Extending SIMP Search to L1L2

Tom Eichlersmith

he/him/his

University of Minnesota

eichl008@umn.edu

September 10, 2024

Premise

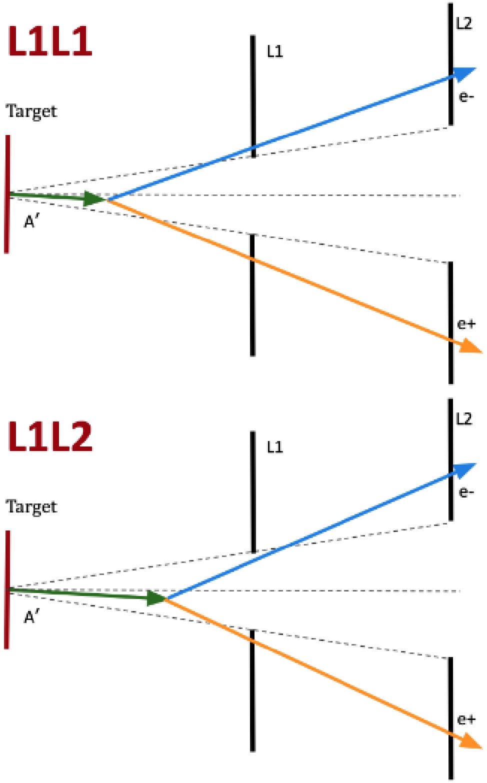


Fig 30 from PRD

Samples

Same data and simulation samples as Alic's SIMP (L1L1) search.

Selections

Rely on Alic's thorough study and validation, copy pre-selection and start with same final selection variables.

Search and Exclusion

- Search for excess in m_{reco} vs $\min(|y_{0,e^-}|, |y_{0,e^+}|)$ space
- Exclude by using OIM on the z distribution after final selections

Technical Changes

OIM Implementation Bug

Found and fixed bug in OIM implementation that was underestimating the maximum allowed signal yield.

Statistical Combination of L1L1 and L1L2 Exclusions

Combining the signal efficiency distributions and observed data z distributions before calculating the expected signal yield and running OIM.

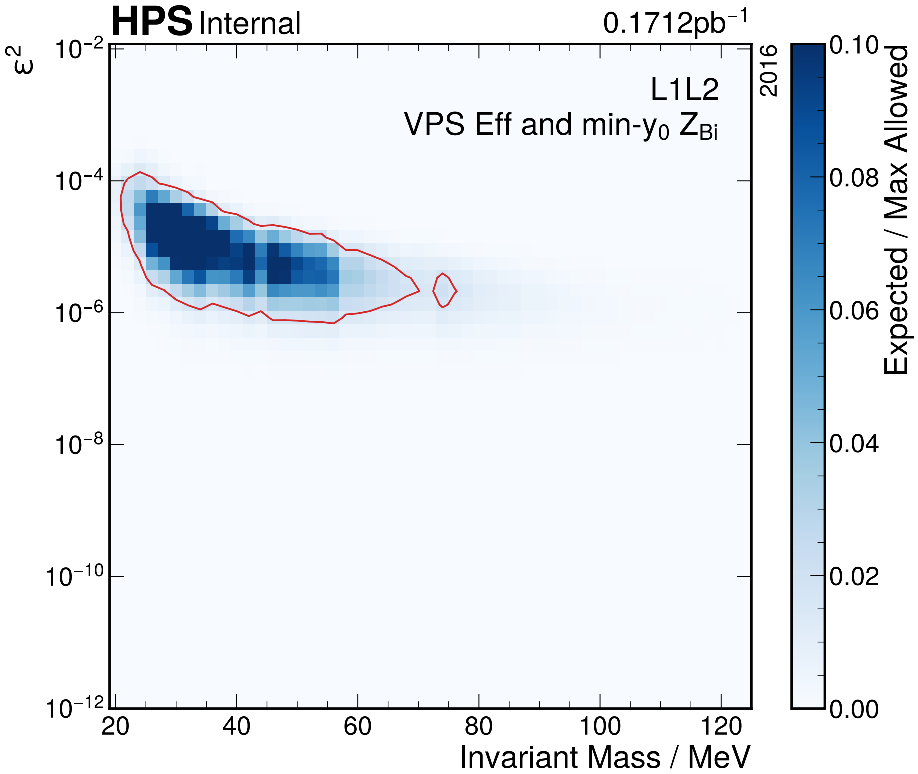
Calculated L1L2 Mass Resolution

Corresponds to a $\sim 15\%$ increase relative to L1L1 mass resolution (as expected from earlier estimates).

