X17 Collab
07/25/2025

Progress Report on X17 Bumphunt Infrastructure

Emrys Peets
Joseph Bailey

Stanford University
SLAC National Accelerator Laboratory

U.S. DEPARTMENT OF ENERGY
Office of Science

Outline

- Introductions
- Personal Project looking at X17 Parameter Space and leptonic coupling
- Background in Bumphunting
- X17 Background Generation
- pyBumpHunter
- Next Steps

https://www.overleaf.com/read/zpzknkkkqrmk#bef878

Internal Note: X17 Collaboration Bumphunt Infrastructure and Methodology

Emrys Peets*[1,2], Joseph Bailey[1]

Quick Introduction

Emrys Peets

- Rising 6th year Stanford Physics PhD Candidate
- Advisors: Tim Nelson (SLAC), Philip Schuster (SLAC, Stanford)

Joseph Bailey

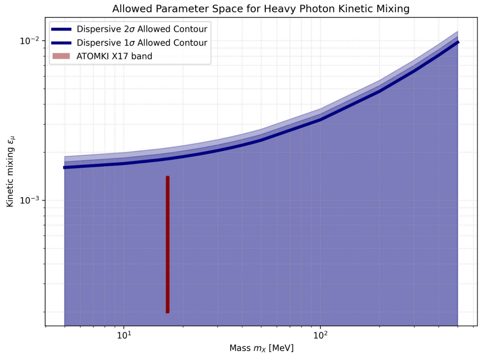
- Rising 3rd year Stanford Undergraduate
- Experience with Python, ROOT, fast simulation

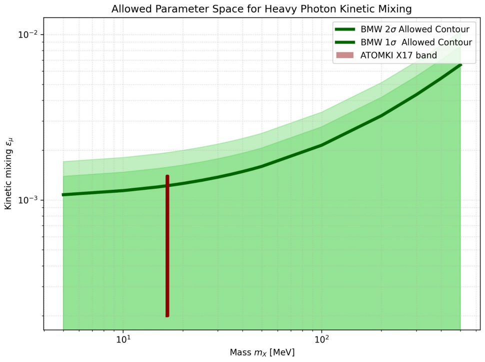
Personal Interest

SLAC

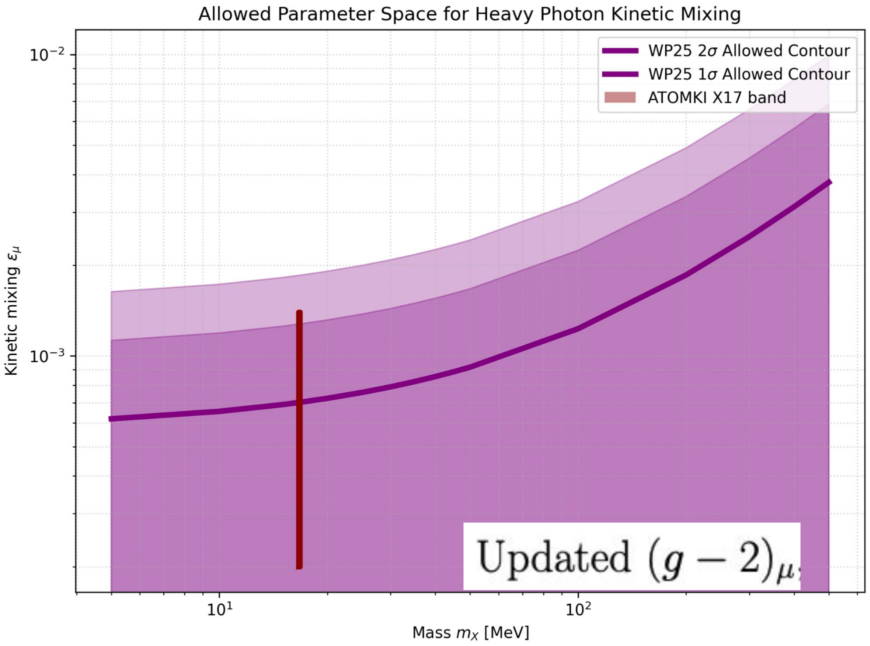
Updated $(g-2)_\mu$, $(g-2)_e$ and PADME-Favored Couplings Narrowly Compatible with the Preferred Region of ATOMKI X17, Given a Protophobic Interpretation

Emrys Peets*[1,2]





