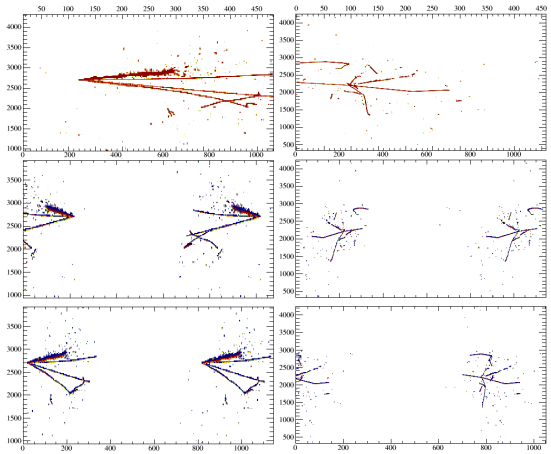
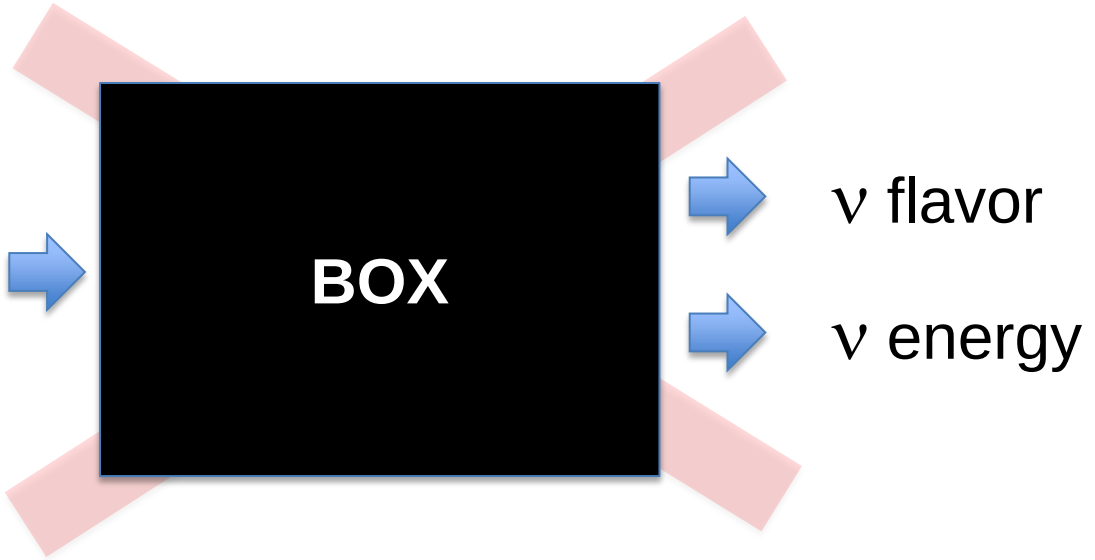
CNN based reconstruction status

D. Stefan, P. Płoński, R. Sulej

The approach (1)





Still a long way to proceed in this way:

- huge effort to ensure understanding of all dependencies (ν energy, noise, …)
- not easy to validate now on data for DUNE
- although this may be likely the final solution

Do partial tasks now:

➔ can be validated on ProtoDUNE (and LArIAT data first) and provide physics results

➔ understand tool properties before applying to complex task

The approach (2)