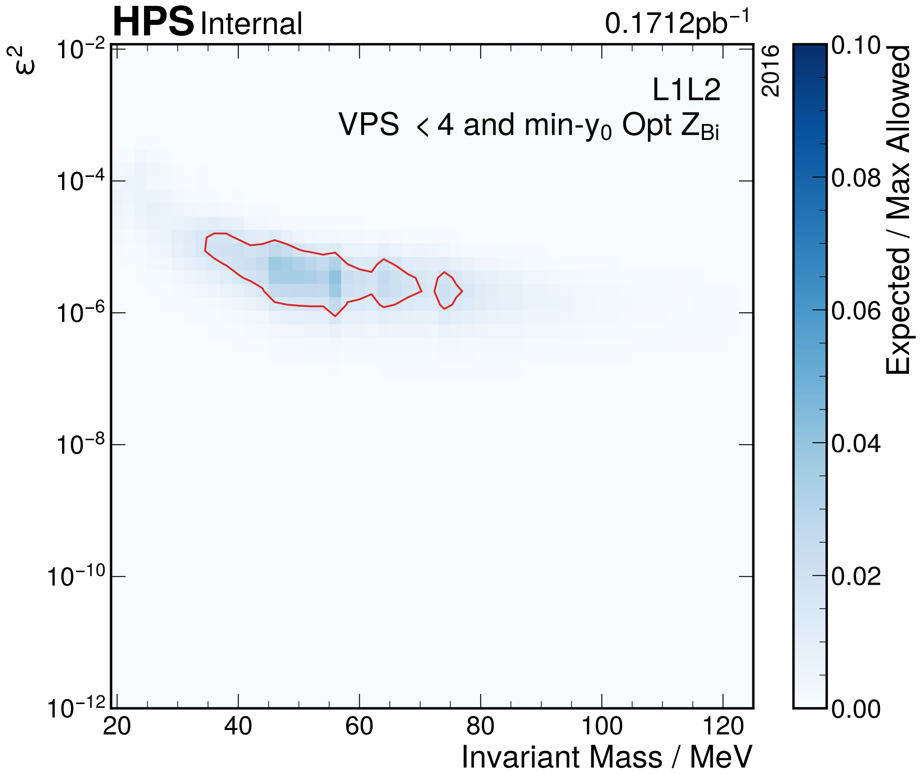
Effect of OIM Bug

Largest Effect on Higher Number of Background Events

With Bug



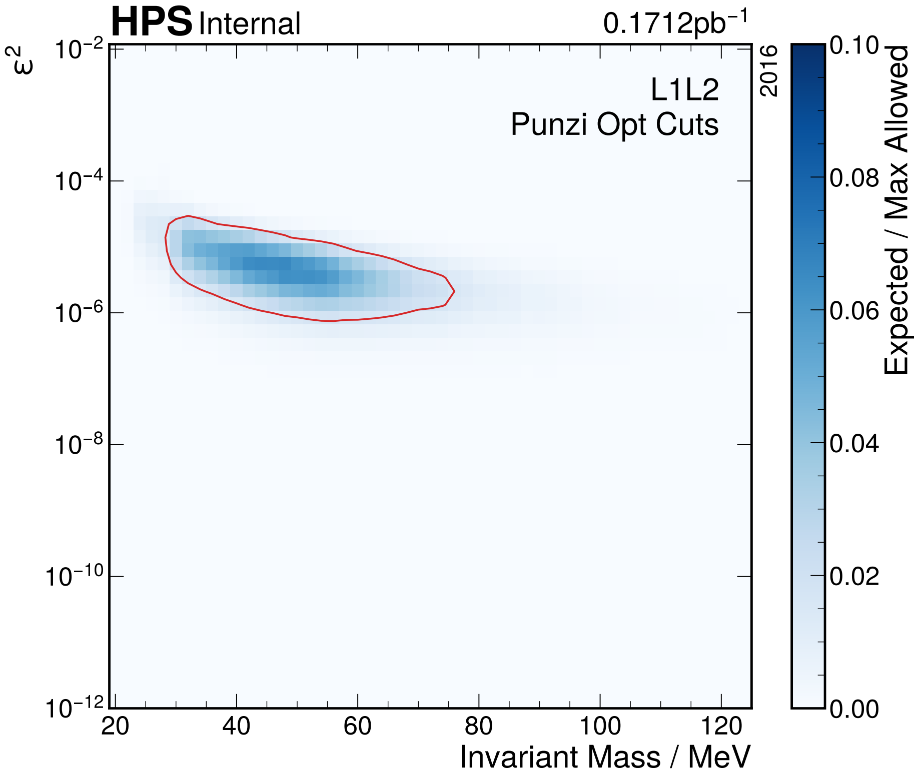
Bug Patched



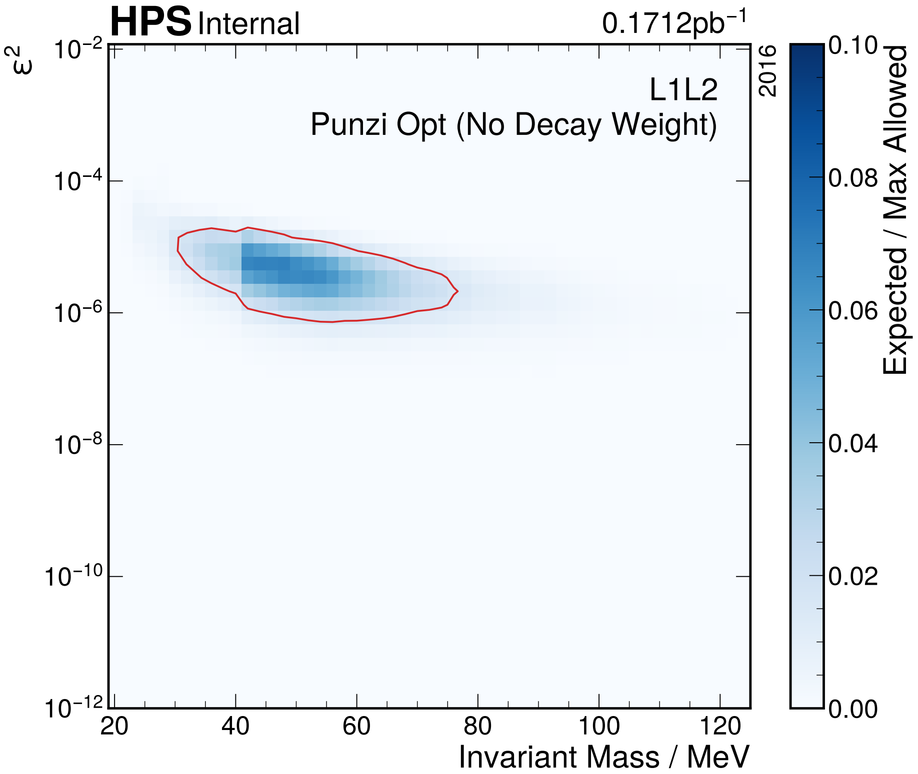
Effect of OIM Bug

Smaller Effect on Lower Number of Background Events

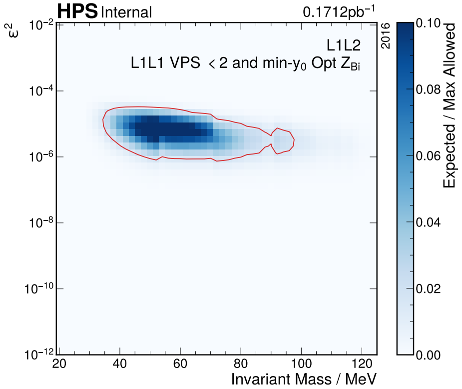
With Bug

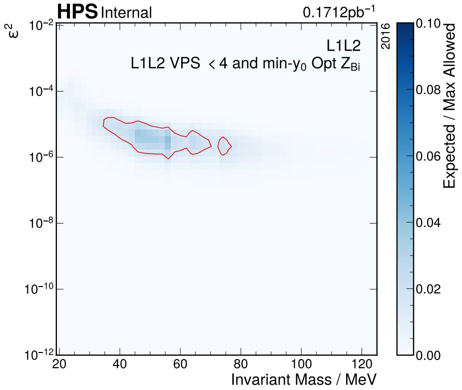


Bug Patched

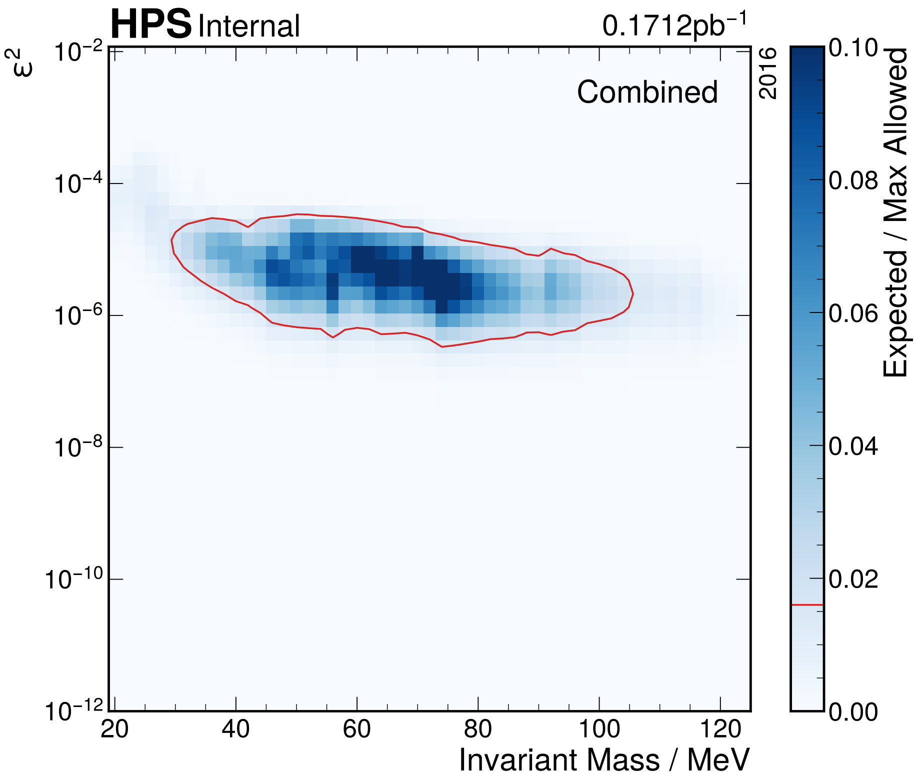


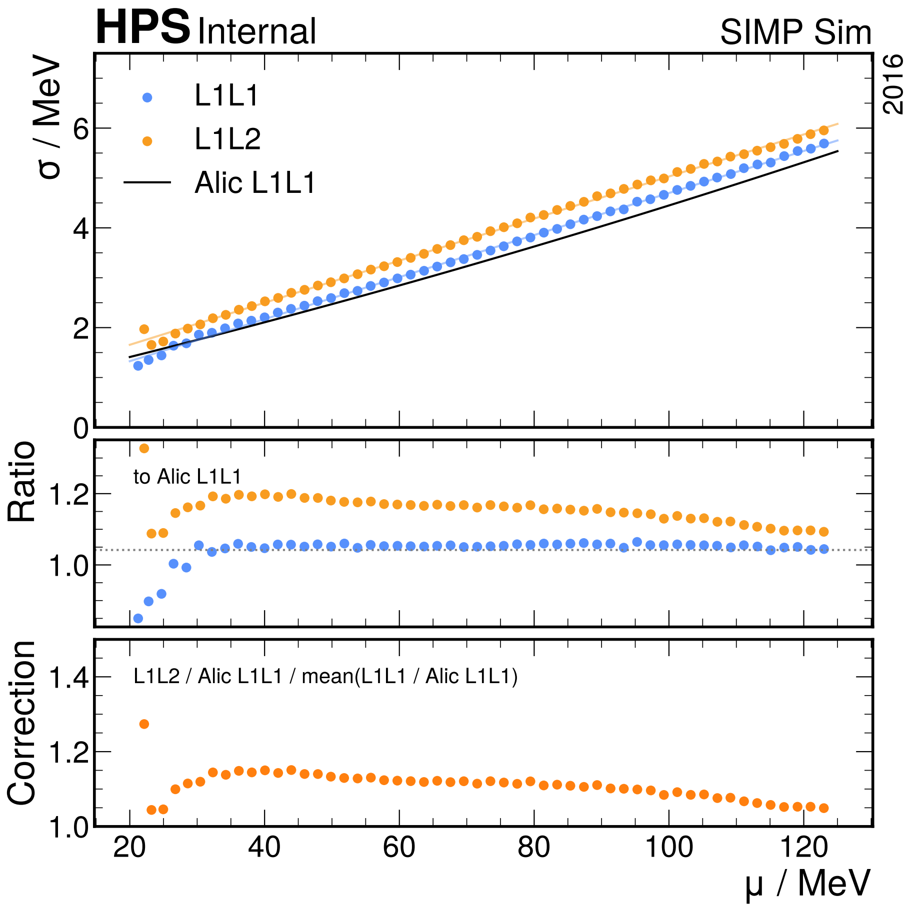
Combined L1L1 and L1L2 Exclusion Estimate



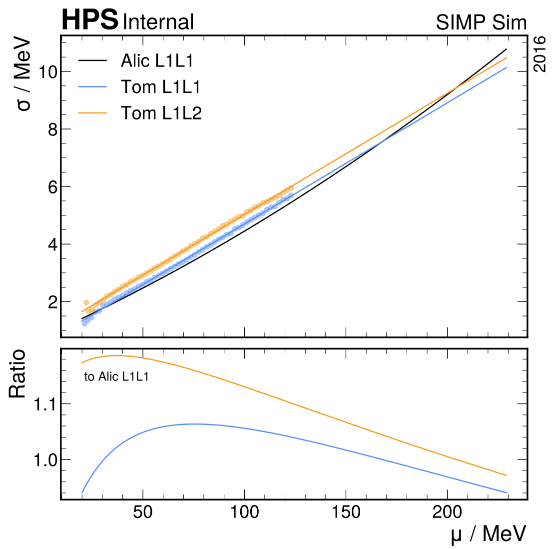


Calculate expected and maximum allowed signal yield *after* combining the signal and data *z* distributions from the two orthogonal channels.





- Alic's L1L1 Mass Resolution function in plot as reference
- Fitting the L1L1 and L1L2 mass resolutions with a line since Alic's quadratic function

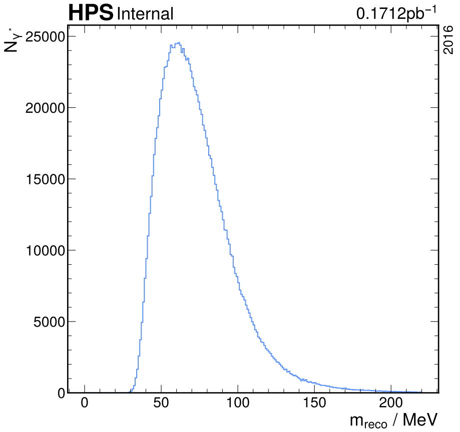


Improving the Exclusion Estimate

Context

Expected Signal

- **TDP** is data-driven, not statistically-bound data-driven ingredient
- Signal efficiencies are from signal simulation and also disconnected from size of data sample



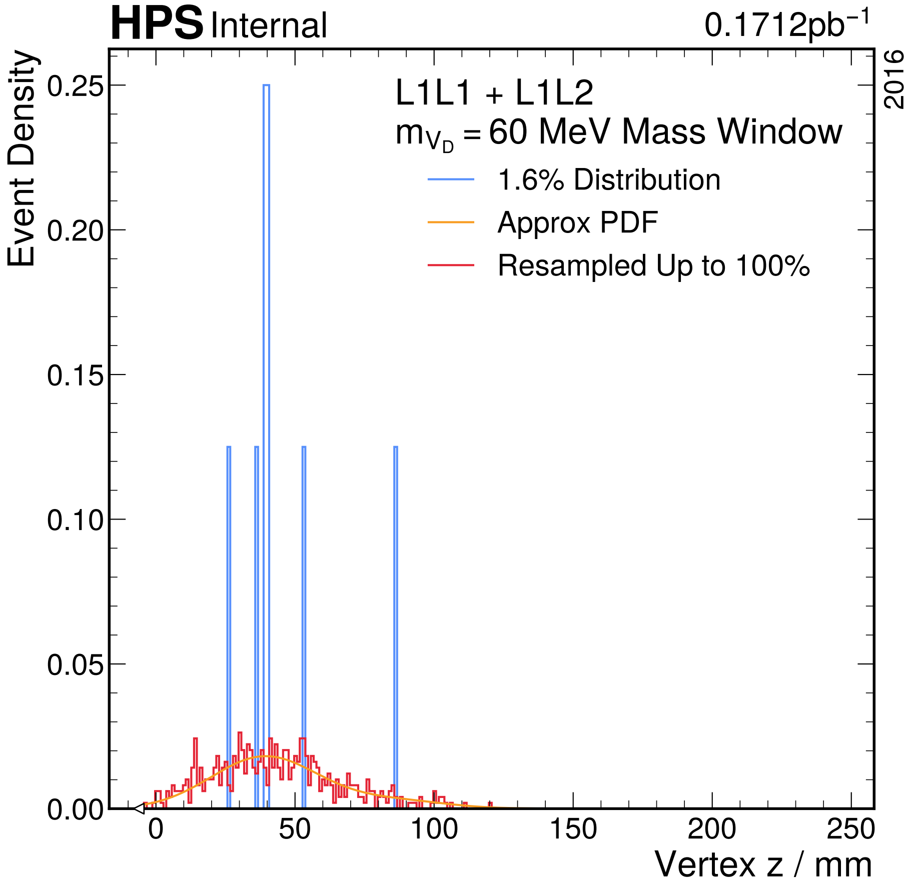
Maximum Allowed Signal

- Use the OIM to interpret a observed z distribution into a maximum allowed signal yield
- Signal differential yield from before is not an issue
- **observed** z distribution is directly limited by size of data sample

Goal

Scale-up observed z distribution while avoiding simply scaling the statistical fluctuations.

