2016 Vertexing Analysis Update

Matt Solt
SLAC National Accelerator Laboratory
May 30, 2019

Introduction

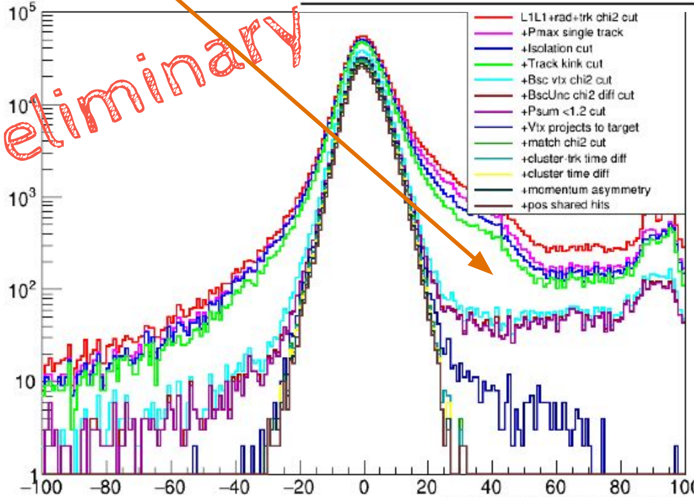
- 2015 Vertex Analysis - summary of 2015 procedure, results, and lessons learned
  - 2015 produced 2 thesis (Sho + Holly), and then some additional work
  - Presented at ICHEP (with proceedings), HPS Analysis Note
- How can we improve future vertex analysis?
- What are the next big steps, where should we focus our energy, and why should we be excited for 2016 analysis?
- Reported in this talk for 2016 analysis - data quality, MC needs, new approaches to old problems, signal yield estimates, and SIMPs and other models

Event Selection

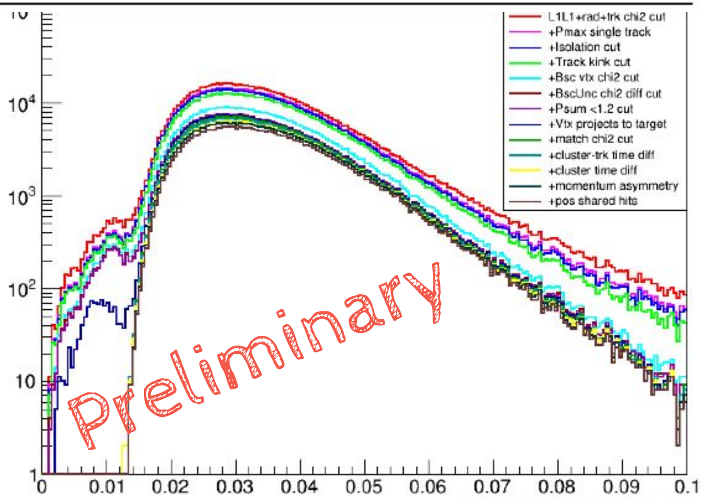
Cut type	Cut	Cut Value	%cut	%cut core	%cut tails
track	Fit quality	track $\chi^2/dof < 6$	24	–	–
track	Max track momentum	$P_{trk} < 75\%E_{beam}$	11	9	20
track	Isolation		4	1	14
track	kinks in L1 and L2	9	7	16	
vertex	beamspot constraint	bsc$\chi^2 < 10$	28	22	57
vertex	beamspot - unconstrained	bscχ^2-unc$\chi^2 < 5$	15	15	15
vertex	maximum P_{sum}	$< 115\%E_{beam}$	0.5	0.5	0.8
vertex	vertex projects to target	elliptical $3\sigma_{x,y}$	2	1	16
ecal	Ecal SVT matching	$\chi^2 < 10$	5	4	10
ecal	track Ecal timing	< 4ns	4	4	5
ecal	2 cluster time diff	< 2ns	6	6	9
physics	momentum asymmetry	< 0.5	3	3	5
event	max shared hits in e^+ track	< 5 shared hits	9	9	10

Goal is to reduce/eliminate backgrounds at large z

Tracking and vertexing work very well!

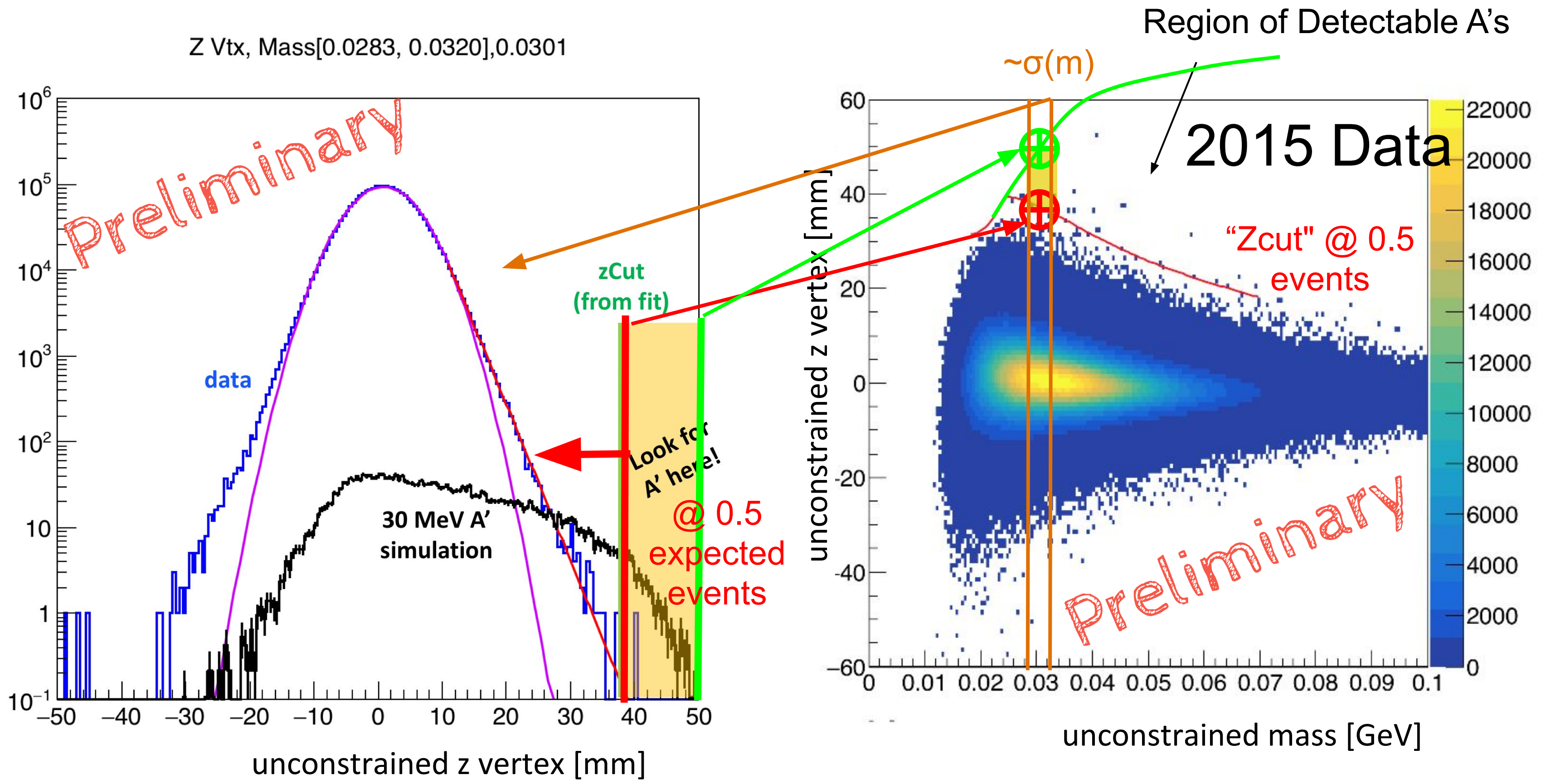


unconstrained z vertex [mm]



unconstrained mass [GeV]

Vertex Analysis Final Selection

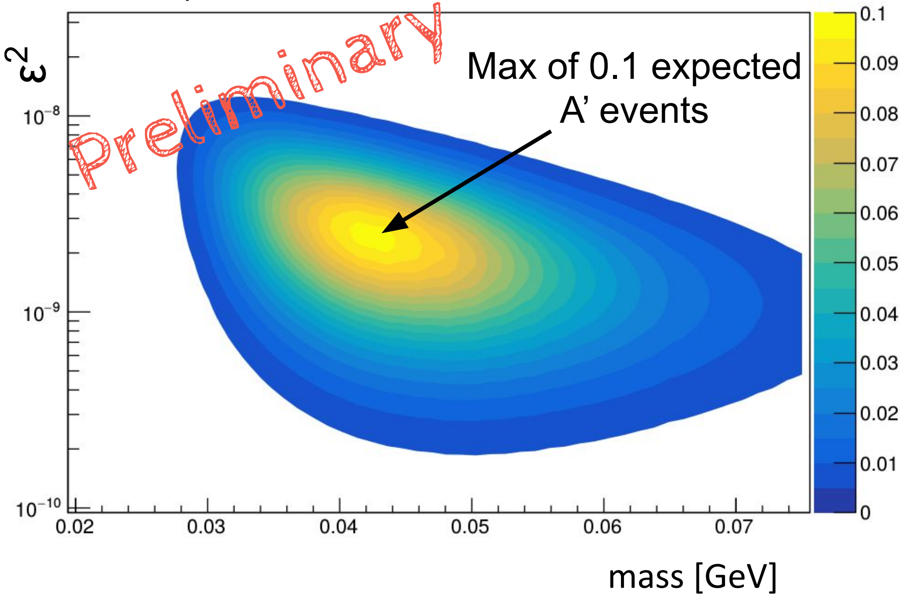


Vertex Analysis Final Results

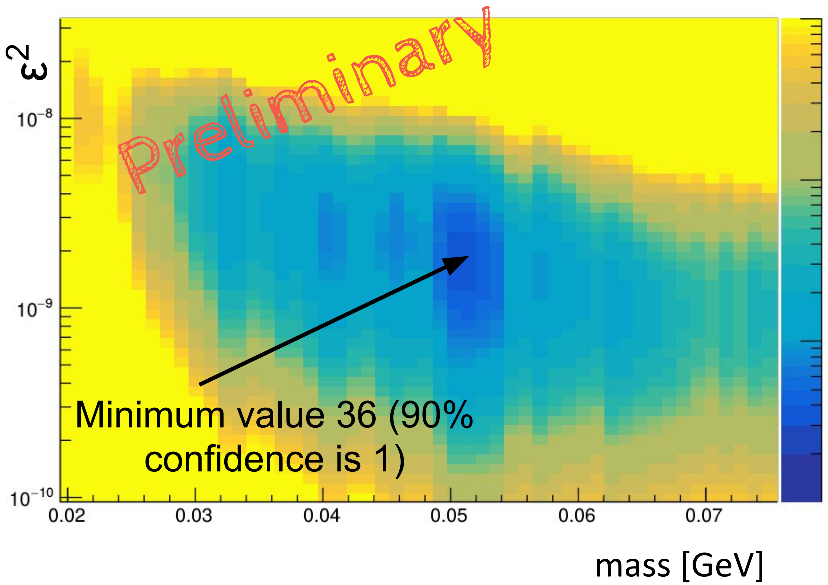
- Optimum Interval Method is ideally used for small signal where signal shapes are known, but background is not sufficiently known (HPS, direct DM detection, etc.)

arXiv:physics/0203002v2

Expected A' Events Detected Past ZCut



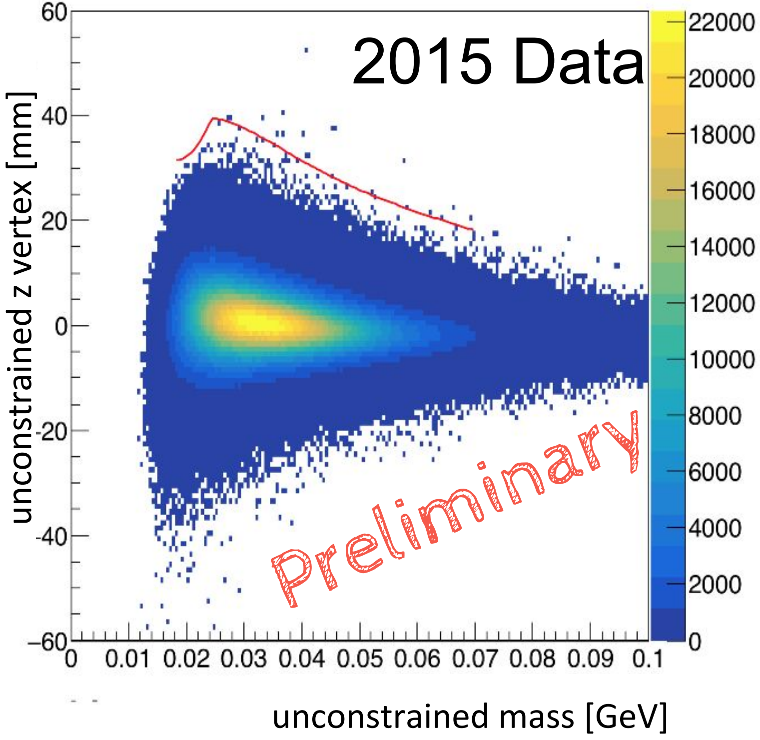
Optimum Interval Method Limit

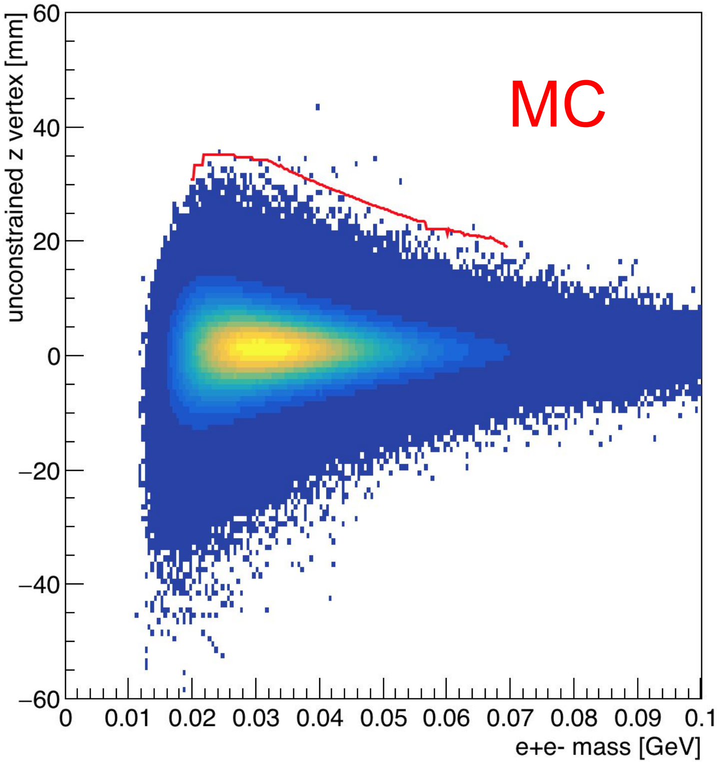


Lesson Learned from 2015 Vertexing (1)

