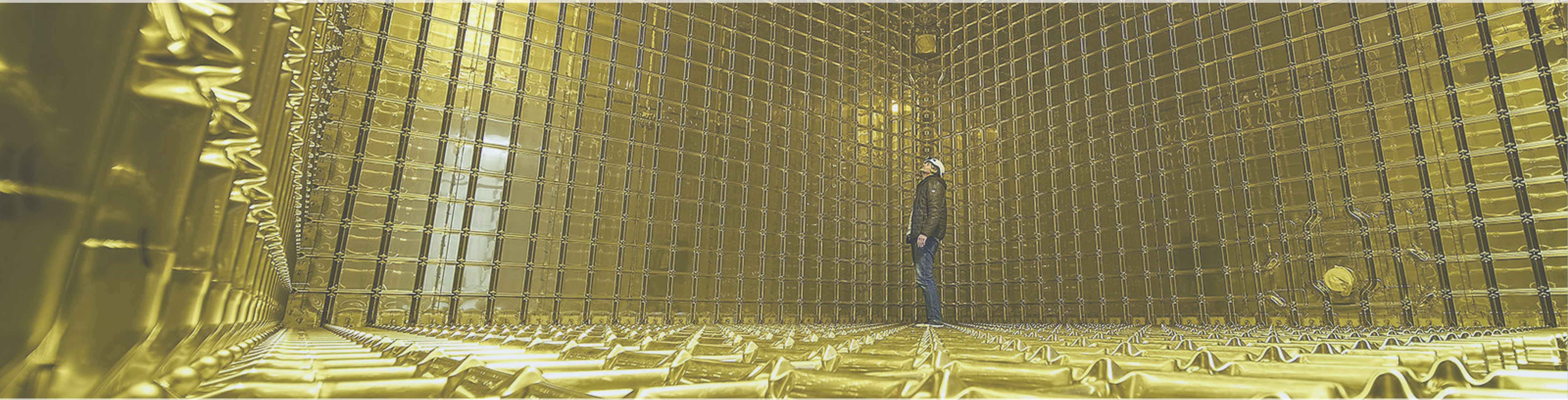
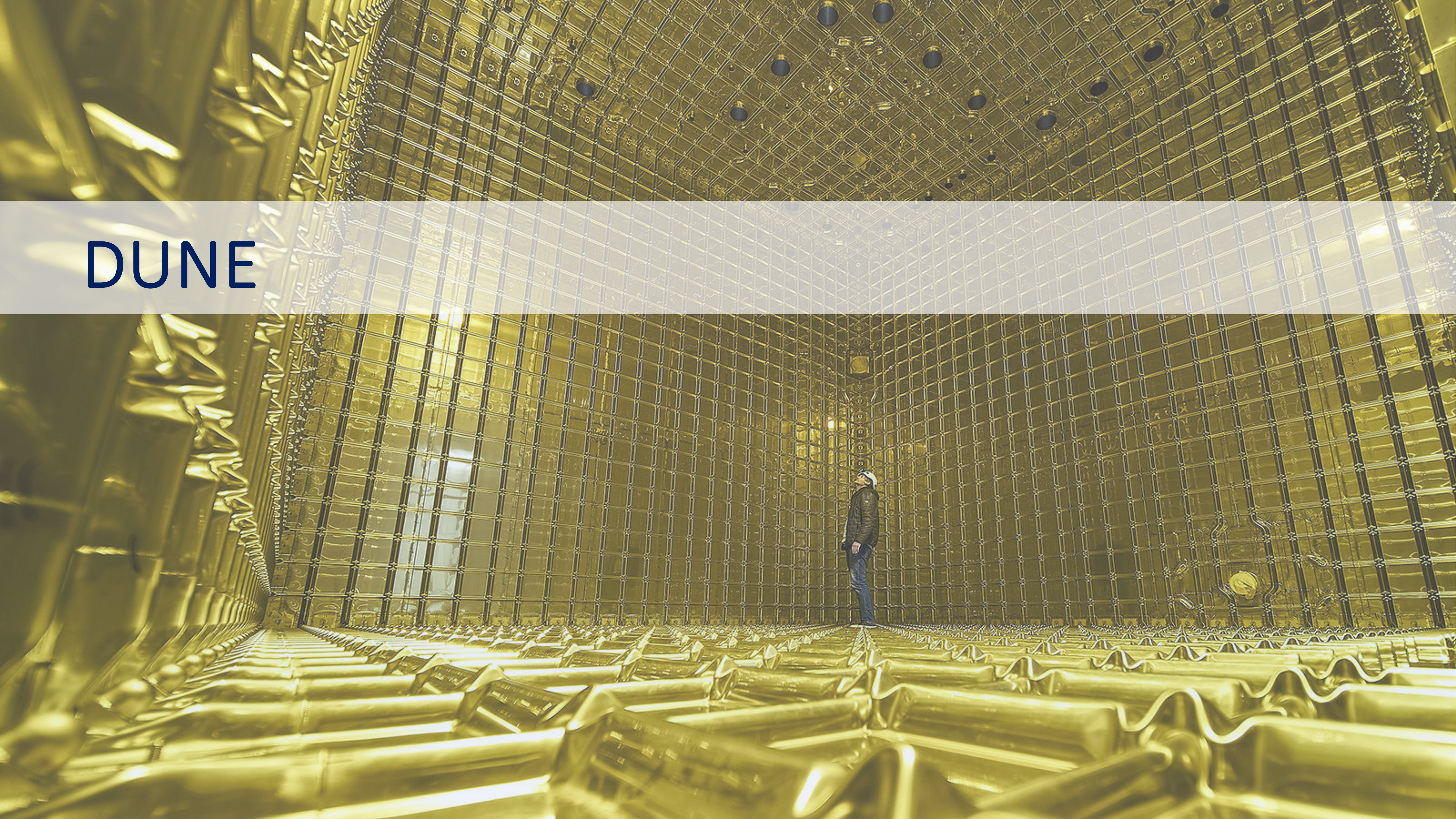
# Machine Learning Approach to a Pandora-based Oscillation Analysis



Isobel Mawby (they/them) *for the DUNE collaboration*

Lancaster University

*Neutrino Physics and Machine Learning @ ETH Zurich*

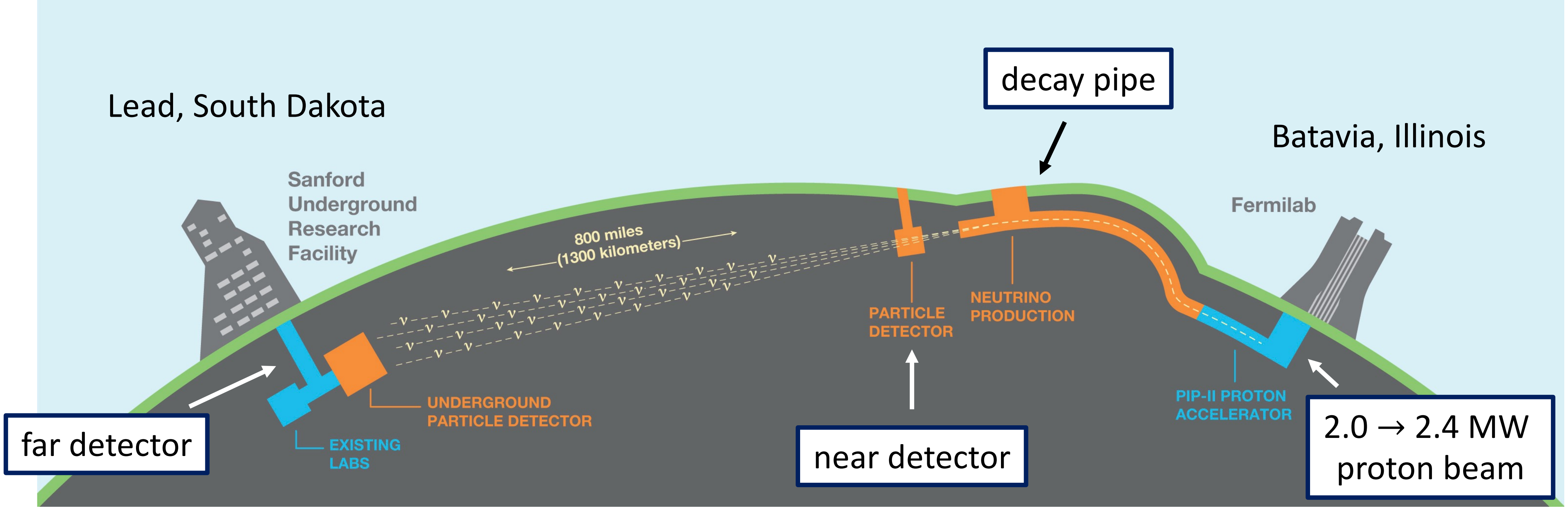

DUNE

# The Deep Underground Neutrino Experiment

## A next generation long-baseline neutrino experiment

**Primary goals:**

- **Precisely** measure the neutrino oscillation parameters
- Search for **beyond the standard model** physics e.g. **proton decay**, and
- Detect **low energy** neutrinos, such as those from a supernova burst

