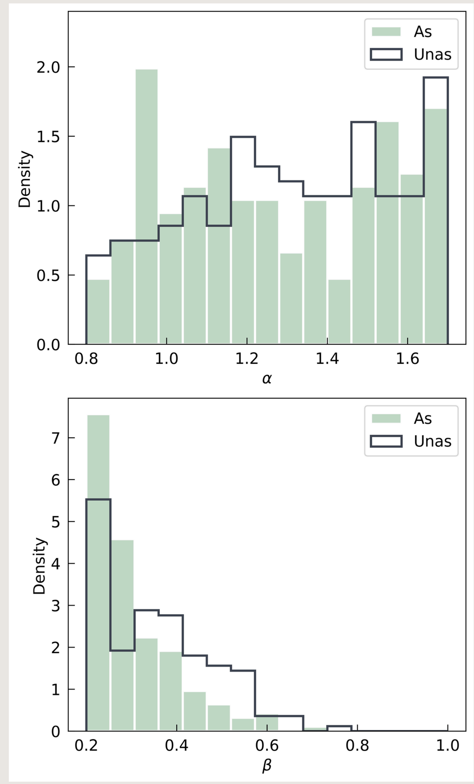


Different classes of sources have **different energy spectrum shapes**, as such we aim to **discriminate them by studying the joint distribution** of the α, β, and integrated flux parameters.
We divide sources into galactic, extragalactic, and DM sources.

Why are some sources still unassociated?
According to [1], several effects could come into play:

- **Lack of statistics** makes the reconstruction of the energy spectrum more uncertain.
- **Closeness to the galactic plane** introduces two effects:
  - Lower signal-to-noise ratio due to the high gamma-ray density (abundance of sources and diffuse galactic emission).
  - Absorption effects due to the galactic gas and dust will affect X-ray and optical measurements, respectively, hindering multi-wavelength association approaches.

The energy spectrum **distribution for unassociated sources is different!**



Astrophysical sources

In machine learning, the joint probability distribution function of observables and classes for the train and test datasets is generally assumed to be the same. This is no longer true when we consider unassociated gamma-ray sources:

$$p_{as}(\mathbf{x}, C_k) \neq p_{unas}(\mathbf{x}, C_k) \qquad p(\mathbf{x}, C_k) = p(\mathbf{x}|C_k)p(C_k) = p(C_k|\mathbf{x})p(\mathbf{x})$$

Where **x** is constituted by the **integral flux**, **α** and **β**, and C_k represents the specific class of an observation. We model this difference according to [2] introducing a prior shift and a covariate shift:

$p_{as}(\mathbf{x}|C_k) = p_{unas}(\mathbf{x}|C_k)$, $p_{as}(C_k) \neq p_{unas}(C_k)$ **prior shift**: due to class imbalance / dark matter

$p_{as}(C_k|\mathbf{x}) = p_{unas}(C_k|\mathbf{x})$, $p_{as}(\mathbf{x}) \neq p_{unas}(\mathbf{x})$ **covariate shift**: due to association bias

We encode these two effects by defining a **mixture model**. The prior weights describe the prior shift, while the covariate shift is modelled using a sigmoid function common to all astrophysical sources:

$$p(\mathbf{x}) = \sum_k \pi_k p_k(\mathbf{x}|\boldsymbol{\theta}_k) \qquad \sigma(x, b, c) = \frac{1}{1 + e^{-b(x-c)}}$$