SLAC

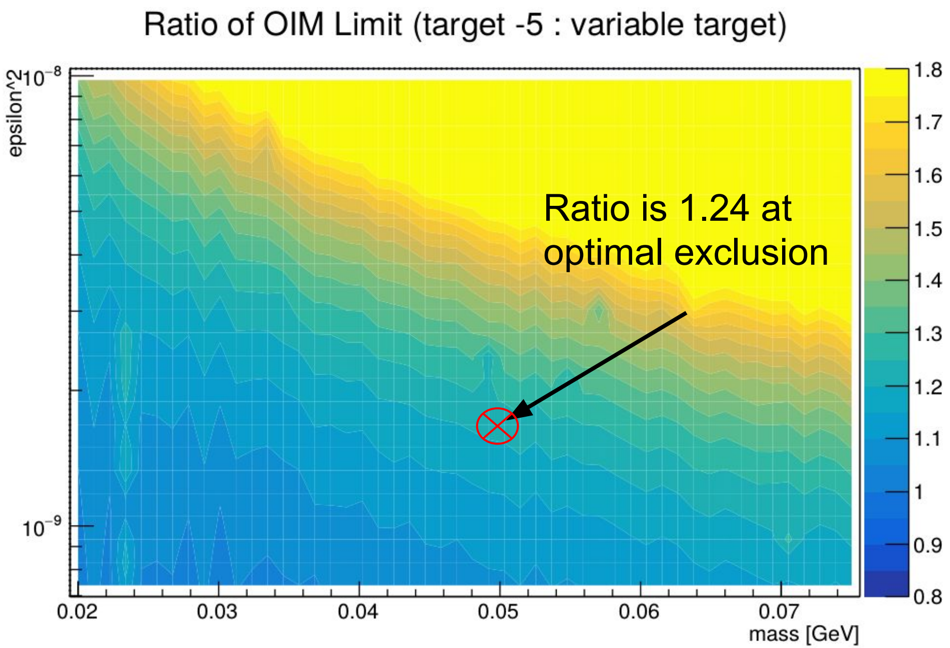
- Global alignment put the target in the wrong place; mass-dependent target

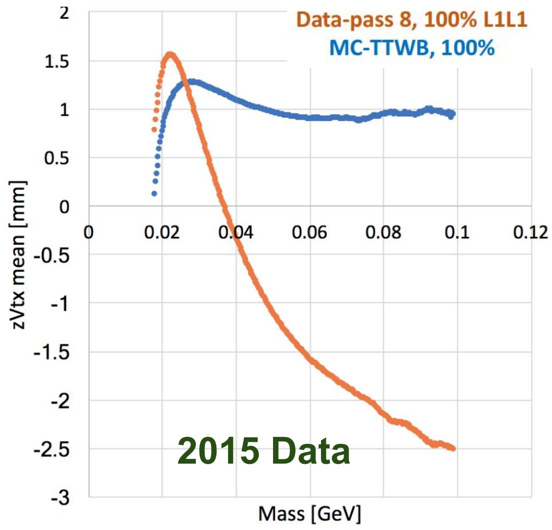




Lesson Learned from 2015 Vertexing (1)

- Resulted in ~5mm discrepancy in target position.
- This is a very large systematic (24% at optimum parameters)



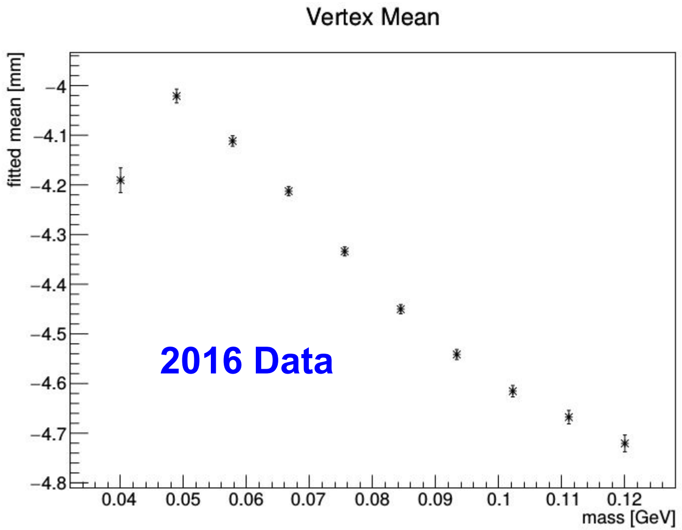


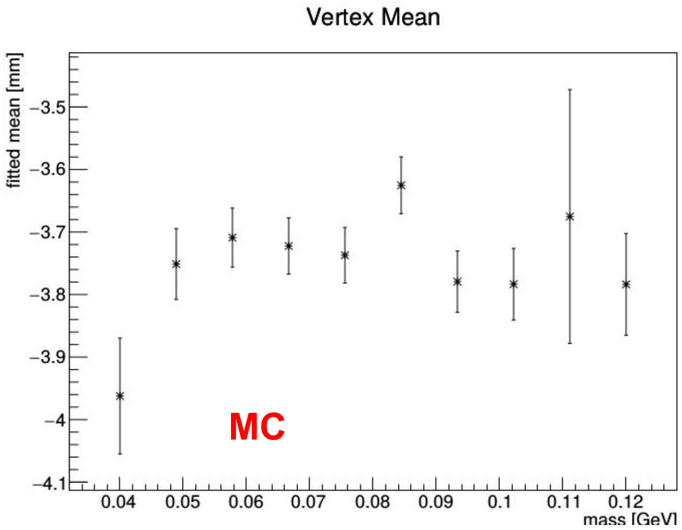
The mean was flat for tweak pass 6 (moved target to -5 mm to fix this effect)

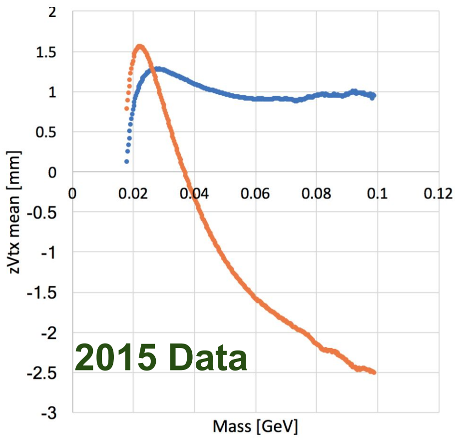
Target Position in 2016

SLAC

- Improving the global alignment procedure in 2016 mitigated these issues
- Target is now at -4.3 mm, consistent with tweakPass6 in 2015
- We still observe some mass dependent target, but the effect is far less





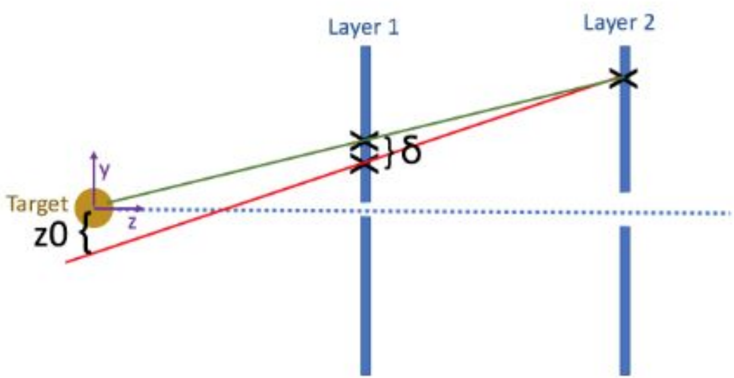


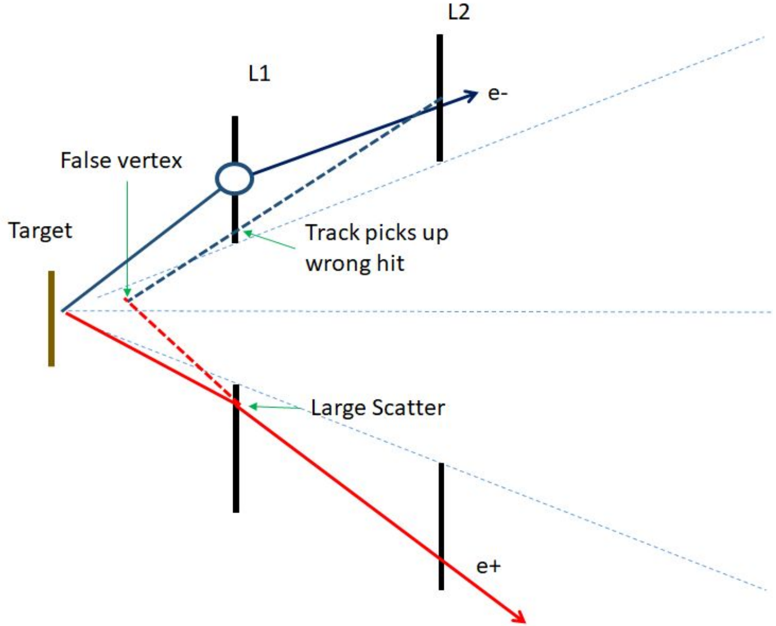
The mean was flat for tweak pass 6 (moved target to -5 mm to fix this effect)

Lesson Learned from 2015 Vertexing (2)

- Bad tracks that pick up a wrong hit in Layer 1; creates a downstream vertex
- We have an isolation cut that gets rid of most of these high z events
- However, this doesn't account for scattering, and some high z backgrounds are left

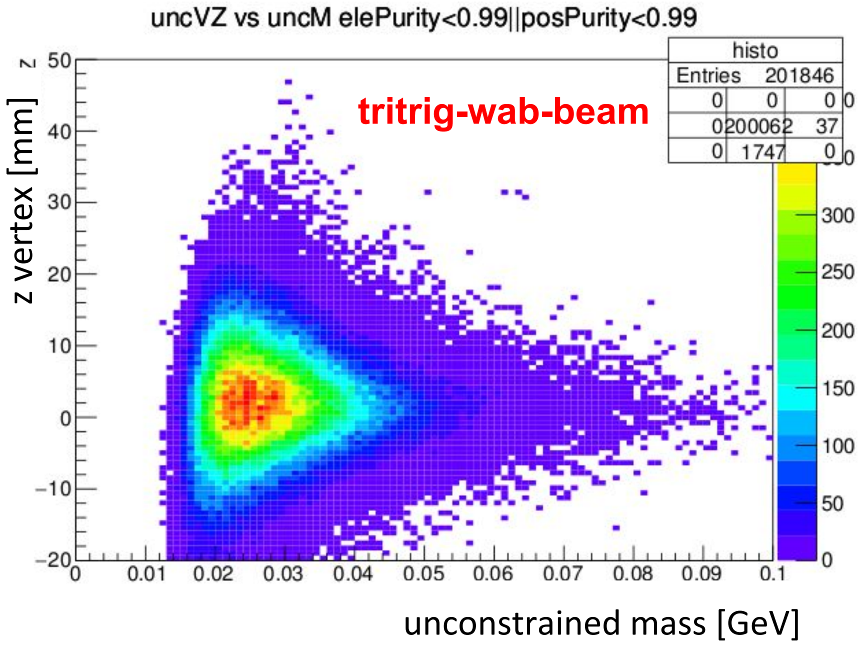
$$2\delta + z0 \times \mathrm{sign}(P_y) > 0$$

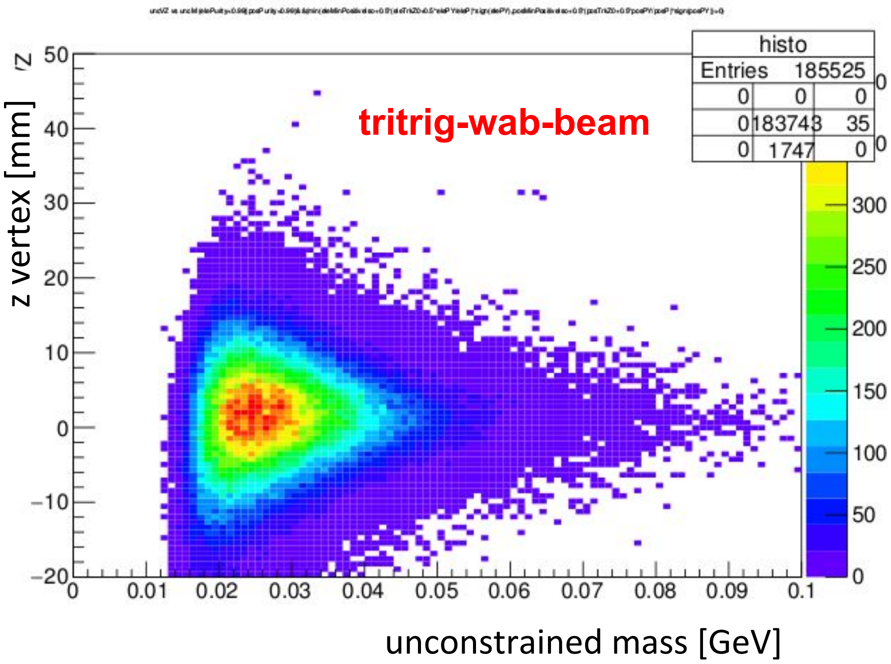




Lesson Learned from 2015 Vertexing (2)

