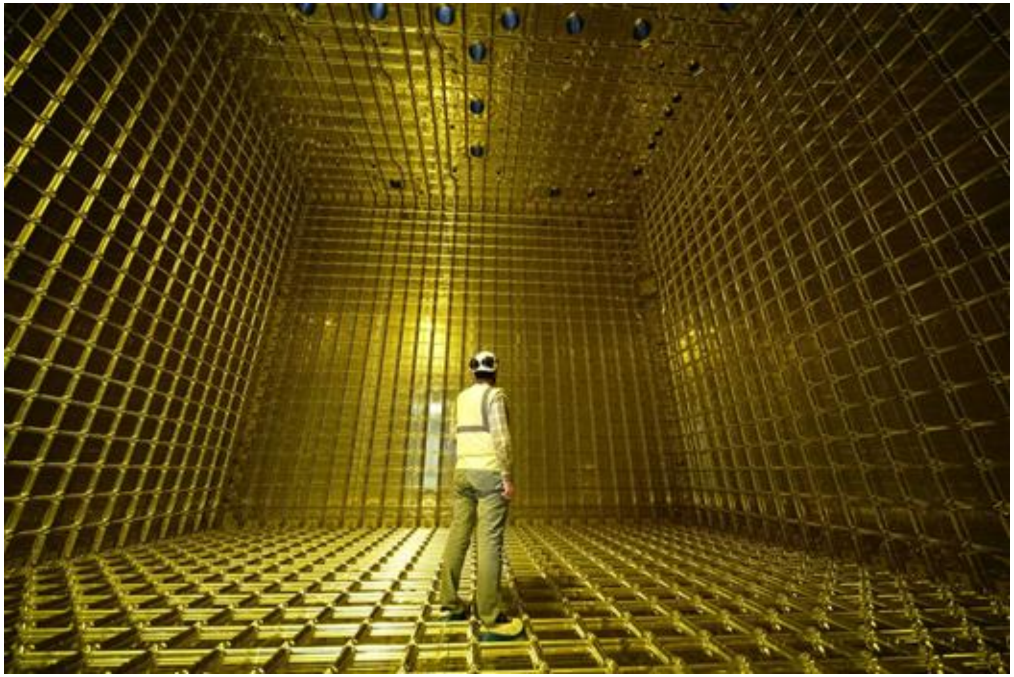
Learning pairwise associations for event slicing in ProtoDUNE-SP

FutureData4EU
Training Future Big Data Experts for Europe



Credit: CERN

Christina Karagianni
University of Ferrara, INFN Ferrara
ML4Jets 2026, Vienna, Austria
Sept 15th, 2026

Co-funded by the European Union

"FutureData4Eu (Grant. Agreement n. 101126733) Co- Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them."

OUTLINE

- **Physics & Detectors**
- Pandora & Event Slicing
- Learning Pairwise Associations
- Preliminary Results
- Summary

Neutrino physics questions
Beyond the Standard Model physics

Credit: Pandey, INSS 2025

Neutrino physics questions
Beyond the Standard Model physics

Neutrino oscillations ($L/E \simeq 1/\Delta m^2$) $\rightarrow$ non-zero mass

Credit: Pandey, INSS 2025

Neutrino physics questions
Beyond the Standard Model physics

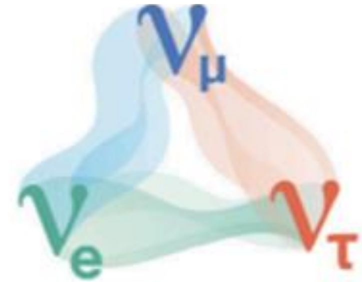
Neutrino oscillations ($L/E \simeq 1/\Delta m^2$) → non-zero mass

Credit: Pandey, INSS 2025

Flavor eigenstates Mass eigenstates

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \underbrace{\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}}_{U_{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

Neutrino physics questions
Beyond the Standard Model physics



Credit: Pandey, INSS 2025

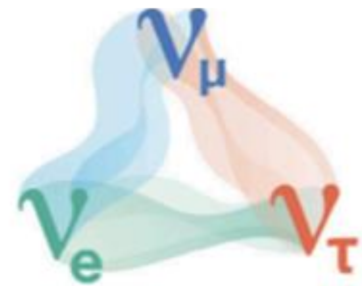
Neutrino oscillations ($L/E \simeq 1/\Delta m^2$) → non-zero mass

- CP violation

Flavor eigenstates Mass eigenstates

$$\begin{pmatrix} v_e \\ v_\mu \\ v_\tau \end{pmatrix} = \underbrace{\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}}_{U_{PMNS}} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

Neutrino physics questions
Beyond the Standard Model physics



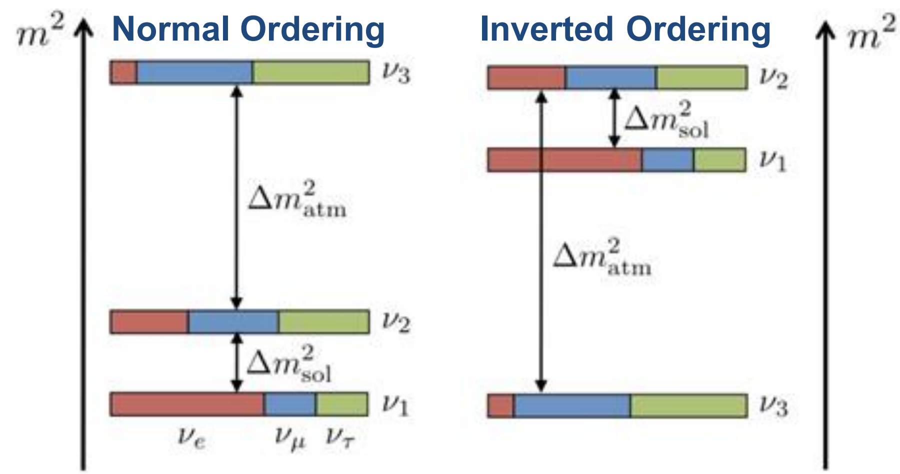
Credit: Pandey, INSS 2025

Neutrino oscillations (L/E ≃ 1/Δm²) → non-zero mass

- CP violation
- Neutrino mass ordering (normal or inverted)

Flavor eigenstates Mass eigenstates

$$\begin{pmatrix}\nu_e\\ \nu_\mu\\ \nu_\tau\end{pmatrix} = \underbrace{\begin{pmatrix}U_{e1} & U_{e2} & U_{e3}\\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3}\\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3}\end{pmatrix}}_{U_{PMNS}} \begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}$$



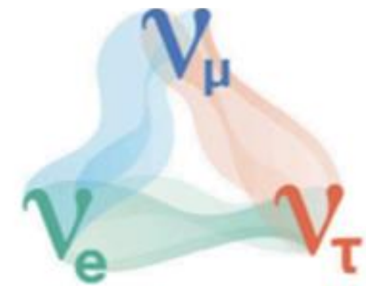
Credit: JUNO Collaboration

Neutrino physics questions
Beyond the Standard Model physics

Neutrino oscillations (L/E ≃ 1/Δm^2) → non-zero mass

- CP violation
- Neutrino mass ordering (normal or inverted)

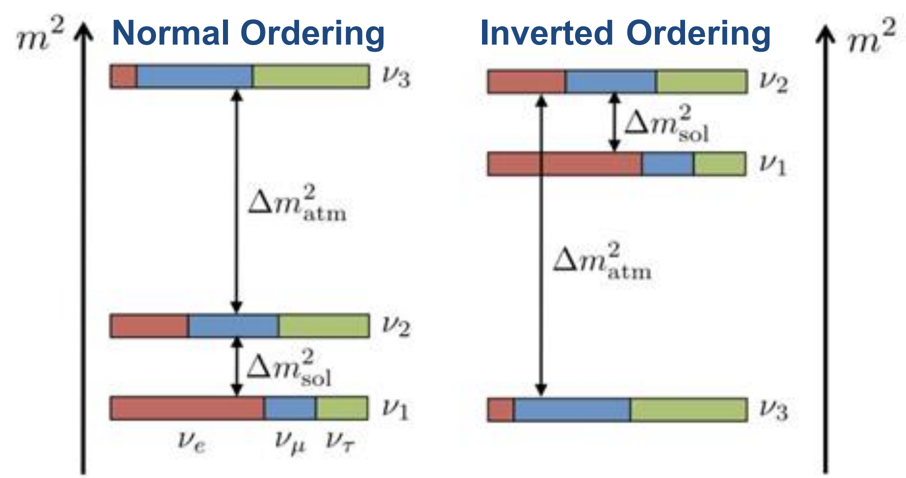
➡ better understand neutrino nature



Credit: Pandey, INSS 2025

Flavor eigenstates Mass eigenstates

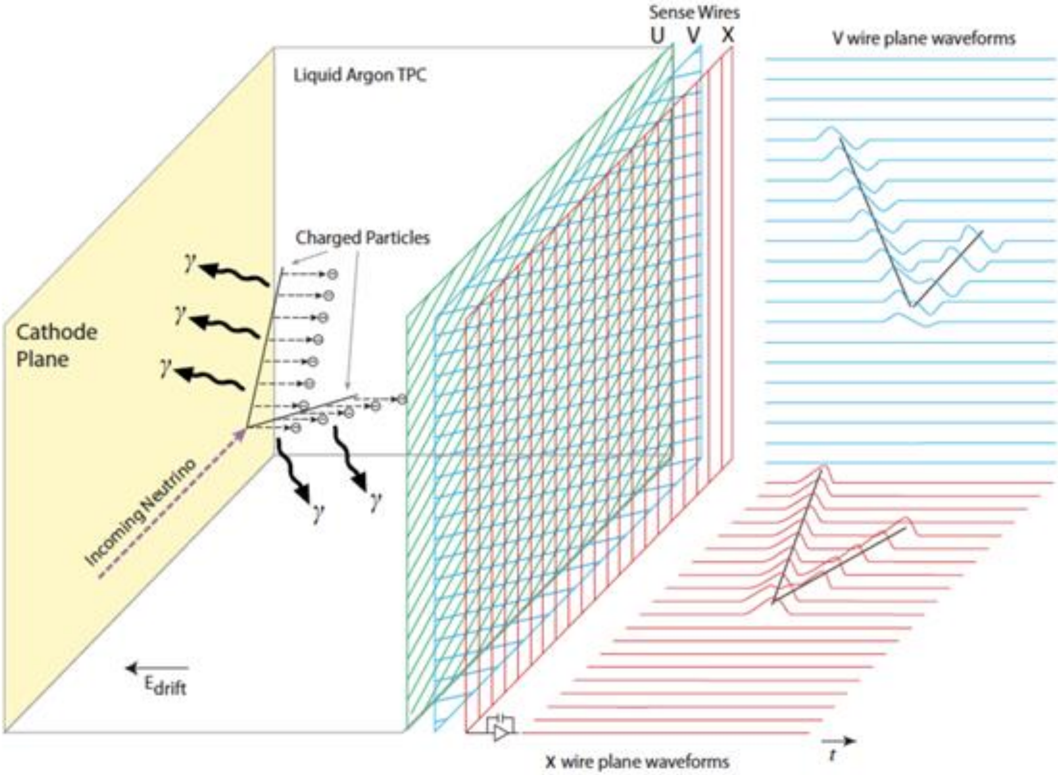
$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \underbrace{\begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix}}_{U_{PMNS}} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$



Credit: JUNO Collaboration

LArTPC operating principle

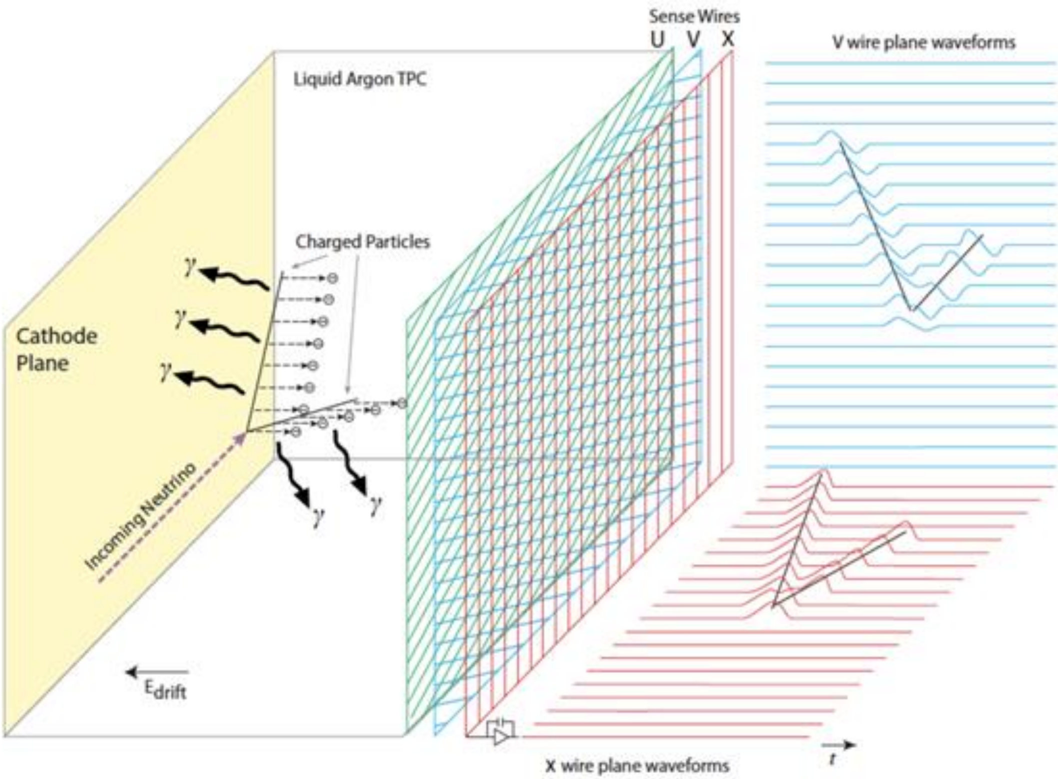
High resolution imaging detector (≈ 1mm)



Source: Abi et al. 2020, JINST 15 T08008

LArTPC operating principle

High resolution imaging detector (≈ 1mm)

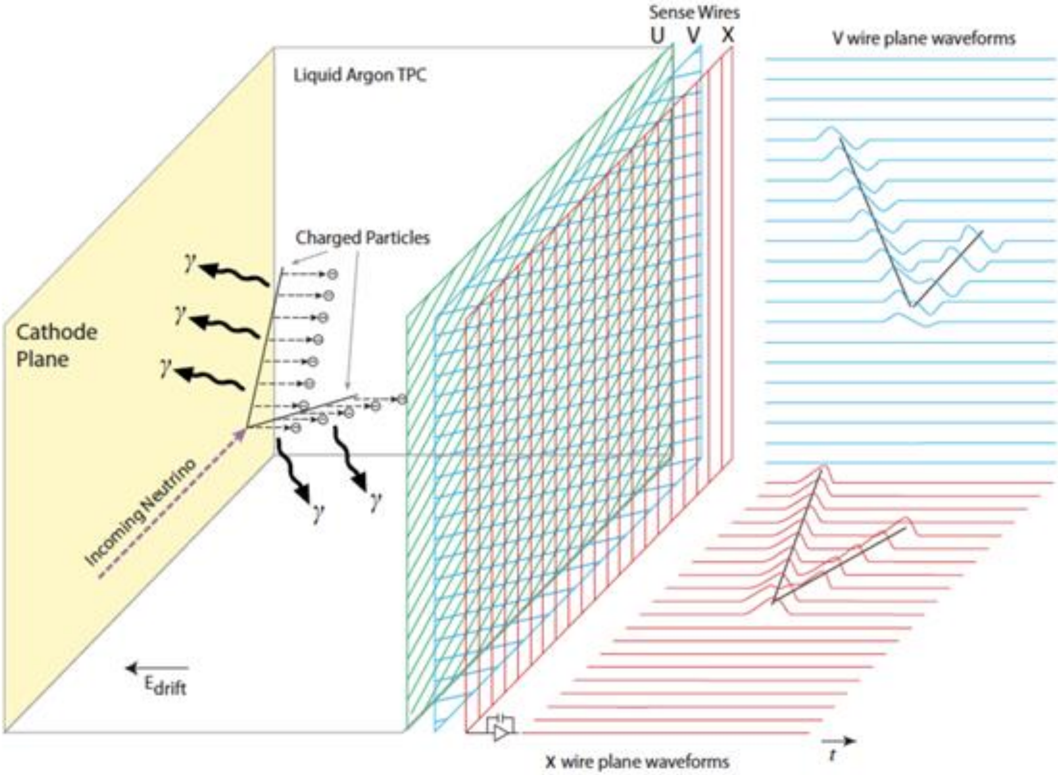


Source: Abi et al. 2020, JINST 15 T08008

- Dense, homogeneous volume of LAr

LArTPC operating principle

High resolution imaging detector (≈ 1mm)

