HPS Collaboration Meeting

Updates on the HPS Prompt A` Resonance Search

Emrys Peets
Stanford University

U.S. DEPARTMENT OF ENERGY
Office of Science

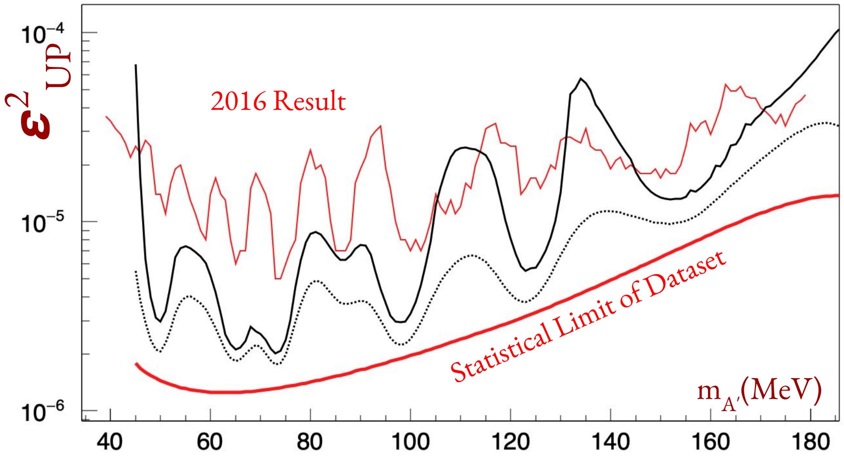
Stanford University

SLAC NATIONAL ACCELERATOR LABORATORY

Road Map

I. Intro
II. Previous Resonance Search Method/Result
III. New "Global Fitting" Analysis Approach
    A. Background Function Creation / Selection
IV. Blinded Procedure
    A. 6.5 % IMD Parameter Projection
    B. Challenges
V. Next Steps / Conclusion
VI. Leveraging our techniques
    A. Soon: 2015 IMD
    B. Later: APEX

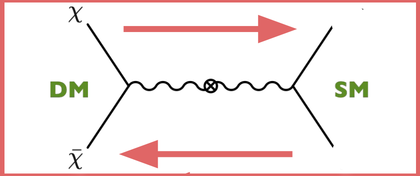
ε^2 Upper Limit Comparison



Freeze Out Thermal Relic Dark Matter Models

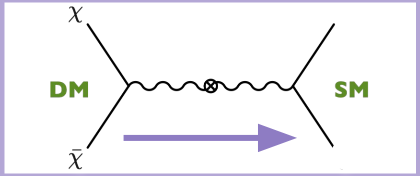
SLAC

Early Universe:
Thermal Equilibrium
Production = Annihilation



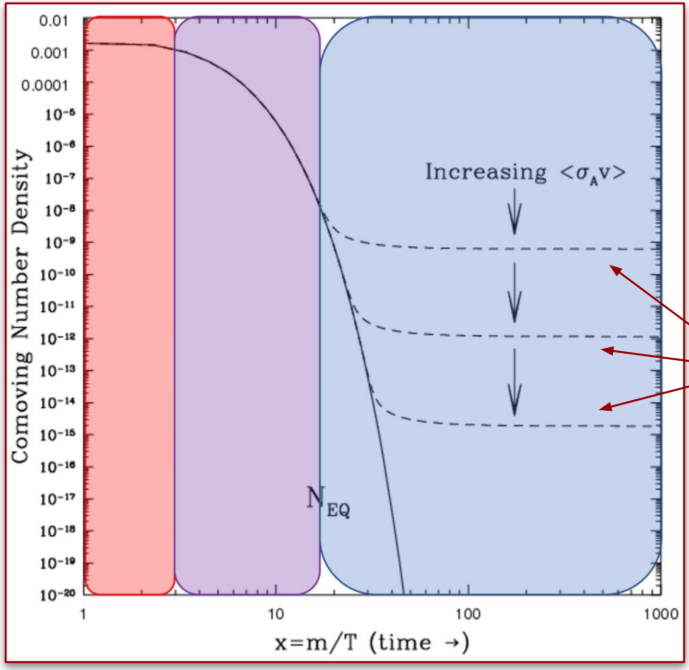
ended as universe cooled

Annihilation
Production < Annihilation

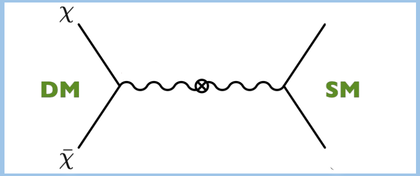


ended as universe expanded

Thermal Relic Density: Ω_χ



Now: Freeze-Out
Relic Density Set by <σ_Av>



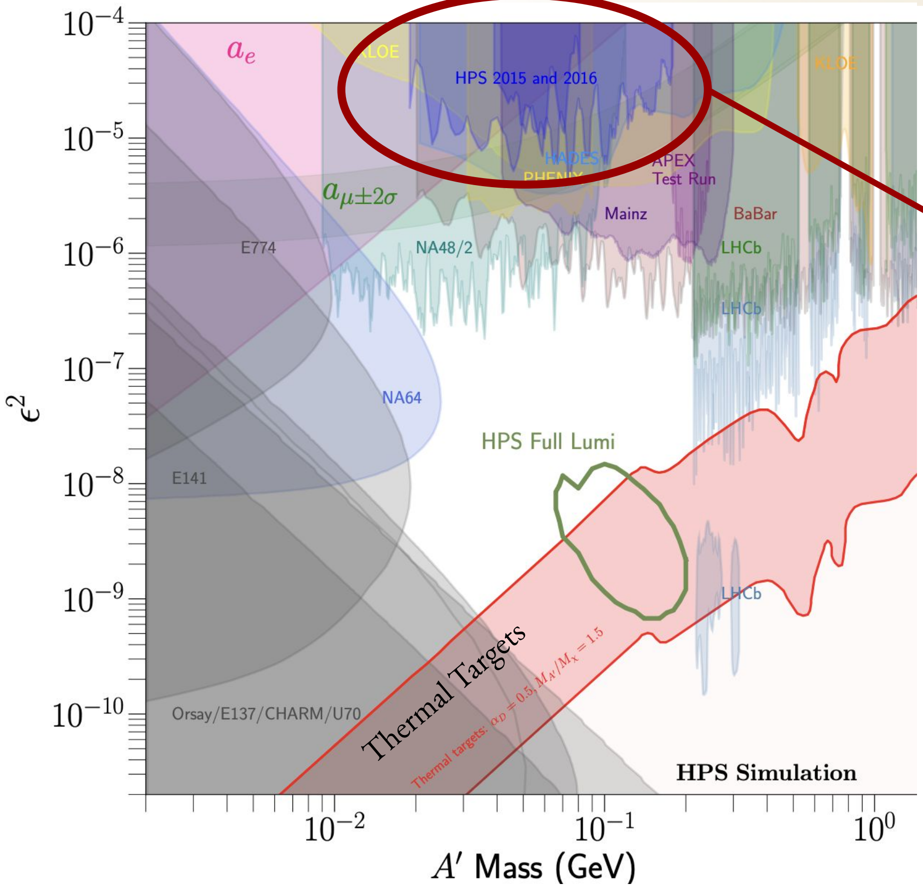
Observed DM abundance can correspond to different number densities depending on the characteristic mass and annihilation cross section.

Model Dependent

$$\Omega_\chi \propto \frac{1}{\langle \sigma v \rangle} \qquad \sigma v \propto \epsilon^2 \alpha_D \frac{m_\chi^2}{m_{A'}^4}$$

Kolb and Turner. The Early Universe (ISBN: 978-0201626742). Colored by Adam Green.

Physics Sensitivity of HPS



HPS has two primary search strategies for the A′ depending on the lifetime / kinetic mixing, or coupling strength, (ε^2).