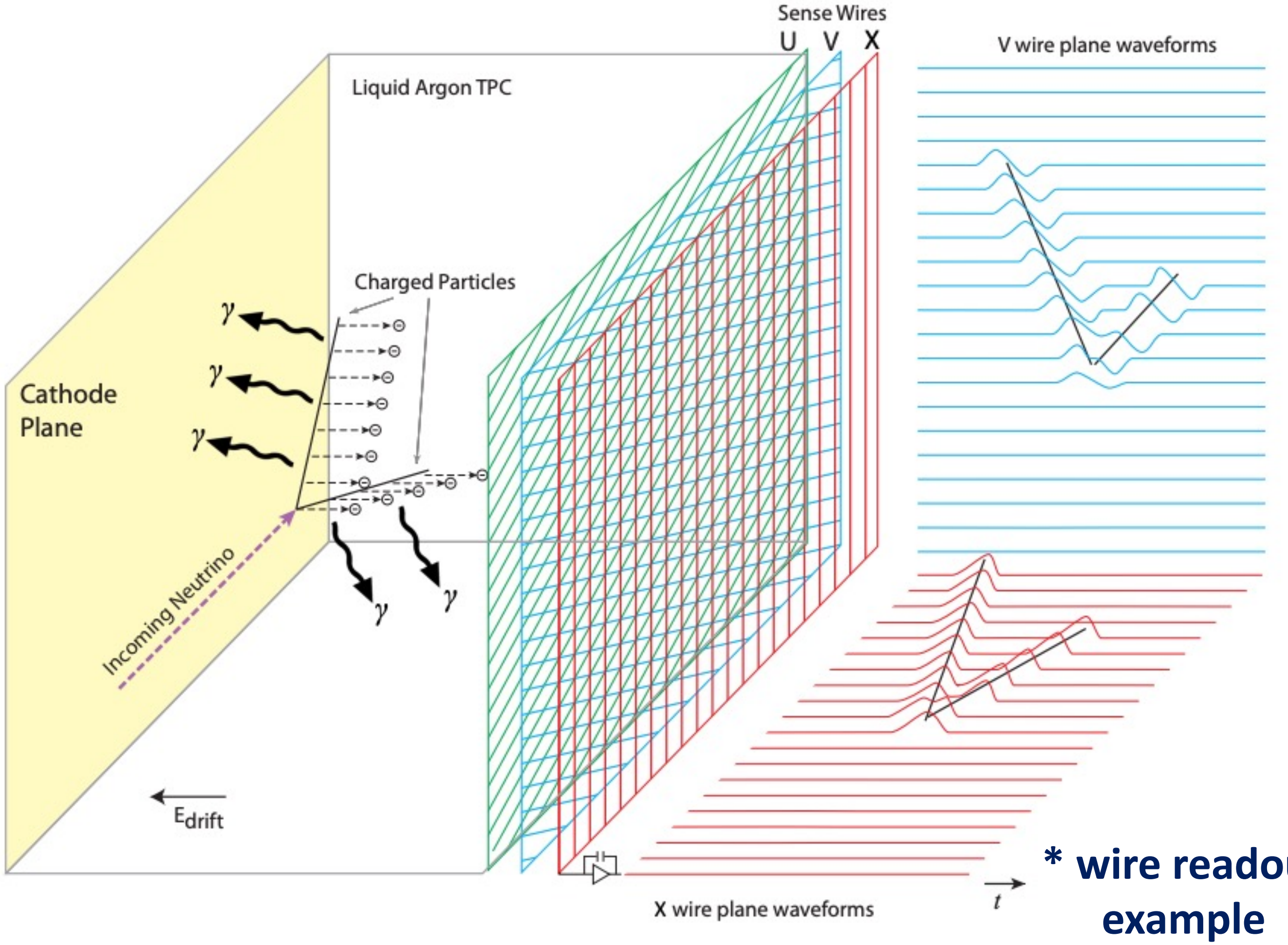


# The DUNE Far Detectors

**Four 10kt (FV) detector modules**

At least three will be Liquid-Argon Time-Projection Chambers (LArTPCs):

- Neutrinos enter the detector and interact with **argon nuclei**
- Outgoing **charged** particles **ionise** the liquid argon as they traverse the detector
- An applied **electric field** drifts the ionisation electrons to a series of **readout planes** where they are detected



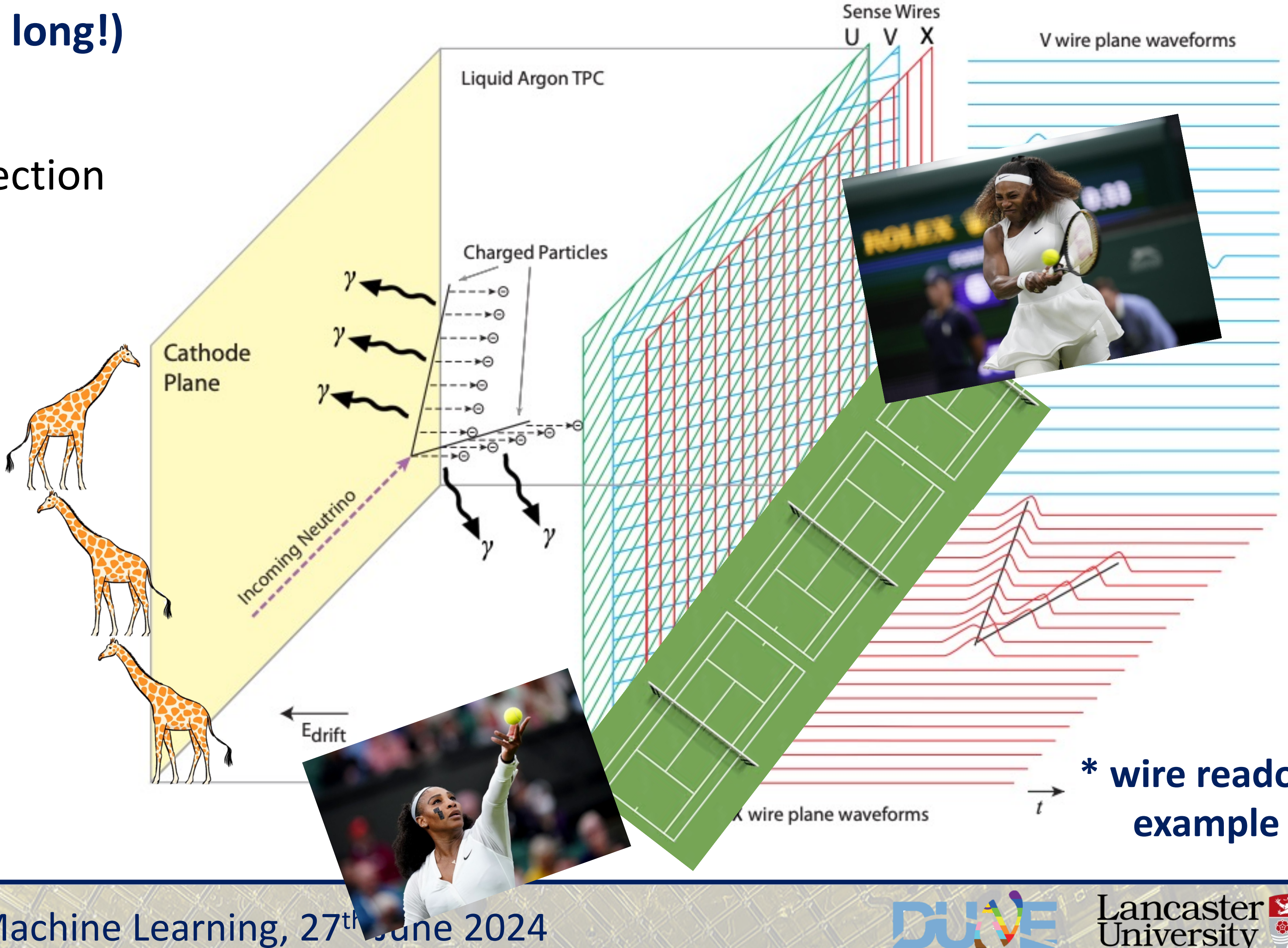

*** wire readout example**

# The DUNE Far Detectors

**Four 10kt (FV) detector modules**
**(approx. 3 giraffes tall and 2.5 tennis courts long!)**

At least three will be Liquid-Argon Time-Projection Chambers (LArTPCs):

- Neutrinos enter the detector and interact with **argon nuclei**
- Outgoing **charged** particles **ionise** the liquid argon as they traverse the detector
- An applied **electric field** drifts the ionisation electrons to a series of **readout planes** where they are detected