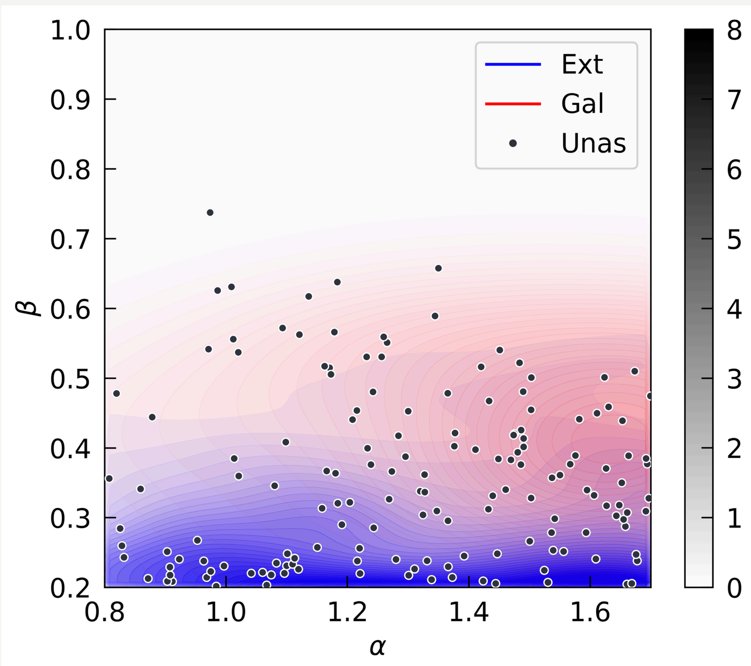
k = Gal, Ext, (DM)

$$p_{unas}(\mathbf{x}) \sim p_{as}(\mathbf{x})\sigma(x, b, c)$$

$$p_{as}(x) \equiv p_{as}^{KDE}(\mathbf{x})$$

The model parameters are optimized for the unassociated sources of the 4FGL-DR4 catalog.
Once trained, the model can be employed for **source classification** evaluating the **posterior probability**:

$$p_i(y = C_k|\mathbf{x}_i) = \frac{\pi_k p(\mathbf{x}_i|\boldsymbol{\theta}_k)}{\sum_j \pi_j p(\mathbf{x}_i|\boldsymbol{\theta}_j)}$$



Contour plot of the individual mixture distributions, MLE for unassociated sources with astrophysical components only. The distributions are marginalized over the flux variable for visualization purposes. Each component is weighted by its prior weight π_k.

Dark matter

We model the dark matter (DM) component of the mixture distribution as follows:

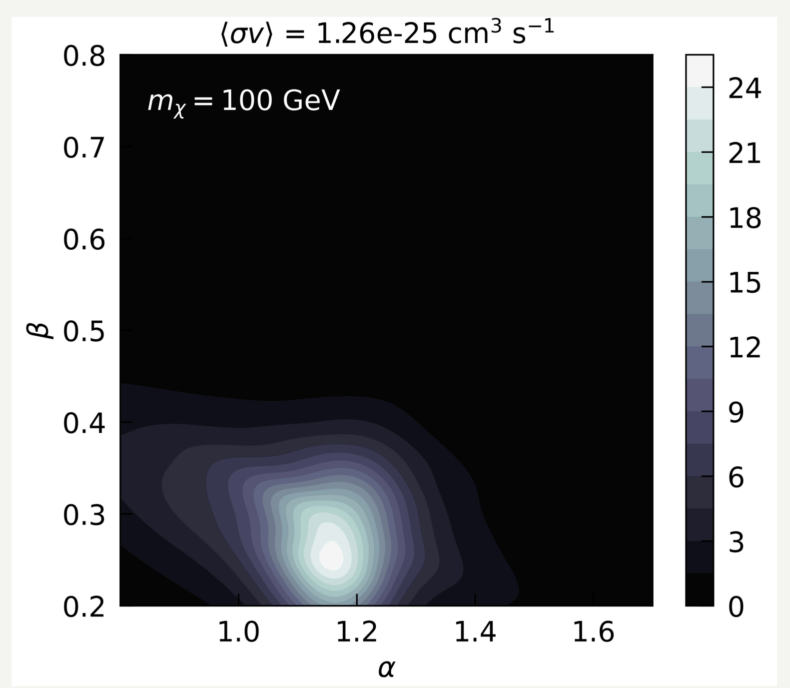
$$p_{DM}(\Phi, \alpha, \beta|\langle\sigma v\rangle, M_{DM}) = p_{DM}(\alpha, \beta|\Phi, M_{DM}) \frac{n(\Phi|\langle\sigma v\rangle, M_{DM})\epsilon(\Phi)}{N_{DM}}$$

$$\Phi = \frac{J\langle\sigma v\rangle}{4\pi M_{DM}^2} \int_{E_1}^{E_2} \frac{dN_\gamma}{dE} dE \qquad p_{eff}(\Phi|\langle\sigma v\rangle, M_{DM})$$

