Detector Design Aided By Machine Learning

A case study of the Charged Pion Polarizability Experiment at Jefferson Lab

David Lawrence, JLab
Robert Johnston, UMass
Ilya Larin, UMass

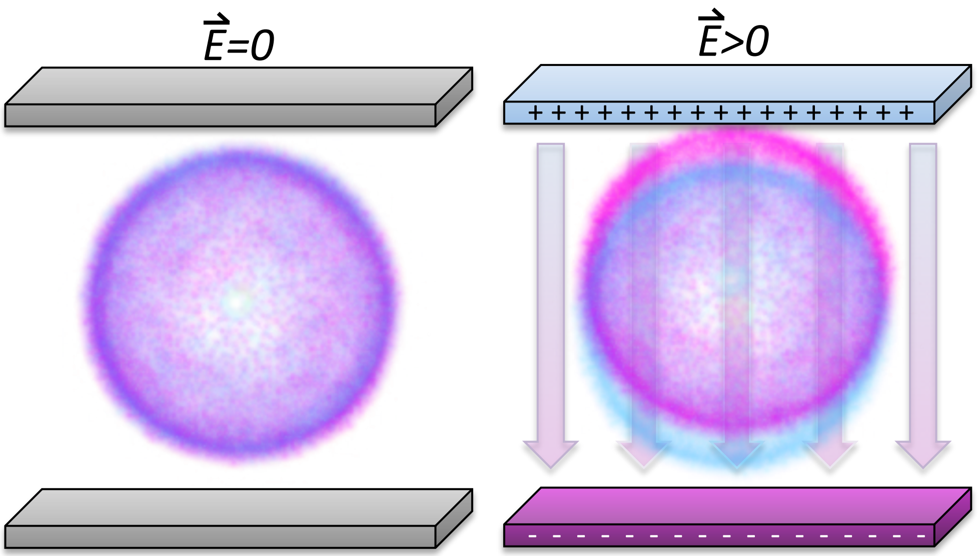
Oct. 25, 2017

U.S. DEPARTMENT OF ENERGY · JSA · Jefferson Lab

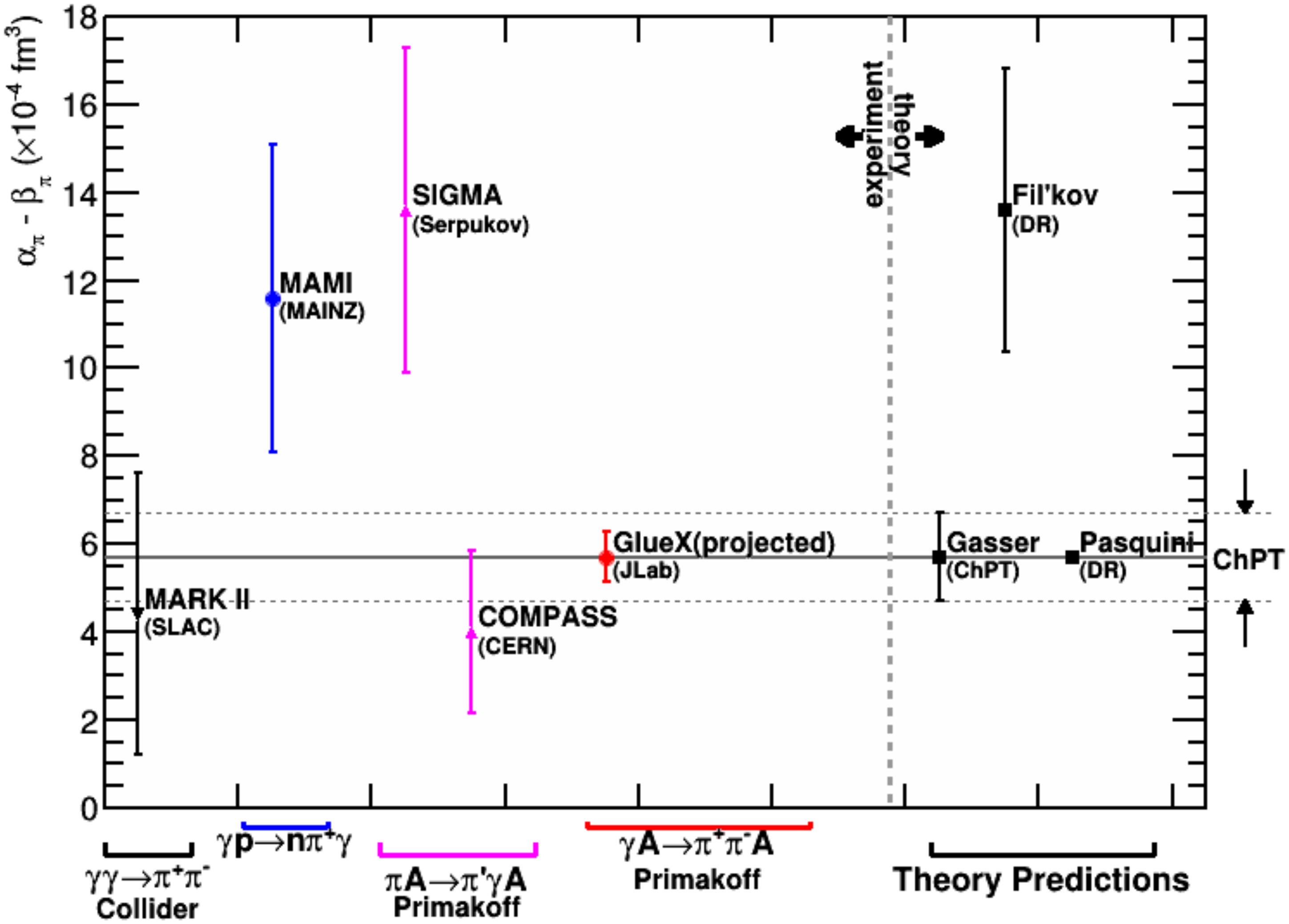
CHARGED π POLARIZABILITY

- Polarizability: Ease in which an external field may induce a dipole moment in a particle

- Property which reflects nature of internal structure of particle

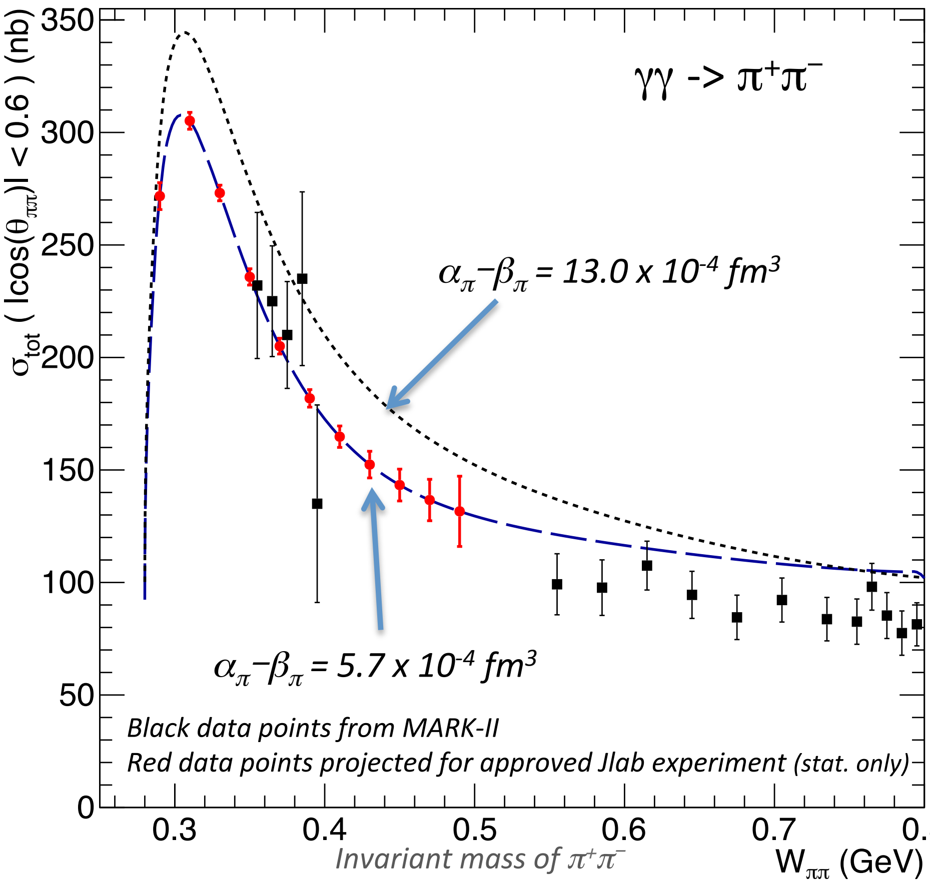


EXPERIMENTAL/THEORETICAL LANDSCAPE



RELATING CROSS-SECTION TO $\alpha_\pi - \beta_\pi$

Curves from figure 5. from Pasquini et al. Phys. Rev. C 77, 065211 (2008)