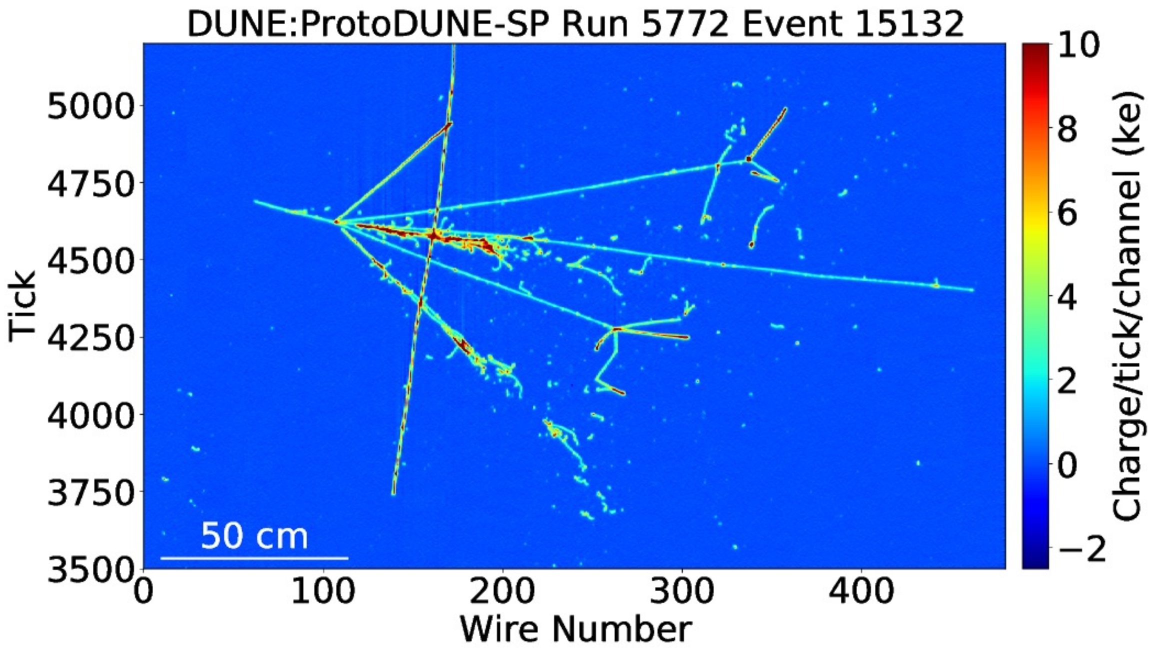



*** wire readout example**

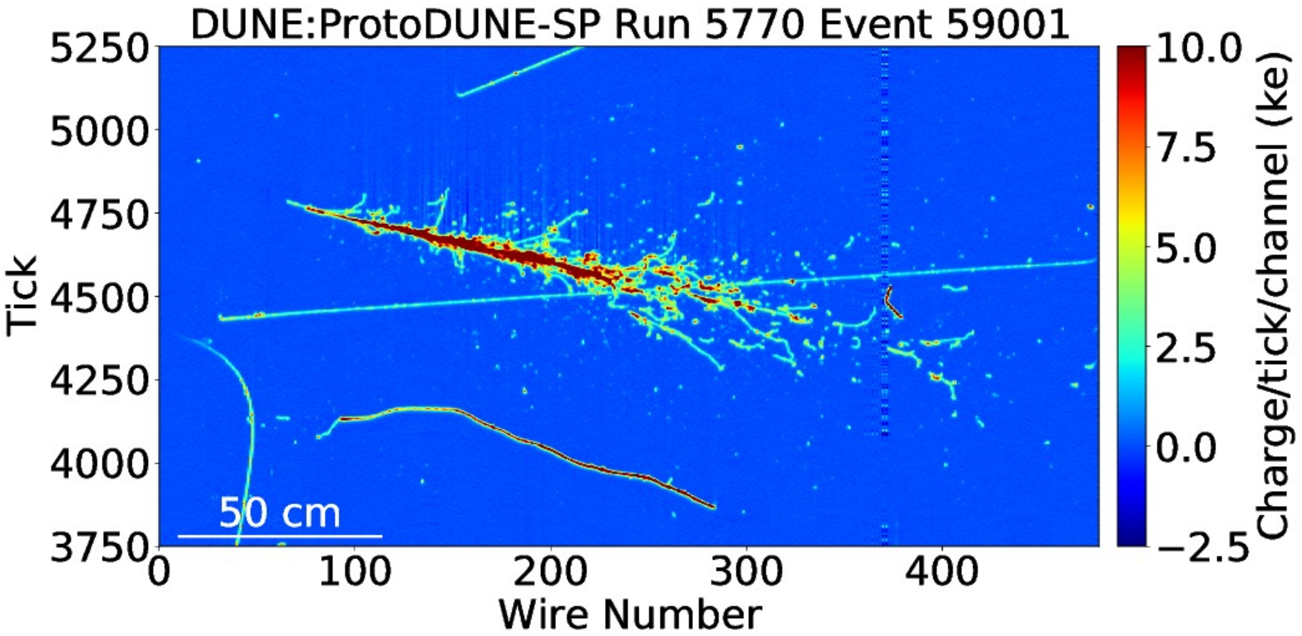
# LArTPC Images

- LArTPC detectors are **fully active and fine grain**
- The images we obtain demonstrate **incredible spatial** and **calorimetric resolution**

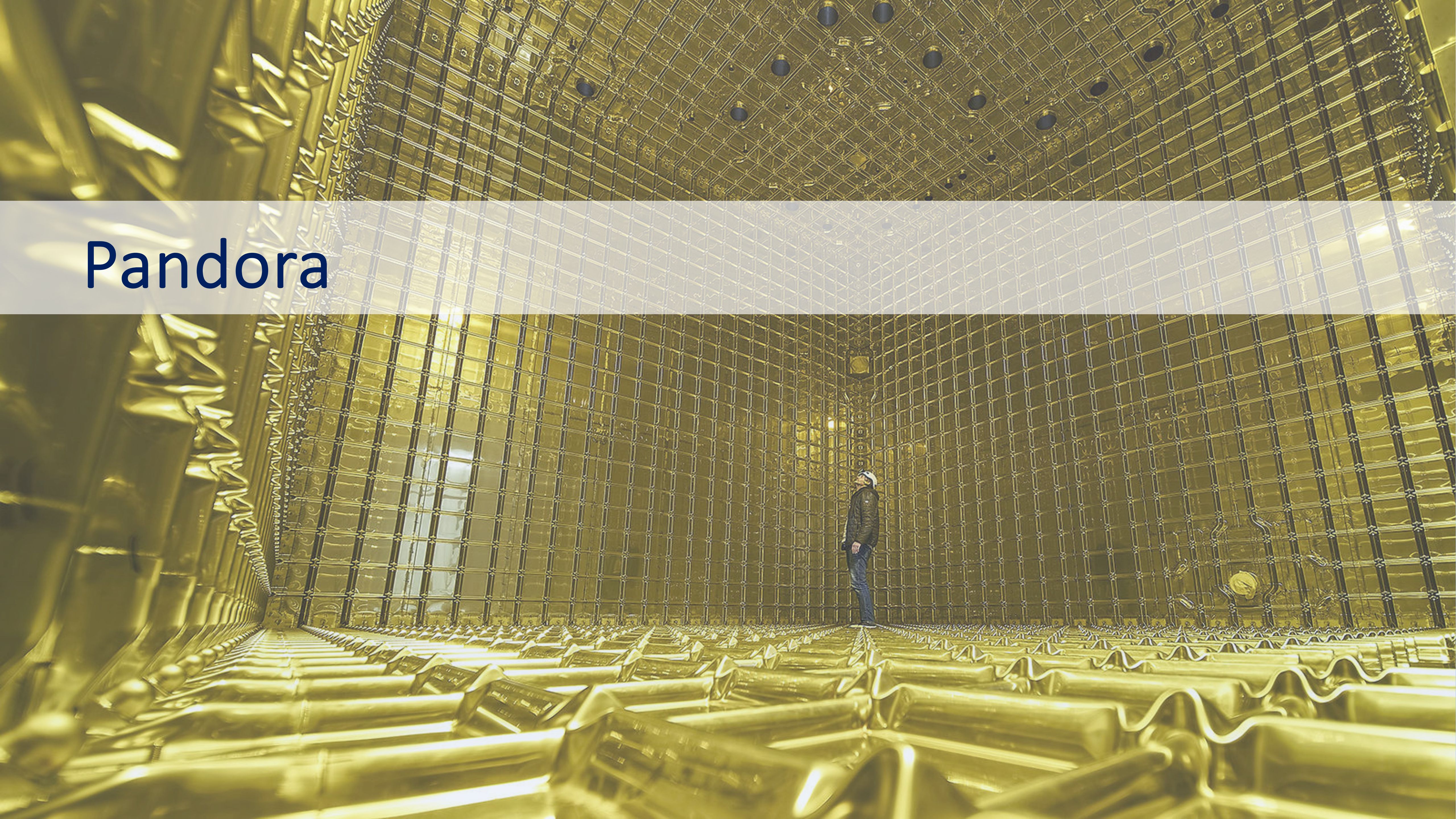
**6 GeV/c pion candidate**




**6 GeV/c electron candidate**

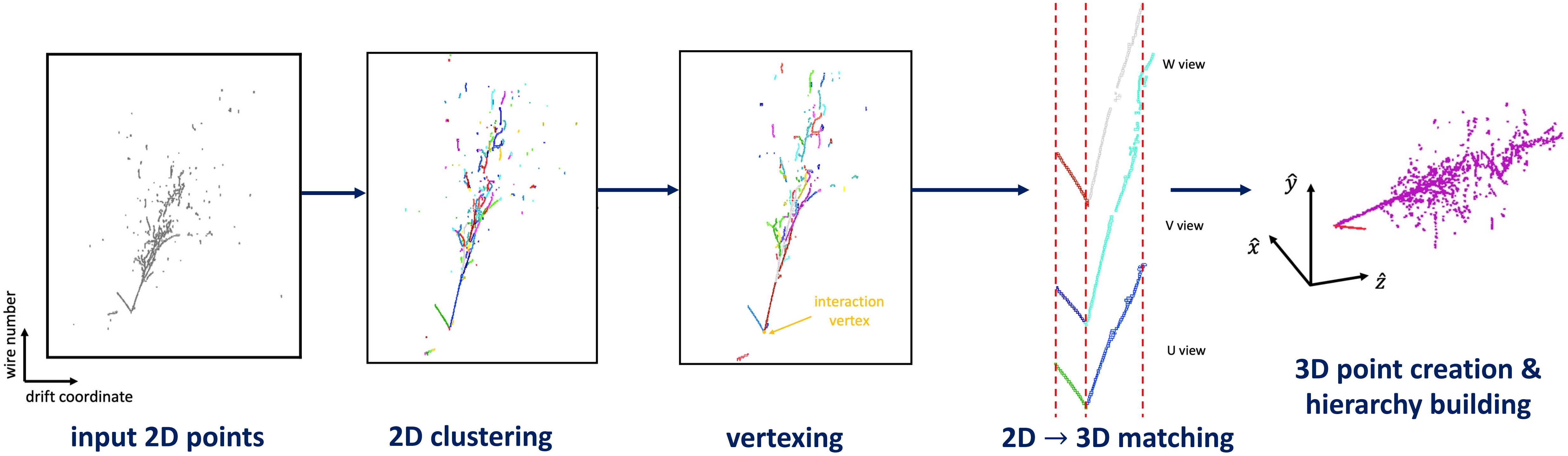


# Pandora



# Pandora – the hope left in the jar

- Pandora is a **pattern recognition software**, used to reconstruct neutrino interactions
- The detail of the **fine-grain images** we obtain from LArTPCs presents a **huge** reconstruction challenge
- Pandora overcomes this with a **'multi-algorithm approach',** where the reconstruction is split into stages composed of many **'hand engineered'** and **machine learning-based algorithms**