➔ to be combined with „standard” reconstruction products

➔ but **DO NOT rely** on inputs from the „standard” reconstruction

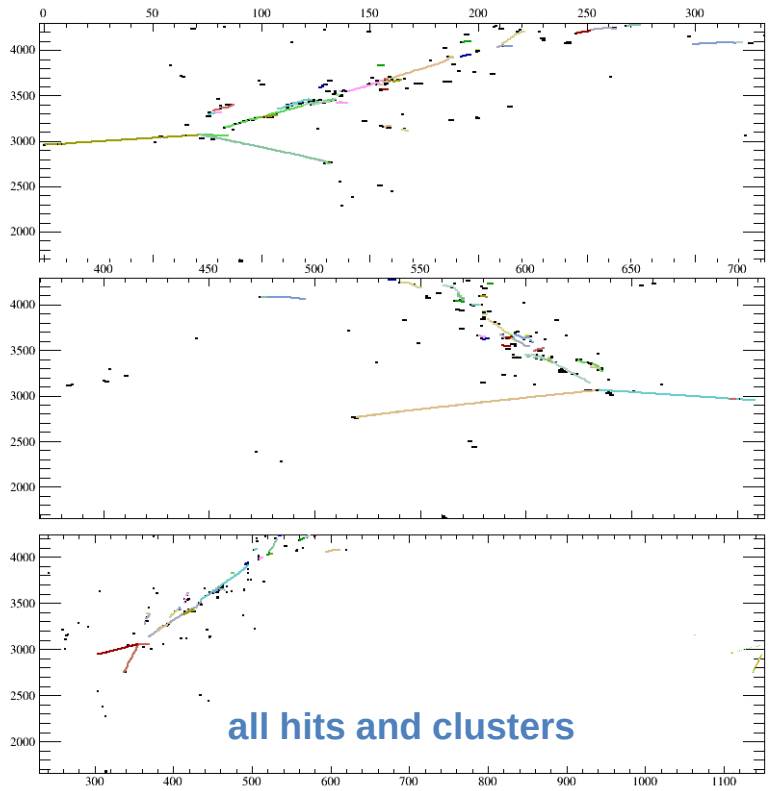
Input: deconvoluted ADC: „2D image”

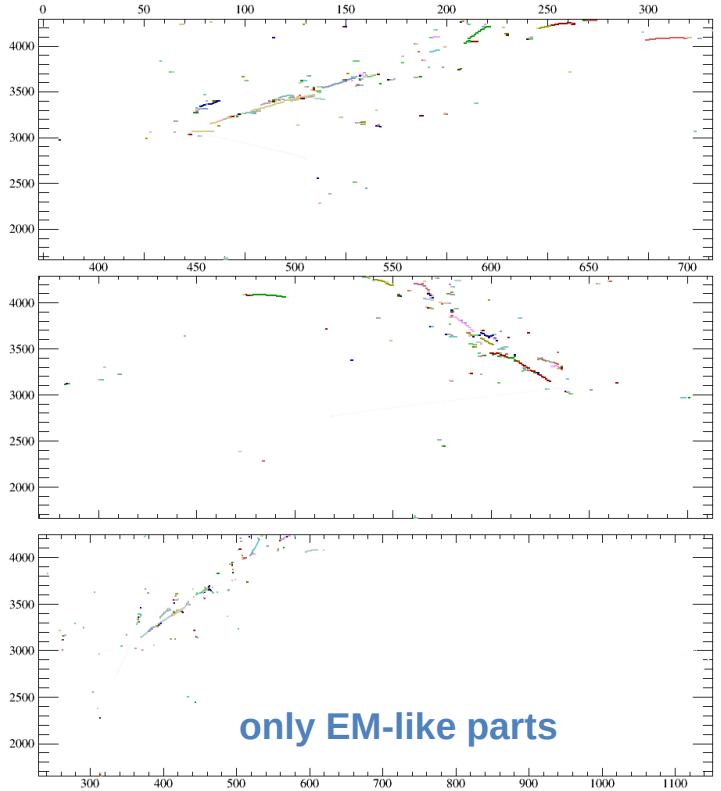
- full projection could be processes, limit to pixels where hits were found is needed to speed up
- look at small patch in high resolution (anyway drift downsampled to ~5mm)
- obviously should use broader context in much lower resolution (fascinating work, but need to go step by step)
- in EM / track separation „broader context” partially substituted with „standard” 2D clustering

Outputs targeted now (most of what we need from CNN for ProtoDUNE):

- EM / track component separation: *well advanced*
- decay vertex identification (before searching for all missed interactions – somewhat test-beam-goals specific application): *ongoing*
- vertex with π^0 identification (for rejection): *next to start work*