Gaussian KDE Smoothing

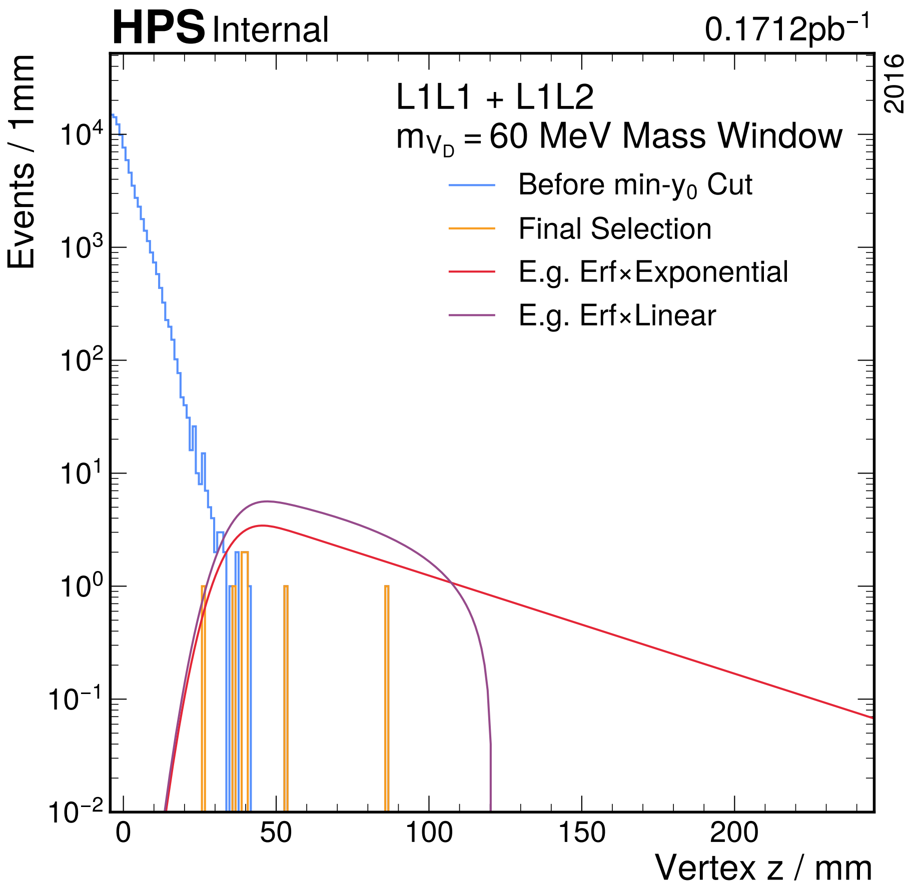


▸ Wikipedia:KDE Approximate parent PDF as sum of normal distributions with a shared width h.

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{N}\left(\frac{x - x_i}{h}\right)$$

Example on left is with $h = 1\,\text{mm}$

- ✓ Single stage, few assumptions to validate
- ✗ Location heavily biased by observations
- ✗ What to do in zero-observed situations?

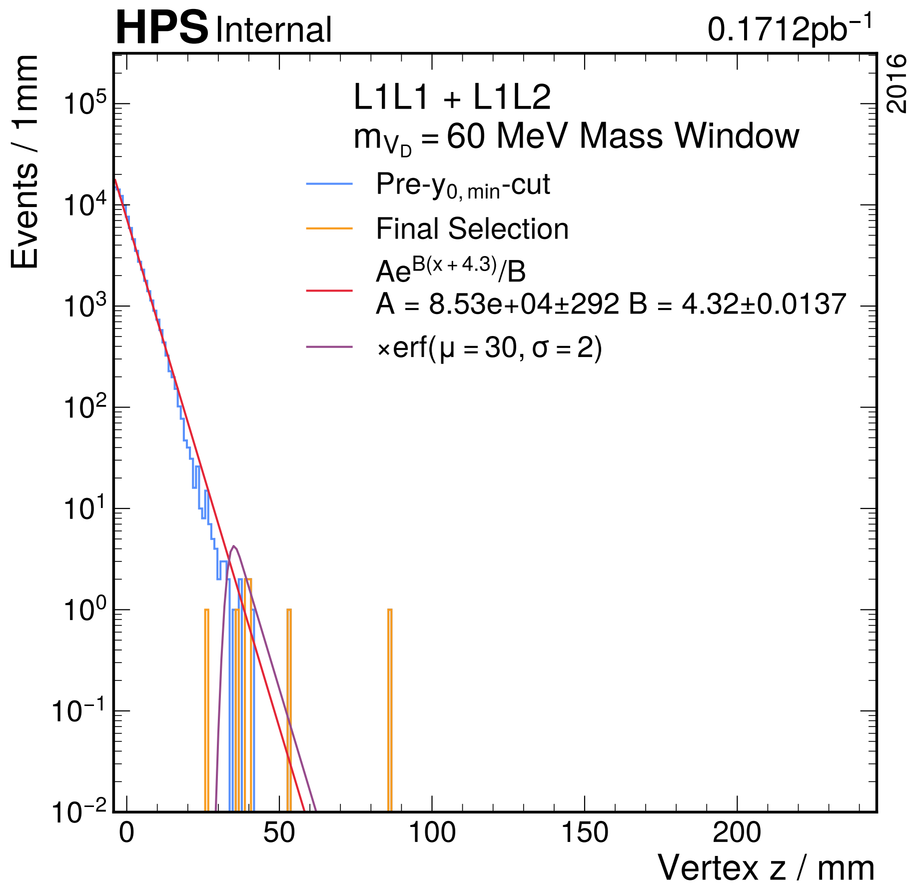


Approximate parent PDF as PDF of pre-$y_{0,\mathrm{min}}$-cut distribution multiplied by an error function ($y_{0,\mathrm{min}}$ and z are highly correlated).

$$P_{\mathrm{final}}(z) = P_{\mathrm{pre}}(z) \times \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{z - z_{\mathrm{cut,eff}}}{\sigma_z}\right)\right)$$

$z_{\mathrm{cut,eff}}$ and σ_z could be gleaned from $y_{0,\mathrm{min}}, z$ 2D distribution

- ✓ physically motivated shape
- ✓ can be applied to zero-observed situations
- ✗ multi-stage, more room for user error
- ✗ what should shape of pre-cut distribution P_{pre} be?



Attempt #1: Exponential

Not getting $z_{\text{cut,eff}}$ or σ_z from data (yet).

- ✓ procedure relatively quick to implement
- ✓ seems to fit bulk of distribution well
- ✗ yield estimate of displaced vertices appears to be optimistic
- ✗ hard to expand complexity of fit function with so few events in tail

Future Work

Optimization Strategy

- Could optimize within each z bin (cuts are functions of z)

