Low Energy Reconstruction with Pandora: Handling backgrounds

Matthew Osbiston

Wednesday 1st May | LEWG Meeting

Moving to **phase 2**

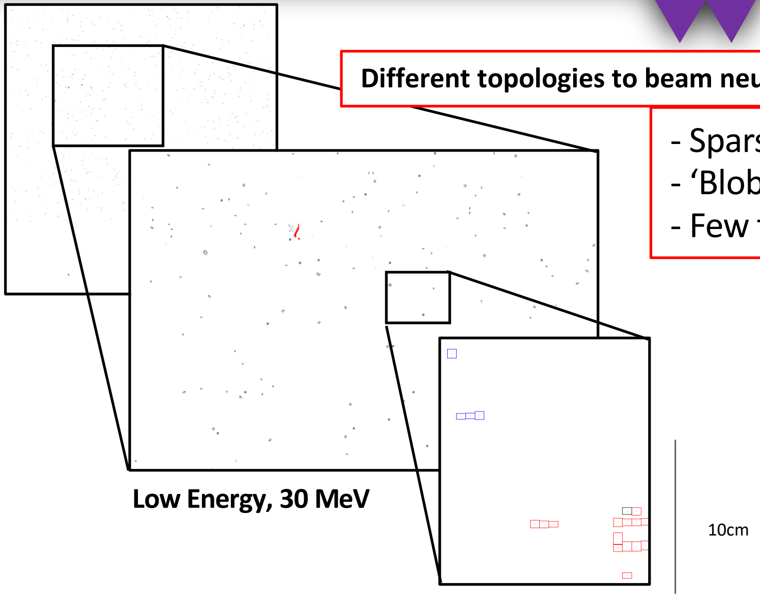
Previous Pandora Development: Beam neutrino interactions

Now: Low energy neutrino interactions

1. Improve the pattern recognition for **individual low energy neutrino** events

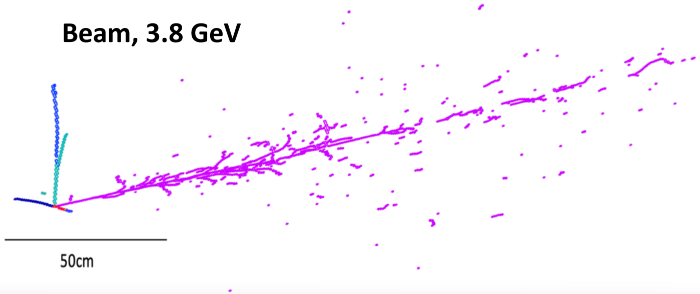
2. Adapt reconstruction to handle a more realistic readout from FD **with backgrounds**

3. Explore **analysis-reconstruction continuum** - exploit tailored reconstruction and use physics focused analysis to refine performance



Different topologies to beam neutrinos

- Sparse hits
- 'Blob-like'.
- Few tracks



Low Energy Reconstruction Algorithms: Work plan

1. Trained a **DL neural network** for **vertex identification** using MARLEY samples (5-30 MeV) at high stats (50k)

2. Retuned **2D to 3D cluster matching** algorithms to match low energy topologies and reconstruct particles – lowered threshold criteria

3. Wrote new algorithm using a **BDT** to drive **2D clustering decisions**, trained on low energy samples – separate **photons** and **electrons** in to individual objects

Now:

1. Trained a **DL neural network** for **signal vs background separation** using new MARLEY samples (5-70 MeV) at low stats (3.7k)

DLSignal Algorithm Development: New Sample production

LArSoft Version

```
version = v09_82_02d01
```

Config file fhicls: ['-c', '/exp/dune/app/users/osbiston/POMS/config.cfg'] ← Find more on input here

```
gen_fcl =  prodmarley_nue_flat_radiological_decay0_dunevd10kt_1x8x14_3view_30deg.fcl
g4_stage1_fcl =  supernova_g4stage1_dunevd10kt_1x8x14_3view_30deg.fcl
g4_stage2_fcl =  standard_g4stage2_dunevd10kt_1x8x14_3view_30deg.fcl
detsim_fcl= standard_detsim_dunevd10kt_1x8x14_3view_30deg.fcl
reco1_fcl= standard_reco1_dunevd10kt_1x8x14_3view_30deg.fcl
reco2_fcl= matt.fcl
```

matt.fcl

```
#include "standard_reco2_dunevd10kt_1x8x14_3view_30deg.fcl"
process_name: RecoPandora
# Use custom settings file
physics.producers.pandora.ConfigFile: "MyPandoraSettings_Master_Standard.xml"
physics.reco: [ @sequence::dunefd_vertdrift_tpc_reco2 ]
```

MyPandoraSettings_Master_Standard.xml

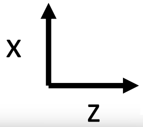
```
<pandora>
    <!-- GLOBAL SETTINGS -->
    <IsMonitoringEnabled>true</IsMonitoringEnabled>
    <ShouldDisplayAlgorithmInfo>false</ShouldDisplayAlgorithmInfo>
    <SingleHitTypeClusteringMode>true</SingleHitTypeClusteringMode>

    <!-- ALGORITHM SETTINGS -->
    <algorithm type = "LArEventWriting">
        <EventFileName>marley_flat_1x8x14_5_70MeV_run5.pndr</EventFileName>
        <ShouldWriteEvents>true</ShouldWriteEvents>
        <ShouldOverwriteEventFile>true</ShouldOverwriteEventFile>
        <ShouldWriteMCRelationships>true</ShouldWriteMCRelationships>
        <ShouldWriteTrackRelationships>true</ShouldWriteTrackRelationships>
        <GeometryFileName>Pandora_Geometry.xml</GeometryFileName>
        <ShouldWriteGeometry>false</ShouldWriteGeometry>
        <ShouldOverwriteGeometryFile>true</ShouldOverwriteGeometryFile>
        <LArCaloHitVersion>2</LArCaloHitVersion>
    </algorithm>
```

3730 (20 GB) events total

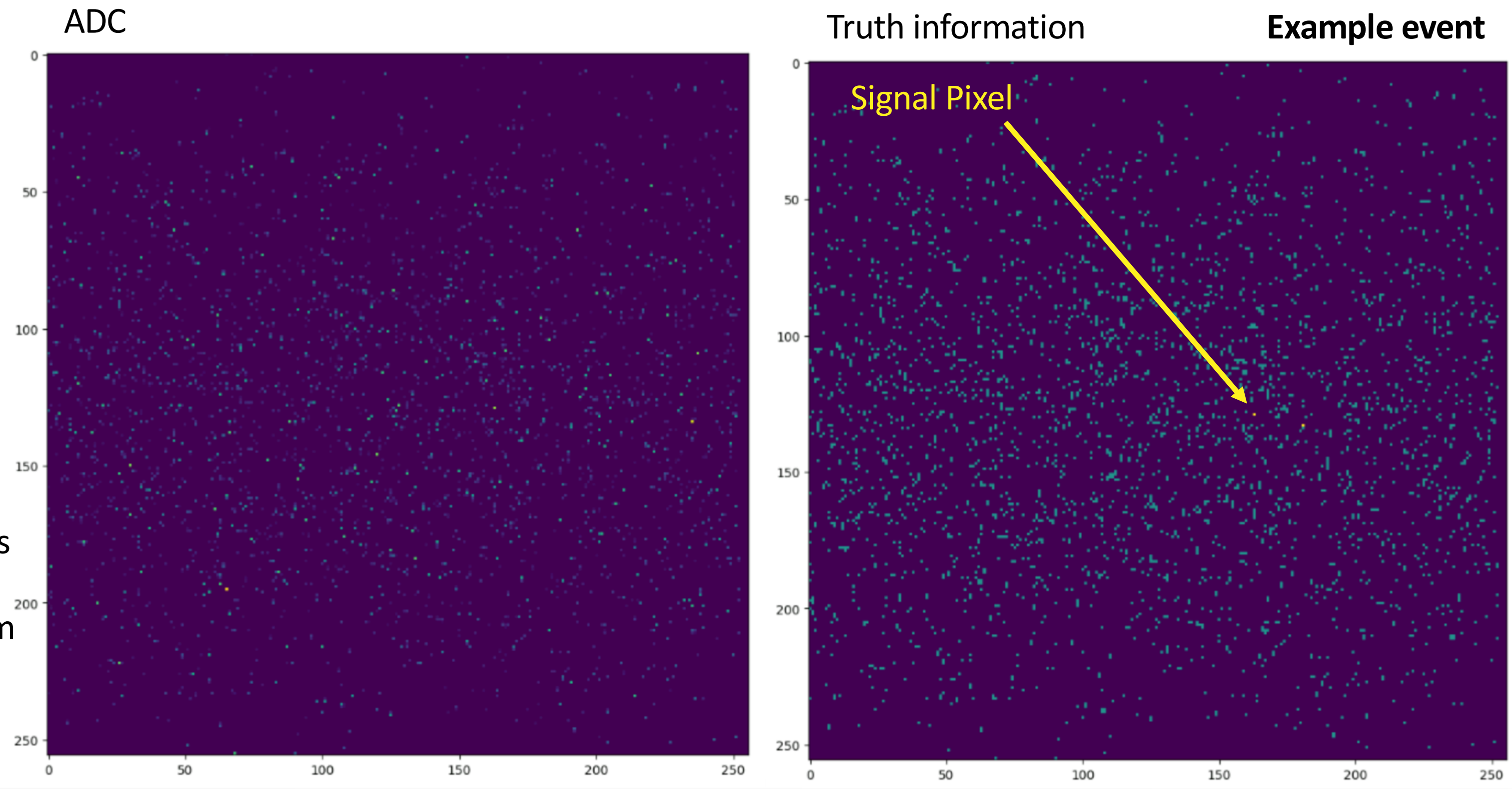
MARLEY 5-70 MeV with CC and ES int. in 1x8x14 VD geometry with backgrounds (decay0)