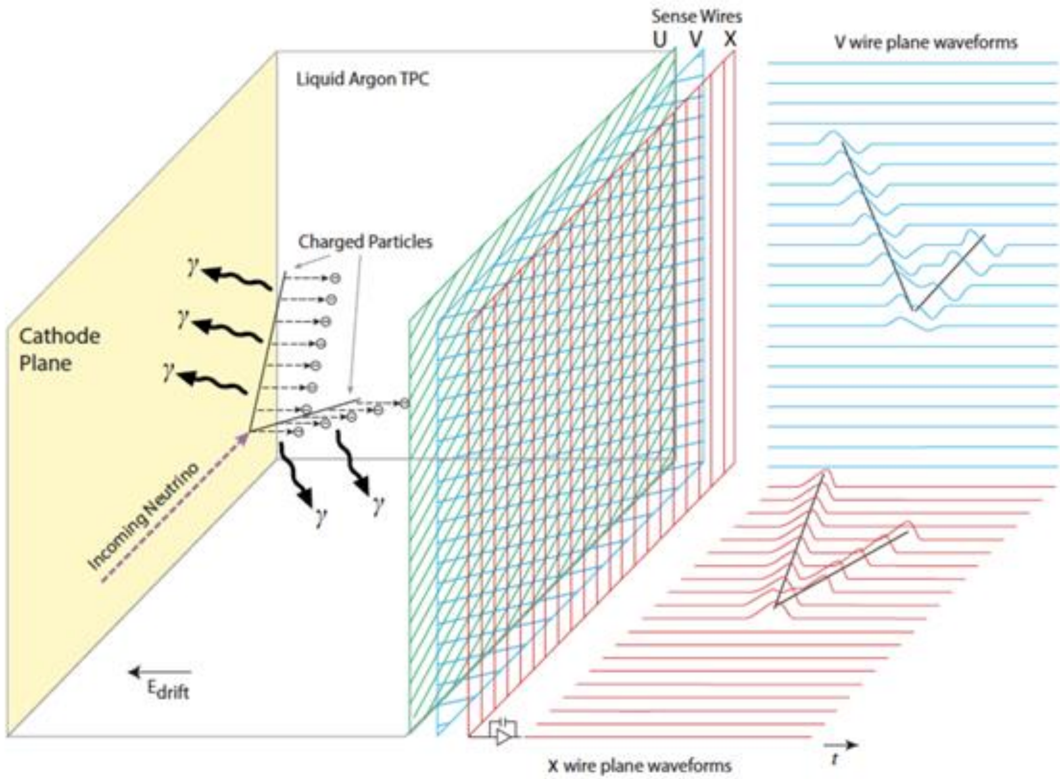


Source: Abi et al. 2020, JINST 15 T08008

- Dense, homogeneous volume of LAr
- Uniform electric field applied in the active region

LArTPC operating principle

High resolution imaging detector (≈ 1mm)

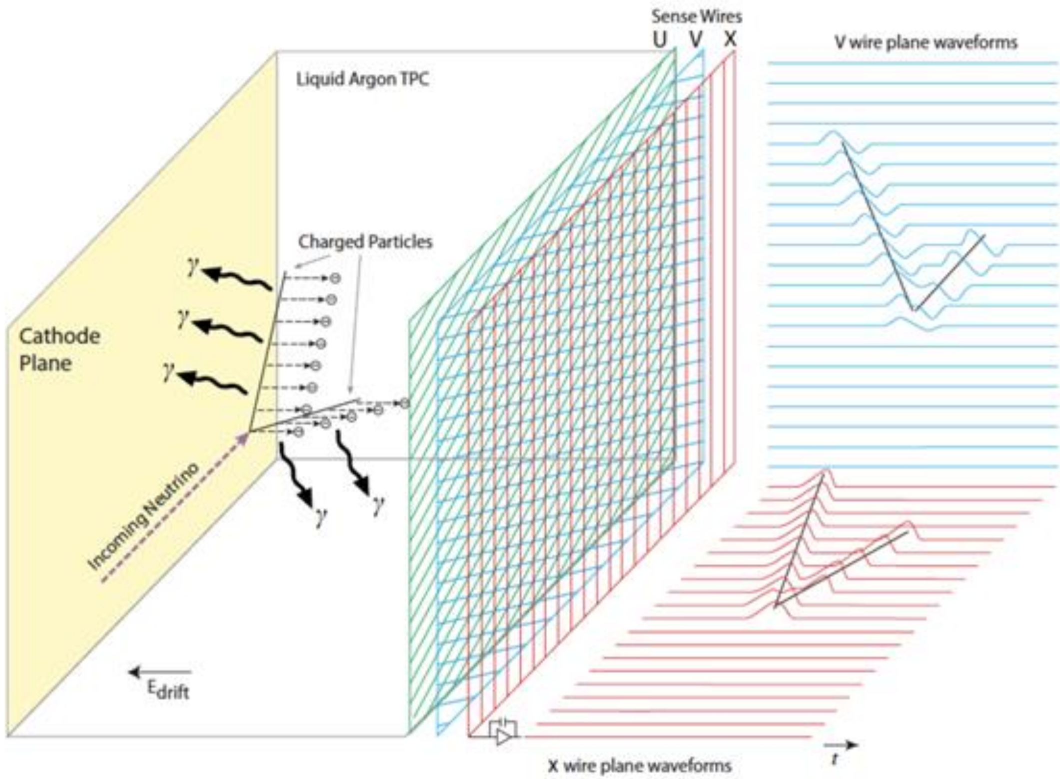


Source: [Abi et al. 2020, JINST 15 T08008]()

- Dense, homogeneous volume of LAr
- Uniform electric field applied in the active region
- Ionization electrons are drifted toward the set of anode wire planes

LArTPC operating principle

High resolution imaging detector (≈ 1mm)

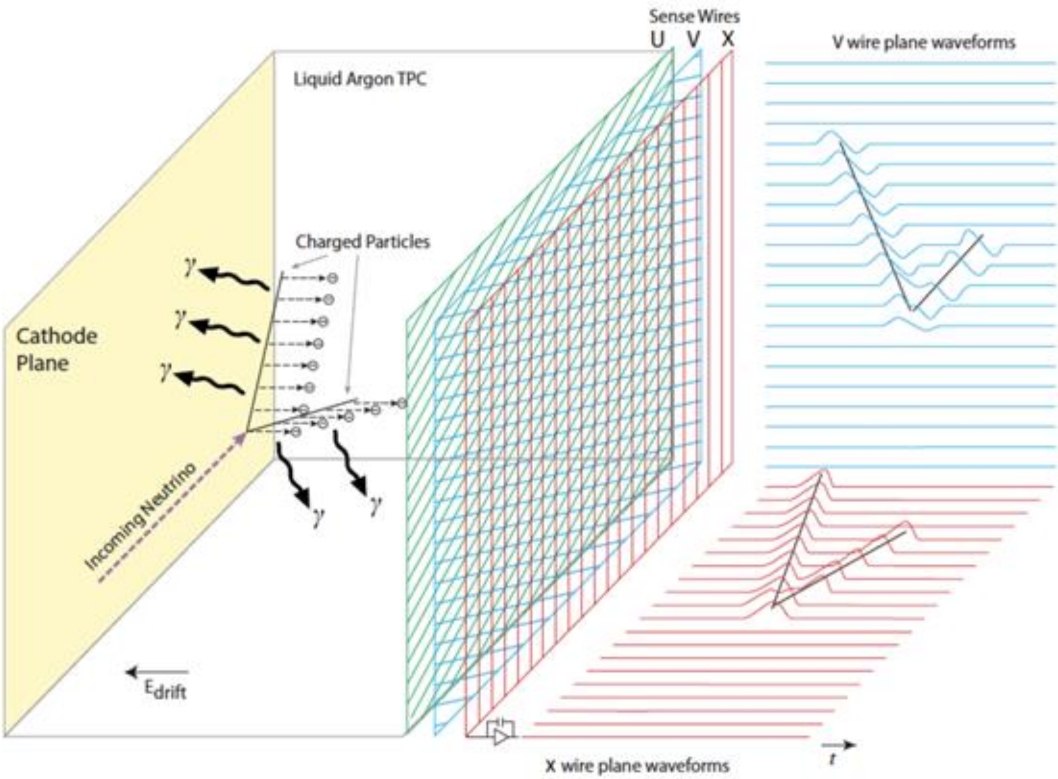


Source: Abi et al. 2020, JINST 15 T08008

- Dense, homogeneous volume of LAr
- Uniform electric field applied in the active region
- Ionization electrons are drifted toward the set of anode wire planes
- Three readout planes (induction planes – U and V - and collection plane – X)

LArTPC operating principle

High resolution imaging detector (≈ 1mm)

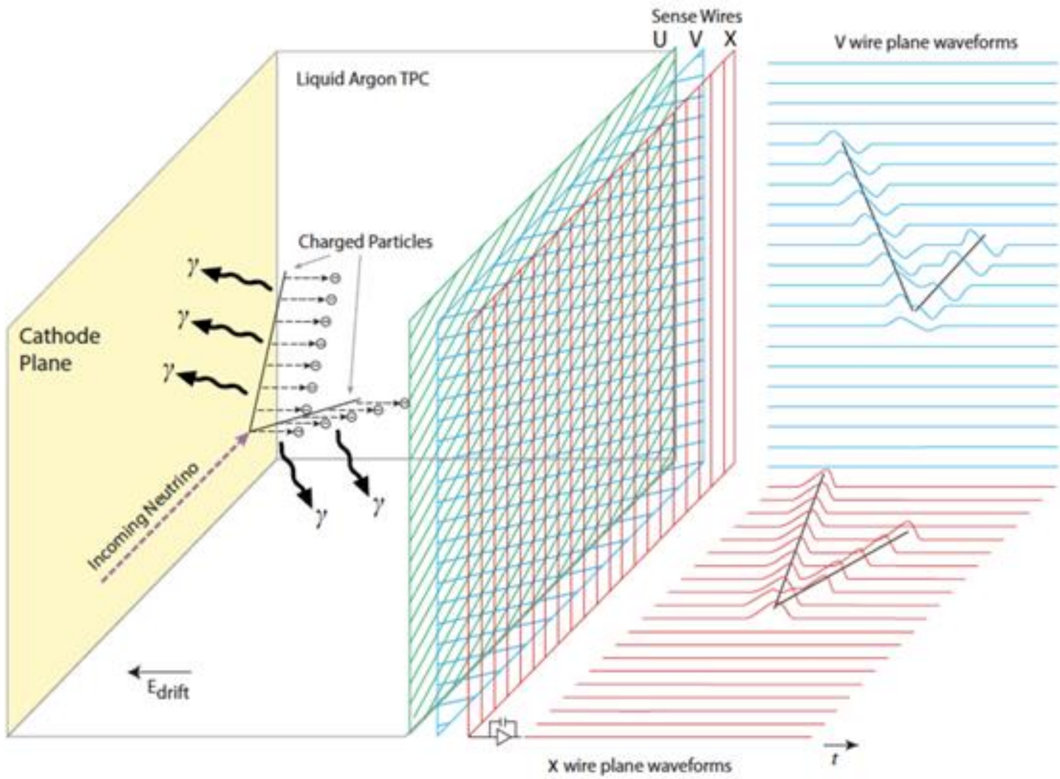


Source: Abi et al. 2020, JINST 15 T08008

- Dense, homogeneous volume of LAr
- Uniform electric field applied in the active region
- Ionization electrons are drifted toward the set of anode wire planes
- Three readout planes (induction planes – U and V - and collection plane – X)
- Multiple 2D views ➔ 3D image of the event

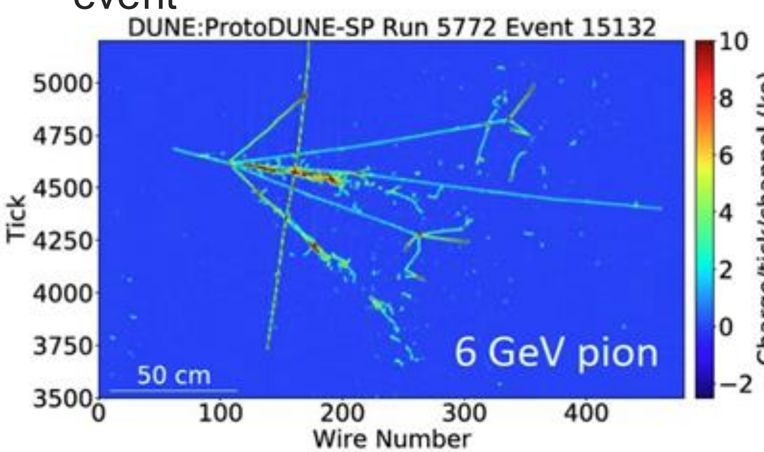
LArTPC operating principle

High resolution imaging detector (≈ 1mm)



Source: Abi et al. 2020, JINST 15 T08008

- Dense, homogeneous volume of LAr
- Uniform electric field applied in the active region
- Ionization electrons are drifted toward the set of anode wire planes
- Three readout planes (induction planes – U and V - and collection plane – X)
- Multiple 2D views ➔ 3D image of the event



Source: Abi et al. 2020, JINST 15 P12004

ProtoDUNE-SP

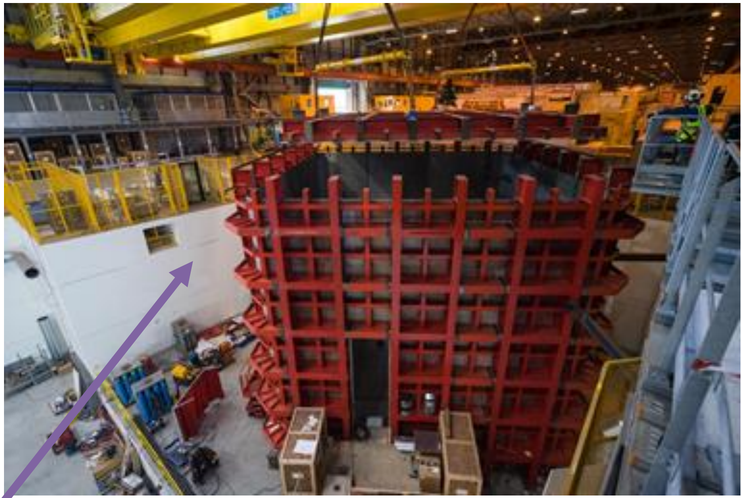
A prototype for the DUNE far detector

- Constructed and operated (2018-2020) on the CERN Neutrino Platform
- Largest LArTPC constructed to date
- Exposed to a charged particle test beam

Credit: Fermilab

ProtoDUNE-SP
A prototype for the DUNE far detector

- Constructed and operated (2018-2020) on the CERN Neutrino Platform
- Largest LArTPC constructed to date
- Exposed to a charged particle test beam



Credit: CERN

Credit: Fermilab

OUTLINE

- Physics & Detectors
- **Pandora & Event Slicing**
- Learning Pairwise Associations
- Preliminary Results
- Summary

