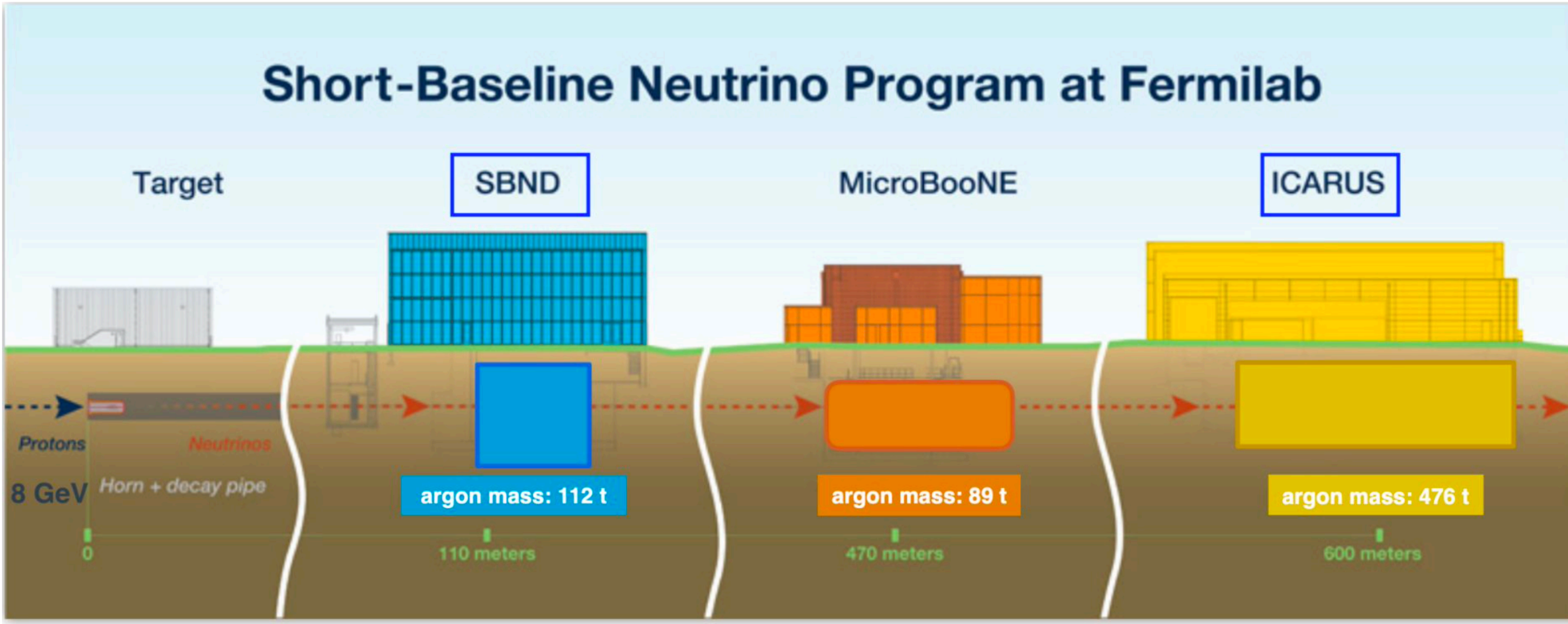
Machine-Learning-Based Data Reconstruction Chain for SBND

NPML 2024

June 26, 2024

B. Carlson - *bcarlson1@ufl.edu*

Short-Baseline Near Detector (SBND)



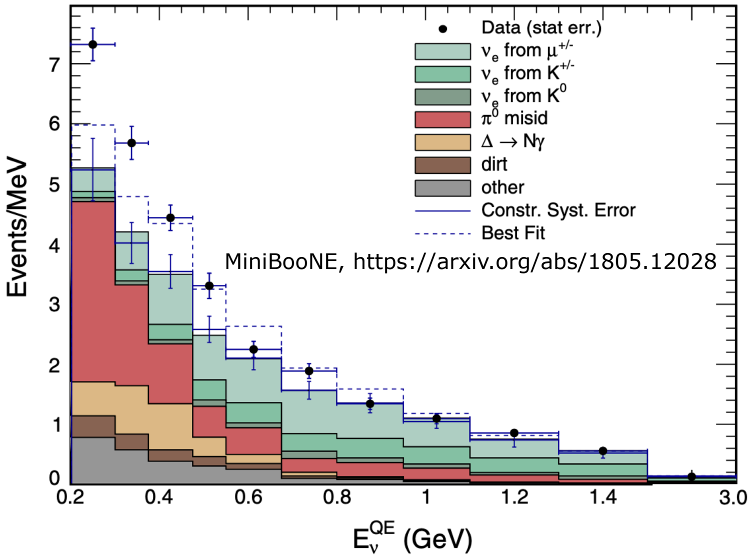
Booster Neutrino Beam (BNB)

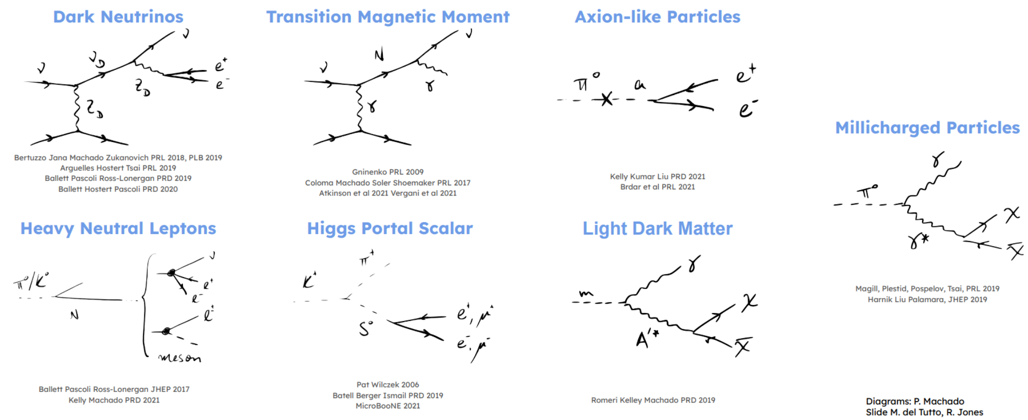
2015-2021
Large production of scientific results

Beam composition:
ν_μ (93.6%)
$\bar{\nu}_\mu$ (5.9%)
ν_e+$\bar{\nu}_e$ (0.5%)

Mean ν_μ energy: ~0.8 GeV

Short-Baseline Near Detector (SBND) - Physics



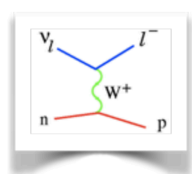


Cross section

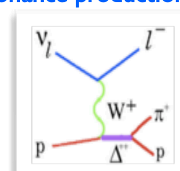
Low energy excess (oscillations?)

Beyond standard model

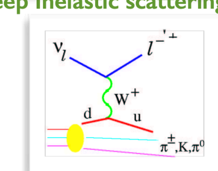
Quasi-elastic scattering (QE)

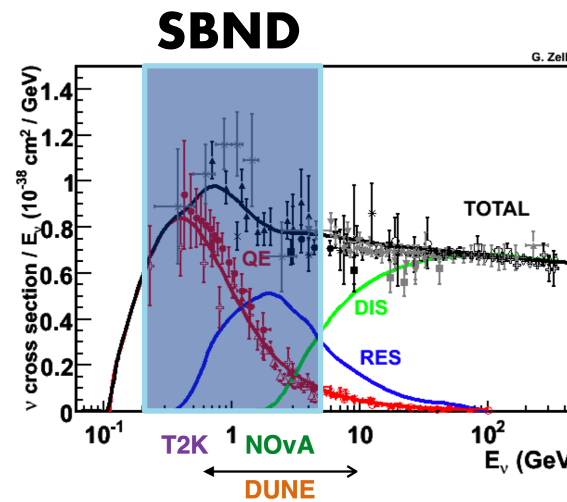


Resonance production (RES)



Deep Inelastic scattering (DIS)



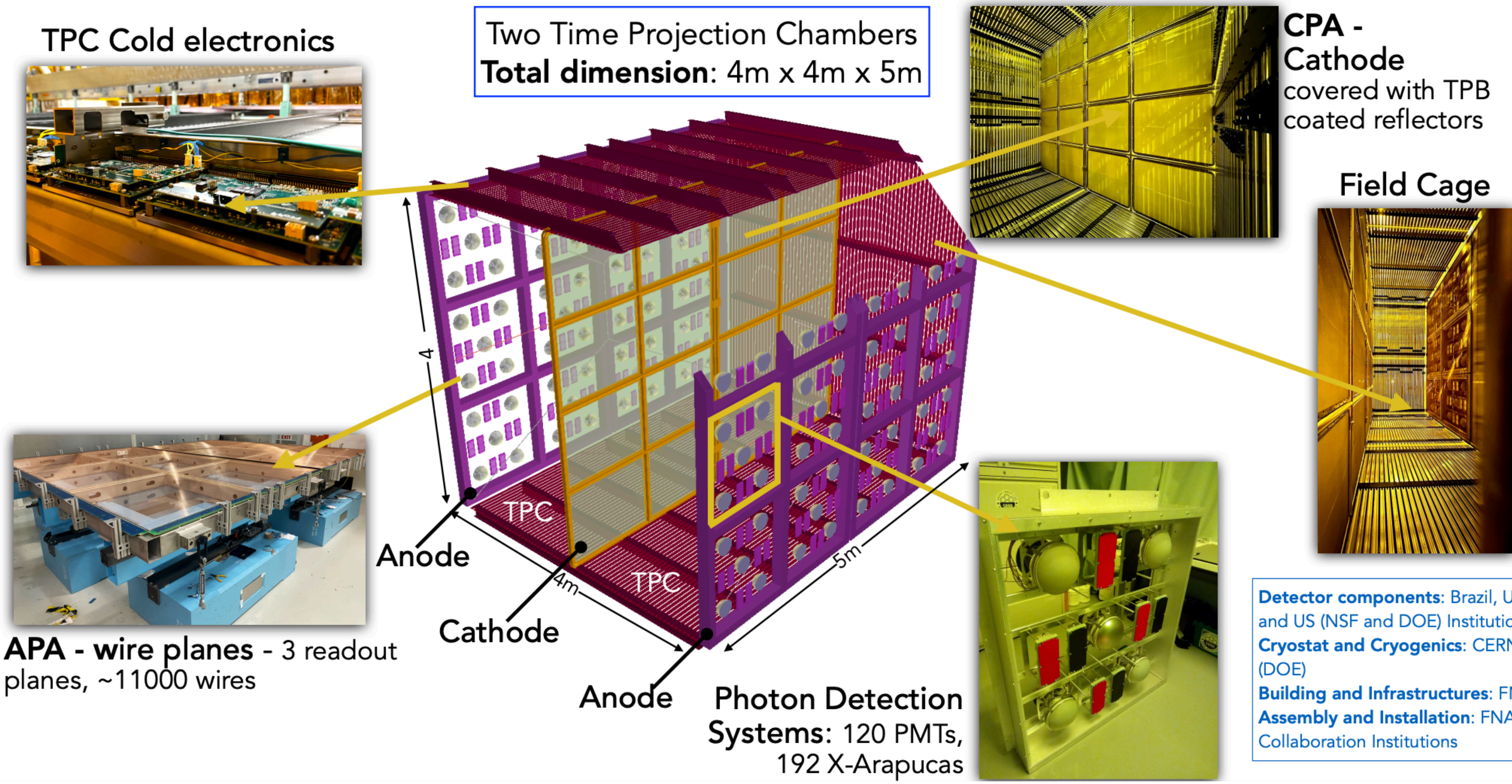


J.A. Formaggio, G. Zeller, Reviews of Modern Physics, 84 (2012)

14

Short-Baseline Near Detector (SBND) - Detector

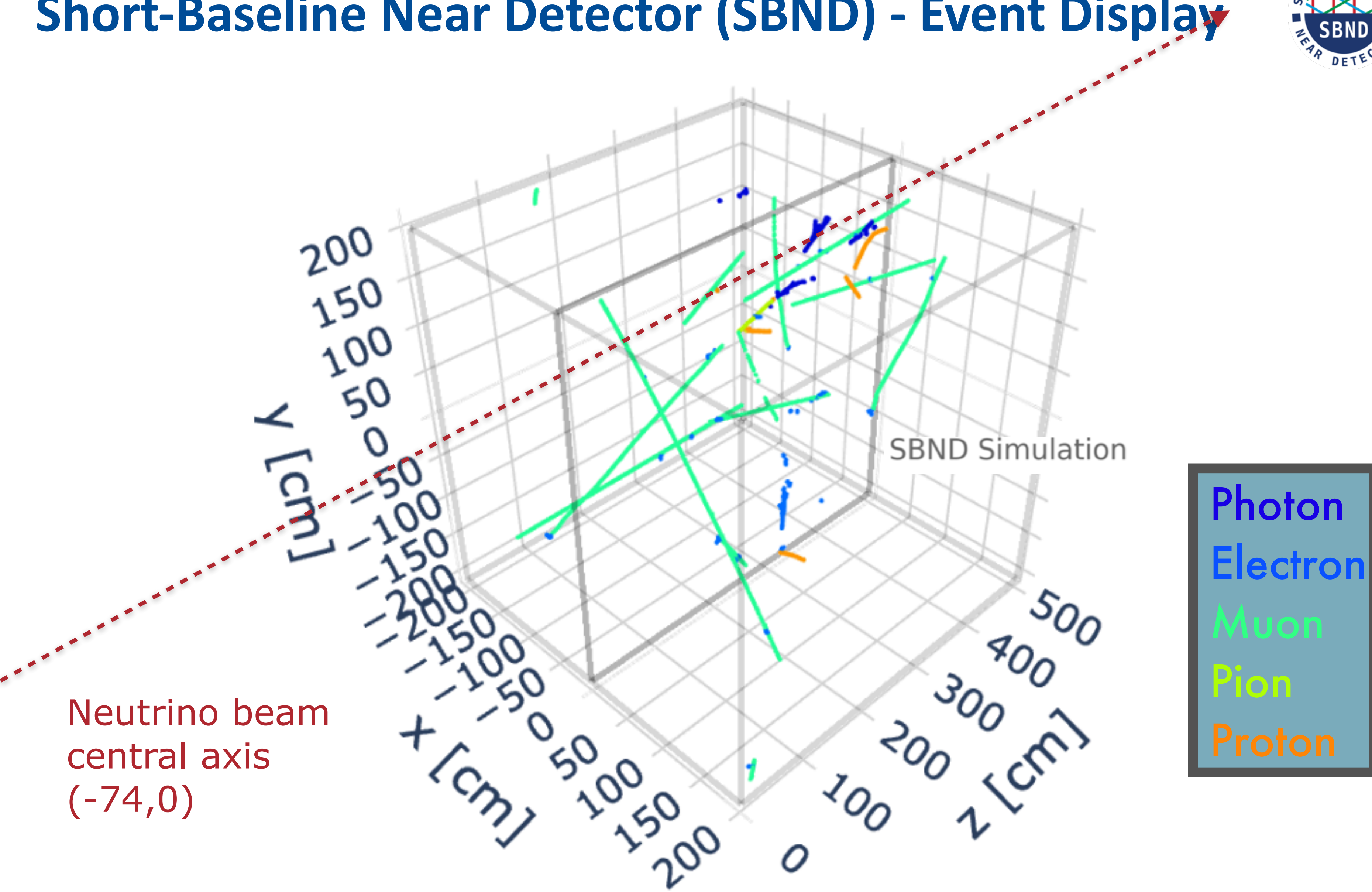
CRT provides 4π cosmic coverage (not shown)



Detector components: Brazil, UK, Switzerland and US (NSF and DOE) Institutions
Cryostat and Cryogenics: CERN and FNAL (DOE)
Building and Infrastructures: FNAL (DOE)
Assembly and Installation: FNAL (DOE) and Collaboration Institutions

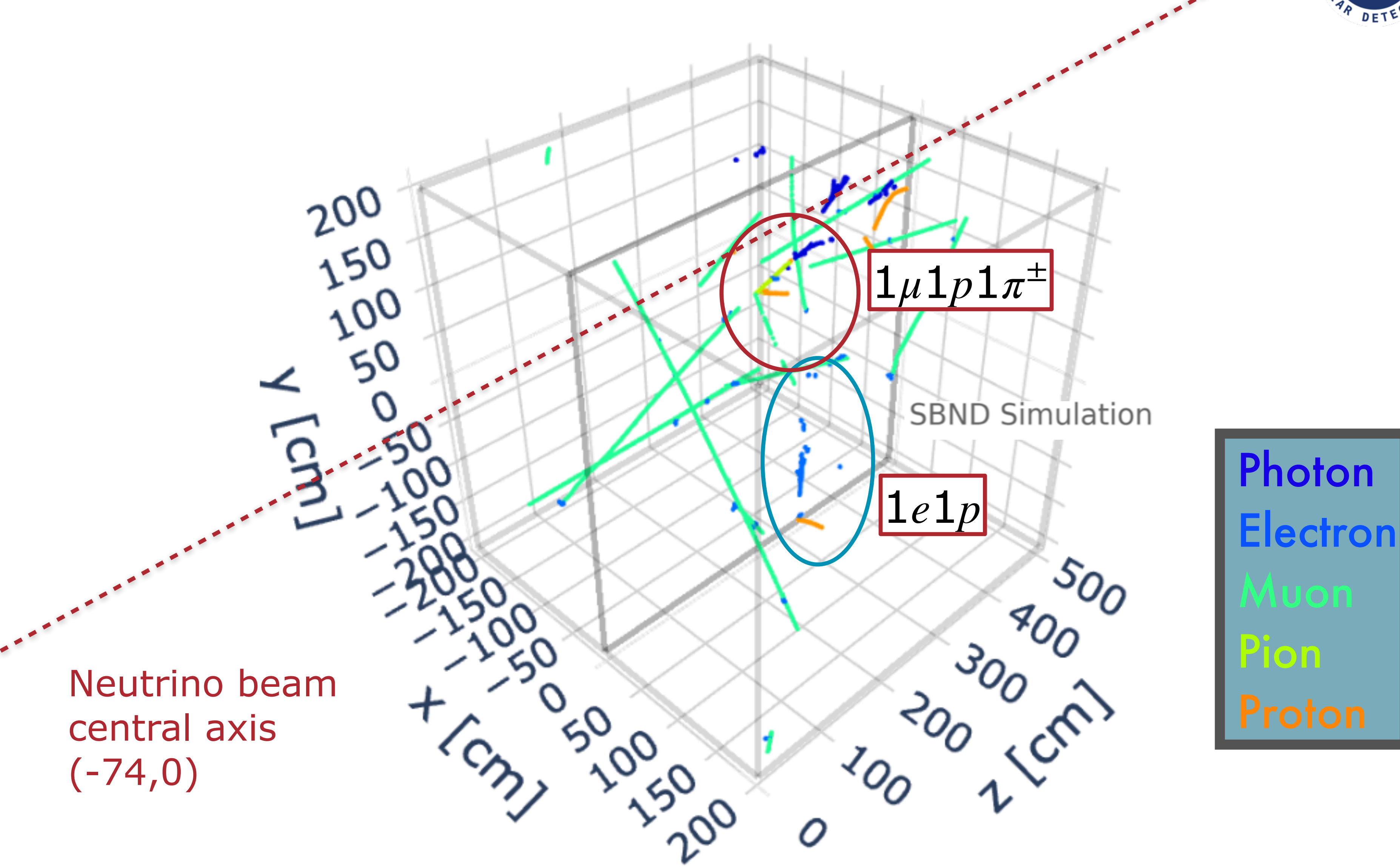
Credit - O. Palamara

Short-Baseline Near Detector (SBND) - Event Display



Neutrino beam central axis (-74,0)