DLSignal Algorithm Development: New Sample production

- Event shown in 256x256 pixel display
- Detector dimensions: 650cm x 2090.9cm
- Wire Pitch: 0.51cm in x 0.76cm in z
- Each pixel here represents several hits
- Pixel: 2.54cm x 8.17cm



Example event



DLSignal Algorithm Development : NN Training

The network:

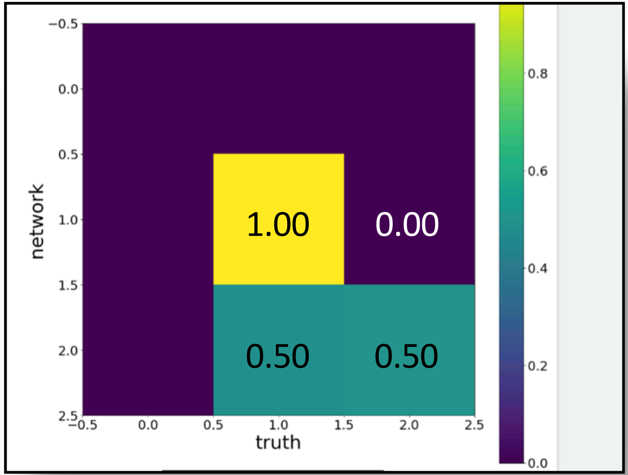
- Same architecture as vertexing network
- Using Andy Chappell's original notebook structure
- Classes: 0 – Null, 1- Background, 2 – Signal
- Weighting to account for uneven representation for signal/background/null pixels

Initial result:

N

B

S



Null Background Signal

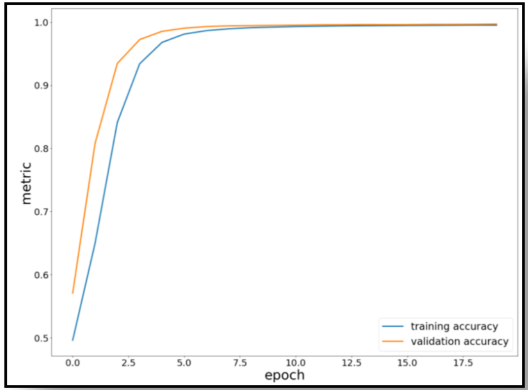
Network Training:

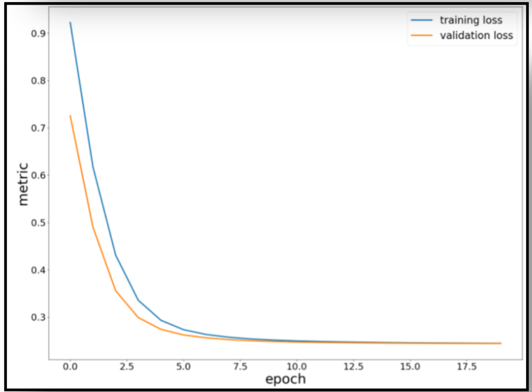
Null: 54,200,000 pixels

Background: 2,730,00 pixels

Signal: 6,400 pixels

Total events: 730



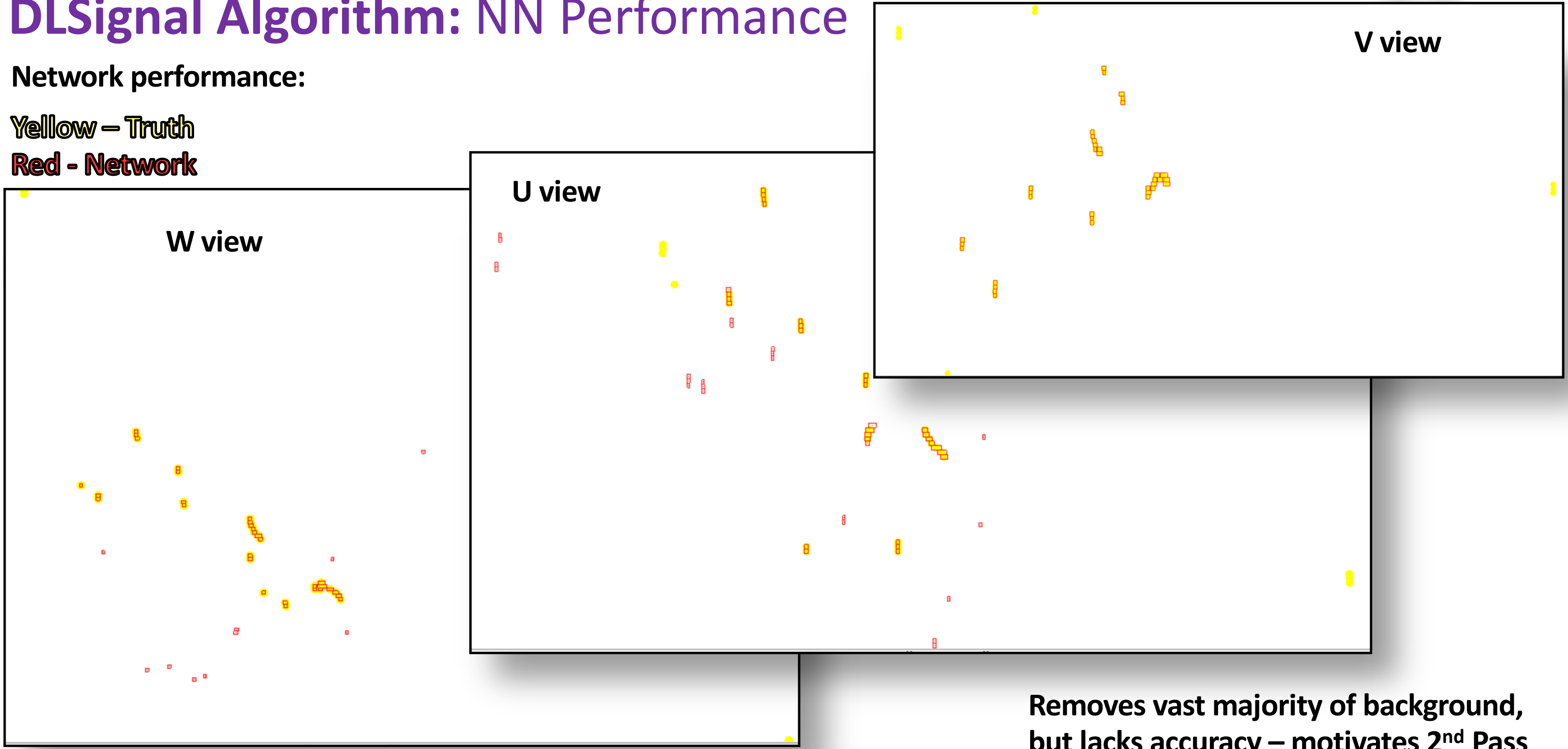


DLSignal Algorithm: NN Performance

Network performance:

Yellow – Truth

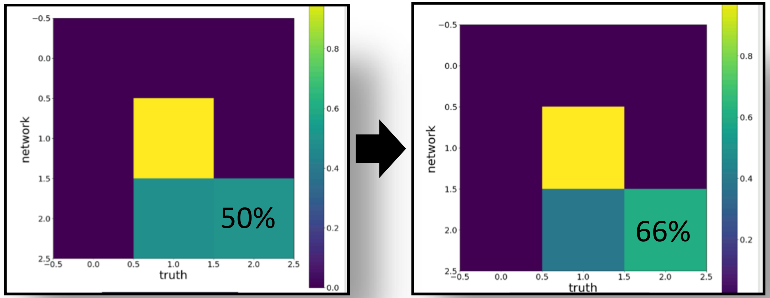
Red - Network

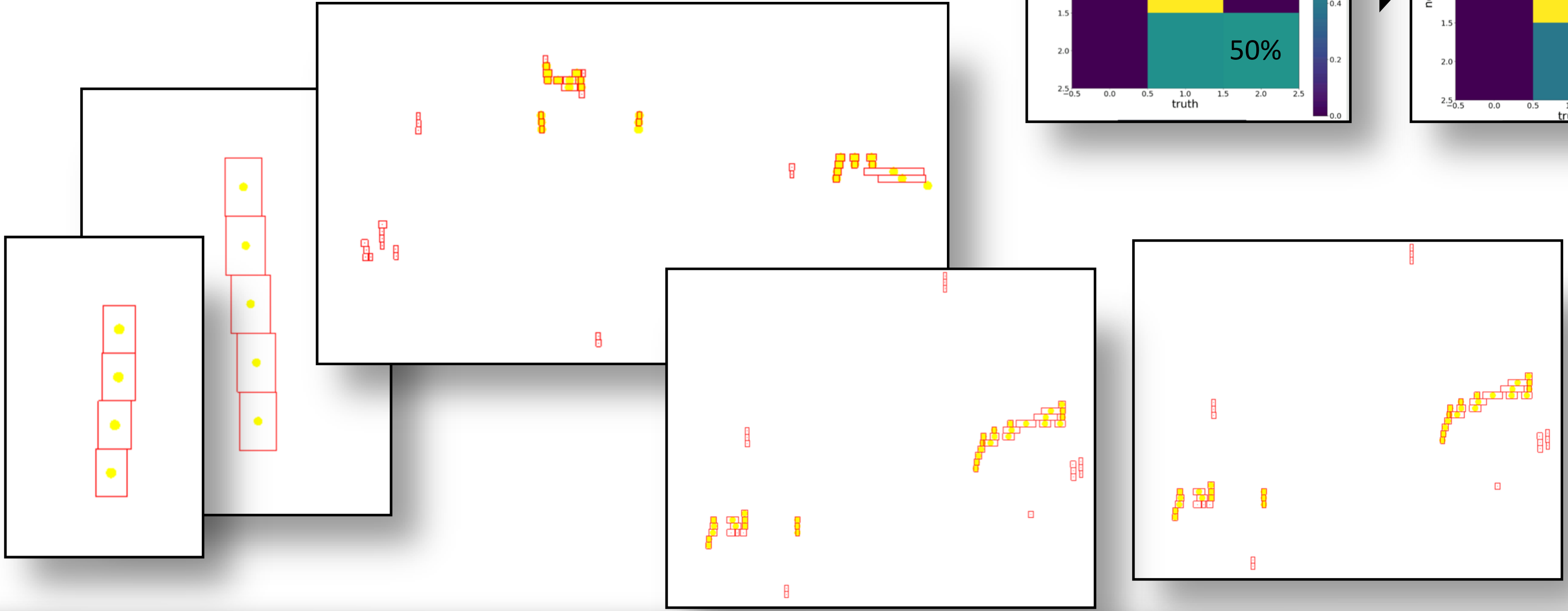


Removes vast majority of background, but lacks accuracy – motivates 2nd Pass

DLSignal Algorithm: NN Performance

Refinement 1:
Train on higher stats *(3.7k vs 640)* and adjusting the weight function *(upweights lower represented classes)* for equal representation