Pandora Software Development Kit

A multi-algorithm approach

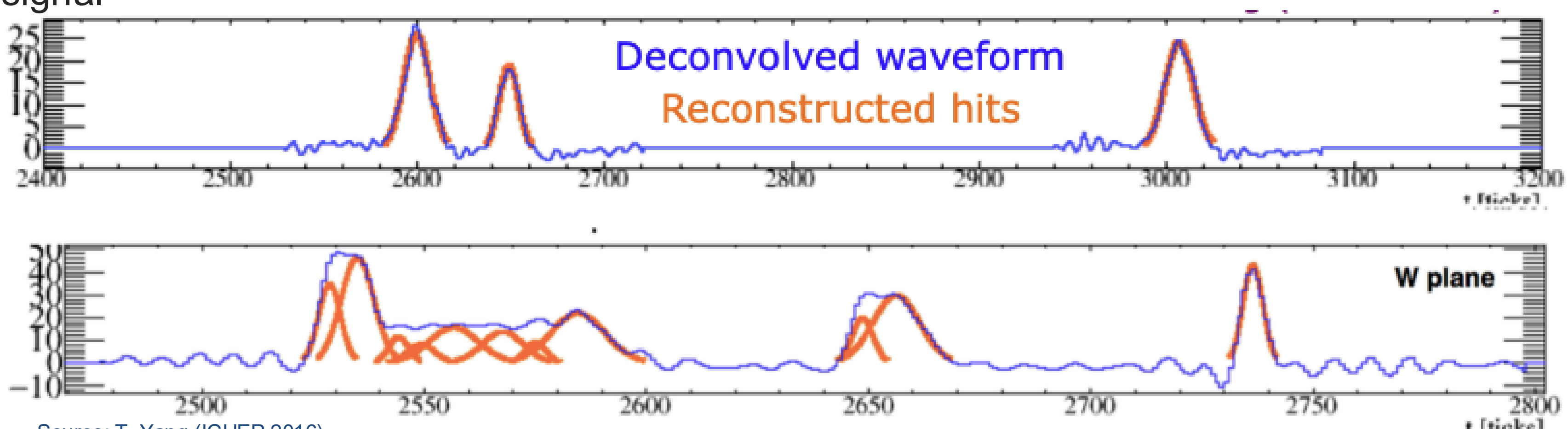
over 100 decoupled algorithms

Pandora Software Development Kit
A multi-algorithm approach

over 100 decoupled algorithms

input hits

Gaussian fitted peaks in the cleaned-up wire signal

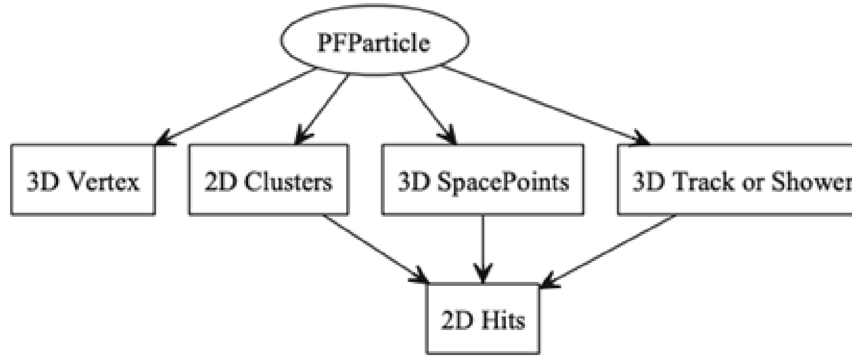


Source: [T. Yang (ICHEP 2016)]

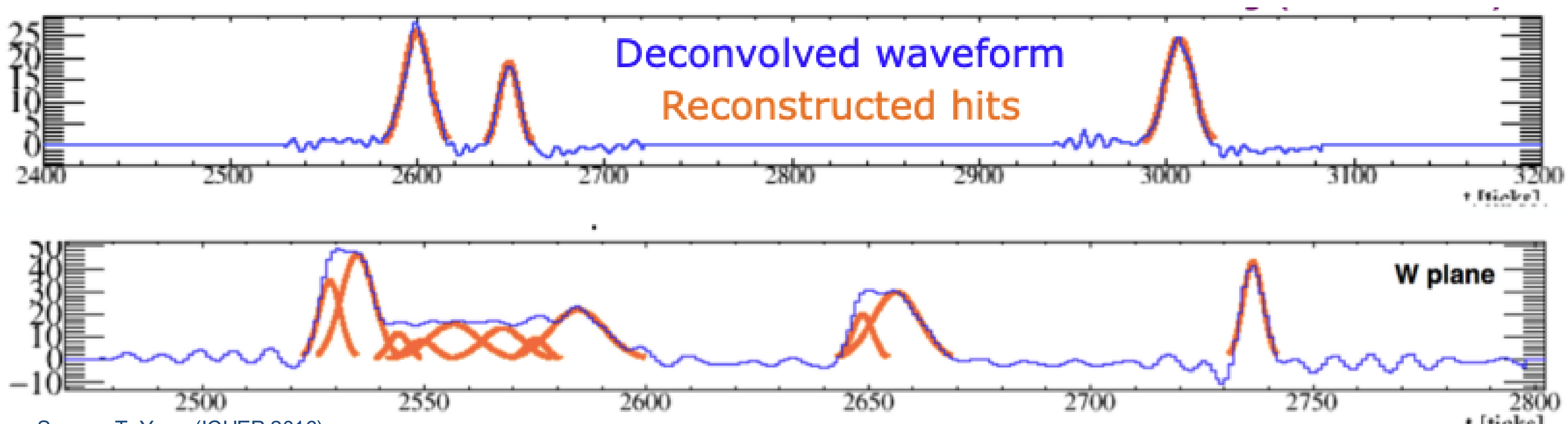
Pandora Software Development Kit
A multi-algorithm approach

over 100 decoupled algorithms

input hits → hierarchy of reco 3D particles



Source: [Acciarri et al. 2018, Eur. Phys. J. C 78, 82]

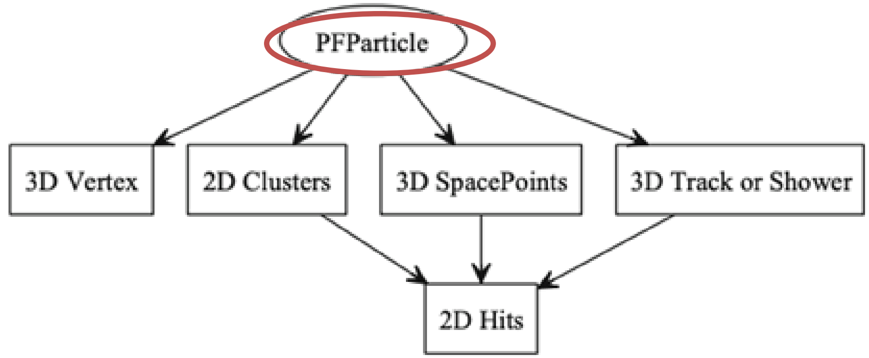


Source: [T. Yang (ICHEP 2016)]

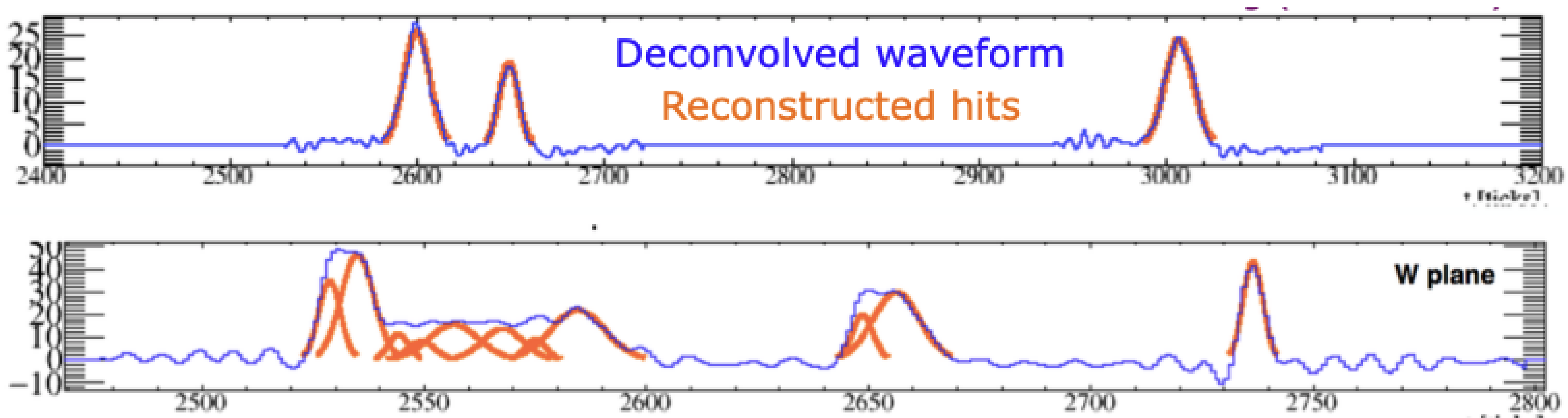
Pandora Software Development Kit
A multi-algorithm approach

over 100 decoupled algorithms

input hits → hierarchy of reco 3D particles



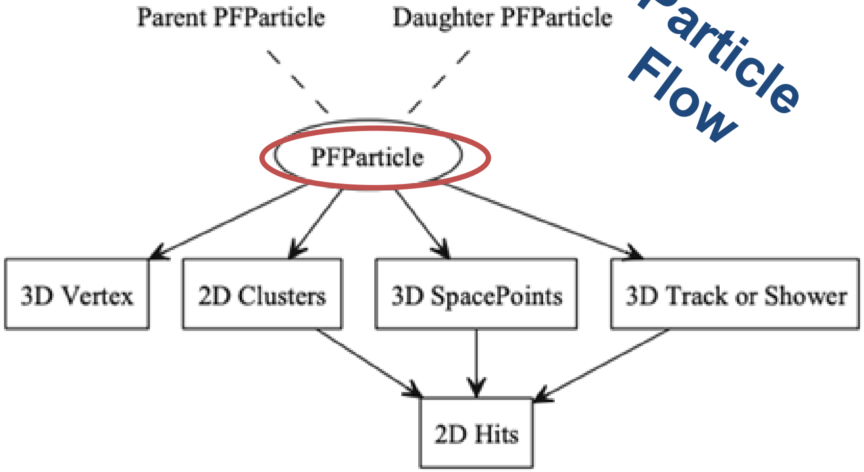
Source: Acciarri et al. 2018, Eur. Phys. J. C 78, 82



Source: T. Yang (ICHEP 2016)

Pandora Software Development Kit

A multi-algorithm approach

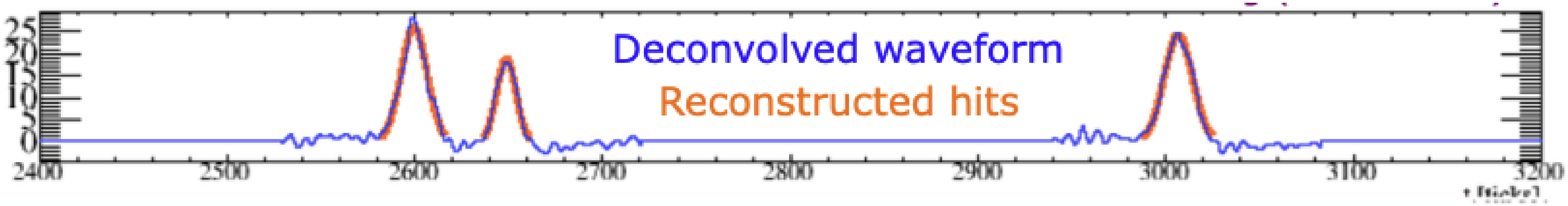


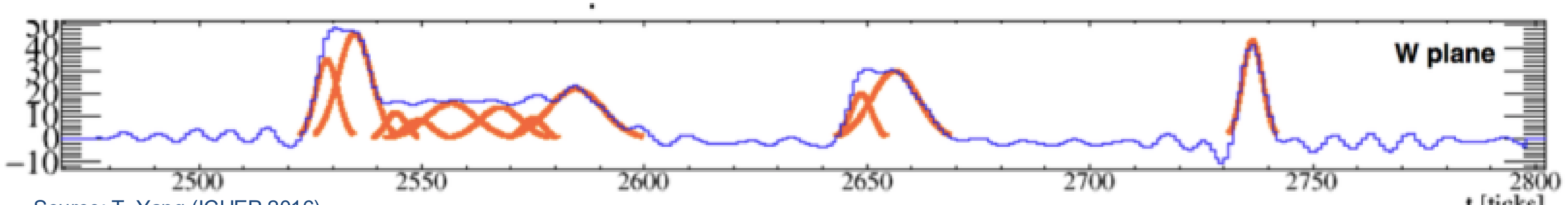
Particle Flow

over 100 decoupled algorithms

input hits → hierarchy of reco 3D particles

Source: Acciarri et al. 2018, Eur. Phys. J. C 78, 82





Source: T. Yang (ICHEP 2016)

DEEP UNDERGROUND NEUTRINO EXPERIMENT

Università degli Studi di Ferrara

INFN FERRARA

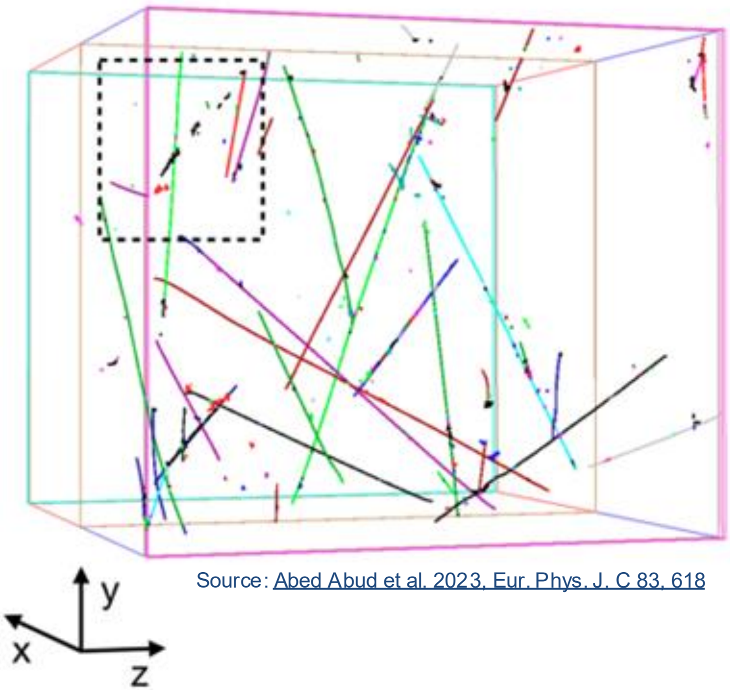
FutureData4EU

Training Future Big Data Experts for Europe

Pandora consolidated reconstruction
A multi-algorithm approach

- Exposed to O(50) cosmic rays per readout (with one beam hierarchy hidden)

ProtoDUNE-SP data run 5387

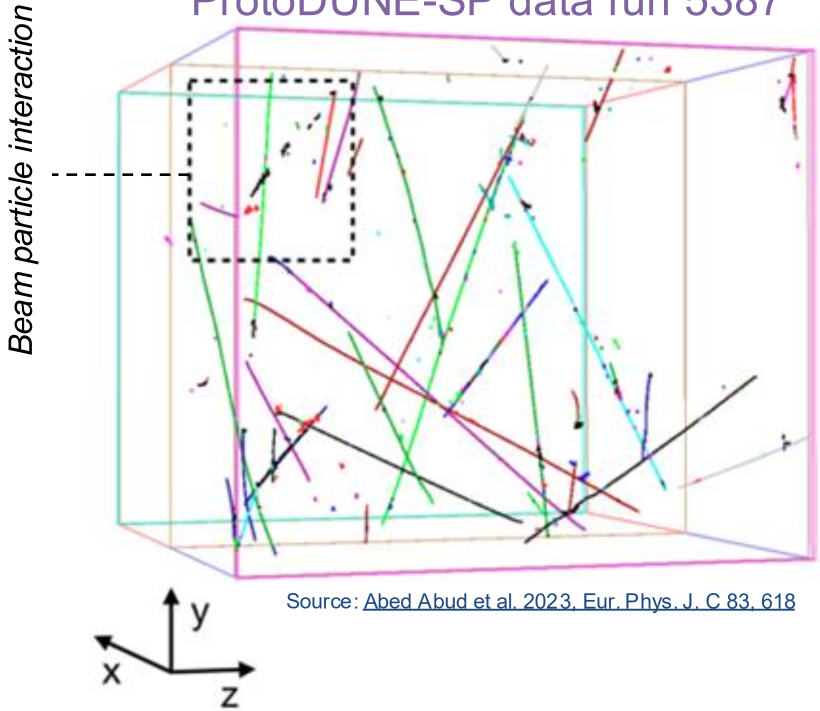


Source: Abed Abud et al. 2023, Eur. Phys. J. C 83, 618

Pandora consolidated reconstruction
A multi-algorithm approach

- Exposed to O(50) cosmic rays per readout (with one beam hierarchy hidden)