Short-Baseline Near Detector (SBND) - Event Display



SBN Working Group

- SPINE Working Group led by **F. Drielsma** and **K. Terao** (SLAC)
  - LArTPC Experts: **T. Usher** (SLAC) and **M. Mooney** (CSU)
  - ML Experts: **D.H. Koh** and **Y-J. Jwa** (SLAC)
  - SBND Group: **B. Carlson, C. Fan** (UF), **N. Oza** (Columbia), **R. LaZur** and **L. Paudel** (CSU)

F. Drielsma

K. Terao

M. Mooney
Data/sim

T. Usher
Signal Proc.

D.H. Koh
ML

Y-J. Jwa
ML

B. Carlson
Porting

N. Oza
Michel Sel.

C. Fan
dE/dx studies

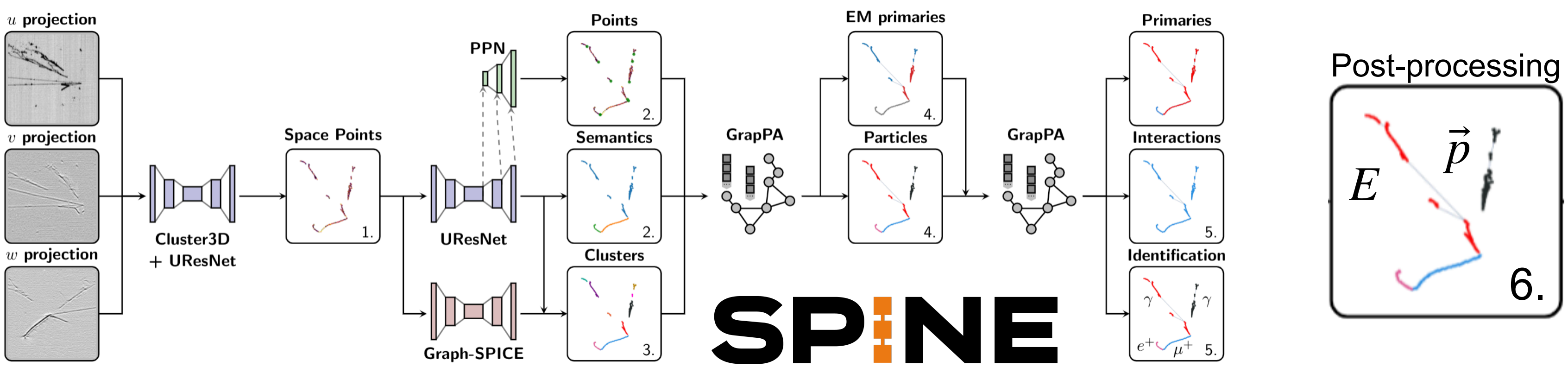
R. LaZur
Onboarding

L. Paudel
Onboarding

SPINE Overview

- End-to-end trainable reconstruction chain that **aggregates 3D space-points** into super structures (particles, interactions) and **identifies types** (particle ID/semantic type)

SPINE = Scalable Particle Imaging with Neural Embeddings [1]



- Post-processing is a non-ML reconstruction that handles **energy reconstruction, particle direction**, etc. after event is inferred using full chain

[1] https://github.com/DeepLearnPhysics/spine

Training

- Multi-particle vertex multi-particle rain (MPVMPR) sample - 3 generators
  - Out-of-time rain (MPR) - trains for **out-of-time cosmic** activity
  - In-time rain (MPR) - trains for **in-time cosmic** activity
  - Vertex (MPV) - trains for **neutrino** activity
- 278k training, validation + ~50k testing

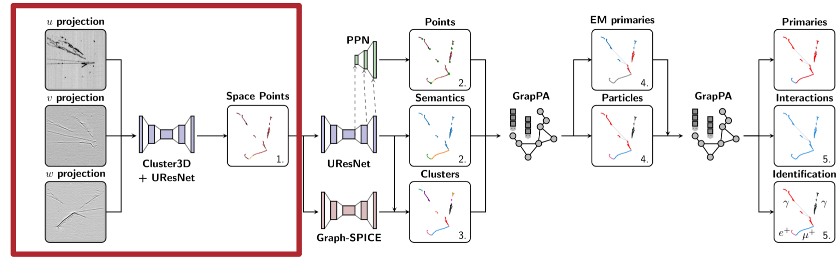
$e^{\pm}$ $\mu^{\pm}$ π^{0} $\pi^{\pm}$ p γ

GeV → KERange

```
MultiMax       : 7
MultiMin       : 2
ParticleParameter: {
    PDGCode      : [   [-11,11,-13,13], [111], [211,-211],   [2212],   [22]]
    MinMulti     : [            0,          0,          0,       0,      0]
    MaxMulti     : [            1,          2,          2,       4,      2]
    ProbWeight   : [            3,          1,          1,       3,      1]
    KERange      : [ [0.0,3.0], [0.0,1.0], [0.0,1.0],  [0.0,1.0], [0.0,1.0]]
    MomRange     : []
}
```

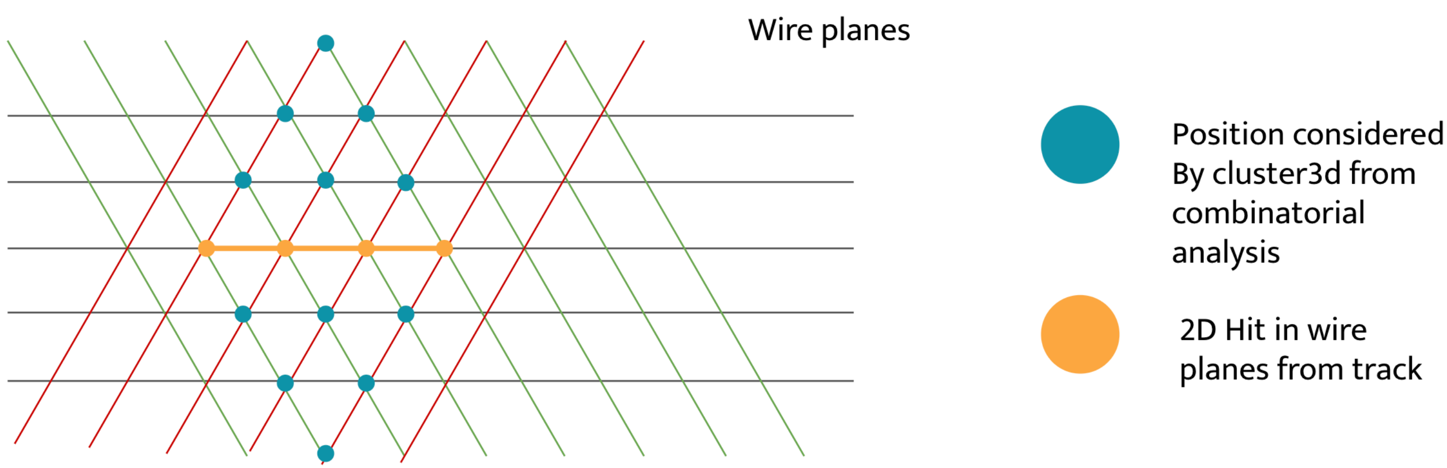
MPV v01 parameters

Cluster3D



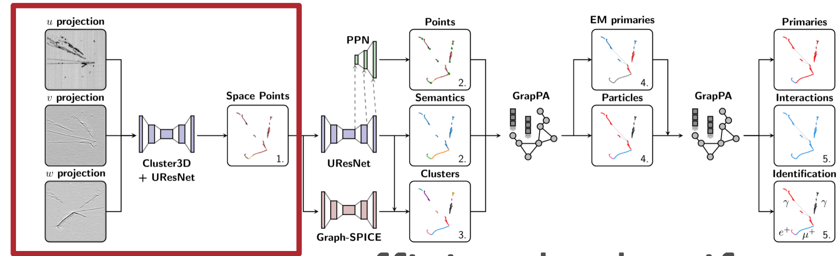
Cluster3D consumes 2D hits in each of 3 projections

- Finds pairs of hits compatible within a time threshold
- Forms a **triplet** point from 3 wires where 3 hits are compatible in time to form candidate space-points

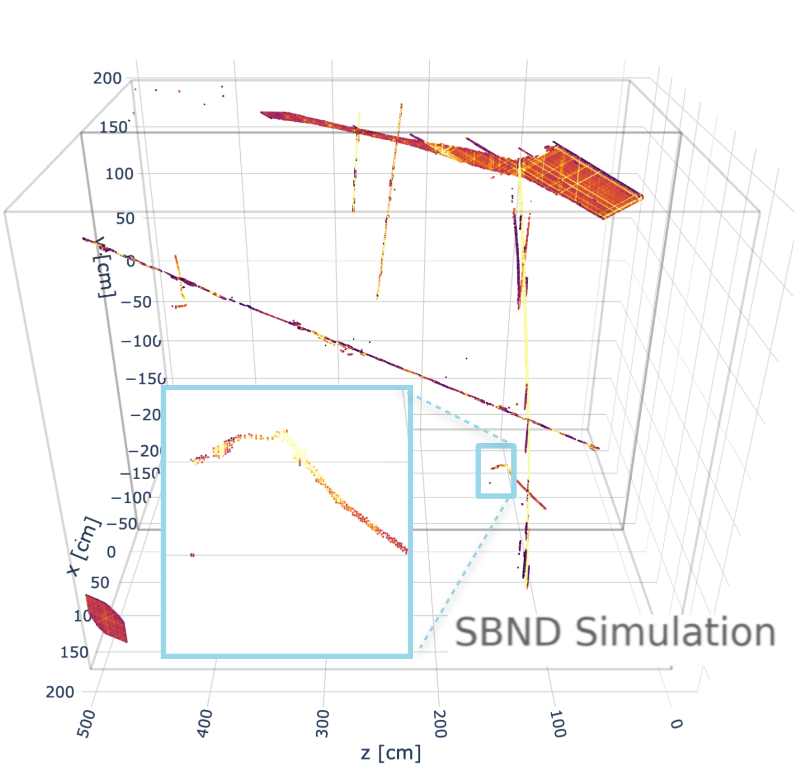


- False hits create <u>*ghost*</u> points, which are de-ghosted using a UResNet CNN

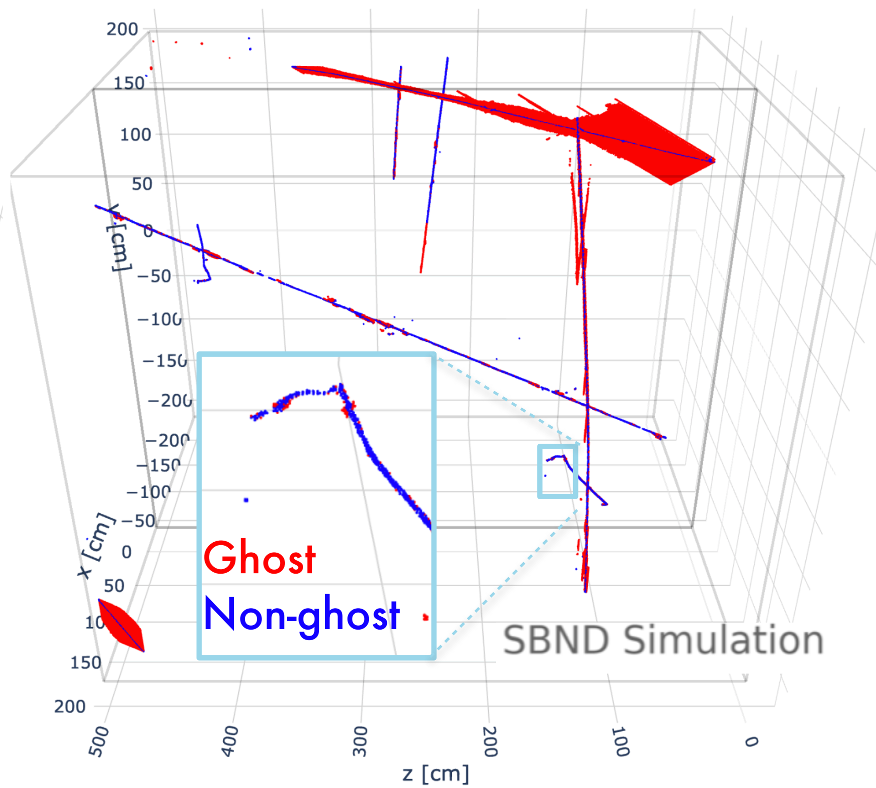
De-ghosting



- UResNet CNN architecture with cross entropy loss to efficiently classify *ghost* points
- Overall **94.3%** de-ghosting accuracy



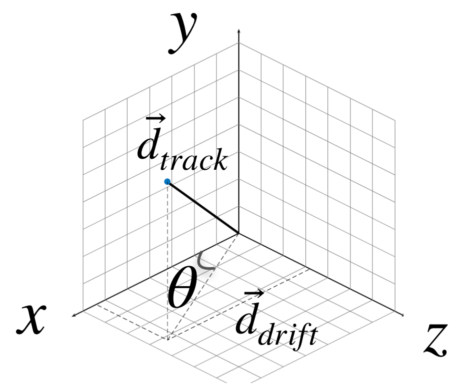
Input raw charge



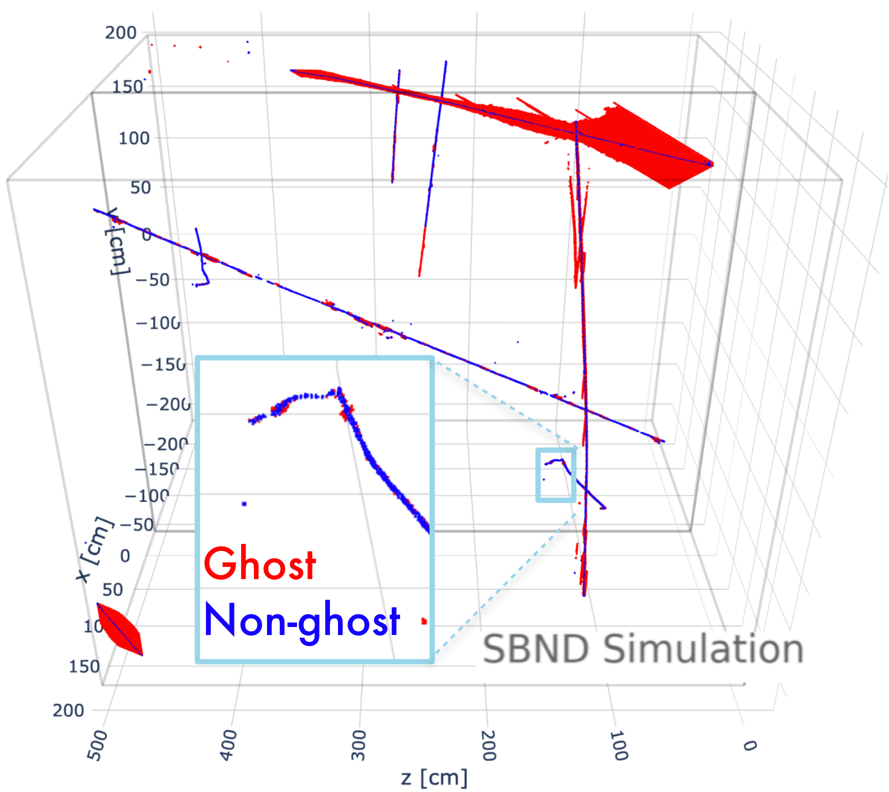
De-ghosted labels

De-ghosting

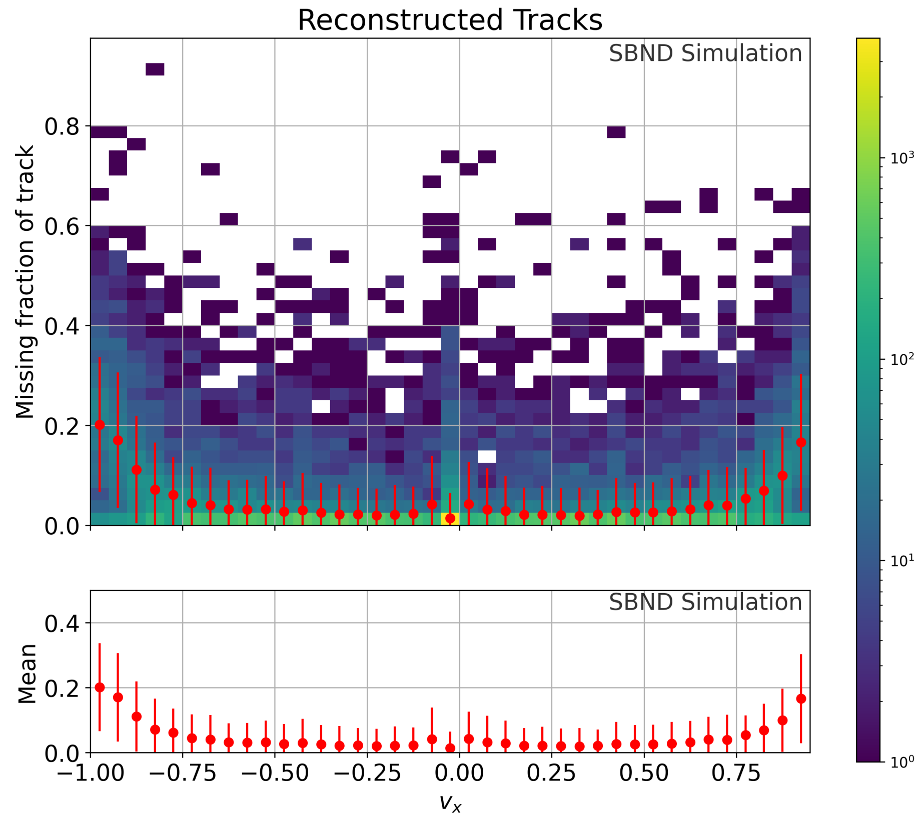
- Hits missing for high v_x cause gappy tracks