HPS Prompt Resonance Search Result

For higher coupling strengths (lower lifetime), A′s are expected to decay extremely fast at the target and a signal is expected as a "bump" in the reconstructed e^+e^- invariant mass distribution (**IMD**).

HPS Displaced Vertex Search

Not discussed in this talk.

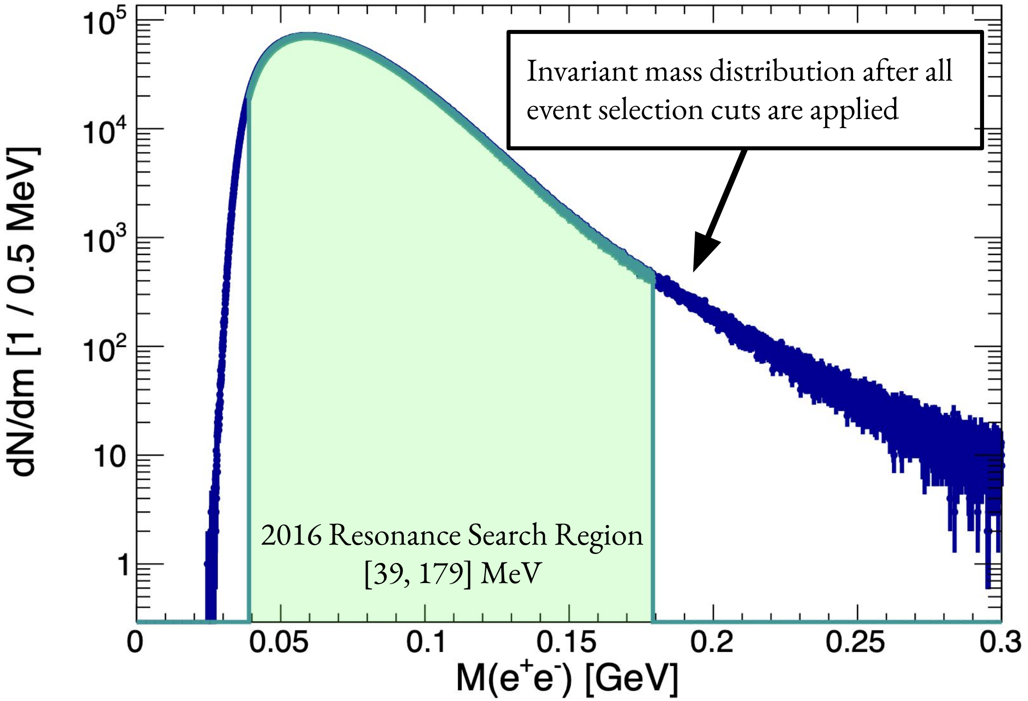
HPS 2023 Publication

HPS 2016 Reconstructed e^+e^- Invariant Mass Distribution

Data collected during 2016 engineering run with total integrated luminosity of **10, 608 nb^{-1}**.

- 67.2 mC or ~7 billion triggered events.

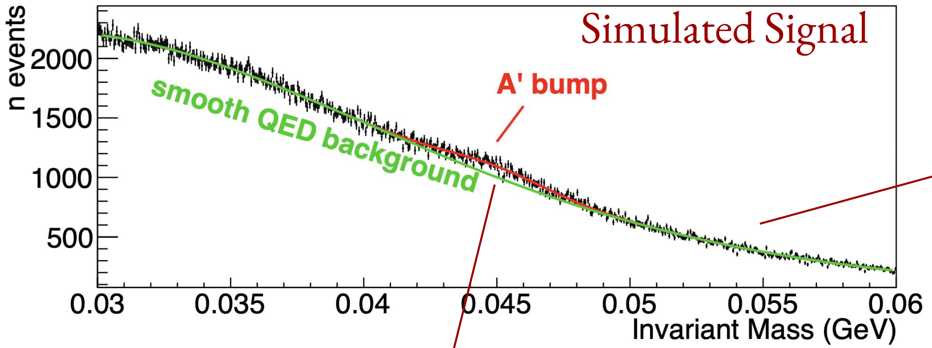
Raw data from the detector and simulation are cleanly reconstructed to (e^+e^-) pairs with shared vertices.



Event selection methodology / figure described in full in 2016 Physics Result

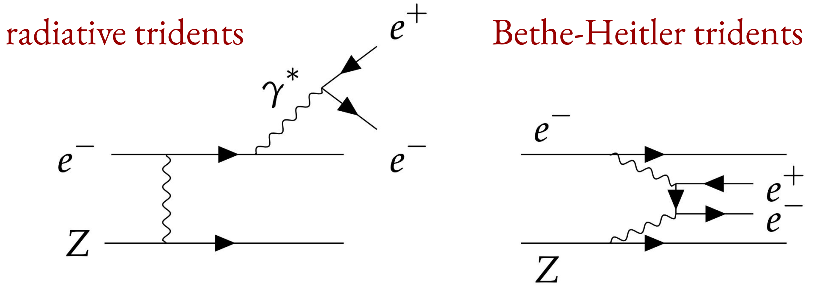
Prompt A′ Signal Model and Backgrounds

If A′ exists within the acceptance of HPS, it will present itself as a gaussian excess above background in the IMD.



natural width of A′ << detector resolution
observed signal width = experimental mass resolution

Primary Backgrounds



Additional background includes converted e^+e^- pairs from wide angle Bremsstrahlung photons inside the target or in the first two layers of the SVT.

Determining Upper Limits for each Mass Hypothesis

Signal Yield Upper Limit Statistical Test

Consists of two likelihood tests

I. background only and bkg + signal fit with a **fixed non-zero signal yield**

II. bkg + signal fit with signal yield left floating and a bkg + signal fixed signal fit

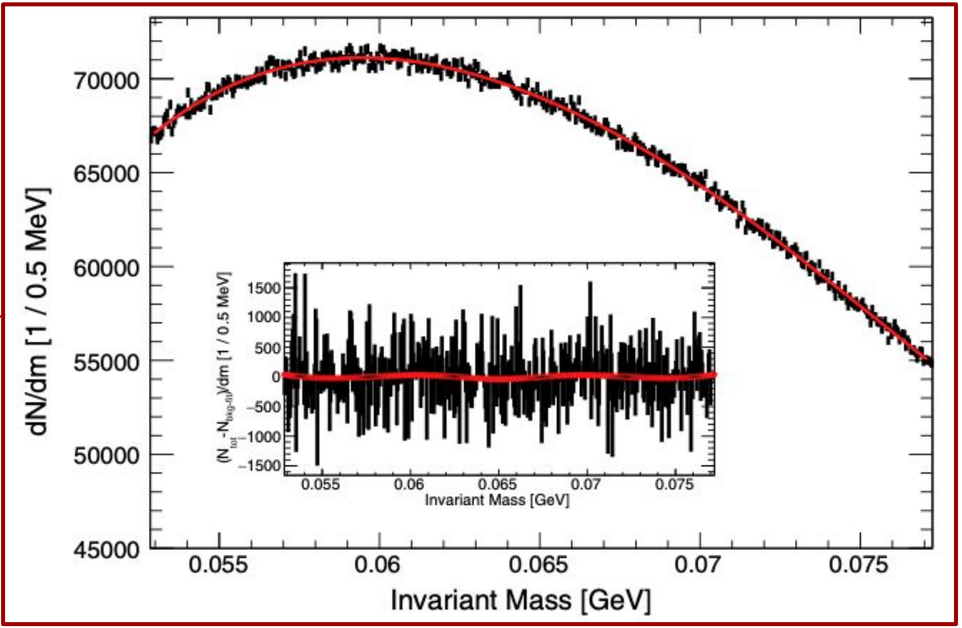
Iteratively done to find maximum fixed signal yield necessary to hit target confidence level threshold.

$$\mathrm{CL}_s(\mu) = \frac{p_\mu}{1 - p_b} \qquad \mathrm{CL}_s(N_{sig}^{up}) = 0.05$$

The upper limit on the signal yield is then propagated into the ε^2 upper limit equation.