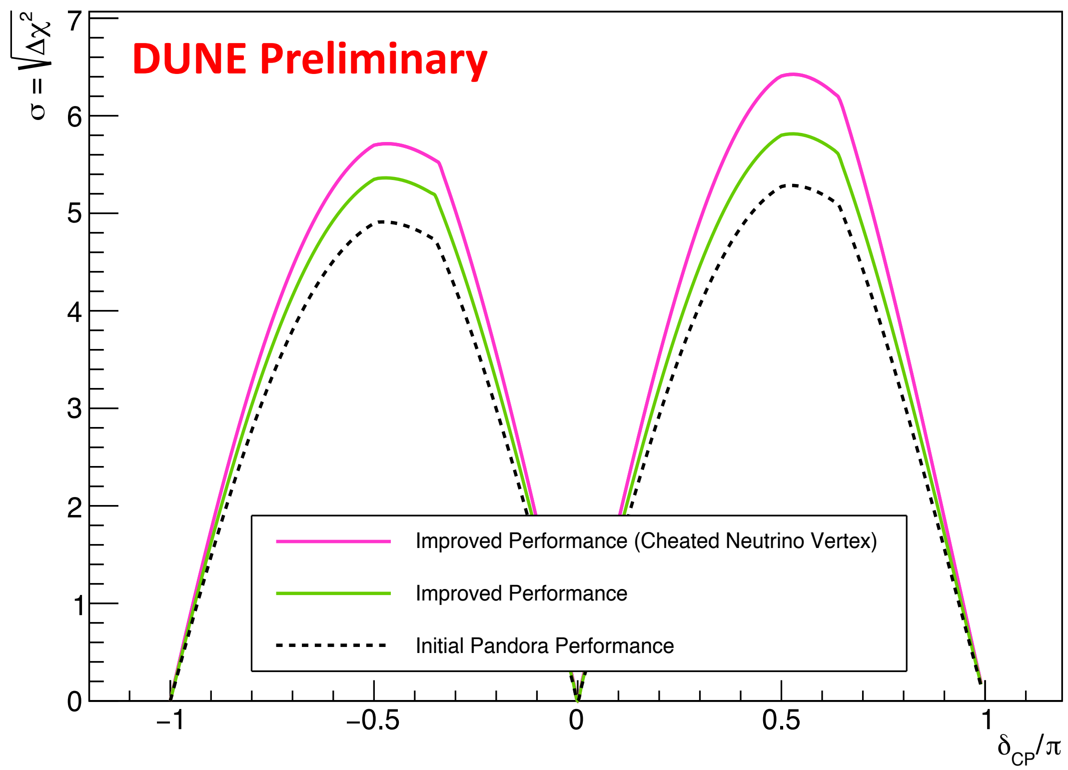


# The Reconstruction → Analysis Continuum

- Pandora uses the **reconstruction → analysis continuum** approach to development:

***i.e. what does the reconstruction need to get right, for an analysis to be optimal?***

- Efforts are therefore focused on the reconstruction improvements **that are important for physics**
- In this talk I'll focus on a Pandora-based measurement of the **CP-violation in neutrino oscillations**

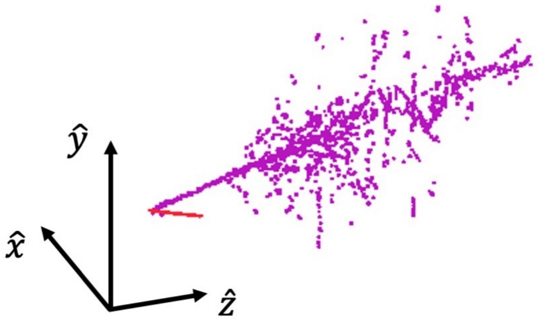


# A Pandora based CP-violation analysis

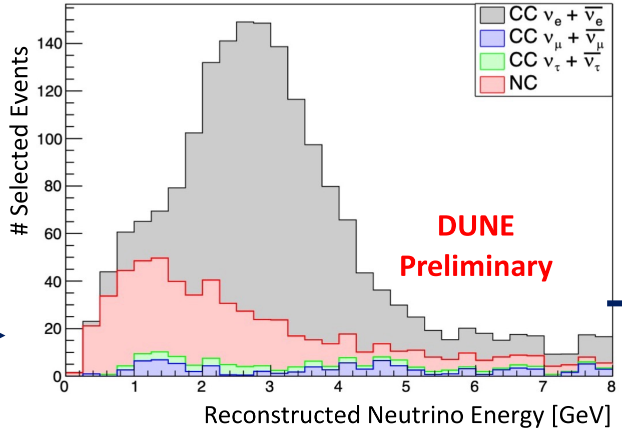
# A Pandora CP-violation Analysis

Pandora Pattern Recognition → Particle Characterisation → $\nu_e/\nu_\mu$ Selection → Neutrino Energy Estimation → CP-Violation Sensitivities

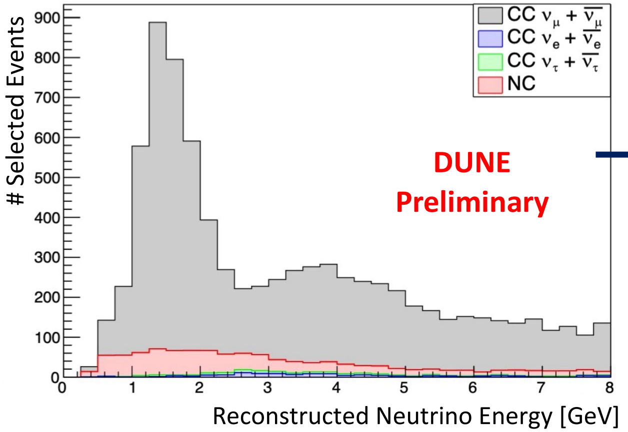
**reconstruction** ← | → **analysis**

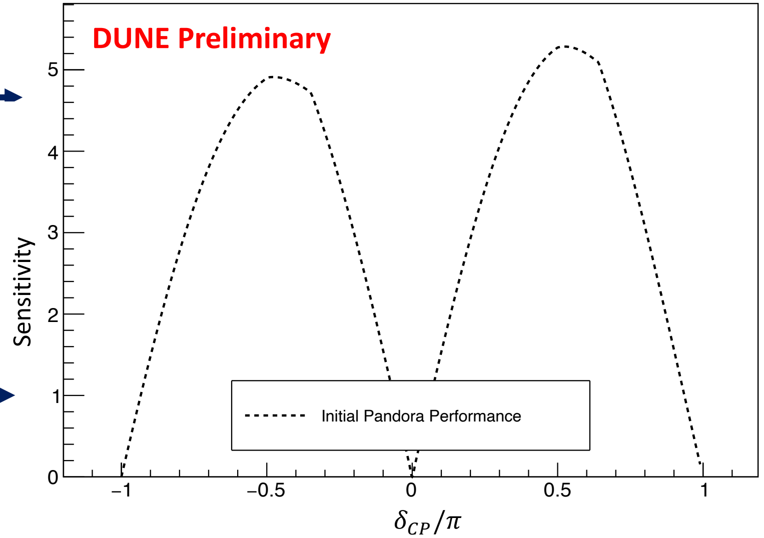



→ Is $\nu_e$ selected? →



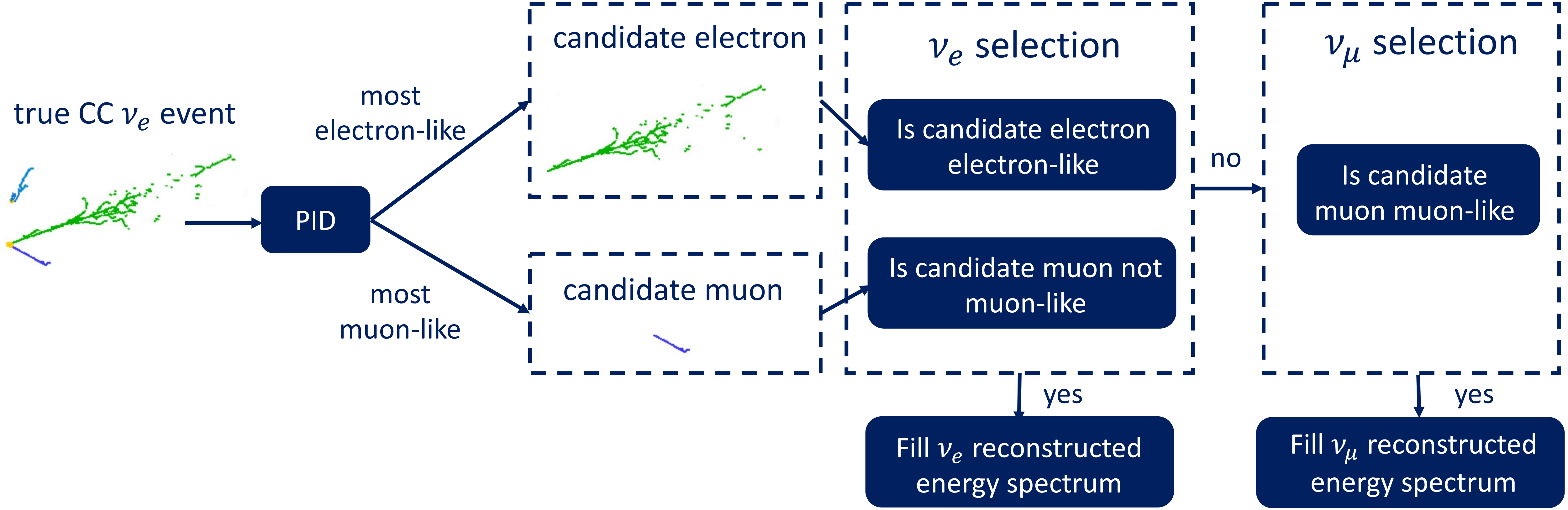

→ Is $\nu_\mu$ selected? →



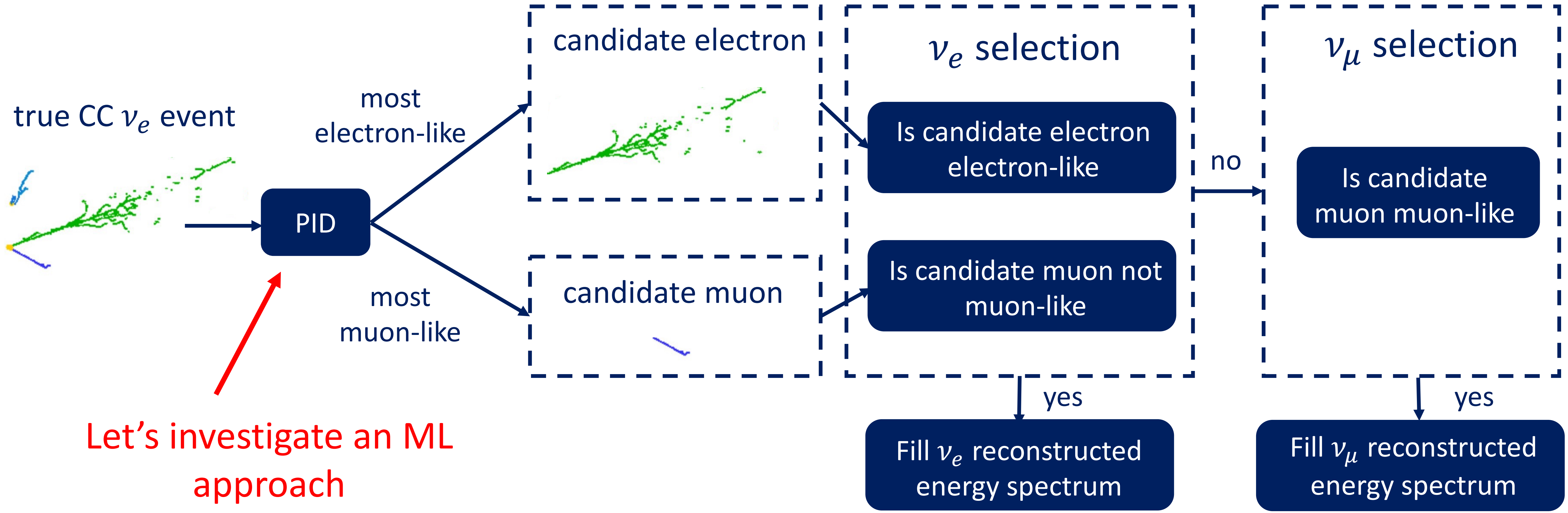
# The Pandora-based $\nu_e/\nu_\mu$ Selection