Previous Theoretical Models



Full Allowable Leptonic Coupling

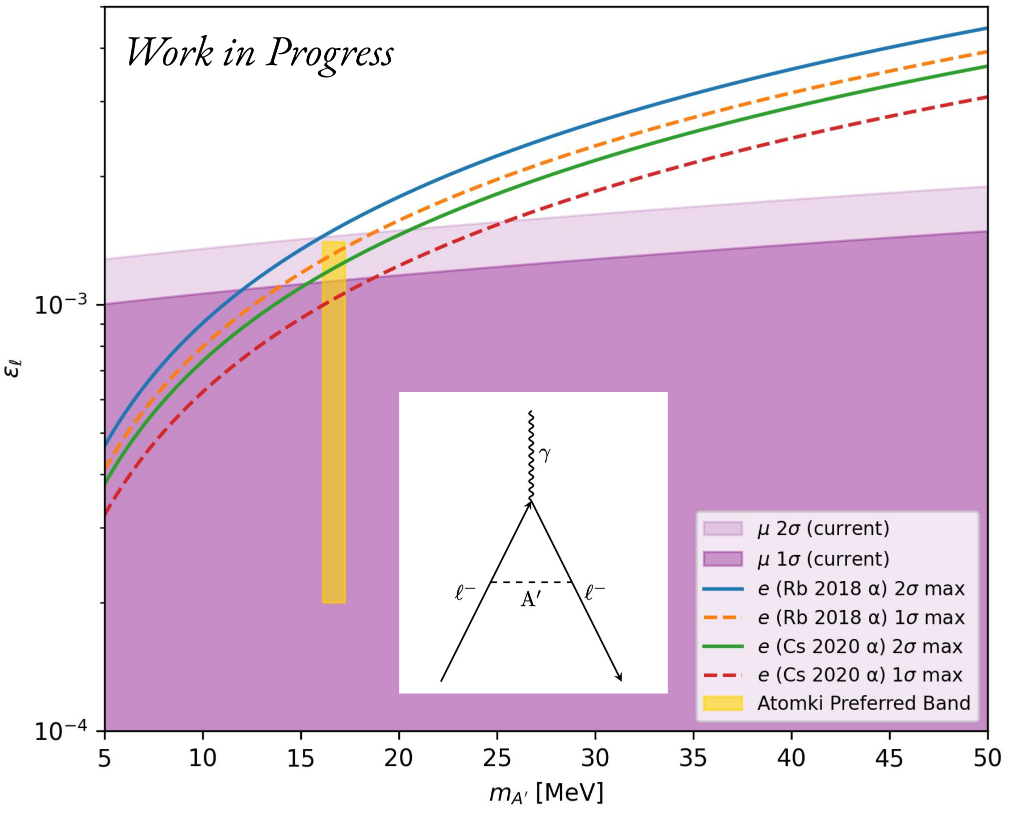
- Narrow overlap with favored electron coupling from most recent fine structure measurements
- NA48/2 necessitates protophobic coupling

NA48/2 Decay Chain

$$\pi^0 \to \gamma A' \to \gamma e^+ e^-$$

$$4.2 \times 10^{-4} \lesssim \varepsilon \lesssim 5.6 \times 10^{-4}$$

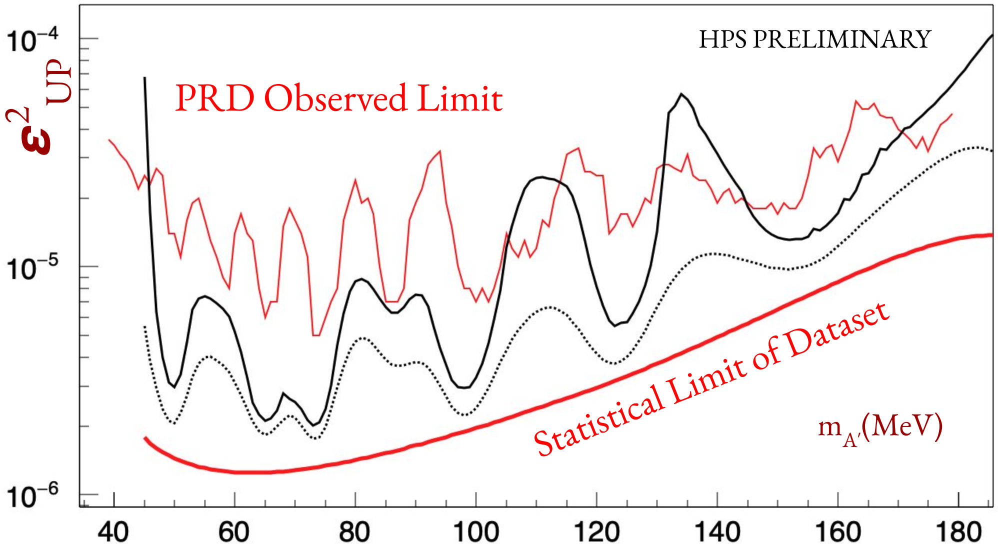
NA64 Padme



Bumphunting Experience

- Primary Prompt A` Resonance Search Analyst for Heavy Photon Search Collaboration
- Performing Global Functional Form Fitting Technique as Core Methodology
- Mentoring one Stanford Undergrad, and one UCLA undergrad on Gaussian process regression bumphunt techniques