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

Published method used a sliding background model **centered around each mass hypothesis** with fit window width determined by the respective mass resolution and shape from 3rd or 5th order Legendre polynomials.



Determining Upper Limits for each Mass Hypothesis

SLAC

Signal Yield Upper Limit Statistical Test

Consists of two likelihood tests

I. background only and bkg + signal fit with a **fixed non-zero signal yield**

II. bkg + signal fit with signal yield left floating and a bkg + signal fixed signal fit

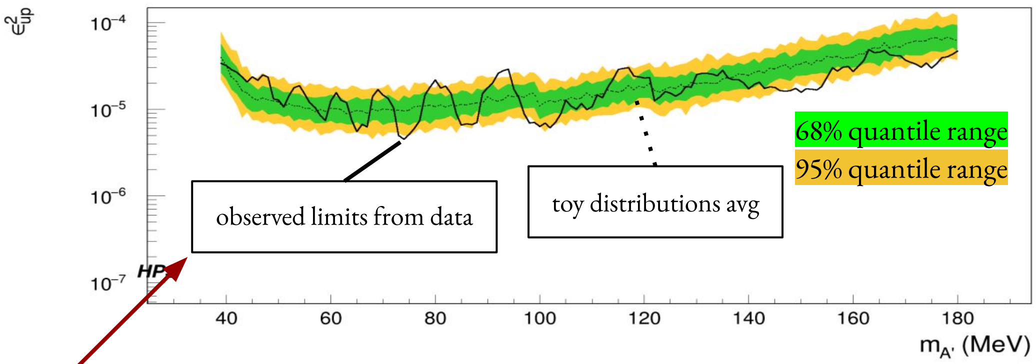
Iteratively done to find maximum fixed signal yield necessary to hit target confidence level threshold.

$$\mathrm{CL}_s(\mu) = \frac{p_\mu}{1 - p_b} \qquad \mathrm{CL}_s(N_{sig}^{up}) = 0.05$$

The upper limit on the signal yield is then propagated into the ε^2 upper limit equation

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

ε^2 Upper Limit Published Result



As published in Phys. Rev D, the HPS resonance search was conducted over the reconstructed (e^+e^-) invariant mass distribution **between 39 MeV and 179 MeV**, and found, in agreement with other searches, a limit of $\boldsymbol{\varepsilon^2 \geq 10^{-5}}$.

Changing the Background Model

Motivations

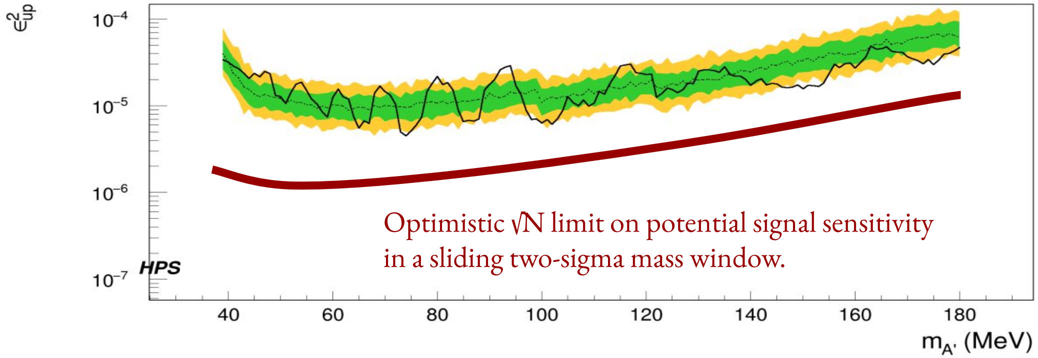
- Based on the statistical uncertainty only limit, there is roughly an order-of-magnitude improvement in sensitivity possible for our background model.

- Flexibility of background model chosen to minimize signal yield bias comes at cost to signal sensitivity.

- Background model not orthogonal to signal model as initially thought. Implies the absorption of signal-like events in each search window.

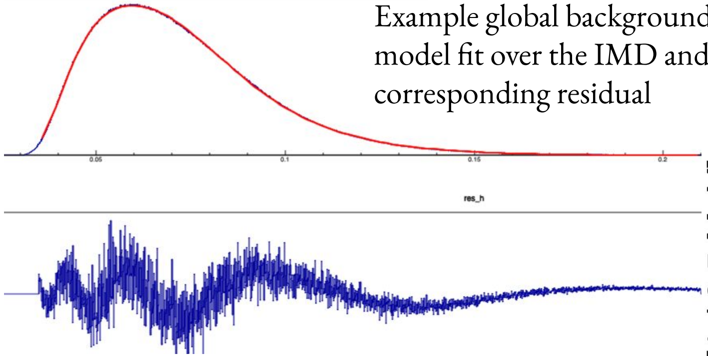
$$10^{\mathrm{L}_N(m_{\mathrm{e}^+\mathrm{e}^-}|\vec{t})}$$

ε^2 Upper Limit Published Result



Study supporting this conducted on significance of even ordered polynomial terms in background and background + signal fits.

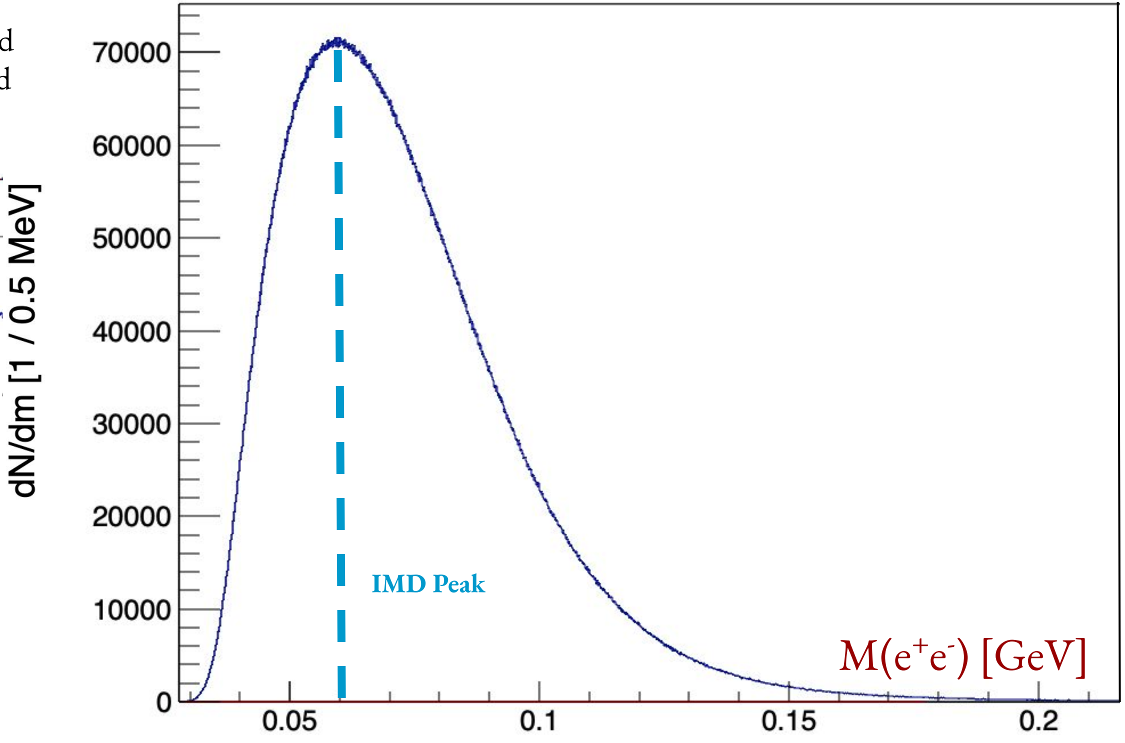
Improving the Background Model



Example global background model fit over the IMD and corresponding residual

A **global background model** fit over the entire IMD is being investigated in order to reduce background shape flexibility and uncertainty to improve the exclusion limit.