dotted: subtracted DR calculation with $\alpha_\pi-\beta_\pi = 13.0$
dashed: subtracted DR calculation with $\alpha_\pi-\beta_\pi = 5.7$

Cross-section for $\gamma\gamma \to \pi^+\pi^-$ calculated based on two values of $\alpha_\pi-\beta_\pi$:

$\alpha_\pi-\beta_\pi = 13.0 \times 10^{-4}\ fm^3$ (top, dotted line)

$\alpha_\pi-\beta_\pi = 5.7 \times 10^{-4}\ fm^3$ (solid and dashed lines)

Cross-section varies by ~10% for factor of 2 variation in $\alpha_\pi-\beta_\pi$

Need measurement of $\sigma(\gamma\gamma \to \pi^+\pi^-)$ at few percent level

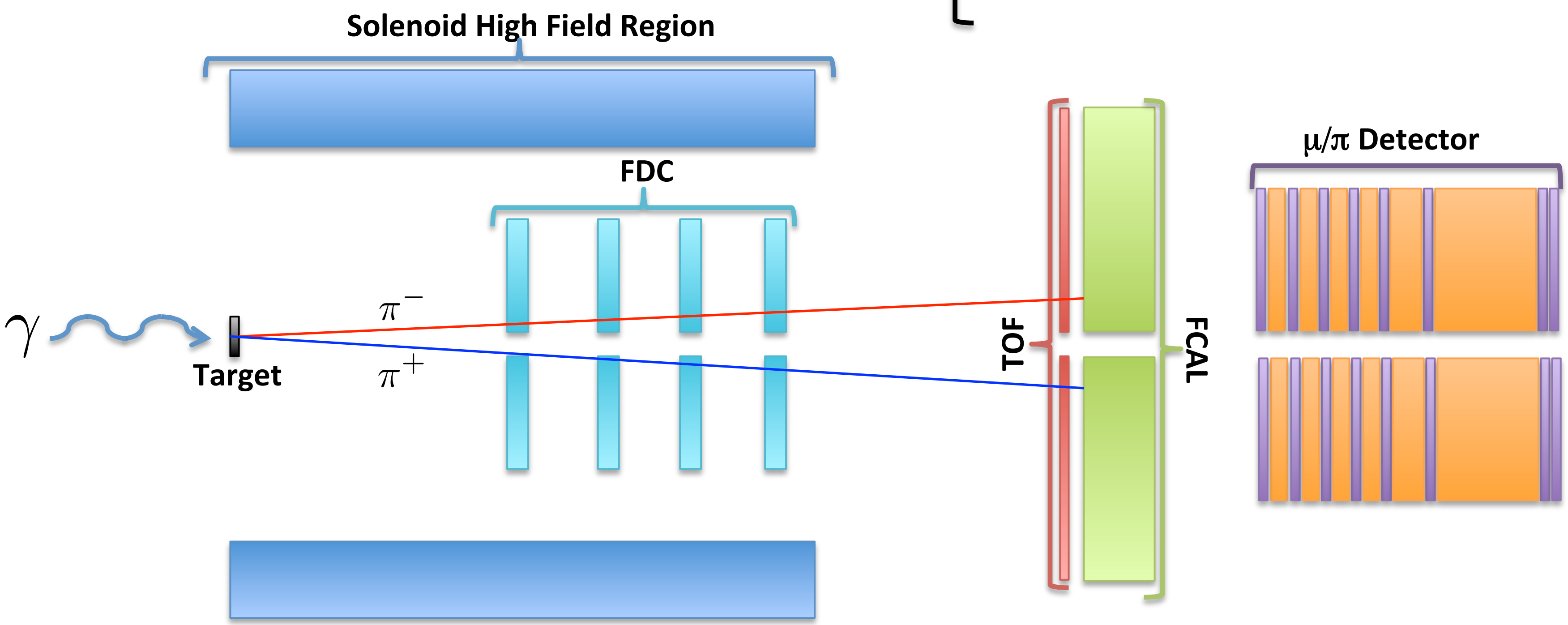
EXPERIMENTAL SETUP

$$\gamma\gamma \underset{Z(\gamma,\pi^+\pi^-)Z}{\longrightarrow} \pi^+\pi^-$$ Signal reaction

$$\gamma\gamma \underset{Z(\gamma,\mu^+\mu^-)Z}{\longrightarrow} \mu^+\mu^-$$ Normalization

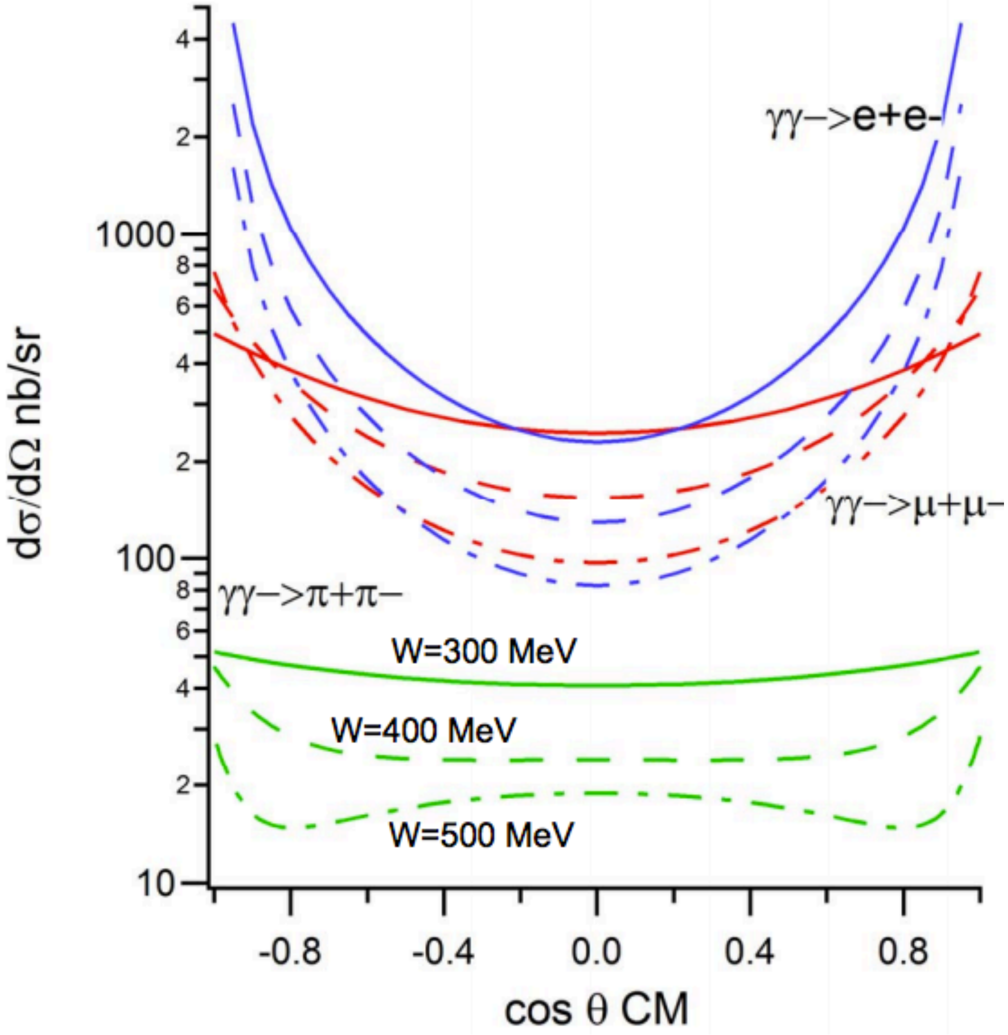
$$\gamma\gamma \underset{Z(\gamma,\pi^o)Z}{\longrightarrow} \pi^o$$ Beam polarization

