Reconstruction at DUNE

Ben Morgan

Workshop on Efficient Computing for HEP
Edinburgh 17-18 February 2020

DUNE and SBN in a Nutshell

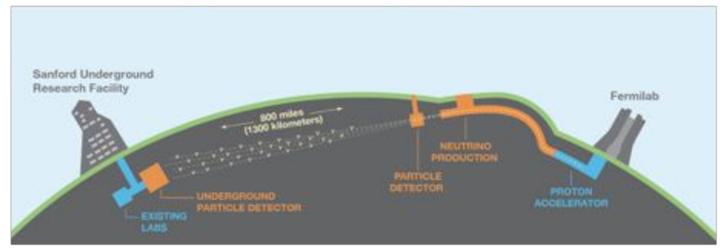
One of the key technologies in the current and future neutrino physics programmes is the Liquid-Argon Time-Projection Chamber (LArTPC)

Short-baseline programme

ICARUS **MicroBooNE** **SBND**

- Three LArTPC detectors located along the Booster Neutrino Beam (BNB) at Fermilab
- Main goal is to investigate the potential sterile neutrino signals from LSND and MiniBooNE
- Precision cross-section measurements for neutrino interactions on argon

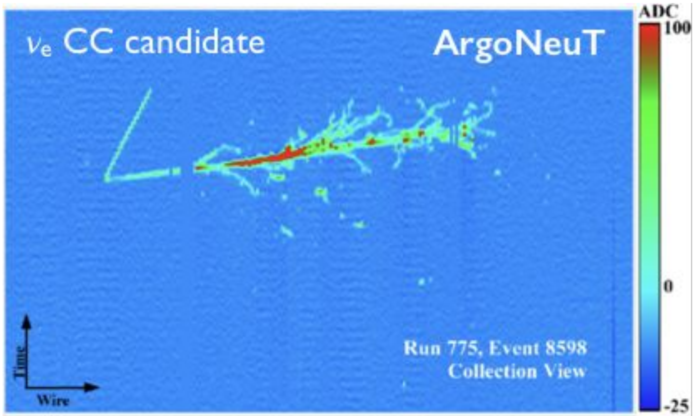
Long-baseline programme

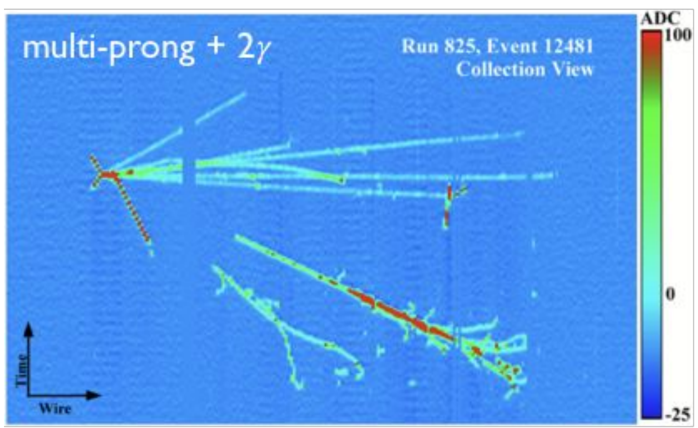


ProtoDUNEs **DUNE**

- Neutrino oscillation physics:
  - Discover CP violation in the leptonic sector
  - Resolve mass hierarchy
  - Test three-flavour paradigm
  - Precision parameter measurement
- Proton decay
- Supernova neutrinos

Why Liquid Argon?





R. Acciari *et al*, Phys. Rev. D 95, 072005 (2017)

- LArTPC detectors are fully active and fine grain, offering superb spatial and calorimetric resolution:
  - Reconstruction of multi-prong final states.
  - Particle identification:
    - $\mu/p/K$ in particle tracks
    - e/γ in electromagnetic showers
- Potential for high efficiency and low backgrounds in most channels
- Scalable to multi-kiloton masses.

LAr TPC data volumes

- The first far detector module will consist of 150 **Anode Plane Assemblies (APAs)** which have 3 planes of wires with 0.5 cm spacing. Total of **2,560 wires per APA**
- Each wire is read out by 12-bit ADC's every 0.5 microsecond for 3-6 msec. Total of **6-12k** samples/wire/readout.
- Around 40 MB/readout/APA uncompressed with overheads → **6 GB/module/readout**
- 15-20 MB compressed/APA → **2-3 GB/module/readout**
- Read it out ~5,000 times/day for cosmic rays/calibration → **3-4PB/year/module (compressed)**

(x 4 modules x stuff happens x decade) = ….