2016 Invariant Mass Distribution



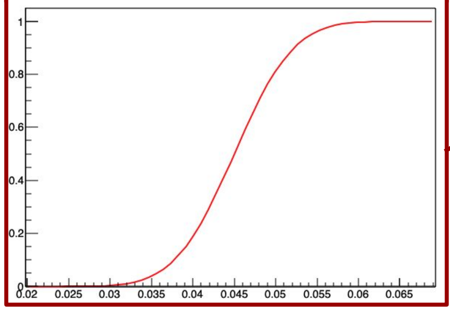
Looking for a Global Background Model

The general strategy for finding functions to fit the IMD is by modeling the broader scale features of the distribution.

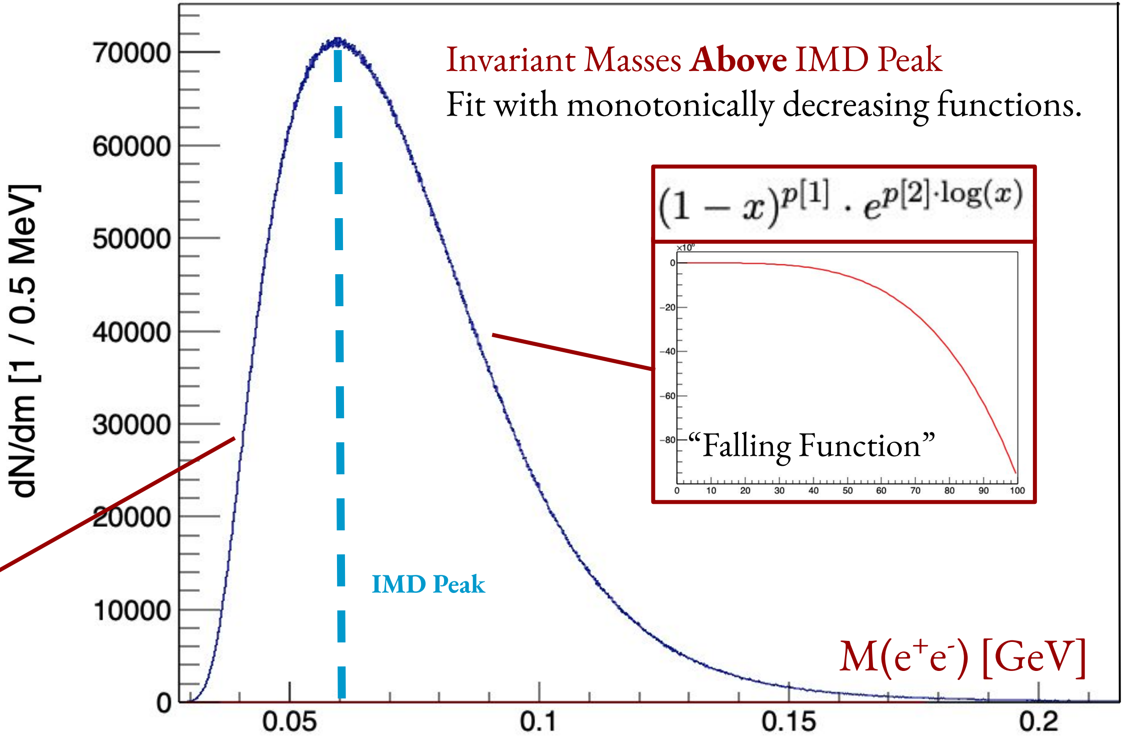
The shape of the IMD is complicated by the complex geometric acceptance of the SVT.

Invariant Masses **Below** IMD Peak
Fit with monotonically increasing functions.

$$\mathrm{Er}_i(x) = \frac{1}{2}\left(\mathrm{Erf}\left(\frac{(x - i[1])}{i[2]}\right) + 1\right)$$



2016 Invariant Mass Distribution



Looking for a Global Background Model

The general strategy for finding functions to fit the IMD is by modeling the broader scale features of the distribution.

The shape of the IMD is complicated by the complex geometric acceptance of the SVT.

Generic Functional Form

$$\mathcal{F} = \sum_i \left(\mathrm{Er}_i \cdot \mathrm{FF}_i \right)$$

Proof of concept conducted from multi stage fitting and selection procedure.

Initial Falling Functions

$$f_{dijet1}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2}} \qquad f_{dijet2}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\log(x)}}$$

$$f_{dijet3}(x) = \frac{p_0(1-x)^{p_1}}{x^{p_2+p_3\log(x)+p_4\log^2(x)}} \qquad f_{ATLAS1}(x) = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2}}$$

$$f_{ATLAS2}(x) = \frac{p_0(1-x^{1/3})^{p_1}}{x^{p_2+p_3\log^2(x)}} \qquad f_{UA2_1}(x) = p_0 x^{p_1} e^{p_2 x}$$

$$f_{UA2_2}(x) = p_0 x^{p_1} e^{p_2 x + p_3 x^2} \qquad f_{UA2_3}(x) = p_0 x^{p_1} e^{p_2 x + p_3 x^2 + p_4 x^3}$$

$$f_{cmsBH1}(x) = \frac{p_0(1+x)^{p_1}}{x^{p_2\log x}} \qquad f_{cmsBH2}(x) = \frac{p_0(1+x)^{p_1}}{x^{p_3+p_2\log x}}$$

$$f_{ATLASBH1}(x) = p_0(1-x)^{p_1} x^{p_2\log(x)} \qquad f_{ATLASBH2}(x) = p_0(1-x)^{p_1}(1+x)^{p_2\log(x)}$$

$$f_{ATLASBH3}(x) = p_0(1-x)^{p_1} e^{p_2\log(x)} \qquad f_{ATLASBH4}(x) = p_0(1-x^{1/3})^{p_1} x^{p_2\log(x)}$$

$$f_{ATLASBH5}(x) = p_0(1-x)^{p_1} x^{p_2 x)} \qquad f_{ATLASBH6}(x) = p_0(1-x)^{p_1}(1+x)^{p_2 x}$$

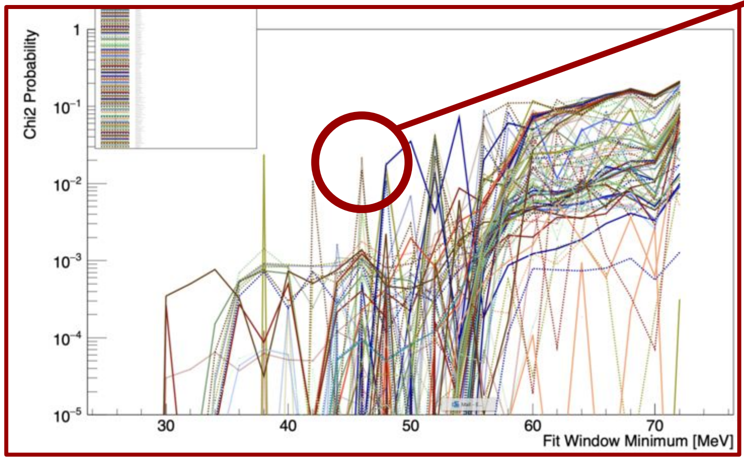
C. Bravo.