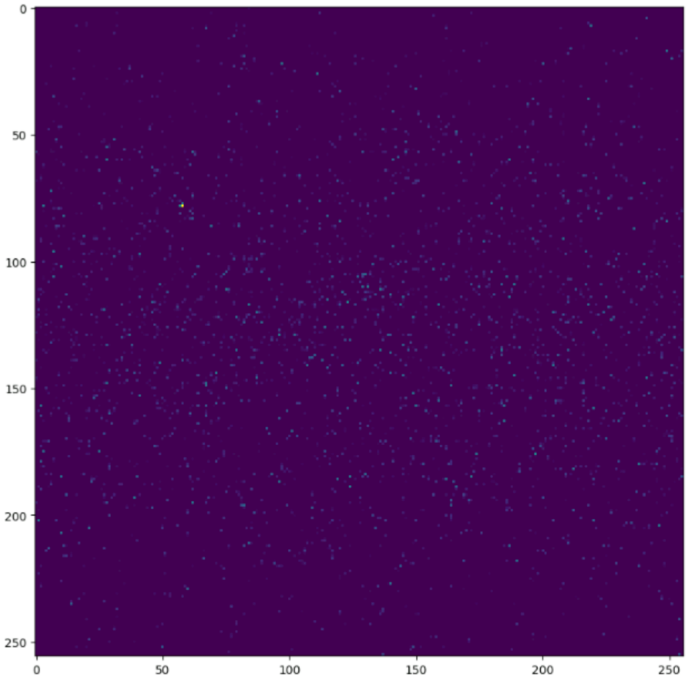


DLSignal Algorithm: NN Training two pass approach

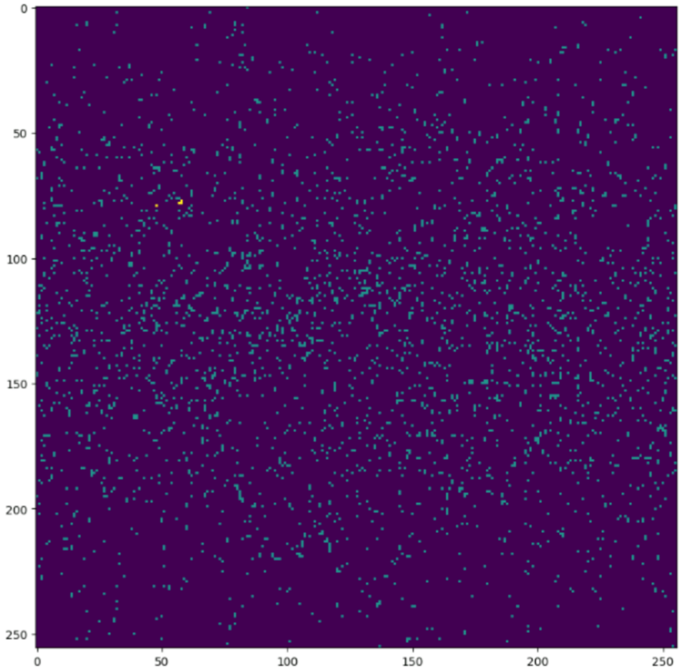
Pass 1

Identifies Region of Interest

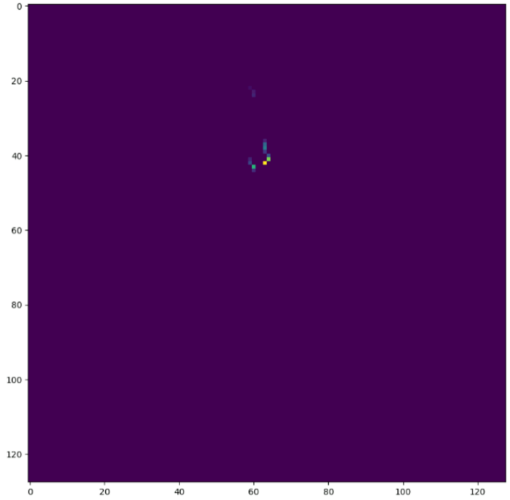
ADC



Truth



Network Identified "Signal" Output Pixels (Hits)



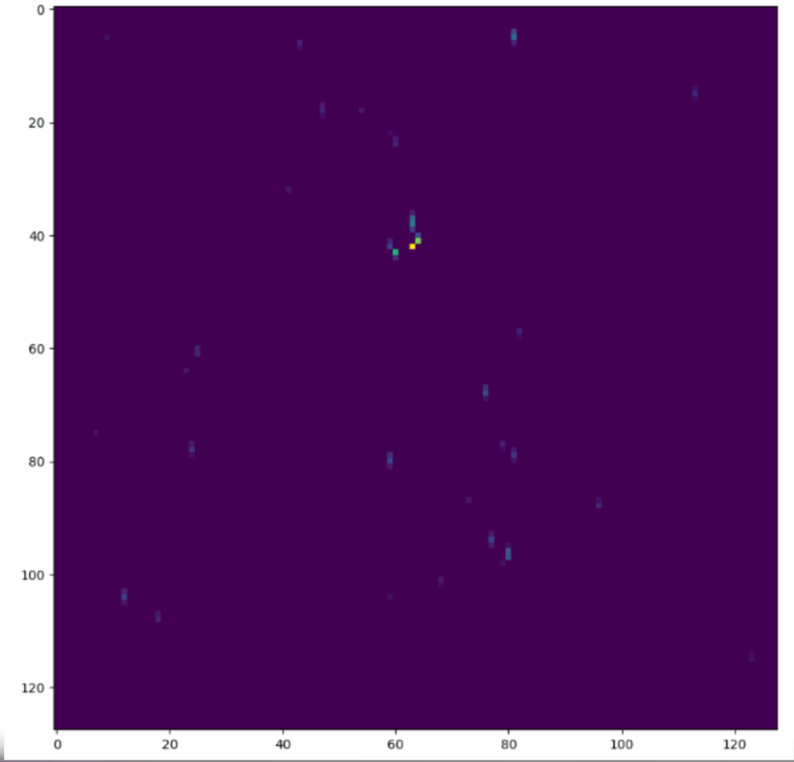
DLSignal Algorithm: NN Training two pass approach

Pass 2

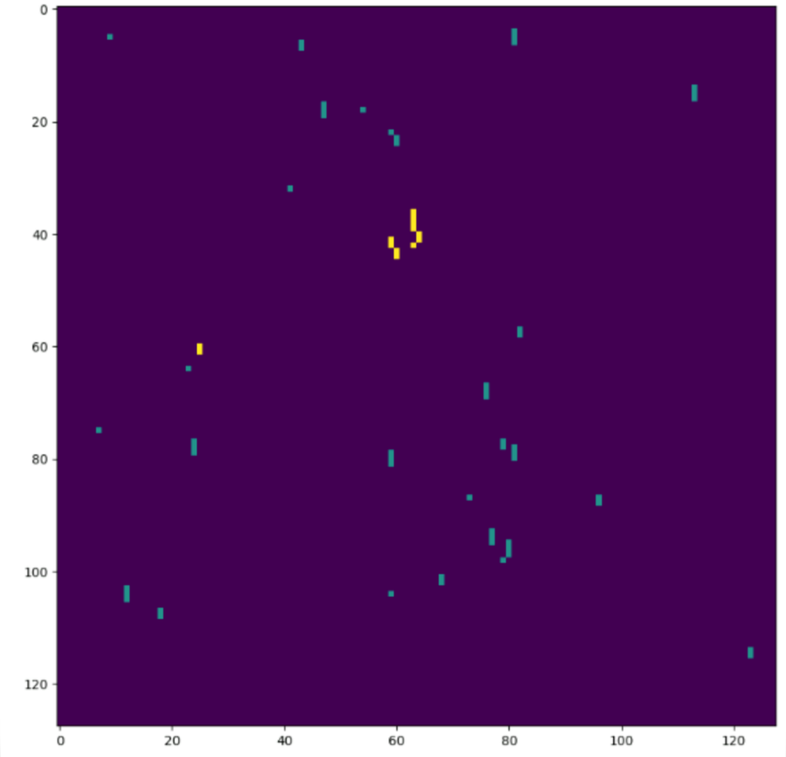
Zooms in for direct hit-to-pixel correspondence

Note: All hits in proximity of area of interest fed back in to Pass 2 to try to collect as many sparse signal hits as possible