1 APA – 2,560 channels
150 of these per FD module

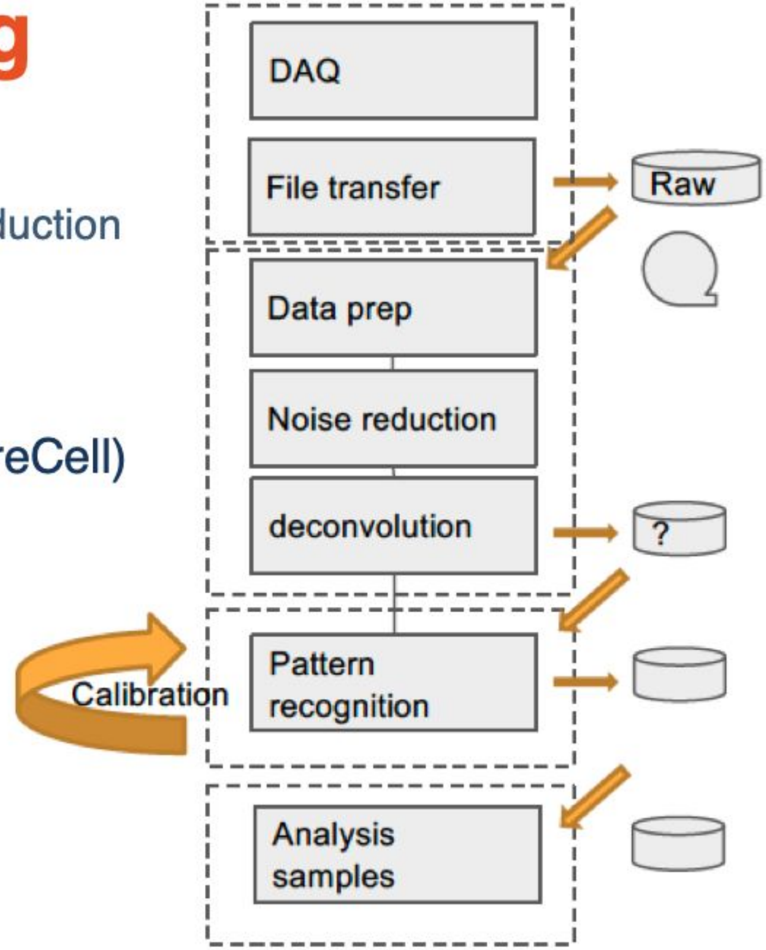
Collaborative Computing Needs for DUNE
Heidi Schellman, CHEP2019, https://indico.cern.ch/event/773049/contributions/3581360/

Data and Event Sizes

Event size is significant, even compressed!

LAr TPC data processing

- hit finding and deconvolution
  - **x5 (ProtoDUNE) -100 (Far Detector)** data reduction
  - Takes 30 sec/APA
  - Do it 1-2 times over expt. lifetime
- Pattern recognition (Tensorflow, Pandora, WireCell)
  - Some data expansion
  - Takes ~30-50 sec/APA now
  - Do it ? times over expt.
- Analysis sample creation and use
  - multiple^2 iterations
  - Chaos (users) and/or order (HPC)