Observed ε^2 Upper Limits ***Proof of Principle***



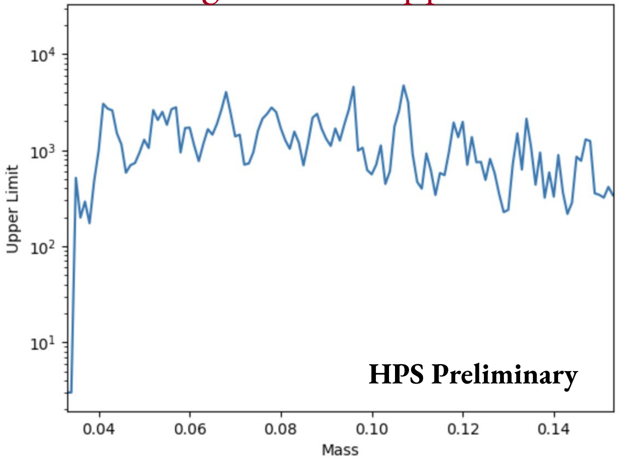
Application of GPR to HPS Datasets

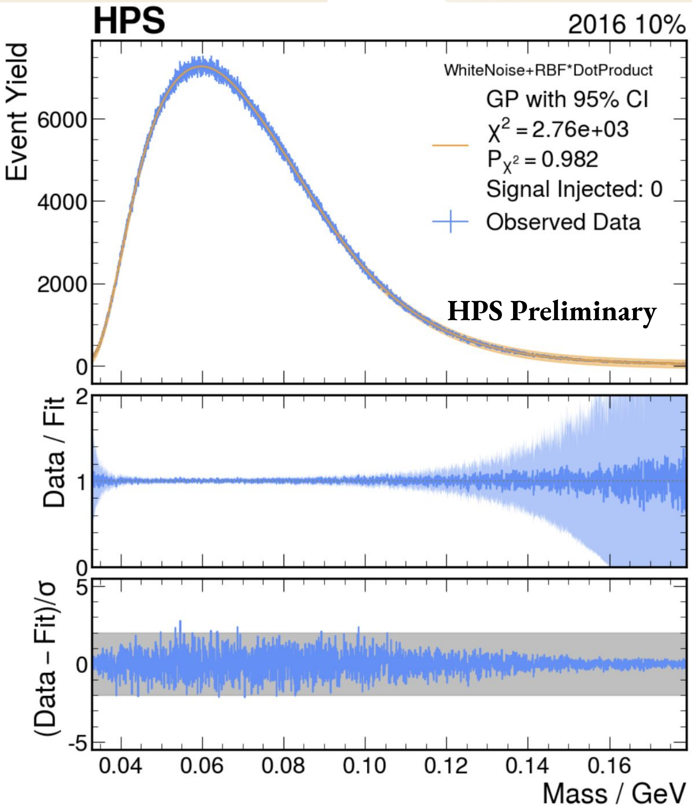
The GP model provides a strong fit to the datasets with well-defined uncertainty estimates.

Preliminary Upper Limits determined to be competitive with functional form fitting.

Kernel Choices:
WhiteNoise - models broad noise
RBF Kernel - models local correlations

10% Signal Yield Upper Limit





from Aidan Hsu's work

Progress on X17 Collaboration Bumphunt Infrastructure

Internal Note Documenting Progress:

- https://www.overleaf.com/read/zpzknkkqrmk#bef878

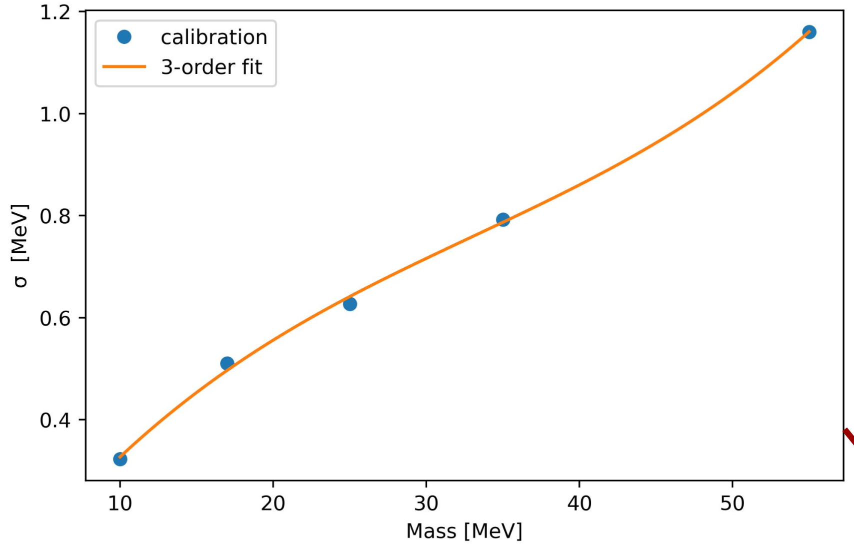
Generated Toy Distribution with Signal Injected at 17 MeV and 40 MeV

- Gaussian Signal Injection [exaggerated signals for illustrative purposes]
- Poissonian Sampling (stat. variance of square root of predicted value)
- added bin-by-bin jitter (gaussian of 10% predicted value)
    - will remove this for next stage

pyBumpHunter Software Package

- proof of principal results gotten, will refine
- promising methodology!