ADC



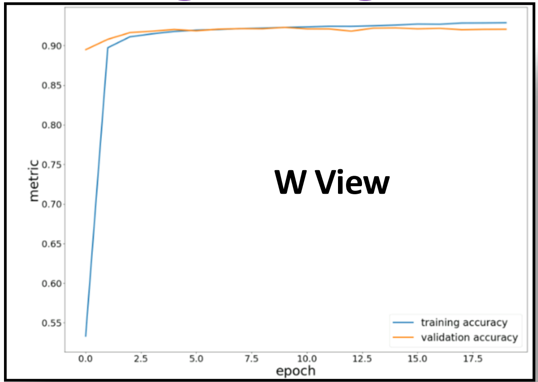
Truth

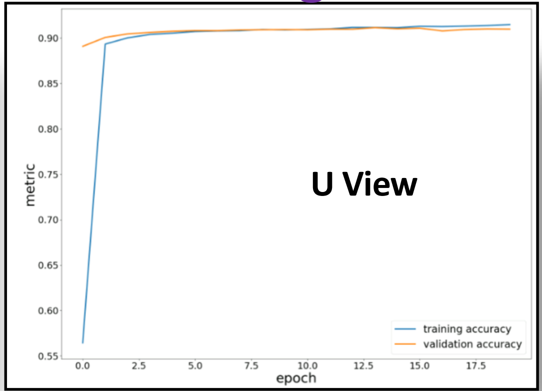


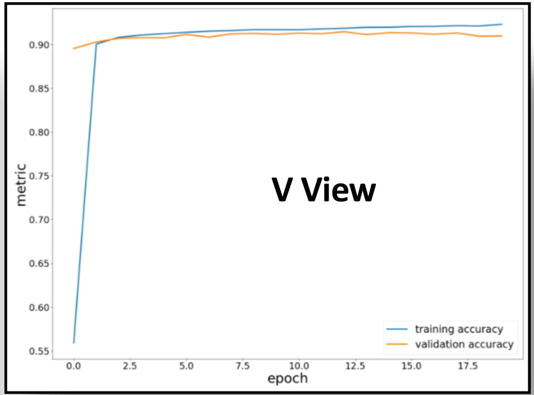
DLSignal Algorithm: NN Training Pass 2

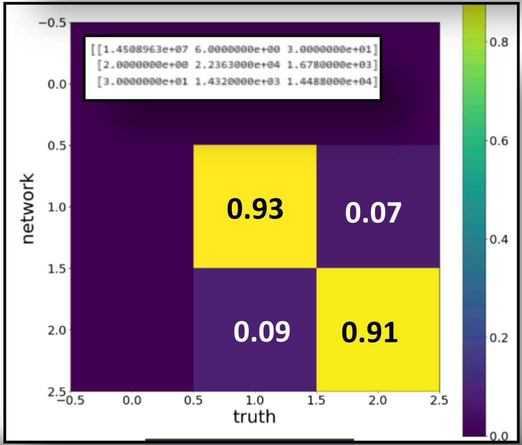
Much better network performance

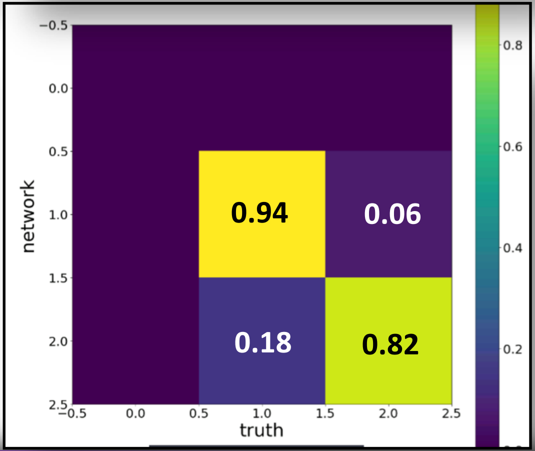
Pass 2

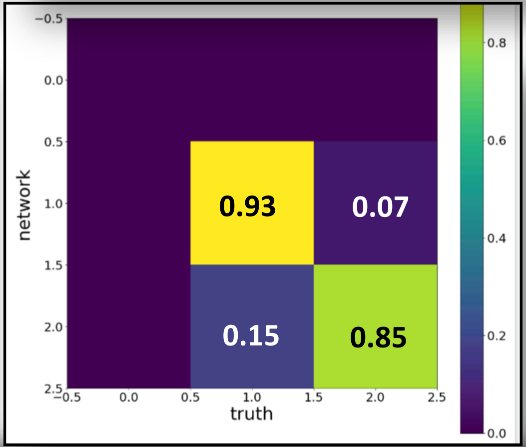












DLSignal Algorithm Development:

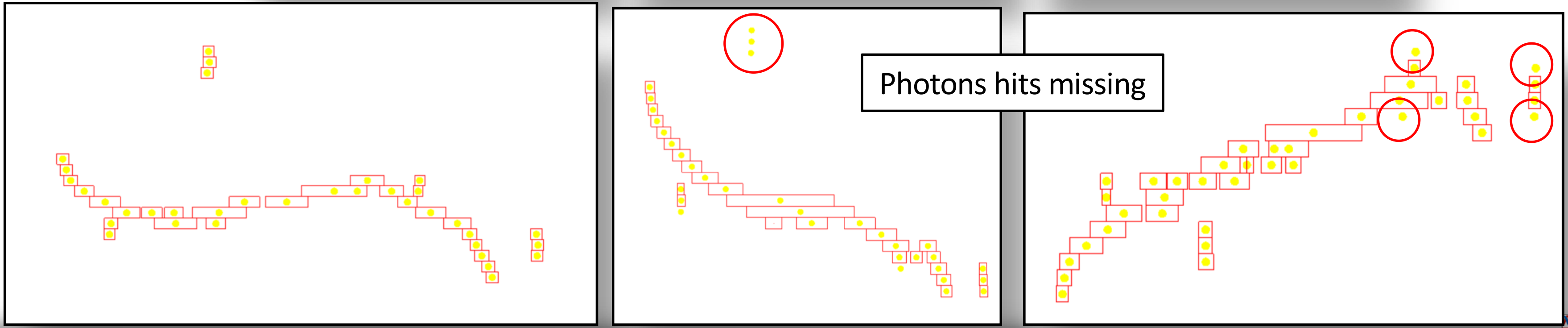
- Can this be pushed further?
- Seems like lots of the photons aren't identified?
- Perhaps create a distinct class for photons to allow network to learn more about them

Yellow – Truth
Red - Network

Pass 1

Background hits remaining

Pass 2



DLSignal Algorithm: NN Training Pass 2

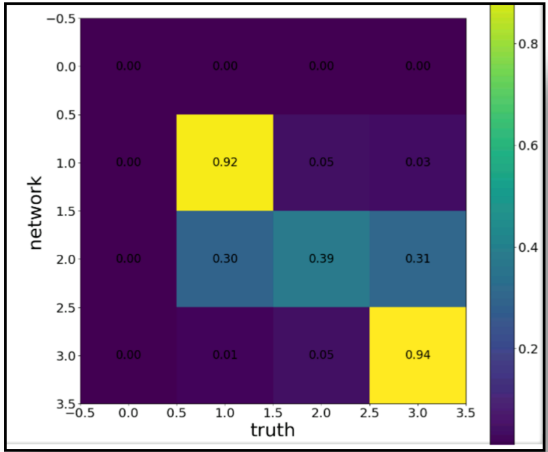
Update: Modified pass 2 to include electron (class 3) and photon (class 2) class. Both of these classes are added to the same CaloHitList

Pass 2

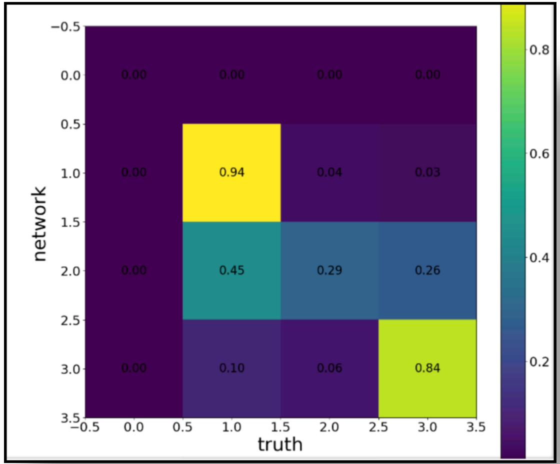
W View	Null	Background	Photon	Electron
Weights	0.94	4.69	8.03	5.47
Stats	39,000,000	63,400	7,320	34,100

U View	Null	Background	Photon	Electron
Weights	0.99	4.08	8.29	5.90
Stats	40,100,000	139,000	8,180	31,700

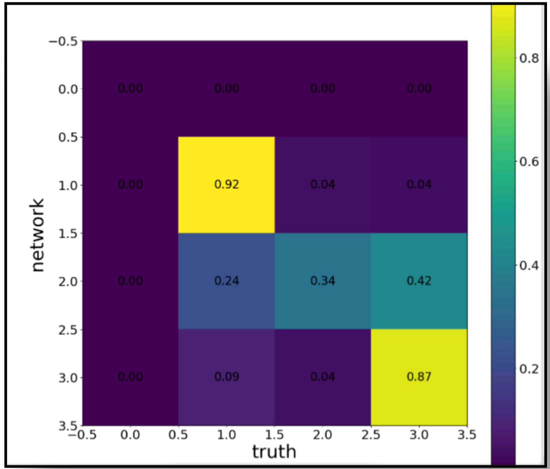
V View	Null	Background	Photon	Electron
Weights	1.01	4.26	8.31	5.90
Stats	38,000,000	120,000	8,340	32,800




```
[[12940283.       13.        4.       27.]
 [       1.    21166.     1055.      800.]
 [       0.      651.      859.      683.]
 [      26.      125.      494.     9941.]]
```