- All occur via the Primakoff effect (interaction with the Coulomb field of nucleus)
- All result in very forward going particles
- Low t ($-t < 0.005$ GeV^2)



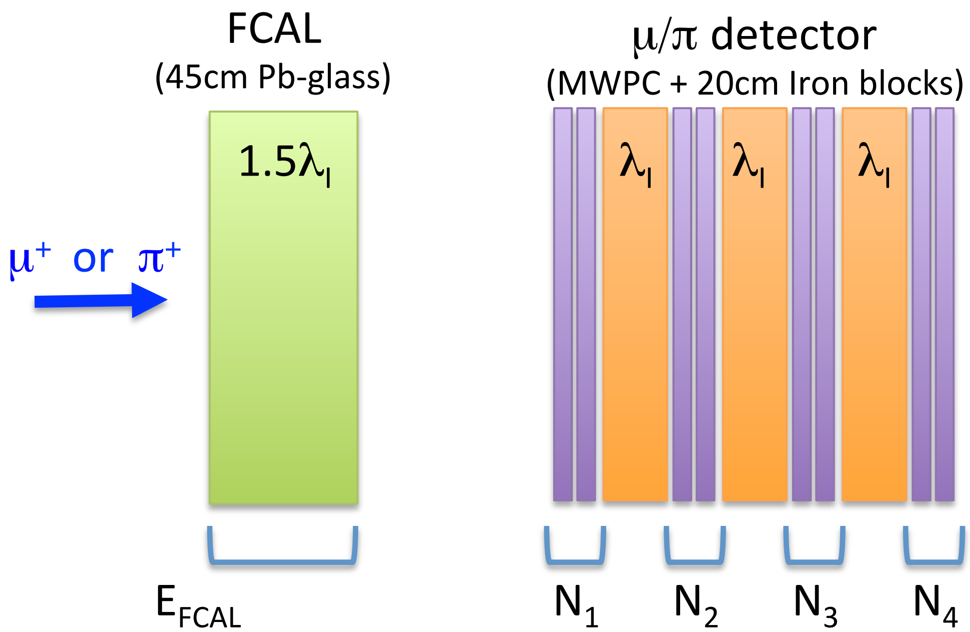
The Problem

- μ+μ- (background) are produced ~10x more often than π+π- (signal)
- To measure cross section to few percent means reducing a 10x bigger background to less than a few 1/10 of a percent
- GlueX detector has no way of distinguishing μ+μ- and π+π- events at this level
- A new detector that works in tandem with GlueX is required



First Attempt

- Simulate single 2GeV particles with simple geometry
- 4 nucl. Interaction lengths ➔ ~2% of πs make it through



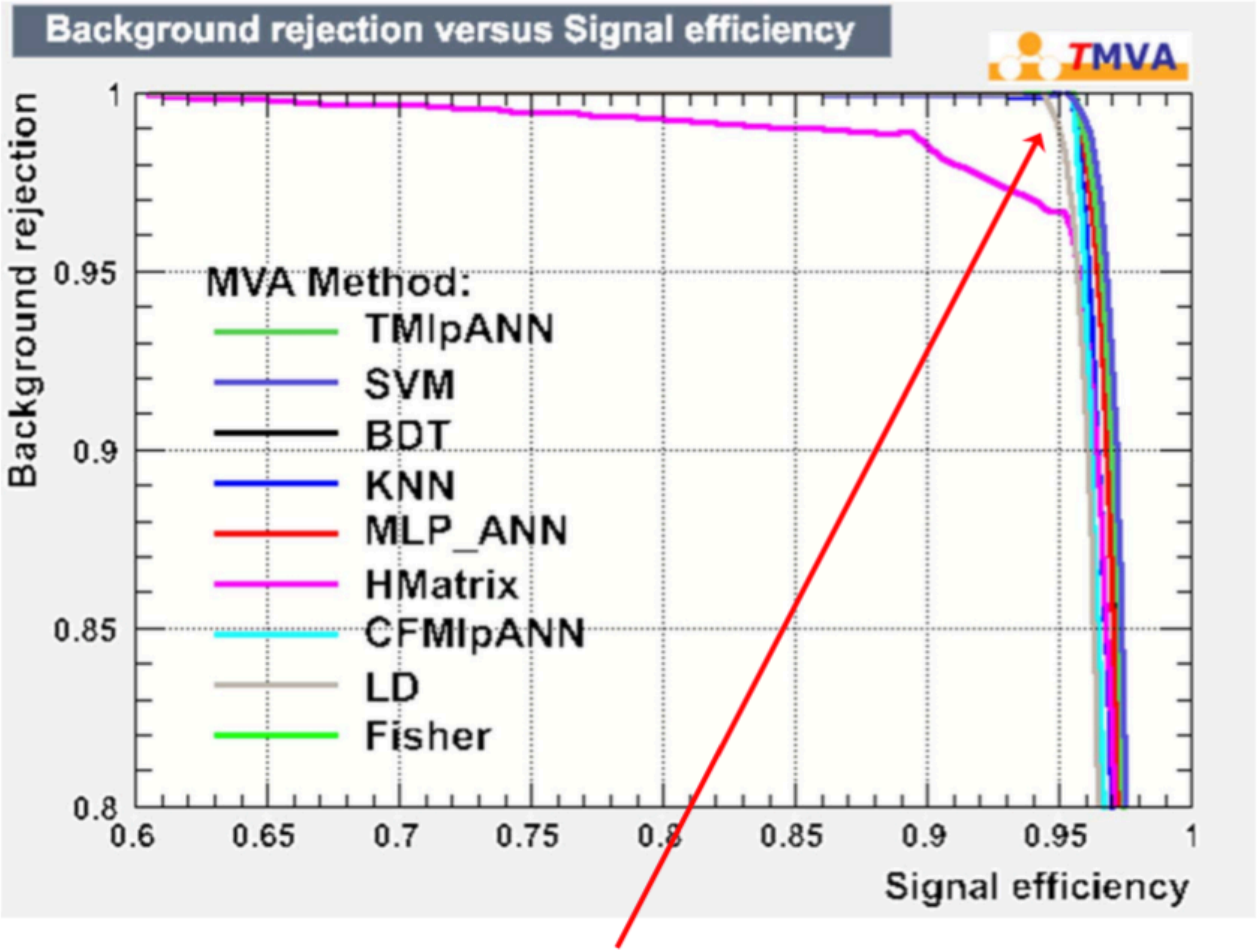
Tell TMVA what input variables to use

```
factory->AddVariable( "Nmwpc1", 'I' );
factory->AddVariable( "Nmwpc2", 'I' );
factory->AddVariable( "Nmwpc3", 'I' );
factory->AddVariable( "Nmwpc4", 'I' );
factory->AddVariable( "Efcal := Sum$(Edep)", 'F' );
```