Base Mass Resolution



Can fit using different shape as necessary

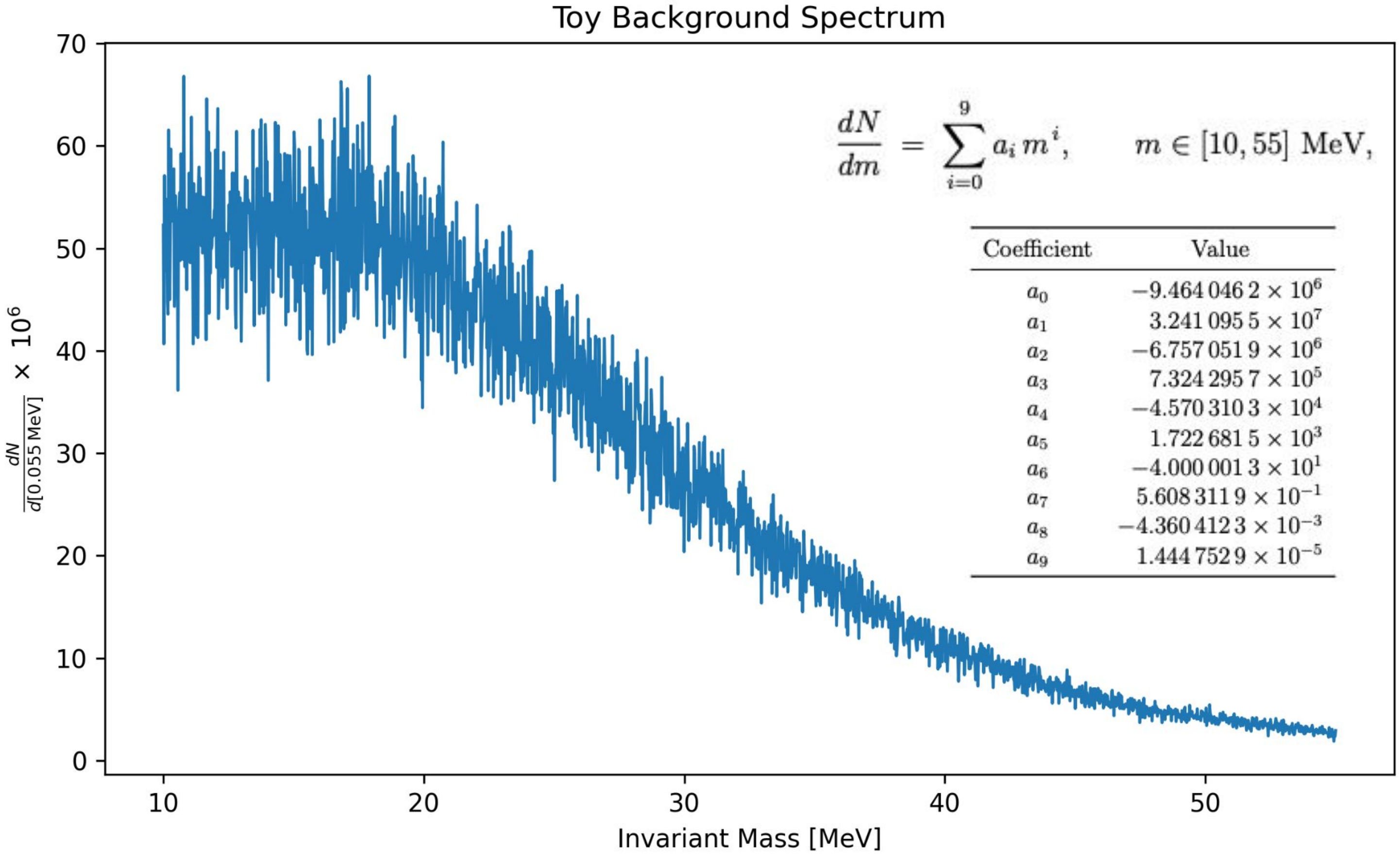
Initial Calibration Values (From Rafo)

Invariant mass m_i [MeV]	1σ mass resolution σ_i [MeV]
10	0.3225
17	0.5100
25	0.6270
35	0.7925
55	1.1600

Assuming Natural Width of X17 << Mass Resolution

$$S(m;\, m_0, \sigma) = \frac{N_{\mathrm{sig}}}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{m-m_0}{\sigma}\right)^2\right\}$$