$$v_x = \vec{d}_{drift} \cdot \vec{d}_{track} = \cos\theta$$



De-ghosted labels

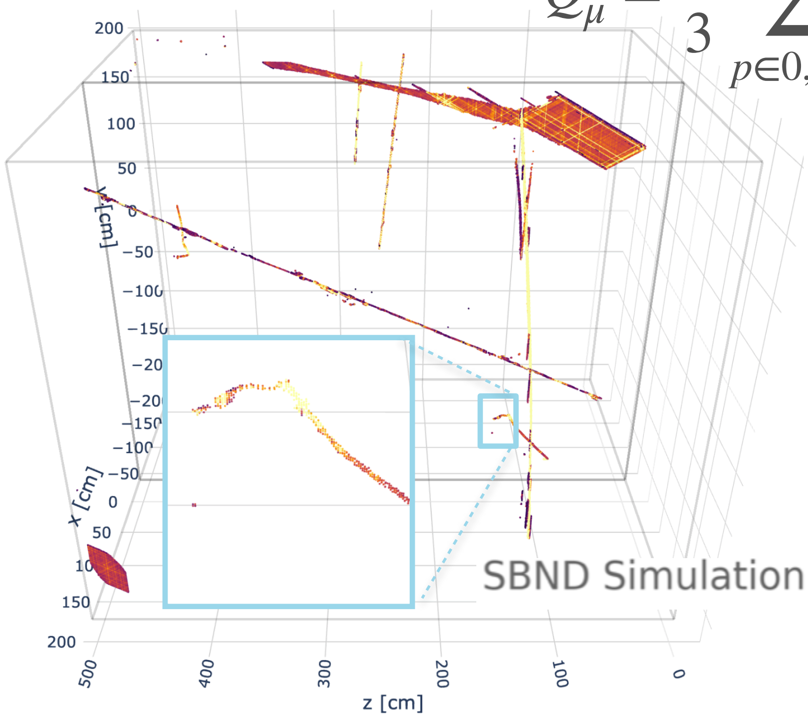


Track Completeness

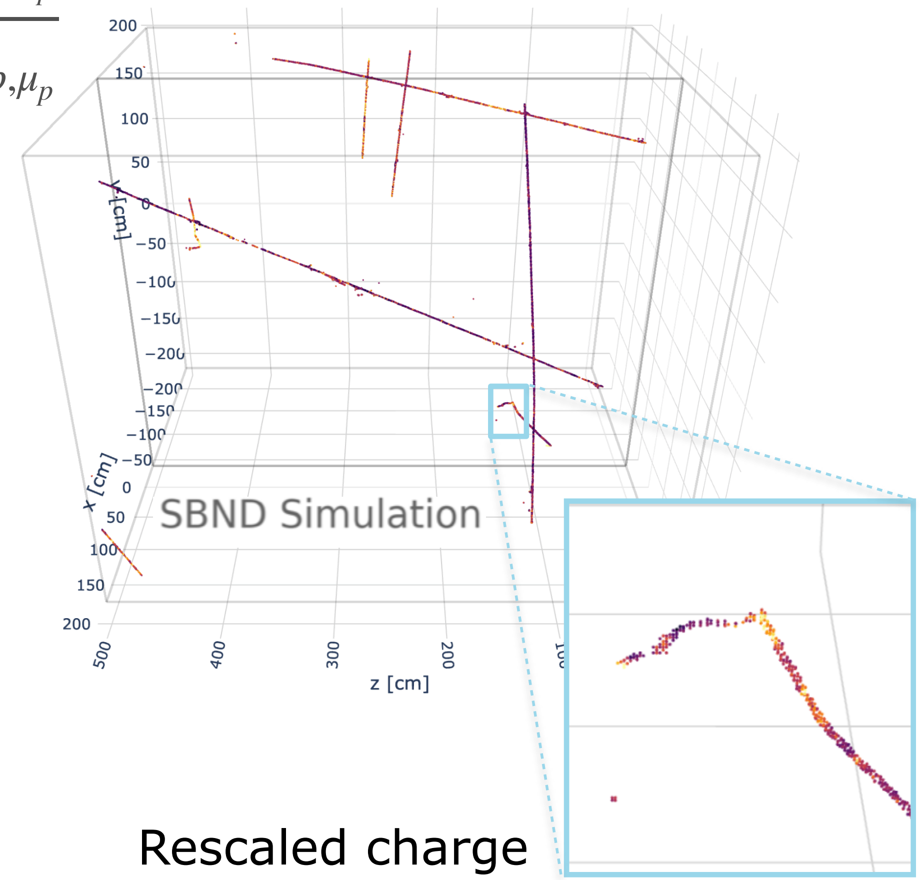
Charge Rescale

- Count the number of times $n_{p,i}$ each hit $h_{p,i}$ is used in the de-ghosted voxels
- Recompute the corrected charge of all voxels, which account for hit multiplicity
- Removes angular dependence of charge

$$Q_\mu = \frac{1}{3} \sum_{p \in 0,1,2} \frac{q_{p,\mu_p}}{n_{p,\mu_p}}$$



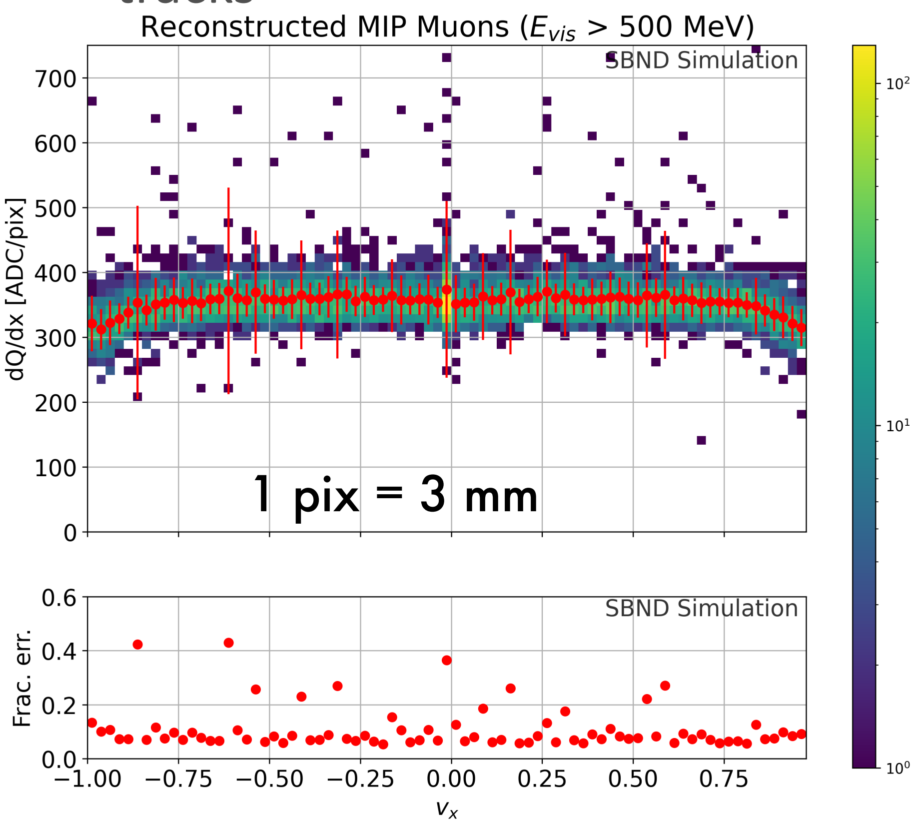
Raw charge



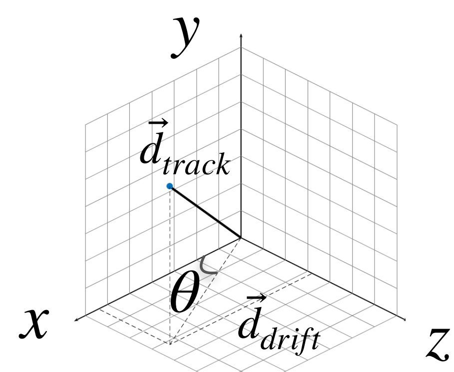
Rescaled charge

Charge Rescale

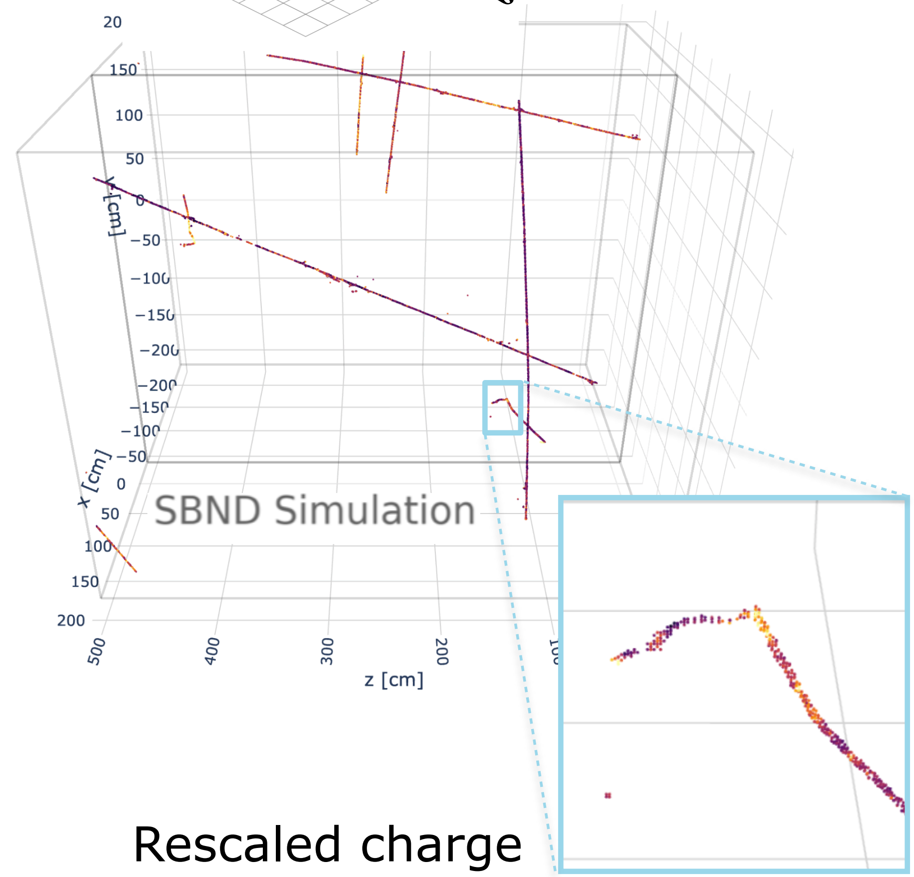
- Charge rescaling has low angular dep.
- Hits missing for high v_x cause gappy tracks



dQ/dx

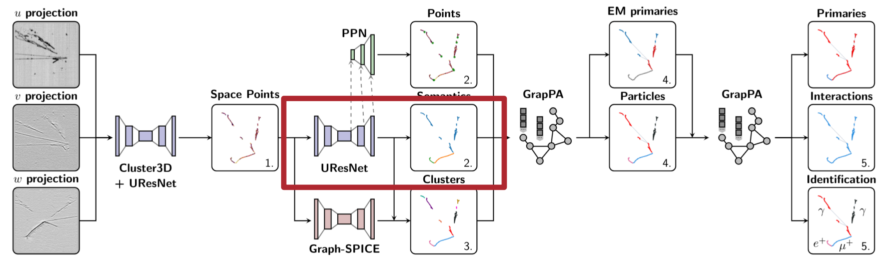


$$v_x = \vec{d}_{drift} \cdot \vec{d}_{track} = \cos\theta$$

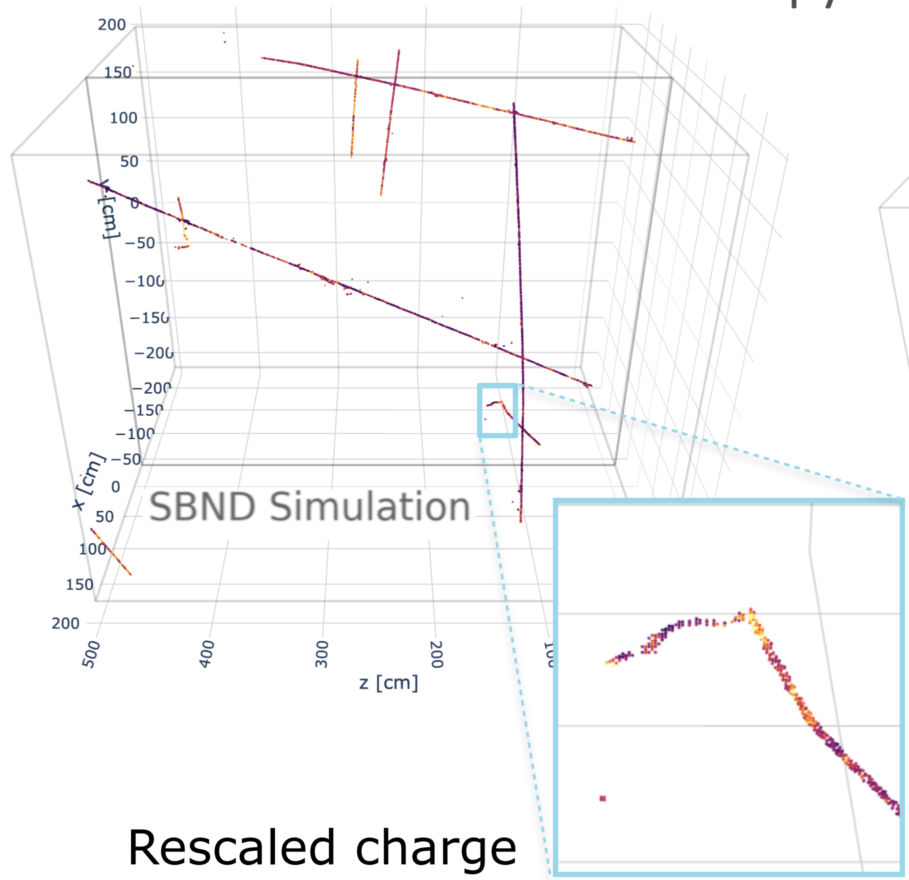


Rescaled charge

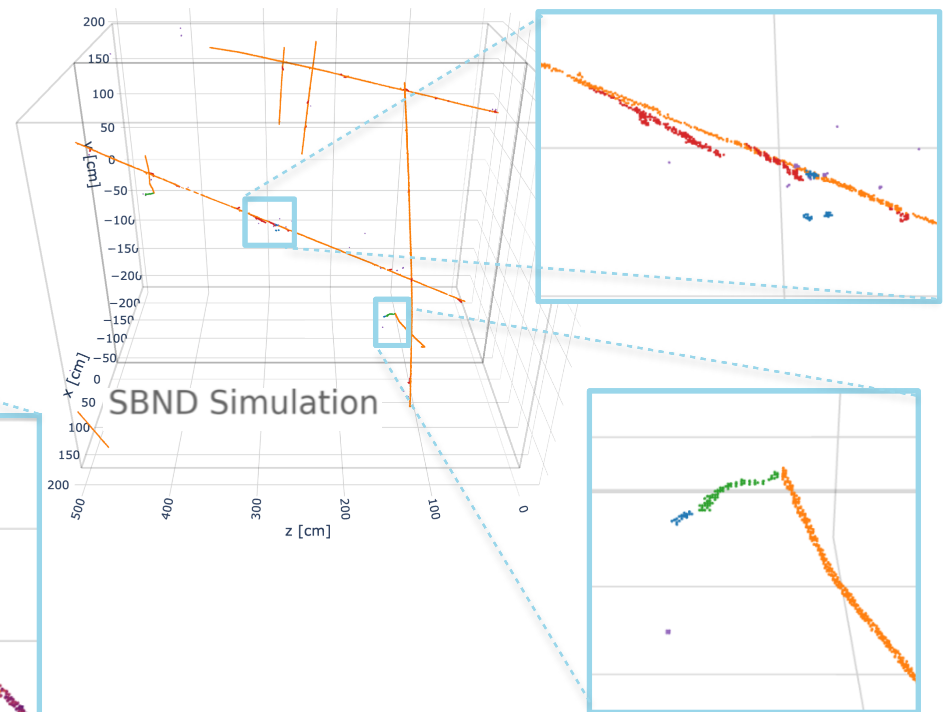
Semantic Segmentation



- Classify space point type as tracks, showers, michels, deltas, LEs
- UResNet CNN with cross entropy loss, one hot encoding for type

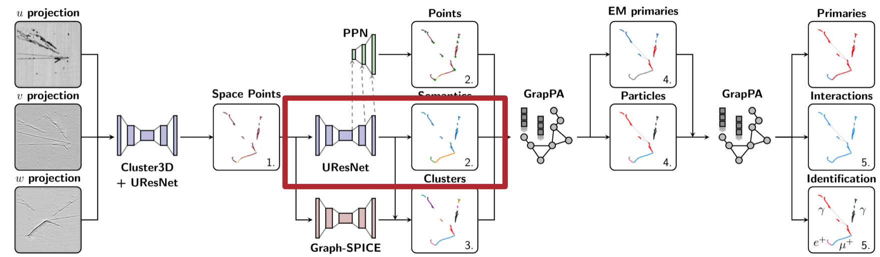


Rescaled charge

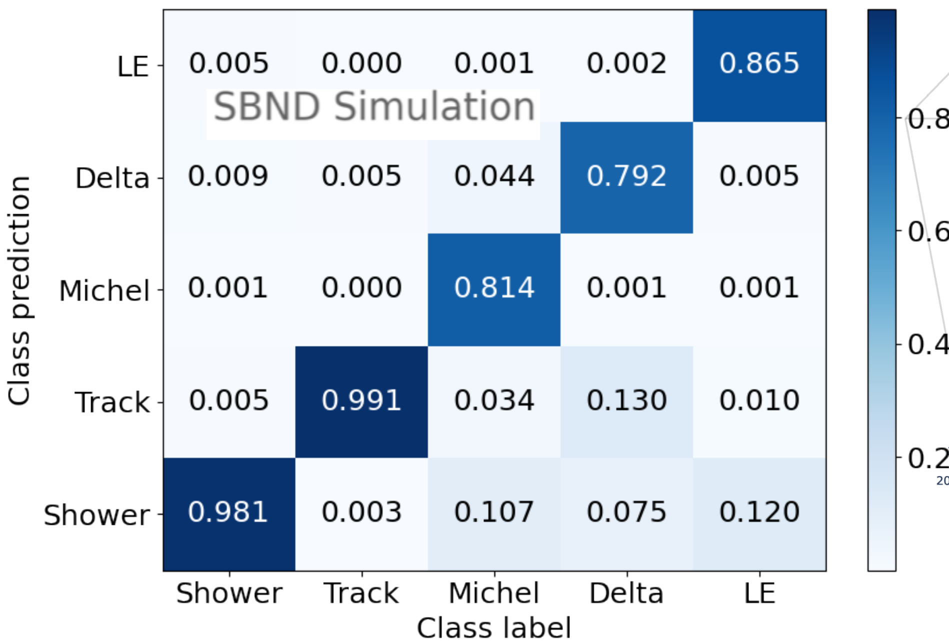


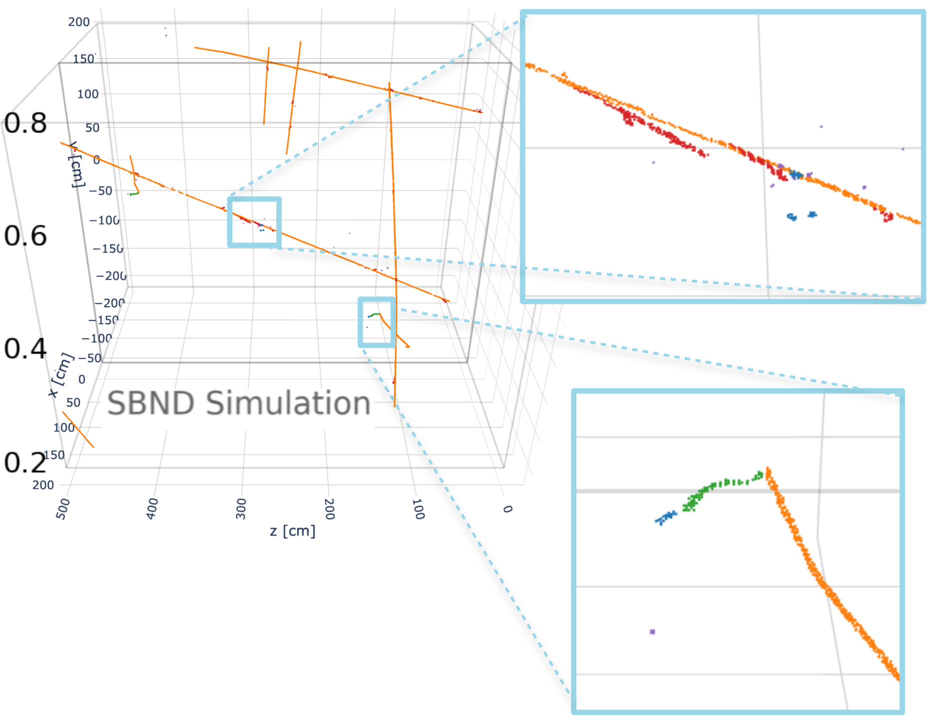
Semantic labels

Semantic Segmentation



- Overall **97.7%** semantic accuracy
- Mistakes primarily from classifying michels, deltas, LEs as showers
- Another class of mistakes comes from merging deltas into tracks