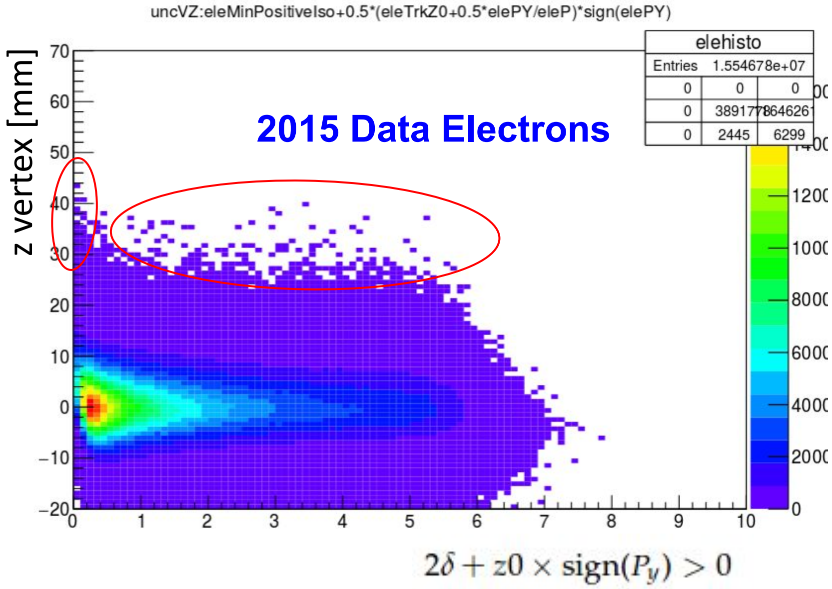
- Left: V0s with track that picks up the wrong hit
- Right: Same as left but WITH current iso cut (i.e. events we need to work harder to get rid of)

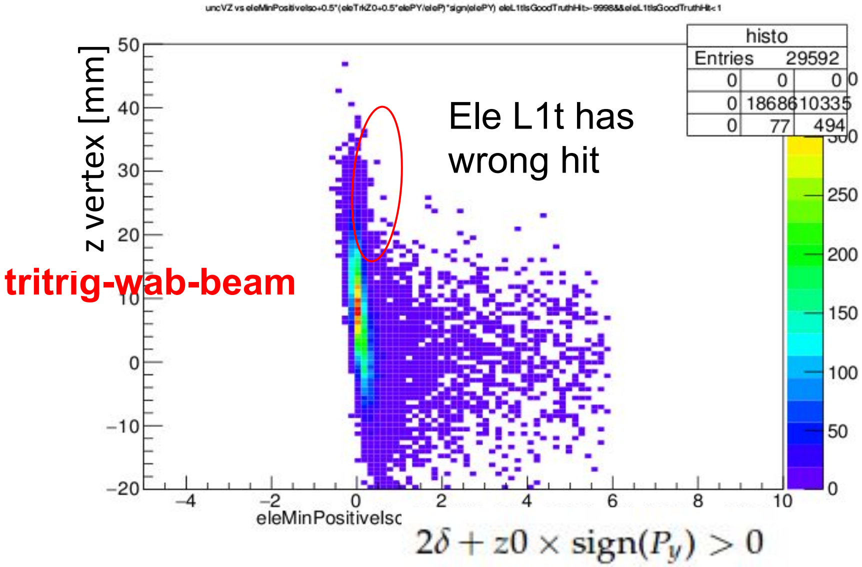




Isolation Cut Data

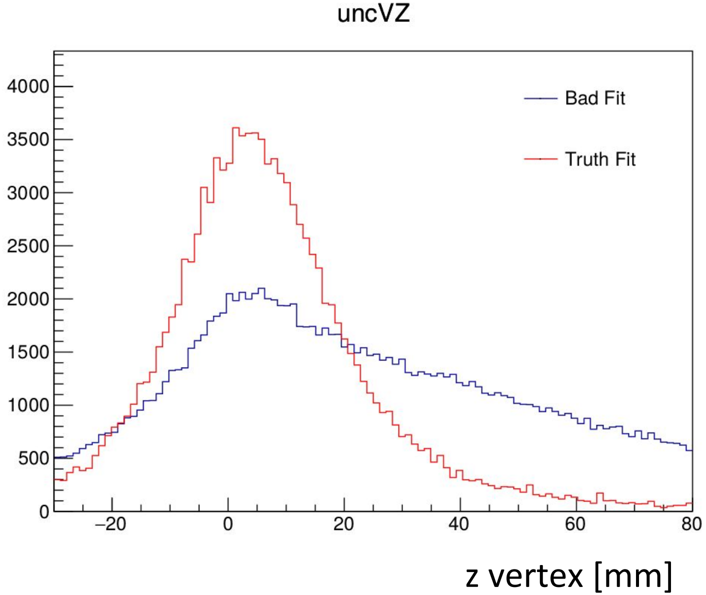
- Data shows similar structure with high z events near cut values
- Can we tighten up isolation cut? What is the effect on signal?

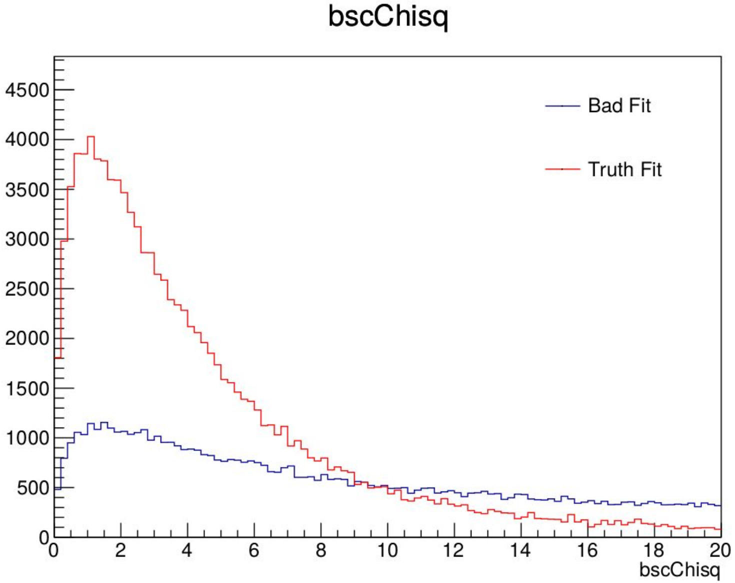




Re-thinking the Isolation Cut

- Match tracks to truth particle, select bad tracks (tracks with at least one bad hit)
- Force track to refit to truth hits, compare re-reconstructed vertices. Details here
- There are clear differences between bad/truth fit in a variety of vertexing/tracking variable. This method looks promising, but needs work

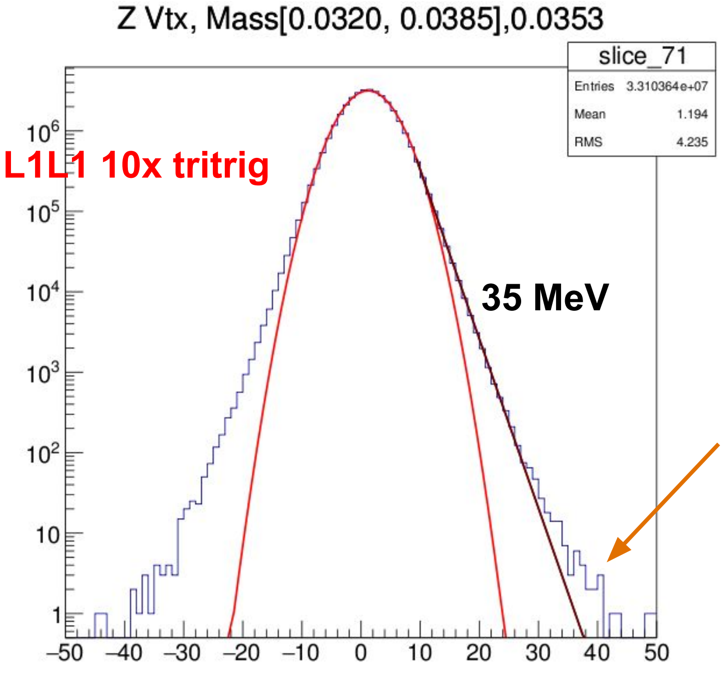


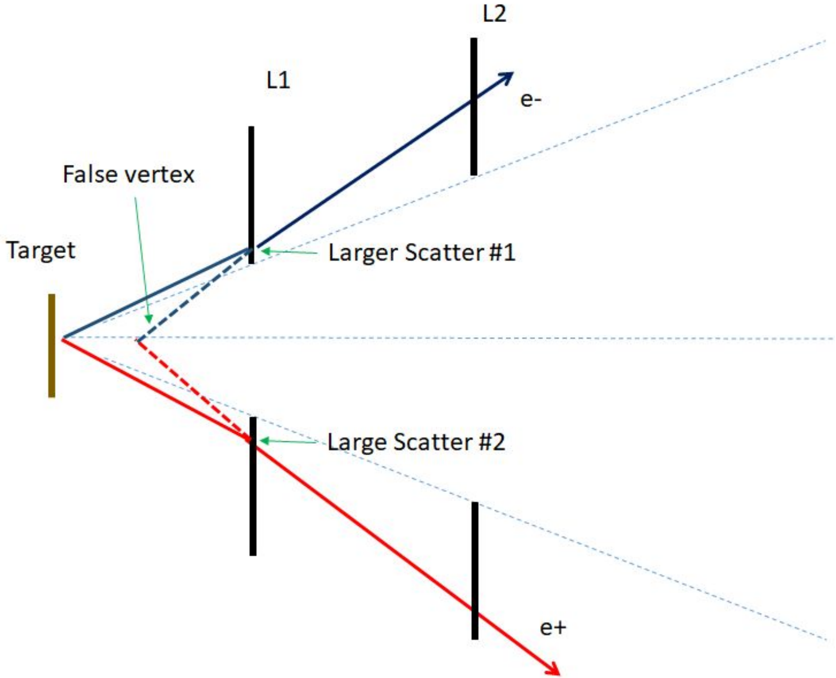


Lesson Learned from 2015 Vertexing (3)

SLAC

- Secondary tail likely due to single Coulomb scatters seen in 10x trident sample

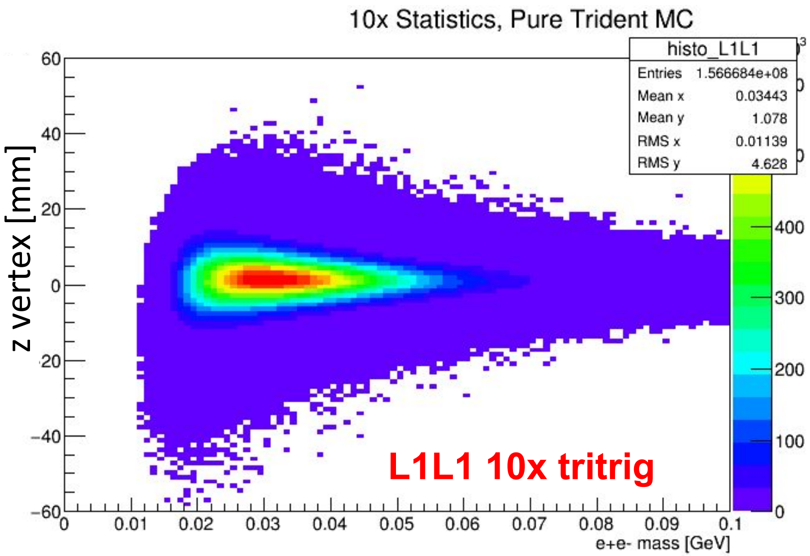


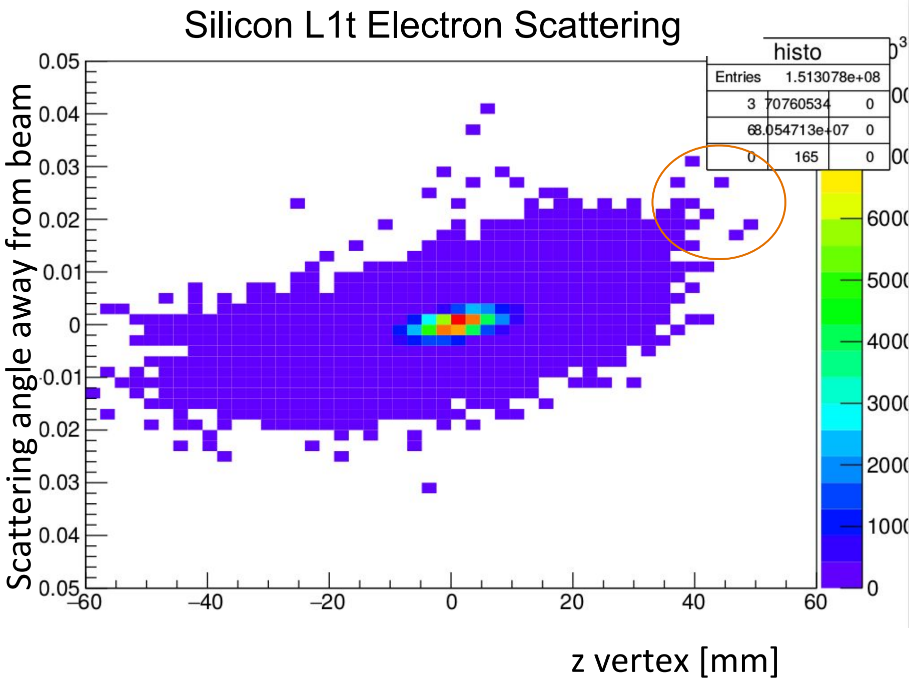


"If there were no scattering, this experiment would be a lot easier." - Matt Solt, future PhD

Modeling High Z Backgrounds

- Very high Z events appear to be from L1 scattering as expected
- Evidence of single Coulomb scattering (large scattering in one sensor)
- How do we mitigate these?