Exclusion Estimate

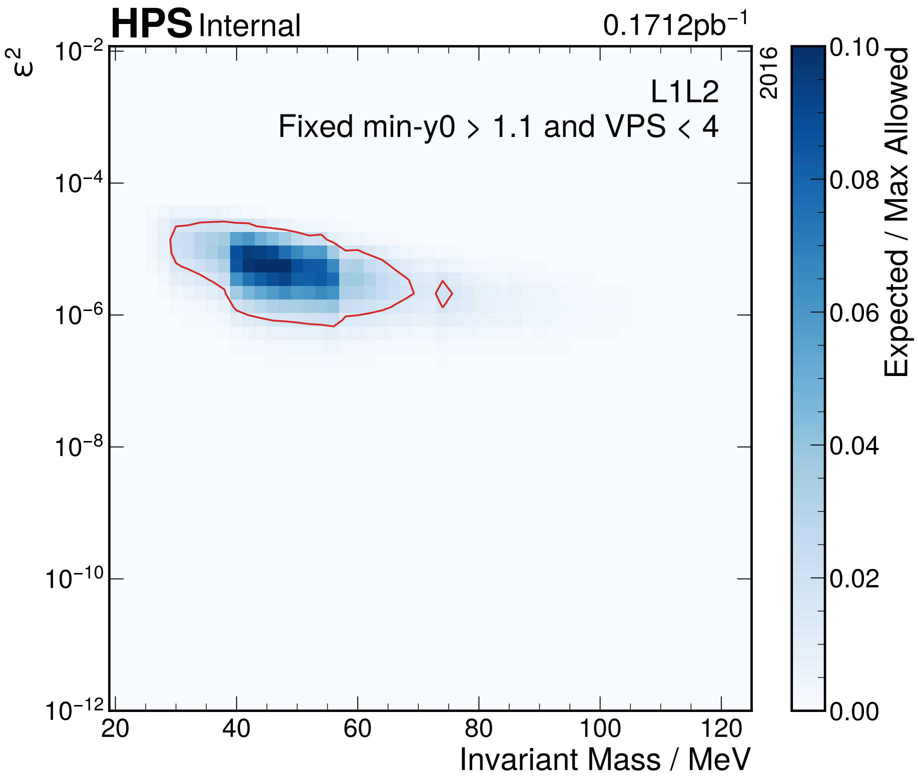
- WIP

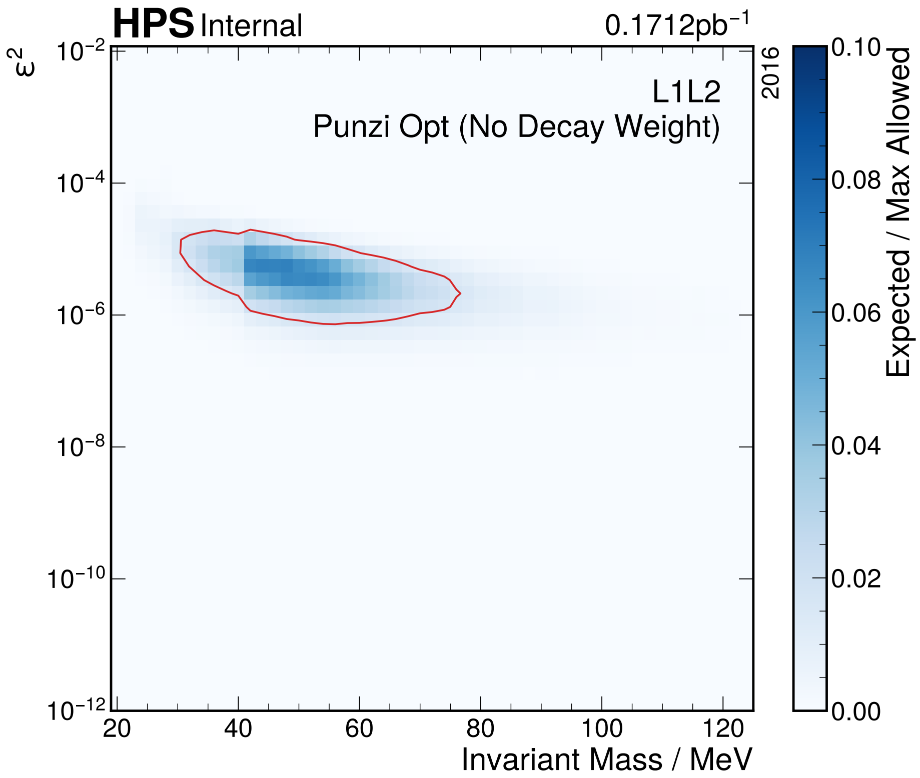
Additional Material

- Various distributions (mass vs z, z vs min-y0) as selections are made

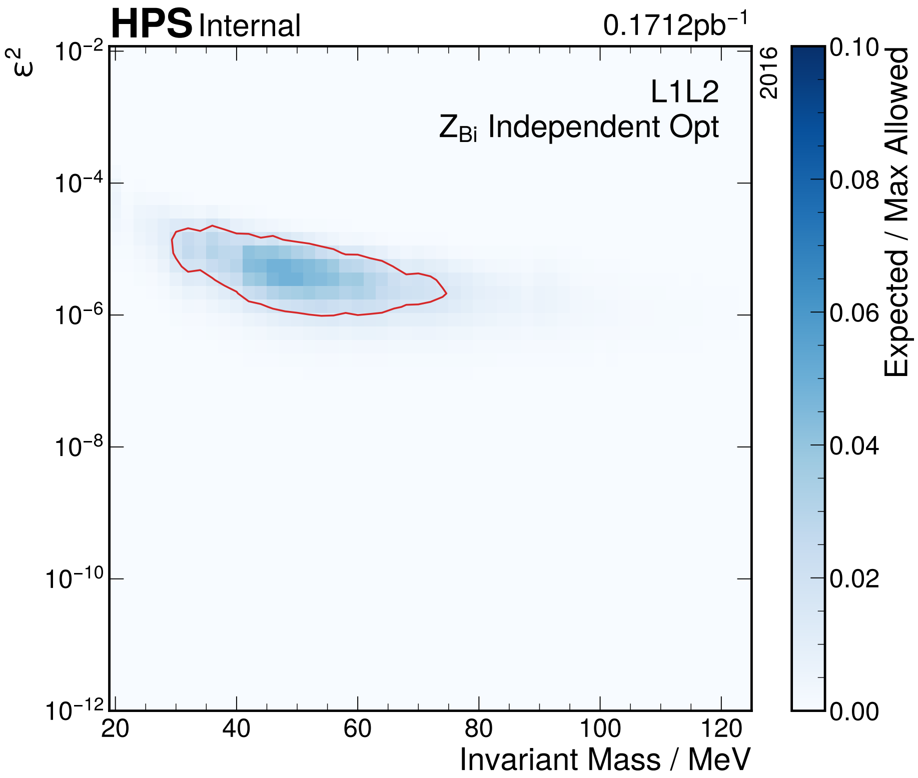
Questions

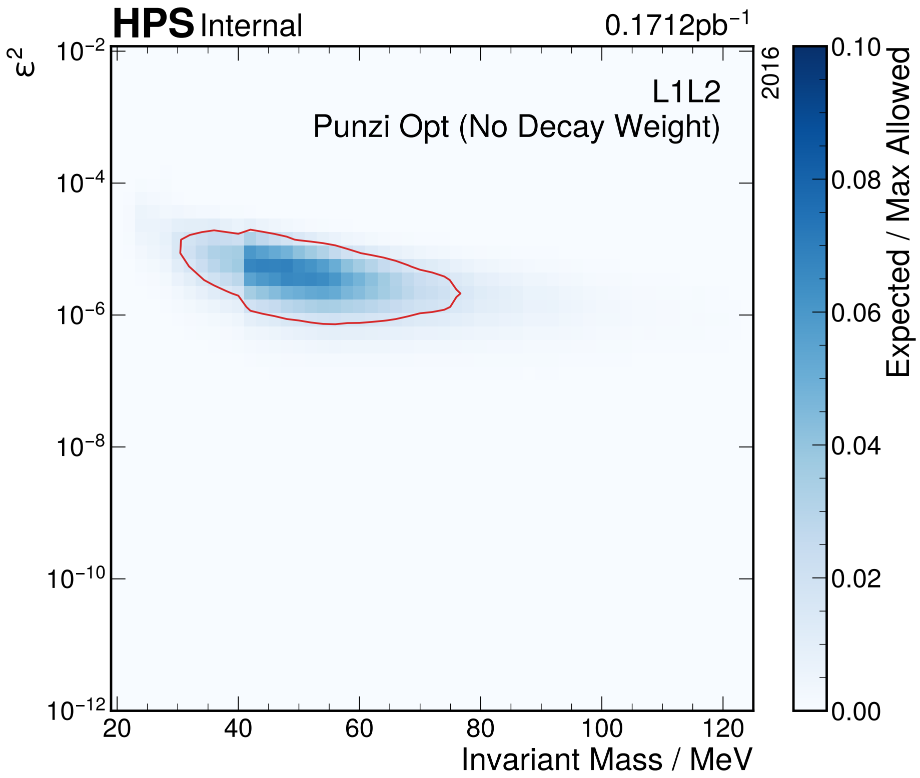
Comparison of Reaches



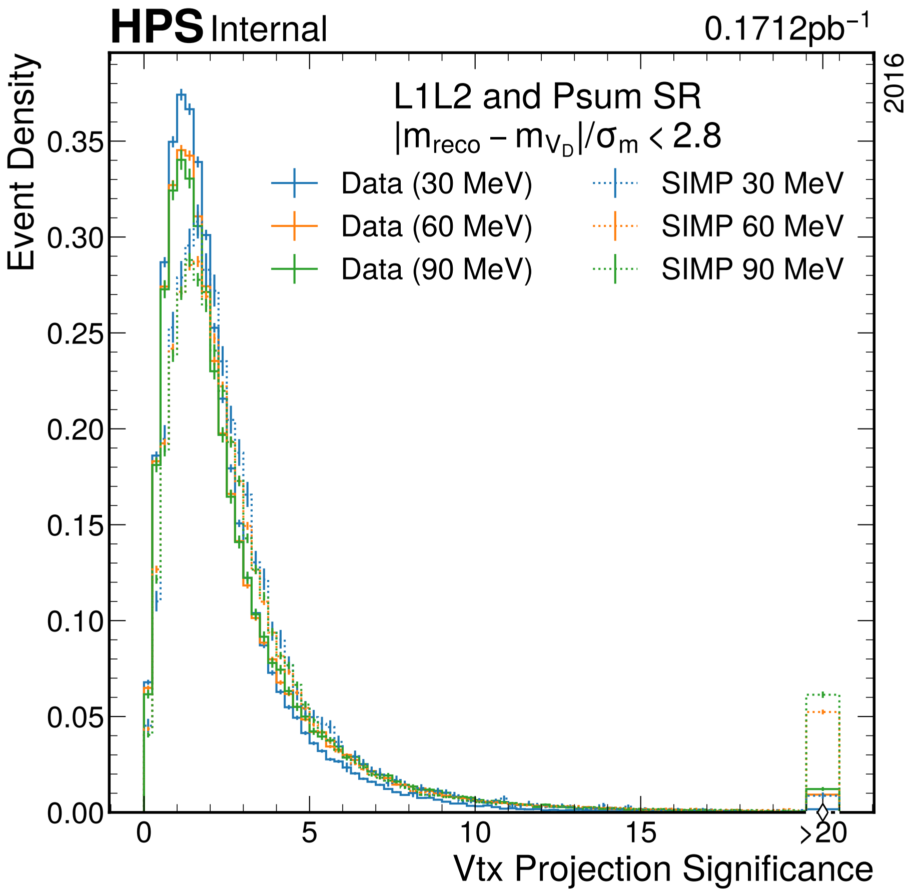


Comparison of Reaches



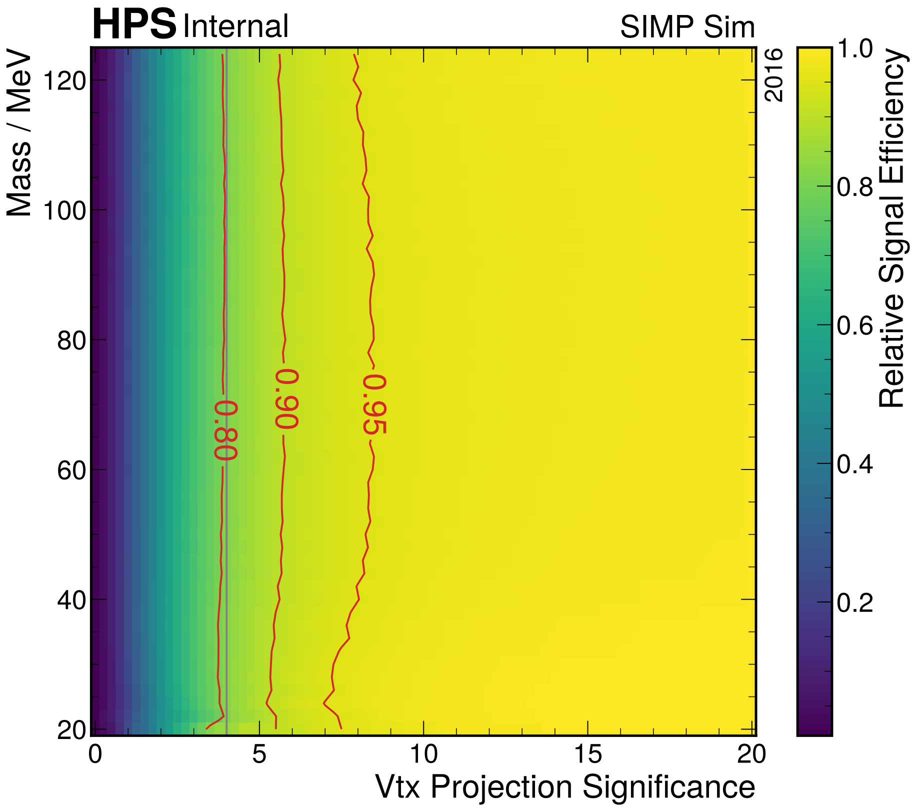


VPS Distribution



- Signal and Background shapes are similar
- Choose cut value by looking at relative signal efficiency (efficiency of only this cut)

Relative Signal Efficiency of VPS Cut



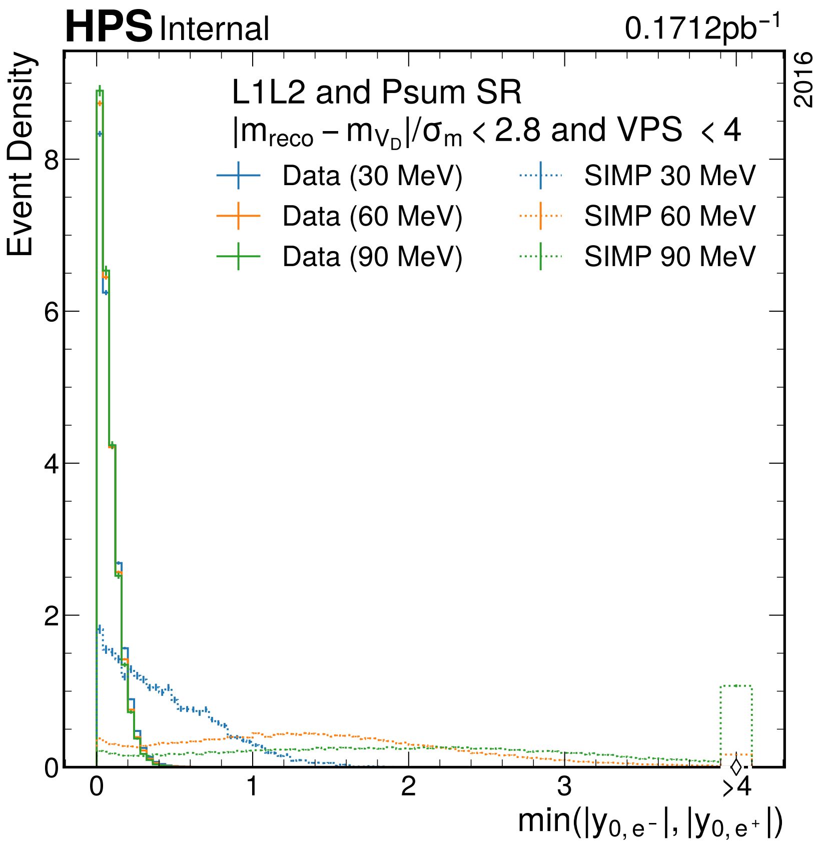
- Red lines are relative-efficiency contours
- Gray line is the cut value of 4 used for further study here

Too Tight?