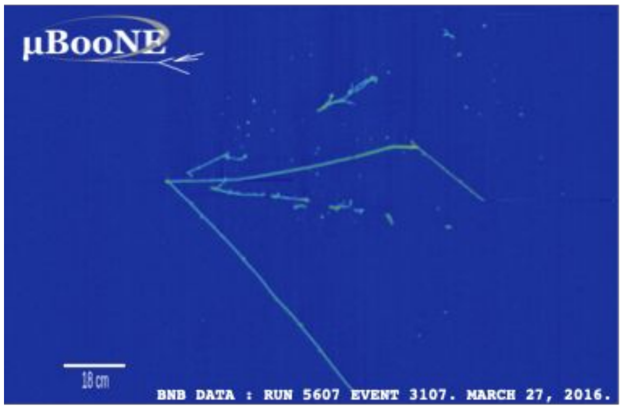


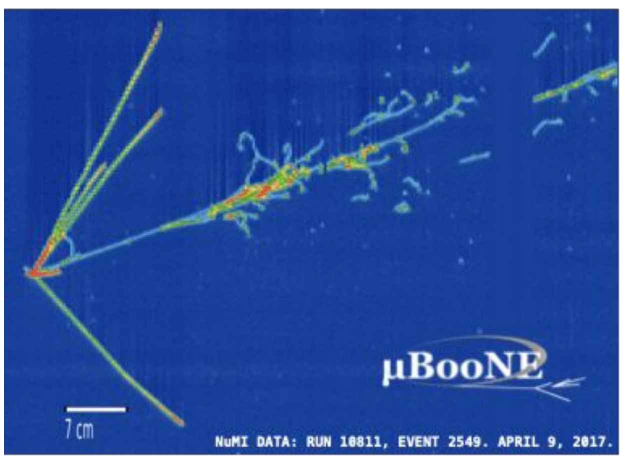
Collaborative Computing Needs for DUNE
Heidi Schellman, CHEP2019, https://indico.cern.ch/event/773049/contributions/3581360/

Data Processing Pipeline

Focus today on the UK Pandora pattern recognition step

LArTPC Event Reconstruction



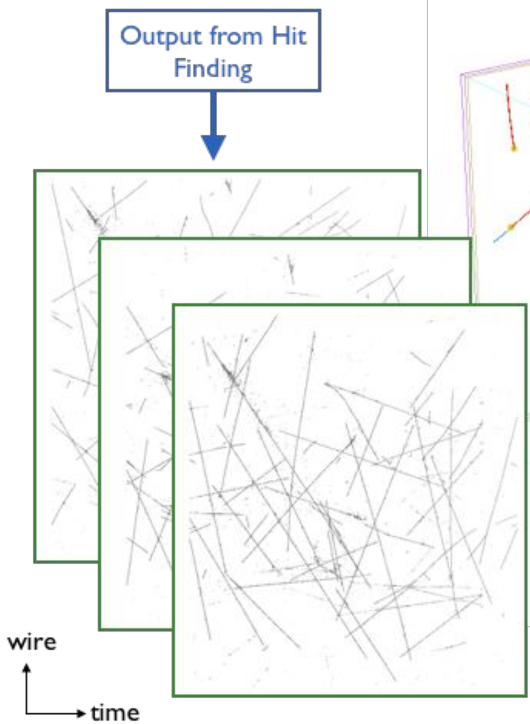


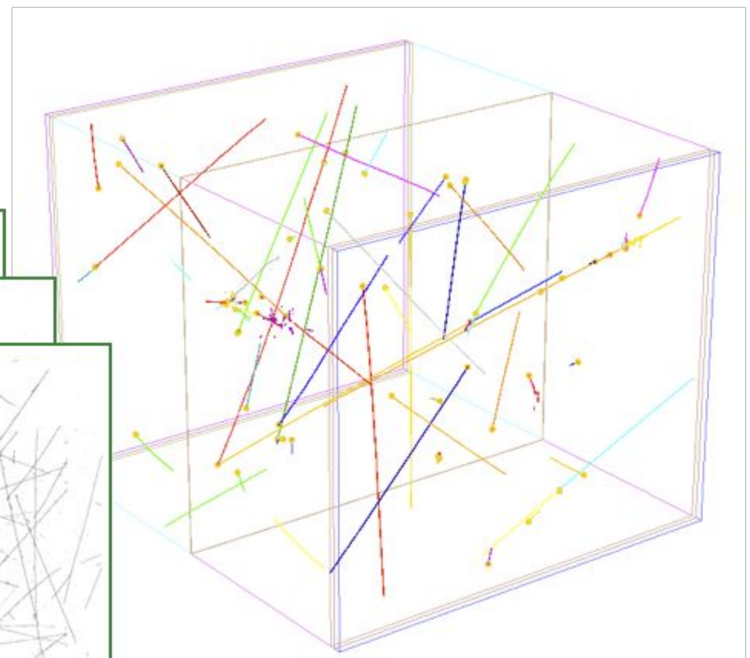
The conversion of raw LArTPC images into analysis-level physics quantities:

