Advanced Particle Flow

- Development of advanced particle flow and pattern recognition algorithms in PandoraPFA
- Application to LHC, LC and neutrino experiments

L. Escudero, University of Cambridge, 6 April 2017

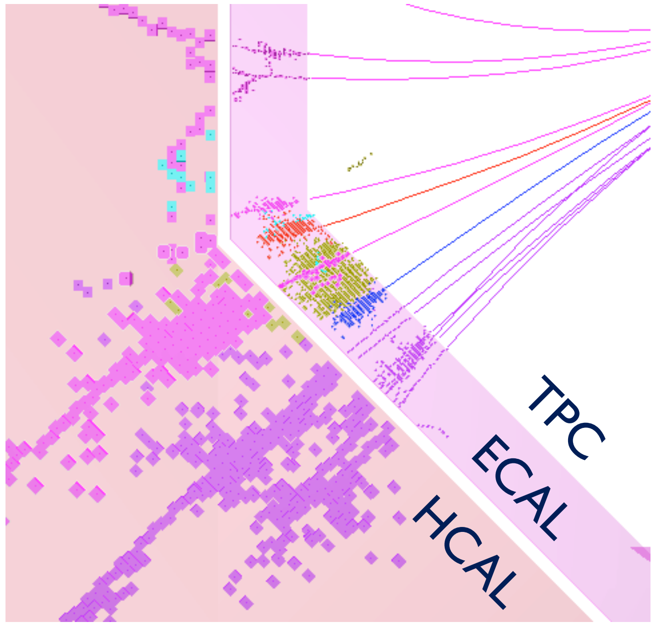
Overview

1. Pandora Particle Flow
    1. Introduction
    2. Multi-algorithm approach
    3. Visualisation
2. Cambridge: LAr TPC
    1. Pandora MicroBooNE
    2. Pandora DUNE
3. Cambridge: LC
4. CERN: DDMarlinPandora and CLIC detector model
5. CNRS-LLR and CNRS-IPNL: Arbor

Pandora Particle Flow

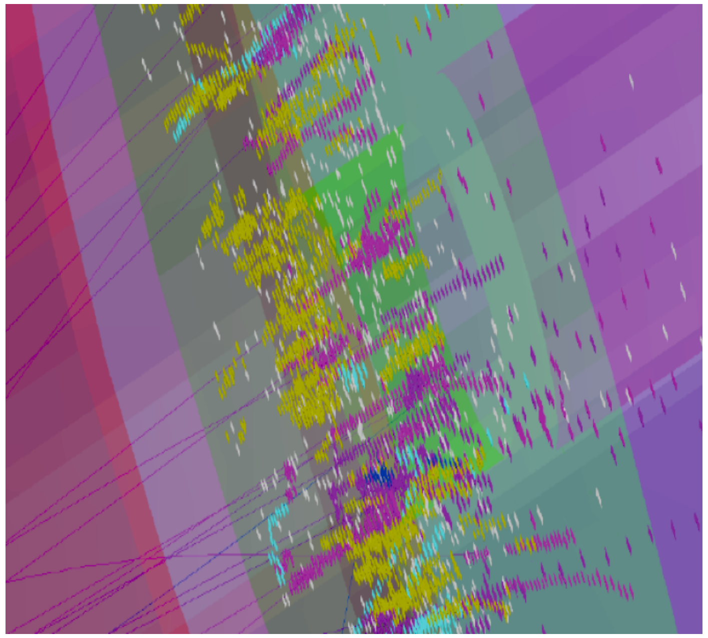
- **Continued support of Pandora Software Development Kit:** aids multi-algorithm approach to pattern recognition, with advanced reclustering and recursion abilities and visualisation.
- Development of new client applications, enabling use of algorithms for different detector concepts and in different software frameworks.
- Development of pattern recognition for both LC (inc. LHC upgrade) and LAr TPC. Continued validation and exploitation of existing algorithms e.g. via detector optimisation studies.