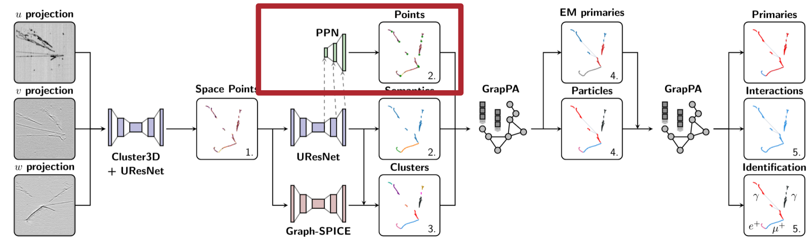




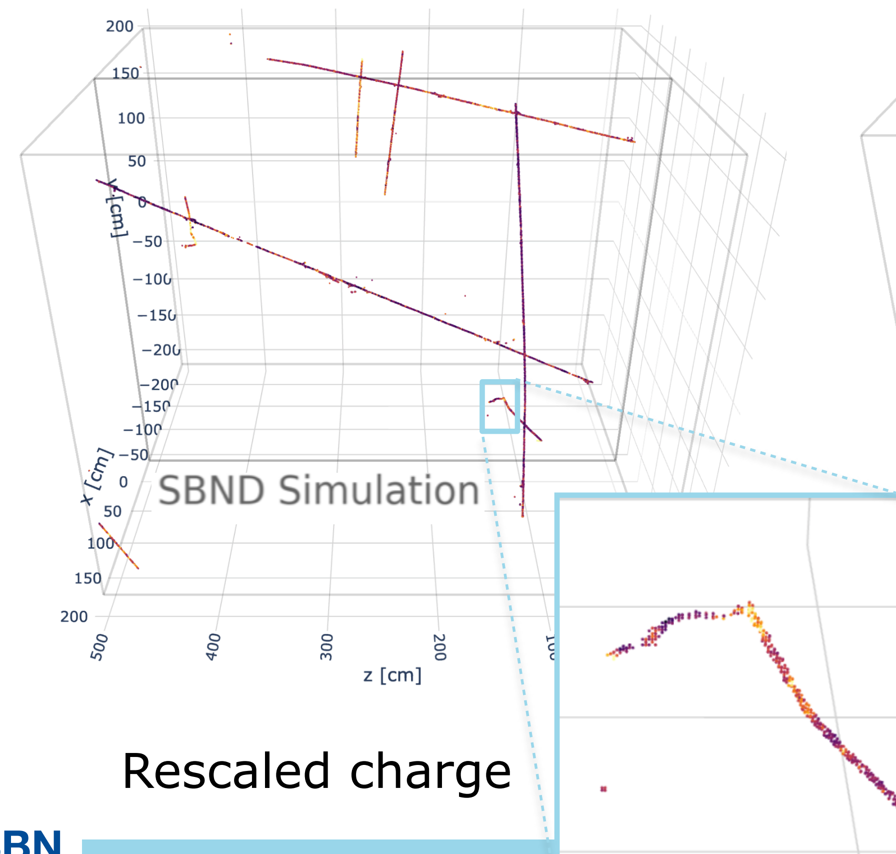
Semantic Confusion

Semantic labels

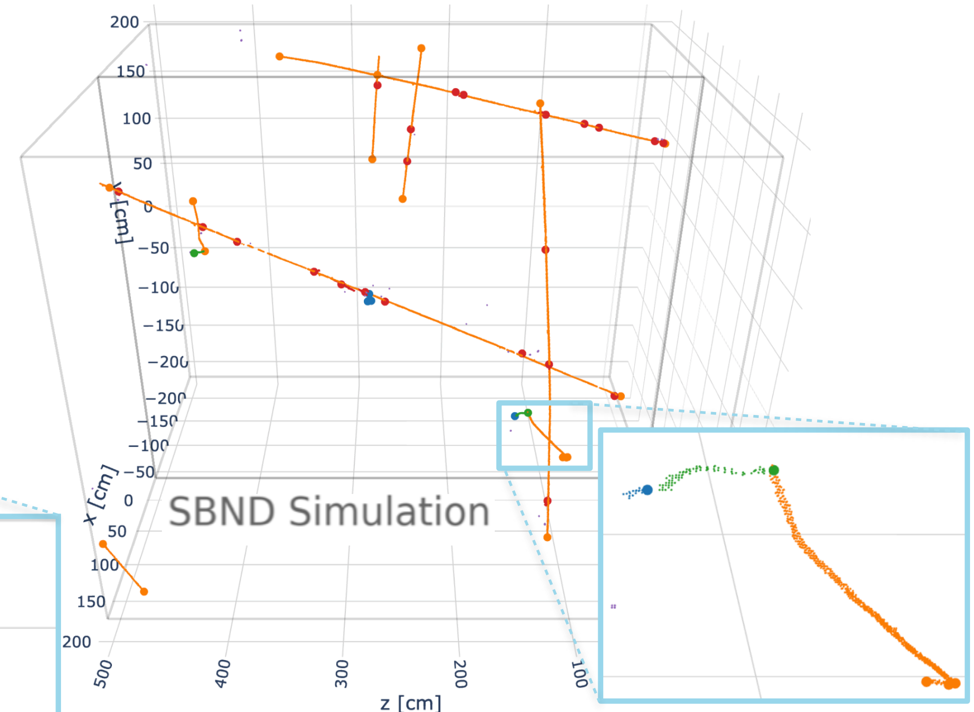
Point proposal network (PPN)



- Predicts track start/end, deltas, michels, and shower starts
- Learned attention mask from decoder UResNet blocks with cross entropy and displacement losses

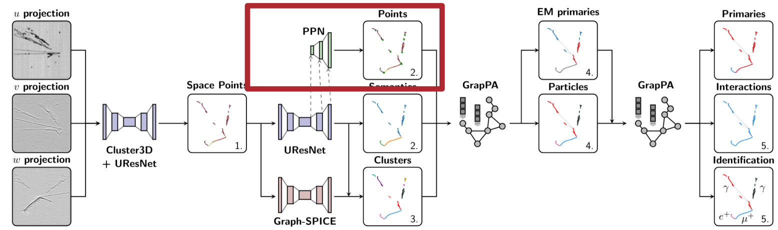


Rescaled charge

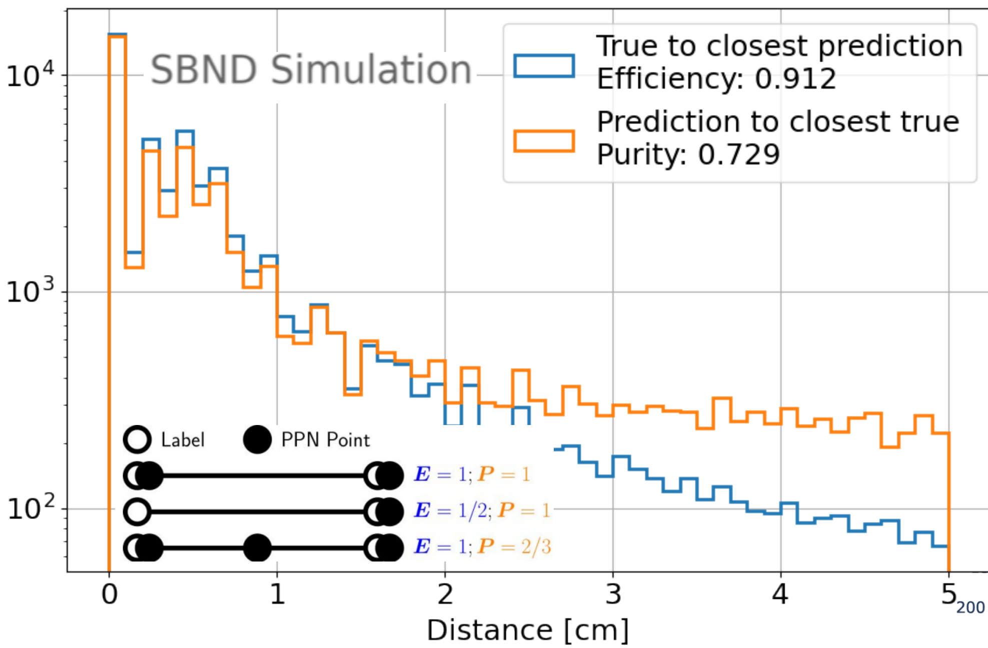


PPN labels

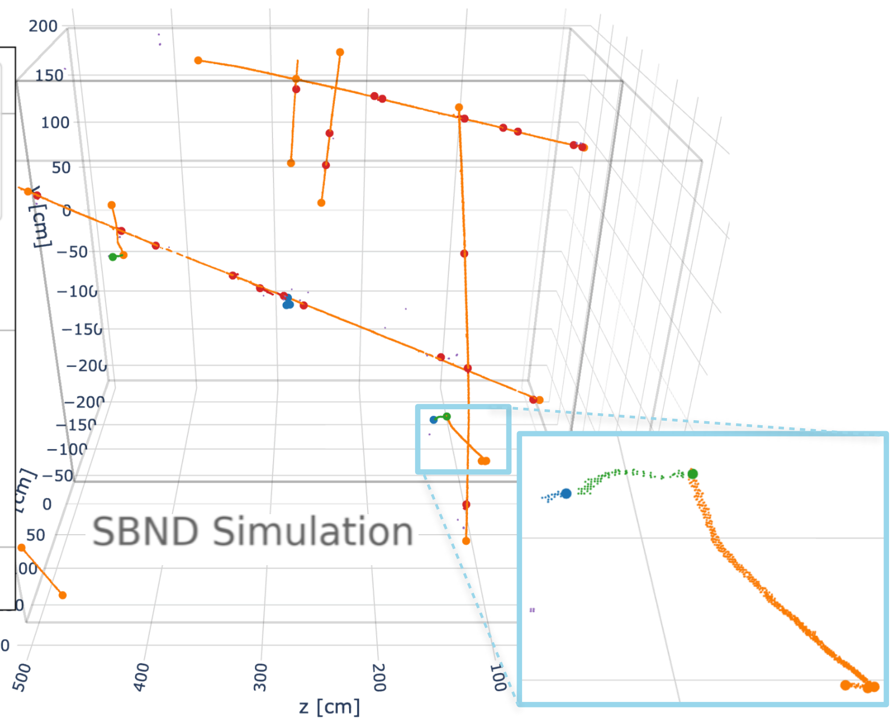
Point proposal network (PPN)



- Predicts track start/end, deltas, michels, and shower starts
- Median distance from true to closest prediction = **0.42 cm**
- Median distance from prediction to closes true = **0.84 cm**

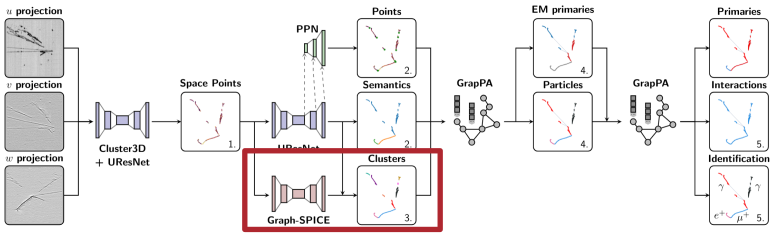


PPN distance metrics

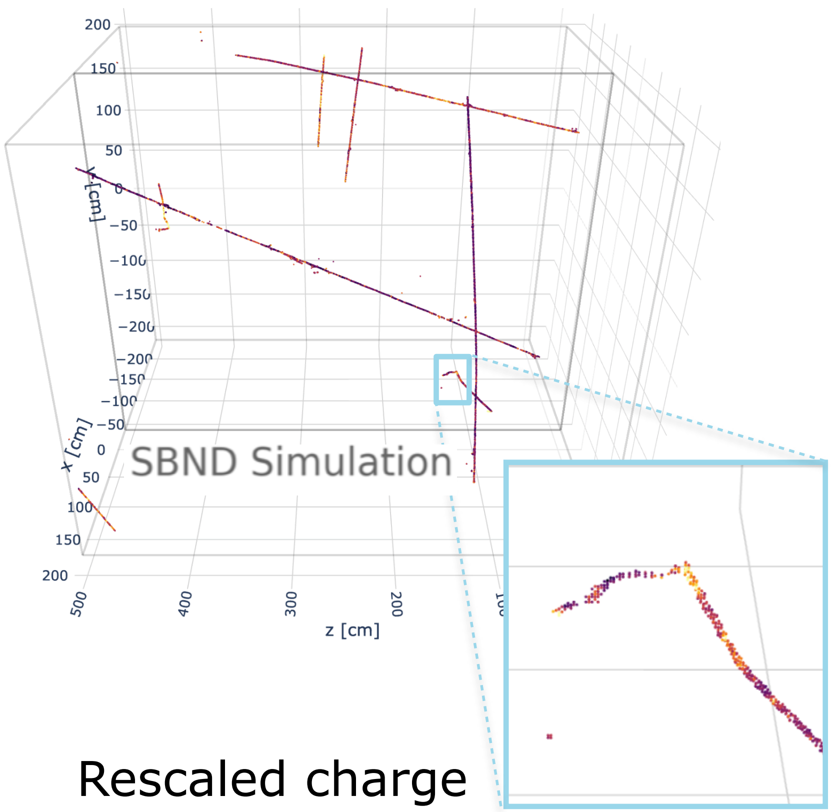


PPN labels

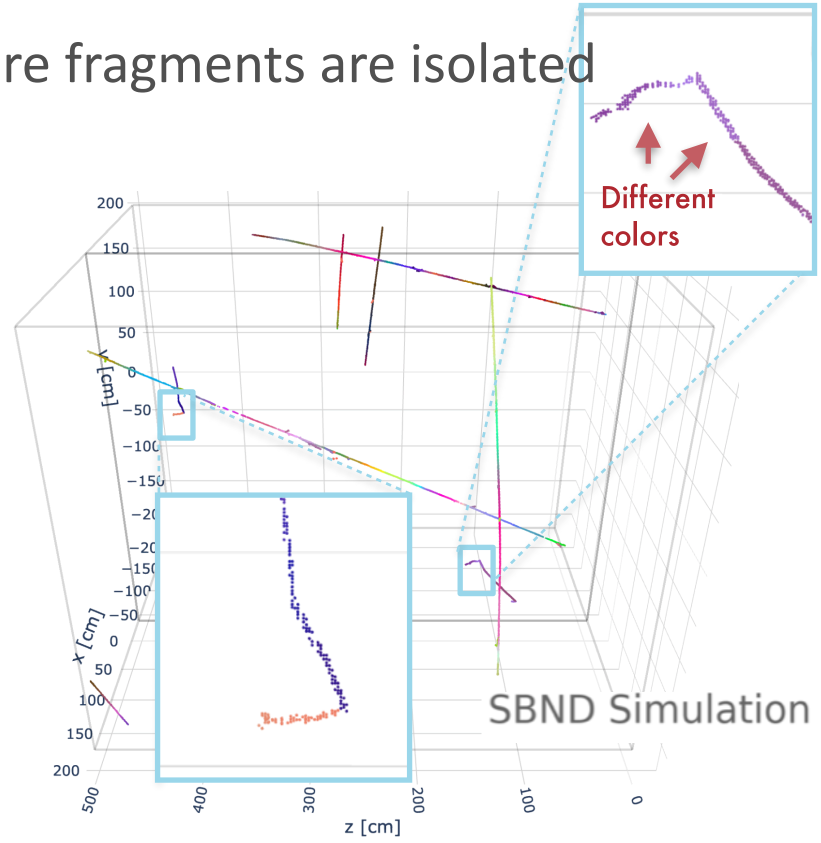
Graph SPICE



- Clusters space points into **fragments** that are aggregated into **particles** by later stages
- Embeds points into state-space where fragments are isolated