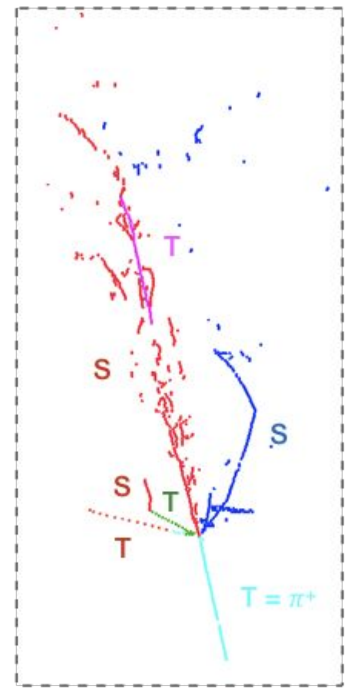
- **Low-level steps:**
  - Noise filtering
  - Signal processing
- **Pattern recognition:**
  - The bit you do by eye!
  - Turn images into sparse 2D hits
  - Assign 2D hits to clusters
  - Match features between planes
  - Output a hierarchy of 3D particles
- **High-level characterisation:**
  - Particle identification
  - Neutrino flavour and interaction type
  - Neutrino energy, etc…

LArTPC Pattern Recognition

- The main aims of the Pattern Recognition step are to:
  - Produce 3D reconstructed particles, based on inputs of 3 x 2D images.
  - Reconstruct the hierarchy of particles resulting from an interaction.







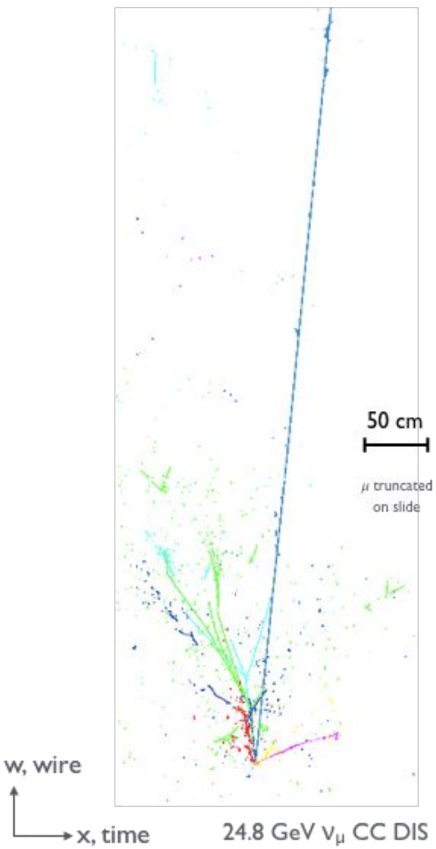
Simulated π^+ Pandora Reconstruction at ProtoDUNE-SP

Spot the event and its vertex!
Few clues: homogeneous detector volume

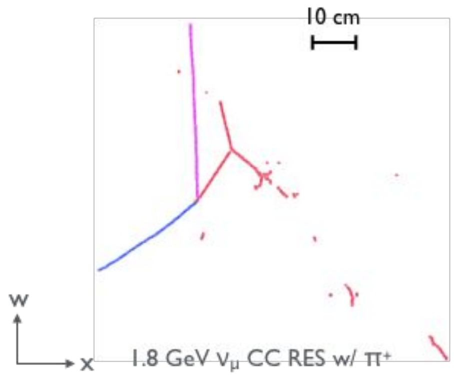
Challenges for Pattern Recognition

It is a significant challenge to develop automated, algorithmic LArTPC pattern recognition

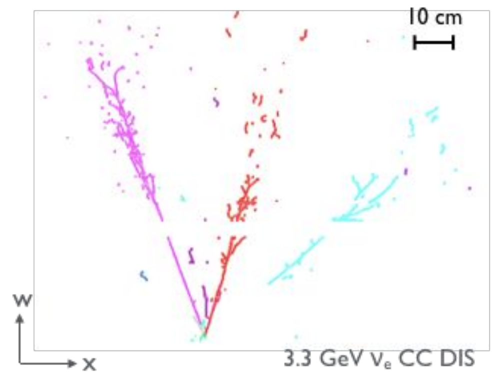
- Complex, diverse topologies:



24.8 GeV ν_μ CC DIS

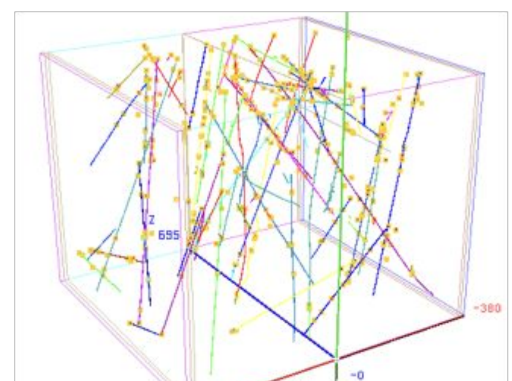


1.8 GeV ν_μ CC RES w/ π^+



3.3 GeV ν_e CC DIS

- Also, LArTPCs have long exposures, due to lengthy drift times (up to few ms).
- Significant cosmic-ray muon background in surface-based detectors.