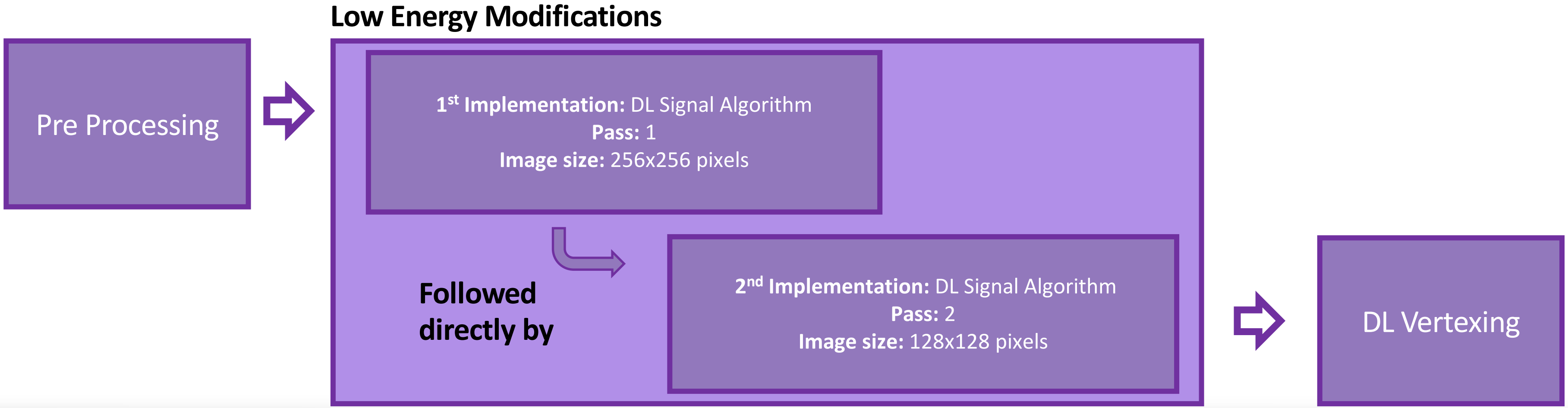

```
[[13443685.       22.        3.       32.]
 [       9.    41779.     1912.     1592.]
 [       0.      204.      293.      365.]
 [      49.      900.      465.     9106.]]
```




```
[[12656922.       32.        7.       29.]
 [      11.    41929.     1584.     1320.]
 [       0.      734.      481.      425.]
 [       7.     1053.      604.     8846.]]
```

DLSignal Algorithm : Implementation in reconstruction chain

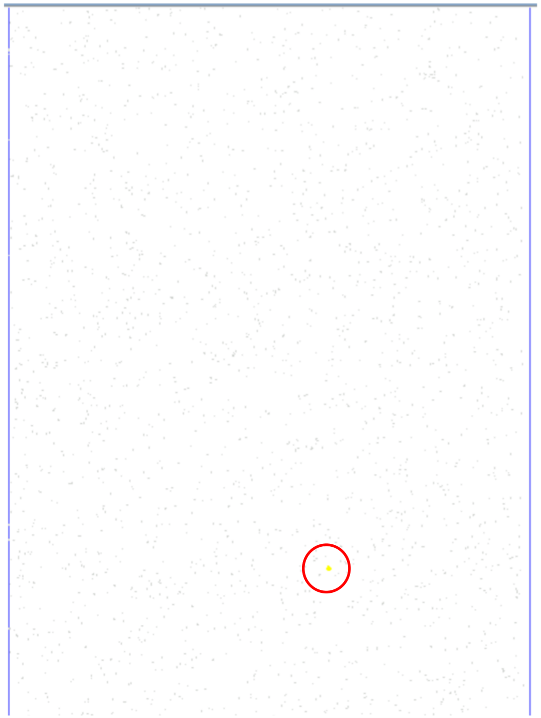
- Implemented at beginning of reconstruction chain
- Runs in two separate instances
- First instances runs pass one training of the NN over 256x256p and outputs a new CaloHitList
- Second Instance instances runs pass two training of the NN over 128x128p and outputs a new CaloHitList for each view and background CaloHitList

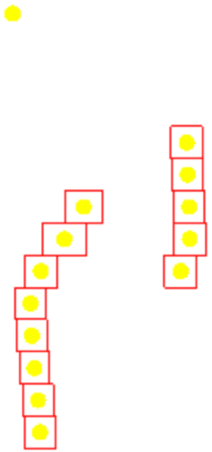


DLSignal Algorithm : Results and refinement

DLSignalAlg in action!

Yellow – Truth
Red - Network



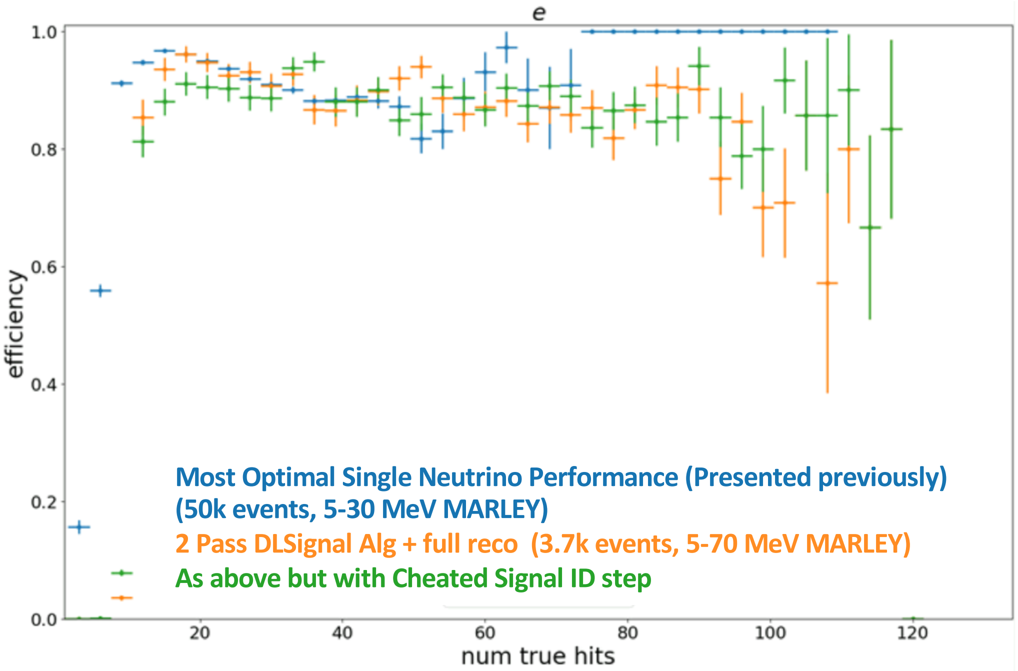


Reconstruction with LowE modifications shown before

DLSignal Algorithm : Final Metric Results

- Impressive efficiency for electron reconstruction – 3.7k events included

(note 50% purity and 50% completeness requirement for reach reconstructed particle)



Reconstruction – Analysis Continuum: What comes next?

- **Aim:** Maximise separation of signal and background energy deposition in detector

- Conduct **Low Energy Analysis** using *remastered reconstruction* to drive **real-world considerations**
  - Make fits of reconstructed samples compared to supernova model expectations

Developed in conjunction with Low Energy Working Group

EXAMPLE ANALYSIS

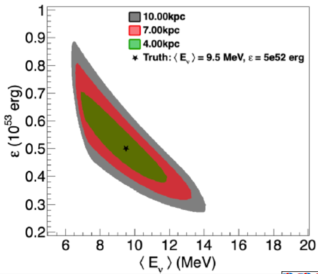
Astrophysics of core collapse

- Supernova spectral parameter fits
  - Fit to the SN spectral parameters

$$\phi(E_\nu) = \mathcal{N}\left(\frac{E_\nu}{\langle E_\nu\rangle}\right)^\alpha \exp\left[-(\alpha+1)\frac{E_\nu}{\langle E_\nu\rangle}\right]$$

~ luminosity   'pinching' parameter

  - SNOwGLoBES to model signals described by the pinched-thermal form.
  - Focus on ν_e flux and ν_e-CC interactions.



Backup

```
---MC-PARTICLE-MONITORING-----------------------------------------------------------

BeamNeutrinos:

--Primary 0, MCPDG 11, Energy 0.05, Dist. 14.63, nMCHits 75 (26, 30, 19 , Process: 1)
MCPDG 11, Energy 0.05, Dist. 14.63, nMCHits 75 (26, 30, 19)

--Primary 1, MCPDG 22, Energy 0.00, Dist. 21.44, nMCHits 11 (3, 5, 3 , Process: 1)
MCPDG 22, Energy 0.00, Dist. 21.44, nMCHits 11 (3, 5, 3)
------------------------------------------------------------------------------------
```

Cheating ID

File 1: Event 3

```
View done
=== MC Interaction : PDG 12 Energy: 0.053603 Type:
   MC 11 hits 75
   Matched 38 out of 38 with purity 1.00 and completeness 0.51
=== MC Interaction : PDG 12 Energy: 0.053603 Type:
   MC 11 hits 75
   Matched 31 out of 31 with purity 1.00 and completeness 0.41
      Unmatched
   Neutrino Interaction Summary:
   Good final state particles: 0 of 2 : 0.00%
```

W View

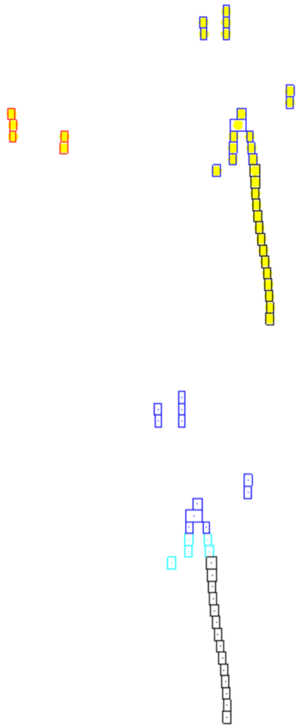
V View

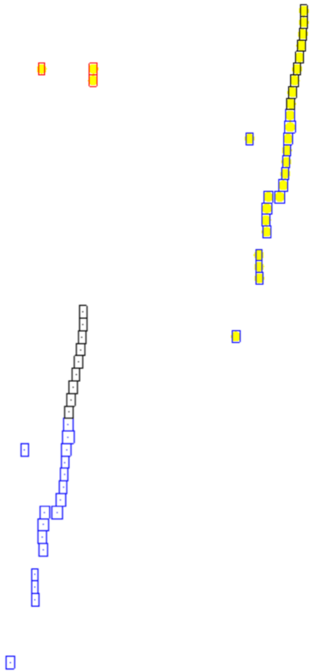
U View

Red hits, missed

Truth Info

PFO





Dark Blue, Black 2 PFOs, Light Blue is missed MC not included in PFO

PFO Split in to Track and Shower

```
---MC-PARTICLE-MONITORING------------------------------------------------------------------------

BeamNeutrinos:

--Primary 0, MCPDG 11, Energy 0.05, Dist. 14.63, nMCHits 75 (26, 30, 19 , Process: 1)
MCPDG 11, Energy 0.05, Dist. 14.63, nMCHits 75 (26, 30, 19)

--Primary 1, MCPDG 22, Energy 0.00, Dist. 21.44, nMCHits 11 (3, 5, 3 , Process: 1)
MCPDG 22, Energy 0.00, Dist. 21.44, nMCHits 11 (3, 5, 3)
------------------------------------------------------------------------------------------------
```