NIMA.2009.09.009, NIMA.2012.10.038



arXiv:1307.7335, 1506.05348

LHCC-P-008

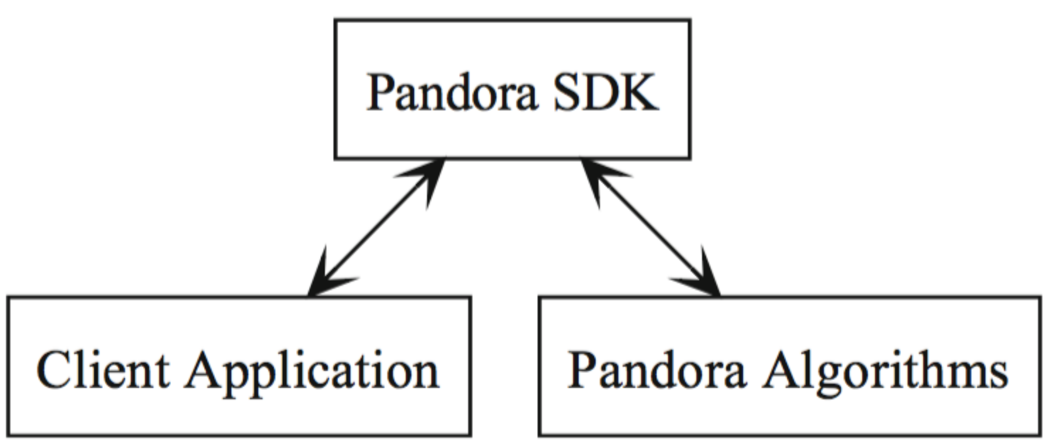


Pandora Introduction

A multi-algorithm event reconstruction can be difficult to implement. The **Pandora SDK** has been carefully engineered to provide a software environment in which:

1. It is easy for users to provide the building-blocks that define a pattern recognition problem.
2. Logic required to solve pattern recognition problems is cleanly implemented in algorithms.
3. Operations to access or modify building-blocks, or to create new structures, are requested by algorithms and performed by the Pandora framework.

It actively promotes use of large numbers of algorithms, carefully building-up a picture of an event.