**In the Pandora-based analysis, events are selected based on the identity of the leading lepton (if it exists)**



# The Pandora-based $\nu_e/\nu_\mu$ Selection

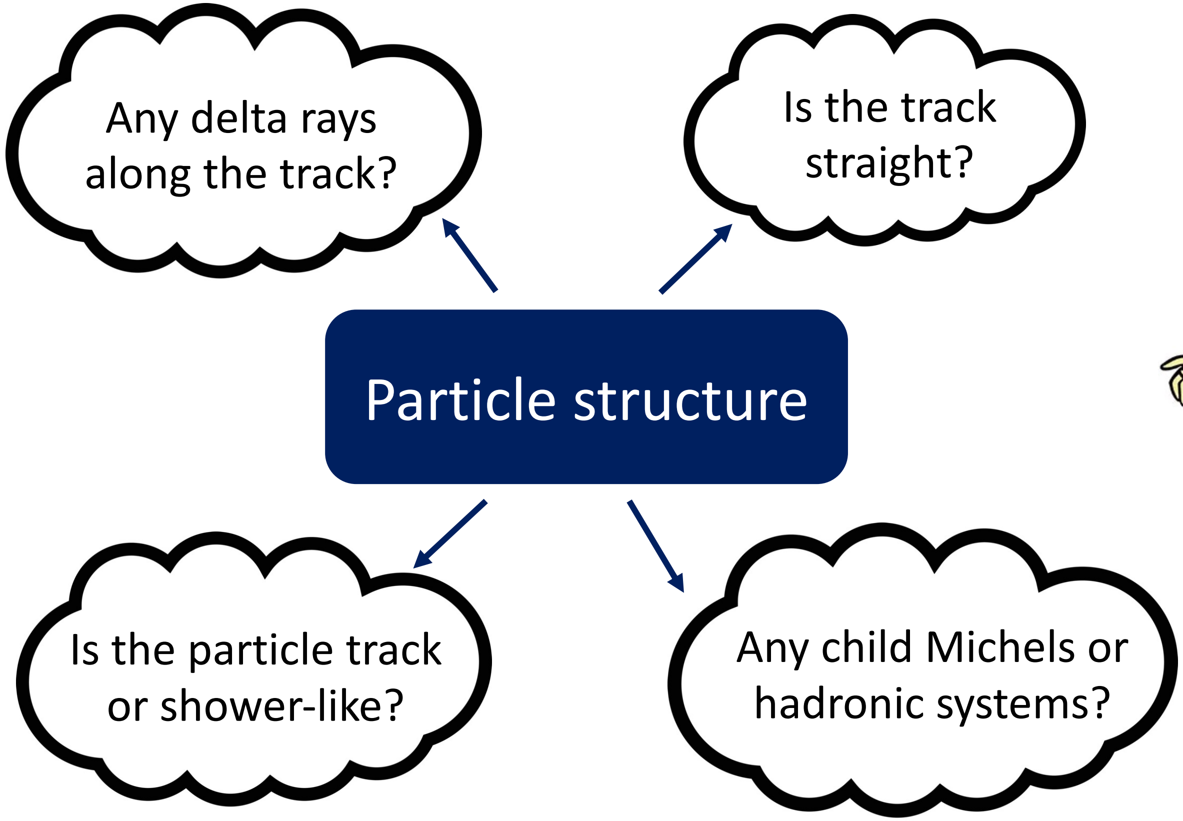
**In the Pandora-based analysis, events are selected based on the identity of the leading lepton (if it exists)**



# Building a Network

# Network Input

- Arguably the most **important choice is the network's input** – we need to think about this carefully
- So, what do we (as physicists) look for when trying to perform PID?




Our input should capture the spatial structure of our particle trajectories

⟹ Take the form of a 2D/3D grid

# Network Input

- Arguably the most **important choice is the network's input** – we need to think about this carefully
- So, what do we (as physicists) look for when trying to perform PID?

*dE/dx* can be used to separate some tracks

Is there a Bragg peak?

Calorimetry

Initial *dE/dx* separates showers

Energy of child particles can separate Michels from hadrons

Our input should contain energy information, with a particular focus on the start and end regions

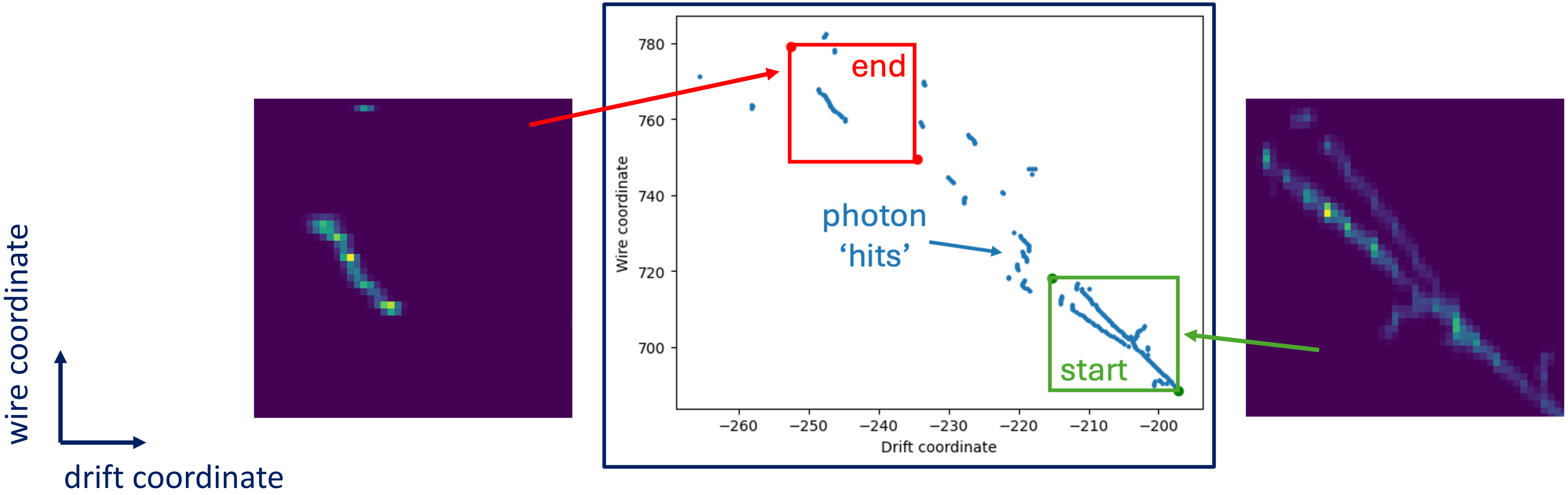
⟹ Two grids, weighted by energy deposits

# Network Input

- 2D networks are the natural choice, so we need our inputs to be 2D grids

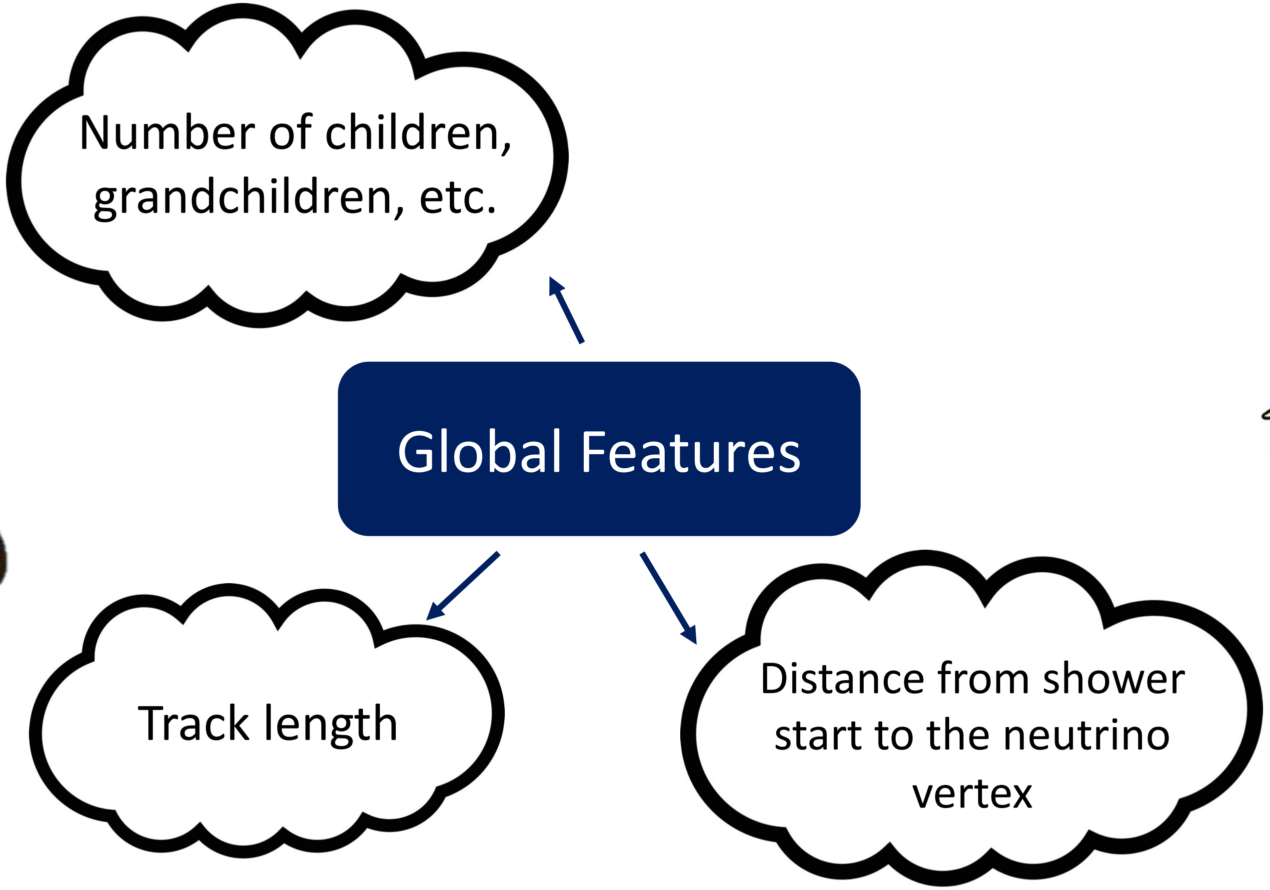
**⟹we will build a start and end grid for the energy deposits of each U, V and W Pandora view**