Function Selection Procedure (1/3)

Prototyping Stage

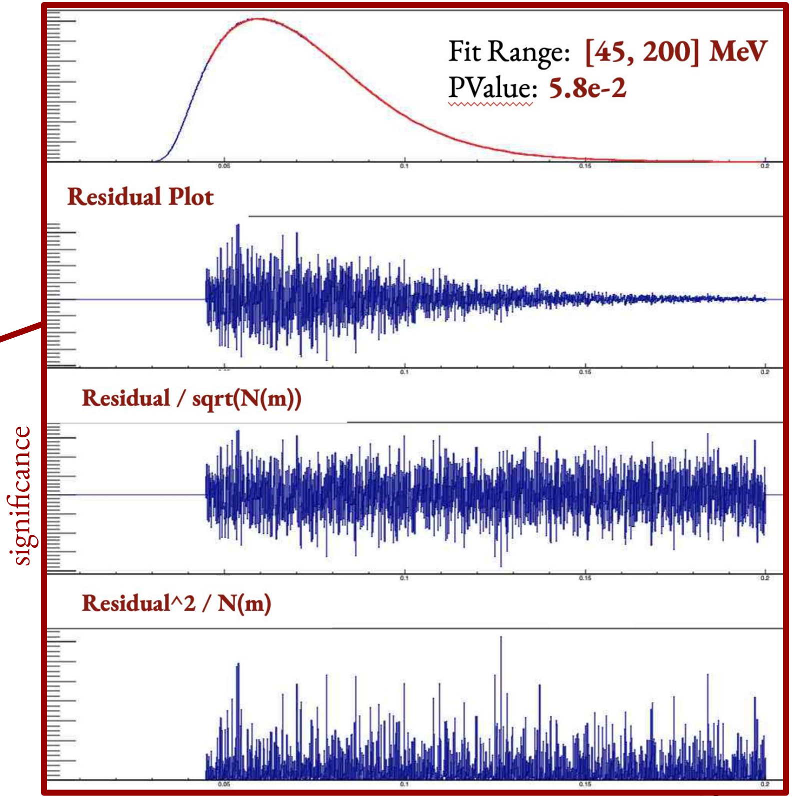
- construct and test functions + displays
- use MINUIT to find rough initial parameters
- develop global fitting analysis infrastructure

χ^2 Probability Compilation



Display tool used to compare function performances across a range of fit windows.

Fit Displays



Function Selection Procedure (2/3)

Preliminary Fitting and Filtering

- All functions are fit over a single invariant mass range with dynamic changes in seeding of initial function parameters.
- Store results of all functions meeting a goodness-of-fit requirement.

Dynamic Parameter Seeding Strategy for Background Model Fitting

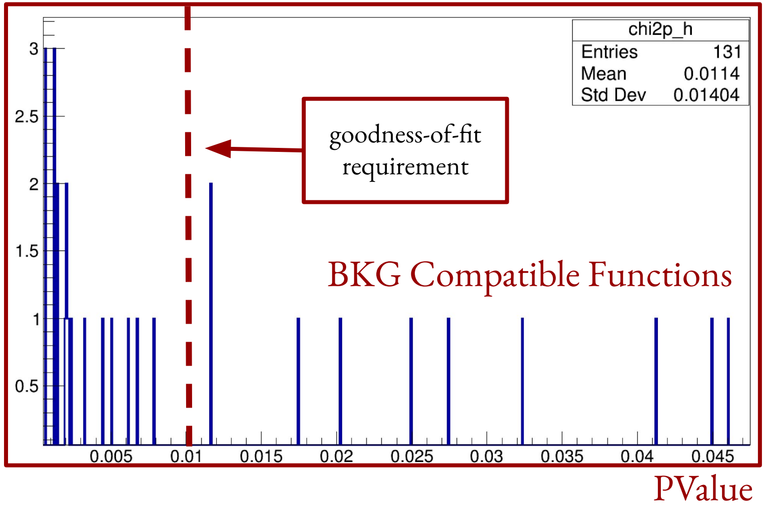
I. For each fit, randomly generate initial fit parameters using a gaussian with a mean equal to stored parameters and width 1% of the mean.

$$\text{Gaussian}(\mu = [\text{Stored Parameter}], \sigma = .01 \cdot \mu) \rightarrow [\text{Initial Fit Parameter}]$$

II. For a specified number of fits (1 count)
   A. Better pvalue found: save fit/parameter info and use for successive fits
   B. Better pvalue not found: increase gaussian width using the general form

$$\sigma = .01 \cdot \mu + .01 \cdot \text{counter} \cdot \mu$$

1D Pvalue Distribution: Fit Range[45, 198 MeV]

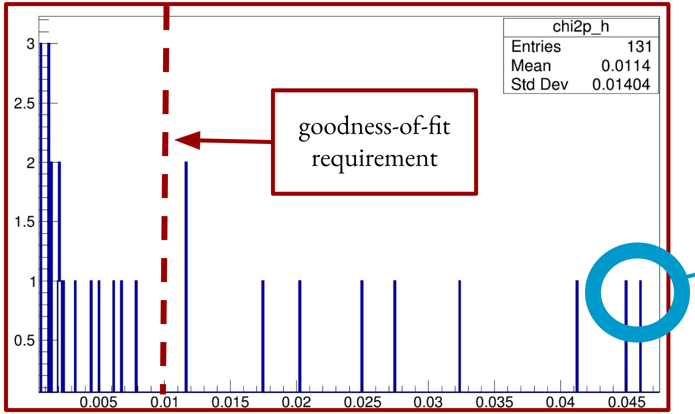


Function Selection Procedure (3/3)

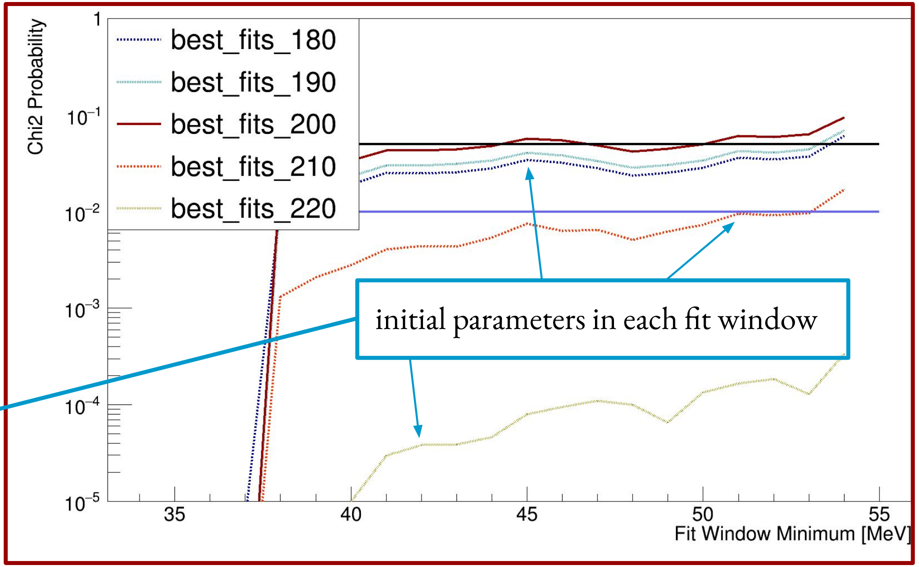
Iterative Fit Window Scanning

- Use previous stage parameter seeds and fit across varying window ranges for candidate functions.