https://www.github.com/PandoraPFA

EPJC.75.439

The Pandora Software Development Kit for Pattern Recognition

J. S. Marshall[a,*], M. A. Thomson[a]

[a] *Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom*

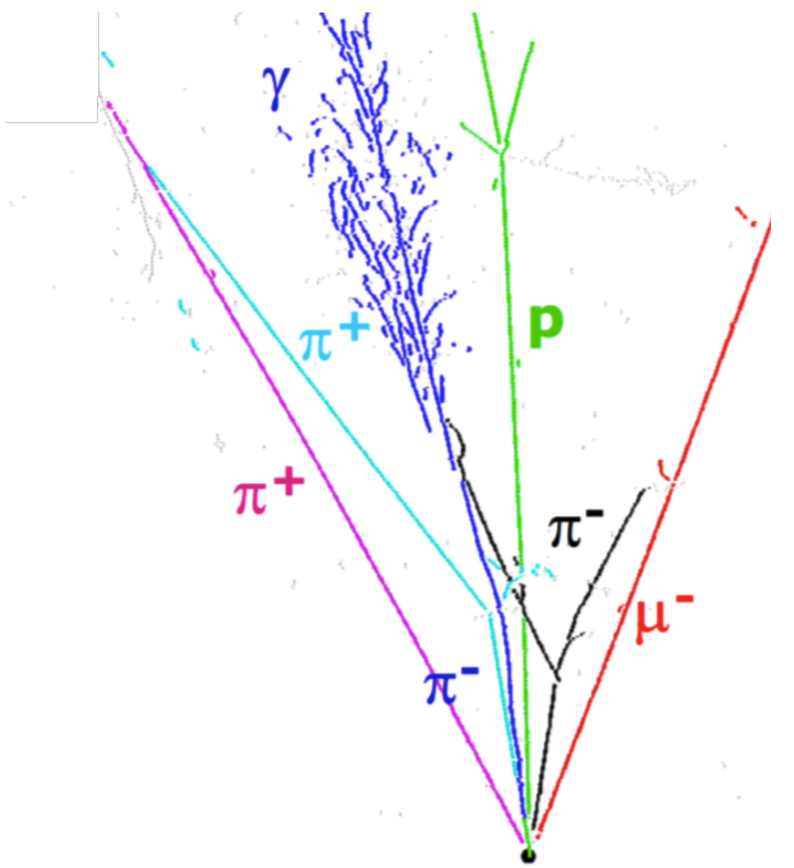
Abstract

The development of automated solutions to pattern recognition problems is important in many areas of scientific research and human endeavour. This paper describes the implementation of the Pandora Software Development Kit, which aids the process of designing, implementing and running pattern recognition algorithms. The Pandora Application Programming Interfaces ensure simple specification of the building-blocks defining a pattern recognition problem. The logic required to solve the problem is implemented in algorithms. The algorithms request operations to create or modify data structures and the operations are performed by the Pandora framework. This design promotes an approach using many decoupled algorithms, each addressing specific topologies. Details of algorithms addressing two pattern recognition problems in High Energy Physics are presented: reconstruction of events at a high-energy e^+e^- linear collider and reconstruction of cosmic ray or neutrino events in a liquid argon time projection chamber.

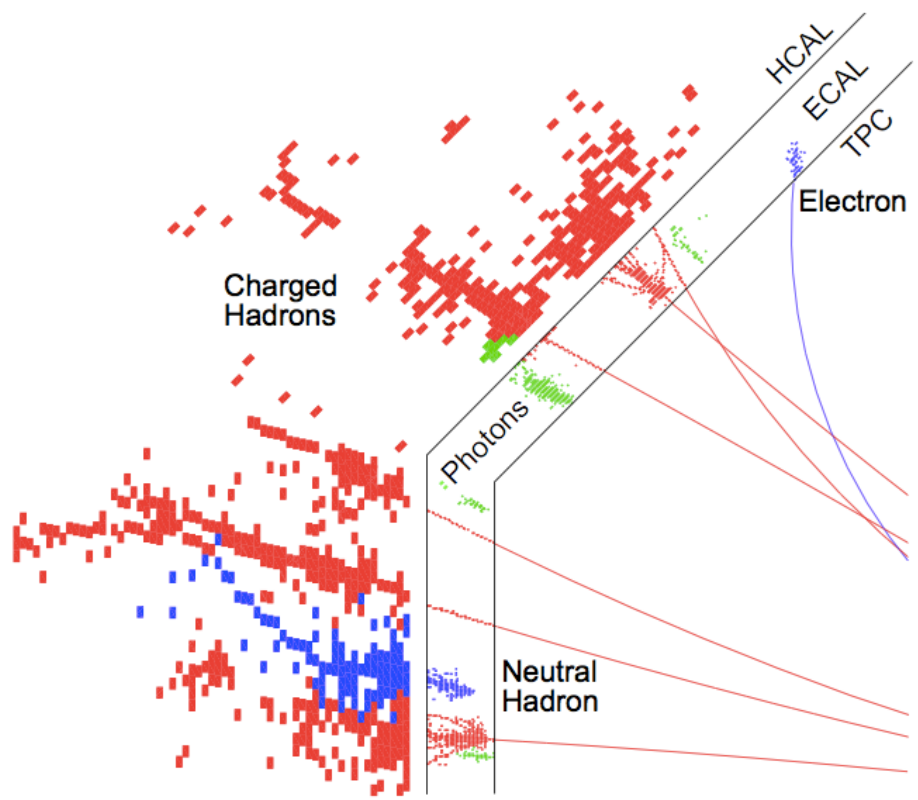
Keywords: Software Development Kit, Pattern recognition, High Energy Physics

Pandora Multi-algorithm

- Pandora algorithms provide the step-by-step instructions for finding patterns in the provided data.
- They use the APIs to access objects and to request the Pandora Managers to make new objects or modify existing objects.