- **DM distribution** for α and β, conditioned on the integrated flux from 100MeV to 1 TeV. We estimate it performing photon-count simulations of the theoretical DM energy spectrum given by [3]. For a fixed flux, we simulate approximately 20.000 sources and for each one we fit the energy spectrum to a log parabola, deriving the corresponding distribution.
- **Detection efficiency**, determined by computing the TS of each source. It is the fraction of sources with TS ≥ 25.
- Dark matter subhalos **J factor number density** [4].
- The normalization constant represents the number of DM subhalos potentially observable by Fermi, given by integrating the number density times the Fermi detection efficiency.

The amount of DM present among the unassociated sources determines the **prior DM weight** (fixed given a DM mass, annihilation channel and ⟨σv⟩)

$$\pi_{DM} = \frac{N_{DM}(M_{DM}, \langle\sigma v\rangle)}{N_{tot}}$$

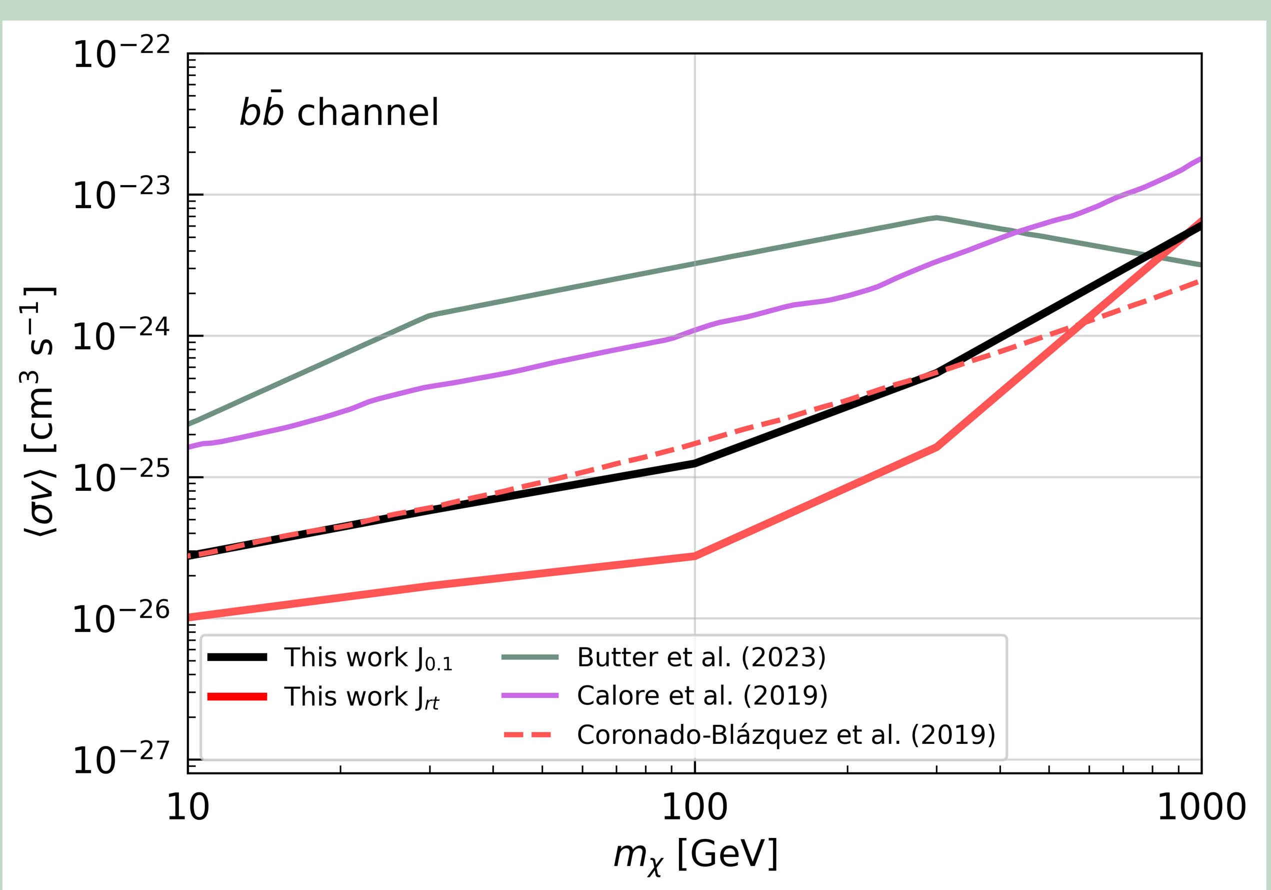


Dark matter distribution computed for DM mass = 100 GeV, $b\bar{b}$ annihilation channel, and annihilation cross section close to the value of our bound.

Bounds

We derive bounds on the thermally averaged annihilation cross-section of dark matter using Wilks' theorem and adopting a 95% confidence level.

$$\mathcal{L} = \prod_{i \in data} \sum_k \pi_k p_k(\mathbf{x}_i|\boldsymbol{\theta}_k) \qquad TS = -2\log\left(\frac{\mathcal{L}(\langle\sigma v\rangle)}{\mathcal{L}_{max}}\right) \leq TS_{crit} \qquad \mathcal{L}_{max} = \text{argmax}_{\langle\sigma v\rangle}[\mathcal{L}(\langle\sigma v\rangle)]$$



Plot with our **bounds** for the thermally averaged annihilation cross-section of dark matter for the $b\bar{b}$ channel.
In **black**, are our main results. In **solid red**, we represent our results with the same J factor employed by Coronado-Blázquez et al., for comparison.

Conclusions