Tell TMVA what methods to use

```
factory->BookMethod( TMVA::Types::kMLP,      "MLP_ANN",  "H:!V" );
factory->BookMethod( TMVA::Types::kFisher,   "Fisher",   "H:!V" );
factory->BookMethod( TMVA::Types::kKNN,      "KNN",      "H:!V" );
factory->BookMethod( TMVA::Types::kHMatrix,  "HMatrix",  "H:!V" );
factory->BookMethod( TMVA::Types::kCFMlpANN, "CFMlpANN", "H:!V" );
factory->BookMethod( TMVA::Types::kTMlpANN,  "TMlpANN",  "H:!V" );
factory->BookMethod( TMVA::Types::kSVM,      "SVM",      "H:!V" );
factory->BookMethod( TMVA::Types::kLD,       "LD",       "H:!V" );
```

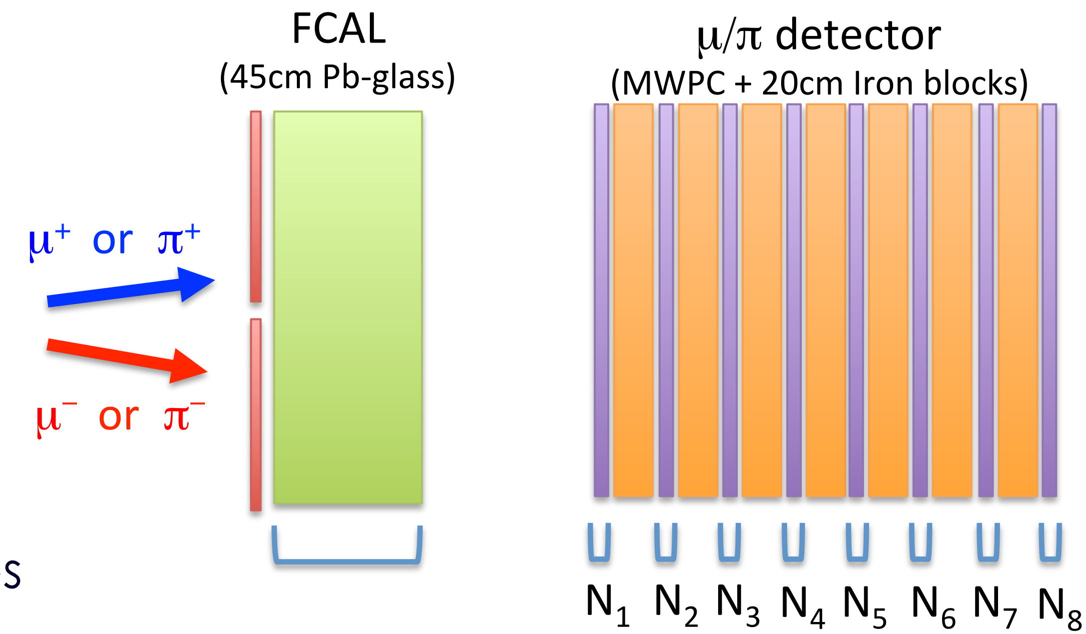
Multi-Variate Analysis for 2 GeV π+ and μ+



μ rejection at 0.998, π efficiency at 95%

PHASE II

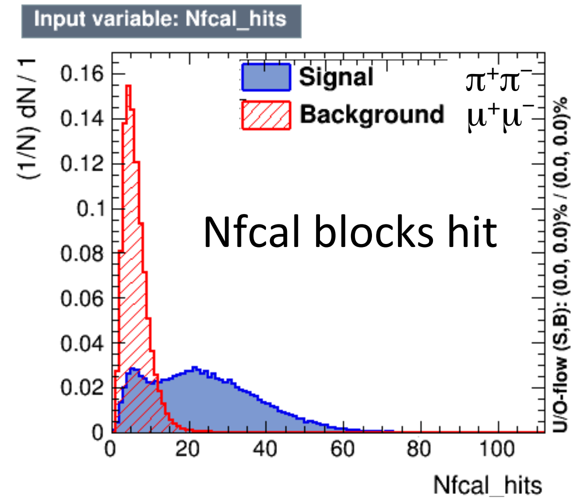
- $\pi^+\pi^-$ or $\mu^+\mu^-$ pairs with realistic momentum distributions
- Addition of beam background
- Additional input variables

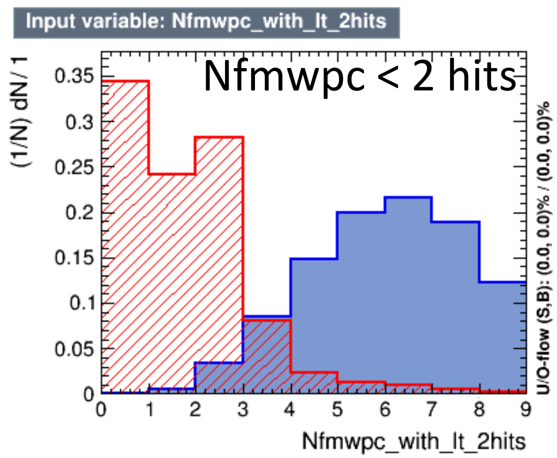


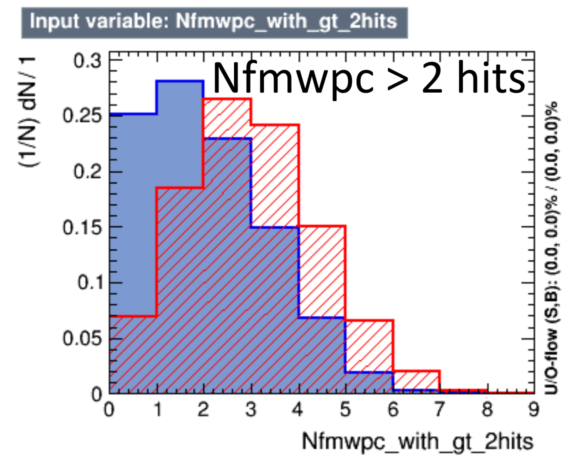

```
    fac->AddVariable("Ntracks",               'I');
    fac->AddVariable("Ntof",                  'I');
    fac->AddVariable("Nfcal_hits",            'I');
    fac->AddVariable("Nfcal_clusters",        'I');
    fac->AddVariable("Efcal_clusters",        'F');
    fac->AddVariable("Nfmwpc",                'I');
    fac->AddVariable("Nfmwpc1",               'I');
    fac->AddVariable("Nfmwpc2",               'I');
    fac->AddVariable("Nfmwpc3",               'I');
    fac->AddVariable("Nfmwpc4",               'I');
    fac->AddVariable("Nfmwpc5",               'I');
    fac->AddVariable("Nfmwpc6",               'I');
    fac->AddVariable("Nfmwpc7",               'I');
    fac->AddVariable("Nfmwpc8",               'I');
    fac->AddVariable("Nfmwpc_with_2hits",     'I');
    fac->AddVariable("Nfmwpc_with_lt_2hits",  'I');
    fac->AddVariable("Nfmwpc_with_gt_2hits",  'I');
```

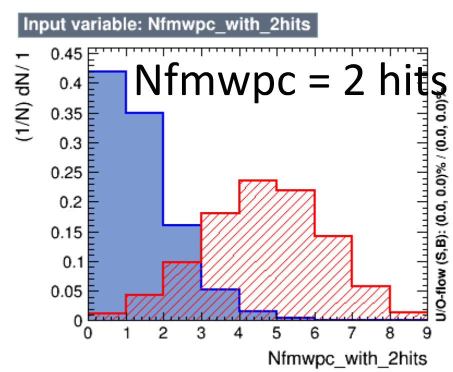
PHASE II

```
: Ranking input variables (method specific)...
: Ranking result (top variable is best ranked)
: -------------------------------------------------------------
: Rank : Variable                : Variable Importance
: -------------------------------------------------------------
:    1 : Nfcal_hits              : 9.746e-02
:    2 : Nfmwpc_with_lt_2hits    : 9.556e-02
:    3 : Nfmwpc6                 : 7.841e-02
:    4 : Nfmwpc8                 : 6.998e-02
:    5 : Nfmwpc                  : 6.997e-02
:    6 : Nfmwpc5                 : 6.676e-02
:    7 : Ntof                    : 5.887e-02
:    8 : Nfmwpc7                 : 5.870e-02
:    9 : Nfcal_clusters          : 5.715e-02
:   10 : Nfmwpc_with_2hits       : 5.478e-02
:   11 : Nfmwpc1                 : 5.187e-02
:   12 : Nfmwpc4                 : 4.410e-02
:   13 : Efcal_clusters          : 4.277e-02
:   14 : Nfmwpc2                 : 4.137e-02
:   15 : Ntracks                 : 4.084e-02
:   16 : Nfmwpc3                 : 3.714e-02
:   17 : Nfmwpc_with_gt_2hits    : 3.429e-02
: -------------------------------------------------------------
```