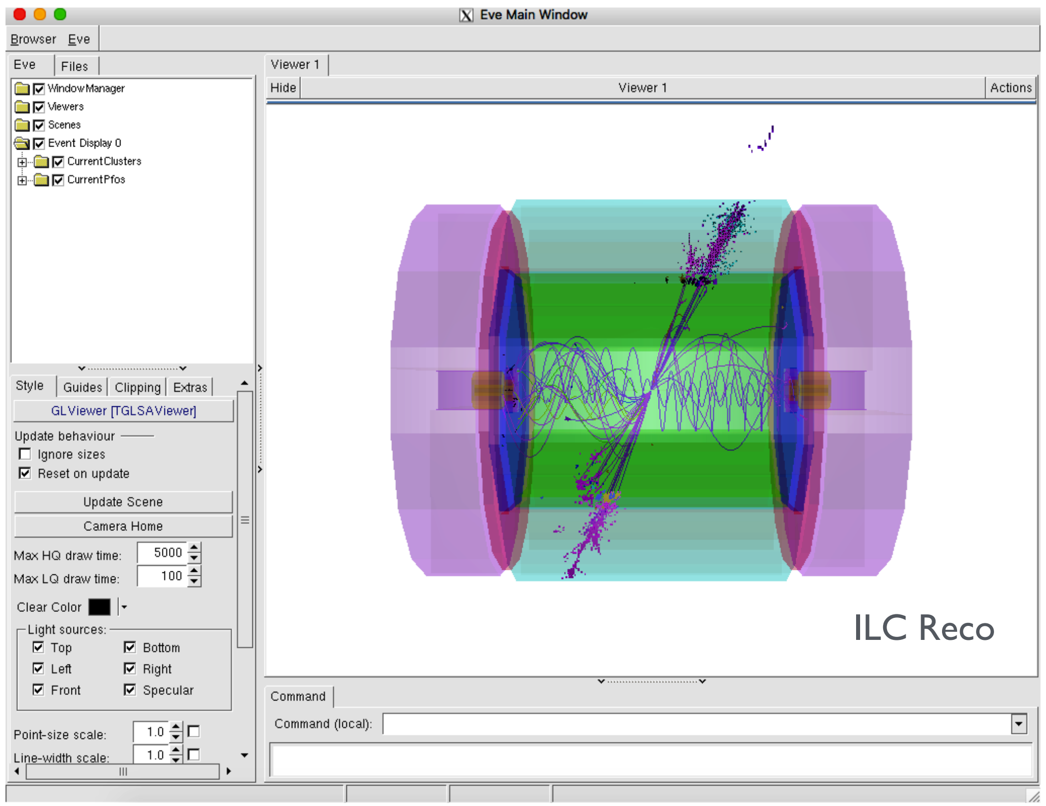
18 GeV ν_μ CC



LC example event

2 very different problems in HEP handled by the same Pandora framework, using a multi-algorithm approach with different algorithms