1D Pvalue Distribution: Fit Range[45, 198 MeV]



Candidate Function χ^2 Probability Compilation



General trend from tested functions

$$\chi^2\text{Prob}(220\text{MeV}) < \chi^2\text{Prob}(210\text{MeV}) < \chi^2\text{Prob}(180\text{MeV}) < \chi^2\text{Prob}(190\text{MeV}) < \chi^2\text{Prob}(200\text{MeV})$$

true muonium rulez

Unblinded ε² Upper Limit Results

Candidate Background Model Functional Form

$$\mathcal{C} \cdot \left[\mathrm{Er}_1 \cdot (1-x)^{p[1]} \cdot e^{p[2]\cdot\log(x)} + \mathrm{Er}_2 \cdot q[1] \cdot (1-x)^{q[2]} \cdot (1+x)^{q[3]\cdot x}\right]$$

Global Normalization Constant

Once candidate function has been determined

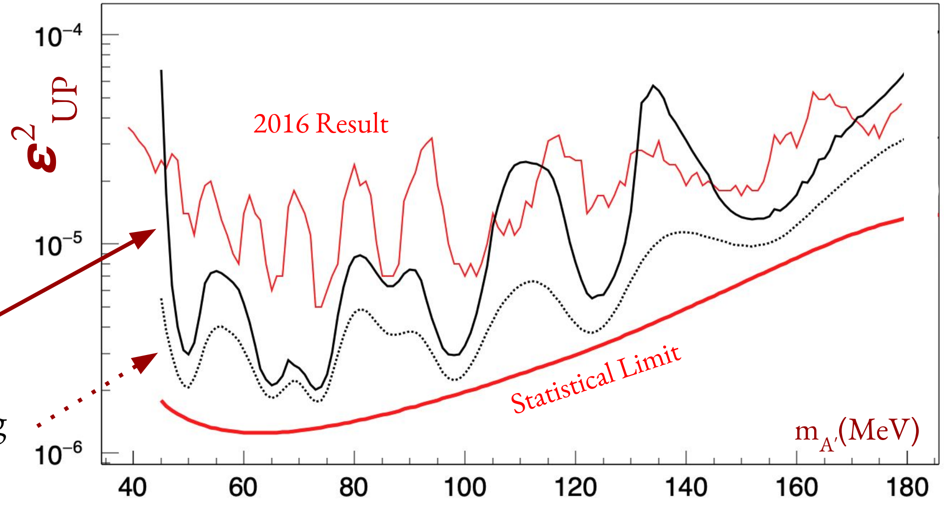
Fit over full IMD using HPS Analysis Software.

Study 1: all background parameters are floating
Study 2: only the global normalization constant is floating

Compute observed upper limits on signal yield and incorporate background + radiative fraction to determine ε^2.

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

ε² Upper Limit Comparison



Background Model Parameterization Studies (1/2)

las3pluslas6 Functional Form

$$\mathcal{C} \cdot \left[\mathrm{Er}_1 \cdot (1-x)^{p[1]} \cdot e^{p[2]\cdot \log(x)} + \mathrm{Er}_2 \cdot q[1] \cdot (1-x)^{q[2]} \cdot (1+x)^{q[3]\cdot x}\right]$$

<u>Initial Freezing Studies with error function parameters</u>

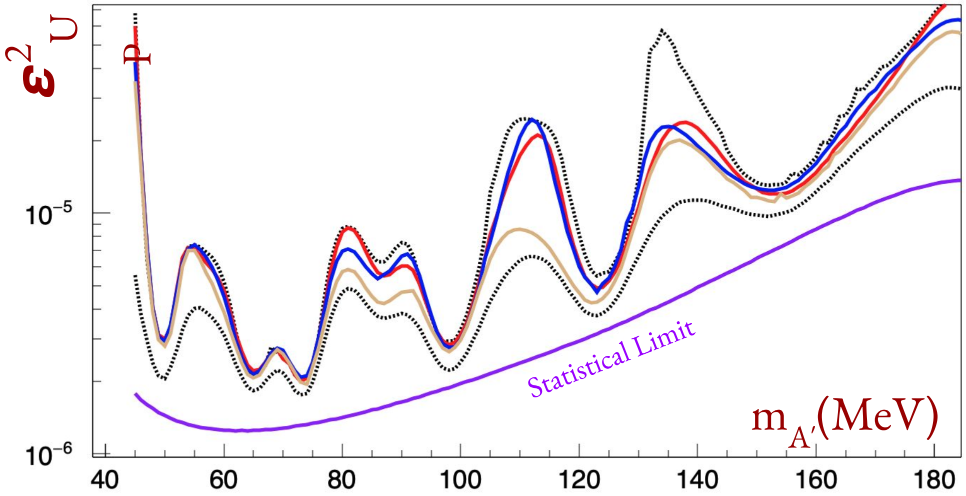
Fixed both "translational parameters" left everything else floating [blue]
Fixed all error parameters left everything else floating [tan]
Fixed one translational/one "scaling" parameter everything else float [red]

translational

$$\mathrm{Er}_i(x) = \frac{1}{2}\left(\mathrm{Erf}\left(\frac{(x-i[1])}{i[2]}\right)+1\right)$$

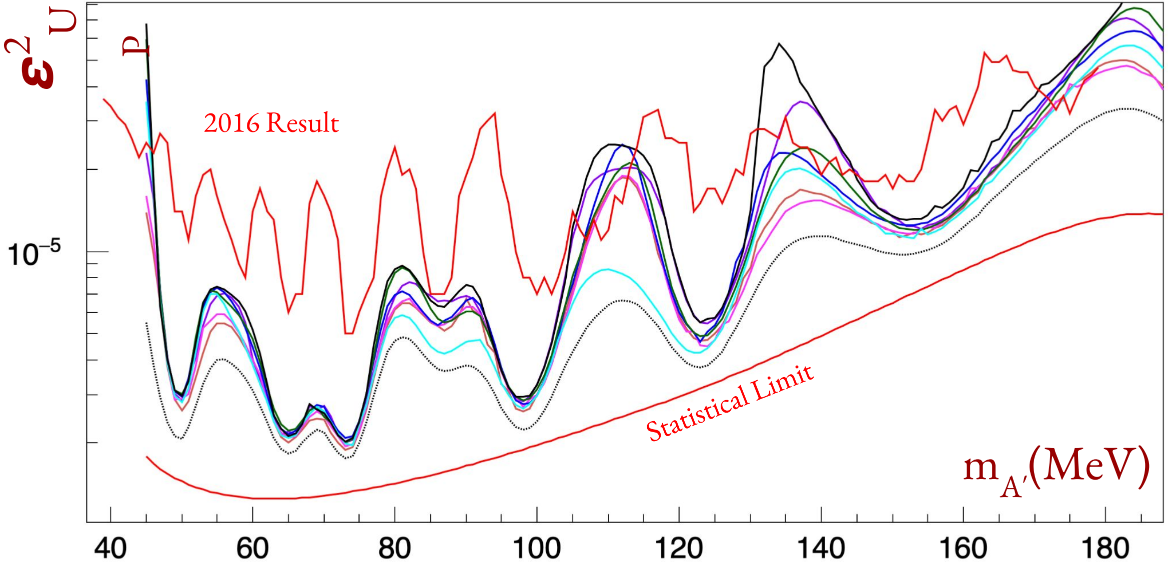
scaling

ε^2 Upper Limit Comparison



Background Model Parameterization Studies (2/2)

ε^2 Upper Limit Comparison



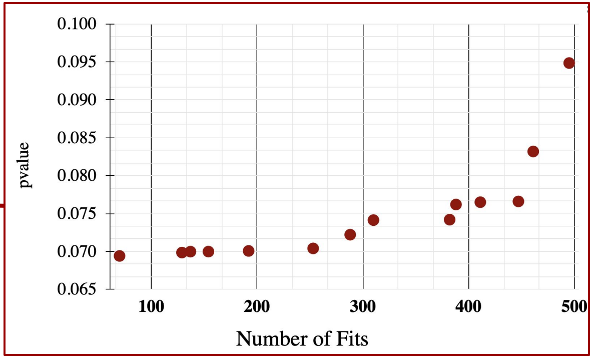
Blinded Procedure: 6.5% —> 100%

SLAC

Candidate Function Fit Procedure for the 6.5% IMD

I. Initial candidate function parameters are left floating
II. Dynamic seeding parameter selection strategy used
   A. For fixed amount of time, fit IMD window range of **[45, 200] MeV.**
   B. For each better fit, store parameters. Lowest pvalue fit parameters stored to populate 100% parameters.