A Machine Learning Approach to Vertexing

SLAC

- We have more power in this analysis than square cuts and reconstructed z
- Combine variables into an MLP which outputs a single score from 0-1
- Use x10 pure trident sample - 60% training, 30% validation, 10% testing
    - Use slices in mass and only high z events in both signal and background

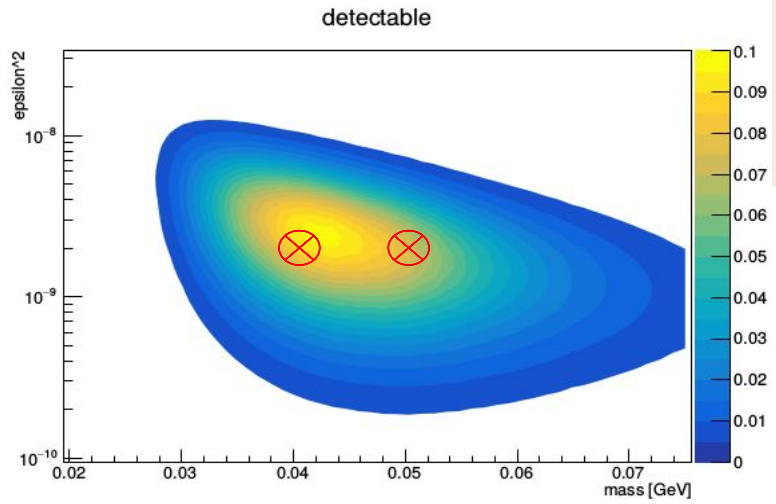
Variables Used	
uncVX & error	ele/pos TanLambda
uncVY & error	ele/pos Track Chisq
uncVZ & error	bscChisq
ele/pos Z0	uncChisq
ele/pos D0	uncM

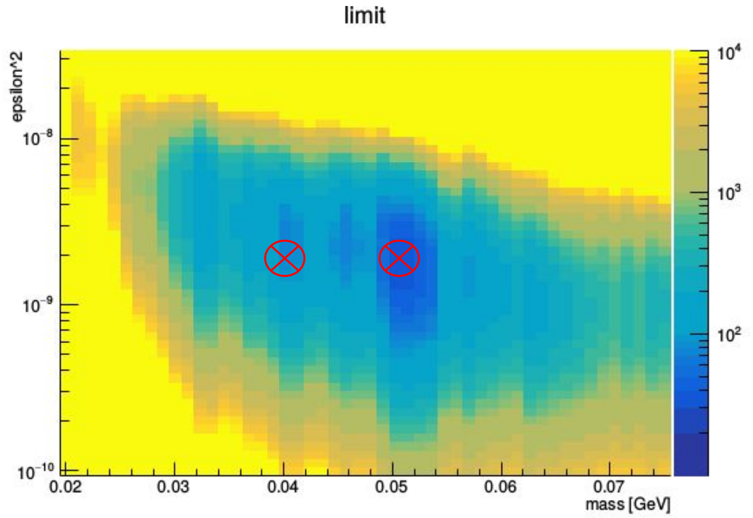
Focus on variables in the measurement direction →

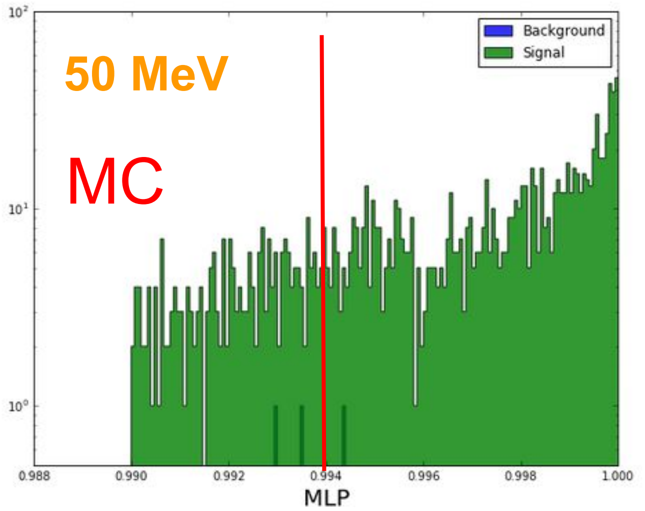
Variables I Should Focus	
uncVY & error	ele/pos TanLambda
ele/pos Z0	V0 y extrap.
uncVZ & error	uncM

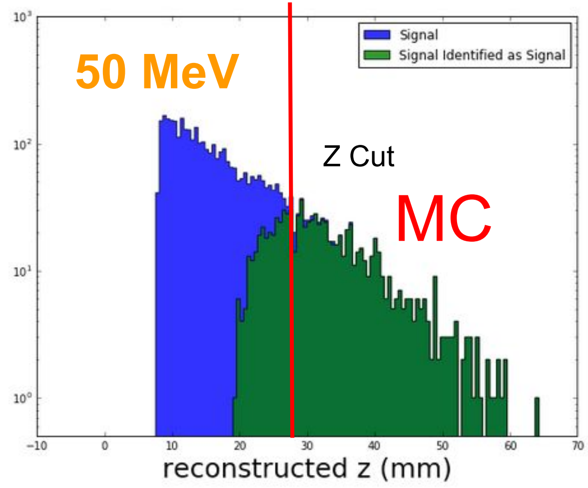
Signal Samples

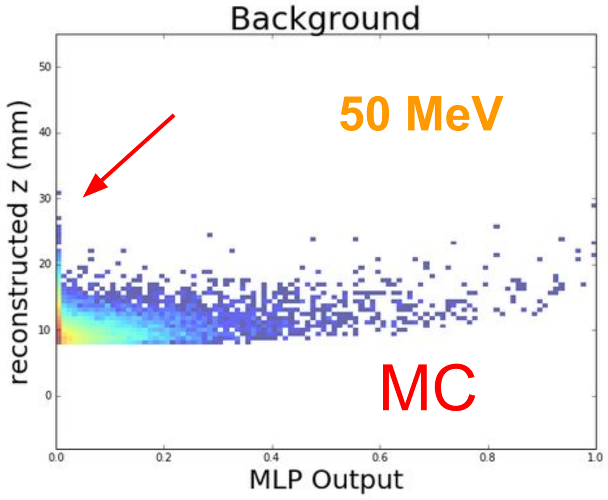
- Choose 2 samples in optimum parameter space for HPS (with correct signal shape)
  - 40 MeV; $\varepsilon^2 = 2 \times 10^{-9}$; $c\tau = 1.0$ mm (optimum A' yield)
  - 50 MeV; $\varepsilon^2 = 2 \times 10^{-9}$; $c\tau = 0.8$ mm (optimum A' limit)
- Use pure tridents for background
- Cut at MLP value where we expect 0.5 background events

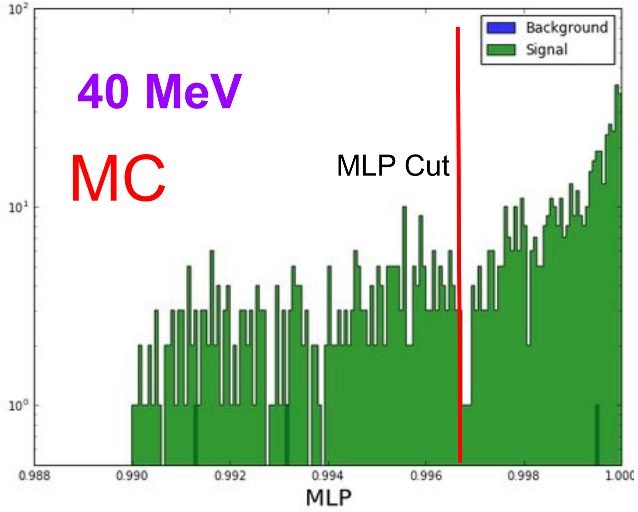


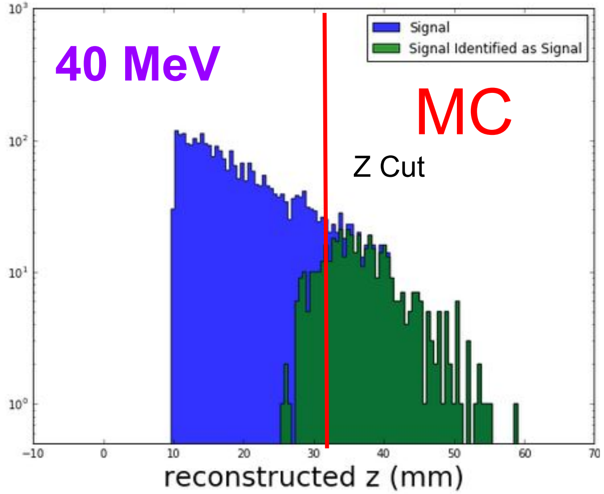


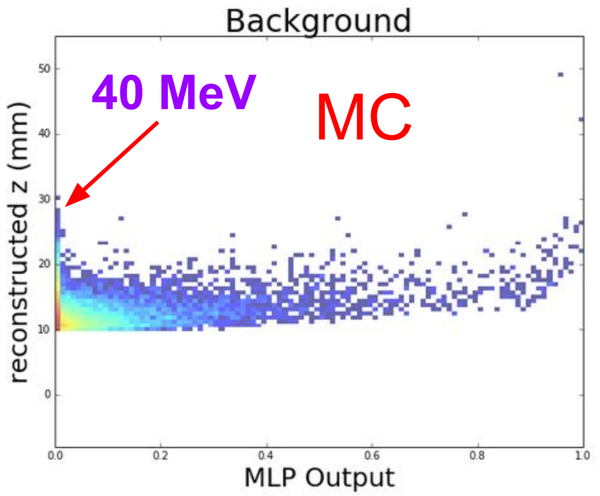

50 MeV
MC
Background
Signal
MLP
0.988
0.990
0.992
0.994
0.996
0.998
1.000

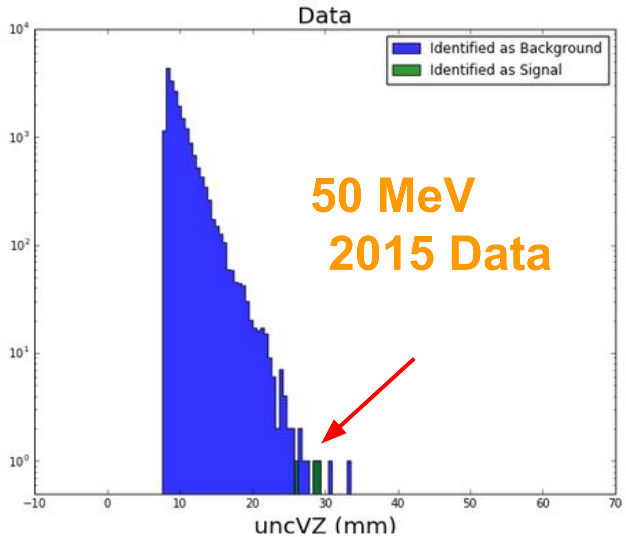

50 MeV
Signal
Signal Identified as Signal
Z Cut
MC
reconstructed z (mm)

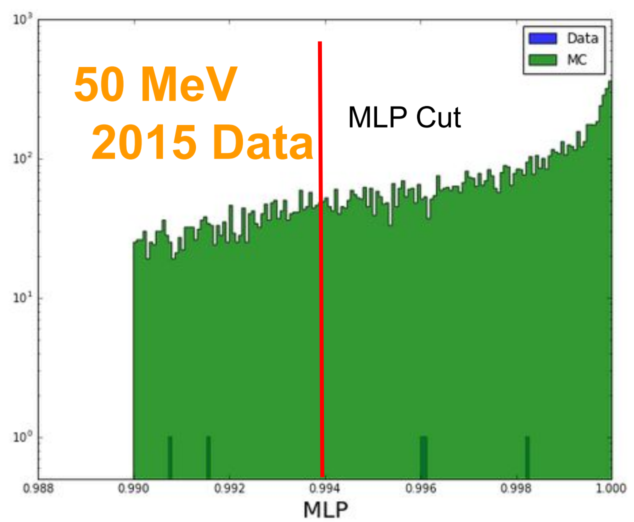

Background
50 MeV
MC
reconstructed z (mm)
MLP Output

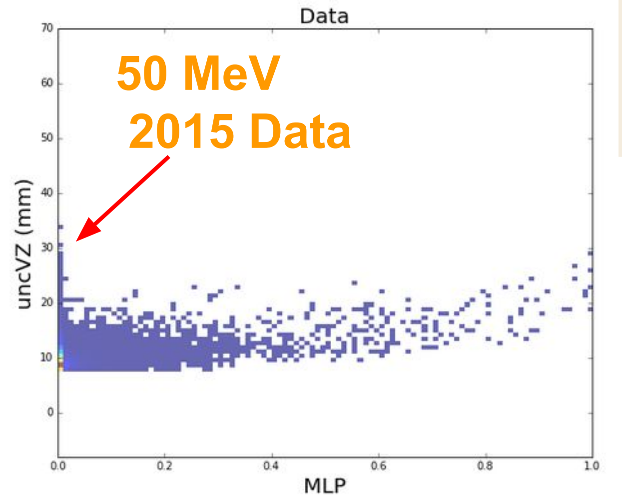

40 MeV
MC
MLP Cut
Background
Signal
MLP

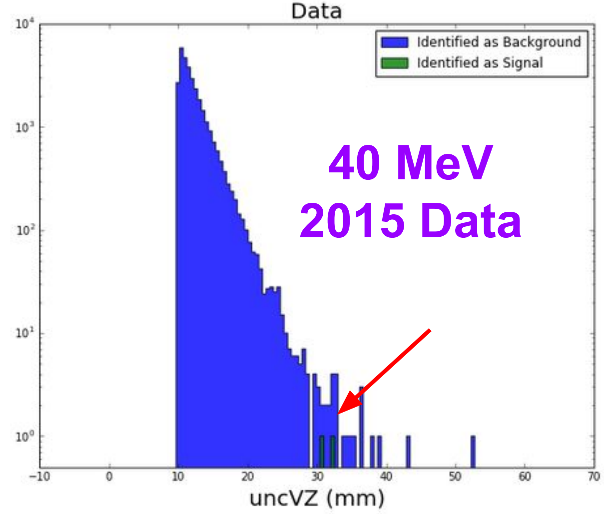

40 MeV
MC
Signal
Signal Identified as Signal
Z Cut
reconstructed z (mm)

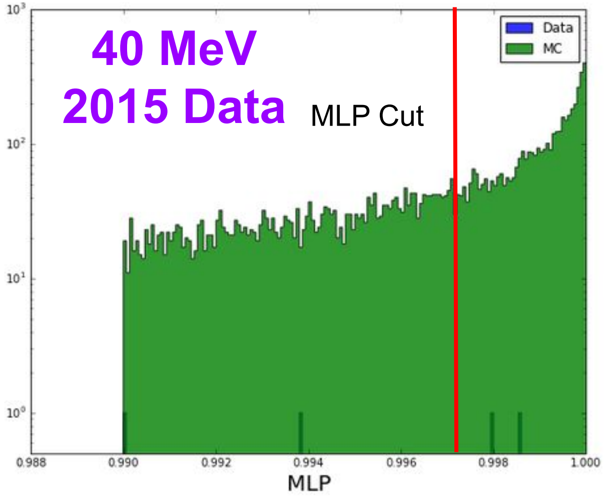

Background
40 MeV
MC
reconstructed z (mm)
MLP Output

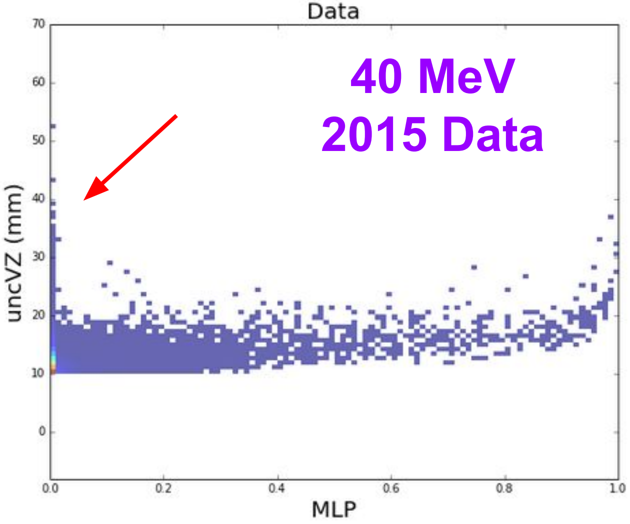

Data
Identified as Background
Identified as Signal
50 MeV
2015 Data
uncVZ (mm)