- The start/end grid triplets will correspond to the **same 3D space,** allowing the network to infer the *dE/dx*



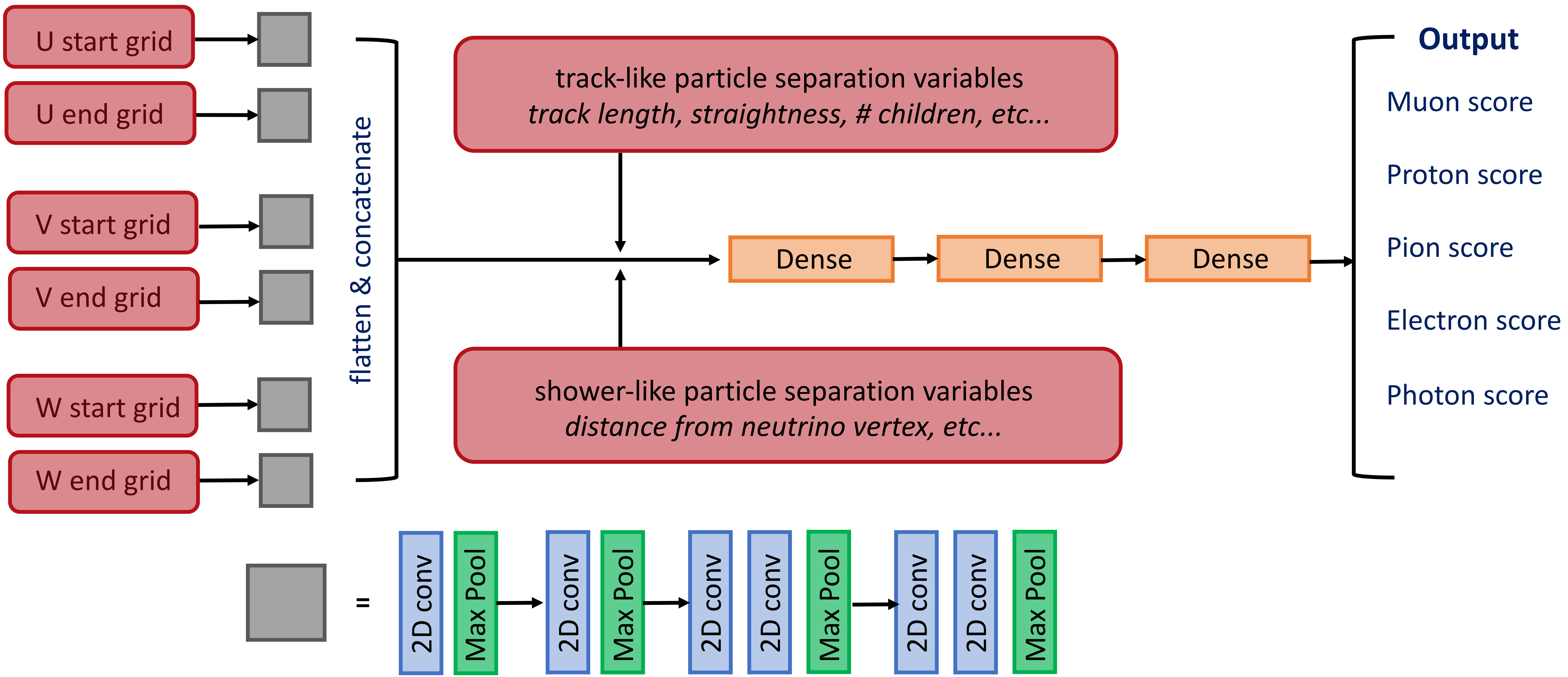
# Network Input

- By focusing on the start and end points, and only using the hits contained within the particle, we miss out some ‘**higher-level’ physics features** that are useful for PID




We should pass these physics variables into our network

# Network Architecture

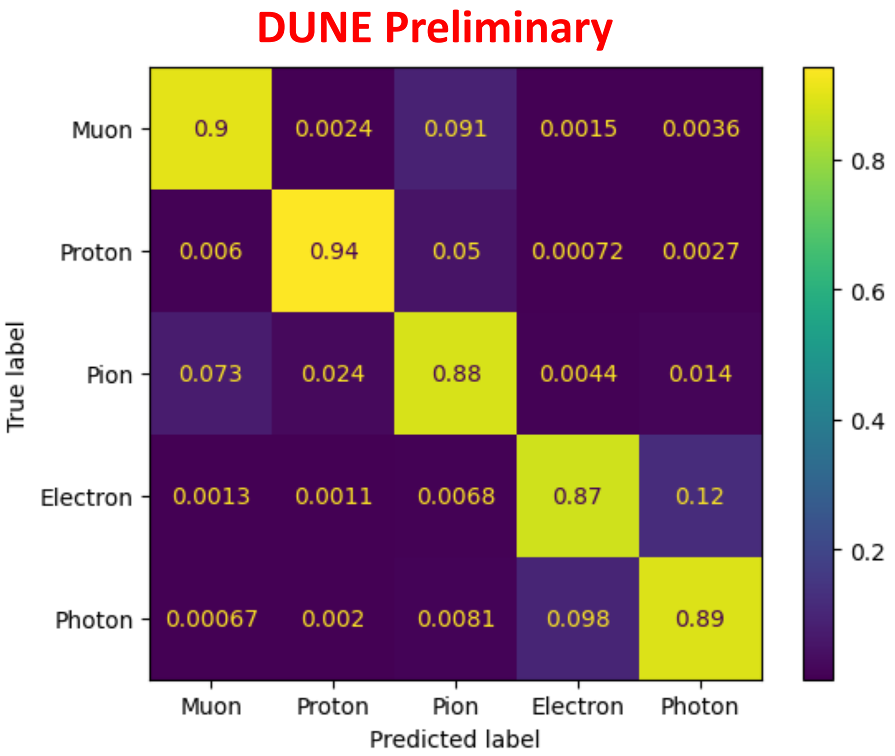


# Performance: Confusion Matrices

- Training (validation) performed on ~**2M (~200,000)** particles, passing reconstruction quality thresholds
- Network trained for 10 epochs, using the **GPUs** of the **Lancaster HEC cluster**

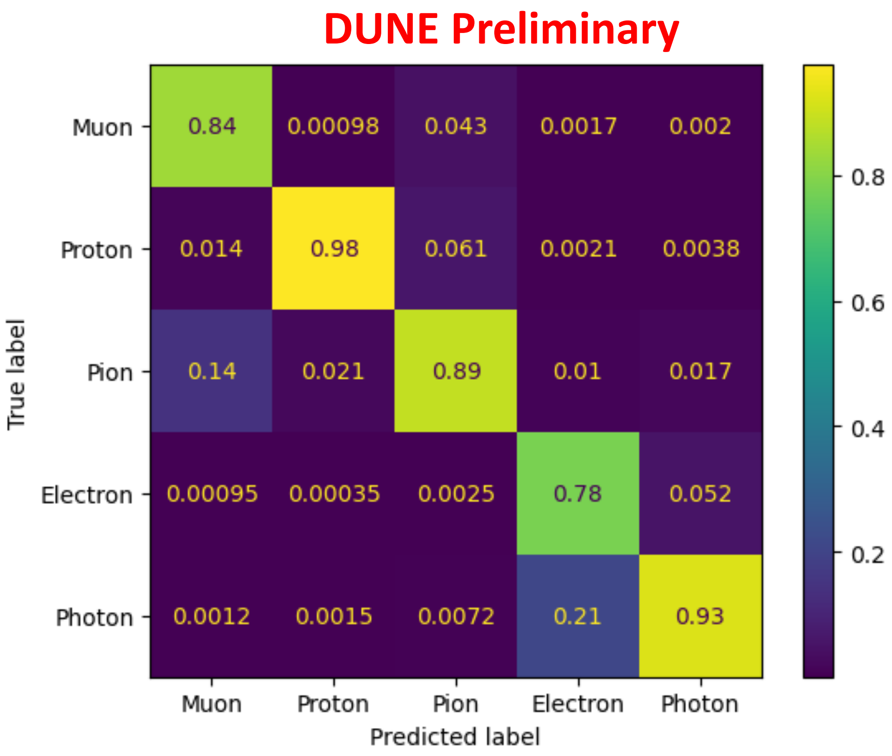
Efficiency normalised:

(read horizontally)