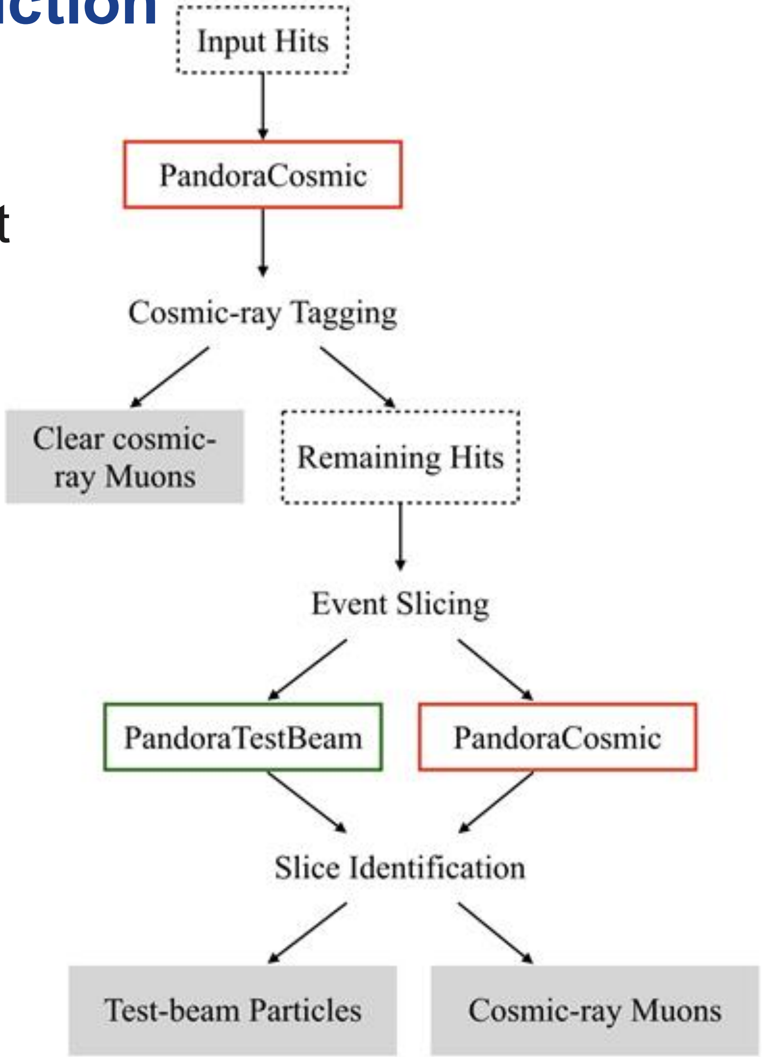
ProtoDUNE-SP data run 5387



Source: Abed Abud et al. 2023, Eur. Phys. J. C 83, 618

Pandora consolidated reconstruction
A multi-algorithm approach

- Exposed to O(50) cosmic rays per readout (with one beam hierarchy hidden)
- Pandora tags and removes obvious cosmic rays



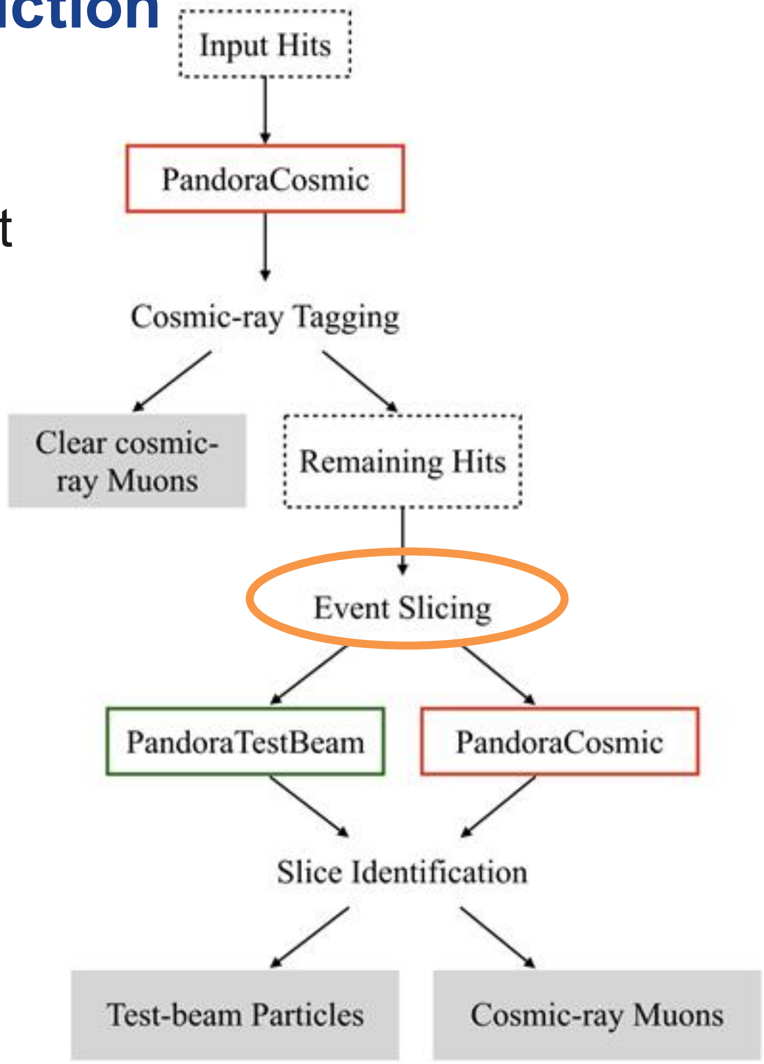
Source: Abed Abud et al. 2023, Eur. Phys. J. C 83, 618

Pandora consolidated reconstruction
A multi-algorithm approach

- Exposed to O(50) cosmic rays per readout (with one beam hierarchy hidden)

- Pandora tags and removes obvious cosmic rays

- Event slicing → hierarchies



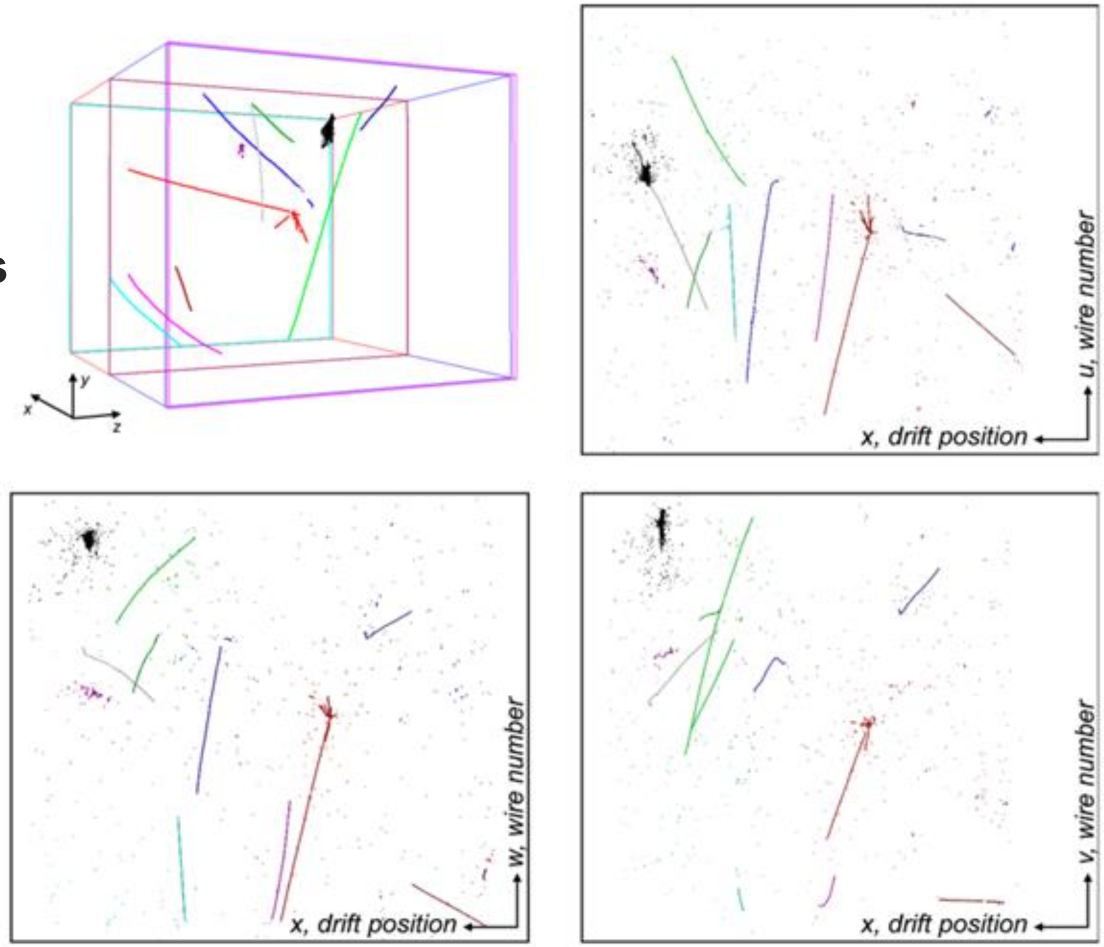
Source: Abed Abud et al. 2023, Eur. Phys. J. C 83, 618

Pandora Event Slicing

Divide up the detector activity into distinct particle hierarchies

3D clusters are divided into slices based on hand-engineered association rules

1. spatial proximity
2. pointing
3. shower cone



Source: Abed Abud et al. 2023, Eur. Phys. J. C 83, 618

OUTLINE

- Physics & Detectors
- Pandora & Event Slicing
- **Learning Pairwise Associations**
- Preliminary Results
- Summary

Learning Pairwise Associations

ProtoDUNE-SP MC data
4 beam energies (0.5, 1, 2, 6 GeV/c)

Energy	# Events	# Pandora slices	# PFPs
0.5 GeV	3,099	269,052	1,607,255
1 GeV	1,500	136,393	760,663
2 GeV	1,500	138,232	781,186
6 GeV	1,500	149,540	915,346
Total	7,599	693,217	4,064,450

Learning Pairwise Associations

ProtoDUNE-SP MC data

4 beam energies (0.5, 1, 2, 6 GeV/c)

Energy	# Events	# Pandora slices	# PFPs
0.5 GeV	3,099	269,052	1,607,255
1 GeV	1,500	136,393	760,663
2 GeV	1,500	138,232	781,186
6 GeV	1,500	149,540	915,346
Total	7,599	693,217	4,064,450

Learning Pairwise Associations

Edge classification problem – Should two reco PFPs belong to the same slice?

ProtoDUNE-SP MC data

4 beam energies (0.5, 1, 2, 6 GeV/c)

Energy	# Events	# Pandora slices	# PFPs
0.5 GeV	3,099	269,052	1,607,255
1 GeV	1,500	136,393	760,663
2 GeV	1,500	138,232	781,186
6 GeV	1,500	149,540	915,346
Total	7,599	693,217	4,064,450

Learning Pairwise Associations

Edge classification problem – Should two reco PFPs belong to the same slice?

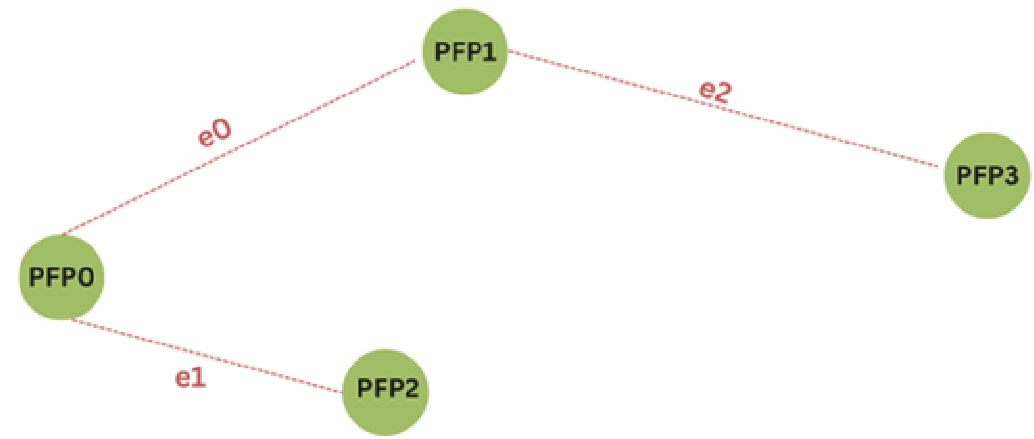
ProtoDUNE-SP MC data

4 beam energies (0.5, 1, 2, 6 GeV/c)

Energy	# Events	# Pandora slices	# PFPs
0.5 GeV	3,099	269,052	1,607,255
1 GeV	1,500	136,393	760,663
2 GeV	1,500	138,232	781,186
6 GeV	1,500	149,540	915,346
Total	7,599	693,217	4,064,450

Each readout is treated as a graph:
- node = reconstructed PFP
- edge = candidate PFP pair
- edge label = same_true_hierachy (0 or 1)

4 PFPs -> 4 nodes, 3 candidate pairs -> 3 edges



Learning Pairwise Associations

Edge classification problem – Should two reco PFPs belong to the same slice?