50 MeV
2015 Data
MLP Cut
Data
MC
MLP


Data
50 MeV
2015 Data
uncVZ (mm)
MLP


Data
Identified as Background
Identified as Signal
40 MeV
2015 Data
uncVZ (mm)


40 MeV
2015 Data
MLP Cut
Data
MC
MLP


Data
40 MeV
2015 Data
uncVZ (mm)
MLP

Machine Learning Methodology

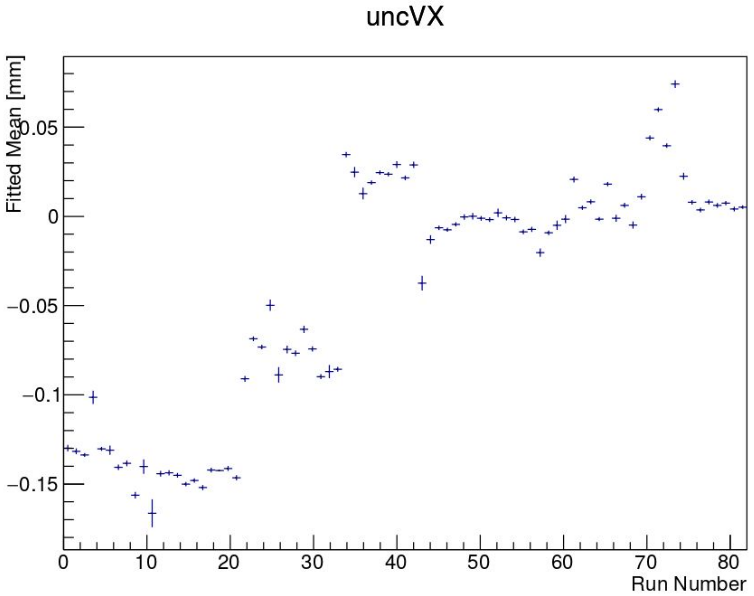
- Various methods have shown both increase in signal yield (10%-100%) and dramatic background reduction
- This specific method (MLP) is not yet justified. Method must be:
    - Economic - completion of analysis needs to be done in a reasonable time
    - Computationally feasible (so far MC production is the limiting factor)
    - Done in such a way that future analysis can use and build upon
- Each variable must be thoroughly justified, I should focus on tracking/vertexing variables in the measurement direction to start
- We need to think of a method to set a limit - Optimum Interval Method on the MLP output?
- talk at ML @ SLAC 2019

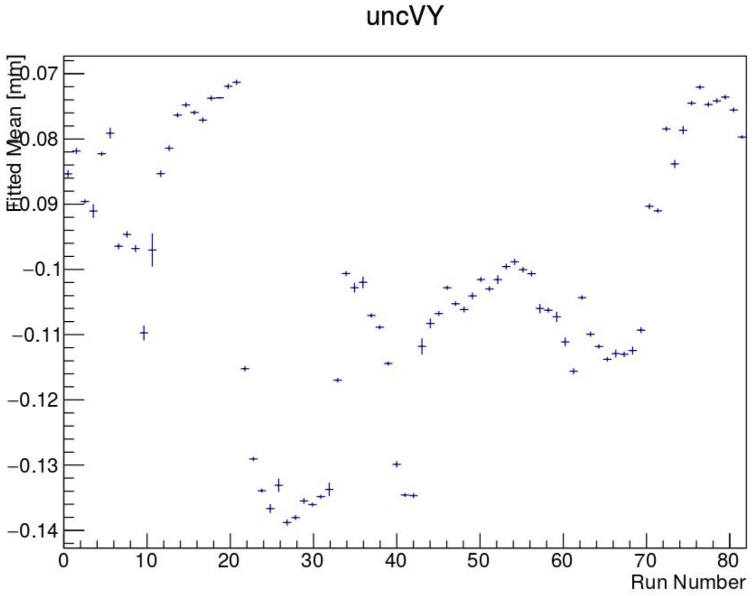
2016 Data Quality

- Major issues (details here)
  - Incorrect corrections applied to Ecal Cluster energy. Bug found and fixed (Nathan & Norman)
  - Major discrepancies in data/MC timing (Matt G.)
  - bscChisqProb MOUSE cuts were cutting out high z signal (Matt S.)
- 2nd order effects (things to keep in mind for 2019 analysis)
  - Target position as function of mass
  - Target position systematic shifts (difference in Mollers and V0s, shift in MC)
  - Resolutions in x are about 50% different between data/MC (D0, uncVX)
  - Minor shift in invariant mass between data/MC
  - Minor discrepancies in Track Chisq Shifts and Bsc Chisq between data/MC
  - Beam size as a function of run

Run-Dependent Beam Position

- We observed run-dependent beam positions in 2016 data
- From Pass2 to Pass4 - run-dependent beam position
- This will be more important at 4.55 GeV with improved resolutions





2016 Current Data/MC status

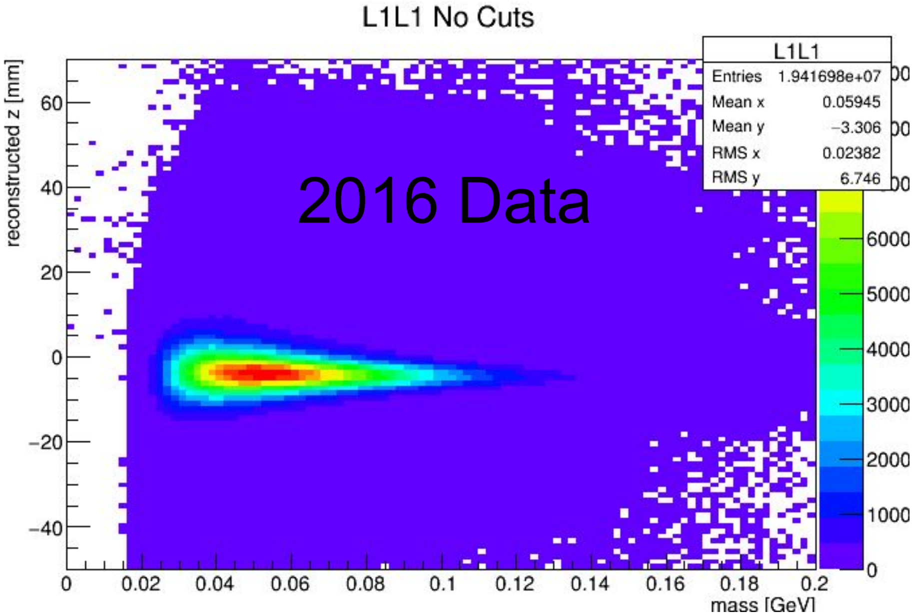
SLAC

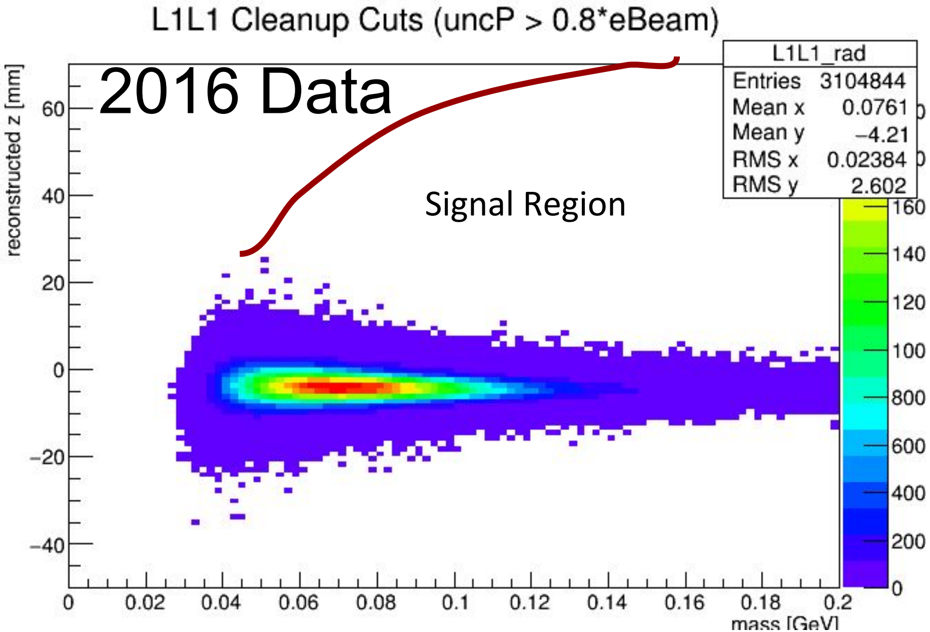
- We are on Pass4c for data
  - This is Pass4b with some changes to beamspot constrained from Matt G.
  - I checked the branch, but no one has looked at the recently reconstructed data
- We have MC samples that are up to date with Pass4c
- We need the large tritrig-wab-beam MC and the very large tritrig MC
- Results from Pass4b with “cleanup cuts” are shown
  - Major cuts missing include isolation cut, V0 extrapolation cut, and kink cuts

Cleanup Cuts	
uncP<1.15*2.3	eleP<0.75*2.3
bscChisq<10	MatchChisq<10
bscChisq-uncChisq<5	abs(eleP-posP)/uncP<0.5
abs(eleClT-posClT)<2	(eleTrkChisq/(2.0*eleNTrackHits-5.0)+posTrkChisq/(2.0*posNTrackHits-5.0))<6.0

L1L1 for 2016 Data

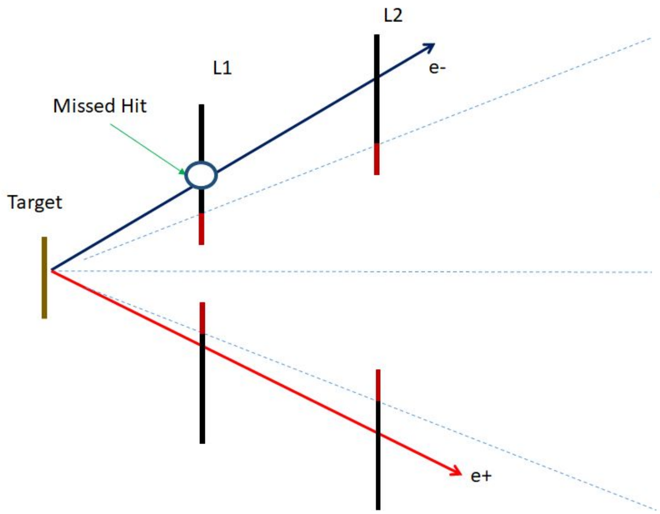
- Require L1 hits and radiative cut in addition to "cleanup cuts"