Pandora Visualisation

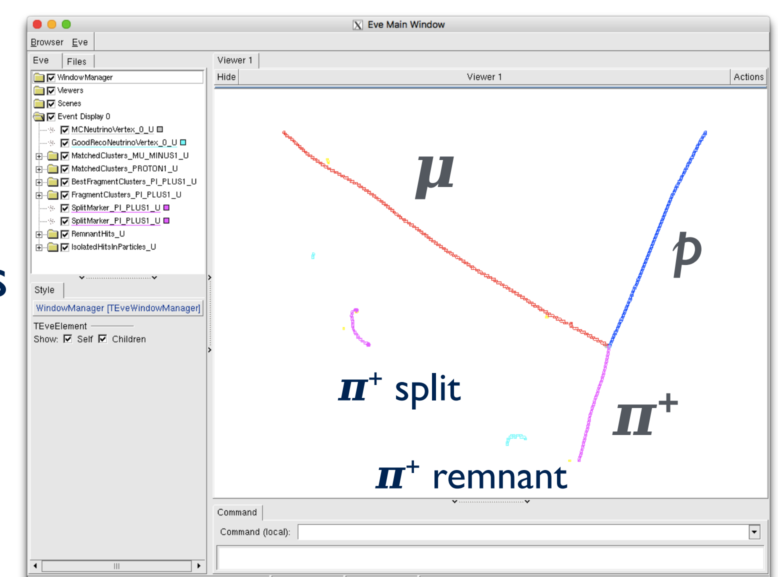



```
. . .
<algorithm type = "LoopingTracks"/>
<algorithm type = "VisualMonitoring">
    <ShowCurrentClusters>true</ShowCurrentClusters>
</algorithm>
<algorithm type = "BrokenTracks"/>
<algorithm type = "VisualMonitoring">
    <ShowCurrentClusters>true</ShowCurrentClusters>
</algorithm>
<algorithm type = "ShowerMipMerging"/>
<algorithm type = "ShowerMipMerging2"/>
<algorithm type = "BackscatteredTracks"/>
<algorithm type = "BackscatteredTracks2"/>
<algorithm type = "ShowerMipMerging3"/>
<algorithm type = "ShowerMipMerging4"/>
<algorithm type = "ProximityBasedMerging"/>
. . .
```

e.g. Add two event display algs to examine changes as reconstruction progresses

- Pandora algorithms can choose to use the Pandora Monitoring library, which has a ROOT dependency.
- Reusable Pandora event display alg can be added to XML to view status of reconstruction at any point.
- Alternatively, algs can use visualisation APIs to provide custom visual debugging - **rewarding way to work.**

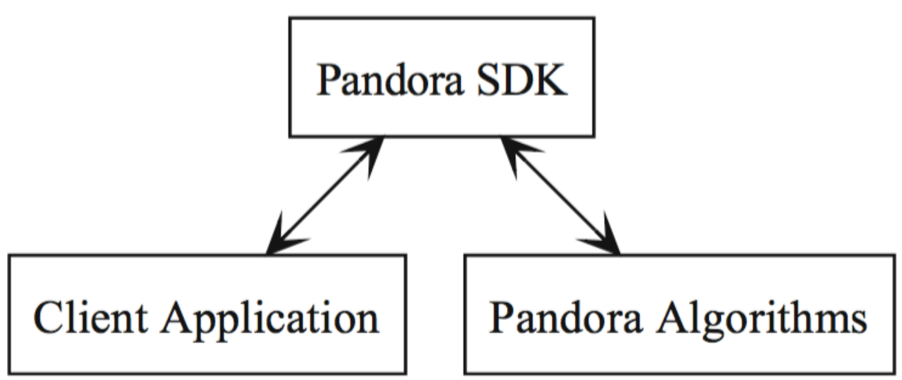
e.g. Colour-coded Particles and feature points