EM / track: ready to use for ProtoDUNE-like events/applications





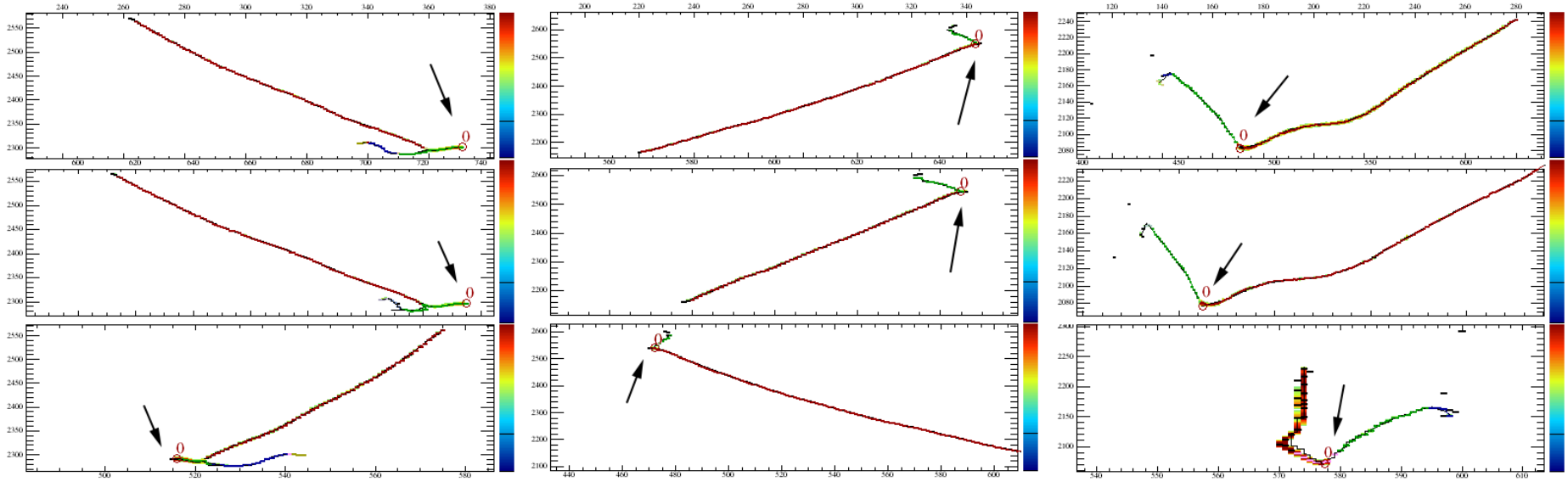
- first model done: >96% correct selections on ProtoDUNE 2GeV/c beam

- model for higher resolution in drift, tests with noise
- model for higher energies (more crowded showers)
- model for ν_e (crowded showers as well)

} now stalled, lack of computing resources

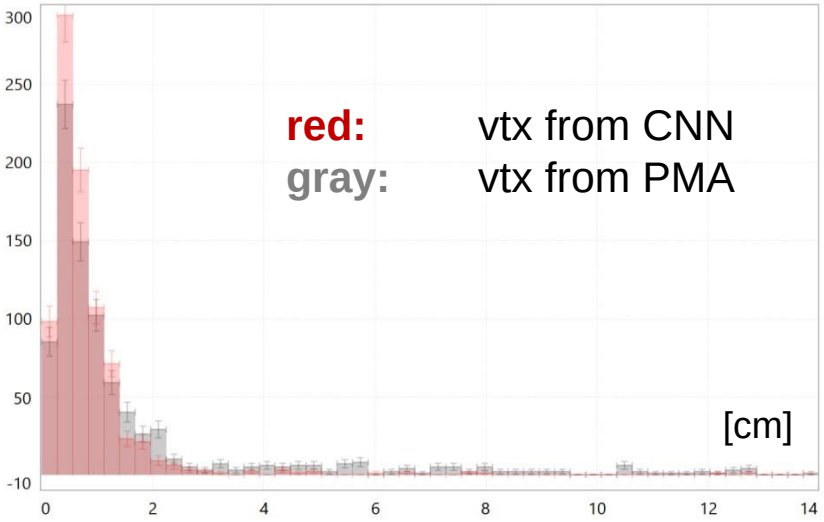
- search for optimal models → see Piotr's slides