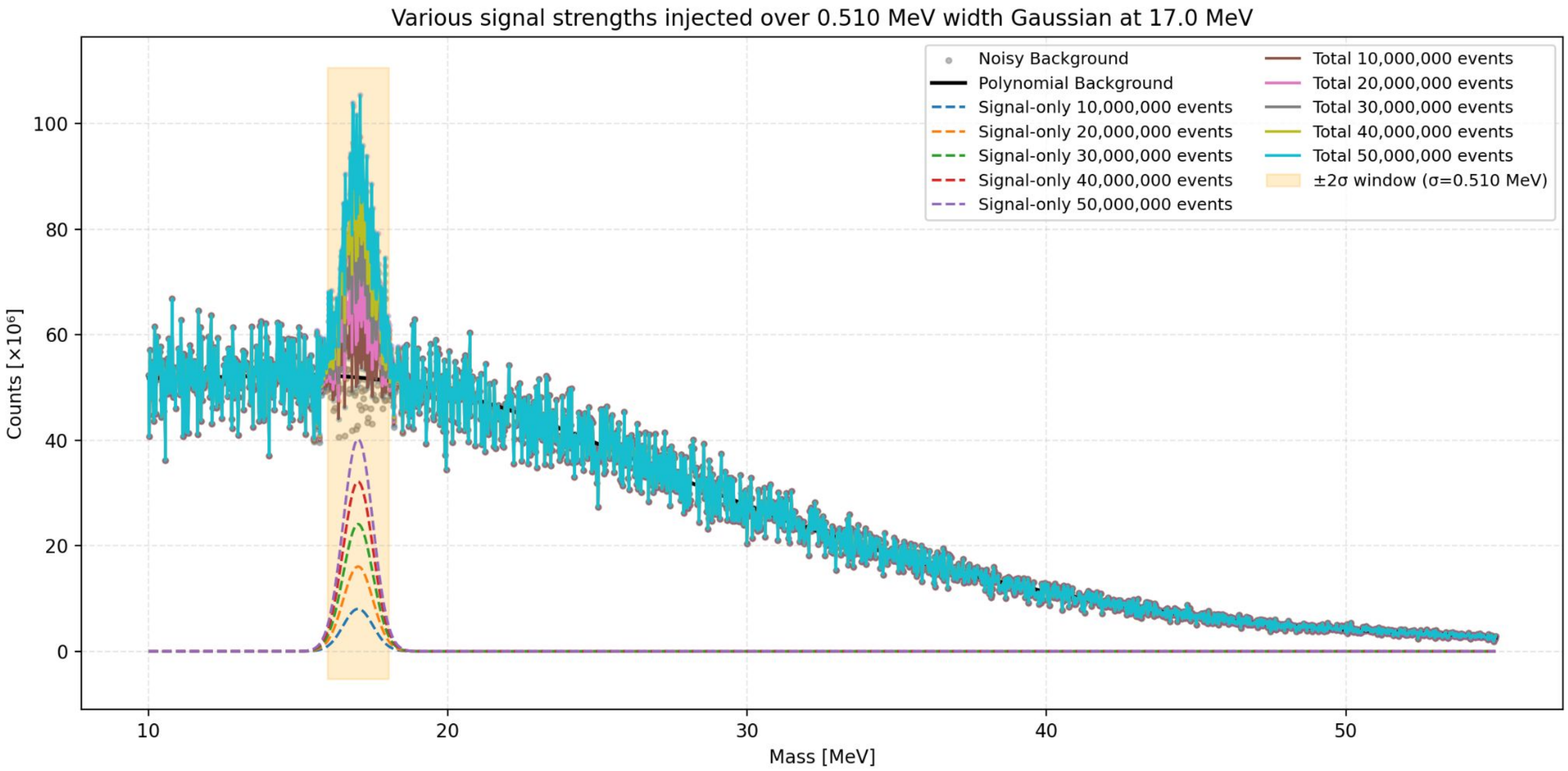
Base Background Distribution



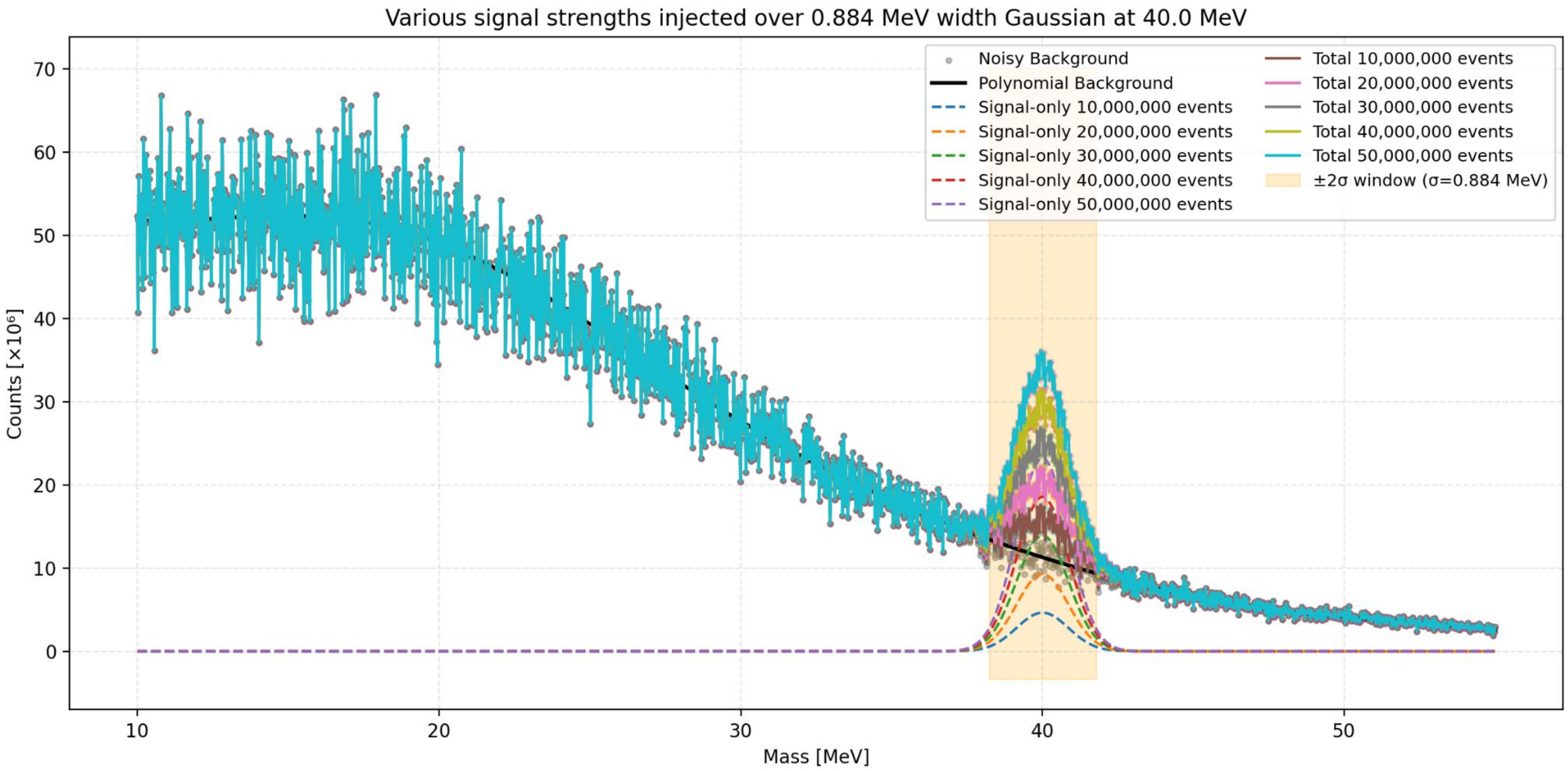
$$\frac{dN}{dm} = \sum_{i=0}^{9} a_i\, m^i, \qquad m \in [10, 55]\ \mathrm{MeV},$$