LAr TPC Reco

Cambridge: Pandora LAr TPC

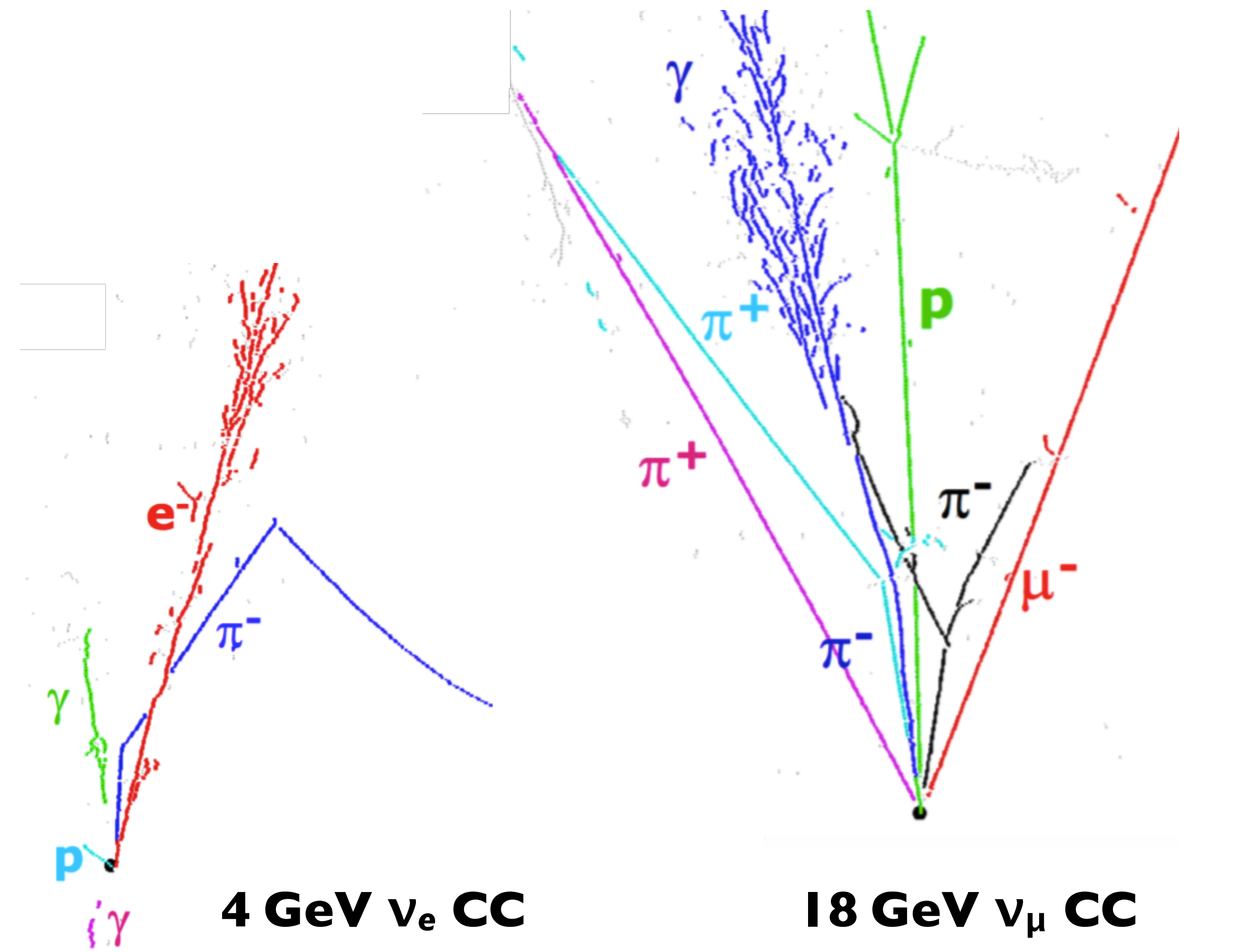
- **Support full breadth of the LAr TPC neutrino programme:**
- Range of neutrino energies:
  - Event topologies vary considerably
- Range of detector configurations:
  - Deal with multiple drift volumes
  - 2x2D views vs. 3x2D views
  - Presence of cosmic rays