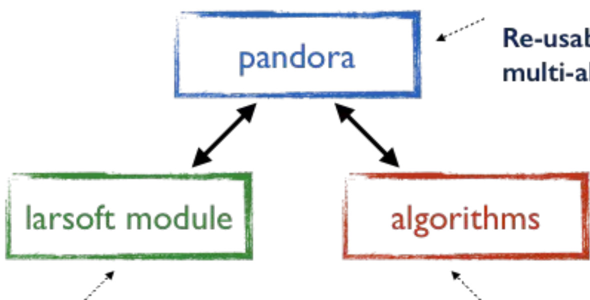

ProtoDUNE-SP

Multi-Algorithm Approach

- Single clustering approach is unlikely to work for such complex topologies:
  - Mix of track-like and shower-like clusters
- Use Pandora multi-algorithm approach to build up events gradually:
  - Each step is incremental - aim not to make mistakes (undoing mistakes is hard…)
  - Deploy more sophisticated algorithms as picture of event develops
  - Build physics and detector knowledge into algorithms

pandora sdk and visualisation

Re-usable libraries support multi-algorithm approach



larpandora

Handles input/output LArSoft↔Pandora

larpandoracontent

137 algorithms and tools, specifically for LArTPC usage

EPJC (2015) 75: 439
EPJC (2018) 78: 82
github.com/PandoraPFA

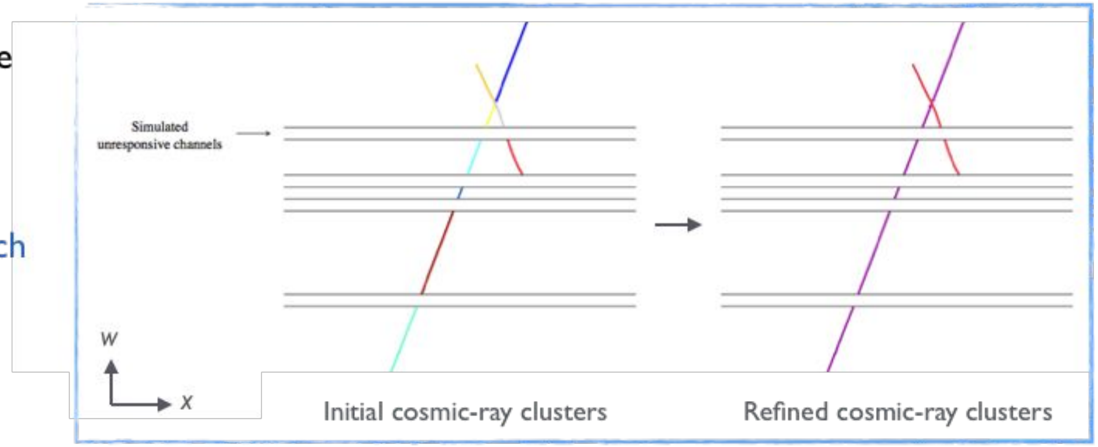
Algorithm 1 Cluster creation pseudocode. The logic determining when to create new Clusters and when to extend existing Clusters will vary between algorithms.

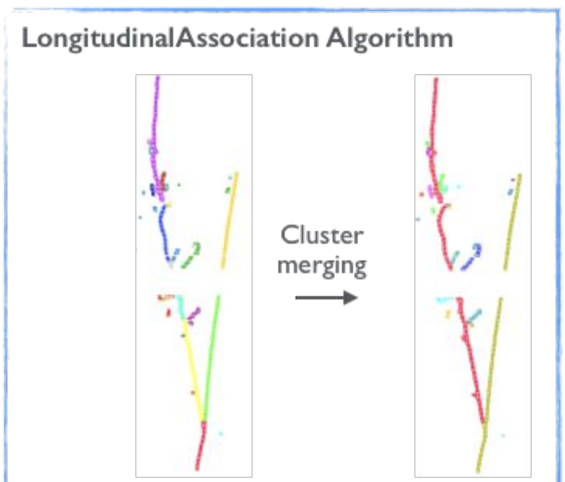
```
procedure CLUSTER CREATION
    Create temporary Cluster list
    Get current CaloHit list
    for all CaloHits do
        if CaloHit available then
            for all newly-created Clusters do
                Find best host Cluster
            if Suitable host Cluster found then
                Add CaloHit to host Cluster
            else
                Add CaloHit to a new Cluster
    Save new Clusters in a named list
```