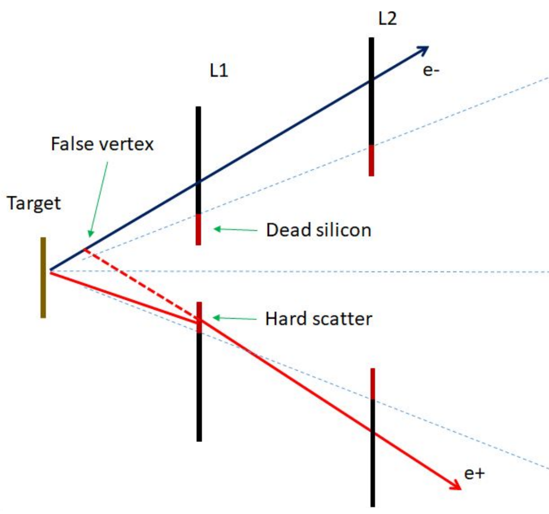


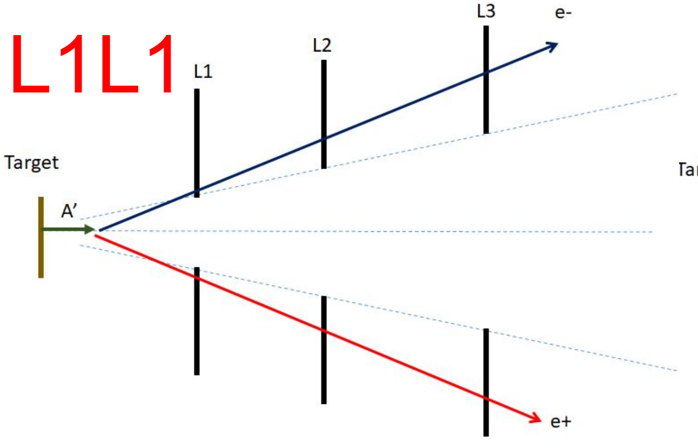


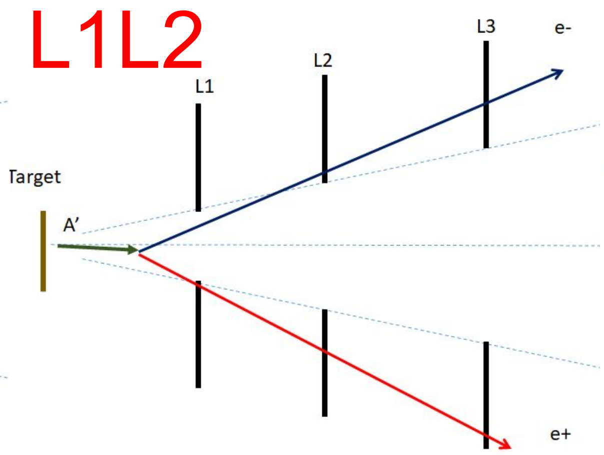
L1L2 and L2L2

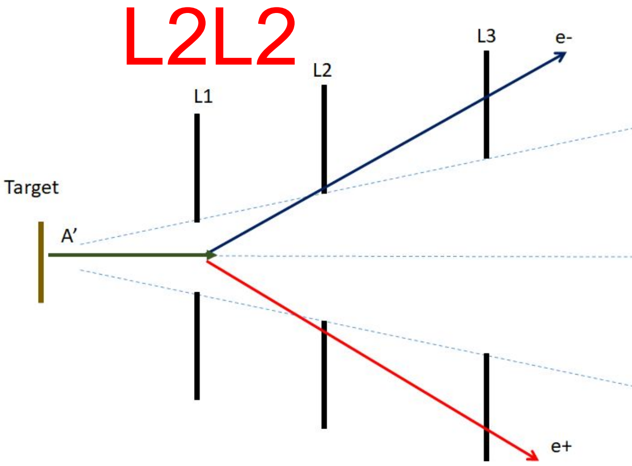
- What if e+ and/or e- misses L1?
- Much tougher backgrounds...





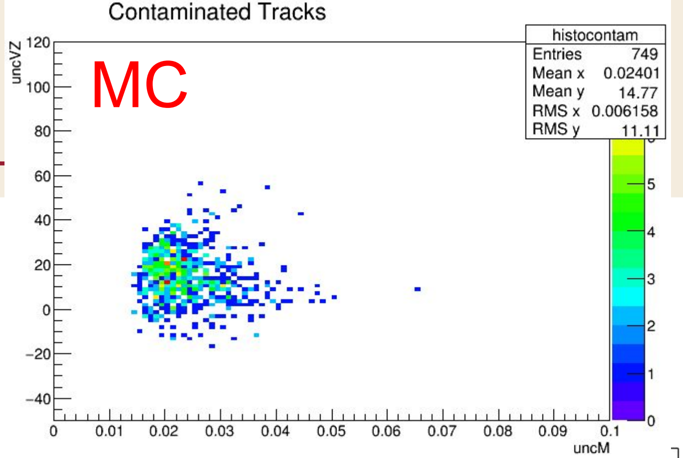


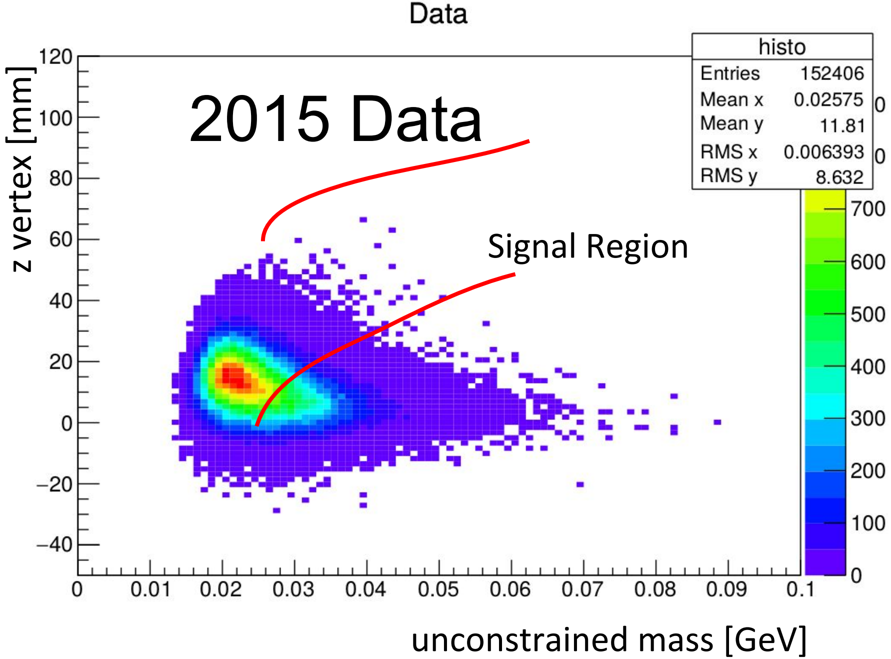


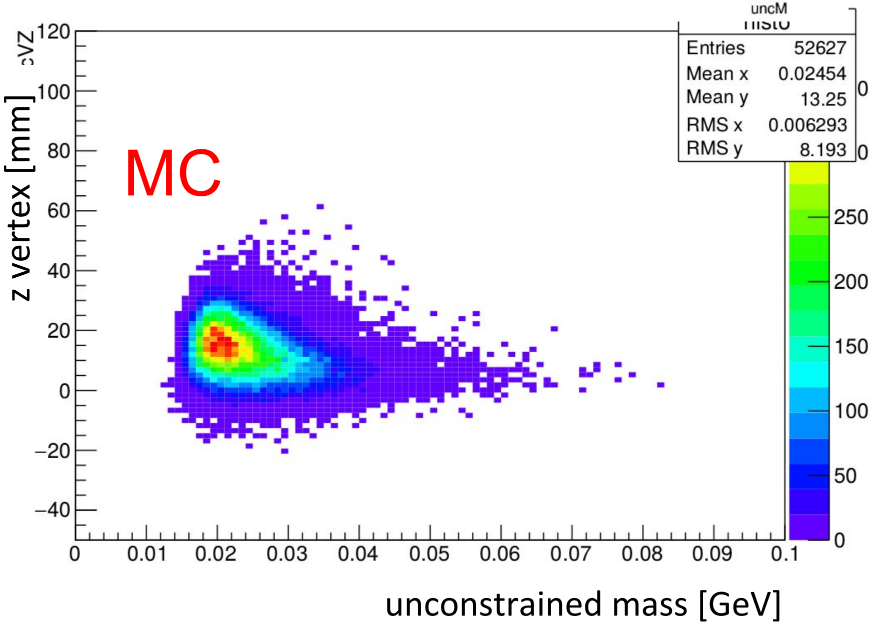


Lessons from L1L2 for 2015 Data

- Bad tracks are a significant source of high z events for L1L2
- Scattering is a secondary source of high z



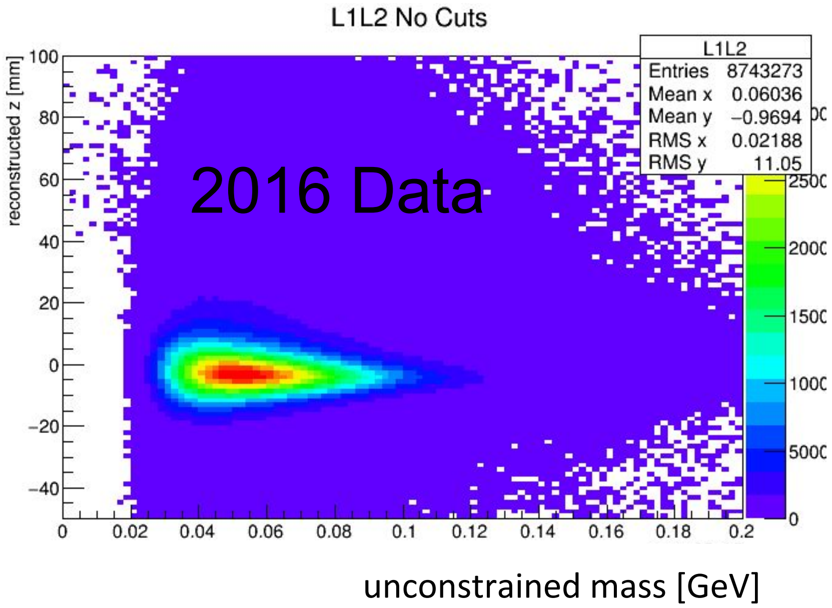


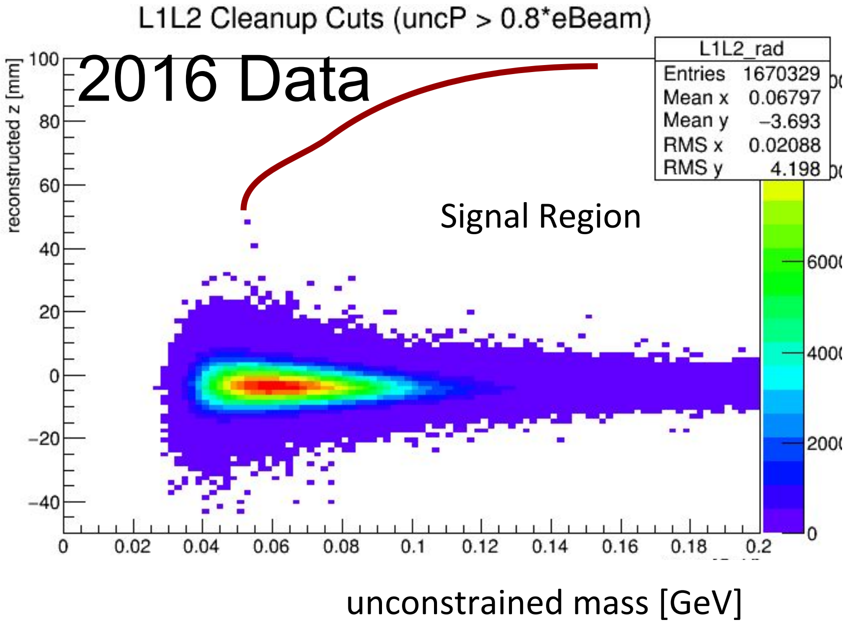


L1L2 for 2016 Data

SLAC

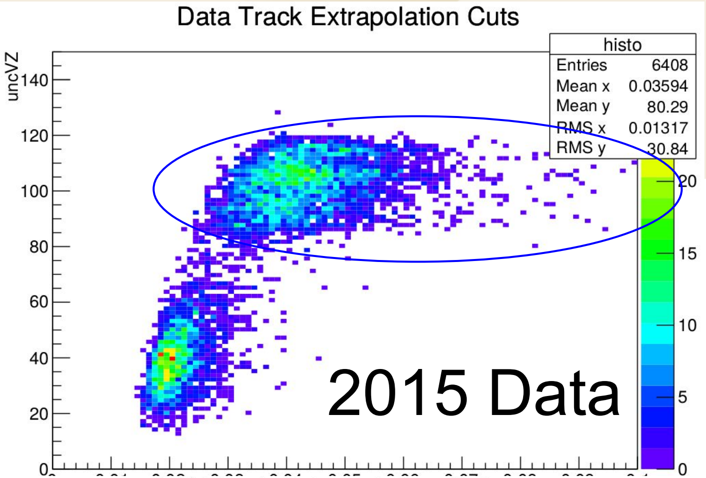
- I did not include a track extrapolation cut to active sensors to eliminate hit inefficiencies (like I did in 2015). This cuts signal and is not a significant source of high z background. We should optimize this cut

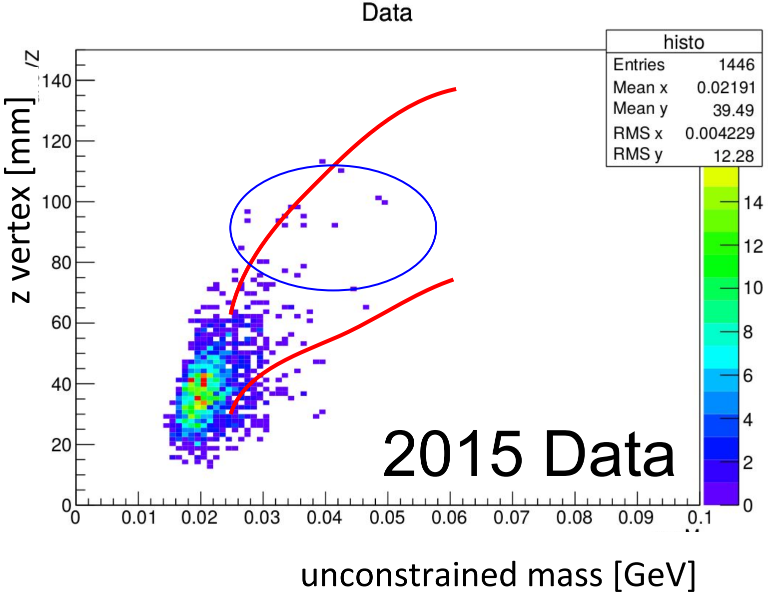


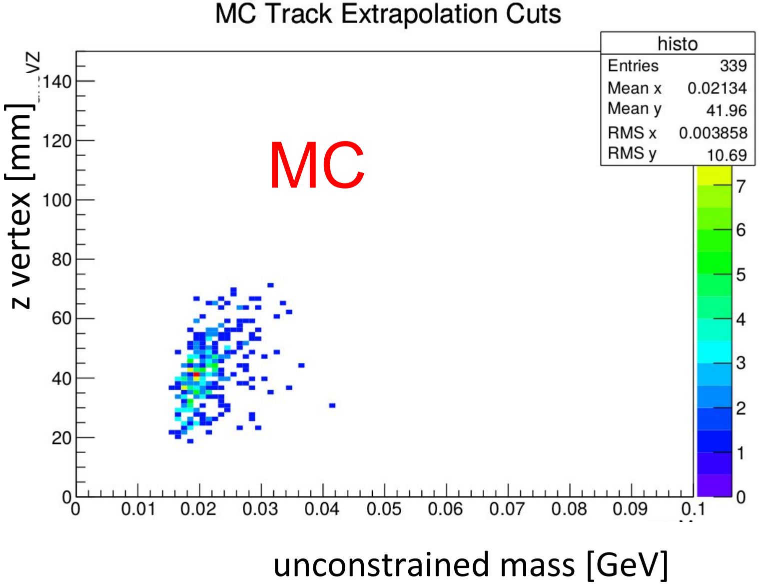


Lessons from L2L2 for 2015 Data

- We observe large amount of trident production in inactive silicon of L1 (most likely from FEEs)
- We probably have bad tracks, but I did not show this in MC yet