Coefficient	Value
a_0	$-9.464\,046\,2 \times 10^{6}$
a_1	$3.241\,095\,5 \times 10^{7}$
a_2	$-6.757\,051\,9 \times 10^{6}$
a_3	$7.324\,295\,7 \times 10^{5}$
a_4	$-4.570\,310\,3 \times 10^{4}$
a_5	$1.722\,681\,5 \times 10^{3}$
a_6	$-4.000\,001\,3 \times 10^{1}$
a_7	$5.608\,311\,9 \times 10^{-1}$
a_8	$-4.360\,412\,3 \times 10^{-3}$
a_9	$1.444\,752\,9 \times 10^{-5}$

17 MeV Toy Distributions



40 MeV Toy Distributions



Capabilities of pyBumpHunter

- pyBumpHunter is a python implementation of the BumpHunter algorithm described in arXiv:1101.0390, G. Choudalakis
- Accounts for the "look-elsewhere effect" by using the BumpHunter test statistic

$$t = -\ln p_{\min},$$

  and comparing this with generated background-only pseudo-experiments
- Can also perform signal injection tests
  - iteratively determine sensitivity given a background distribution

Changes to pyBumpHunter

The GitHub release contains some bugs that prevent pyBumpHunter from running properly

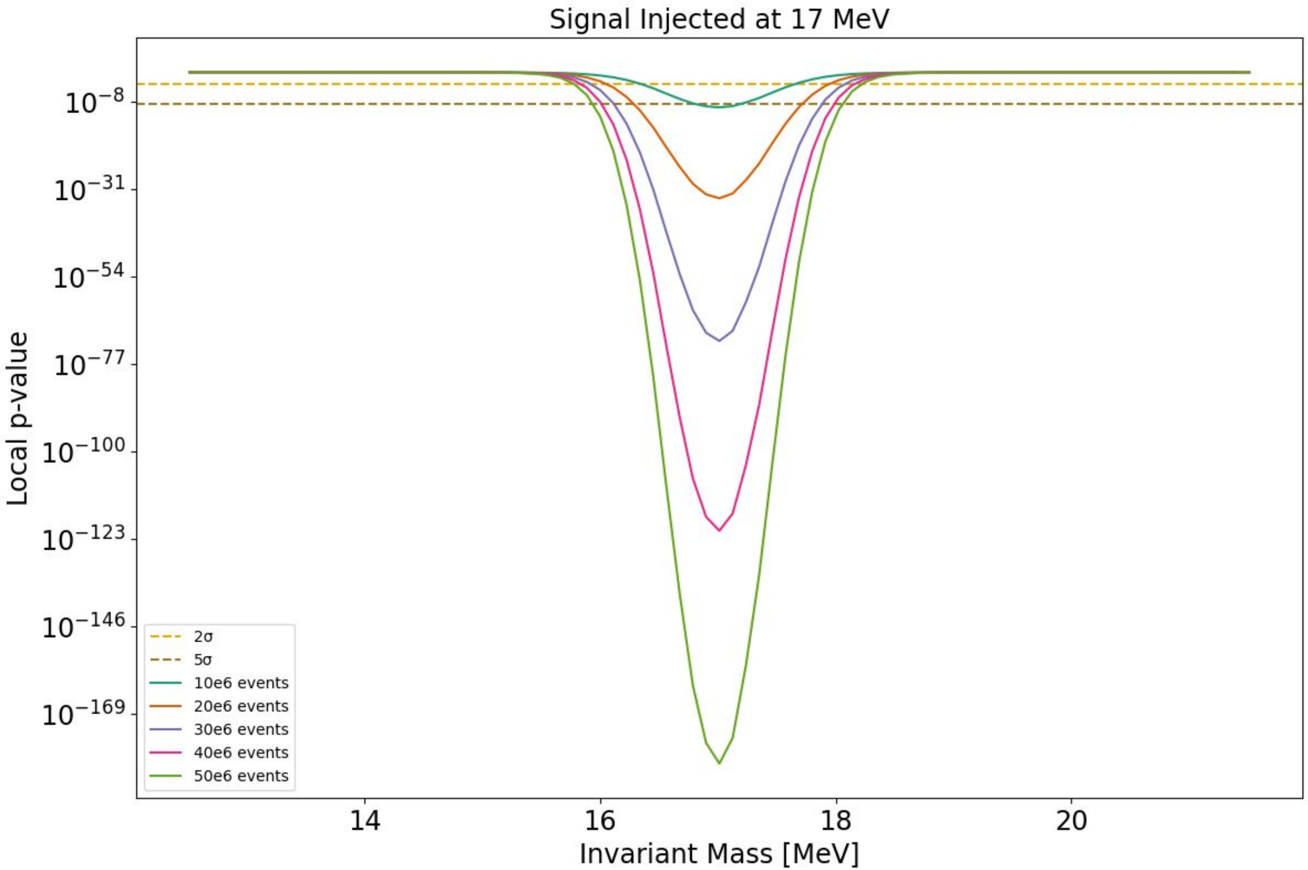
In pyBumpHunter/bumphunter_1dim.py:

- lines 1511, 1983 uses deprecated numpy behavior
- line 1891 needs to copy the looping behavior at line 1320