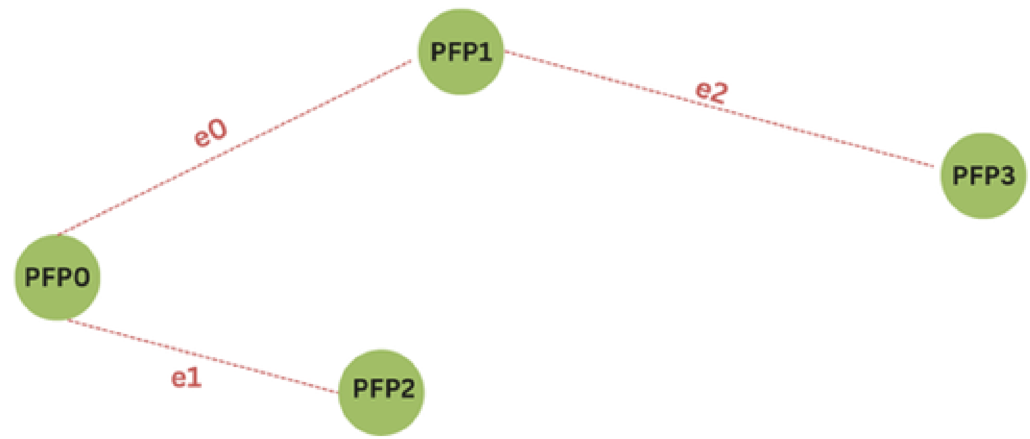
ProtoDUNE-SP MC data

4 beam energies (0.5, 1, 2, 6 GeV/c)

Energy	# Events	# Pandora slices	# PFPs
0.5 GeV	3,099	269,052	1,607,255
1 GeV	1,500	136,393	760,663
2 GeV	1,500	138,232	781,186
6 GeV	1,500	149,540	915,346
Total	7,599	693,217	4,064,450

Each readout is treated as a graph:
- node = reconstructed PFP
- edge = candidate PFP pair
- edge label = same_true_hierachy (0 or 1)

4 PFPs -> 4 nodes, 3 candidate pairs -> 3 edges



Output: the per-edge probability that the two PFPs belong to the same true hierarchy

Learning Pairwise Associations
Node and edge features

N PFPs and M candidate pairs

node_x
[N, 6]
PFP features

edge_attr
[M, 16]
Pair features

y
[M]
Target

edge_index
[2, M]
Connectivity

Learning Pairwise Associations

Node and edge features

N PFPs and M candidate pairs

node_x
[N, 6]
PFP features

edge_attr
[M, 16]
Pair features

y
[M]
Target

edge_index
[2, M]
Connectivity

6 geometric features

- PFP length
- Number of 3D space points
- Is-primary flag
- Is-track-like flag
- Is-shower-like flag
- Number of reconstructed hits

Learning Pairwise Associations

Node and edge features

N PFPs and M candidate pairs

node_x
[N, 6]
PFP features

edge_attr
[M, 16]
Pair features

y
[M]
Target

edge_index
[2, M]
Connectivity

6 geometric features

- PFP length
- Number of 3D space points
- Is-primary flag
- Is-track-like flag
- Is-shower-like flag
- Number of reconstructed hits

16 pairwise features

- Centroid distance (1)
- Direction cosine (1)
- Length, A and B (2)
- Length ratio (1)
- Spacepoint count, A and B (2)
- Spacepoint-count ratio (1)
- Primary flags, A and B (2)
- Track-like flags, A and B (2)
- Shower-like flags, A and B (2)
- Hit counts, A and B (2)

Learning Pairwise Associations
Training details

- 60-20-20% train/val/test event-level splits
- 100 epochs
- trained with class-weighted binary cross-entropy

$$L_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln(\hat{y}_i) + (1 - y_i)\ln(1 - \hat{y}_i)\right]$$

- High-confidence edges (≥0.90) → Slices connected component clustering
- 1× GPU on CINECA Leonardo, 100 GB RAM, 1.7h wall-clock training time

Learning Pairwise Associations
Training details

- 60-20-20% train/val/test event-level splits
- 100 epochs
- trained with class-weighted binary cross-entropy

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)\right]$$

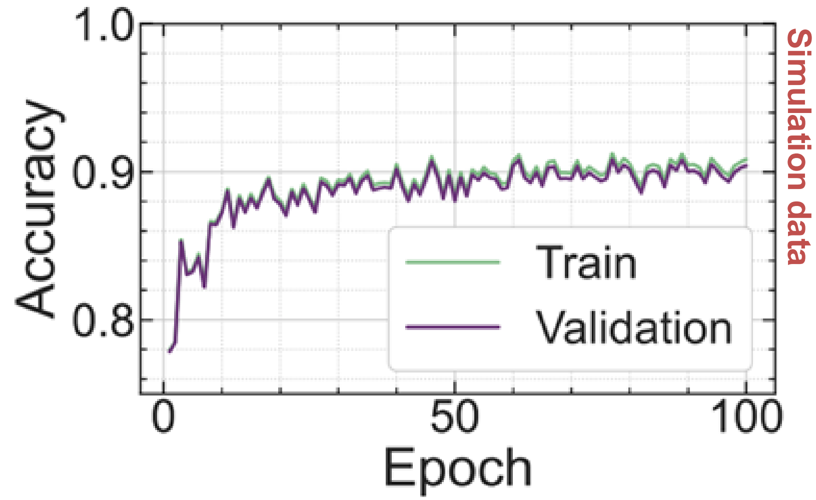
- High-confidence edges (≥0.90) → Slices connected component clustering
- 1× GPU on CINECA Leonardo, 100 GB RAM, 1.7h wall-clock training time

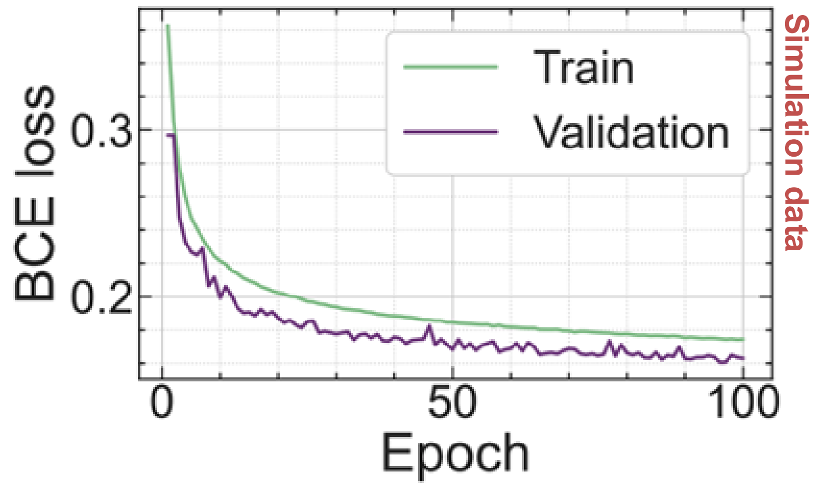
Learning Pairwise Associations
Training details

- 60-20-20% train/val/test event-level splits
- 100 epochs
- trained with class-weighted binary cross-entropy

$$L_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \ln(\hat{y}_i) + (1-y_i)\ln(1-\hat{y}_i)\right]$$

- High-confidence edges (≥0.90) → Slices connected component clustering
- 1× GPU on CINECA Leonardo, 100 GB RAM, 1.7h wall-clock training time





OUTLINE

- Physics & Detectors
- Pandora & Event Slicing
- Learning Pairwise Associations
- **Preliminary Results**
- Summary

Preliminary Results

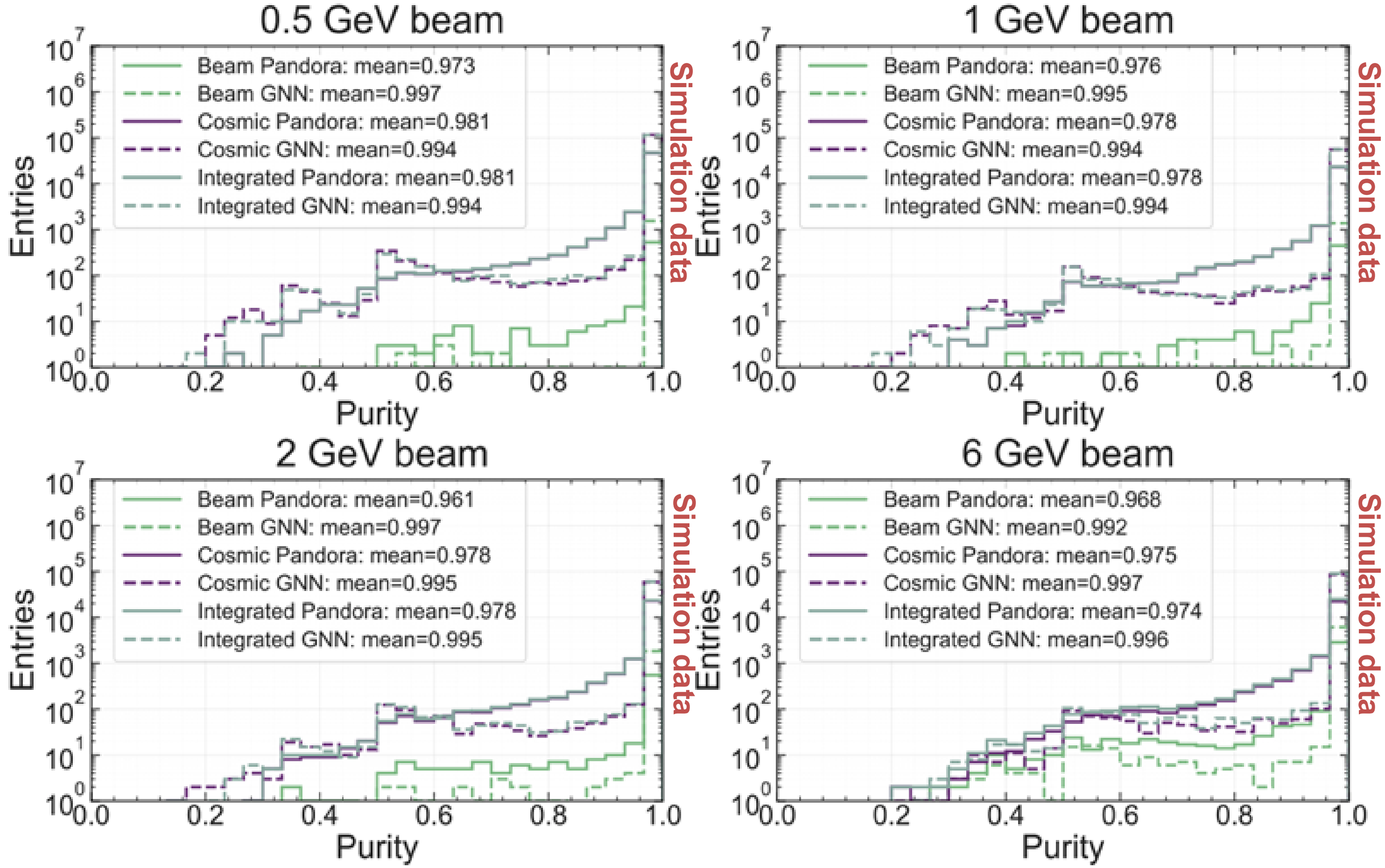
Purity: Does each slice mostly contain hits from one true hierarchy?

Fraction of hits in a reco slice that comes from its dominant true hierarchy

Preliminary Results

Purity: Does each slice mostly contain hits from one true hierarchy?

Fraction of hits in a reco slice that comes from its dominant true hierarchy



Preliminary Results

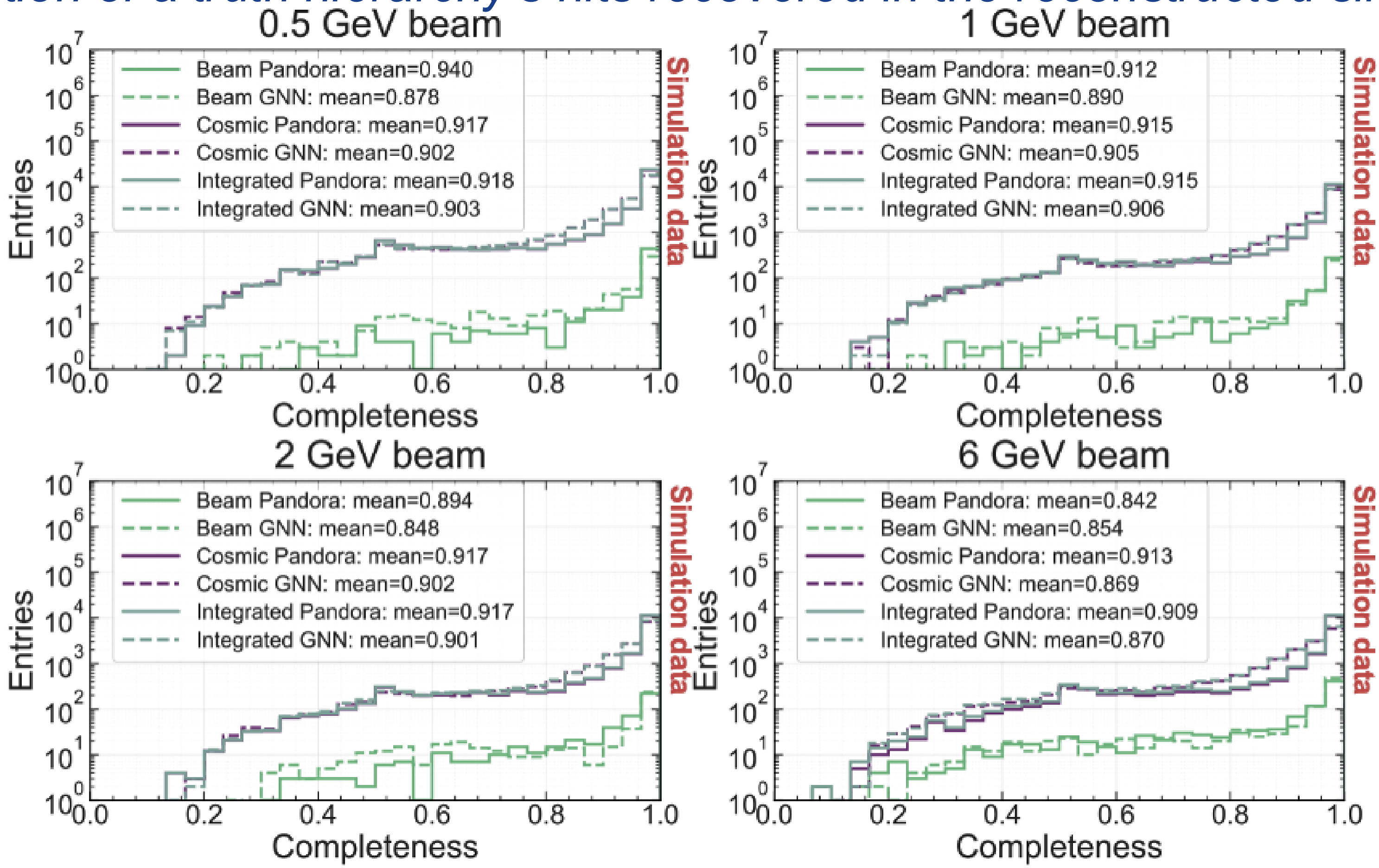
Completeness: Is each true hierarchy kept together in one slice?

Fraction of a truth hierarchy’s hits recovered in the reconstructed slice

Preliminary Results

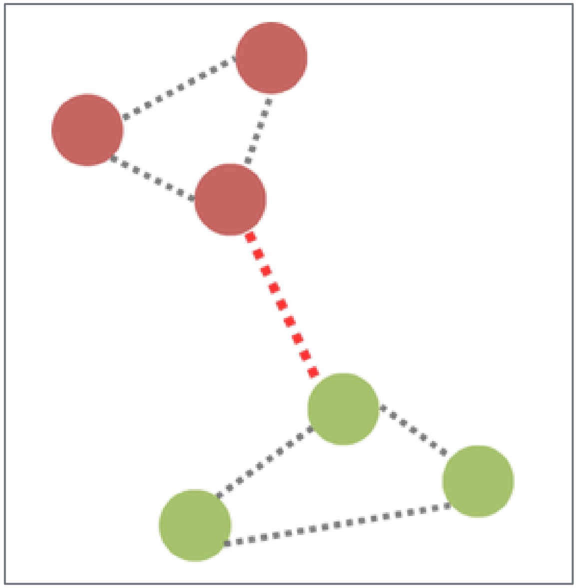
Completeness: Is each true hierarchy kept together in one slice?