Decay vertex: a basic excersise



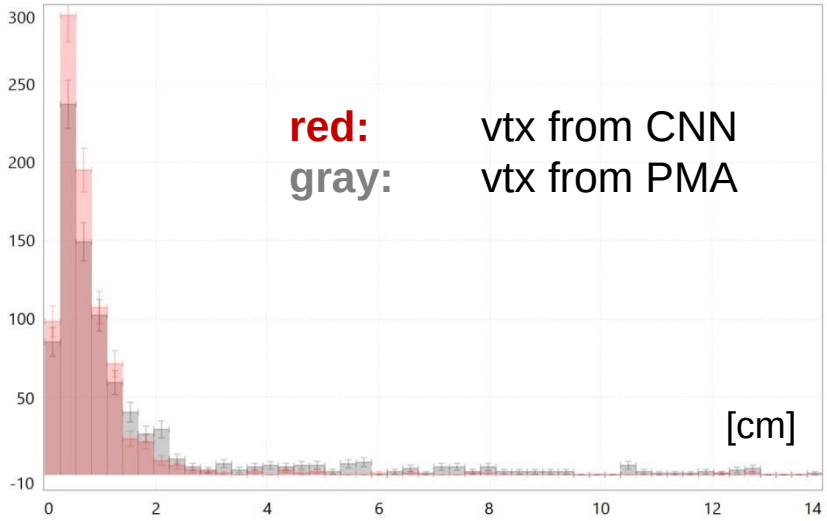
vertex selected using CNN outputs combined from all views

- LArSoft module in place, now inventing best way to use CNN output for vertex creation
- does the right thing: tails in the reco vertex distance to MC truth decay position lower but need to find the most efficient approach of using CNN output

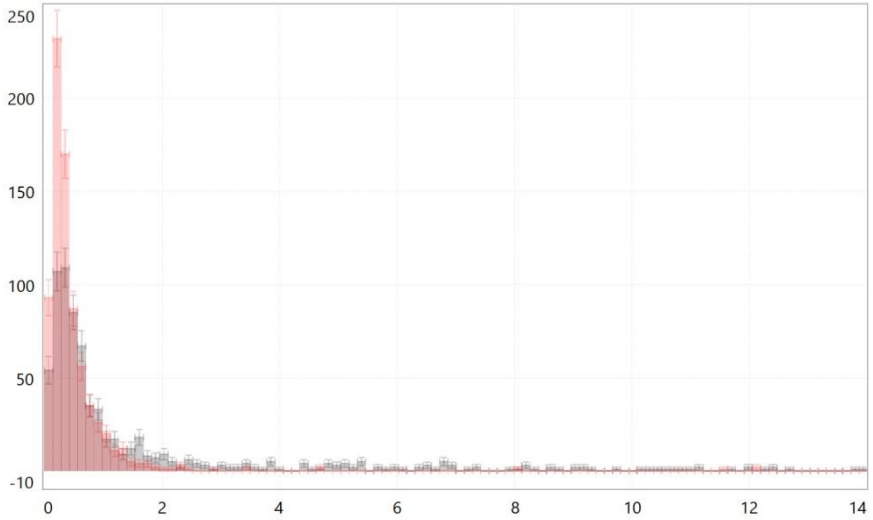


Decay vertex: a basic excersise

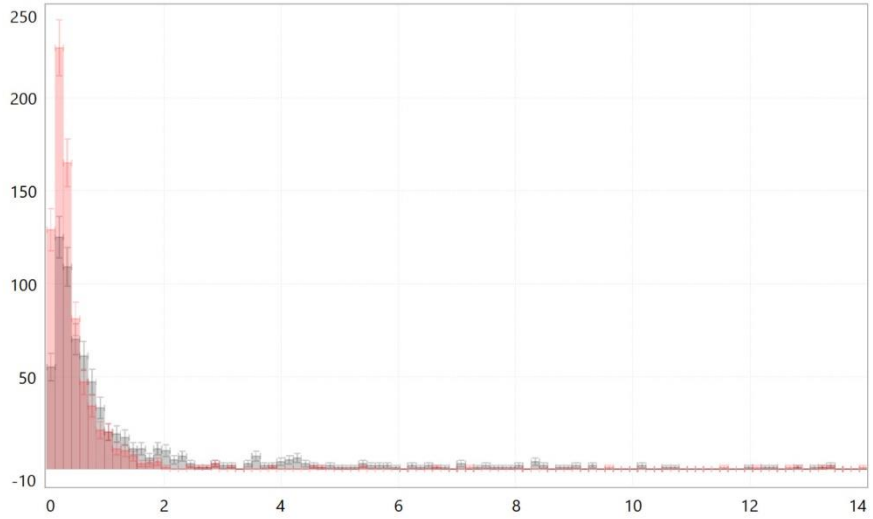
3D



view 0 (2D)



view 1 (2D)