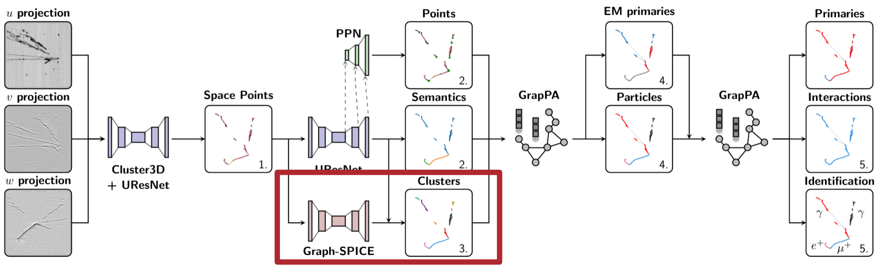


Rescaled charge

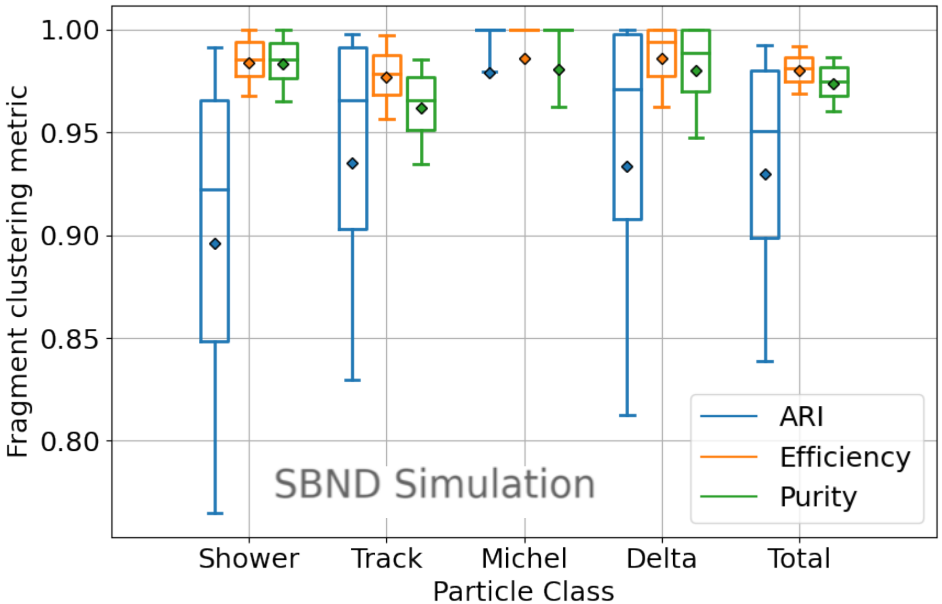


Fragment labels

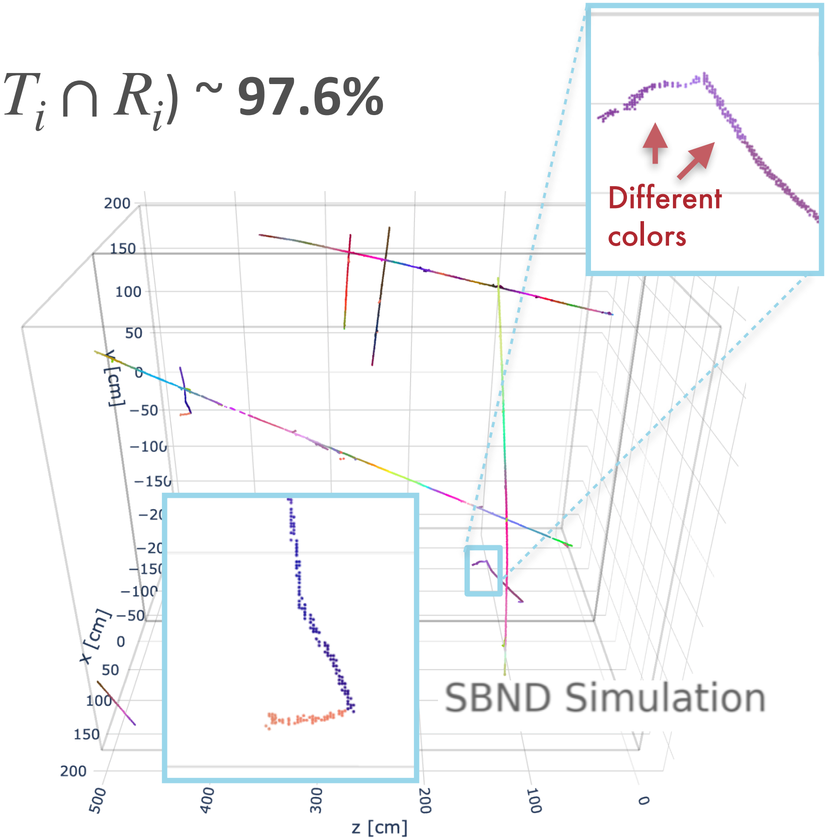
Graph SPICE



- Clusters space-points into **fragments** that are aggregated into **particles** by later stages
- Eff. $(R_i \cap T_i)$ ~ **98.2%** Pur. $(T_i \cap R_i)$ ~ **97.6%**

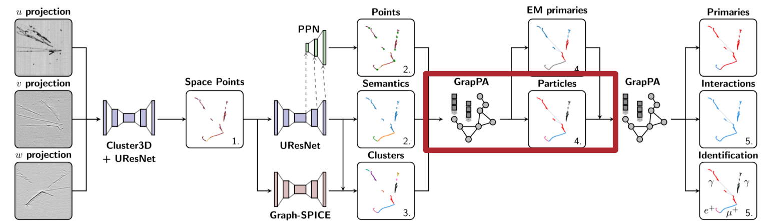


Fragment clustering for each sem. type

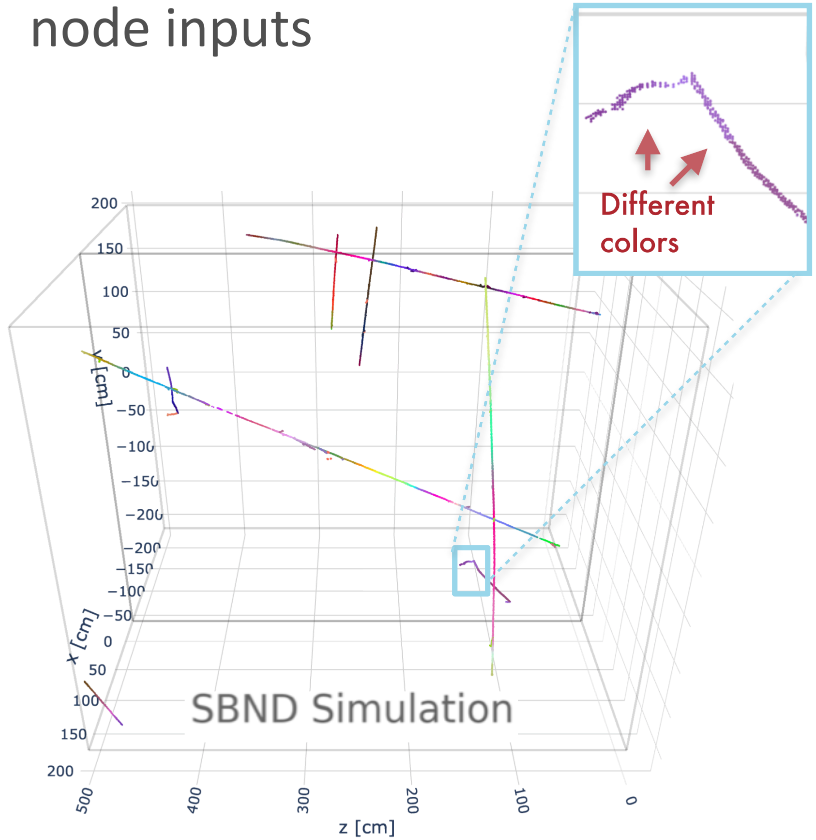


Fragment labels

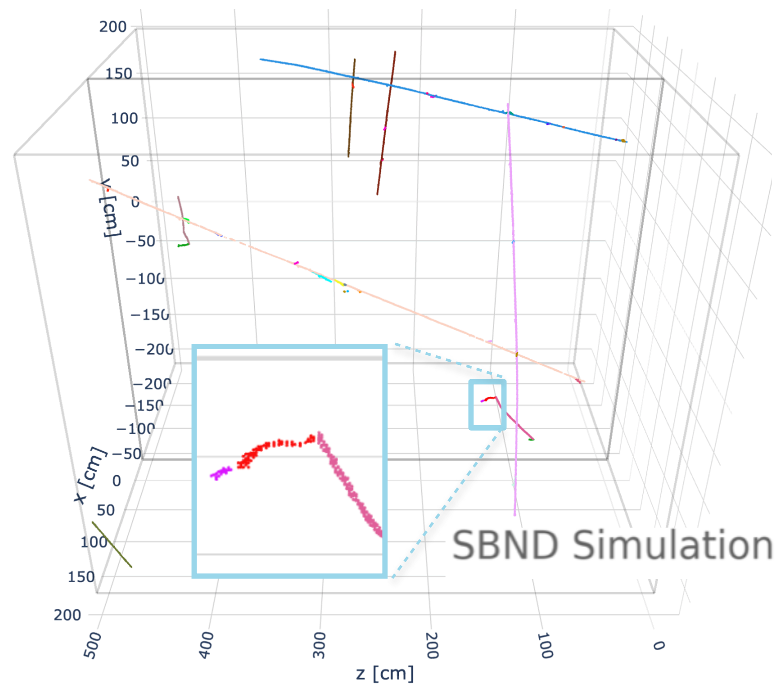
Graph particle aggregator (GrapPA)



- Aggregates **fragments** into **particles** and identifies **shower primaries**
- **Fragment** GNN with geometric edge inputs and charge, PCA, and PPN node inputs

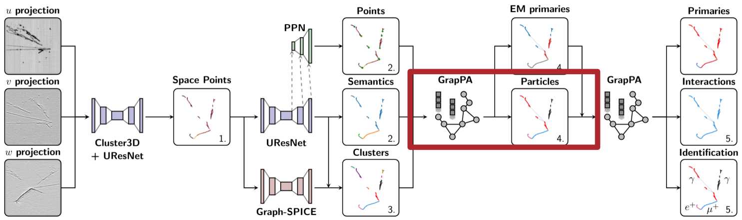


Fragment labels

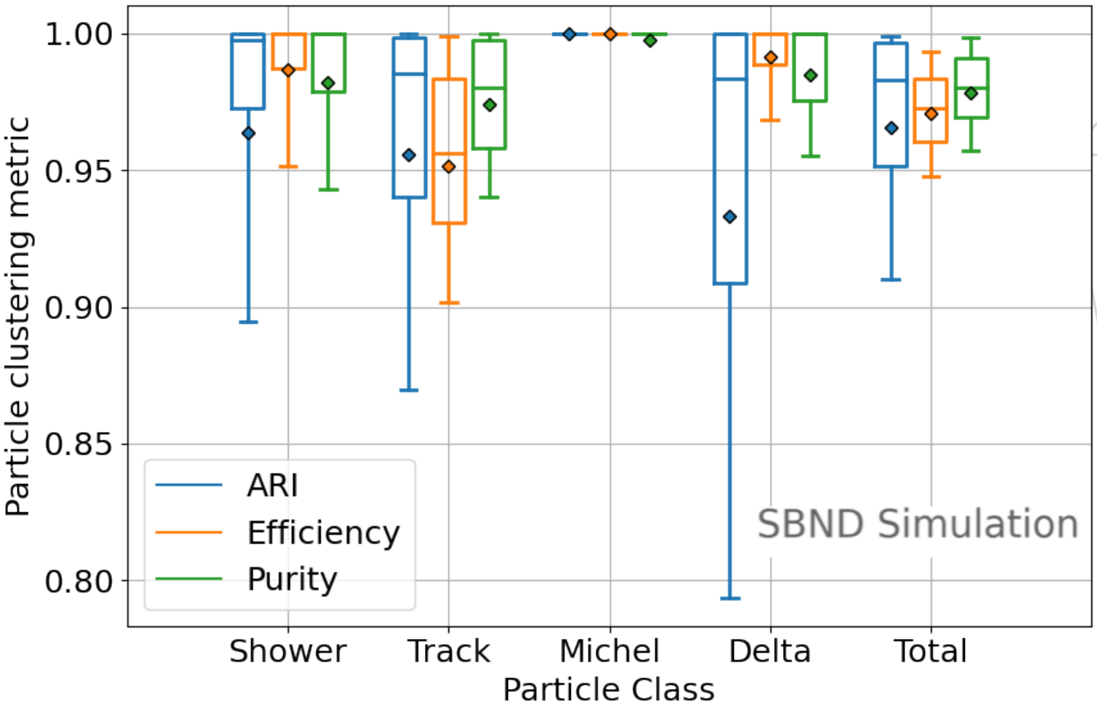


Particle Labels

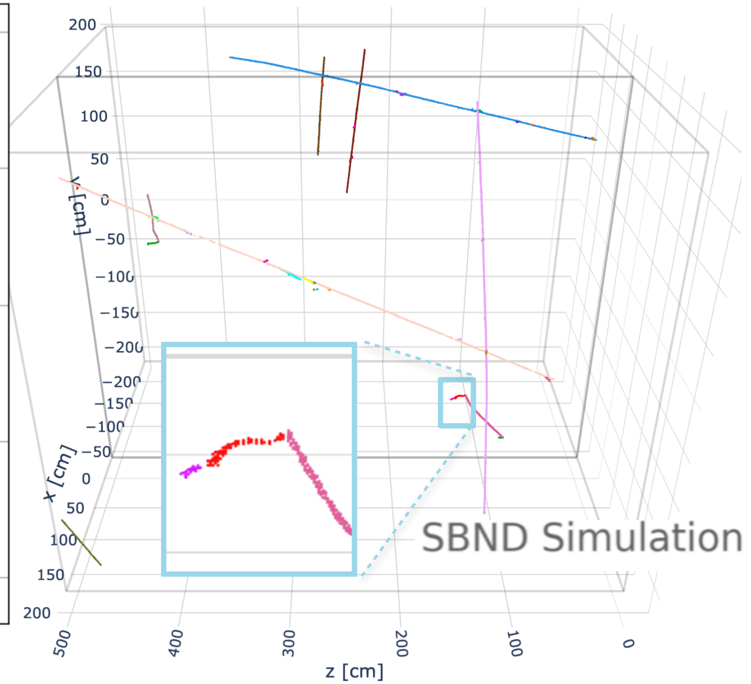
Graph particle aggregator (GrapPA)



- Aggregates **fragments** into **particles** and identifies **shower primaries**
- Eff. $(R_i \cap T_i)$ ~ **98.0%** Pur. $(T_i \cap R_i)$ ~ **98.3%**

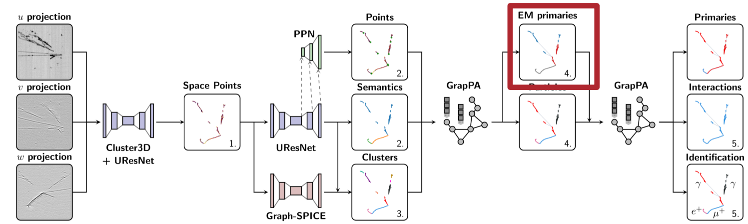


Particle clustering for each sem. type

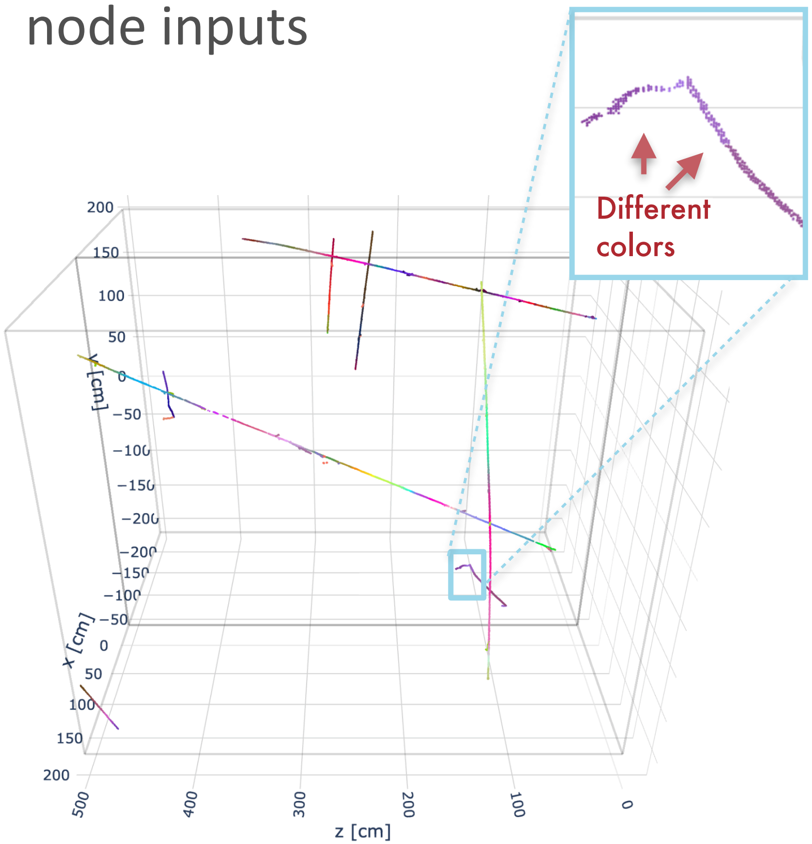


Particle Labels

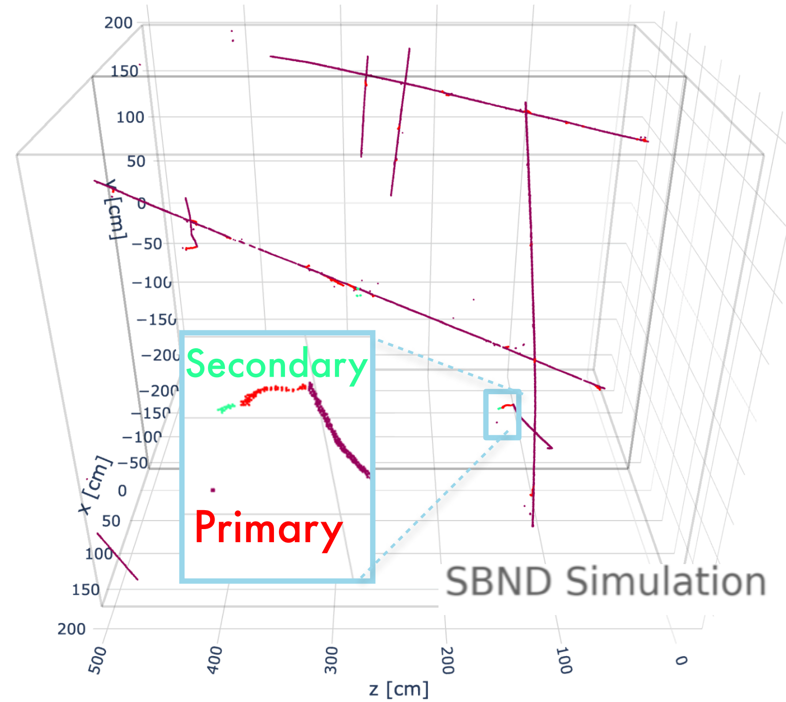
Graph particle aggregator (GrapPA)



- Aggregates **fragments** into **particles** and identifies **shower primaries**
- **Fragment** GNN with geometric edge inputs and charge, PCA, and PPN node inputs