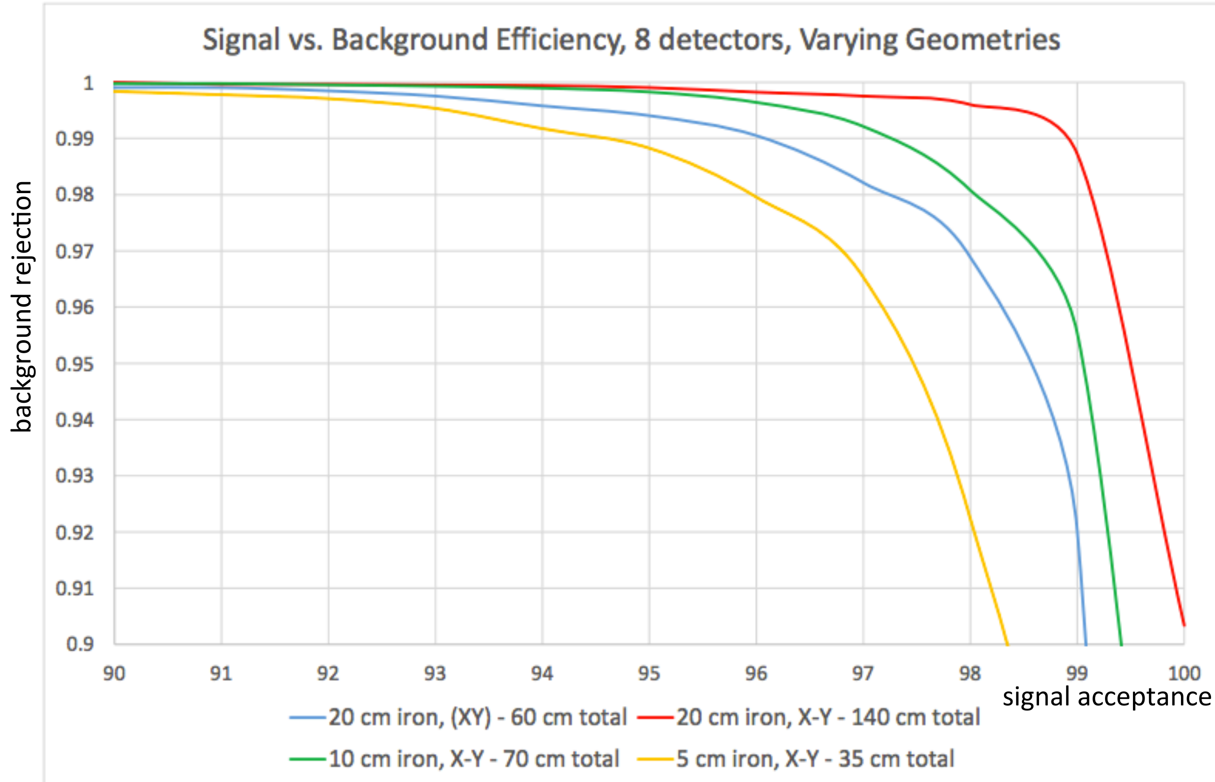






More Iron Is Better

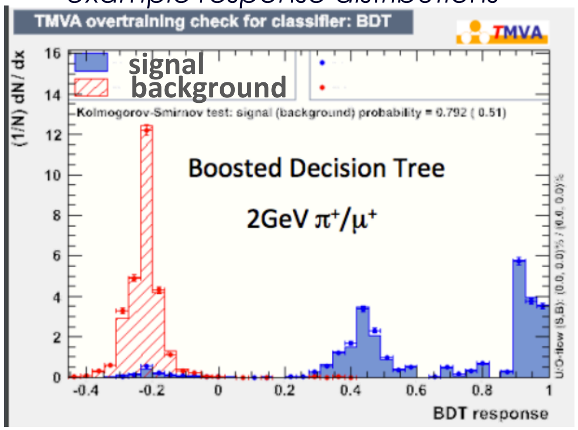
Question 1: Should we have 'xy-iron-xy-...', or 'x-iron-y-iron-x-iron-y- ... ?



Compare blue (xy-iron-xy...) and green (x-iron-y-iron...) : Green is better.
To do: adjust so total amount of iron is identical.

- Realistic simulation with both particles present and distribution of incident photon energies
- An MVA can be biased by your choice of inputs, but not by your interpretation

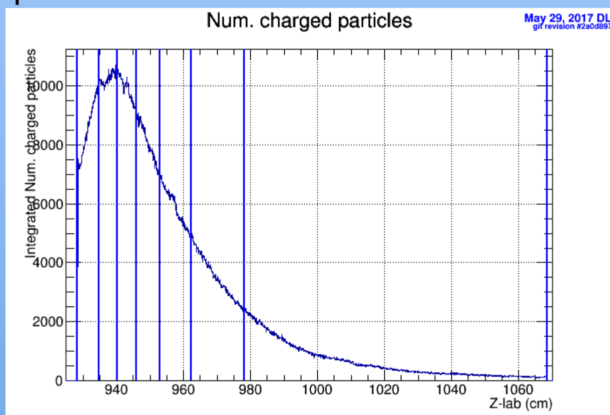
example response distributions

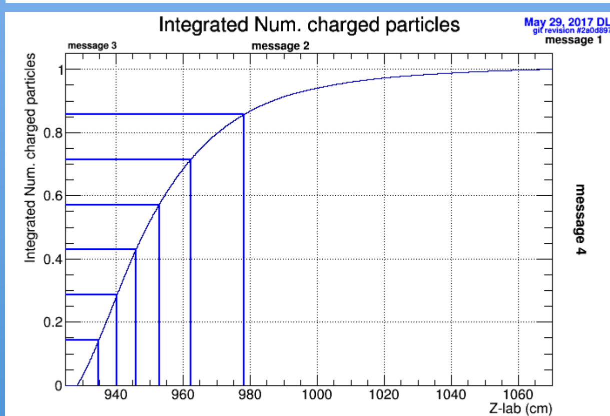


Study done by UMass undergrad. Bobby Johnston

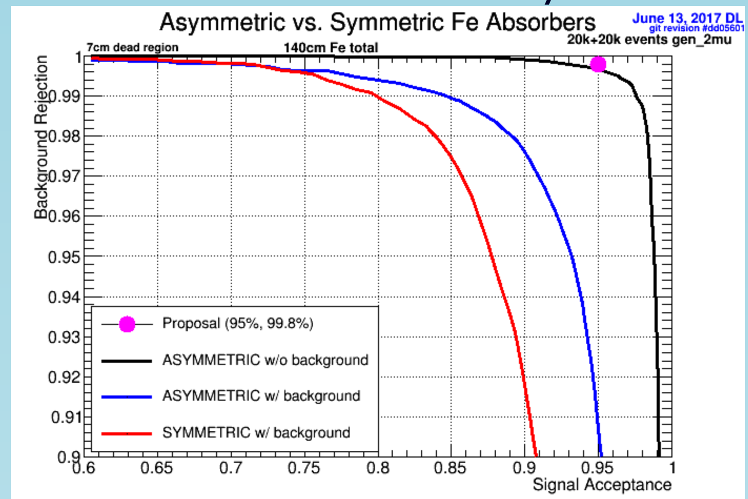
Symmetry of Iron Absorber Thickness

- ## Human Derived Concept
  - Integrate number of particles as function of depth in Iron for $\pi^{\pm}$ showers
  - Split Iron so sections contain equal number of particles