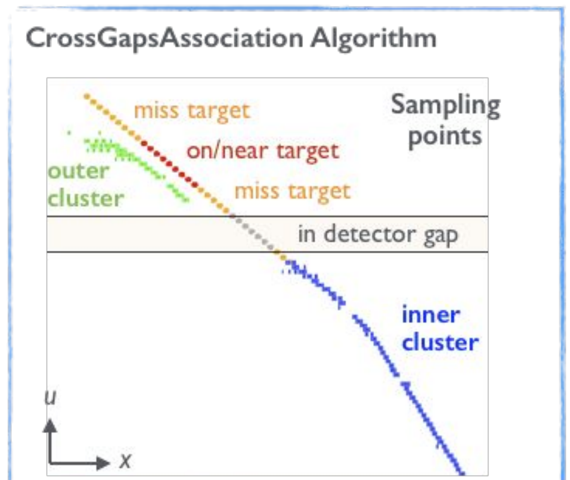
Simplified **algorithm** implementation

Algorithms: “Traditional”

- For each wire plane, create a list of 2D clusters that represent continuous, unambiguous lines of hits:
  - Separate clusters for each structure, with clusters starting/stopping at any branch or ambiguity.



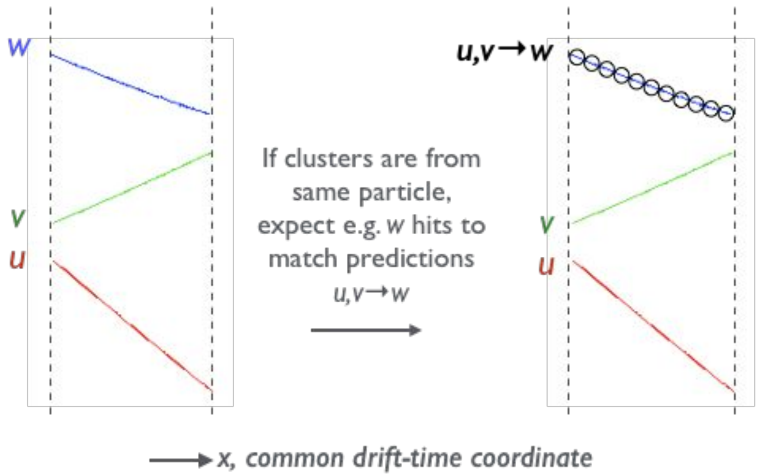




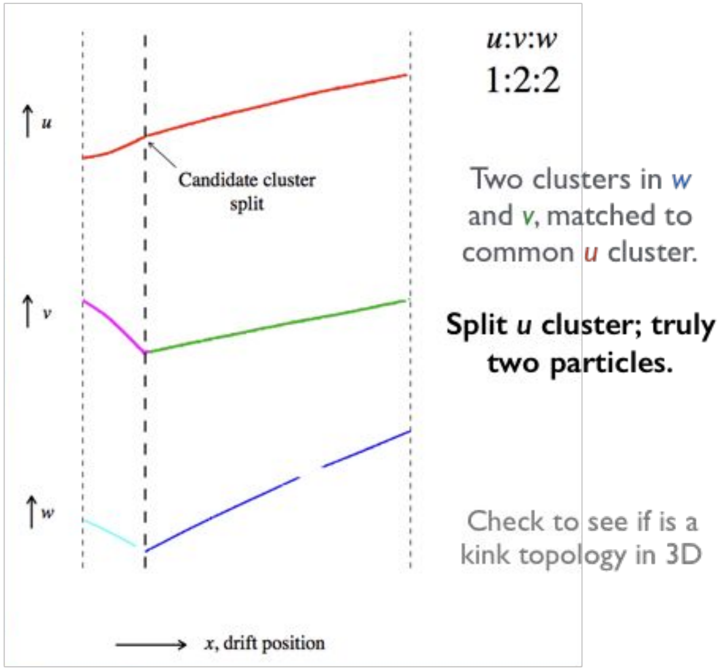
Initial clusters are refined by a series of cluster-merging and cluster-splitting algorithms that use topological info.