No Cheating ID

File 1: Event 3

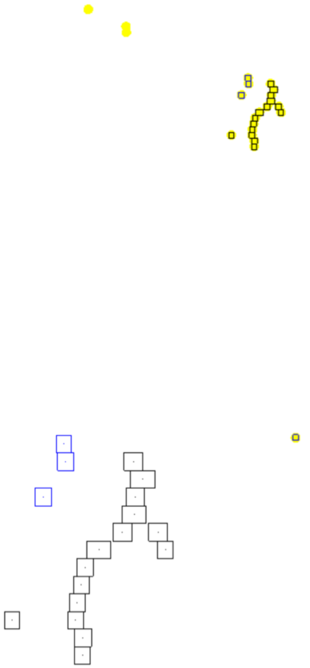
```
View done
=== MC Interaction : PDG 12 Energy: 0.053603 Type:
   MC 11 hits 75
   Matched 56 out of 59 with purity 1.00 and completeness 0.75
=== MC Interaction : PDG 12 Energy: 0.053603 Type:
   MC 11 hits 75
   Matched 11 out of 11 with purity 1.00 and completeness 0.15
      Unmatched
   Neutrino Interaction Summary:
   Good final state particles: 0 of 2 : 0.00%
```

Yellow: Signal Hits,

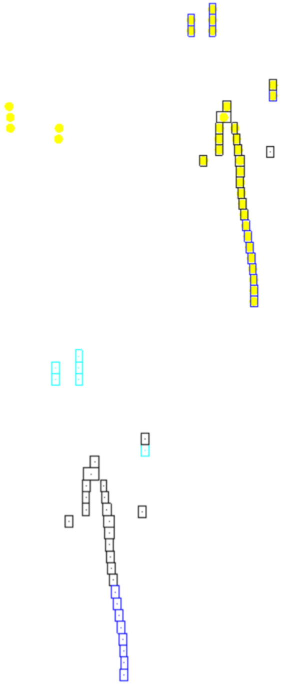
Black Squares: Network ID'd hits

Track and Shower PFO Produced

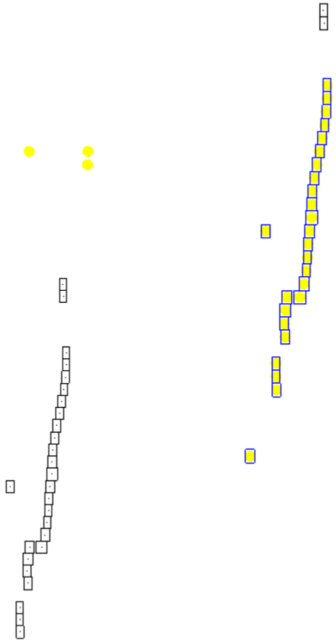
W View



V View



U View



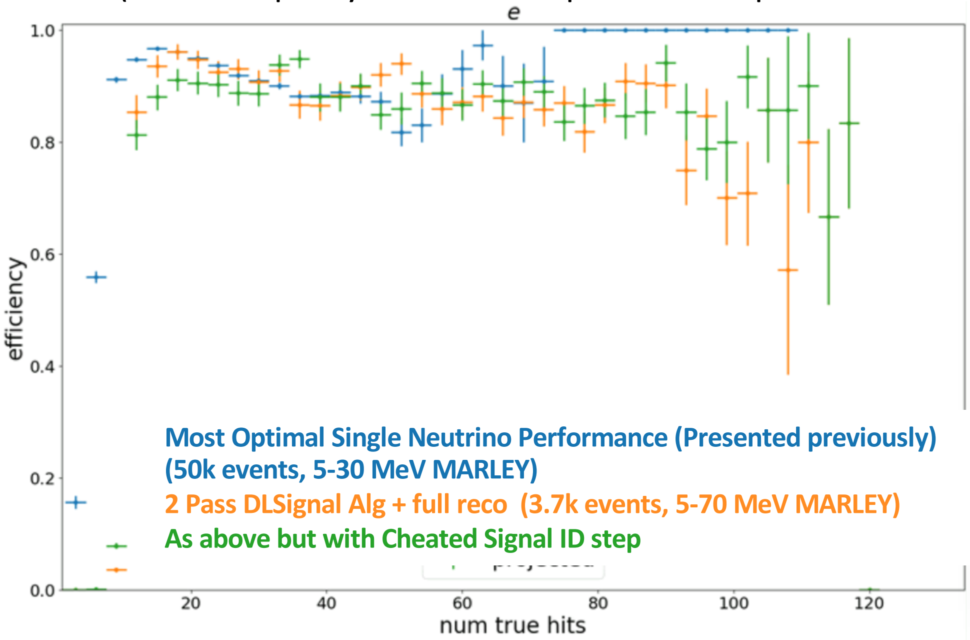
PFO

Blue Track, Black Shower, Light Blue missed E MC

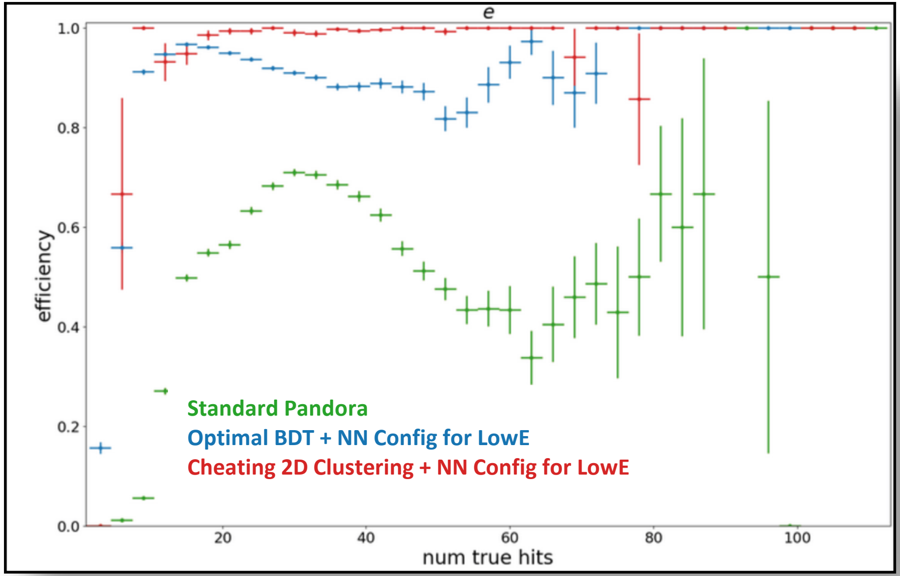
DLSignal Algorithm : Final Metric Results

- Impressive efficiency for electron reconstruction – 3.7k events included

(note 50% purity and 50% completeness requirement for reach reconstructed particle)



Signal plus background



Signal neutrino events only

Background/Signal Work

Implementation:

- New code in larpandoradlcontent/LArSignalId
  - DlSignalAlgorithm.h
  - DlSignalAlgorithm.cc

- Uses similar structure to DlVertexing
- Much more simplified structure/functions
- Converts CaloHit positions to a relative 256x256 pixel
- Runs event 'pixels' through network
- Creates mapping between signal pixels and CaloHits and adds to new SCaloHitListX, which is returned

PandoraSettings_Neutrino_DUNEFD_VD_LowE.xml

```
<algorithm type = "LArDLSignal">
    <TrainingMode>false</TrainingMode>
    <CaloHitListNames>CaloHitListW CaloHitListU CaloHitListV</CaloHitListNames>
    <SignalListNameU>SCaloHitListU</SignalListNameU>
    <SignalListNameV>SCaloHitListV</SignalListNameV>
    <SignalListNameW>SCaloHitListW</SignalListNameW>
    <SignalListName2D>SCaloHitList2D</SignalListName2D>
    <ModelFileNameU>PandoraNetworkData/PandoraUnet_backgrounds_1_U.pt</ModelFileNameU>
    <ModelFileNameV>PandoraNetworkData/PandoraUnet_backgrounds_1_V.pt</ModelFileNameV>
    <ModelFileNameW>PandoraNetworkData/PandoraUnet_backgrounds_1_W.pt</ModelFileNameW>
    <Visualise>false</Visualise>
</algorithm>
```

- Runs after LArPreProcessing and before LArDlVertexing
- Outputs new 'signal list with Prefix 'S' – Used by subsequent algs

Function Calls:

Run

- PrepareTrainingSample
- Infer
  - MakeNetworkInputFromHits
  - GetMCToHitsMap
  - CompleteMCHierarchy
  - GetHitRegion

Low Energy Reconstruction: Previous Work

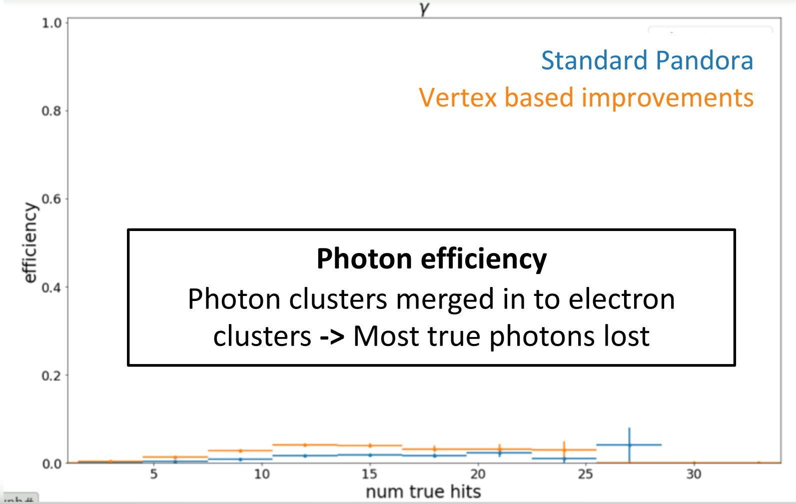
- **Deep Learning Vertex Position**
  - Improved vertex efficiency leads to greater reconstruction efficiency
  - Vertex position used in logic in downstream algorithms
- **Electrons with 40 – 80 hits**
  - Despite more accurate vertex, reconstructed clusters are still split
  - 2D to 3D cluster matching creates split particles below threshold