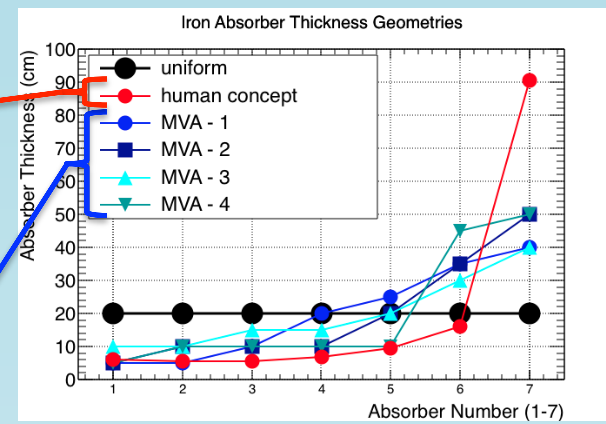


- ## Confirmation by MVA

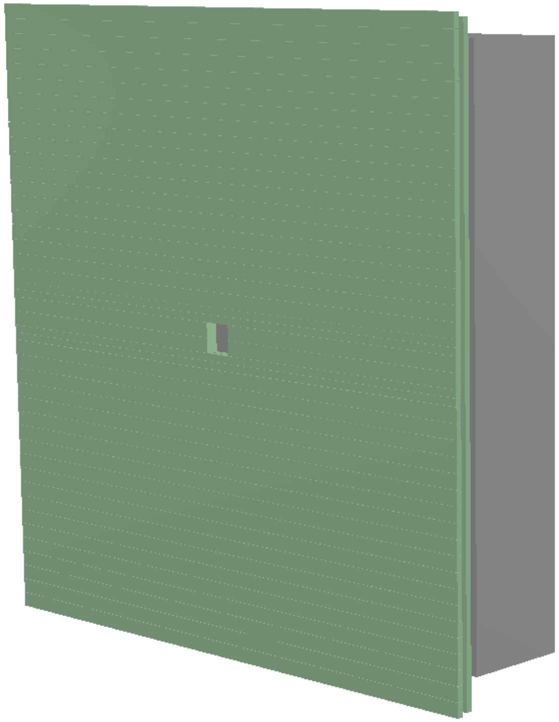


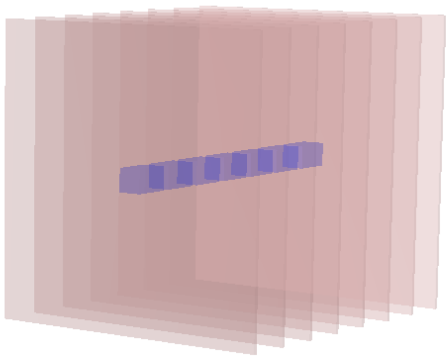
(human)