The VPS$<$ 4 cut has a $\sim$ 80% relative signal efficiency which makes me think it is too tight; however, a cut value of $\sim$ 8 seems too loose as well.

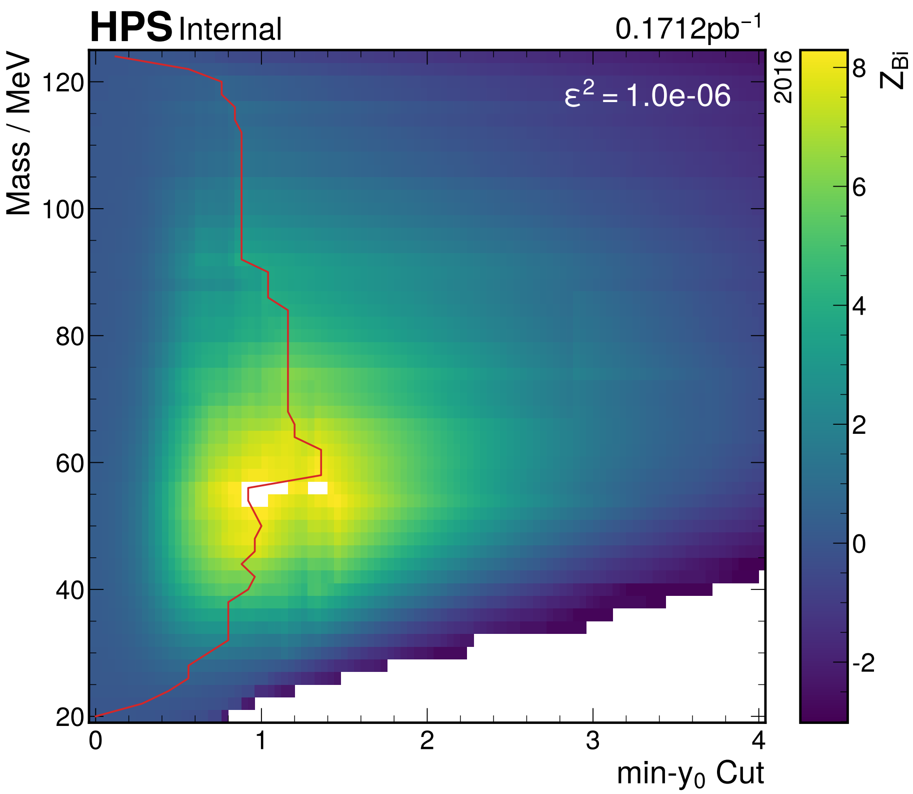
Remember The VPS is *roughly* the number of standard deviations a vertex's project to the target is from the beam spot at the target.

min-y_0 Distribution



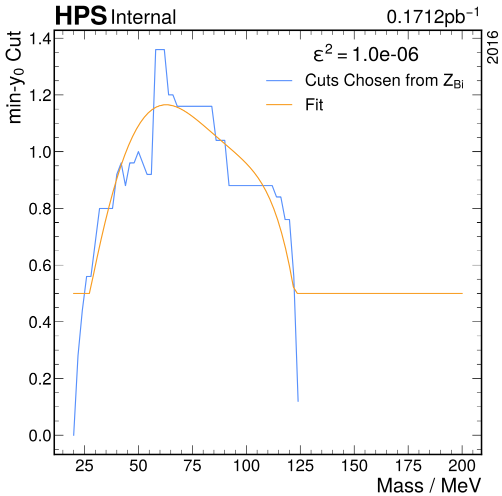
- Signal and Background shapes are different
- Choose cut value by maximizing binomial significance

Binomial Significance of min-y_0 Cut

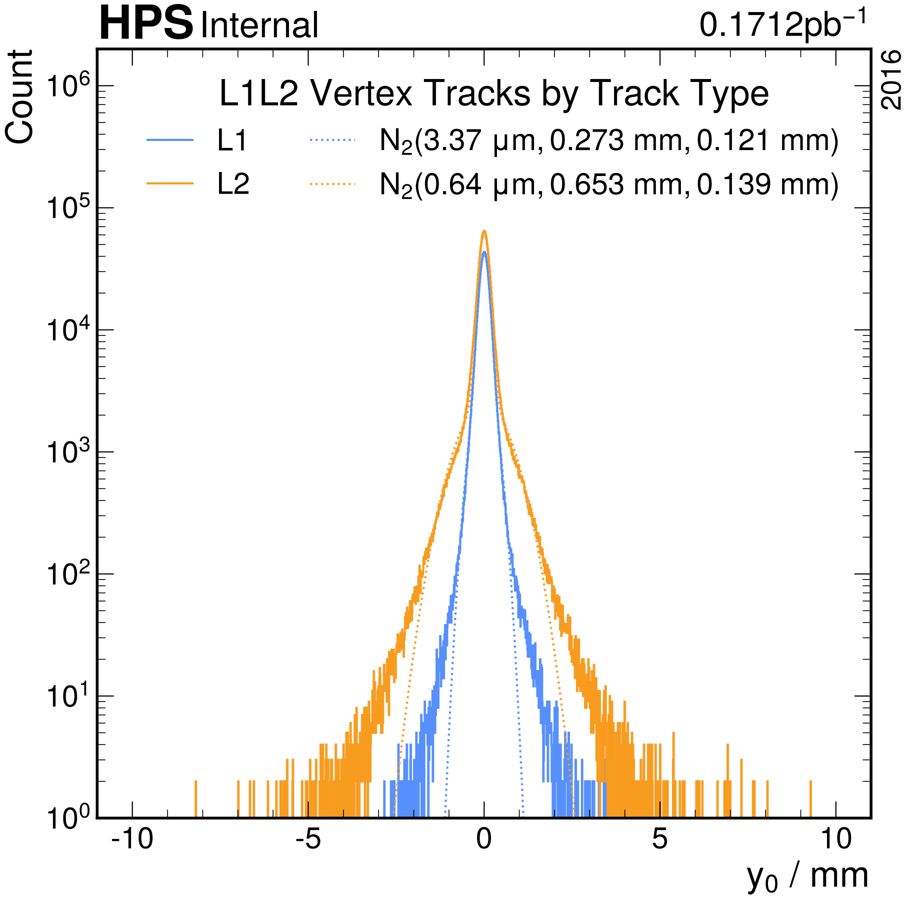


- Scaled signal by $1/\epsilon$ to get signal yield into an appropriate range
- Red line is chosen cut (obtaining maximum Z_{Bi} for that mass)

Fit Like Before



y_0 Distribution by Track Type



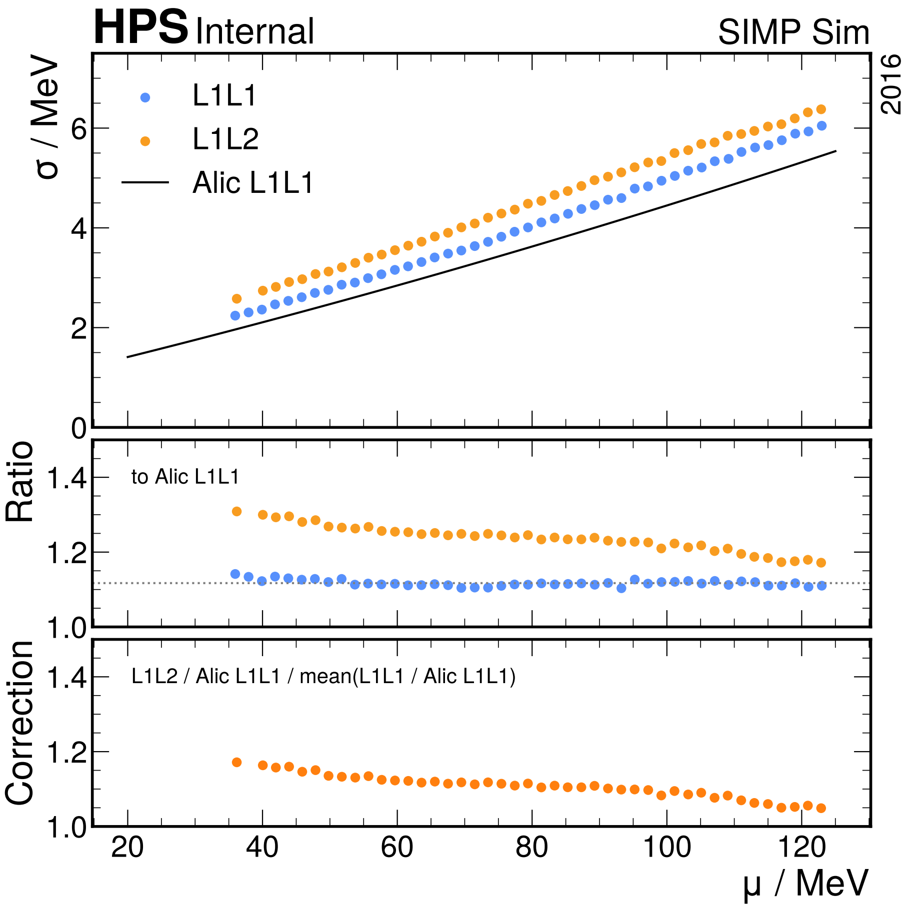
Selection

Pre-selected vertices and required to be L1L2 (i.e. tracks for L1L1 or L2L2 vertices are not included in this plot).

$\mathrm{N}_2(\mu, \sigma_1, \sigma_2)$ is the sum of two normal distributions that share the same mean μ.

- L2 tracks slightly broader than L1 tracks (expected)
- Both centered on 0 within 5 μm
- Width of core y_0 distribution is only ~ 18 μm larger for L2 tracks compared to L1 tracks (while the tail ends up being $\sim 3\times$ wider)

Mass Resolution Comparison



Selection

Pre-selected vertices – no truth matching done (hence the separation between L1L1 here and Alic's L1L1 in black)

- Seeing L1L2 widening in mass resolution by $< 20\%$ across mass range
- Separation shrinks as mass is increased

Not Used

The following reach comparisons use "Alic L1L1" in order to maintain comparability with prior estimates. Updating to an L1L2 mass resolution would not be difficult but is left for the future when the cuts and optimization strategy have been more established.