Next step:

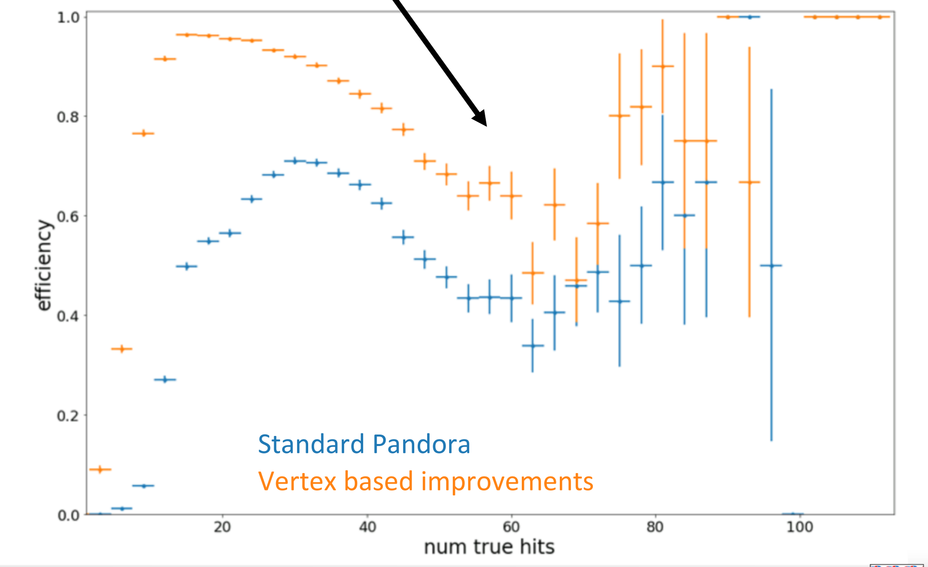
- Examine event displays to understand failures

Middle region (40 – 80 hits)
True particle split in to multiple PFOs -> incomplete PFOs, drop below completeness cut

Particle Flow Object (PFO)



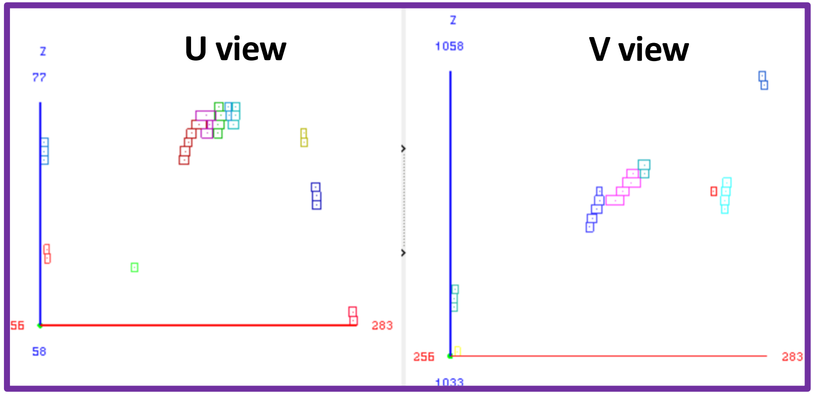
Electron Reconstruction Efficiency



Algorithm Development: Understanding standard Pandora

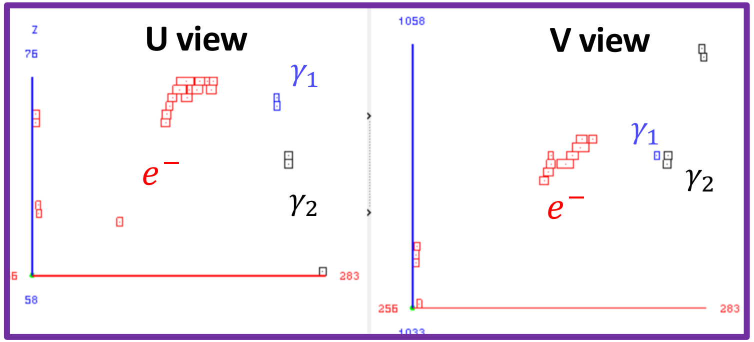
- True particles are being reconstructed as multiple, incomplete PFOs due to fragmented input clusters

Input cluster hits (U, V Views)



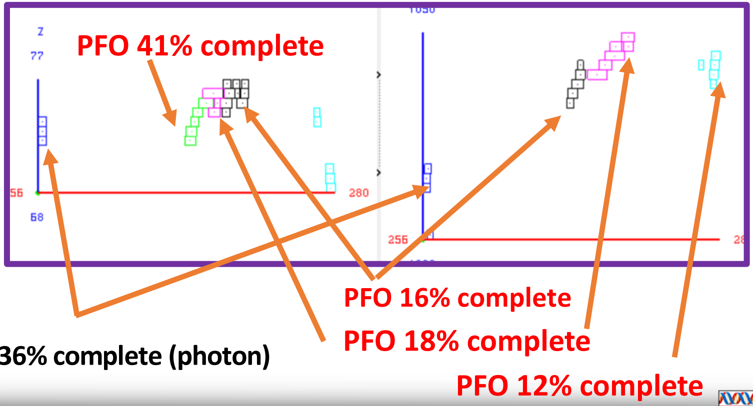
Performance we want

MC True Particles



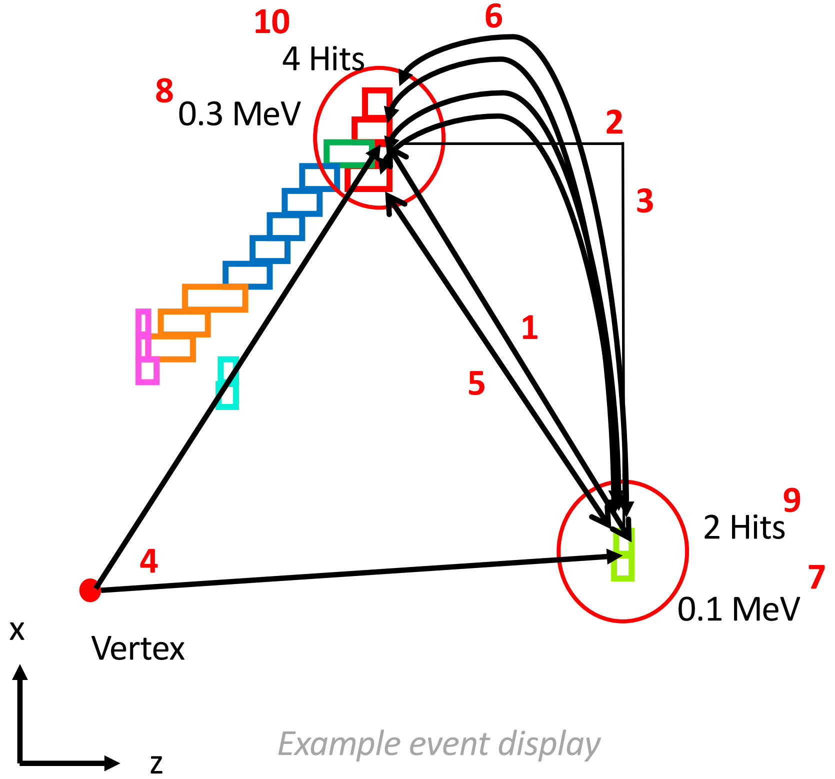
Performance we see

Standard Pandora – End of Reco chain



Distinguishing between electrons and deexcitation photons is a really difficult task!

Cluster Merging Algorithm Design: Parameters



Example event display

12 Parameters

Parameters to compare for each pairwise cluster choice:

1. Distance between centroid of each cluster

2/3. x and z component of vector between centroid of each cluster

4. Angle between centroid and interaction vertex

5. Minimum distance between a hit in each cluster

6. Average distance between each hit in each cluster

7/8. Total ADC of each cluster

9/10. Number of hits in each cluster

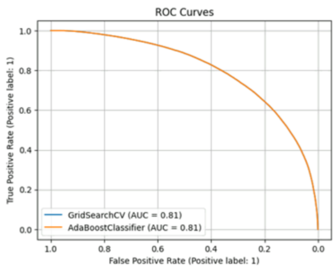
*11. Number of hits in contact with hits in other cluster

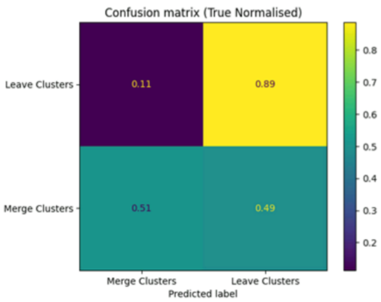
*12. Number of hits in proximity to hits in other cluster