Top 4 MVA results of 741 geometries tested

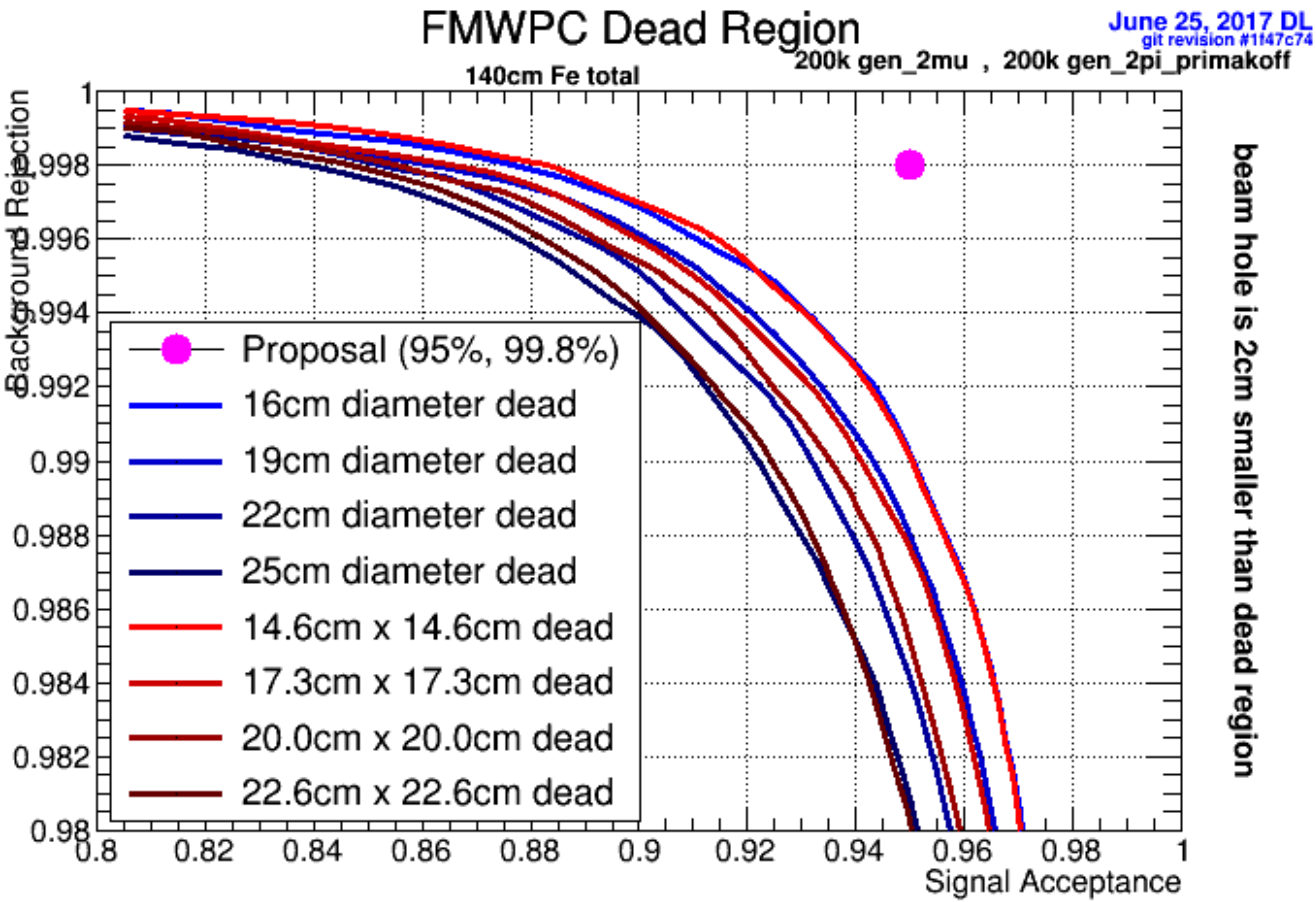


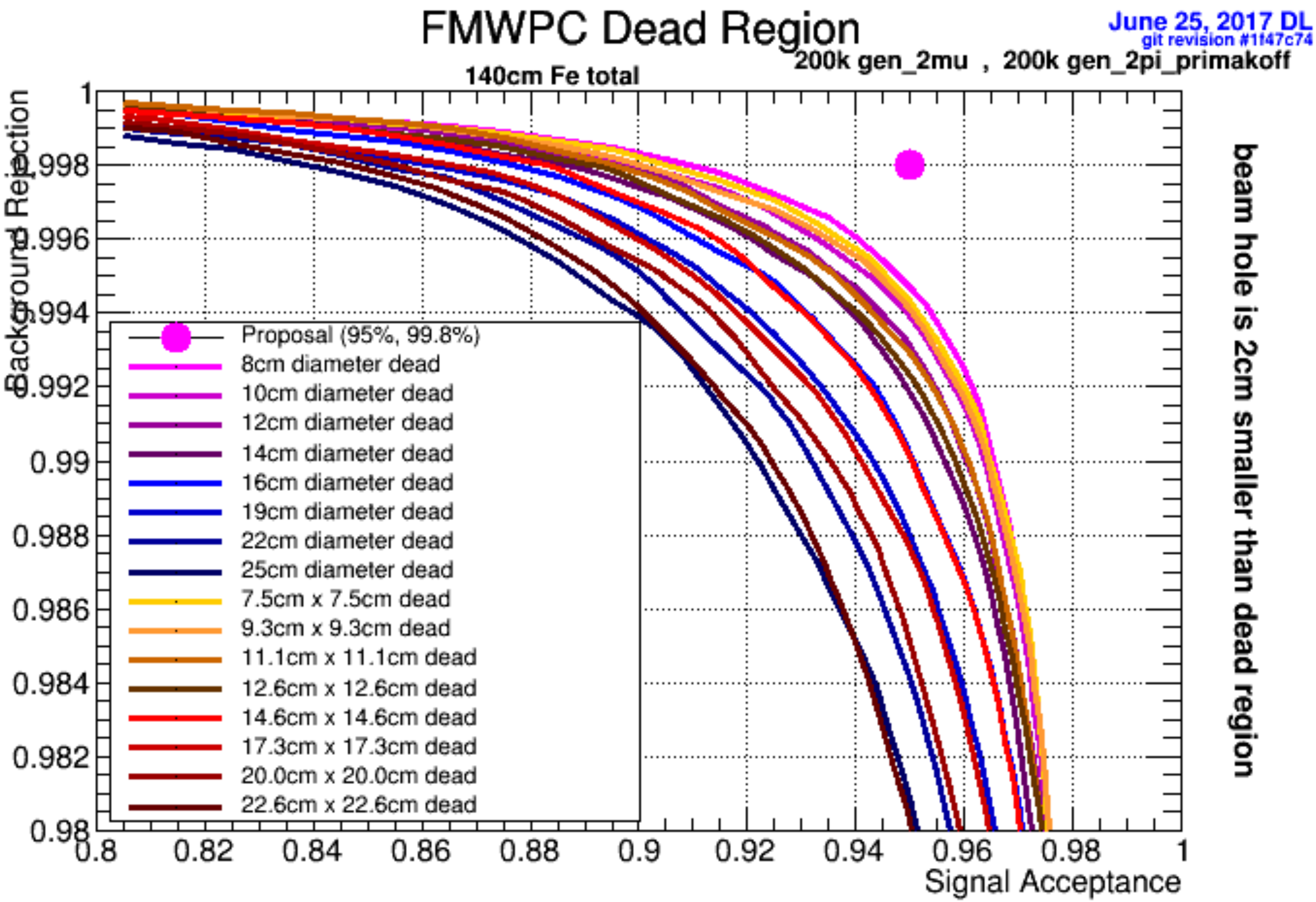
Beam Hole / Dead Region Geometry





- Primary photon beam passes through square holes in center of FCAL and TOF detectors
- Natural geometry of beam is round
- Should hole in Iron absorbers be square to match FCAL or round to match beam? Does it matter?
- What size should it be?
- What about dead region in chambers?


FMWPC Dead Region
June 25, 2017 DL
git revision #1f47c74
200k gen_2mu , 200k gen_2pi_primakoff
140cm Fe total
Background Rejection
Signal Acceptance
Proposal (95%, 99.8%)
16cm diameter dead
19cm diameter dead
22cm diameter dead
25cm diameter dead
14.6cm x 14.6cm dead
17.3cm x 17.3cm dead
20.0cm x 20.0cm dead
22.6cm x 22.6cm dead
beam hole is 2cm smaller than dead region
1
0.998
0.996
0.994
0.992
0.99
0.988
0.986
0.984
0.982
0.98
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1


FMWPC Dead Region
June 25, 2017 DL
git revision #1f47c74
200k gen_2mu , 200k gen_2pi_primakoff
140cm Fe total
Background Rejection
Signal Acceptance
beam hole is 2cm smaller than dead region
Proposal (95%, 99.8%)
8cm diameter dead
10cm diameter dead
12cm diameter dead
14cm diameter dead
16cm diameter dead
19cm diameter dead
22cm diameter dead
25cm diameter dead
7.5cm x 7.5cm dead
9.3cm x 9.3cm dead
11.1cm x 11.1cm dead
12.6cm x 12.6cm dead
14.6cm x 14.6cm dead
17.3cm x 17.3cm dead
20.0cm x 20.0cm dead
22.6cm x 22.6cm dead
1
0.998
0.996
0.994
0.992
0.99
0.988
0.986
0.984
0.982
0.98
0.8
0.82
0.84
0.86
0.88
0.9
0.92
0.94
0.96
0.98
1

AN MVA SURPRISE

- In 2013 we had a software tutorial for GlueX
  - $\gamma p \rightarrow p\pi^+\pi^+\pi^-$
  - **Signal**: specialized generator with selected amplitudes
    - 5343 events
  - **Background**: pythia + selected reactions at low energy with uniform sampling in phase space
    - 8527 events

AN MVA SURPRISE

Original Exercise

```
: -------------------------------------------------------
: Rank : Variable              : Variable Importance
: -------------------------------------------------------
:   1 : PV_r                   : 1.129e-01
:   2 : MissingNeutron_PT      : 9.881e-02
:   3 : PiPlus1__Timing_FOM    : 7.881e-02
:   4 : PiMinus__Timing_FOM    : 7.325e-02
:   5 : Unused__Max_KPlus_FOM  : 7.313e-02
:   6 : Unused__Max_Proton_FOM : 6.612e-02
:   7 : FOM_KinFit             : 6.467e-02
:   8 : PiPlus2__Timing_FOM    : 6.264e-02
:   9 : PiPlus1__NDF_Tracking  : 5.885e-02
:  10 : PiPlus2__NDF_Tracking  : 5.744e-02
:  11 : PiPlus2__DCdEdx_FOM    : 5.722e-02
:  12 : Unused__Max_KMinus_FOM : 5.641e-02
:  13 : PiMinus__NDF_Tracking  : 5.075e-02
:  14 : PiMinus__DCdEdx_FOM    : 5.054e-02
:  15 : PiPlus1__DCdEdx_FOM    : 3.848e-02
: -------------------------------------------------------
```

Added extra variable "just for fun"

```
: -------------------------------------------------------
: Rank : Variable              : Variable Importance
: -------------------------------------------------------
:   1 : PV_r                   : 9.909e-02
:   2 : Entry$                 : 9.189e-02
:   3 : MissingNeutron_PT      : 9.053e-02
:   4 : PiPlus1__Timing_FOM    : 8.114e-02
:   5 : PiMinus__Timing_FOM    : 7.219e-02
:   6 : Unused__Max_KPlus_FOM  : 6.512e-02
:   7 : Unused__Max_Proton_FOM : 6.133e-02
:   8 : PiPlus2__NDF_Tracking  : 5.963e-02
:   9 : FOM_KinFit             : 5.823e-02
:  10 : PiPlus2__Timing_FOM    : 5.629e-02
:  11 : PiPlus1__NDF_Tracking  : 5.361e-02
:  12 : Unused__Max_KMinus_FOM : 5.358e-02
:  13 : PiMinus__NDF_Tracking  : 4.415e-02
:  14 : PiMinus__DCdEdx_FOM    : 4.318e-02
:  15 : PiPlus2__DCdEdx_FOM    : 3.784e-02
:  16 : PiPlus1__DCdEdx_FOM    : 3.221e-02
: -------------------------------------------------------
```

Some advice

1. Any variable could be useful. Just give it to the MVA and let it decide
2. Avoid variables that are strongly correlated with a quantity you're trying to measure
   (e.g. inv. mass of a resonance you're trying to search for)
3. Make any obvious cuts before handing over to MVA so it doesn't have to waste discriminating power on the obvious