Fraction of a truth hierarchy's hits recovered in the reconstructed slice



Preliminary Results
Two failure modes

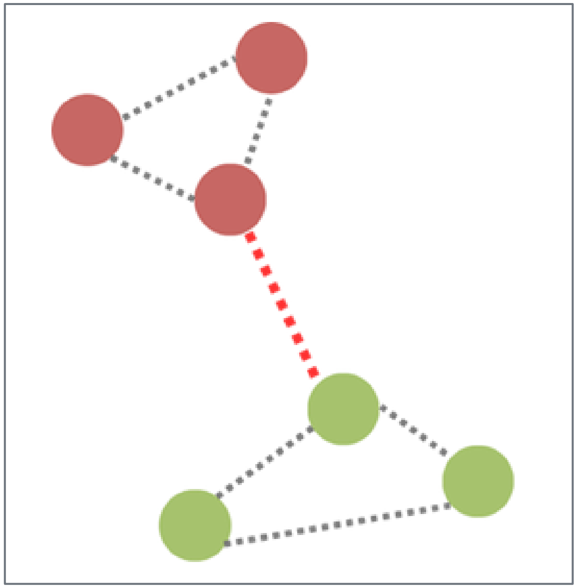
Merging
2 truth hierarchies → 1 predicted slice



false positive rate
two unrelated PFPs are connected

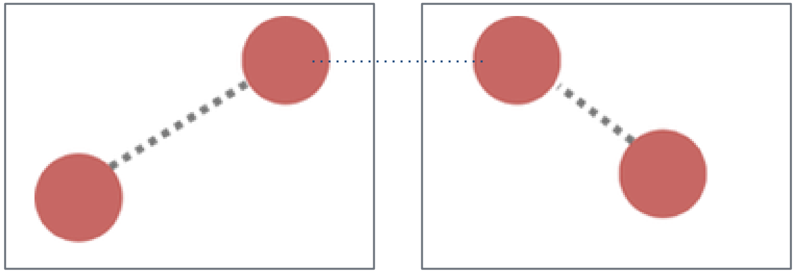
Preliminary Results
Two failure modes

Merging
2 truth hierarchies → 1 predicted slice



false positive rate
two unrelated PFPs are connected

Fragmentation
1 truth hierarchy → 2 predicted slices



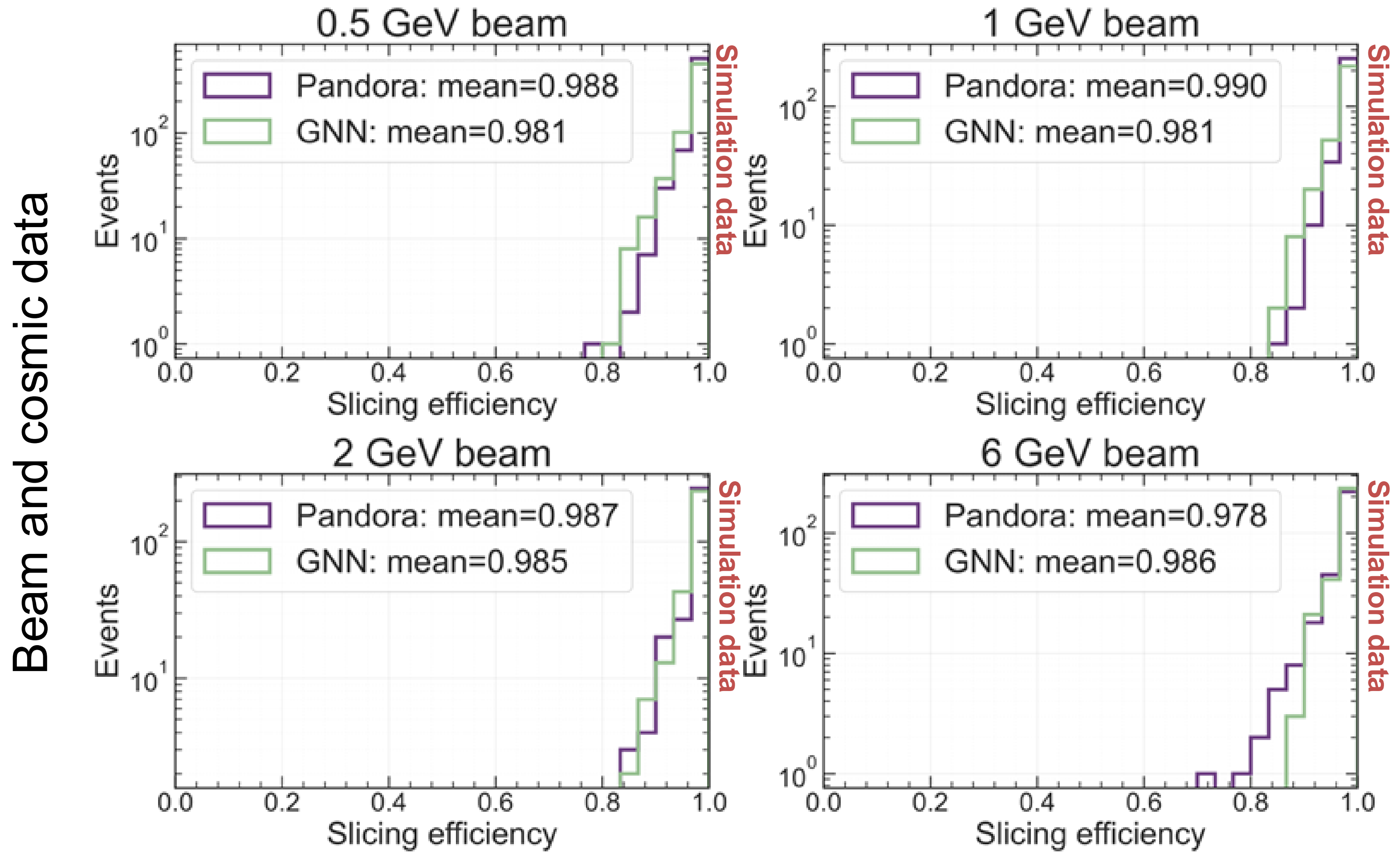
false negative rate
two related PFPs are not connected

Preliminary Results

Efficiency: Fraction of reco truth hierarchies matched to a slice

Preliminary Results

Efficiency: Fraction of reco truth hierarchies matched to a slice

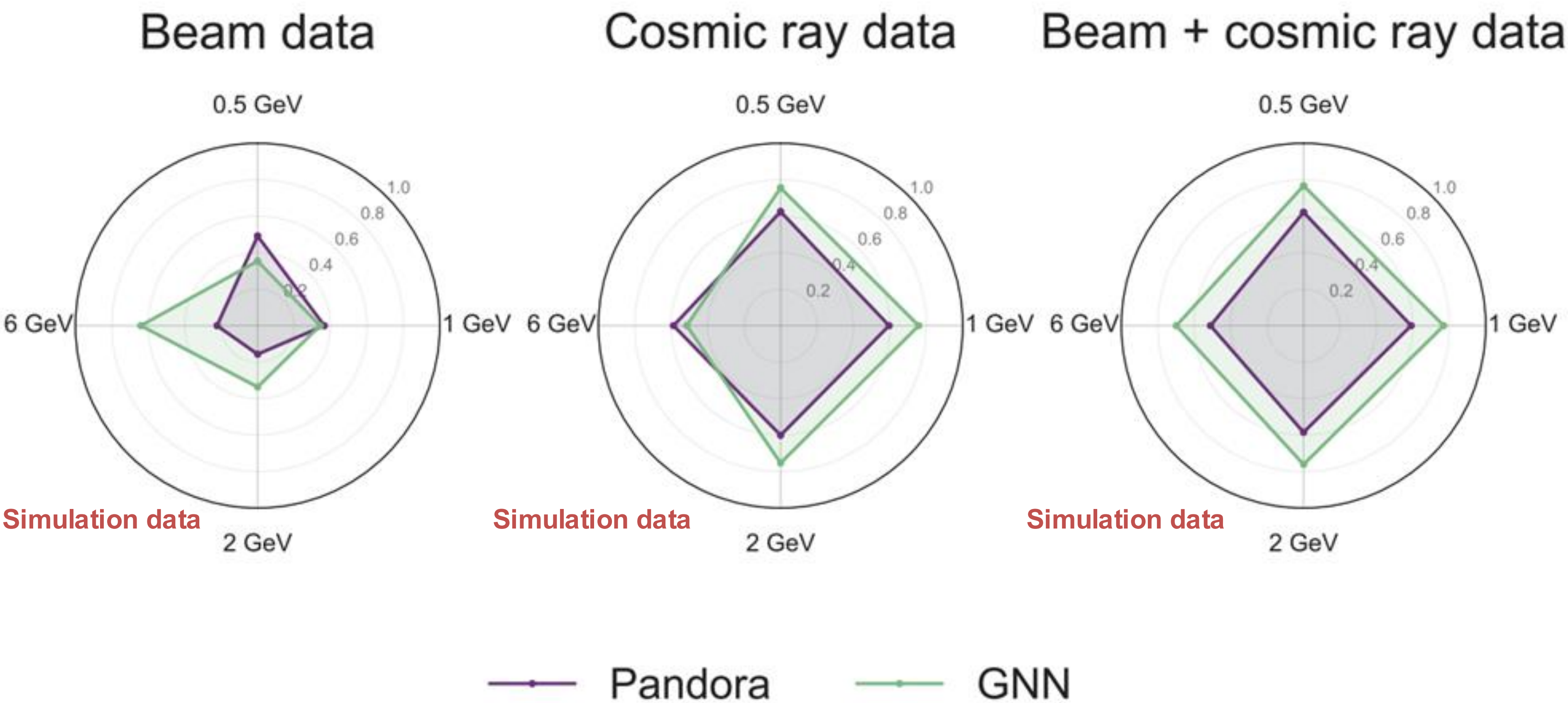


Preliminary Results

Adjusted Rand Index (ARI): Agreement between the predicted slice and the true hierarchy grouping

Preliminary Results

Adjusted Rand Index (ARI): Agreement between the predicted slice and the true hierarchy grouping

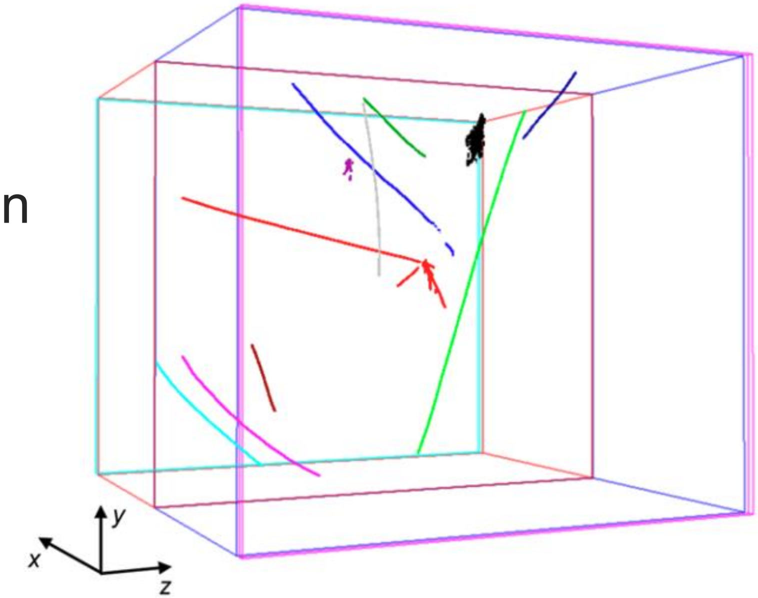


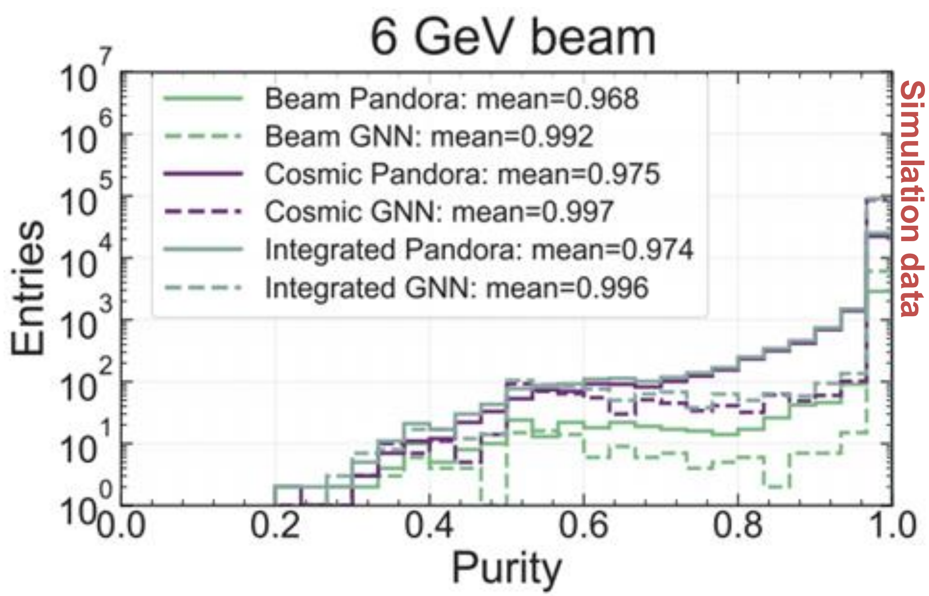
OUTLINE

- Physics & Detectors
- Pandora & Event Slicing
- Learning Pairwise Associations
- Preliminary Results
- **Summary**

Closing remarks

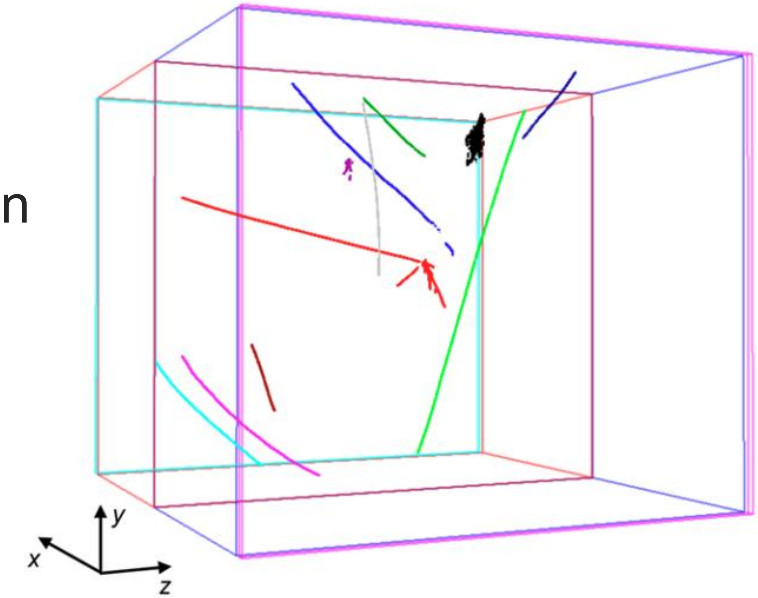
- Event slicing reframed as edge classification
- Pandora favors completeness over purity
- GNN favors purity over completeness
- A proof of principle: learned pairwise association is a viable alternative to hand-engineered slicing rules





Closing remarks

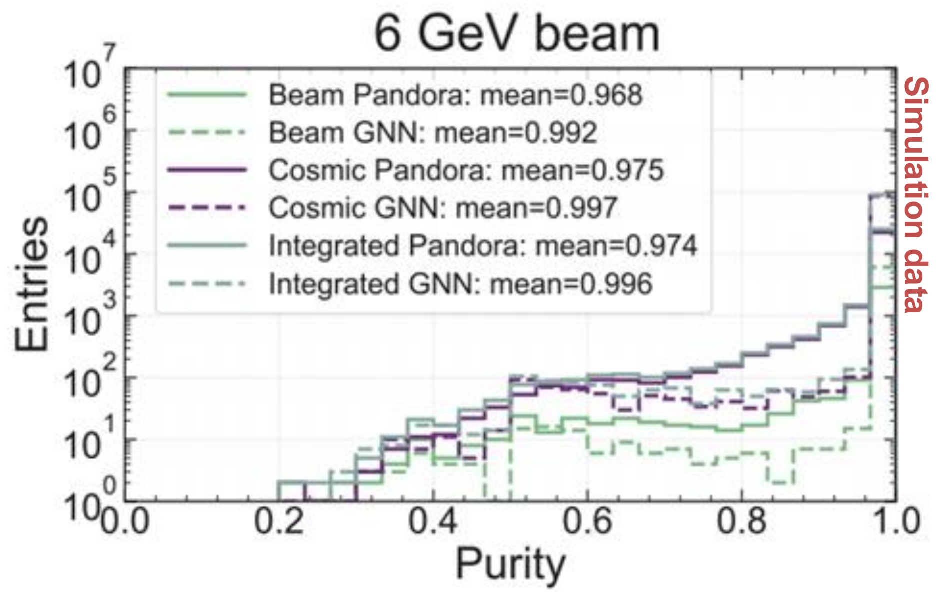
- Event slicing reframed as edge classification
- Pandora favors completeness over purity
- GNN favors purity over completeness
- A proof of principle: learned pairwise association is a viable alternative to hand-engineered slicing rules



Thank You!