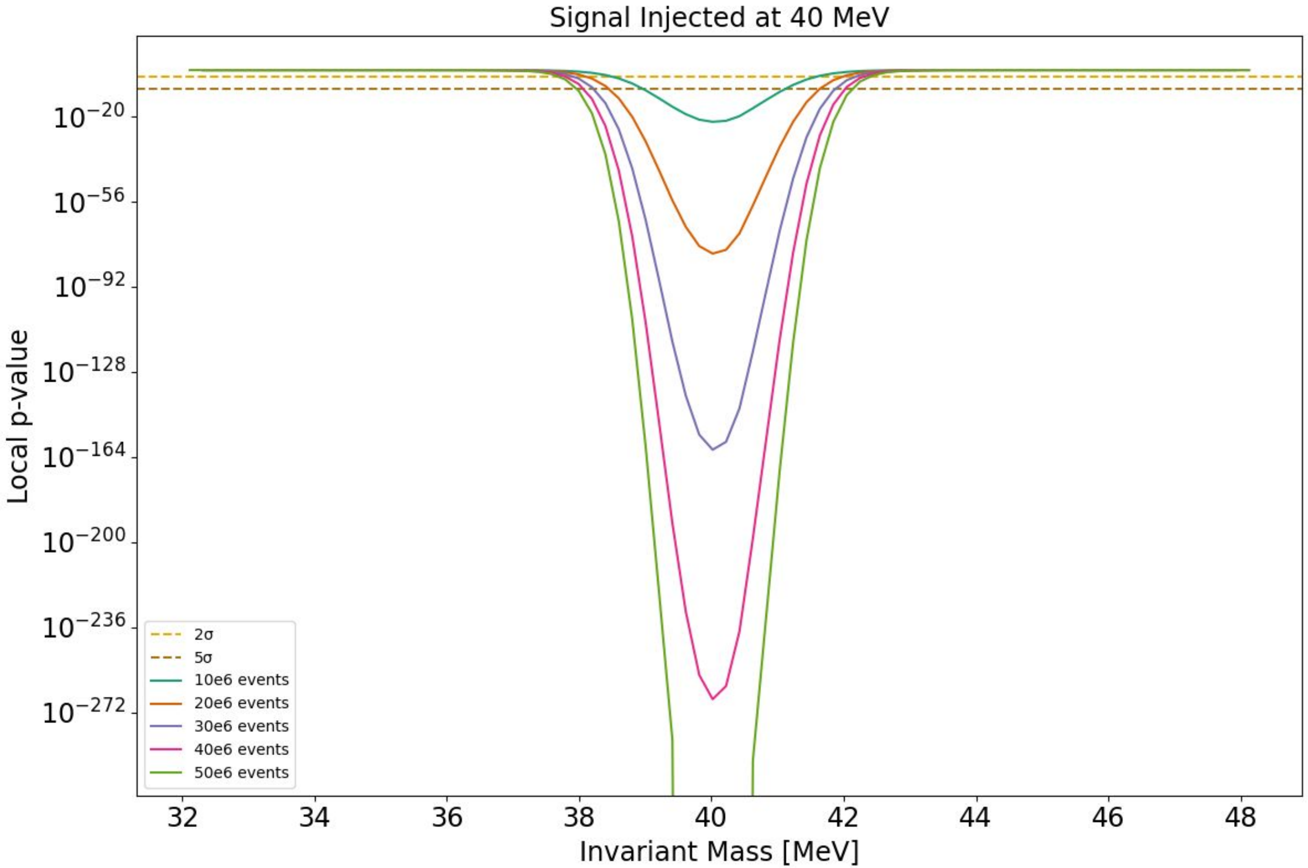
pyBumpHunter Setup and Examples

- We run pyBumpHunter on the toy distributions with signal injected at 17 and 40 MeV.
- The window size is twice the mass resolution and the step size is one fourth the mass resolution.
- The bump hunt finds local p-values for each window, then generates 10,000 pseudo-distributions and compare their BumpHunter statistics with the observed data.

pyBumpHunter Significances [17 MeV]



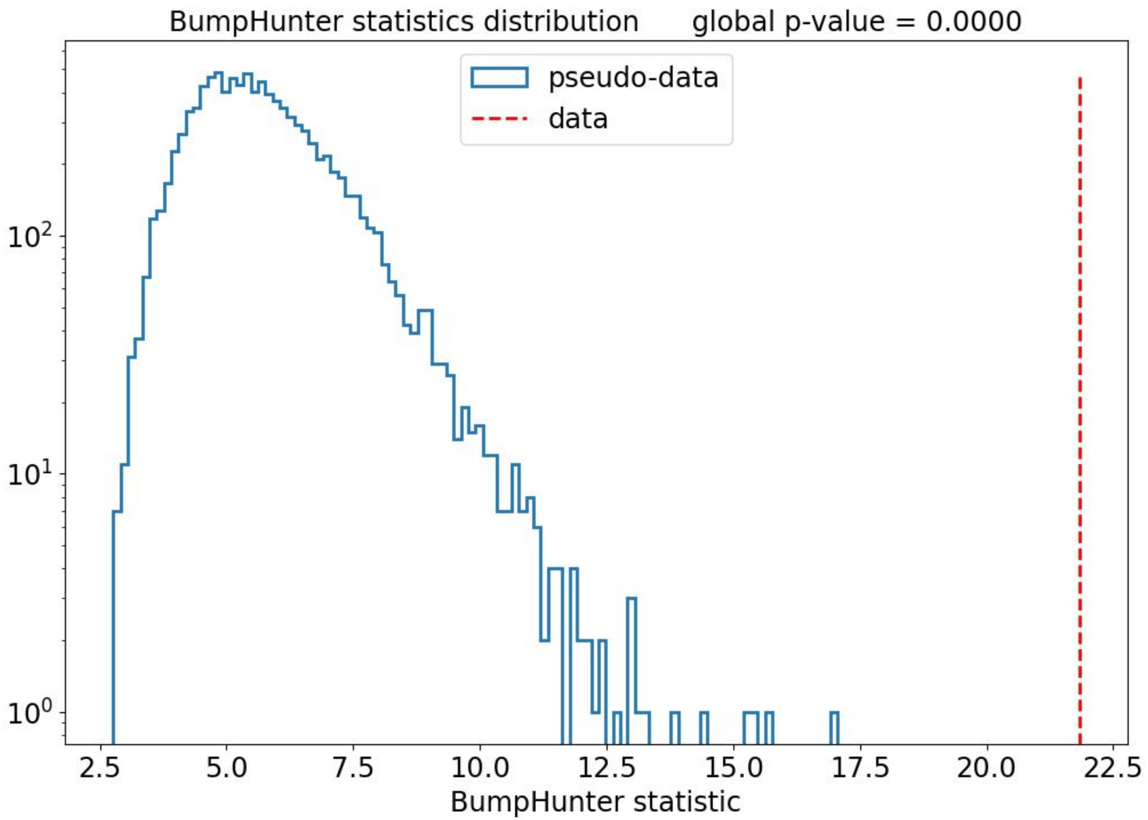
pyBumpHunter Significances [40 MeV]