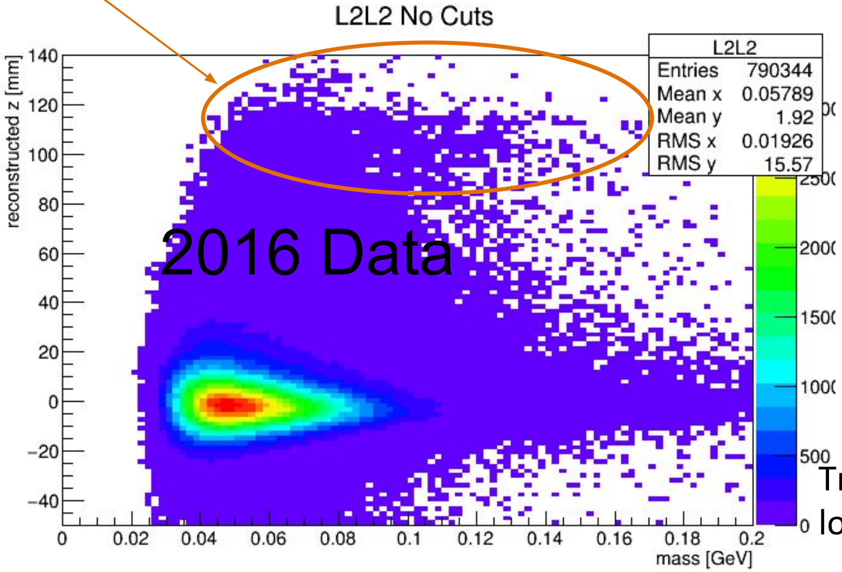




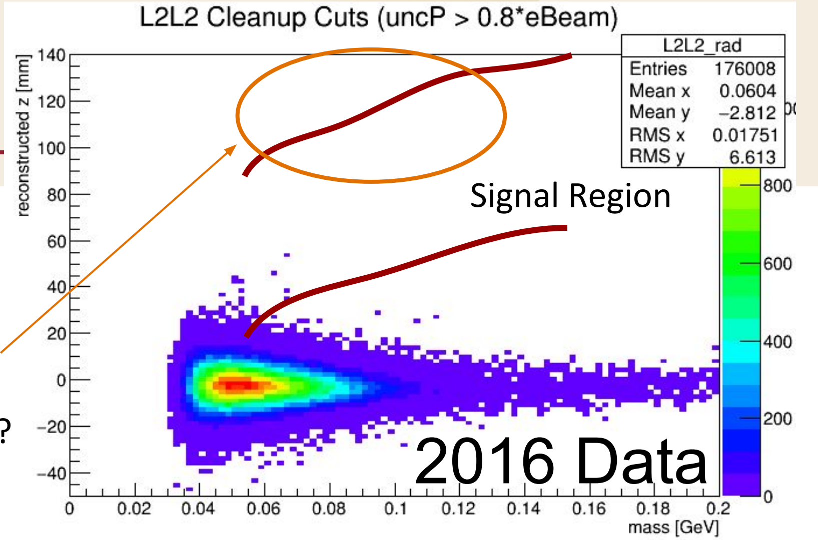
L2L2 for 2016 Data

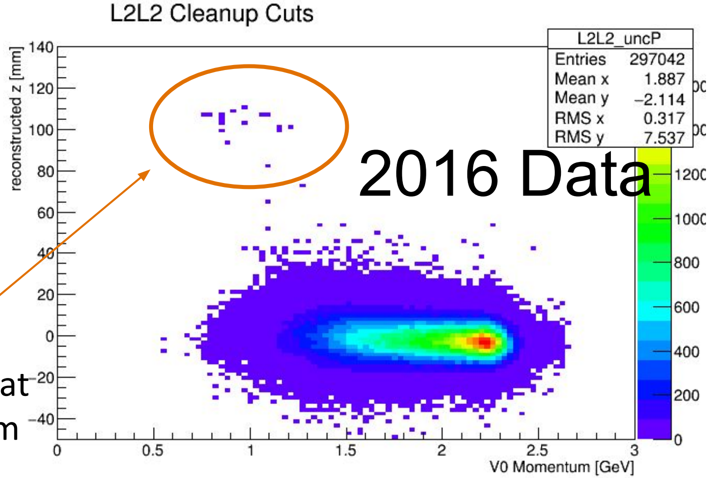
- We observe trident production at L1 silicon, but only at low momentum sum. Why?

Most likely trident production in L1 silicon



But, why not here after cuts?





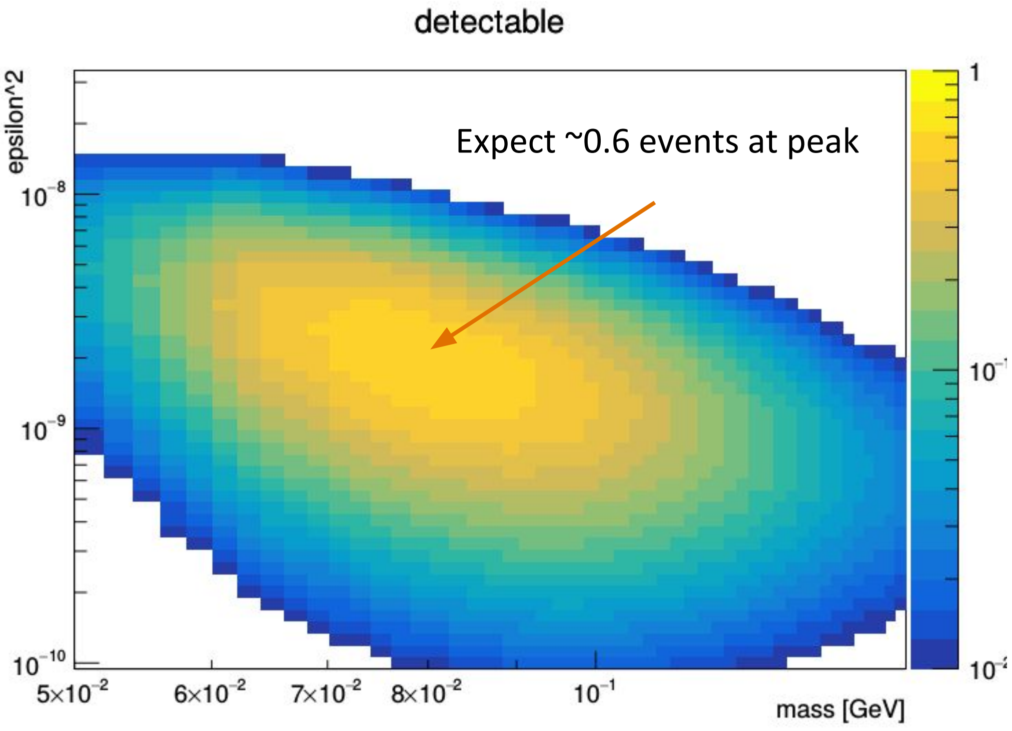
Trident production at low momentum sum

Method for Estimating Signal Yield 2016

- Use full detector simulation for displaced A's
- Use basic cleanup cuts on both A's and data
- Use 10% of pass4b data to fit tails, scale up to 100% to get zcuts as function of mass
  - This does not include a fit to the secondary tail
- Integrate signal yield past zcut, plot as function of mass and ε
- Assume radiative fraction of 10% (from 2015 data), Bradley will show something different for 2016
- Estimate zcut by eye for L1L2 and L2L2 (we need a better method obviously)

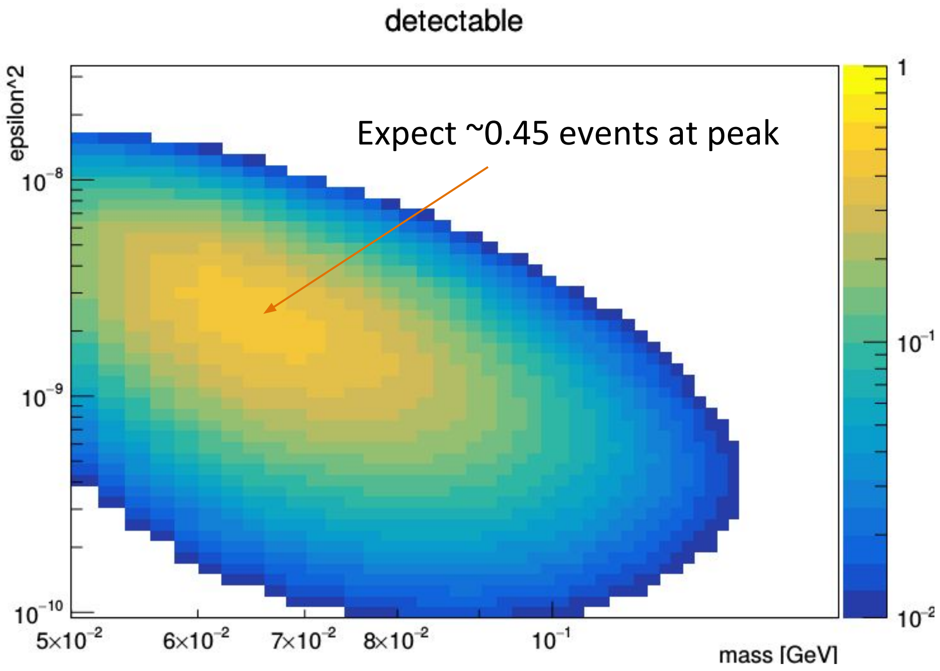
2016 Reach Estimate L1L1

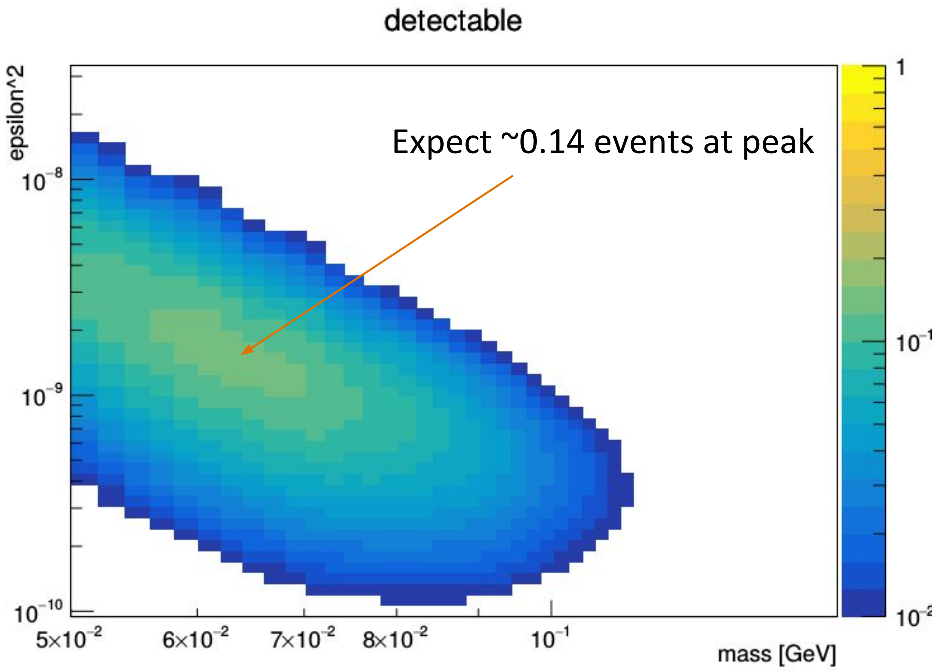
- “Detectable” = integrated expected number of A’s past zcut



2016 Reach Estimate 2016 L1L1 and L2L2

- These projections for L1L2 and L2L2 are probably optimistic, zcut was determined “by eye”. For L2L2, I assume no background at L1 silicon

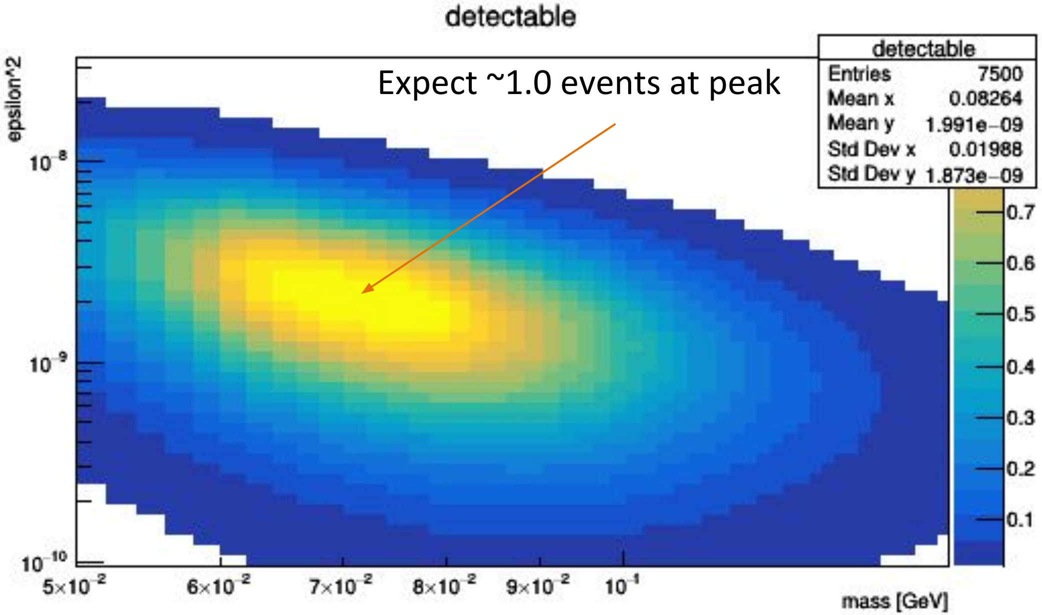




Total Signal Yield

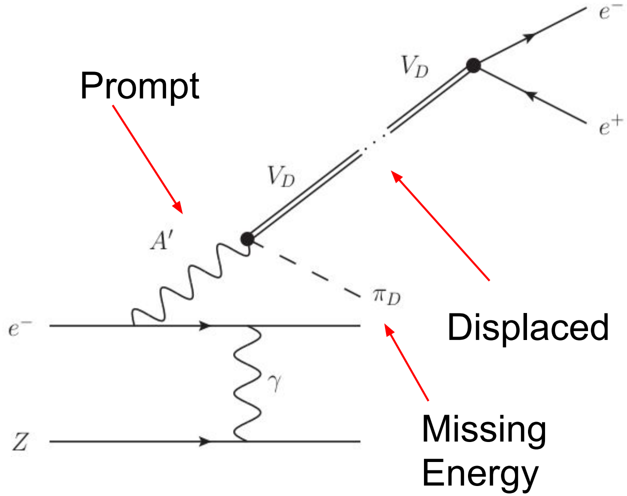
- Order of magnitude improvement from 2015 (which was at 0.1)!
- This includes a radiative fraction of 10% (could be more!)
- Does not include possible signal yield improvement from ML approaches (shown to be ~50% - ~100% improvement)
- We need to get to 2.3 events to have a chance
- Do we have reach? Maybe… if we get *really* lucky...