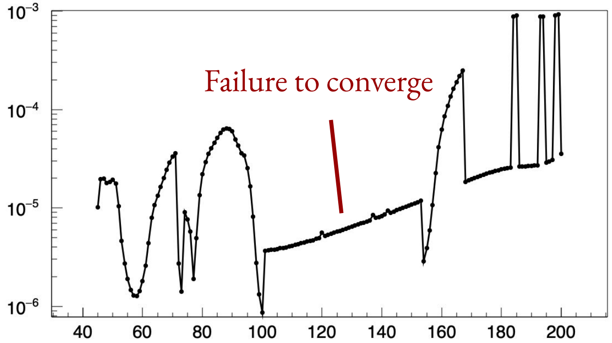
Dynamic fitting and parameter selection process conducted on the 6.5% 2016 IMD



Candidate Function Fit Procedure for the 100 % IMD

I. Determine parameters to fix based on stored parameters
II. Difficult
   A. May need to project parameters taking into account expected statistical uncertainties
   B. May also make function selection more robust

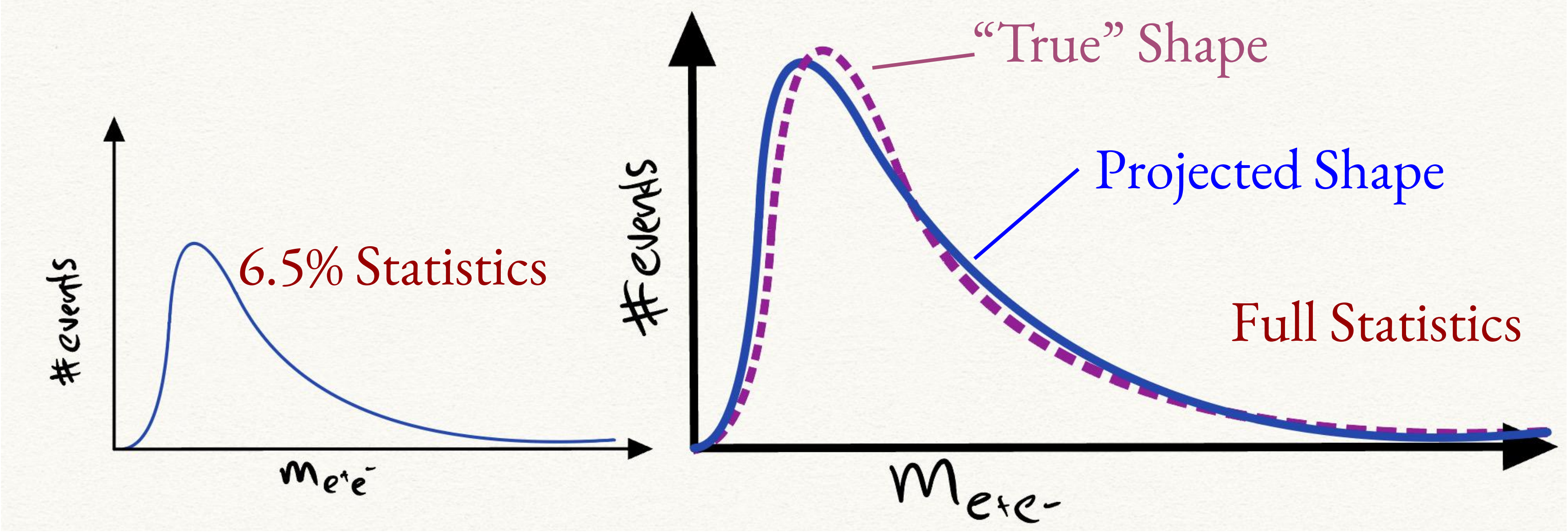
Projected BKG Parameterization of 100%



Challenge of Projecting Parameters

Parameters stored and fixed using subset of data set project a shape which may deviate from the shape of the full distribution.

insert discussion



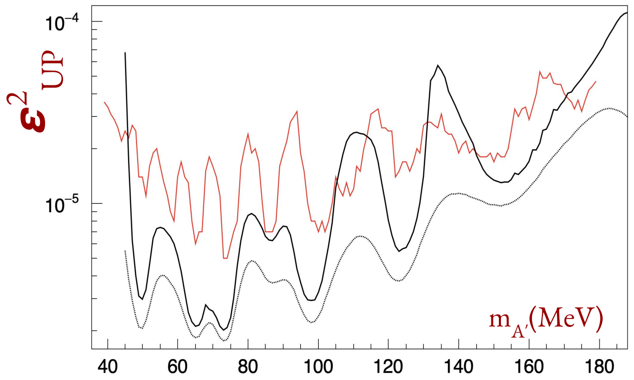
Conclusions and Analysis Framework Moving Forward

Demonstrated proof of principle

- Changing BKG parameterization can improve exclusion results.
- Have found promising global background model candidates.

Blinded fitting procedure difficult but within reach! Larger datasets offer exciting place to utilize techniques.

ε^2 Upper Limit Comparison



Next Steps

Tidy Blinded Procedure
Signal Injection Studies
Signal shape switch to data

Luminosity of Datasets

2016 Luminosity: 10 pb^{-1}
2019 Luminosity: 110 pb^{-1}
2021 Luminosity: 160 pb^{-1}

Calibrations are being finalized for the HPS Physics Runs of 2019/2021.

Global BKG Fitting of 2015 IMD (Re-Re-Analysis)

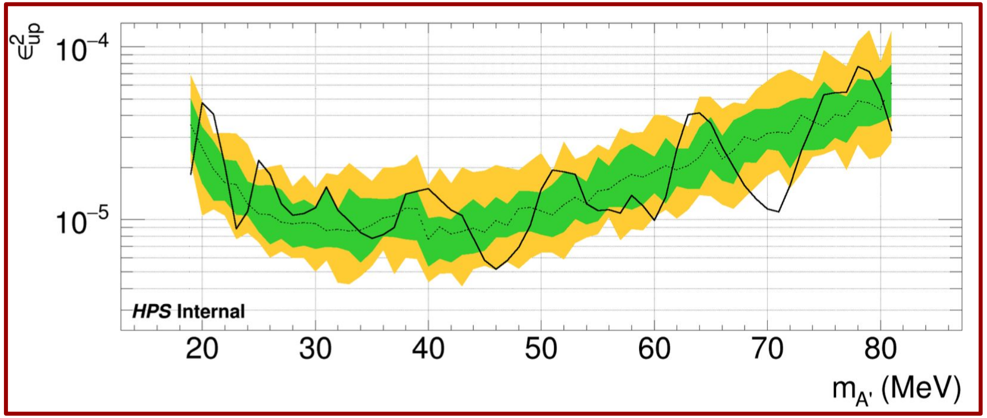
Summer student TJ Britt will work on implementing global fitting techniques on 2015 IMD.

Objectives

- Standardize procedure across datasets.
- Combine 2015/2016 upper limits to create 2019, 2021 template.

Need 2015 histogram.

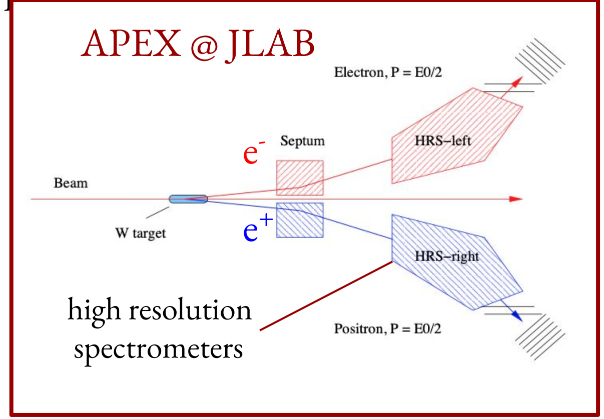
2015 ε^2 Upper Limit Result (re-analysis)



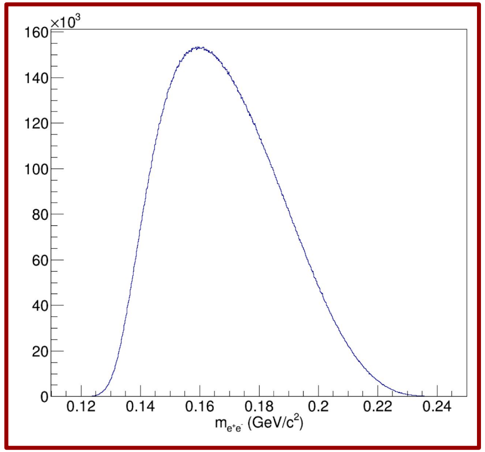
2015 Engineering Run: 1.2 pb^{-1}, 1.06 GeV

Additional Use Case: APEX

APEX, a JLAB fixed target experiment, has nearly identical resonance search methodology to HPS.