- For the first time, we propose a **generative model** describing the **likelihood of unassociated gamma-ray sources** including both astrophysical and dark matter components.
- This model can be employed for the **probabilistic classification of unassociated sources** and as a means to derive **bounds on the annihilation cross-section of dark matter**.
- **We open new ventures for the search of dark matter subhalos** among gamma-ray sources with statistical techniques. We derive **bounds** on the dark matter annihilation cross-section that are **competitive with other analyses** using Fermi-LAT data to search for subhalos.

References & Affiliations

[1] Abdollahi et al., 2020. Fermi Large Area Telescope Fourth Source Catalog, ApJS, 247(1), 33.
[2] Moreno-Torres et al., Pattern Recognition, 45(1), 2012, pp. 521-530.
[3] Arina et al., arXiv:2312.01153
[4] Aguirre-Santaella & Sánchez-Conde, MNRAS, 530(3), 2024, pp. 2496-2511.

[a] Instituto de Física Corpuscular (IFIC), Universitat de València and CSIC, Calle Catedrático José Beltrán 2, 46980 Paterna, Spain
[b] Erlangen Centre for Astroparticle Physics, Nikolaus-Fiebiger-Str. 2, Erlangen 91058, Germany
[c] Department of Information Technology, Escuela Politécnica Superior, Universidad San Pablo-CEU, CEU Universities, Campus Montepríncipe, Boadilla del Monte, Madrid 28668, Spain
[d] Departamento de Física Teórica, Facultad de Ciencias, Mod. 15, Universidad Autónoma de Madrid, E-28049 Madrid, Spain
[e] Instituto de Física Teórica, UAM-CSIC, Calle Nicolás Cabrera 13-15, Campus de Cantoblanco, E-28049 Madrid, Spain

IFIC Institut de Física Corpuscular · CSIC Consejo Superior de Investigaciones Científicas · Vniversitat de València · Financiado por la Unión Europea NextGenerationEU · Gobierno de España, Ministerio de Ciencia e Innovación · Generalitat Valenciana, Conselleria d'Educació, Investigació, Cultura i Esport · Plan de Recuperación, Transformación y Resiliencia · GVA NEXT Fons Next Generation a la Comunitat Valenciana · ASFAE