… and thanks to Leigh Whitehead[1], Dominic Brailsford[2], and Andy Chappell[3]

[1] University of Cambridge,
[2] Lancaster University,
[3] University of Warwick



FutureData4EU

Training Future Big Data Experts for Europe

Funding

COESIONE ITALIA 21-27 EMILIA-ROMAGNA

Co-funded by the European Union

REPVBBLICA ITALIANA

Regione Emilia-Romagna

Partner

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

UNIVERSITÀ CATTOLICA del Sacro Cuore

Università degli Studi di Ferrara

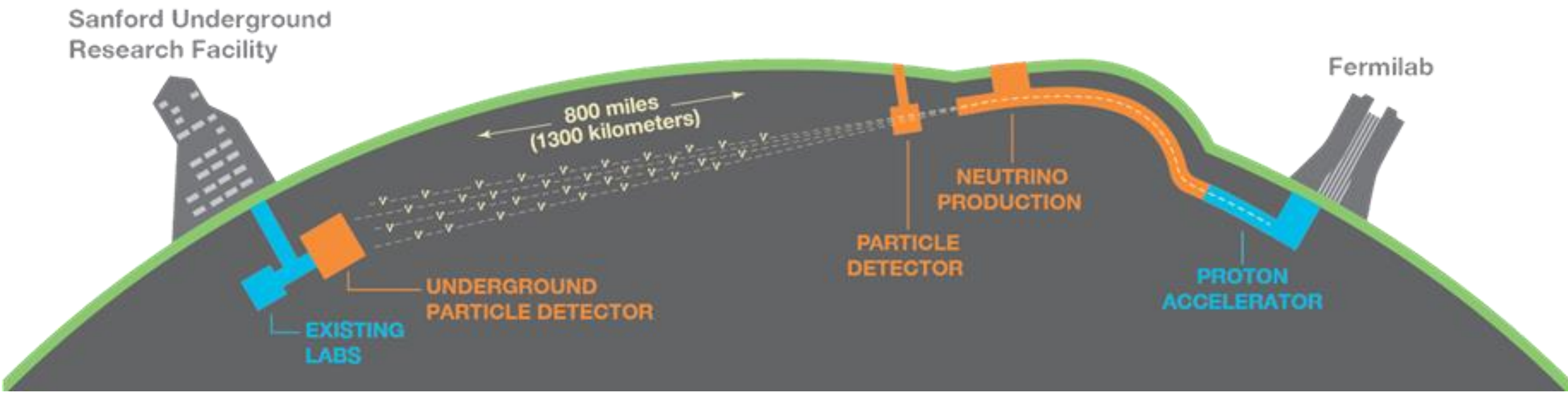
UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

UNIVERSITÀ DI PARMA

POLITECNICO MILANO 1863

Backup slides

The Deep Underground Neutrino Experiment

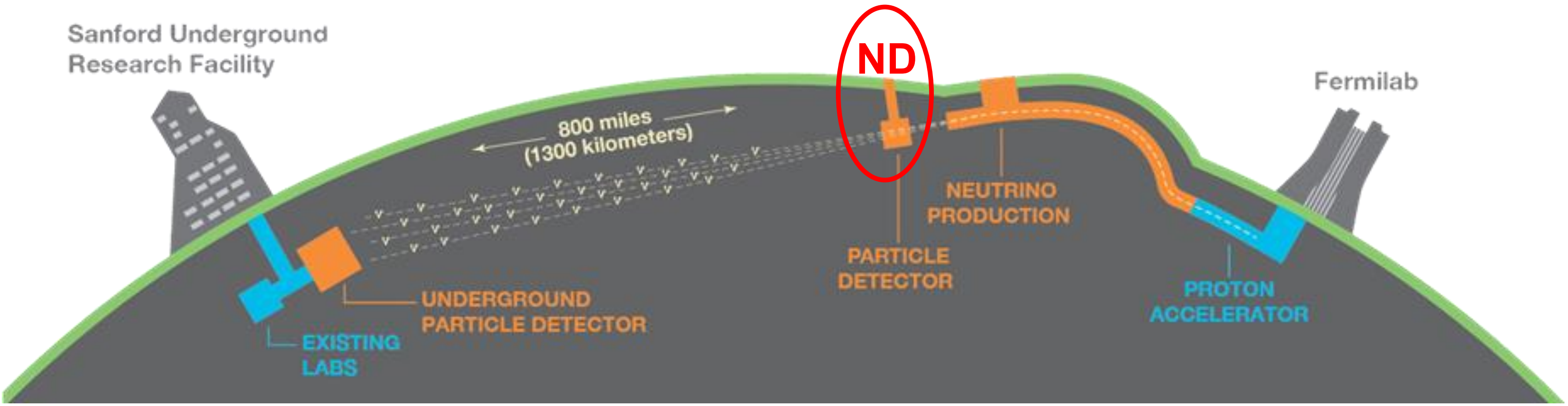


Source: Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE (2020).

<u>Long-baseline neutrino oscillation experiment</u>

- High-power proton beam of 1.2MW upgradable to 2.4MW.
- High-power, wide-range neutrino beam (~GeV energy range)
- Near detector (ND) (575m from the ν source)
- Far detector (FD) in South Dakota (~1300km) and 1.5km deep underground

The Deep Underground Neutrino Experiment

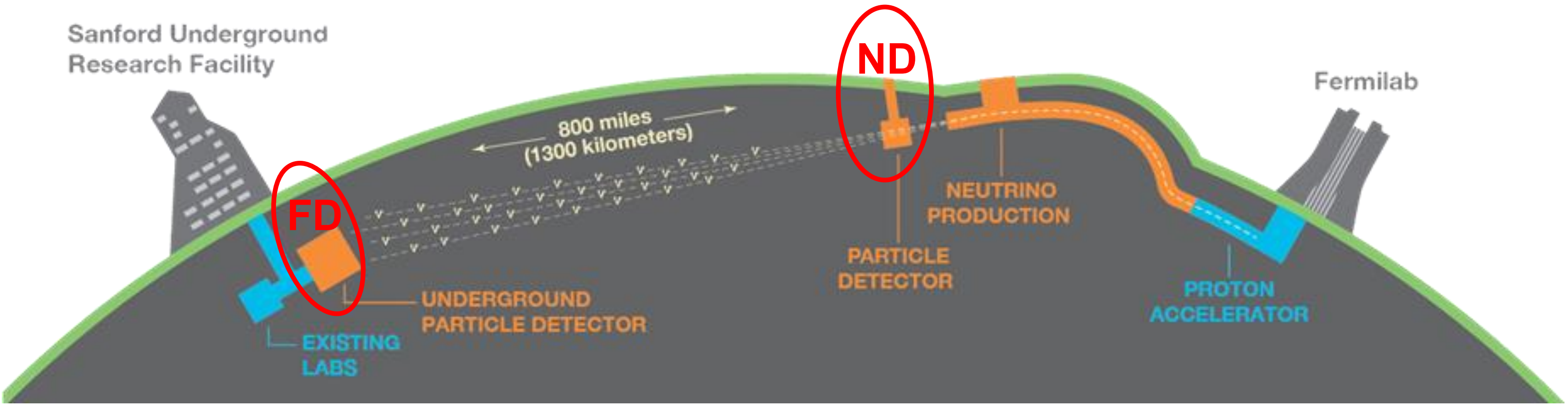


Source: Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE (2020).

<u>Long-baseline neutrino oscillation experiment</u>

- High-power proton beam of 1.2MW upgradable to 2.4MW.
- High-power, wide-range neutrino beam (~GeV energy range)
- Near detector (ND) (575m from the ν source)
- Far detector (FD) in South Dakota (~1300km) and 1.5km deep underground

The Deep Underground Neutrino Experiment



Source: Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume I: Introduction to DUNE (2020).

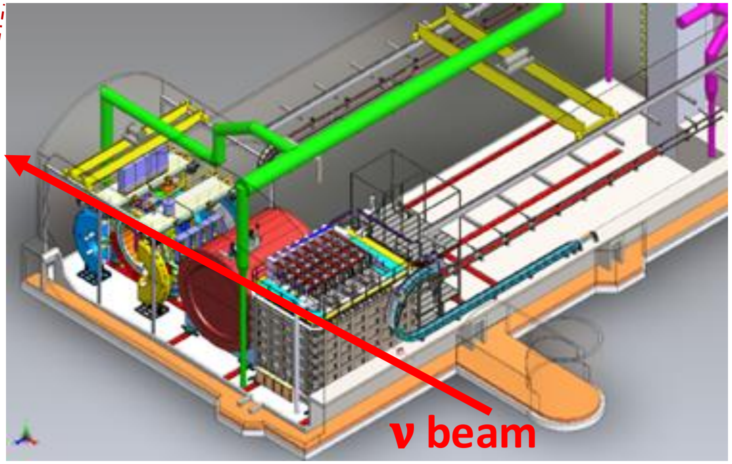
<u>Long-baseline neutrino oscillation experiment</u>

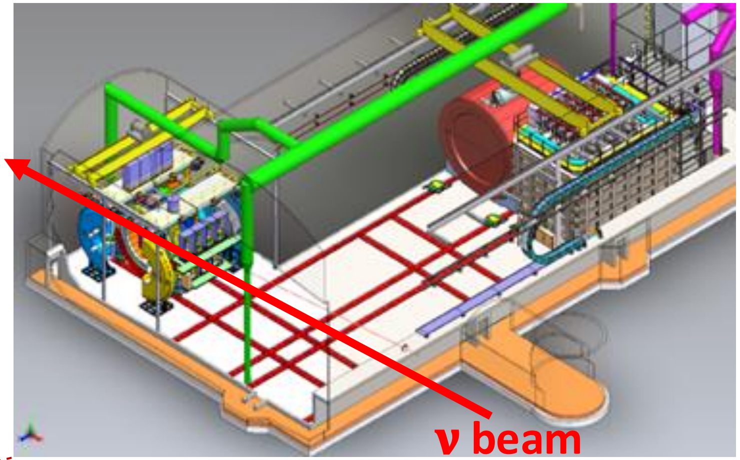
- High-power proton beam of 1.2MW upgradable to 2.4MW.
- High-power, wide-range neutrino beam (~GeV energy range)
- Near detector (ND) (575m from the ν source)
- Far detector (FD) in South Dakota (~1300km) and 1.5km deep underground

DUNE-PRISM ND Complex

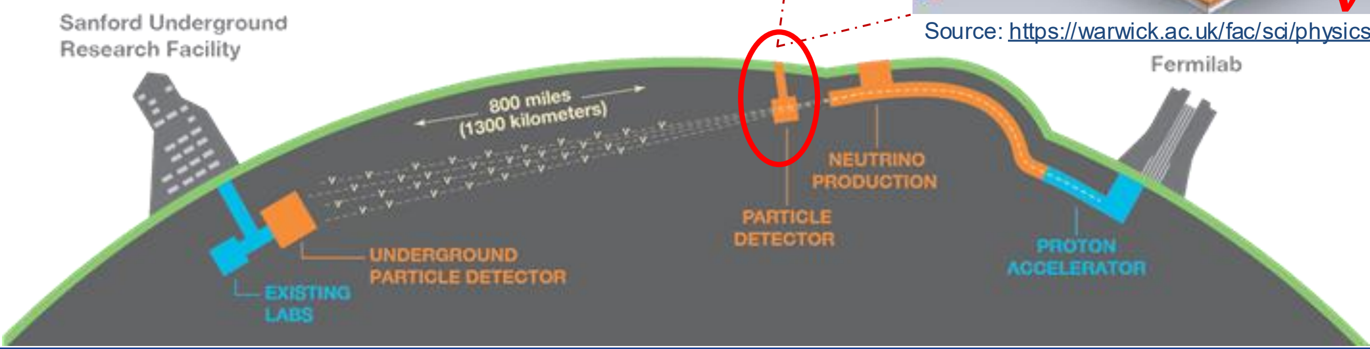
Off-axis beam measurement

- Constrain systematic uncertainties
- Beam monitoring
- Initial neutrino flux measurement before any oscillations



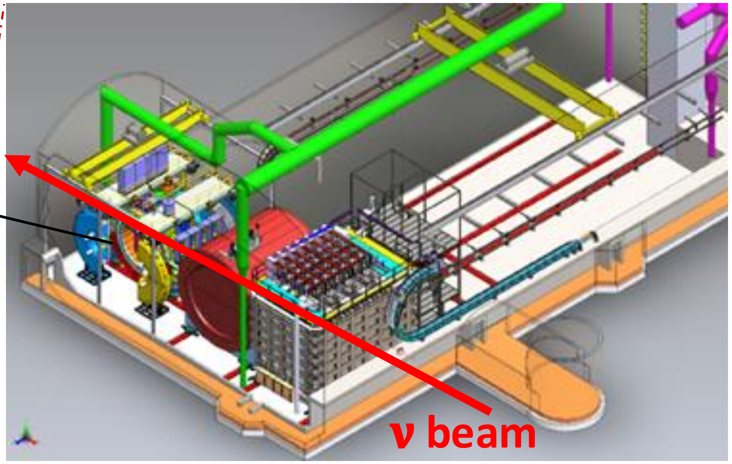


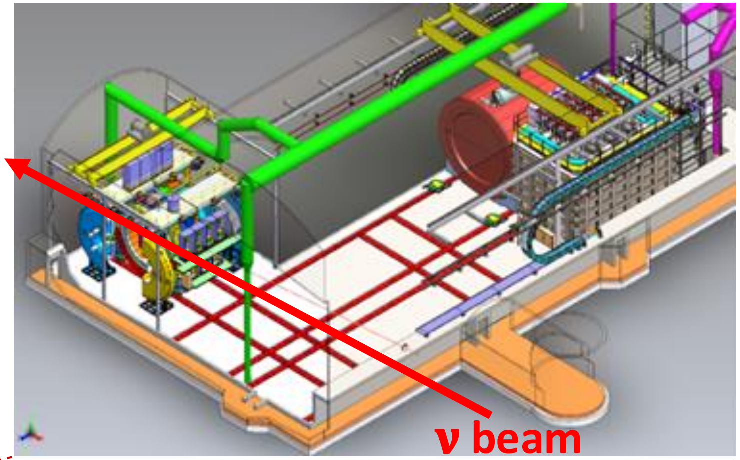
Source: https://warwick.ac.uk/fac/sci/physics/research/epp/exp/dune/



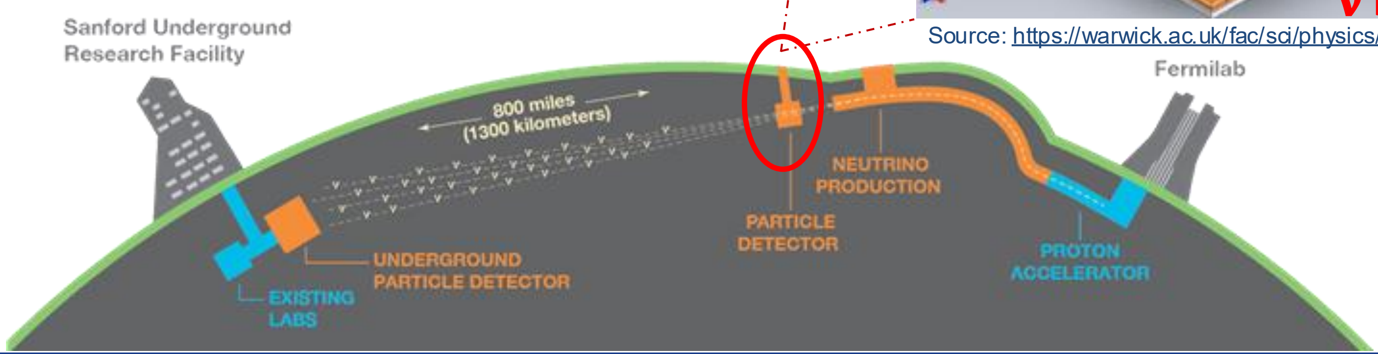
DUNE-PRISM ND Complex
Off-axis beam measurement

- Constrain systematic uncertainties
- Beam monitoring
- Initial neutrino flux measurement before any oscillations