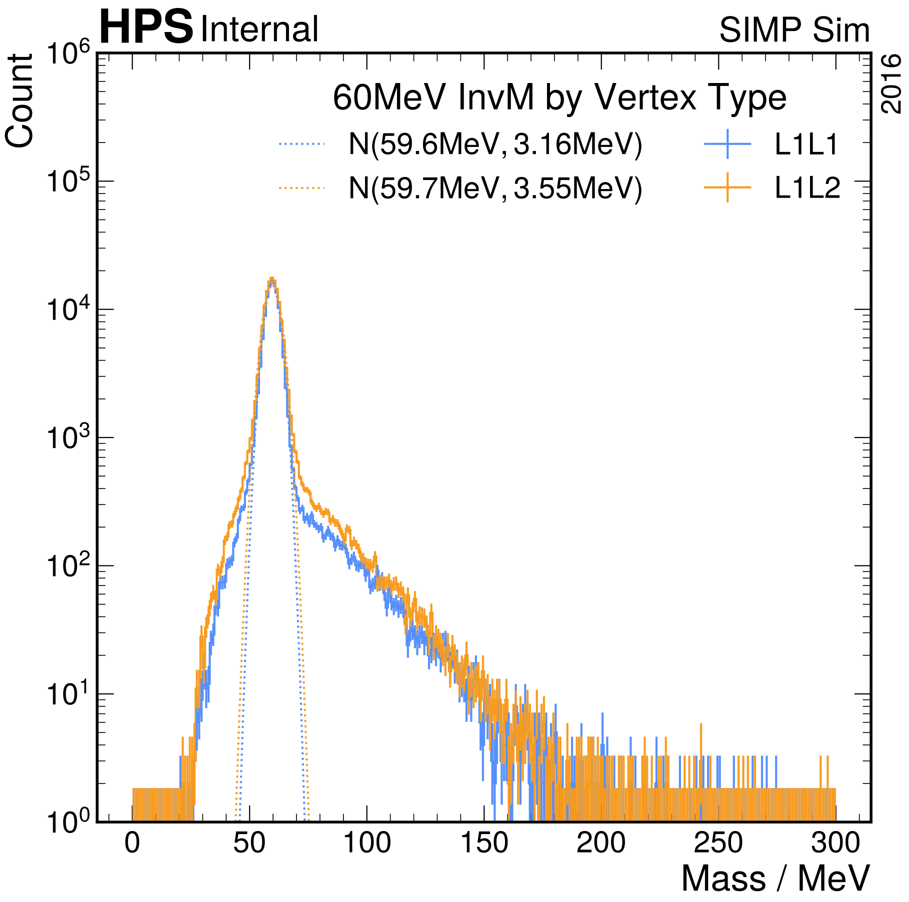
Sample Size Comparison

To help contextualize reach estimates, it is helpful to compare the luminosity of these two reconstruction categories.

Category	N_{7800}	Ratio to L1L1
L1L1	2216982	1.0
L1L2	1445740	0.65

Table: Luminosity comparison between the two reconstruction categories being studied. N_{7800} is the number of pre-selected vertices that correspond to the given reconstruction category. No other selections (for example, on Psum) were made.

Mass Resolution Calculation

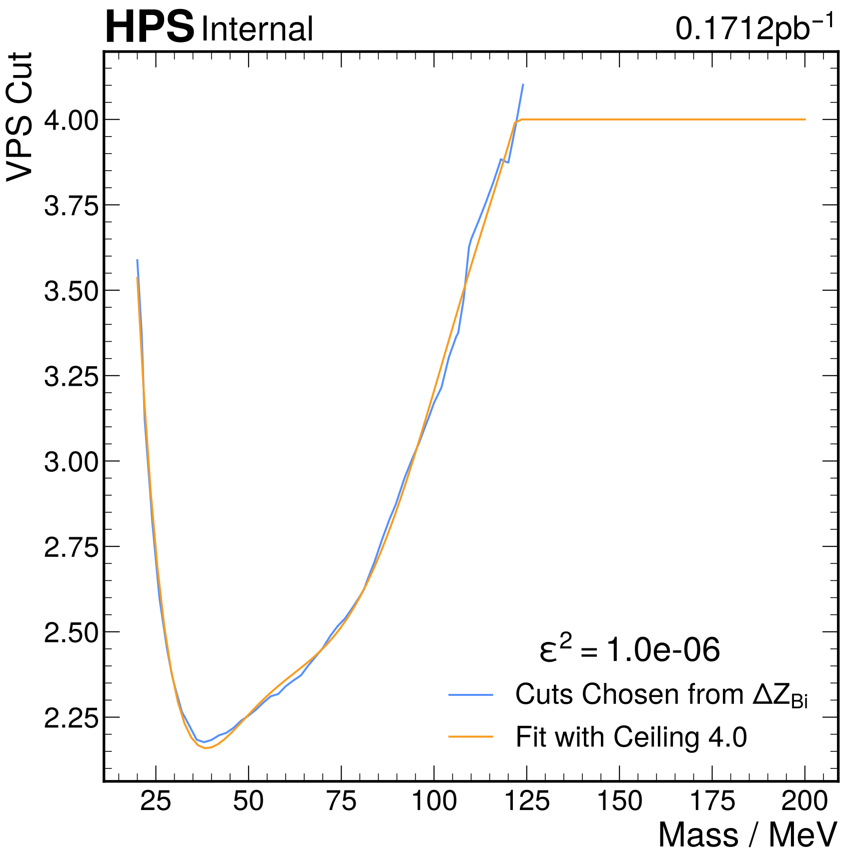


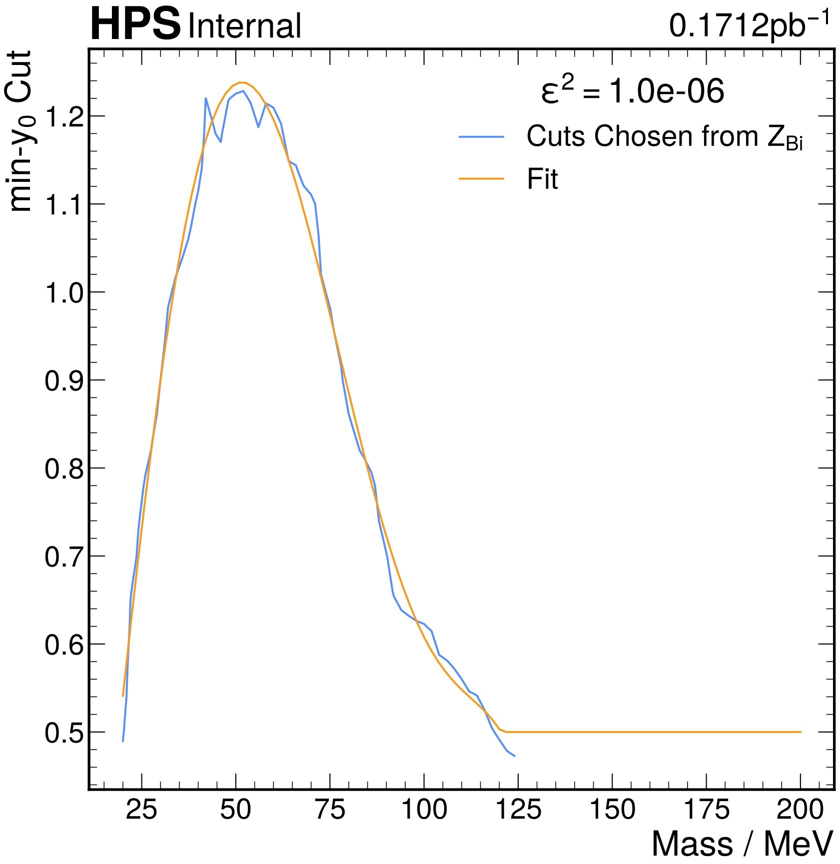
Selection

Pre-selected vertices – no truth matching done (hence the separation between L1L1 here and Alic's L1L1 in black)

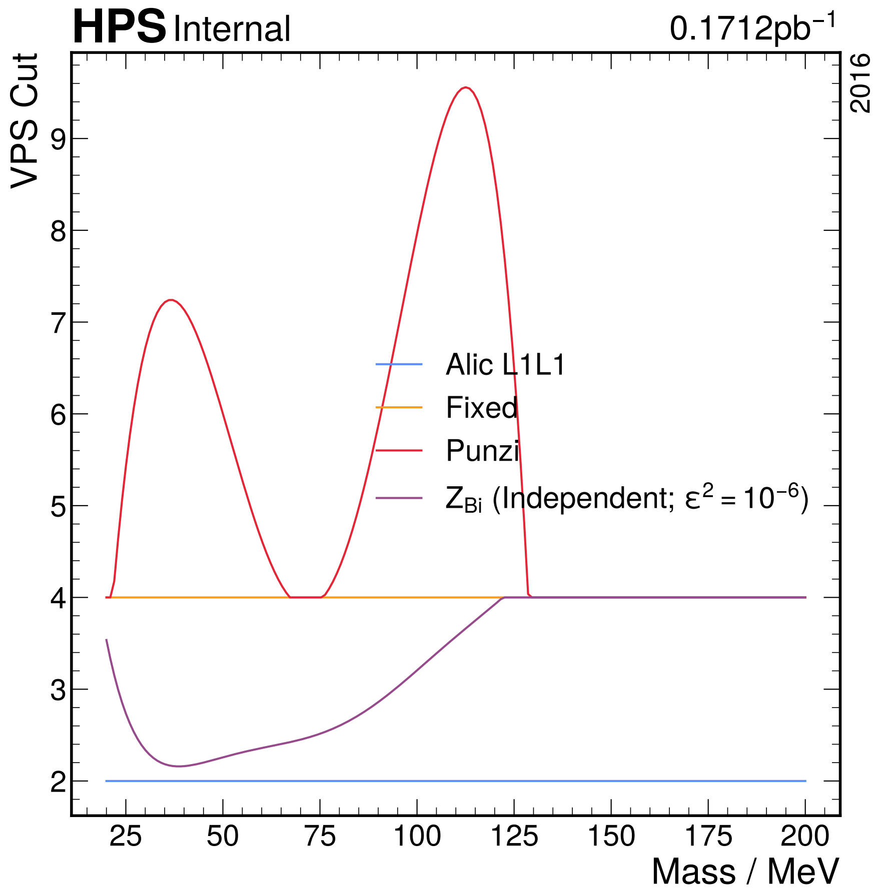
- Select core of distribution by calcuating mean of histogram and dropping bins further than 3σ away from mean
- Fit this core with a normal distribution to obtain μ and σ