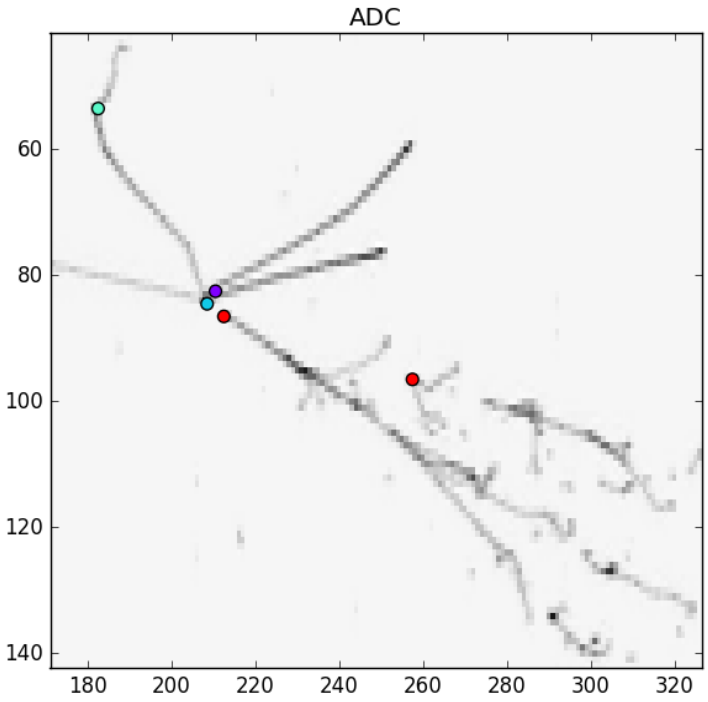


Difference in 2D larger than in 3D

→ need to improve logic and not to spoil what CNN done well

Vertex with π^0

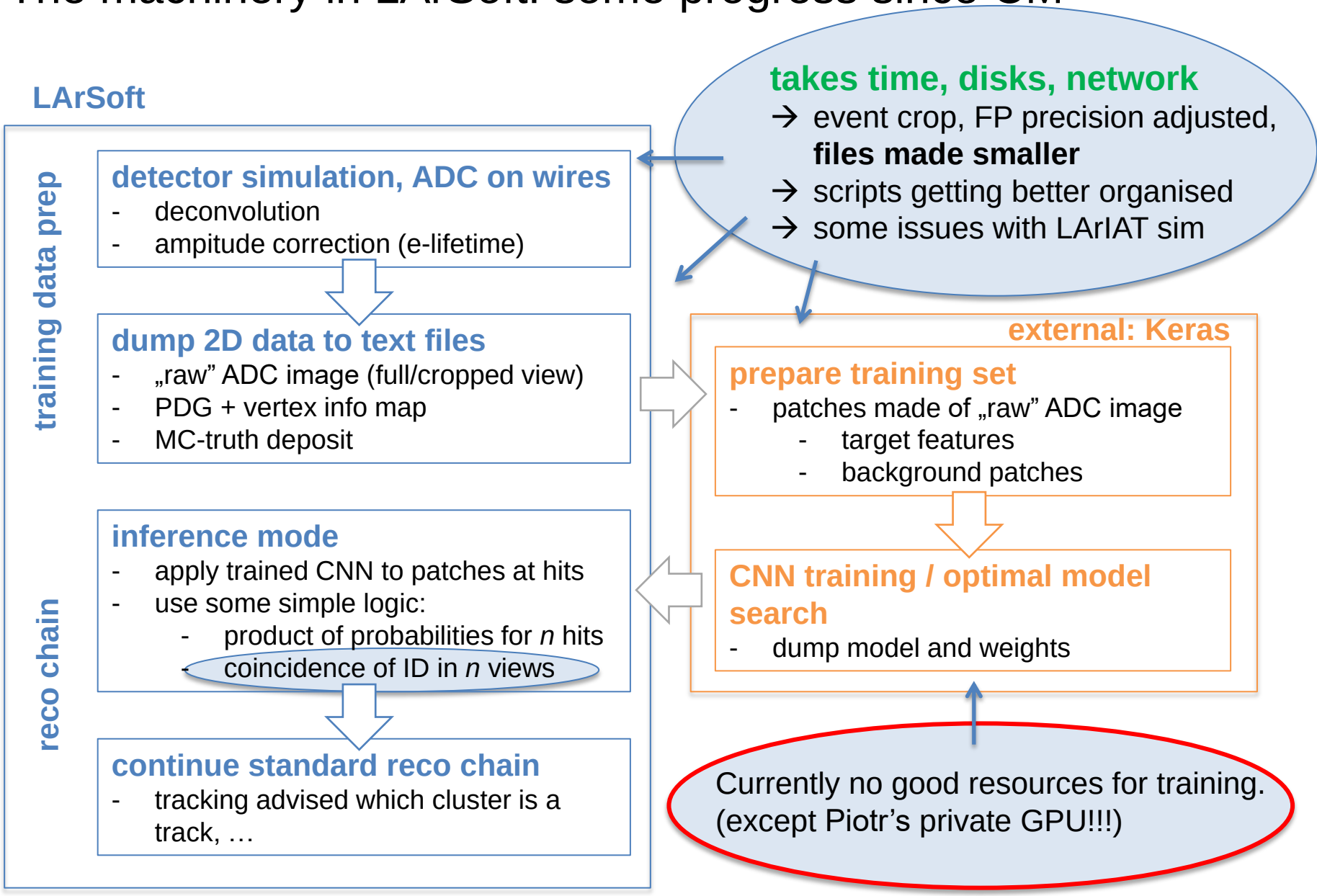
- not started, only the research idea
- two possible training targets:
  - tag interaction vertex which is acompanied with EM showers
  - tag gamma conversion points
- Aim:
  - select beam particle interactions for π^0 rejection study
  - support π^0 reconstruction
- **practice on this task to understand ν_e selection**



Note on vertex ID's:

Only a few training vectors/patches per event ➔ much larger MC samples needed than for EM/track separation (>100 patches per event)

The machinery in LArSoft: some progress since CM



Copying data: ~50-100 GB each: root files, dumps, training sets

I was naively thinking that scp will always work (like for a TB copied between CERN and UTA)… but:

- dCache at FNAL (where /pnfs/dune is) has file system not compatible with scp: you may be lucky or file will be broken
- one should use dcat or xrootd protocols, see howto:
  http://cd-docdb.fnal.gov/cgi-bin/RetrieveFile?docid=5399&filename=PublicdCacheHowTo.pdf&version=4
- at CERN a grid user certificate is needed:
  https://ca.cern.ch/ca/ (see Help there for instructions how to use it with voms-proxy-init)
- and it has to be recognized by FNAL (mapped to your account in GUMS)
  Tom Junk can add you to DUNE VO if needed and may apply for mapping in GUMS
- Scripts handling these mapping had some problems, investigated by FNAL computing support and after a week of trying:

[rosulej@lxplus045 mu_dcy]$ xrdcp cnn_test.nnet
xroot://fndca1.fnal.gov/pnfs/fnal.gov/usr/dune/persistent/users/rnd/cnn/models/
[103.5MB/103.5MB][100%][==========][11.5MB/s]

Red part shows how to form a path to pnfs area on dCache, the rest is standard: persistent, scratch, and your folder there…

- Progressing on the best approach for decay tagging
- More organized tools for data preparation
- Major problem with no stable place for doing the actual research
  - going back to TeV cluster at FNAL (known to be oversubscribed)
  - looking for options at CERN
  - will apply for external resources as well
- **The machinery for model optimization – see Piotr's slides**