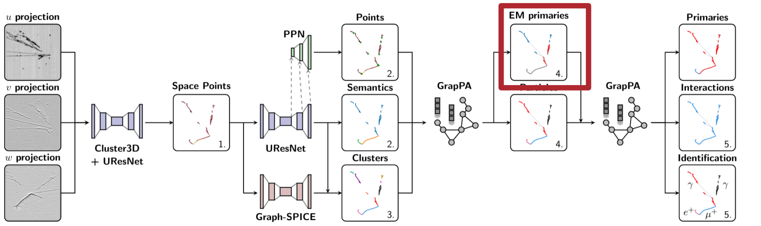


Fragment labels

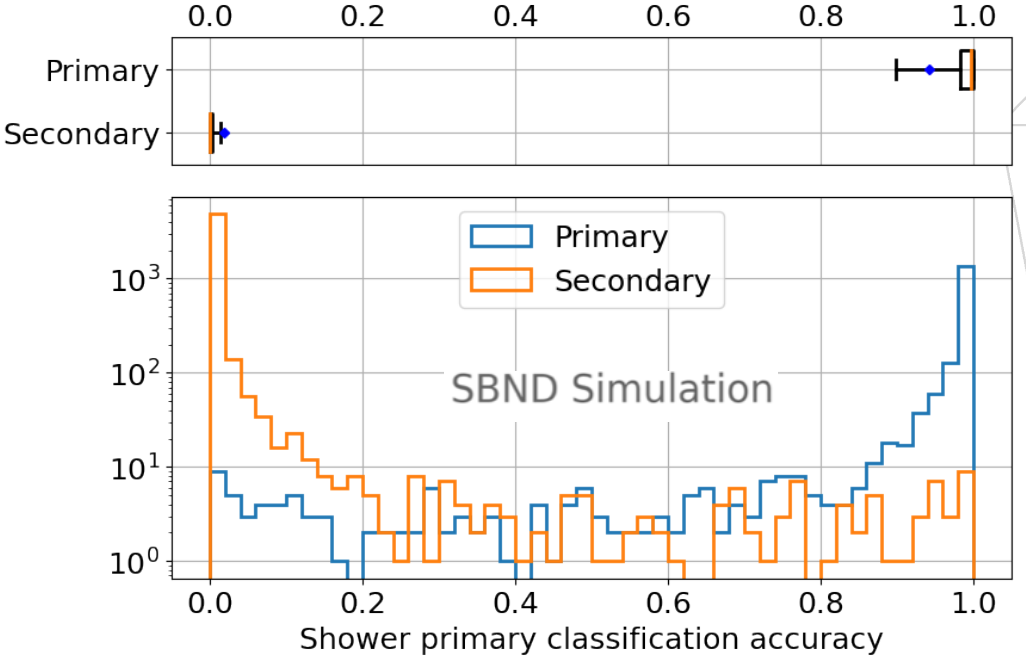


Shower primary labels

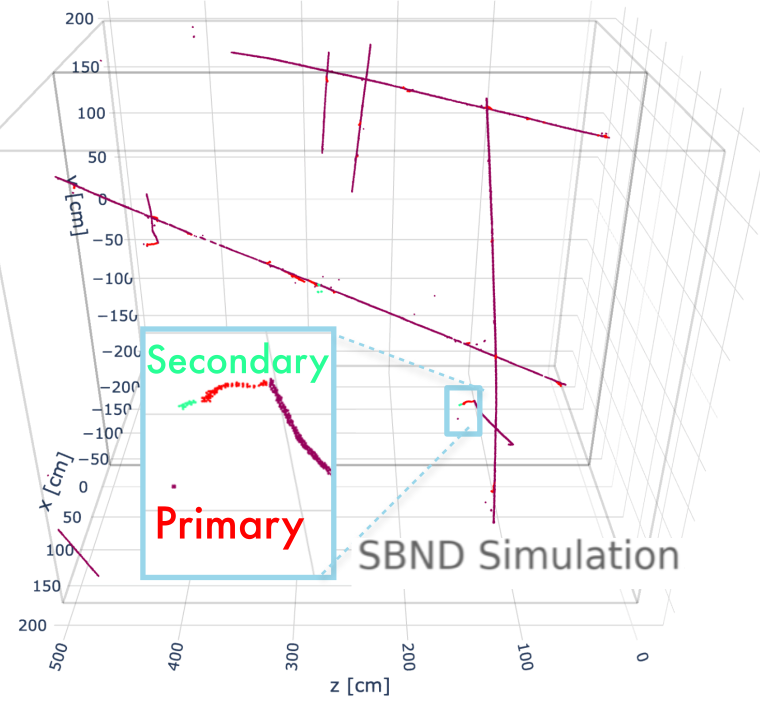
Graph particle aggregator (GrapPA)



- Aggregates **fragments** into **particles** and identifies **shower primaries**
- Overall accuracy **93.3%**, classifies as primaries and secondaries

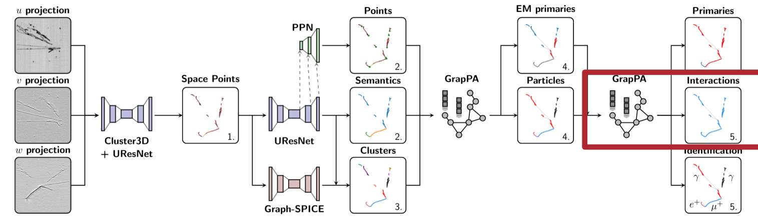


Shower primary classification

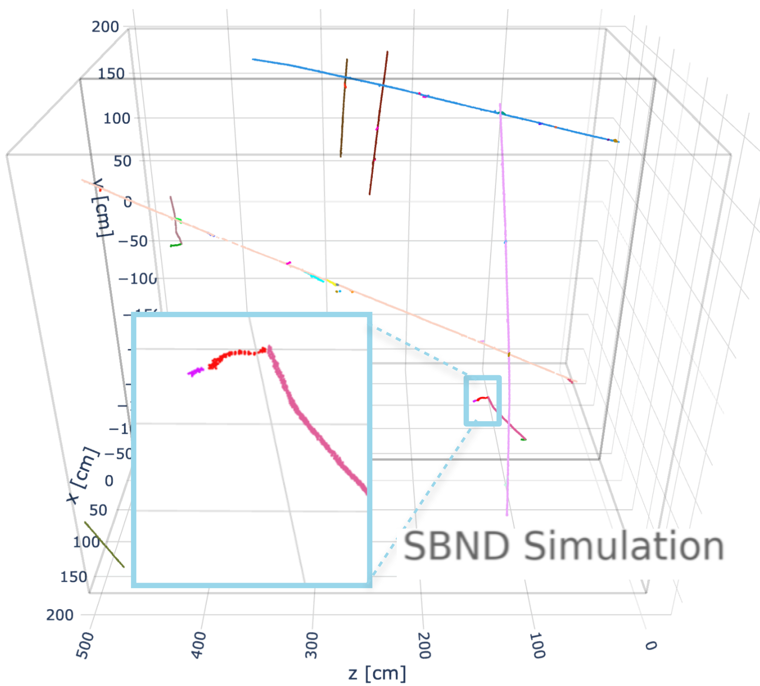


Shower primary labels

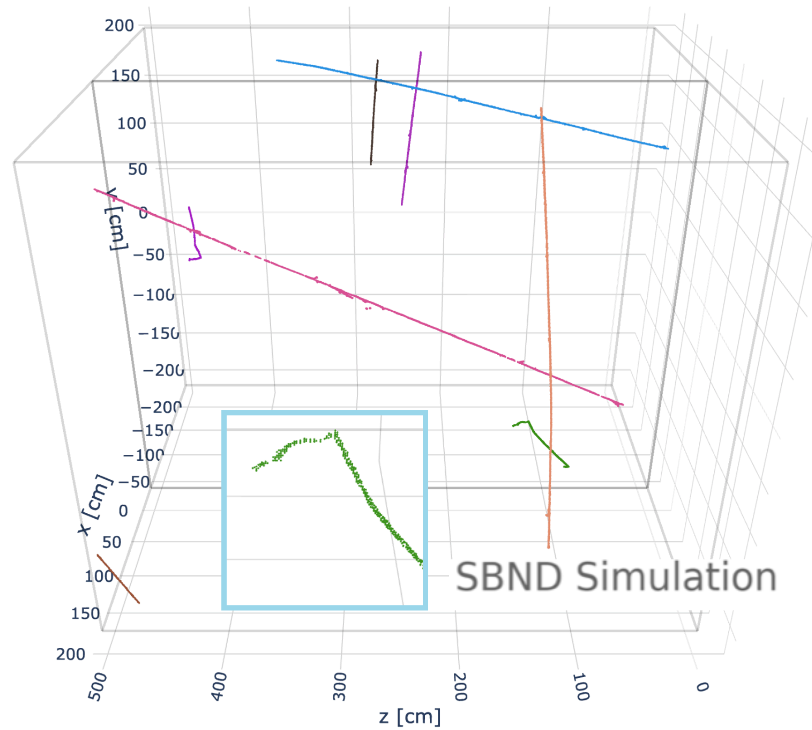
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- **Particle** GNN with geometric edge inputs and charge, PCA, and PPN node inputs

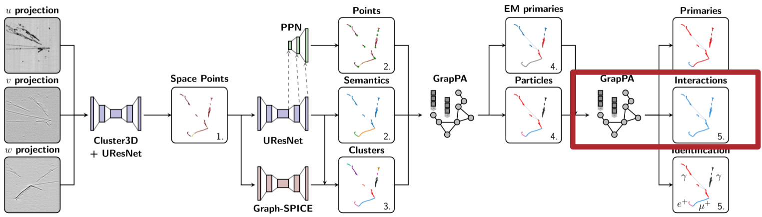


Particle Labels

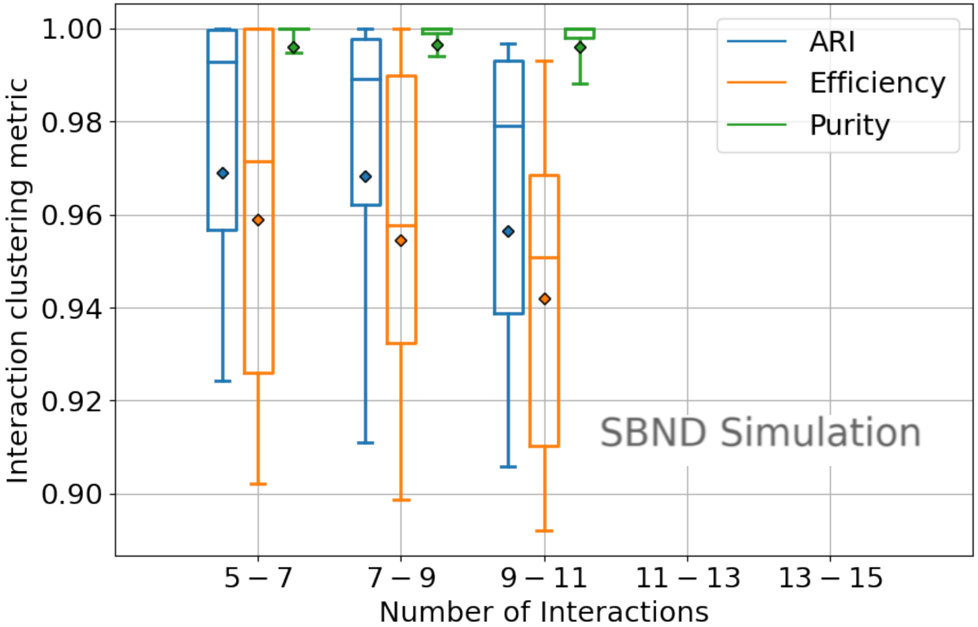


Interaction labels

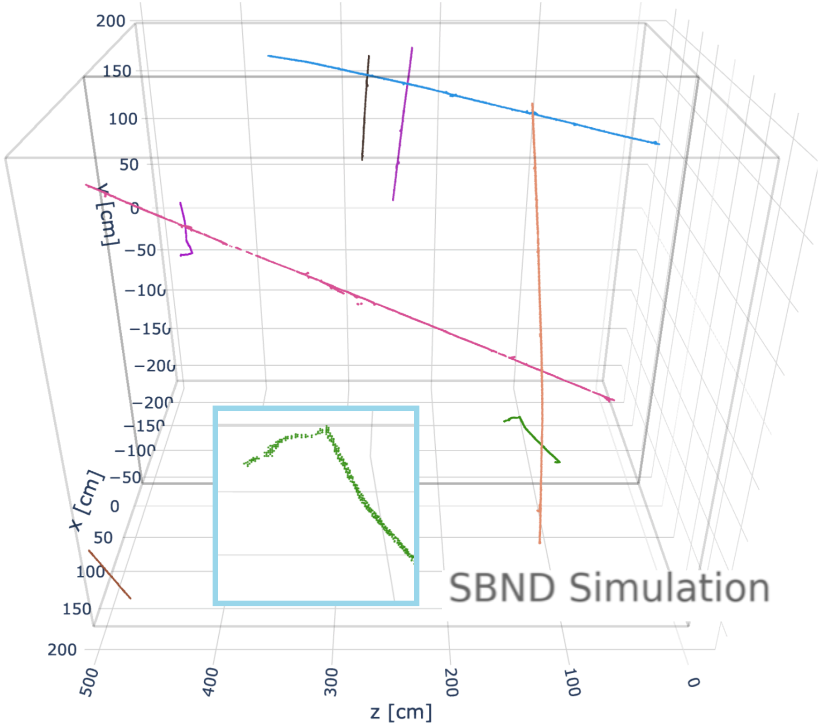
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- Eff. $(R_i \cap T_i)$ ~ **95.1%** Pur. $(T_i \cap R_i)$ ~ **99.6%**

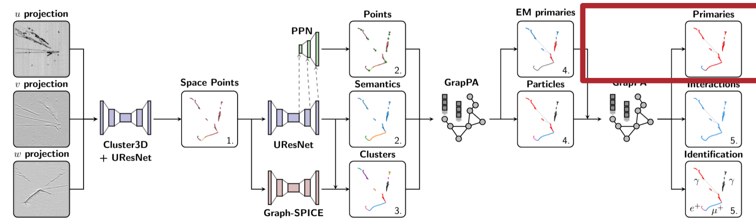


Interaction Clustering Metrics

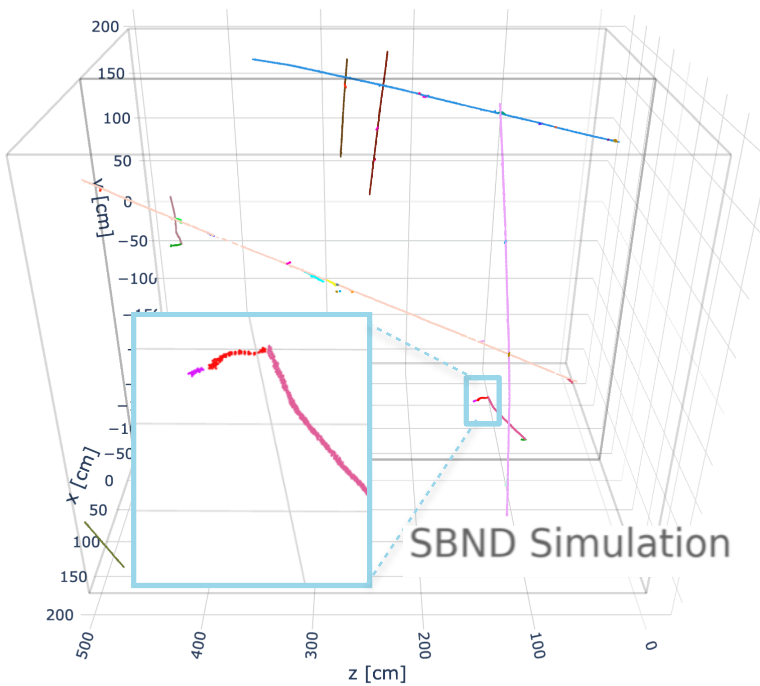


Interaction labels

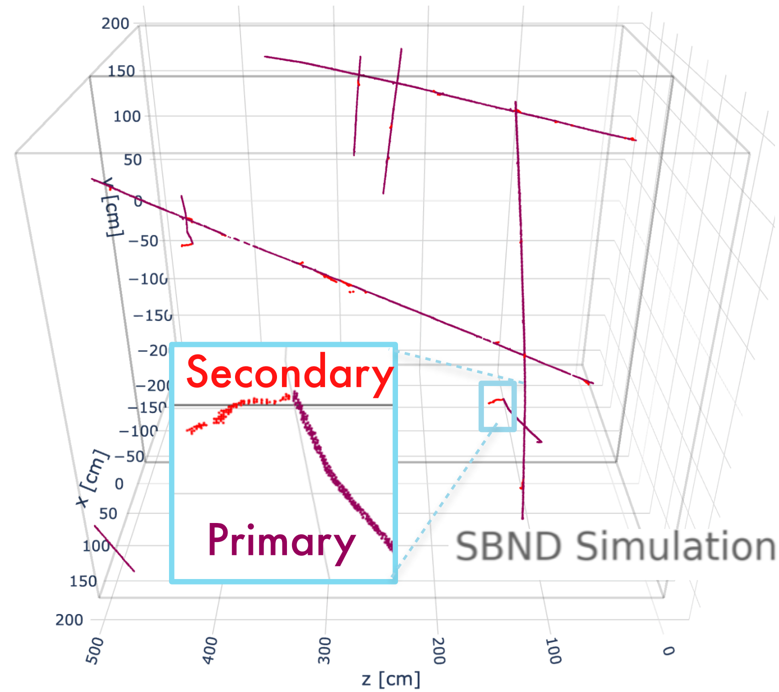
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- **Particle** GNN with geometric edge inputs and charge, PCA, and PPN node inputs

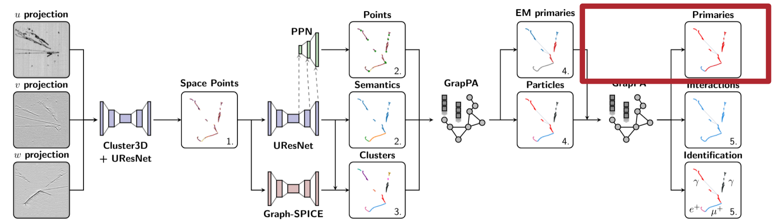


Particle Labels

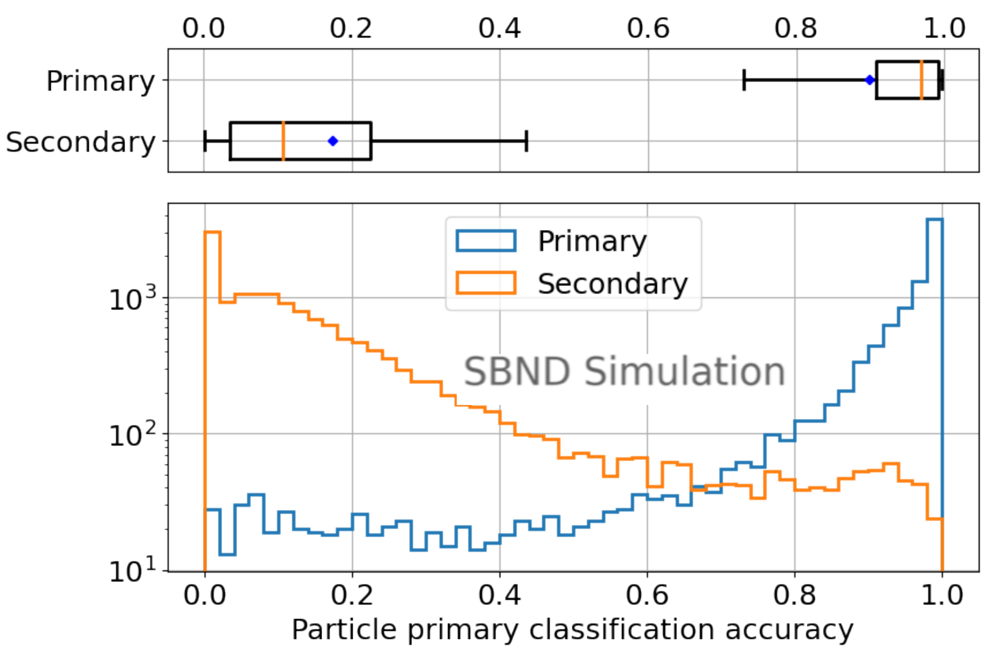


Primary labels

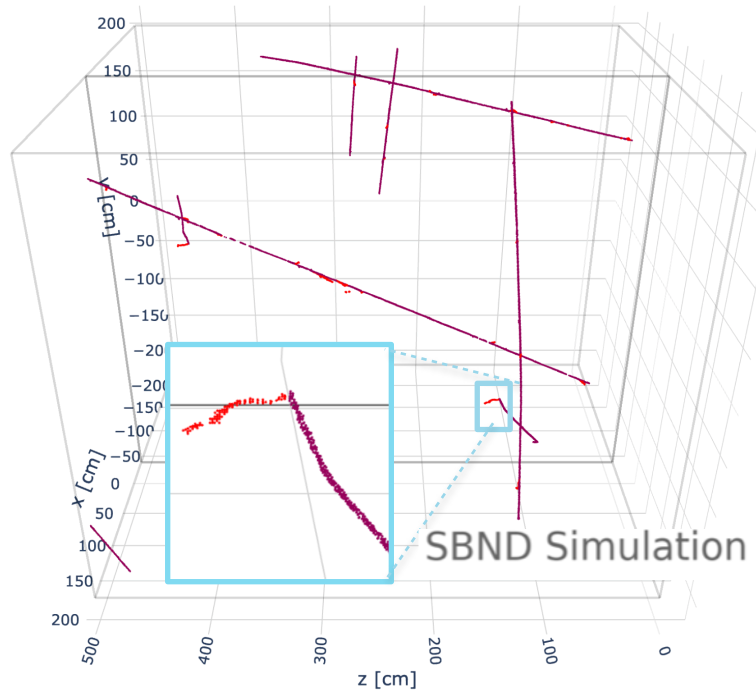
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- Overall accuracy **92.7%**