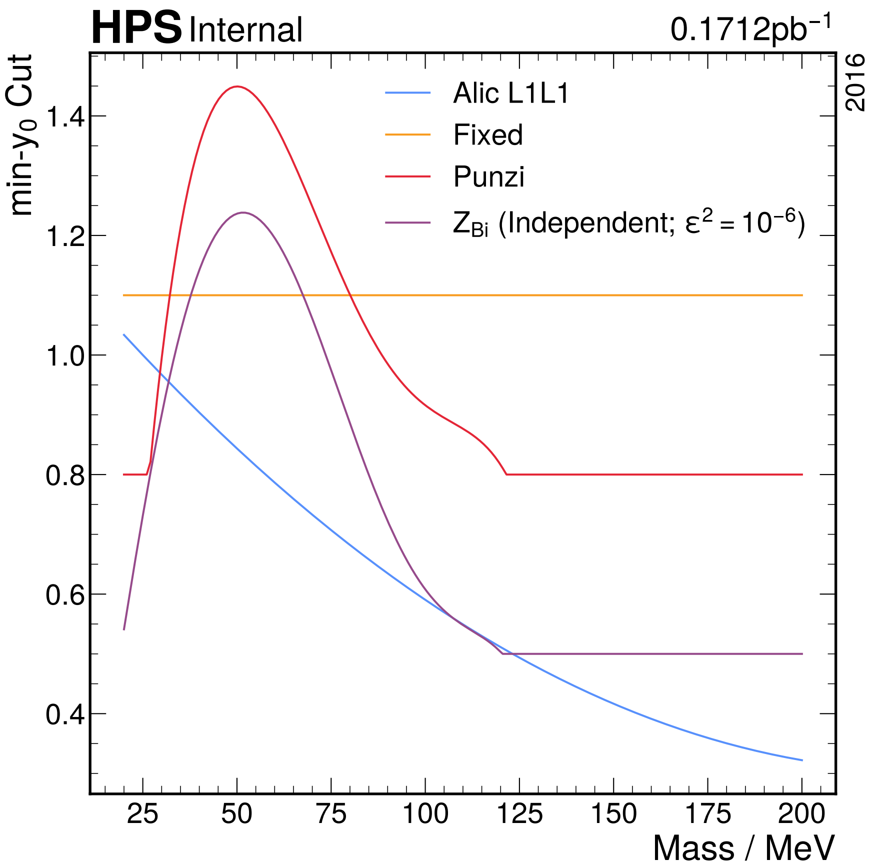
Z_{Bi} Cut Values and Fits



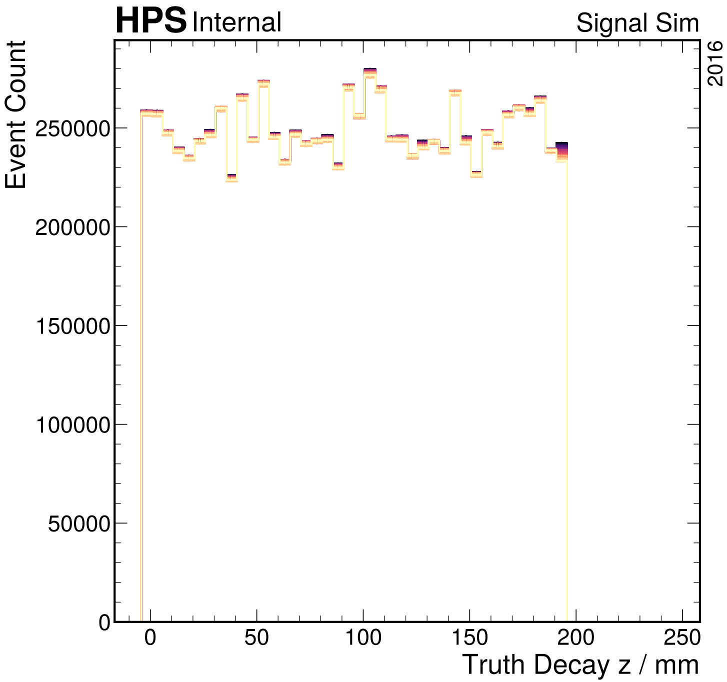


Cut Comparisons





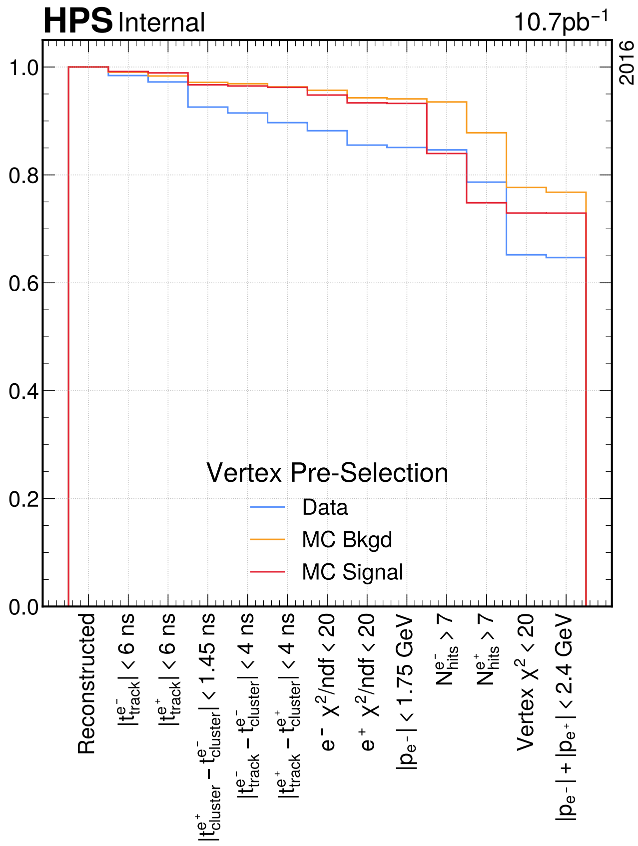
Displacement of SIMP Signal Events



One of the first things I did to make sure samples didn't need to be created.

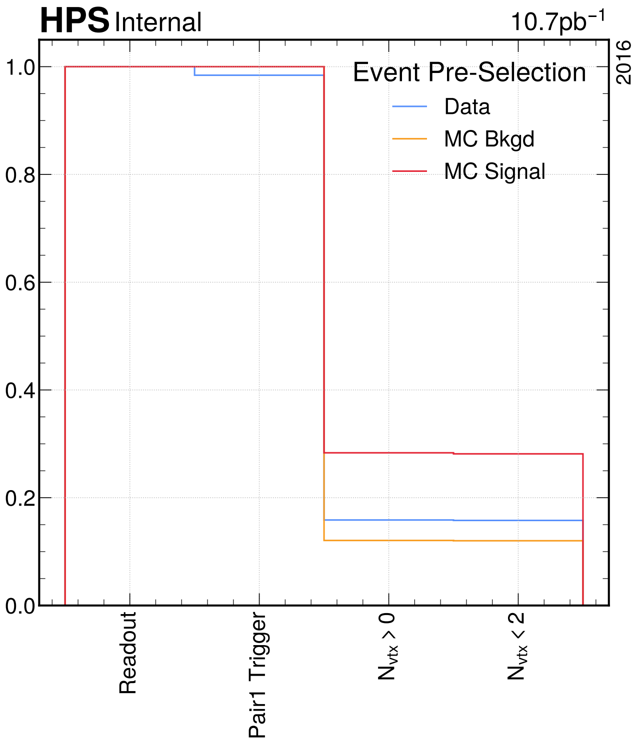
- Different colors correspond to different mass points
- Truth-level decay vertices sampled out until ~ 200 mm
- Close to the same z position as L1, so I think these samples can be faithfully used to study the L1L2 selection
- Idential distribution across colors makes me think that the random seed determining the decay length was not changed, but I think that is okay since we normalize by this distribution during exclusion estimates anyways

Pre-Selection on Vertices



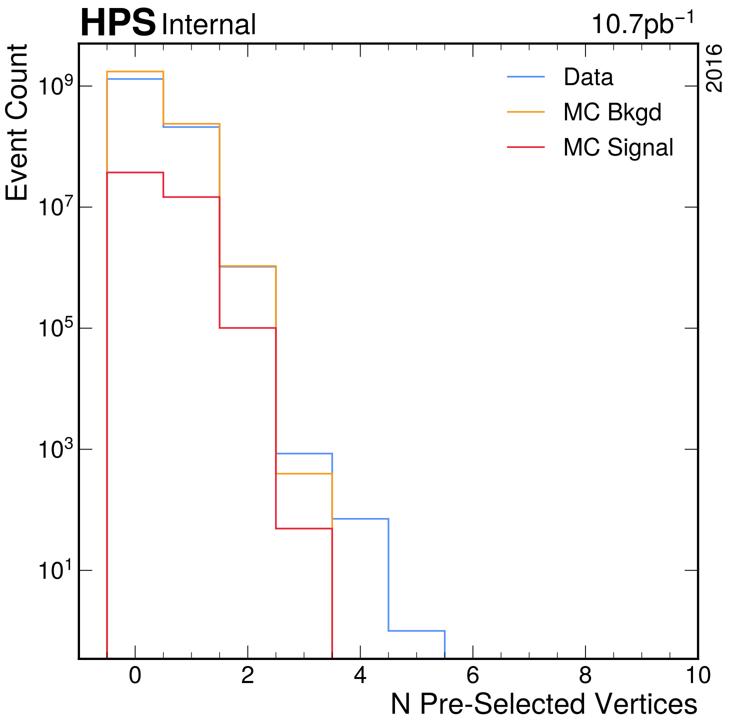
- Same Pre-Selection on vertices as developed and validated by Alic
- Seeing same efficiencies as documented within Alic's SIMP (L1L1) note

Pre-Selection on Events



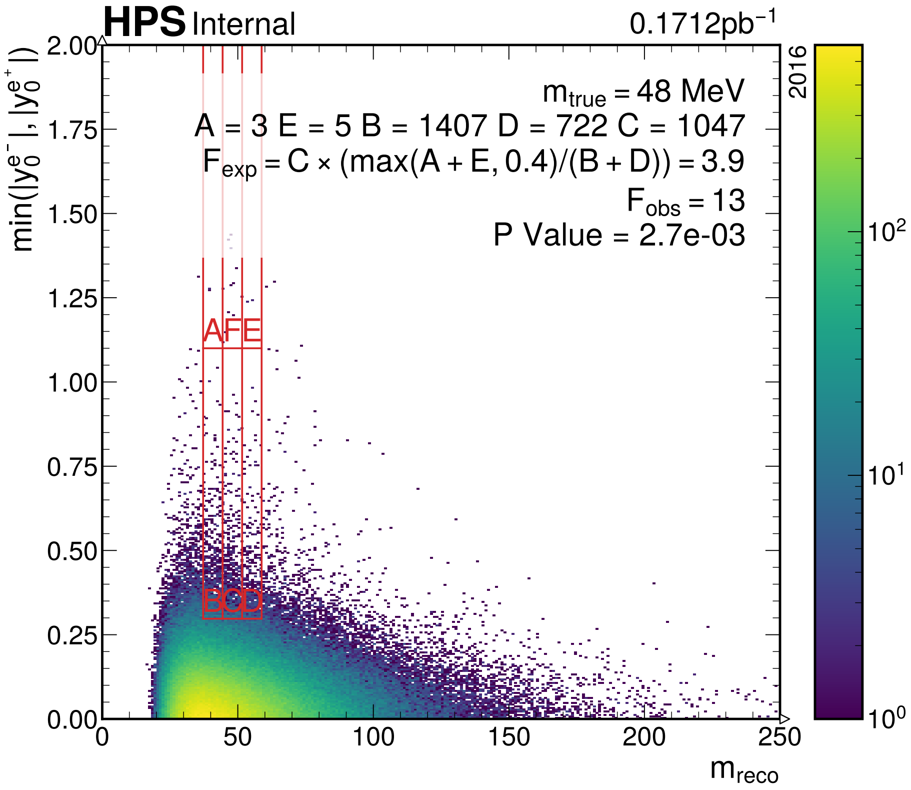
- Similar to first stage of Alic's event selection, although dropping reconstruction category requirement
- Largest effect is requiring at least one pre-selected vertex

Number of Pre-Selected Vertices per Event



Event Pre-Selection basically amounts to choosing the events falling into the $N = 1$ bin.
Data is the only sample which has the additional requirement of the Pair1Trigger which has a small effect.