Algorithms: Detector Physics

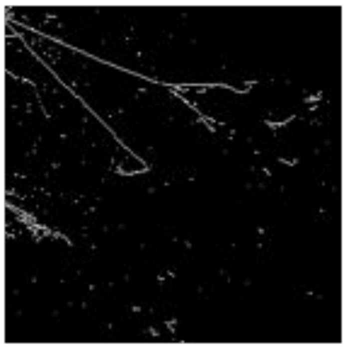
- Our original input was 3x2D images of charged particles in the detector.
- Should now have reconstructed three separate 2D clusters for each particle:
  - Compare 2D clusters from *u*, *v*, *w* planes to find the clusters representing same particle.
  - Exploit common drift-time coordinate and our understanding of wire plane geometry.

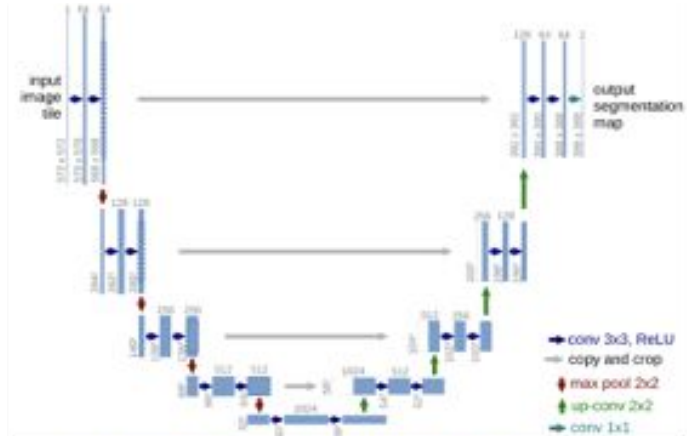


- Approach really comes to life when the 2D clustering "disagrees" between wire planes:
  - Automated detection of 2D PatRec issues, with treatment for specific cases, e.g.:

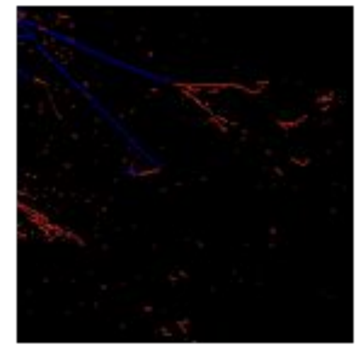


Algorithms: Deep Learning

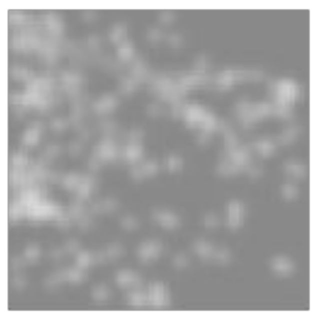


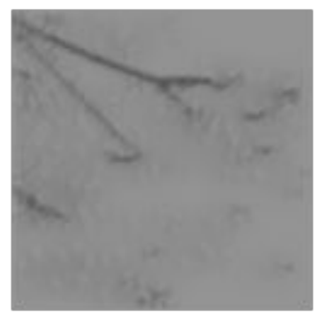


https://arxiv.org/abs/1505.04597



- U-Net down-samples on the left, up-sample on the right
- Skip connections to fill out up-sampling phase





- Convolution filters identify features
- Track/Shower probability map constructed

"Semantic Segmentation": DL methods can use local context to classify hits

Deep Learning Applications

- Feature extraction
  - *Hit classification "semantic segmentation"*
    - *Also for hit finding*
  - *Vertex Identification*
- Particle Identification
- Pandora's multi-algorithm approach provides the framework for developing and refining a mix of traditional ,detector, and deep learning
  - *Easy to split initial singular algorithms into multiple*
  - *Focus algorithms on specific problems, easy to swap in/out new solutions, like deep learning*
  - *Easy to implement reclustering and iteration*

Summary

- DUNE's LArTPC detectors present unique challenges for event reconstruction and physics extraction
  - *Large event sizes…*
  - *… but not ultimately limited by CPU/storage resources*
  - ***Homogeneous, highly detailed data...***
  - ***… requiring development of suitable computational methods to extract physics and realize potential***
- Pandora's multi-algorithm approach provides framework for blending different techniques
  - *Traditional/Detector based*
  - *Deep Learning*
  - *Flexibility to add future developments*