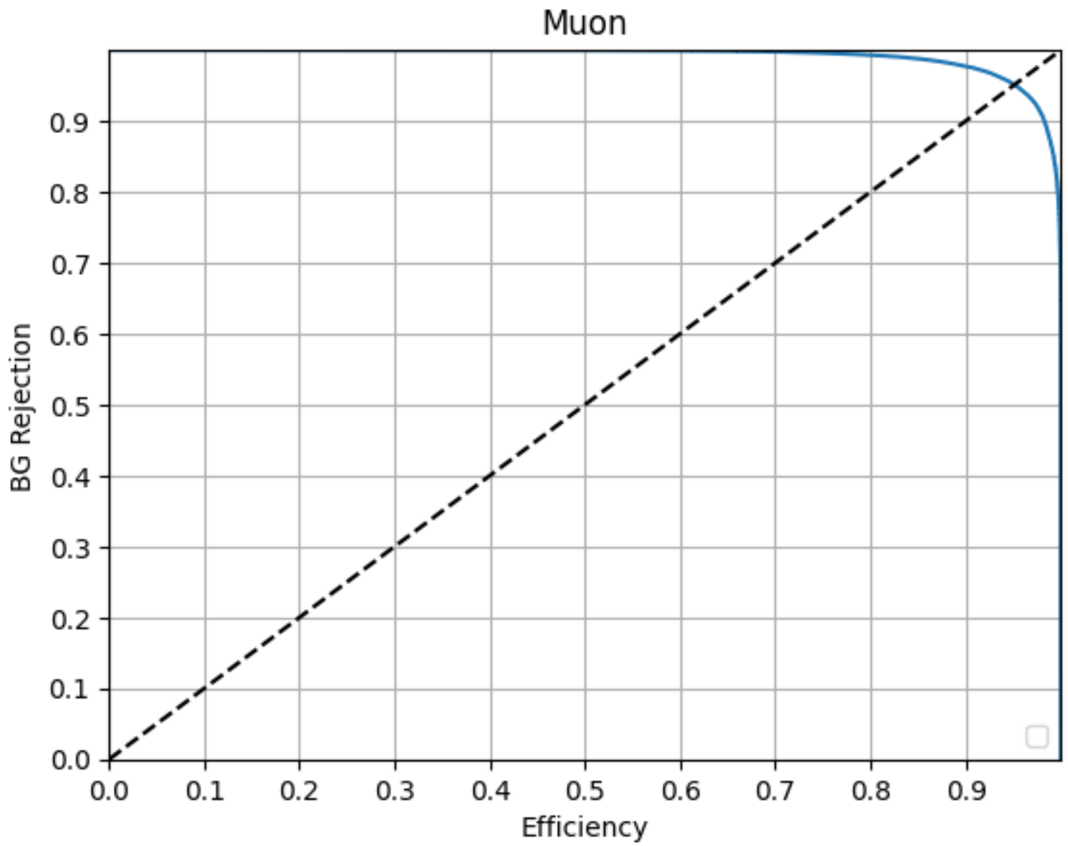


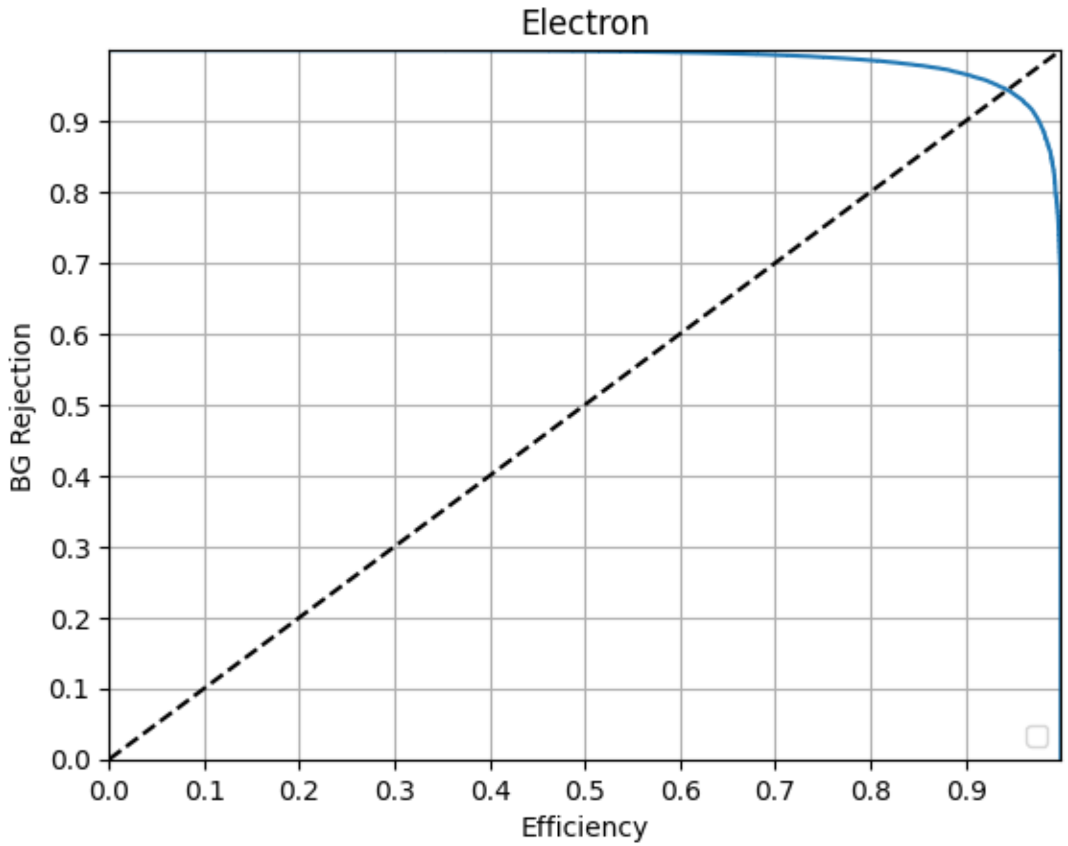

Purity normalised:

(read vertically)



# Performance: ROC Curves




* For particles passing reconstruction quality thresholds

# Performance: Neutrino Selection Metrics

| | Selection Efficiency | Selection Purity |
|---|---|---|
| $\nu_e$ Selection | 67.7% | 72.7% |
| $\nu_\mu$ Selection | 92.8% | 93.2% |

- The $\nu_e$ **selection** is limited by the **reconstruction of electron** showers, particularly their **contamination and incomplete growth**
- The $\nu_\mu$ **selection** is limited by $\mu/\pi$ **confusion**

# Building the Neutrino Hierarchy

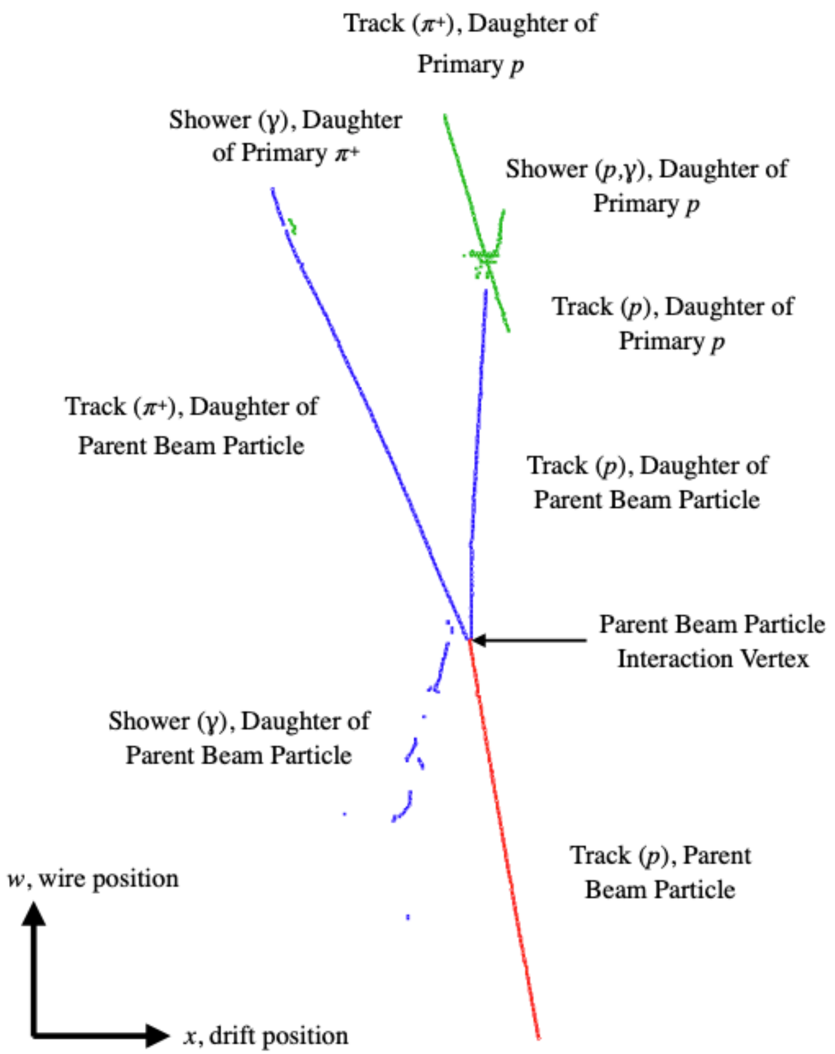
**VERY EARLY DAYS WARNING**

# The Neutrino 'Hierarchy'

VERY EARLY DAYS WARNING

**One of the final algorithms in Pandora constructs the 'neutrino hierarchy'**

- Obviously very important in **cross-section analyses,** but is also a very important stage in our **CP-violation analysis:**
  - Optimise **efficiency** by tagging leading leptons as children of the neutrino
  - Reduce **backgrounds** by tagging neutrino grandchildren (and higher) $\pi$ & $\gamma$ as such
  - Achieve a **better energy reconstruction** by correctly identifying the evolution of particles in neutrino interactions



# Graph Neural Network (GNN) Approach

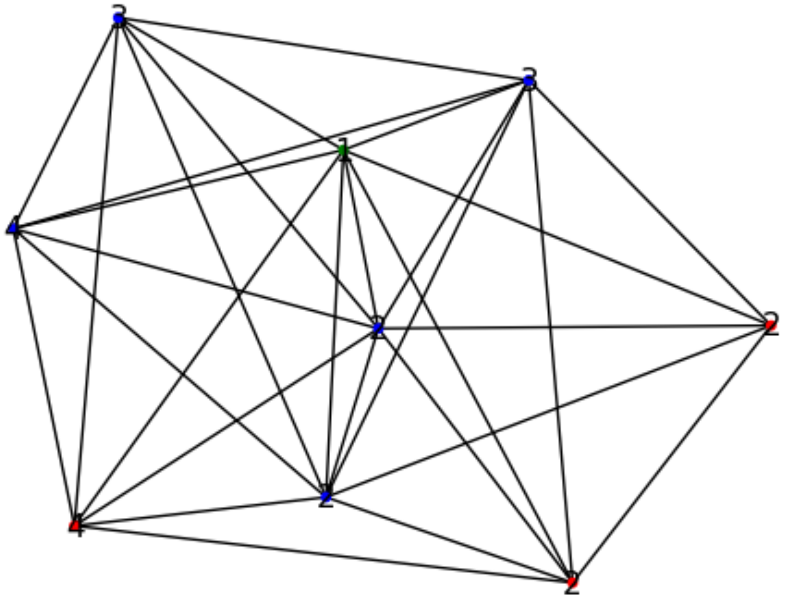
VERY EARLY DAYS WARNING

- Currently Pandora uses **a 'hand engineered' algorithm**, which identifies most particles as neutrino children unless absolutely sure otherwise ⟹ needs to be improved!