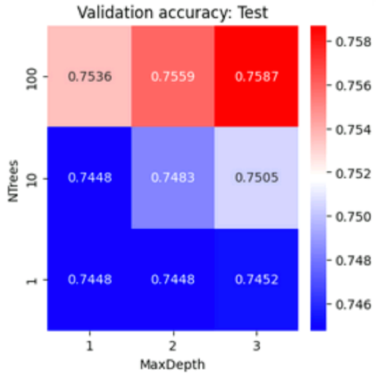
*Not represented

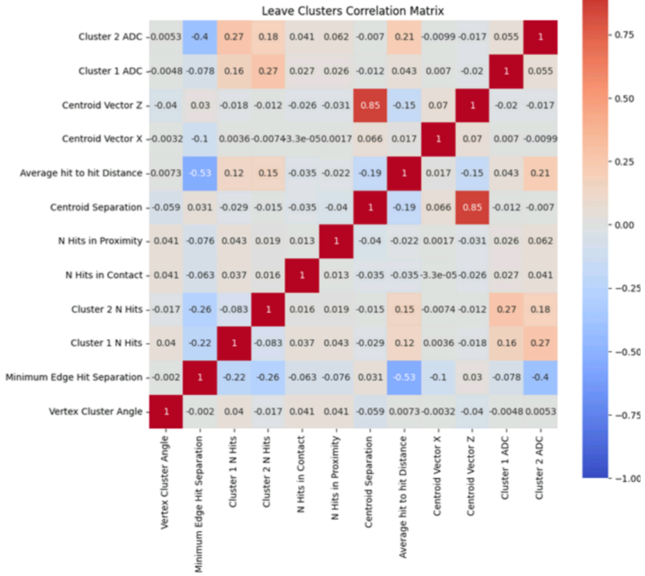
Merging BDT v8

- BDT Training stats
- 12 Parameters

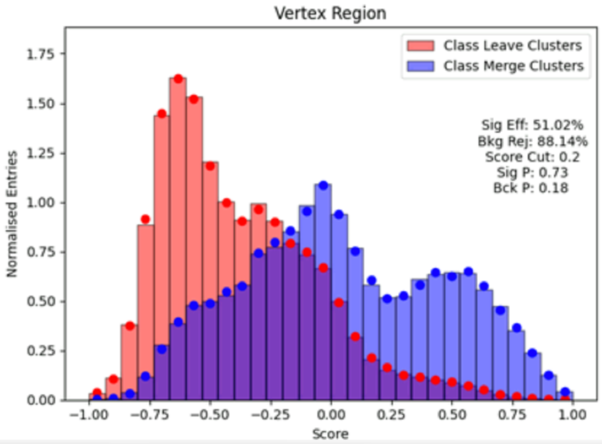




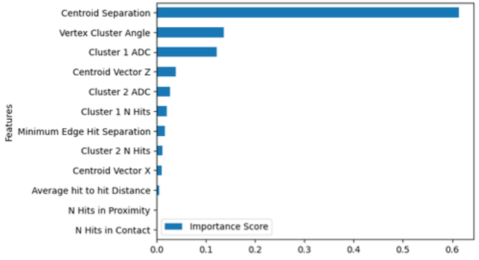




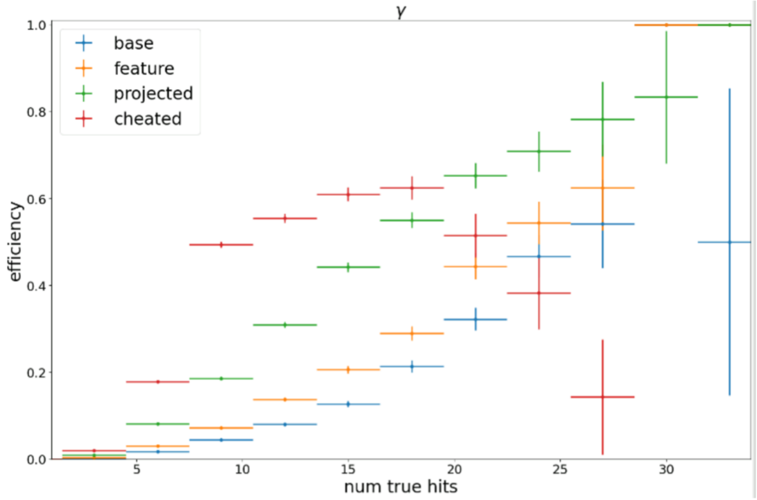

```
(224588, 12)
(224588,)
(224588, 12)
(224588,)
KS Signal:     0.0034986994323501497 with P value: 0.7314153858751964
KS BackGround: 0.004040885018558926 with P value: 0.1793406929159631
```

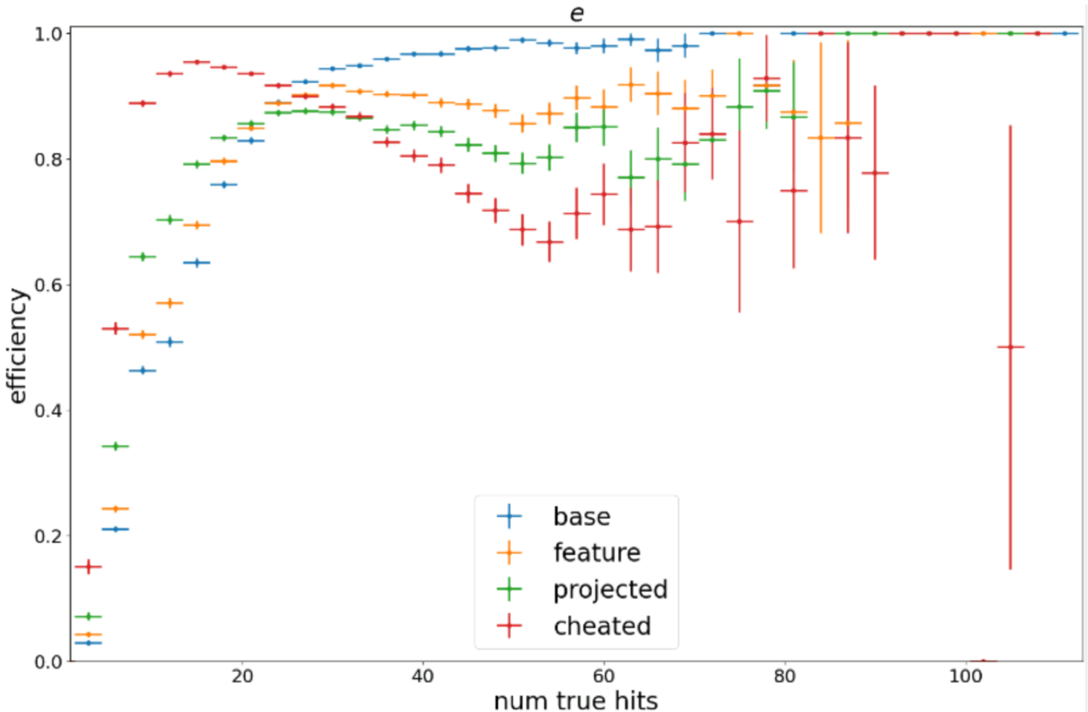


	Features	Importance Score
4	N Hits in Contact	0.000000
5	N Hits in Proximity	0.000000
7	Average hit to hit Distance	0.005595
8	Centroid Vector X	0.010659
3	Cluster 2 N Hits	0.011437
1	Minimum Edge Hit Separation	0.016611
2	Cluster 1 N Hits	0.020066
11	Cluster 2 ADC	0.026429
9	Centroid Vector Z	0.038464
10	Cluster 1 ADC	0.121202
0	Vertex Cluster Angle	0.136543
6	Centroid Separation	0.612994



Merging Algorithm: Particle Efficiency with BDT Cuts





BDT v8 -0.5 cut

BDT v8 -0.35 cut

BDT v8 -0.25 cut

BDT v8 0 cut

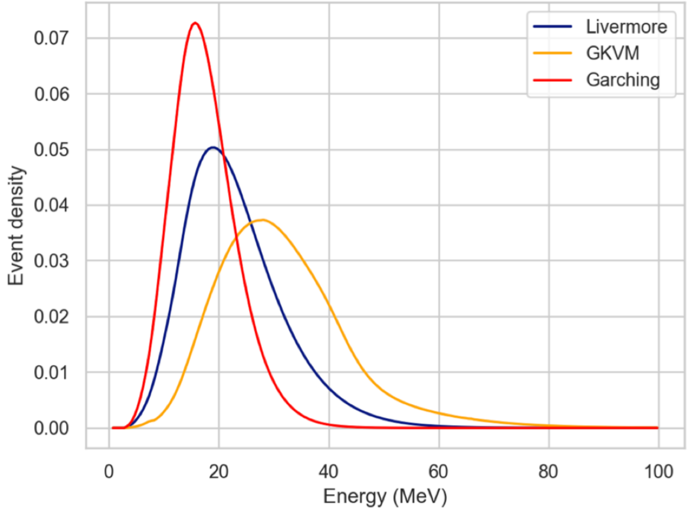
Completing analysis will help answer how important photon efficiency vs electron efficiency is, to help decide cut values

```
validation1 = MCValidation("hierarchy_validation_Mvav80.25_all.root", "mc")
validation2 = MCValidation("hierarchy_validation_Mvav80.325_all.root", "mc")
validation3 = MCValidation("hierarchy_validation_Mvav80.375_all.root", "mc")
validation4 = MCValidation("hierarchy_validation_Mvav8_all.root", "mc")
```

Expected supernova core-collapse signal: SNOwGLoBEs

- Make fits of reconstructed samples compared to supernova model expectations
  - SNOwGLoBEs - fast event rate calculation tool
  - MARLEY – Monte Carlo Event Generator for LAr interactions

Work developed in conjunction with Low Energy Working Group



νe **Charged current events with arbitrary normalisation**

40 kton LAr, 10 kpc SN

Channel	Liver-more	GKVM	Garching
$\nu_e + ^{40}\mathrm{Ar} \rightarrow e^- + ^{40}\mathrm{K}^*$	2648	3295	882
$\overline{\nu}_e + ^{40}\mathrm{Ar} \rightarrow e^+ + ^{40}\mathrm{Cl}^*$	224	155	23
$\nu_X + e^- \rightarrow \nu_X + e^-$	341	206	142
Total	3213	3656	1047

νe **dominant signal**

https://indico.cern.ch/event/1199289/contributions/5447099/attachments/2705699/4697057/DUNE_LEP_CCuesta.pdf