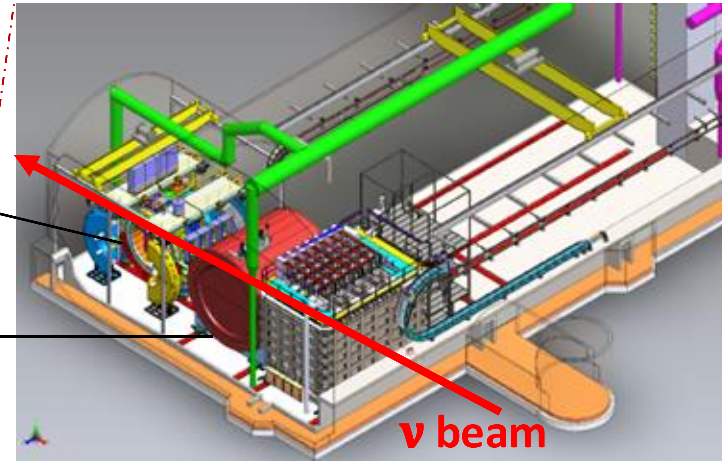
Source: https://warwick.ac.uk/fac/sci/physics/research/epp/exp/dune/

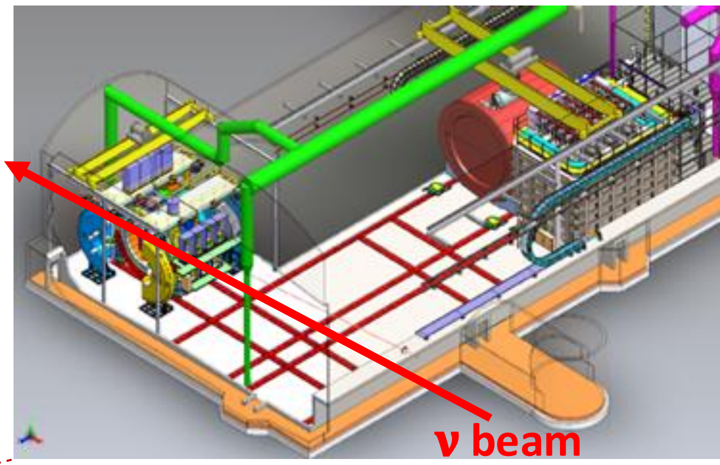


DUNE-PRISM ND Complex
Off-axis beam measurement

SAND

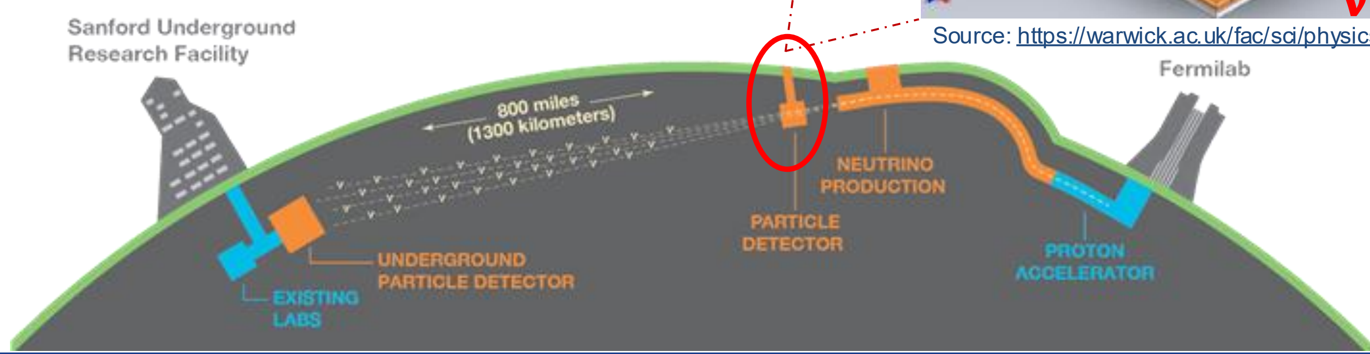
TMS -> ND-GAr





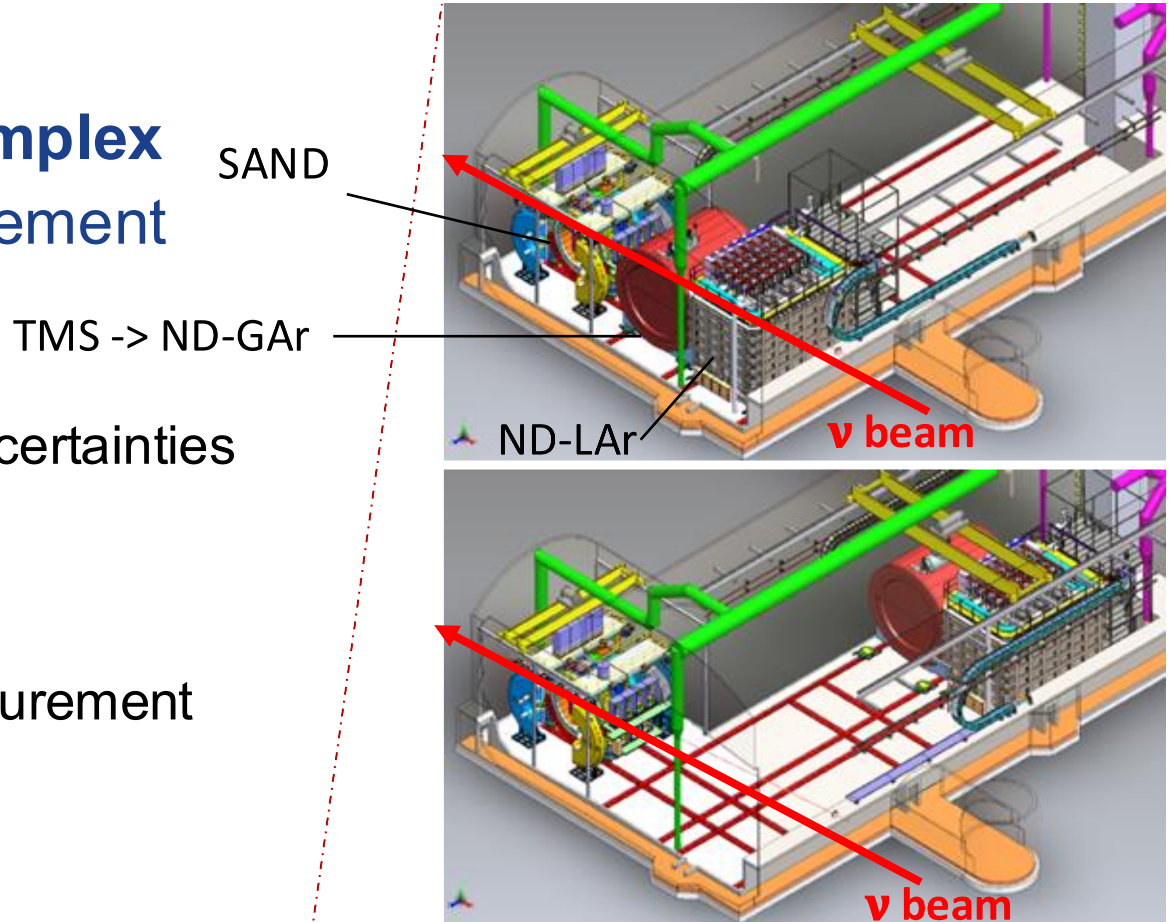
Source: https://warwick.ac.uk/fac/sci/physics/research/epp/exp/dune/

- Constrain systematic uncertainties
- Beam monitoring
- Initial neutrino flux measurement before any oscillations

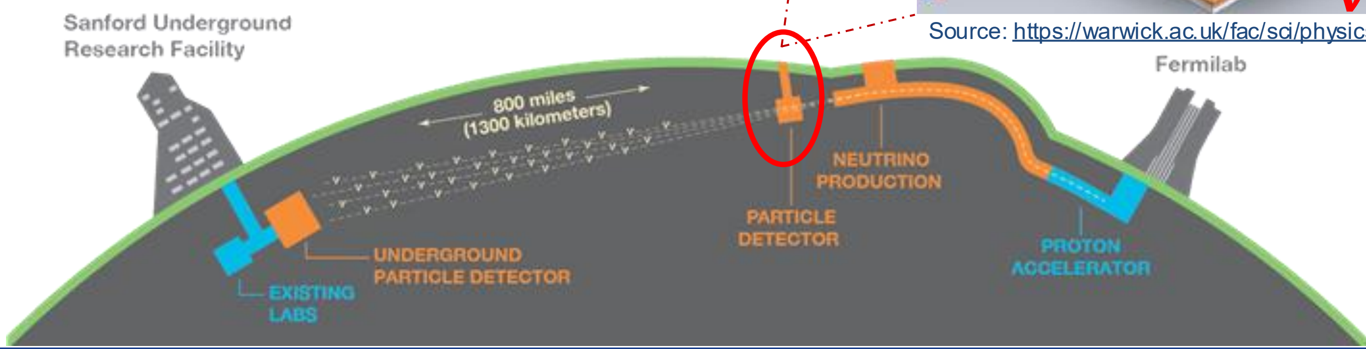


DUNE-PRISM ND Complex
Off-axis beam measurement

- Constrain systematic uncertainties
- Beam monitoring
- Initial neutrino flux measurement before any oscillations

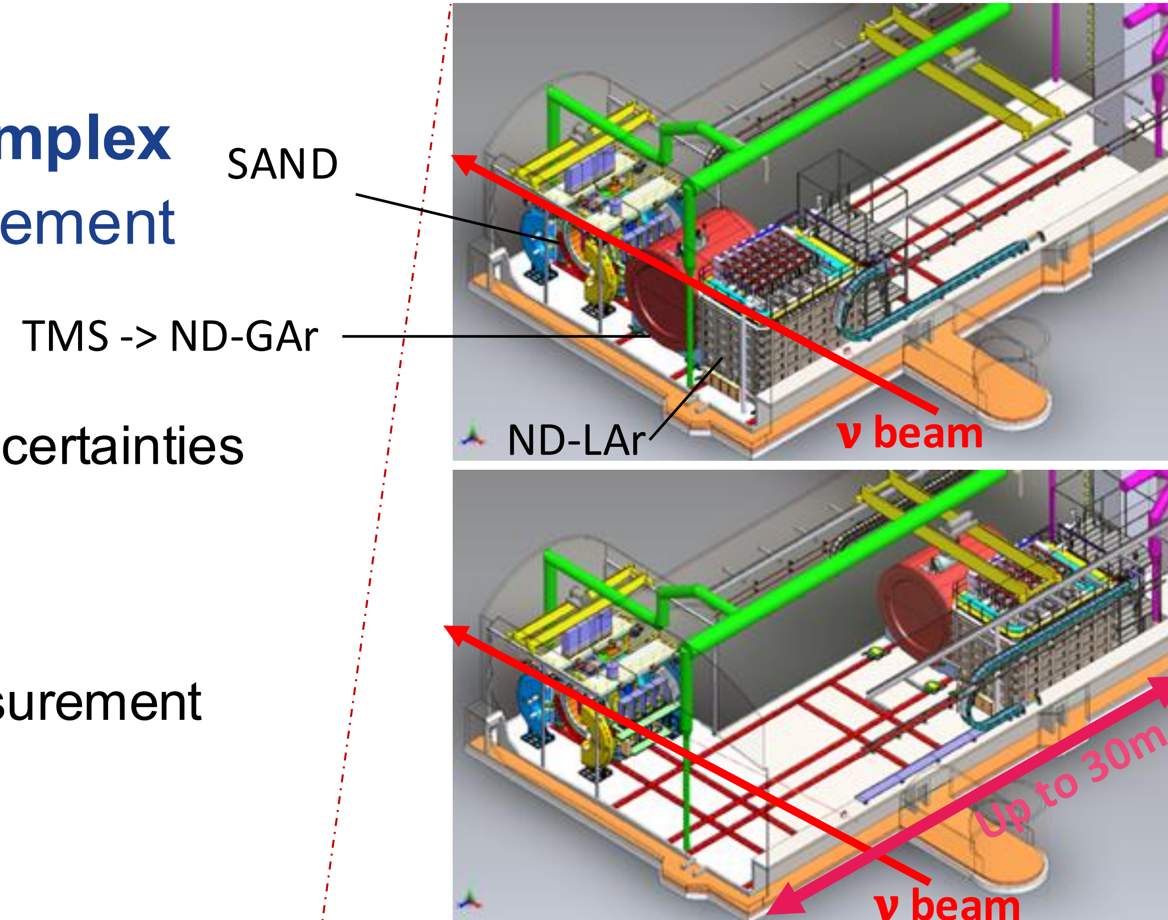


Source: https://warwick.ac.uk/fac/sci/physics/research/epp/exp/dune/

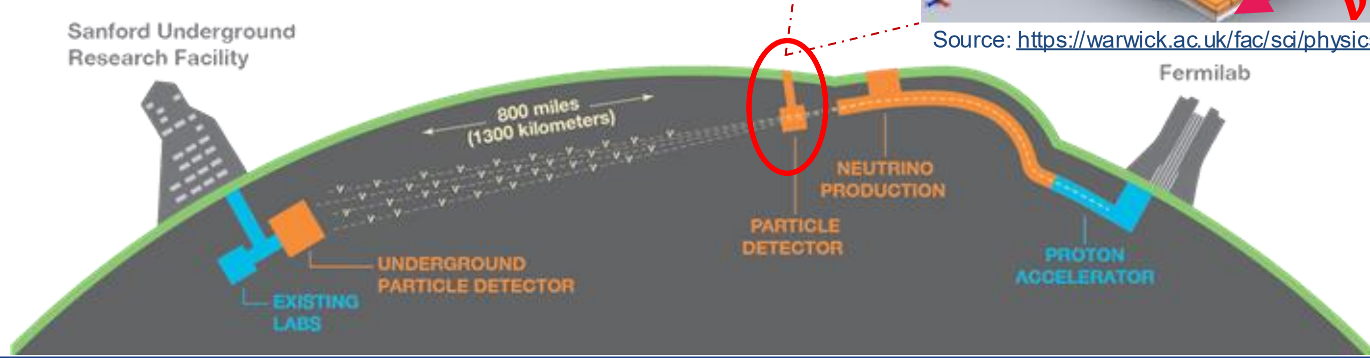


DUNE-PRISM ND Complex
Off-axis beam measurement

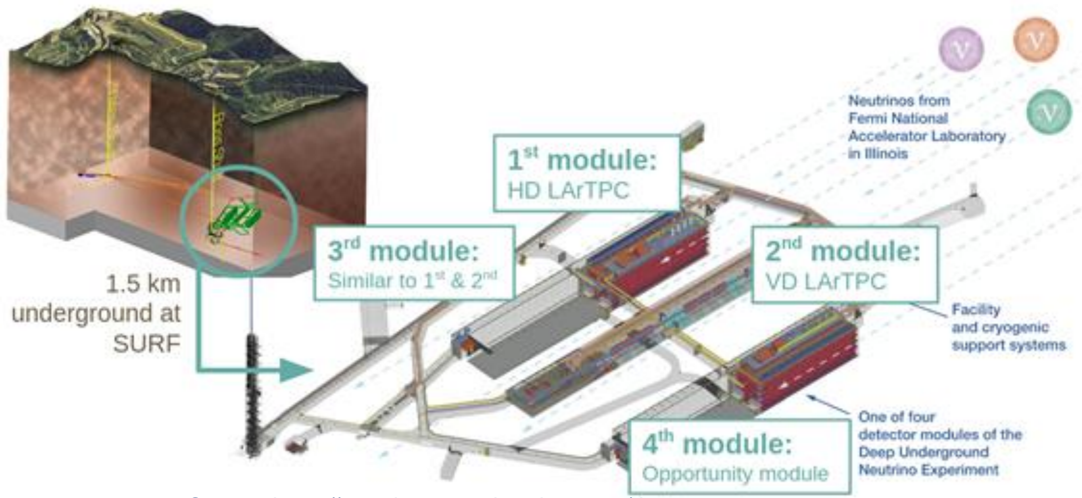
- Constrain systematic uncertainties
- Beam monitoring
- Initial neutrino flux measurement before any oscillations



Source: https://warwick.ac.uk/fac/sci/physics/research/epp/exp/dune/



DUNE: The FD Complex



Source: https://neutrinos.portales.ciemat.es/dune-en

- 2→4 modules
- Fiducial (sensitive) mass of 40kt of liquid argon (LAr)
- Measure flavor-tagged neutrino interaction and compare the results with the ND prediction