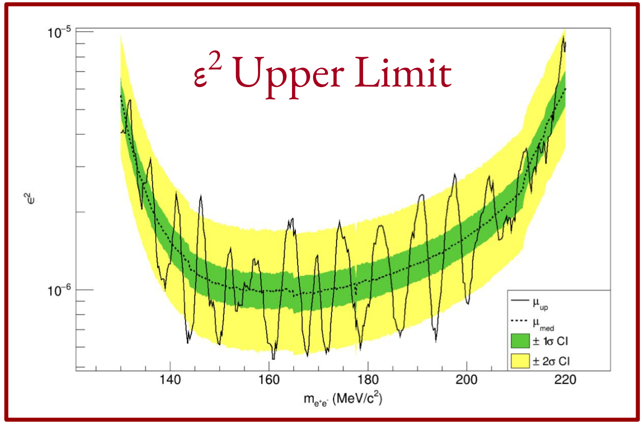


APEX Blinded 10% e^+e^- IMD



Only 10% of 2019 dataset has been analyzed.

Systematic similarities to HPS results.



[APEX: A Search for Dark Photons in Hall A]

[Williamson, John Thomas Austin (2022) *APEX (A' Experiment): The search for a dark photon at Jefferson Lab.* PhD thesis.]

APEX is an opportunity to leverage HPS analysis techniques, improve physics sensitivity in **well motivated parameter space**, and publish a result.

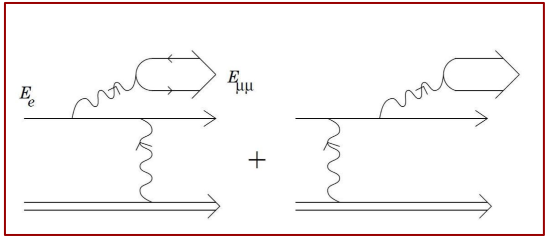
Extra Slides

- Full Previous Resonance Search Procedure
- Local/Global P-Values

Electroproduced Orthodimuonium (13S1) at HPS

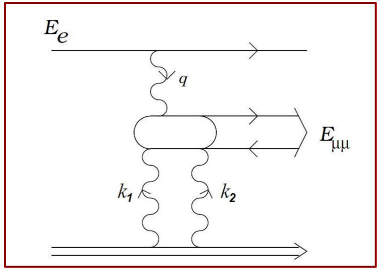
HPS displaced vertex capabilities are sensitive to existing/predicted standard model particles with decay lengths and corresponding (e^+e^-) pair distributions within the detector's acceptance.

Radiative Production



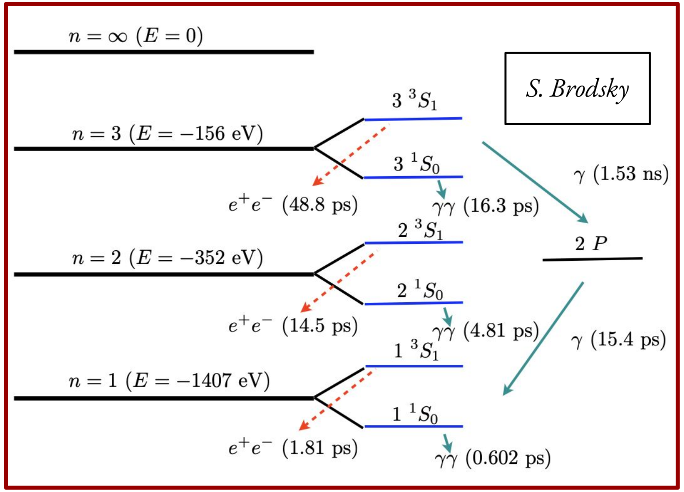
[5.77+ in 2019+2021]

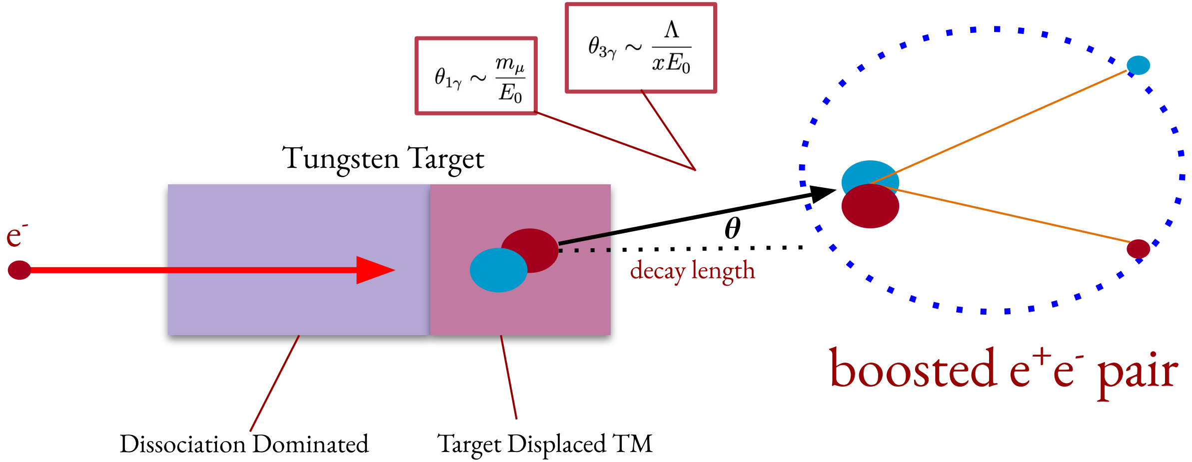
Three Photon Production



[10.37+ in 2019+2021]

True Muonium Level Diagram





Resonance Search Procedure

For each mass hypothesis

I. Fit data in each window with background only and background + signal models
   A. legendre polynomial order and window size determined during background model tuning
      a) determine fit parameters and integral of resultant PDFs (number of data events in window)
   B. T-Test conducted using 1000 background + signal fits
      a) generate mass resolution used for signal from a gaussian with mean set to the experimental resolution and width set to mass resolution uncertainty
      b) use likelihood ratio from each fit to calculate p-value
      c) take **84th percentile** resultant pvalue as the **observed local p-value**
   C. Conduct signal yield upper limit statistical test (eps^2 test done with value determined)
      a) composed of two likelihood tests
         (1) background only and bkg + signal fit with a **fixed non-zero signal yield**
         (2) bkg + signal fit with signal yield left floating and a bkg + signal fixed signal fit
      b) Iteratively done to find maximum fixed signal yield necessary to hit target confidence level threshold
         (1) Limit is the fixed signal yield such that $\mathrm{CL}_s(N_{sig}^{up}) = 0.05$

Local p-values and the Look Elsewhere Effect

Each mass hypothesis has a representative background fit as determined by the 2016 fit selection.

- corresponding χ^2 probabilities are "local" to the fit window
- global pvalues must be determined and take into account statistical fluctuations expected for searching **multiple independent regions**

The Look-Elsewhere Effect defines global p-values as being proportional to the number of independent regions:

$$p_{global} = p_{local} * N_{regions}$$

where $N_{regions} = W / \sigma_{ave}$

- W: total search window size
- σ_{ave}: average mass resolution

In 2016, $\mathbf{N_{regions} \sim 32}$
- implying a sufficiently independent search region *on average* every ~4.4 MeV