Total = L1L1 + L1L2 + L2L2

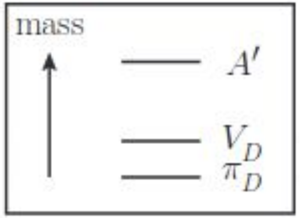


SIMPs at HPS

- Prompt A', Displaced dark V, and Missing Energy π
- Same A' kinematics and cross sections as minimal model
- Produce high rate of A's (large ε), displaced dark V (additional parameters decouple cτ and x-section)



Assume Mass Hierarchy

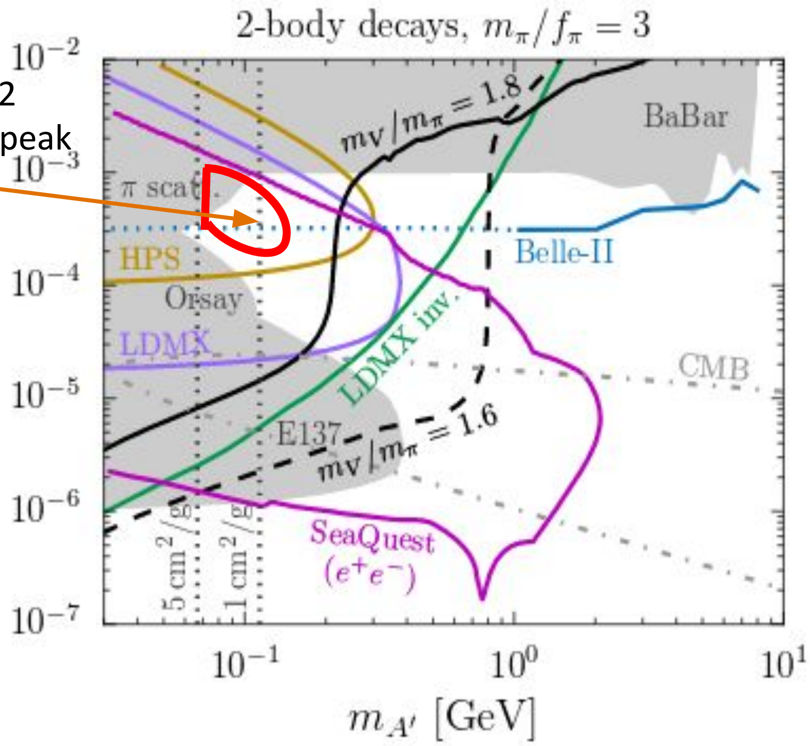


$m_{A'}$
m_V (mass of dark vector)
m_π (mass of dark pion)
ε
α_D (dark coupling constant)
m_π/f_π (dark pion decay constant)

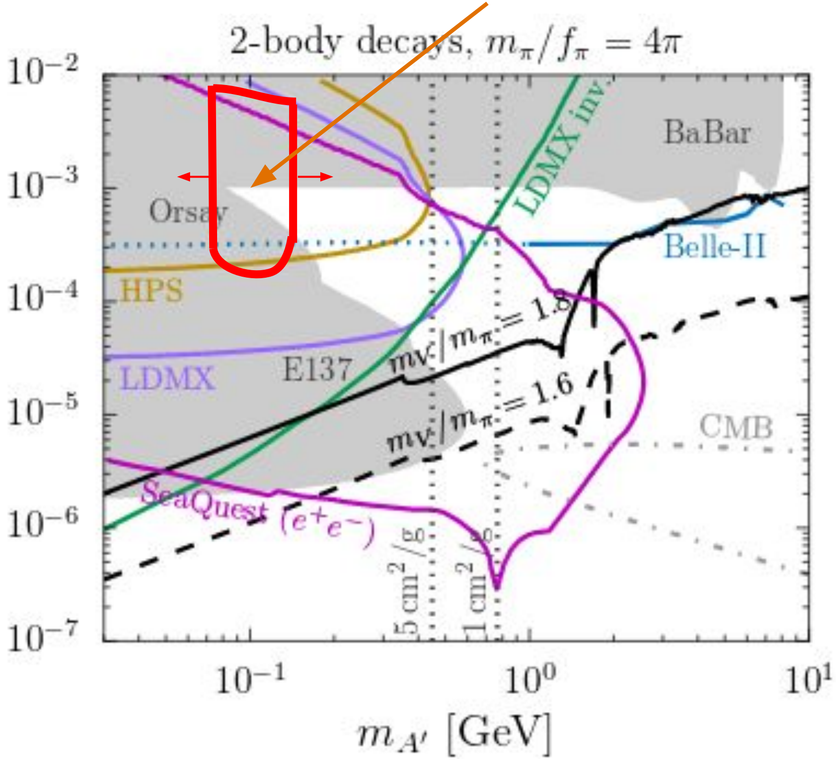
(Old) SIMP Reach Projections

- We have reach in NEW territory with 2016 data!

Expect ~12 events at peak

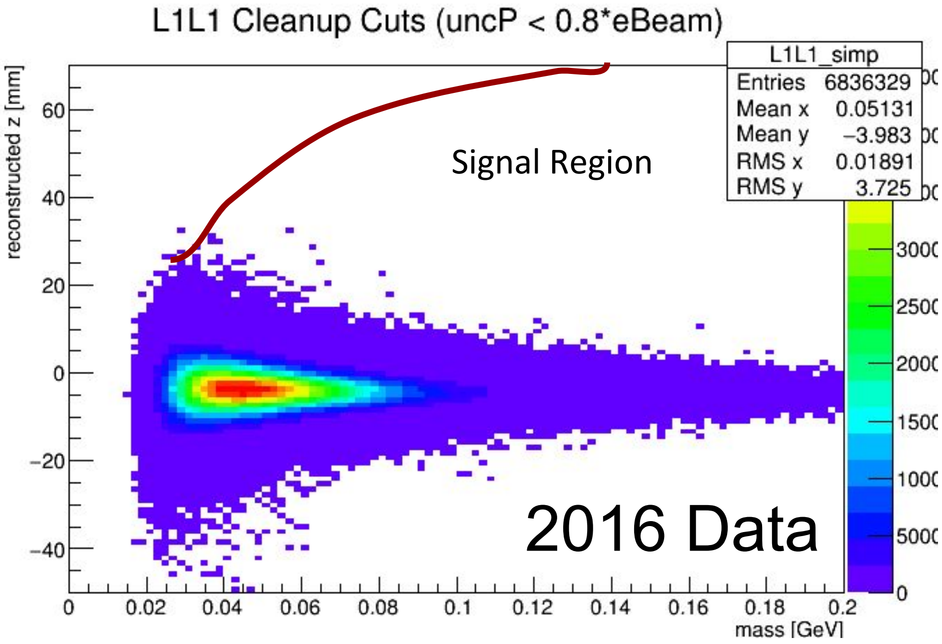


Expect ~1000s of events at peak

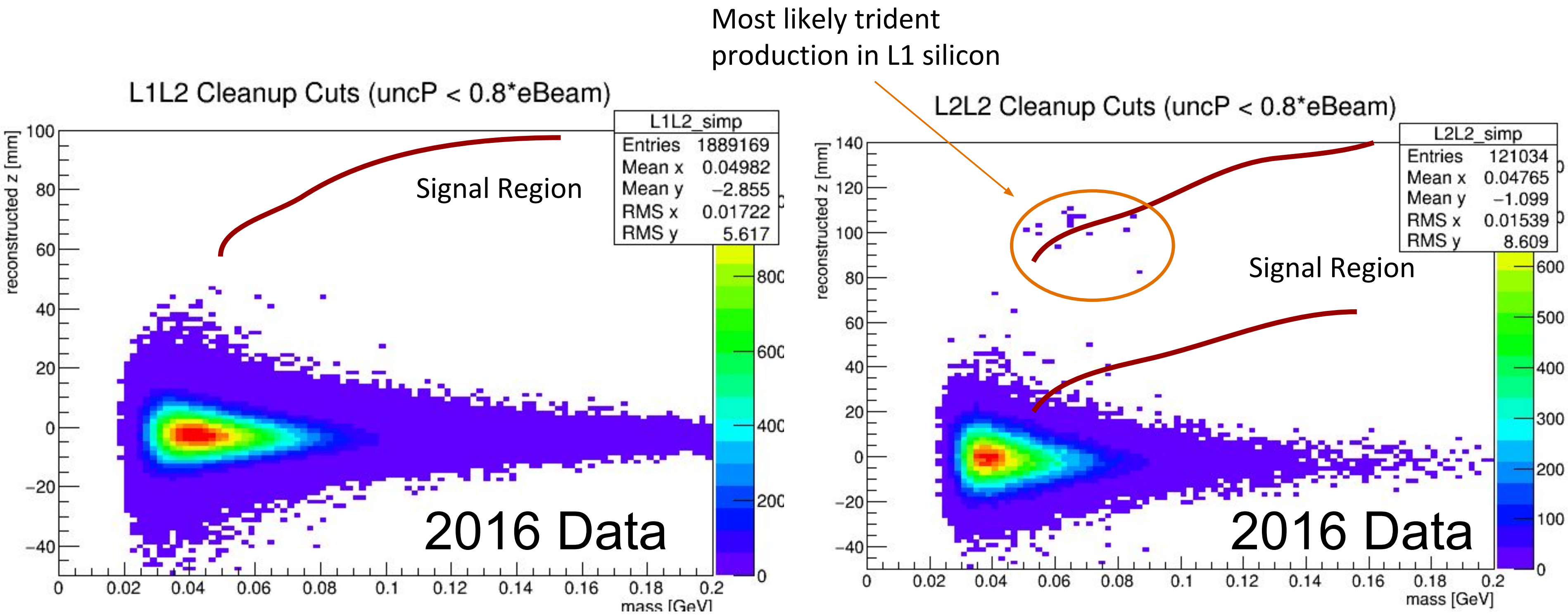


SIMPs in 2016 Data

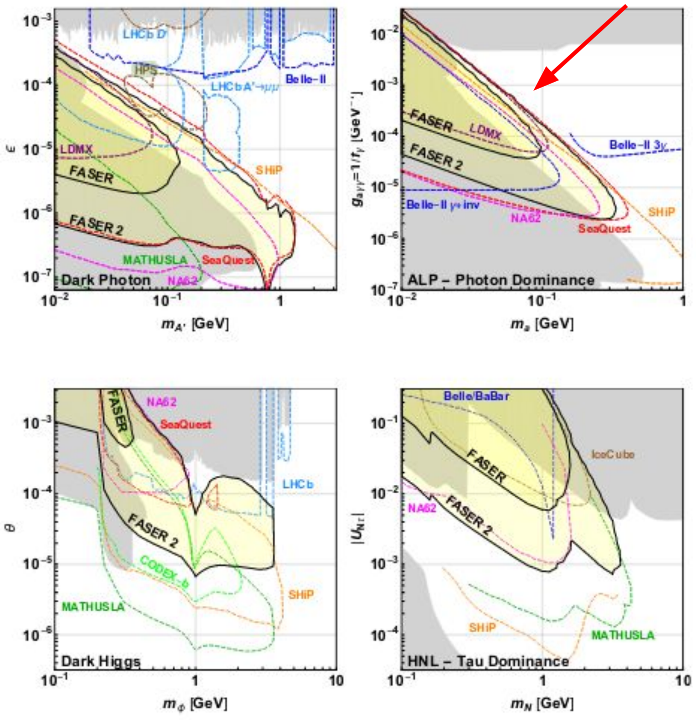
- SIMPs analysis can be done the exact same way as A’ search, except with an “uncP < 0.8 * eBeam” cut
- Requiring L1 hits and “cleanup cuts”, it doesn’t look too crazy
- With 10% 2016, we already seem to exclude new territory for SIMPs (whether this is theoretically motivated is another question)
- We should not be excited/worried about high z events yet, we have many more handles to be used



SIMPs in 2016 Data (L1L2 and L2L2)



Generalized Displaced Vertices



arXiv:1903.04497v1 arXiv:1811.12522v3

- We can measure σ(X)*Br(X->e+e-), cτ, and mass. Many models fit this constraint
- Many possible models presented in FASER reach paper
- Axion-like particles look promising (needs some more thought; production mechanism?)
- In the future, someone can think of >2 body decays

Conclusions

SLAC

- We learned a variety of valuable lesson from 2015 vertexing
  - Backgrounds from scattering and bad tracks; target position and global alignment
- The next big steps for 2016 analysis - L1L2 & L2L2, ML approach, and SIMPs
- Signal yield estimates shown, I project an order of magnitude improvement over 2015 data (expected peak of 1.0 A' events)
  - We still have more handles to use - radiation fraction improvement, ML improvements, and other things we can think of
- Focus for 2016 vertexing analysis right now should be large MC samples
- Plenty of room for people to get involved (especially new people)!