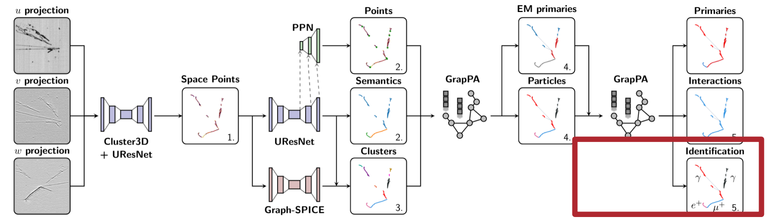


Primary accuracy

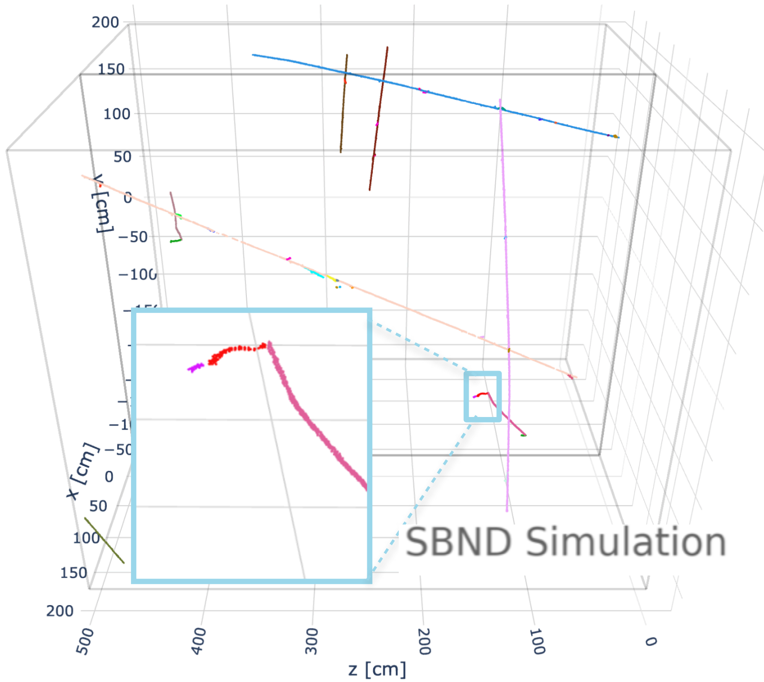


Primary labels

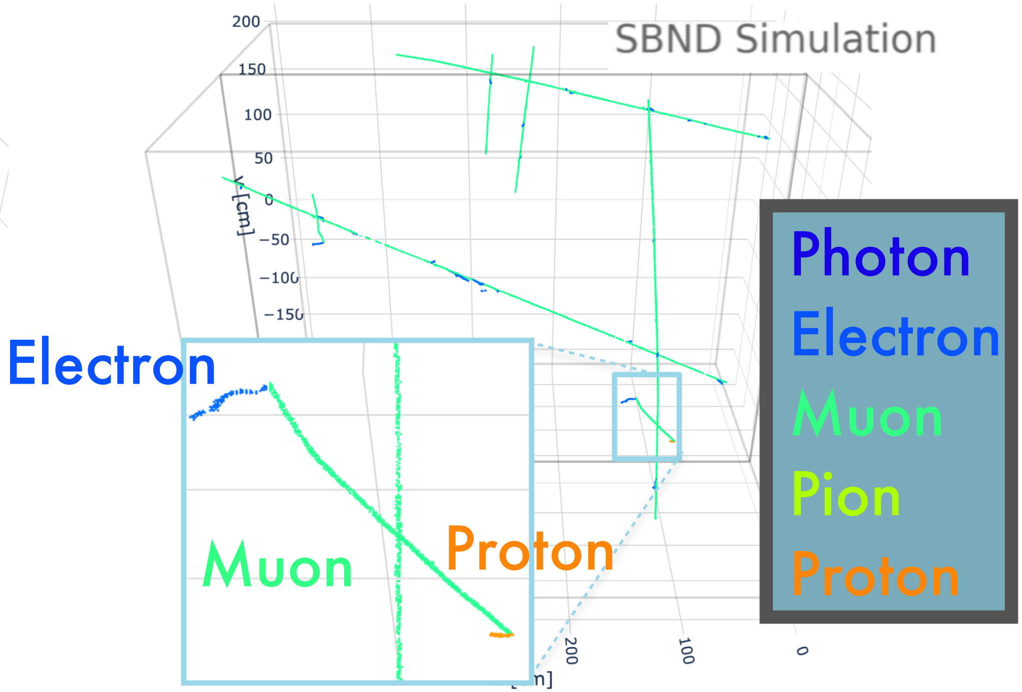
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- **Particle** GNN with geometric edge inputs and charge, PCA, and PPN node inputs

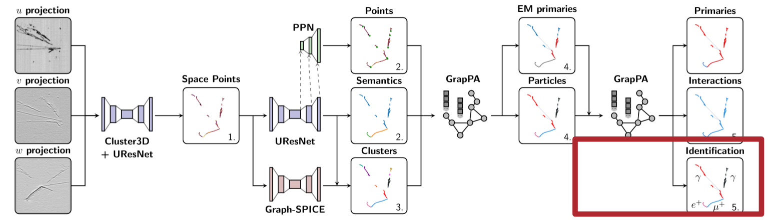


Particle Labels

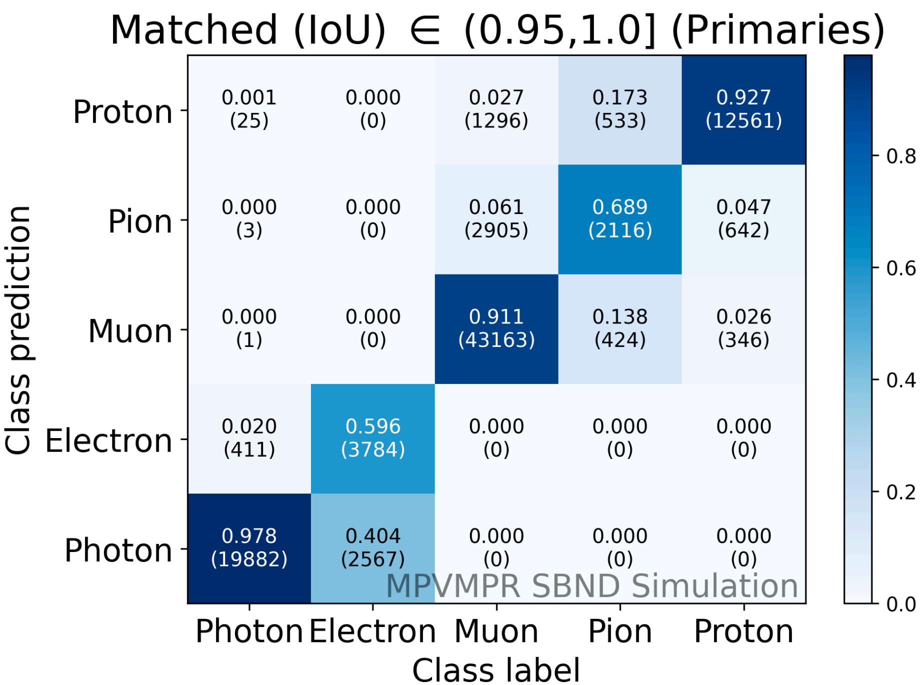


PDG labels

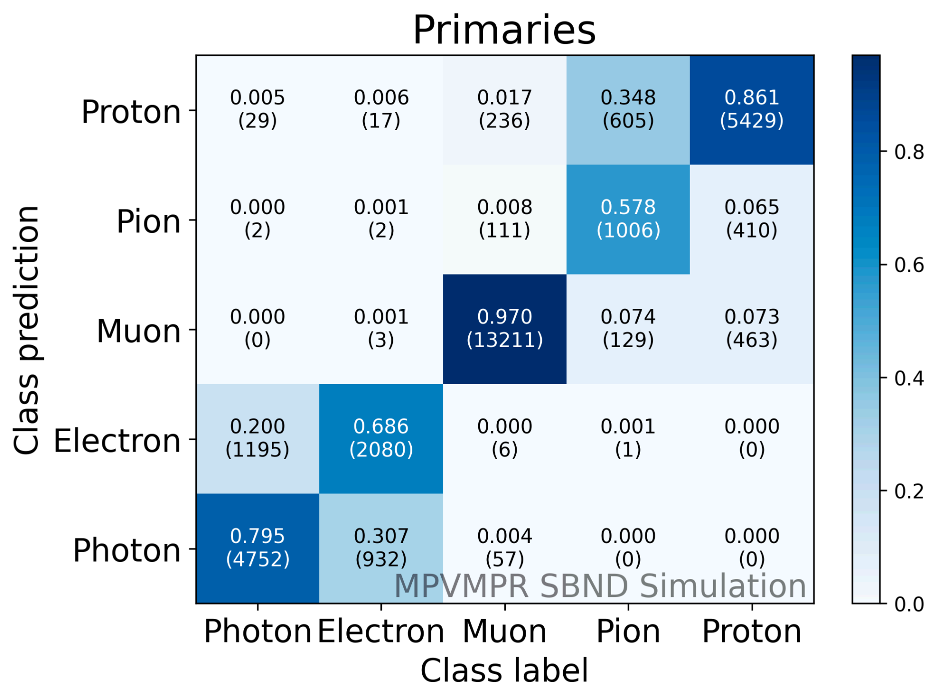
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- Primary PID accuracy **85.5%**
- Electron-photon confusion from poor class balancing during training

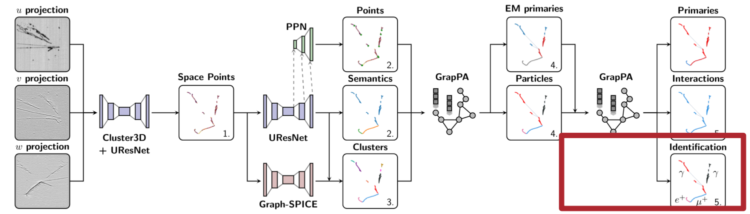


MPVMPR Test v01 (this sample)

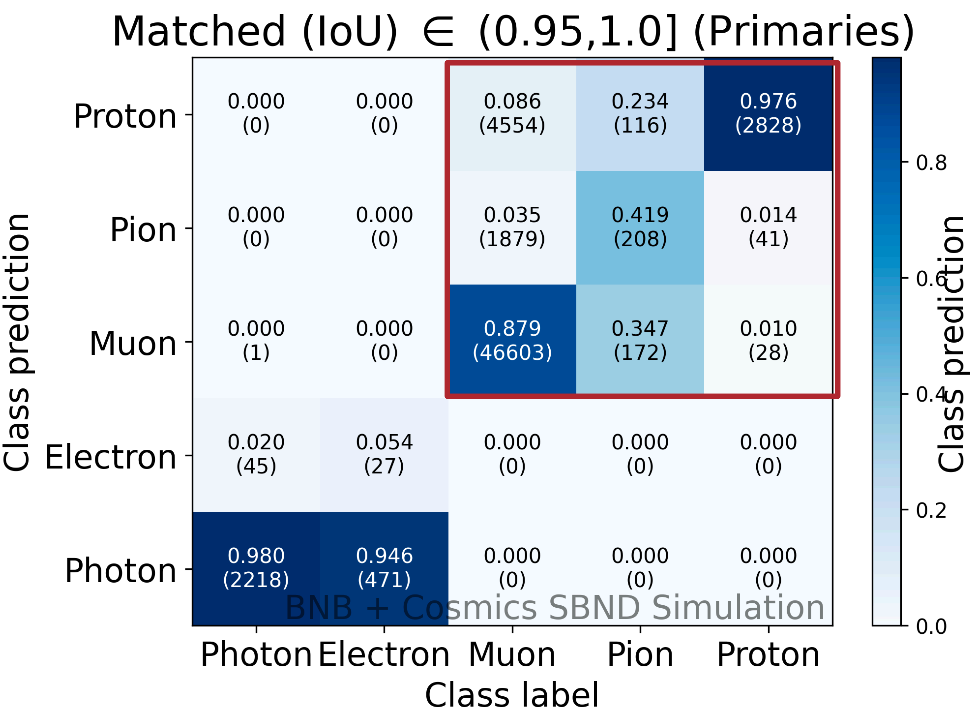


MPVMPR Test v00 (prev. sample)

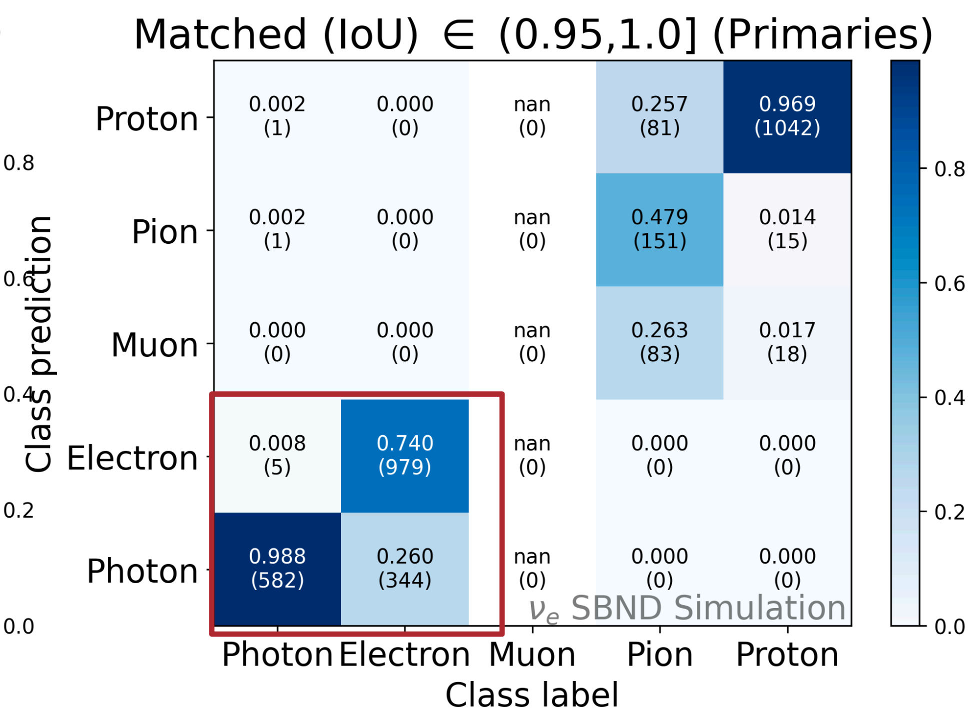
Graph particle aggregator (GrapPA)



- Aggregates **particles** into **interactions** and identifies **primaries** and **PID**
- Primary PID accuracy **85.5%**
- Electron-photon confusion from poor class balancing during training