Injected Signal Statitics

- The BumpHunter statistics for 10 million events injected at 17 MeV.

- 0 / 10,000 pseudo-data distributions have statistics larger than the data, so we record a global p-value of 0.



Next Steps

- Work Through Upper Limit Scans for:
  - Signal Yield
  - Coupling

$$\mathrm{CL}_s(\mu) = \frac{p_\mu}{1 - p_b}$$

$$\mathrm{CL}_s(N_{sig}^{up}) = 0.05$$

$$\boxed{\epsilon^2 = \frac{2\alpha N_{\mathrm{sig}}^{\mathrm{up}}}{3\pi m_{A'} f_{\mathrm{rad}} \frac{\mathrm{d}N_{\mathrm{bkg}}}{\mathrm{d}m}}}$$

- Determine how much run time necessary to hit PADME Target
  - Ideally with estimates at 3, 4, 5 sigma
  - should come from upper limits and number of events

- Iron Out Classical Bumphunting Infrastructure
  - need to create a version where background model is **not known a priori**
    - Data can have unexpected shape!
  - Perhaps merge pyBumpHunter with GP methodologies

Thoughts on Future

With real data, recommend implementation of gaussian process regression techniques.

- Contacted BaBar, APEX, Belle-2 for their invariant mass histograms for global fits in these regions all using a fundamentally similar fit model to what will be used on HPS 2015, 2016, 2019, 2021 datasets

- ^Postdoc proposal, but will pursue over the next several months regardless

Questions / Thoughts / Suggestions?

Thank you for listening!

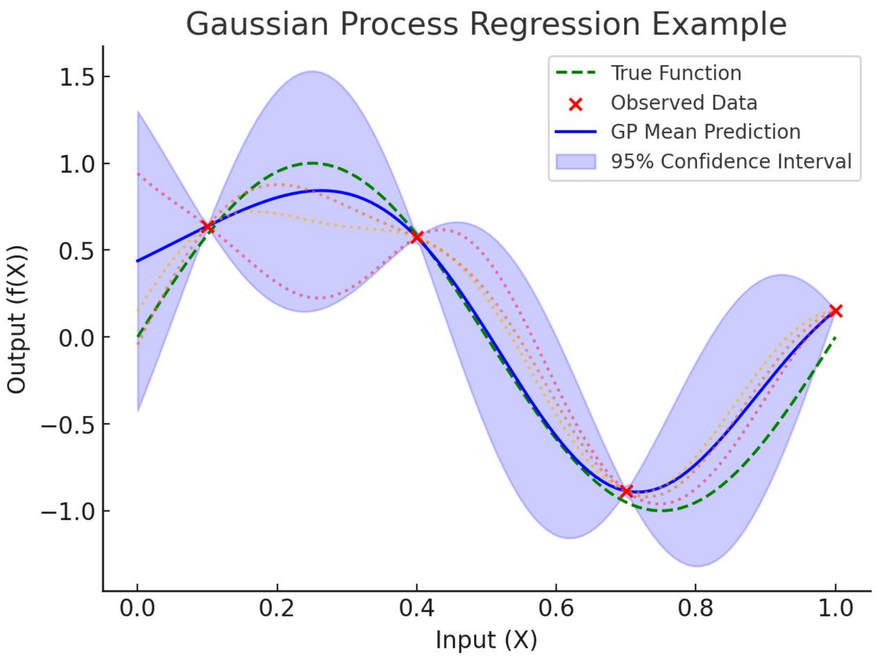
Extra Slide: Understanding Gaussian Process Regression

Understanding Gaussian Process Regression

Collaborative effort with Tom Eichlersmith (Minnesota, PhD), Aidan Hsu (Stanford Undergraduate), Takumi Britt (High School).

What is Gaussian Process Regression (GPR)?

- A **flexible, non-parametric Bayesian approach** that models distributions over functions.
- Unlike traditional regression, **GPR does not assume a fixed set of parameters**—it learns a distribution of possible functions.
- **Built-in uncertainty quantification** makes it ideal for noisy and complex datasets.



The kernel function (covariance function) governs how data points interact and influence one another.

The choice of kernel shapes the model's **smoothness, flexibility, and generalization ability**, making it crucial for capturing underlying data patterns.