- **Particle hierarchies look like graphs!**
  - Nodes represent particles
  - Edges represent parent-child links

- With GNNs, there's often several valid approaches to a given task

- I'm currently considering a **two step link-prediction solution:**



Identify neutrino-particle edges → Identify particle-particle edges

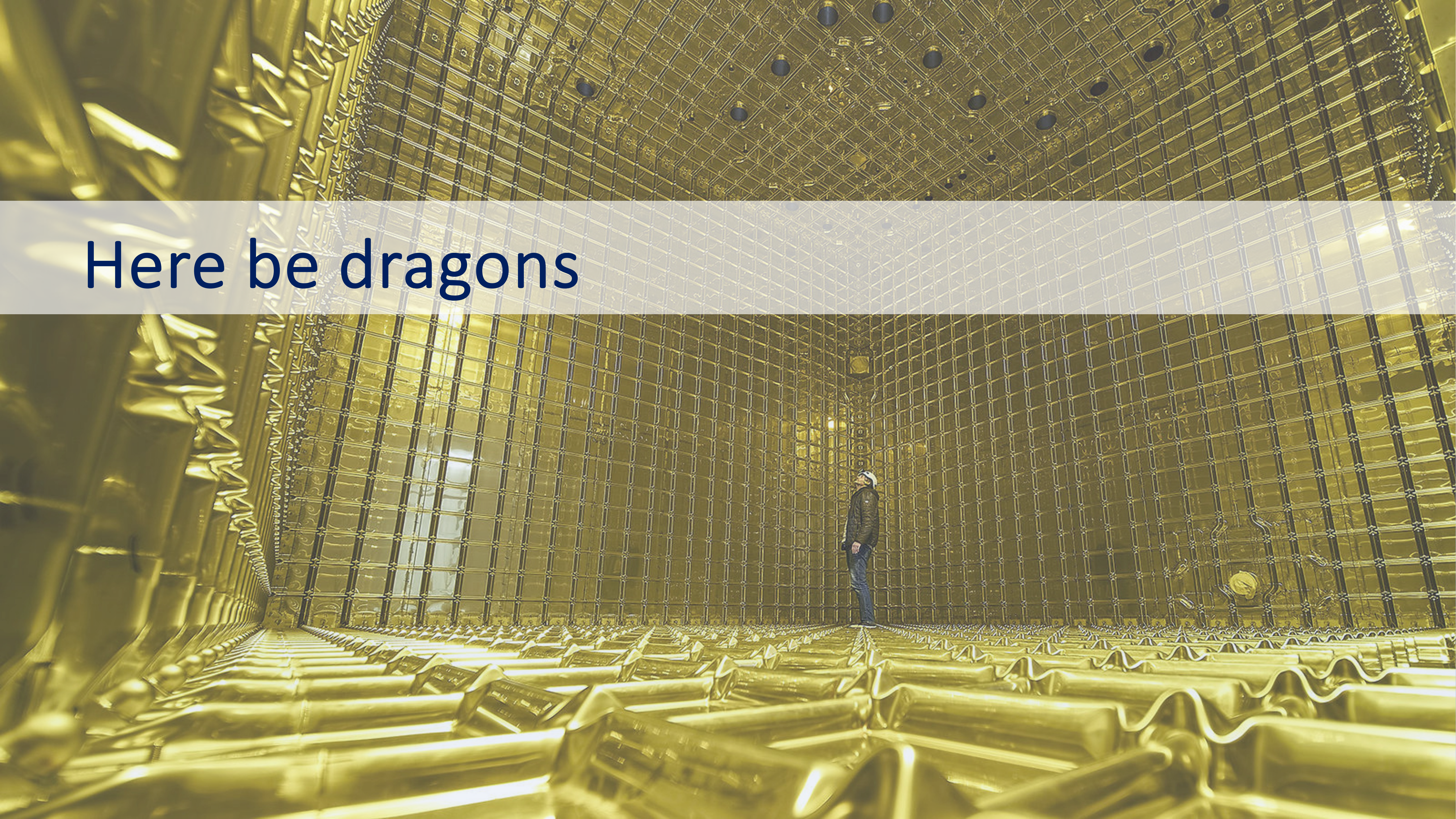
**Off to a good start, hopefully exciting work to come!**

# Conclusions

- Pandora is a **pattern recognition software,** used to reconstruct neutrino interactions
- Our reconstruction performance is best understood in terms of **physics analysis,** and these results are used to **drive developments**
- I've introduced DUNE's flagship analysis: **the search for CP-violation using the DUNE far detectors**
- We have a strong analysis foundation
- Can now turn our attention to **reconstruction improvements,** e.g.
    - Neutrino hierarchy building via GNNs
    - Shower re-clustering (Maria Brigida's talk)
    - Neutrino vertexing (Andy's talk)

**Thank you for listening!**

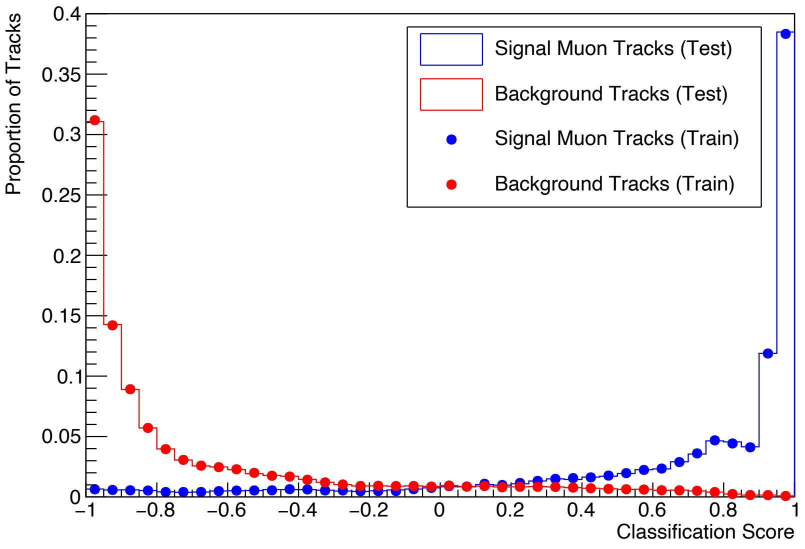
# Here be dragons



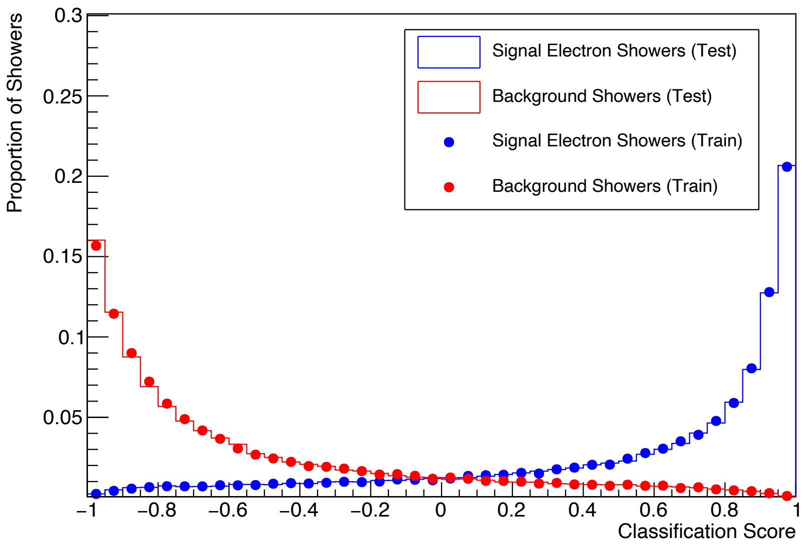
# Pandizzle and Pandrizzle

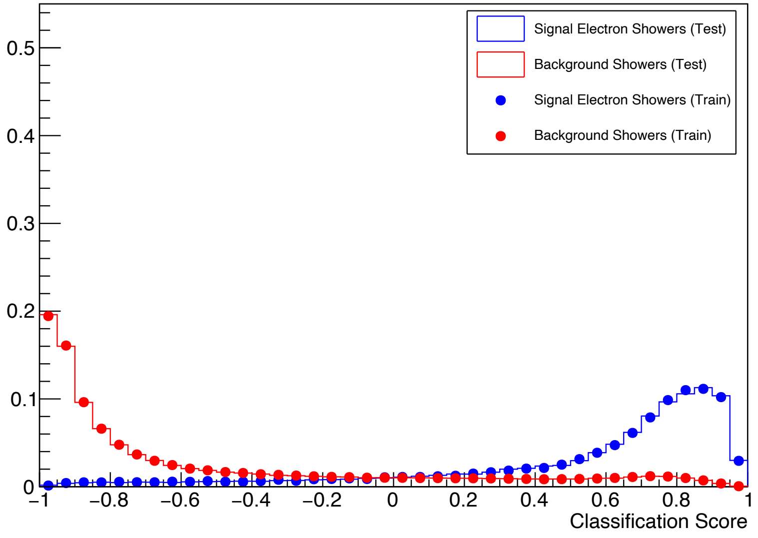
- Currently, we use BDTs for the PID:

## Muon PID: Pandizzle




## Electron PID: Pandrizzle (made up from two BDTs)




- They're very good, but can improve with **more sophisticated machine learning** methods?