SUMMARY

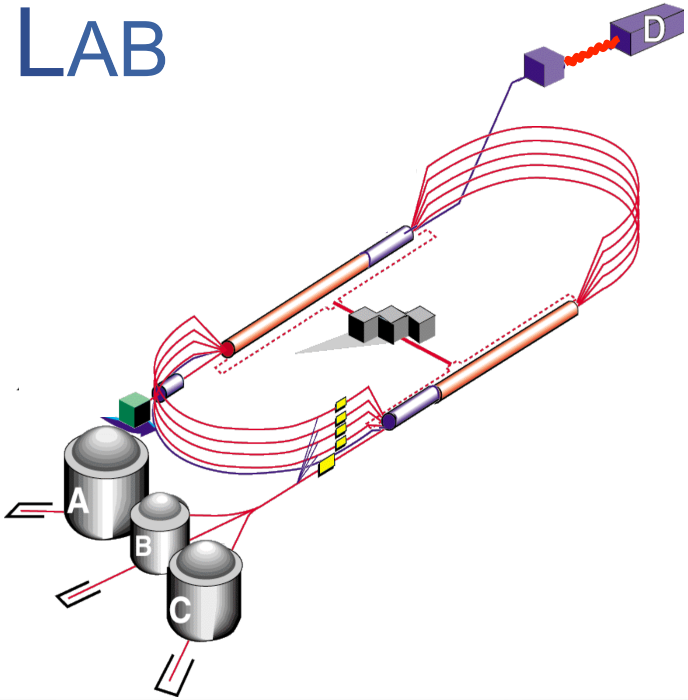
- Using multiple MVA algorithms can allow quick insight into design decisions without much expertise in machine learning
  - The TMVA package in ROOT is a great way to get started with machine learning and gives easy access to several algorithms using a single input format and API
- MVA is a great way to make relative comparisons between different detector designs
- The Charged Pion Polarizability experiment at Jefferson Lab Hall-D has used TMVA to refine several of the π/μ detector design aspects

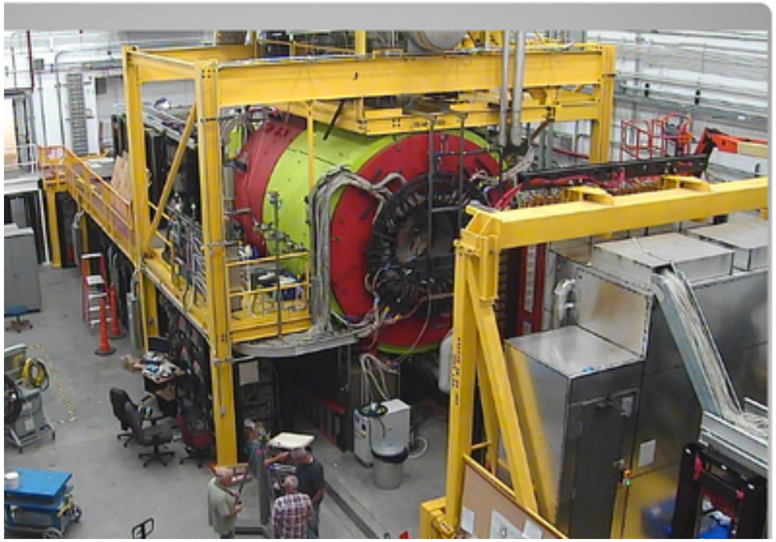
Backups

Hall-D at Jefferson Lab

(home of GlueX)

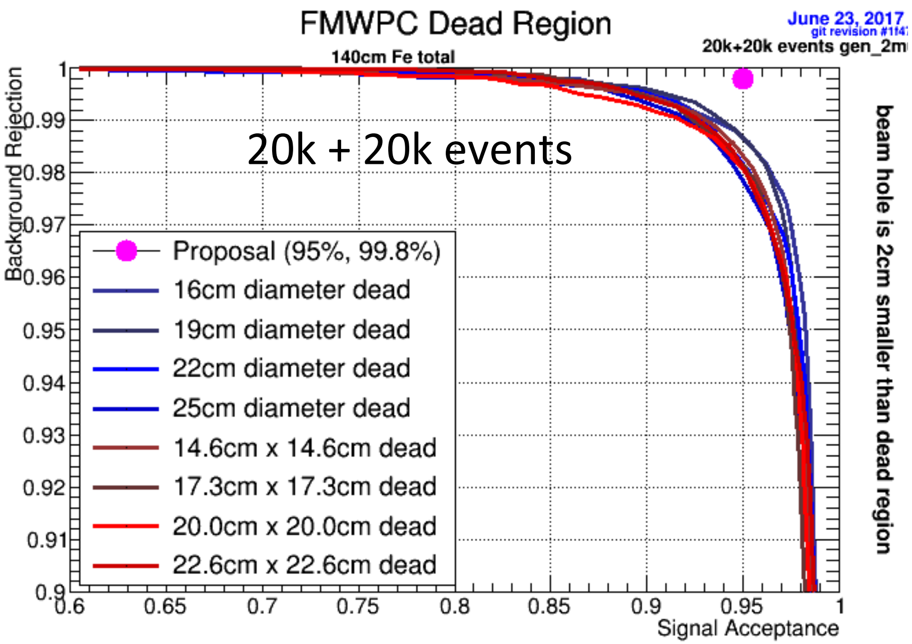
- 12GeV CW(2-4ns) electron beam
- Only high energy photons enter Hall-D

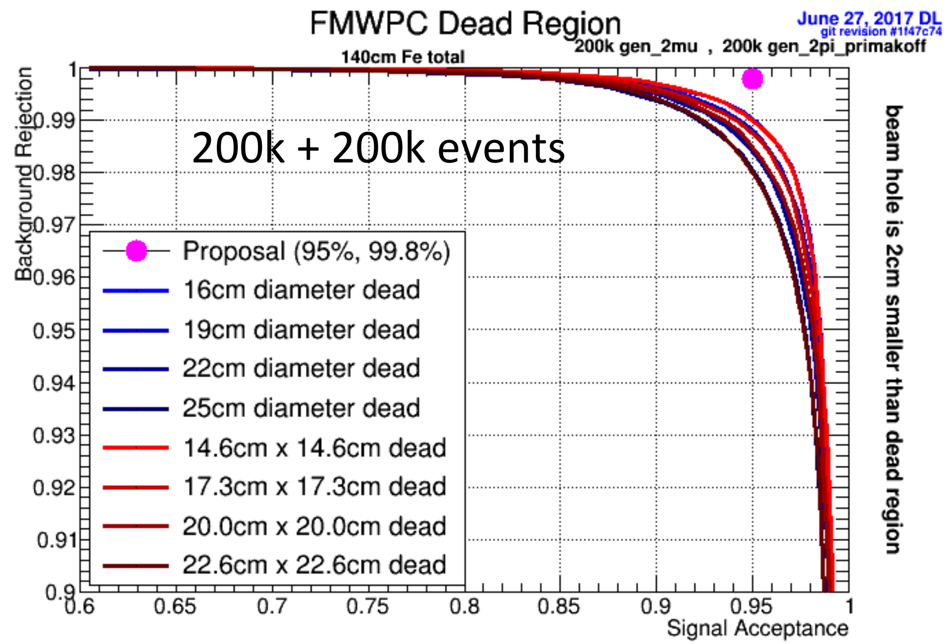




Jefferson Lab

20k -> 200k


FMWPC Dead Region
June 23, 2017
20k+20k events gen_2m
140cm Fe total
20k + 20k events
Background Rejection
Signal Acceptance
Proposal (95%, 99.8%)
16cm diameter dead
19cm diameter dead
22cm diameter dead
25cm diameter dead
14.6cm x 14.6cm dead
17.3cm x 17.3cm dead
20.0cm x 20.0cm dead
22.6cm x 22.6cm dead
beam hole is 2cm smaller than dead region


FMWPC Dead Region
June 27, 2017 DL
git revision #1f47c74
200k gen_2mu , 200k gen_2pi_primakoff
140cm Fe total
200k + 200k events
Background Rejection
Signal Acceptance
Proposal (95%, 99.8%)
16cm diameter dead
19cm diameter dead
22cm diameter dead
25cm diameter dead
14.6cm x 14.6cm dead
17.3cm x 17.3cm dead
20.0cm x 20.0cm dead
22.6cm x 22.6cm dead
beam hole is 2cm smaller than dead region