Search Method



1. Fill histogram with data
2. Set mass edges at 1.5σ and 4.5σ (values optimized by Alic)
3. Set upper min-y_0 edge at cut value
4. Lower other min-y_0 edge (a.k.a. y_0 "floor") from the cut value until there are at least 1k events in region C
5. Calculate expected number of events in F and compare to observed number of events
6. Estimate p-value by throwing toy experiments in A+E (Poission), B+D and C (Normal) and re-calculating F from these toys

Z_{Bi} Optimization

Replicating optimization procedure as developed by Alic – a one-tailed binomial p-value test (with the p-value converted to a significance).

$$Z_{\mathrm{Bi}} = \sqrt{2}\mathrm{erf}^{-1}(1 - 2p_{\mathrm{Bi}})$$

where

$$\begin{aligned} p_{\mathrm{Bi}} &= \sum_{j=n_{\mathrm{on}}}^{n_{\mathrm{tot}}} P(j|n_{\mathrm{tot}}; 1/(1+\tau)) \approx B(1/(1+\tau), n_{\mathrm{on}}, 1+n_{\mathrm{off}})/B(n_{\mathrm{on}}, 1+n_{\mathrm{off}}) \\ &= B_{\mathrm{reg}}(1/(1+\tau), n_{\mathrm{on}}, 1+n_{\mathrm{off}}) \end{aligned}$$

where $B(a, b)$ is the complete beta function, $B(x, a, b)$ is the incomplete beta function, and $B_{\mathrm{reg}}(x, a, b)$ is the regularized incomplete beta function. Used $\tau = 1$, $n_{\mathrm{on}} = S + B$, and $n_{\mathrm{off}} = B$ for this evaluation, so in summary the FoM is

$$Z_{\mathrm{Bi}} = \sqrt{2}\mathrm{erf}^{-1}\left(1 - 2B_{\mathrm{reg}}(0.5, S+B, 1+B)\right)$$

Punzi Figure of Merit

Basis

Proposed in ▸ PHYSTAT2003 by Giovanni Punzi where a FoM is designed to be maximized while improving *both* search and exclusion potential.

$$f_{\text{punzi}} = \frac{E}{\frac{a}{2} + \sqrt{B}}$$

where E is the signal efficiency, B is the background yield, and a is the desired confidence level of search or exclusion (in number of σ, currently using 3).

Two main benefits (from my perspective)

- Does not diverge as $B \to 0$
- Does not require knowledge of absolute rate of signal

Experiments with Decay Weighting

First idea is to simply define a new FoM that includes the decay weighting function.

$$f_{\mathrm{DW}} = \frac{1}{\frac{a}{2} + \sqrt{B(t)}} \int_{z_{\mathrm{target}}}^{\infty} D(z)E(z,t)dz$$

where

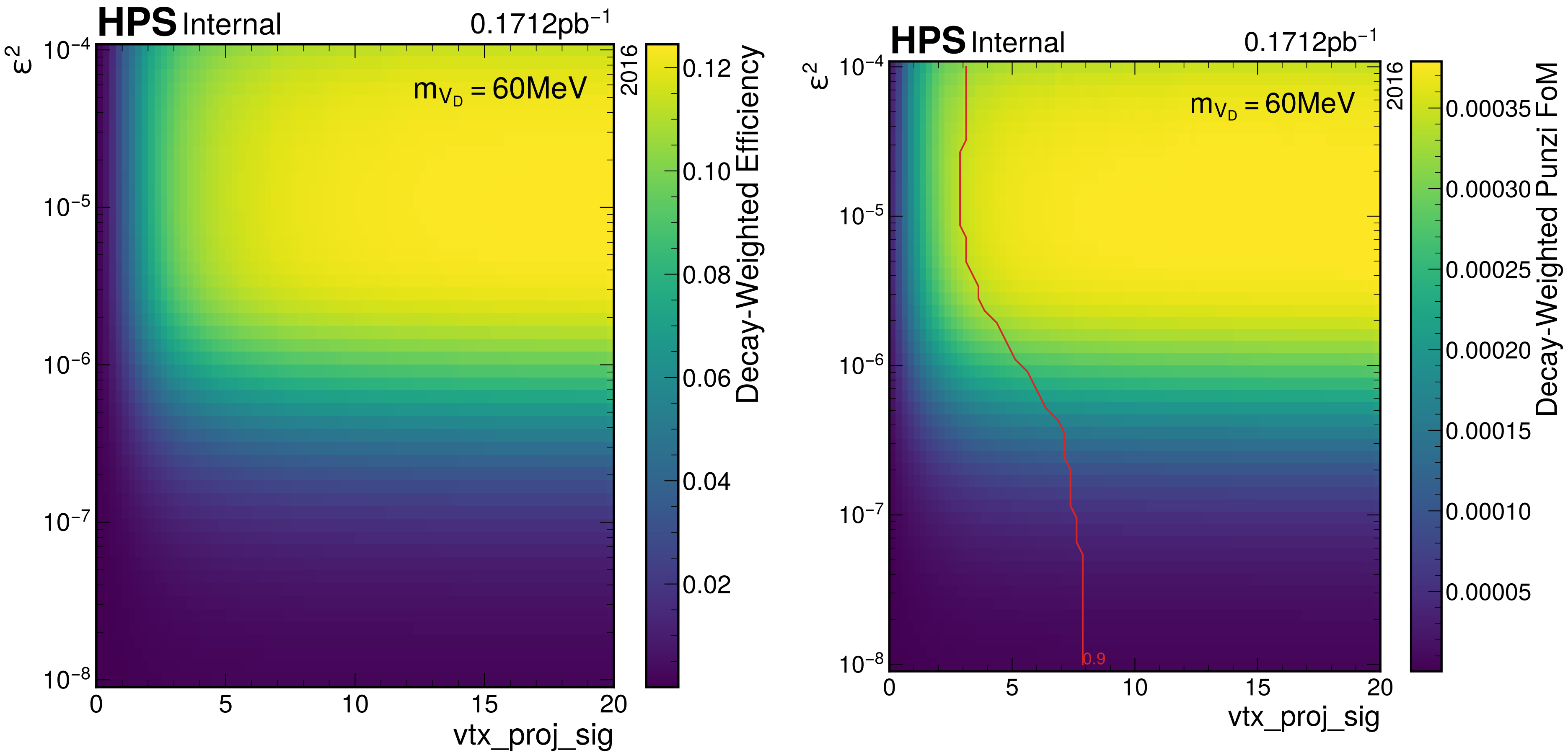
$$D(z) = \sum_{V \in \{\rho_D, \phi_D\}} \mathrm{BR}(A' \to V\pi_D) \frac{\exp((z_{\mathrm{target}} - z)/(\gamma c \tau_V))}{\gamma c \tau_V}$$

This becomes equivalent to f_{punzi} in the $\epsilon \to 0$ limit where $D(z)$ becomes flat and the events are equally weighted along z. Calling this "Decay-Weighted Punzi FoM" and the integral "Decay-Weighted Efficiency".

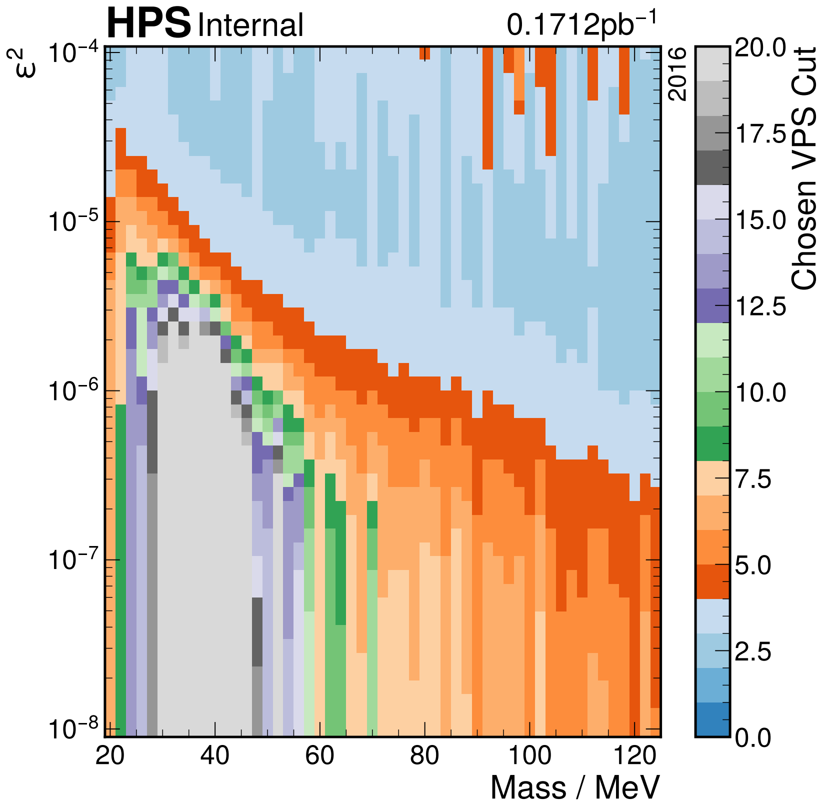
Issues

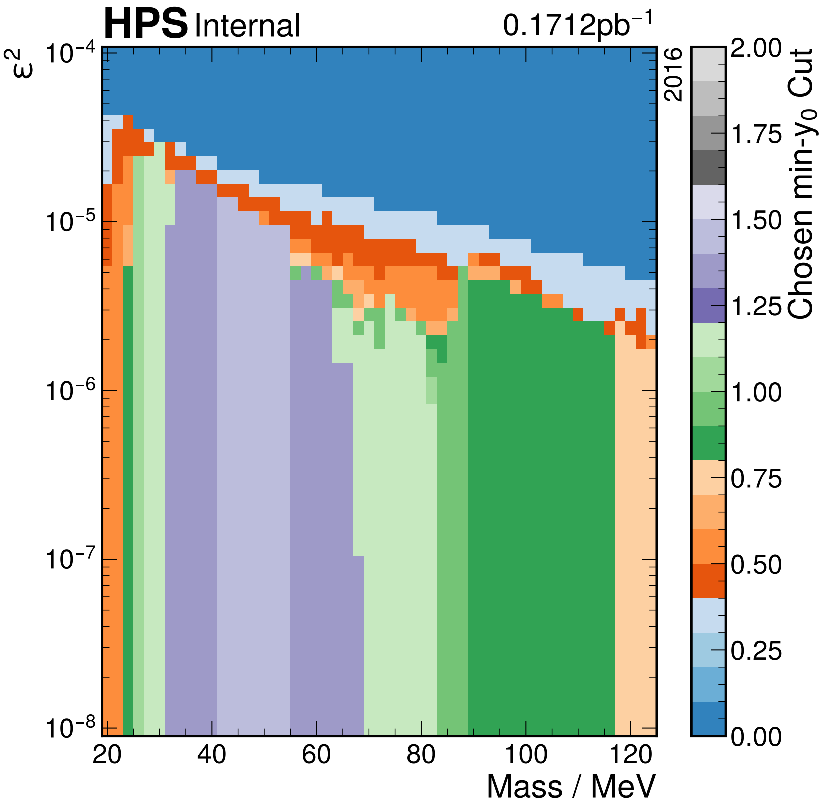
- May have to choose an ϵ^2 value to optimize for
- Pure maximum is often attained by removing the cut $\to$ look for where the FoM "flattens" out (i.e. tigtening the cut does not improve the FoM much anymore) $\to$ choose cut that is the tightest cut getting to 90% of the maximum

Example Decay-Weighted Punzi Calculation



Cut Choices by ϵ^2 and m_{V_D}





Chose $\mathrm{VPS} < 4$ and then applied it to optimize min-y_0.