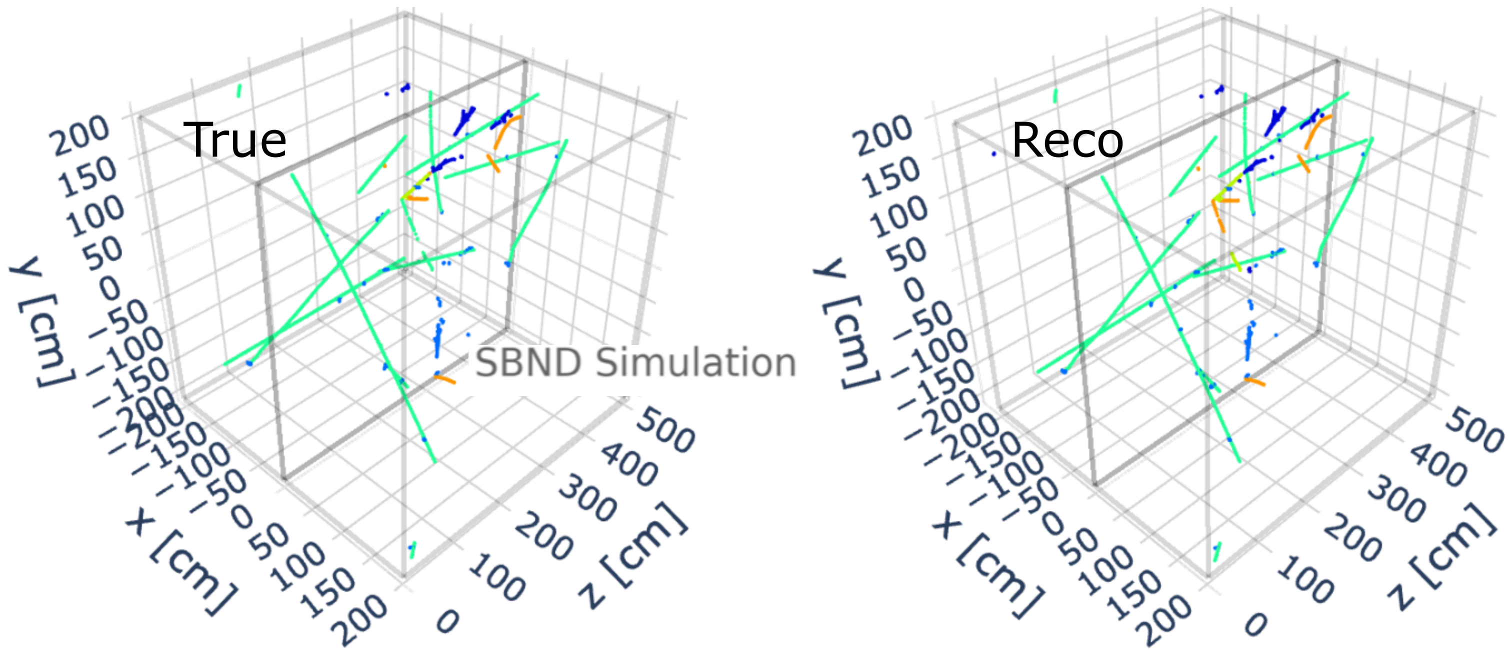
BNB + cosmics + rockbox



Intrinsic nu e

Conclusion

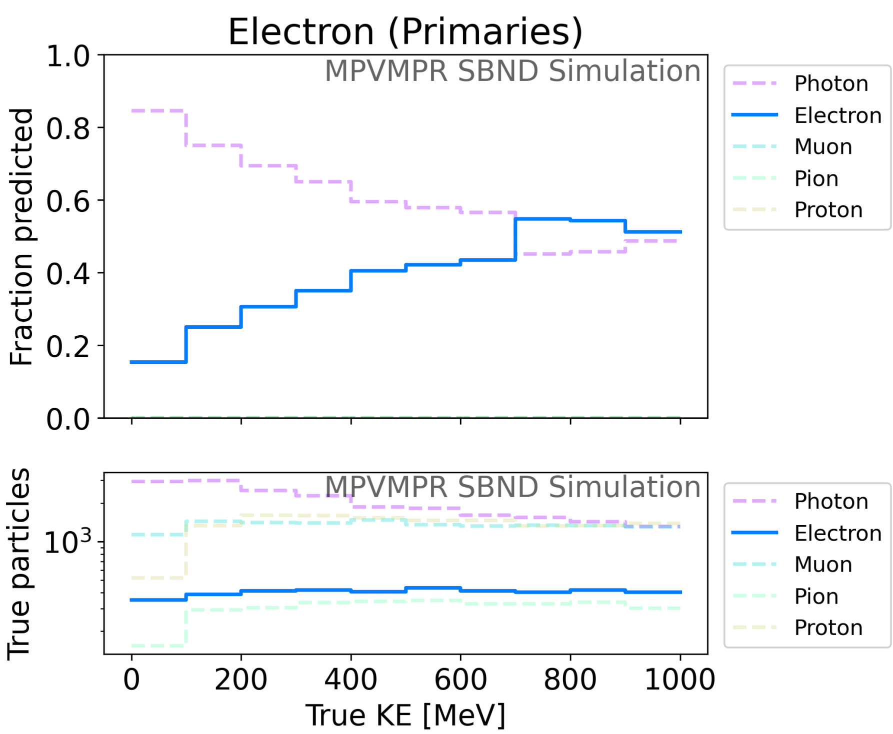
- SBND is able to successfully reconstruct LArTPC wire readouts using SPINE
- Clustering works well, primary identification and PID need deeper studies
- Stay tuned for future SBND analyses
  - C. Fan - ν_e CC selection
  - N. Oza - Detector calibration using michel electrons
  - B. Carlson - ν_μ CC selection

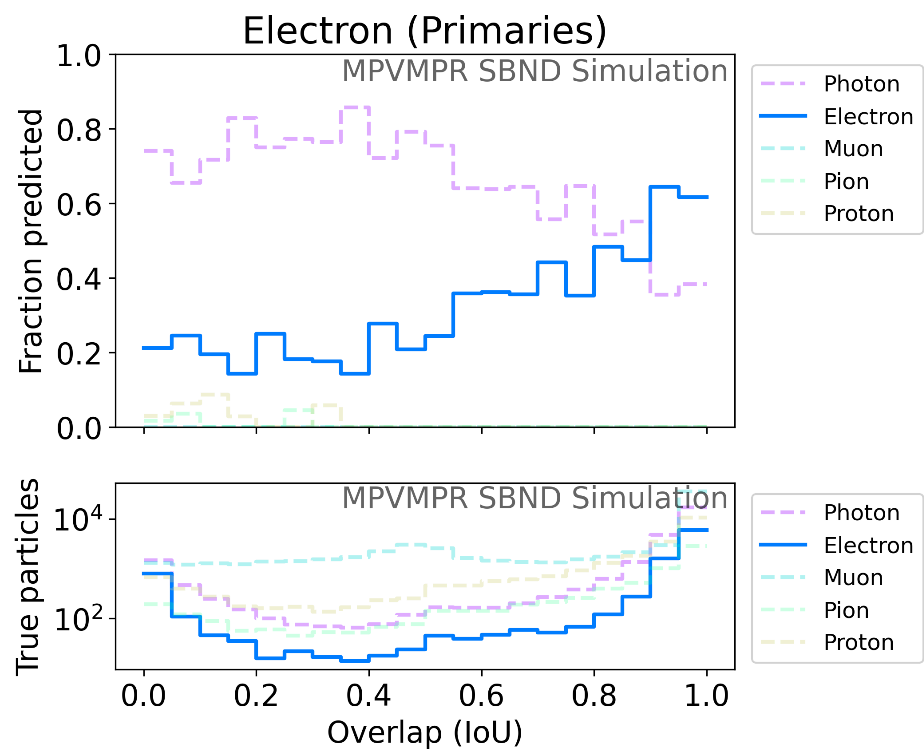


Thanks!

PID - Electrons

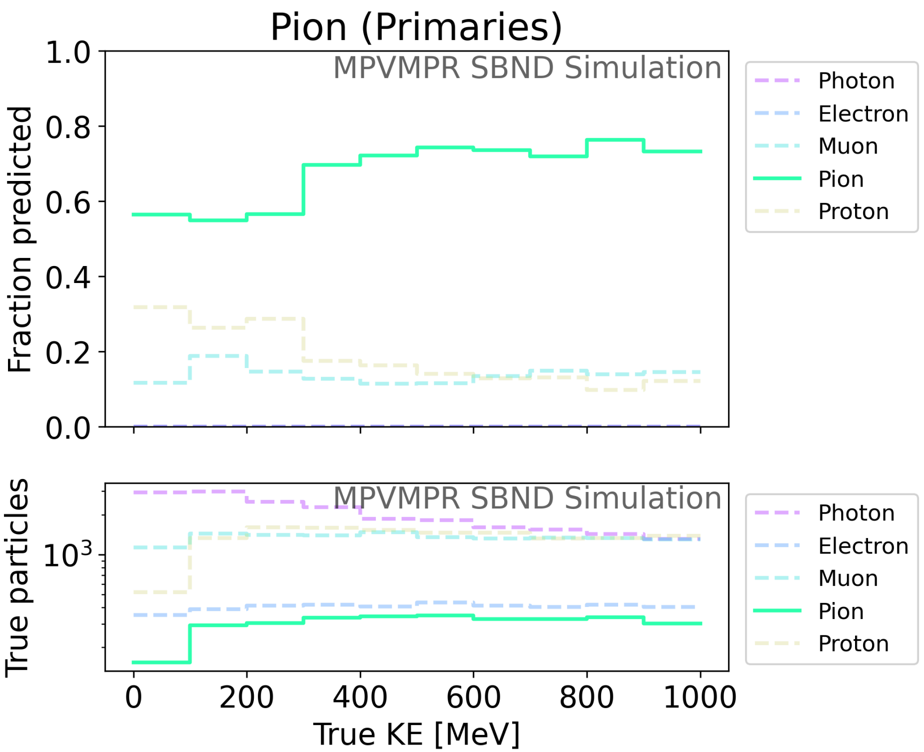
- More photons in mpv sample than electrons for low KE (left)
- Showers with more space-points shared between true and reco (Overlap IoU) leads to better PID
- Low overlap -> missing fragments from showers -> inflated confusion for small fragments of showers

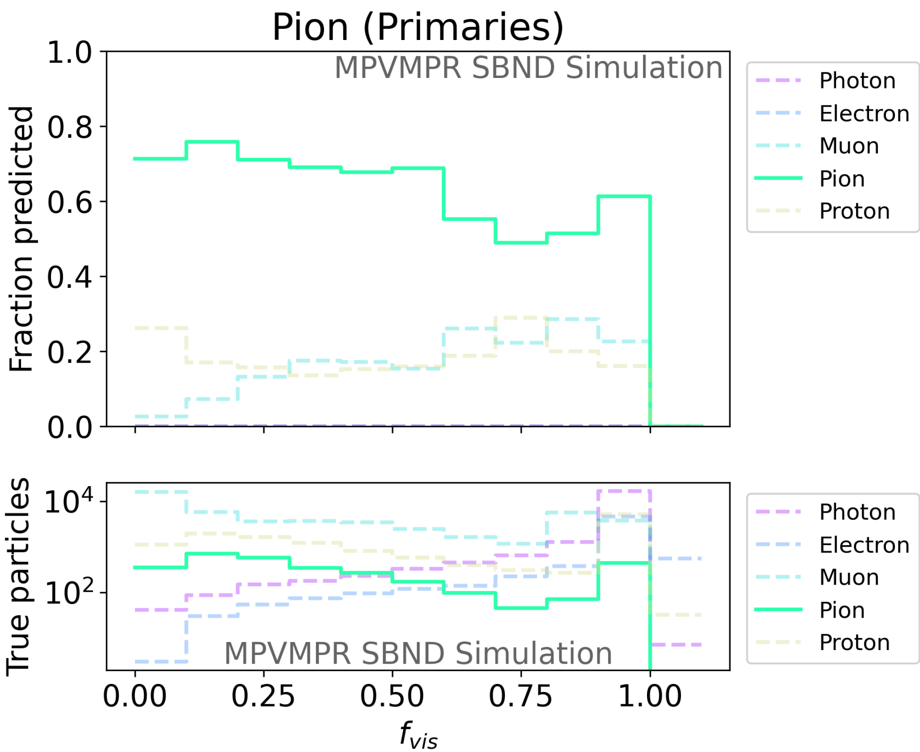




PID - Pions

- Low KE pions classified as protons (left)
- Higher fraction of visible energy $f_{vis} = E_{vis}/E_{tot}$ leads to classification of muons
- Strangely, low f_{vis} leads to better classification

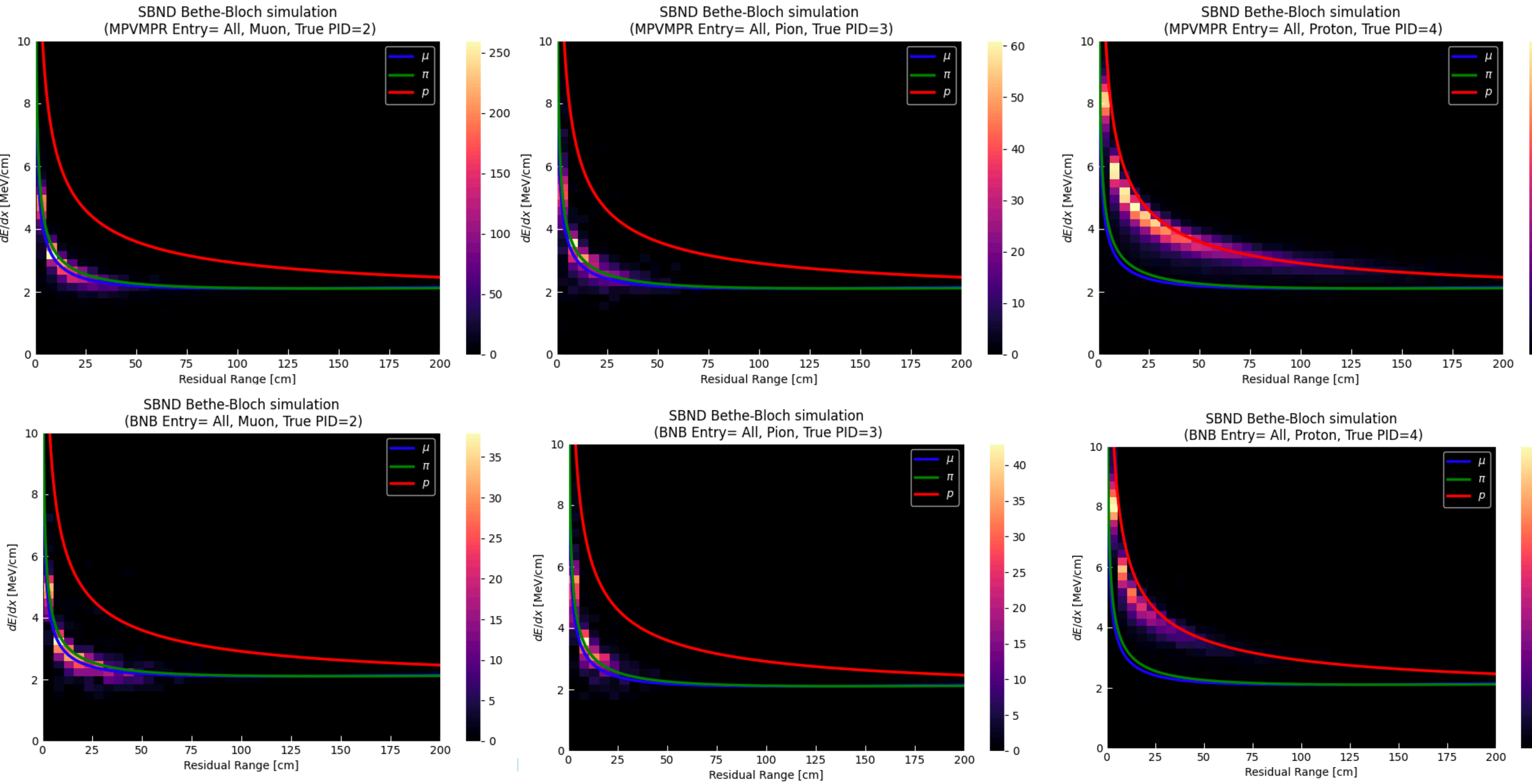




dE/dx Studies

- **C. Fan** investigating dE/dx of stopped tracks
- Clear separation between proton and muon/pion
- <u>*Tiny*</u> separation between muon and pion

C. Fan
dE/dx studies



Training

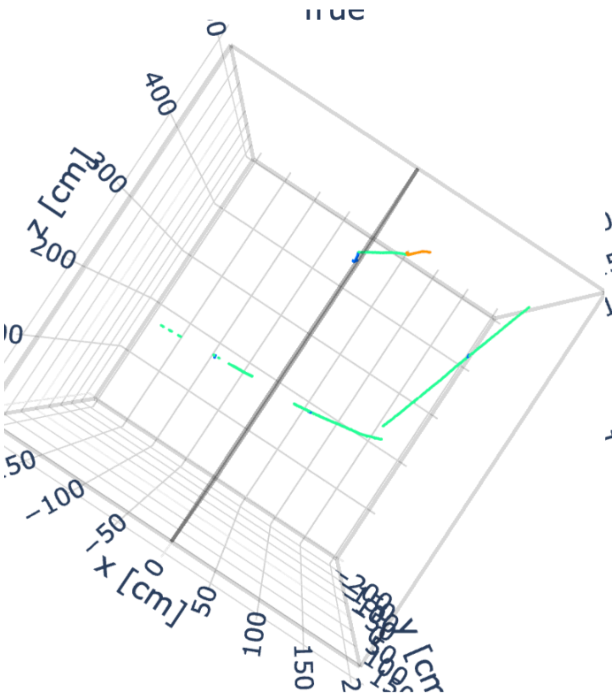
```
MultiMax      : 5
MultiMin      : 3
ParticleParameter: {
    PDGCode       : [   [-13,13],    [-13,13],      [11,-11],      [22],      [2212]]
    MinMulti      : [          0,           0,            0,          0,           0]
    MaxMulti      : [          5,           5,            2,          3,           5]
    ProbWeight    : [          5,           5,            1,          2,           1]
    KERange       : [ [0.0,20.0],   [0.0,2.0],   [0.0,1.0],   [0.0,1.0],   [0.0,1.0]]
    MomRange      : []
}
```

In-time rain v01 parameters

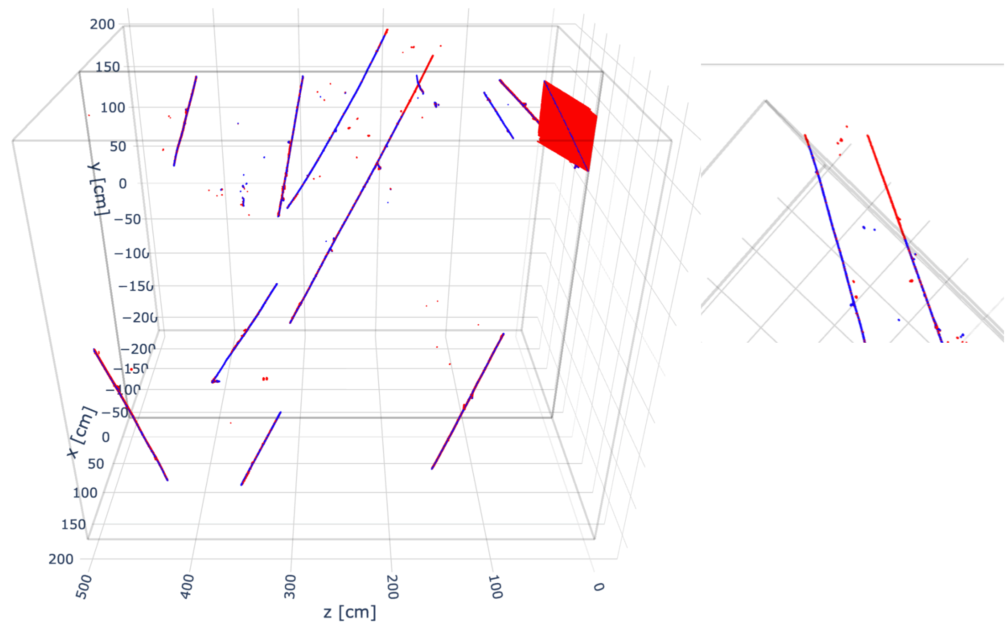
```
MultiPartRain2.MultiMax: 4
MultiPartRain2.MultiMin: 2
MultiPartRain2.ParticleParameter.PDGCode       : [   [-13,13], [-13,13],      [2212]]
MultiPartRain2.ParticleParameter.MinMulti      : [          0,        0,           0]
MultiPartRain2.ParticleParameter.MaxMulti      : [          5,        5,           5]
MultiPartRain2.ParticleParameter.ProbWeight    : [          5,        5,           1]
MultiPartRain2.ParticleParameter.KERange       : [ [0.0,20.0], [0.0,2.0], [0.0,1.0]]
```

Out-of—time rain v01 parameters

Out of time - CPA



Out of time - APA



De-ghosted labels

Intrinsic nu e