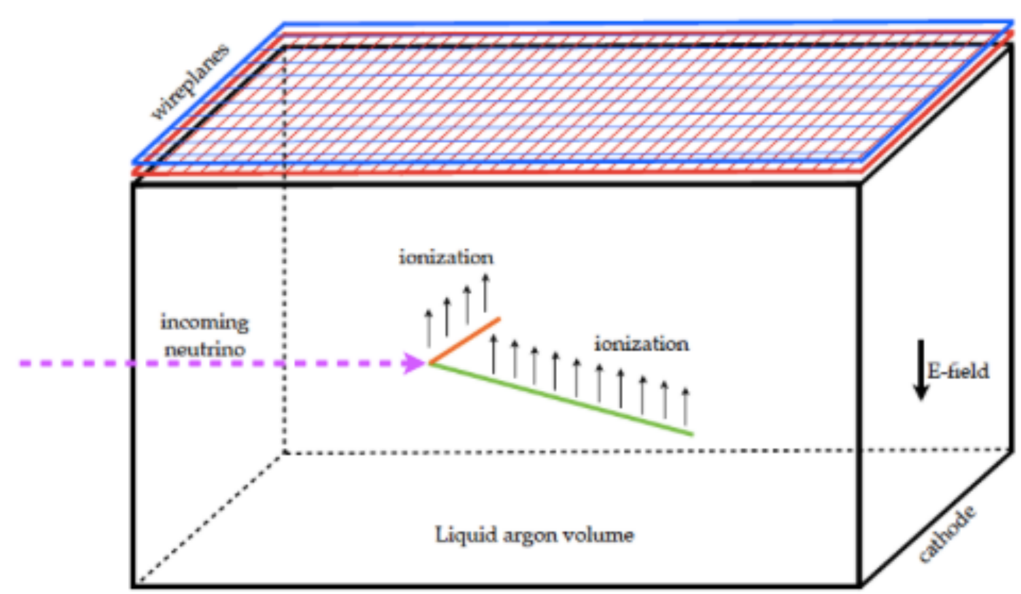


MicroBooNEPandora
DUNE35tPandora
DUNE4APAPandora
ProtoDUNEPandora
LArIATPandora
SBNDPandora

LArPandoraContent

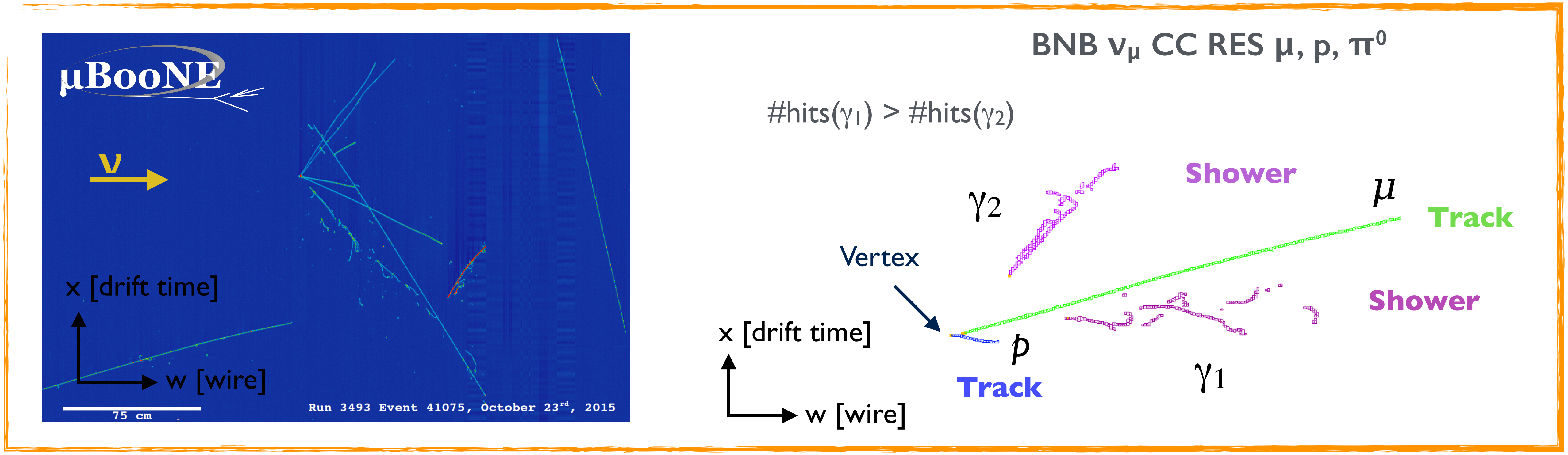
Currently supported



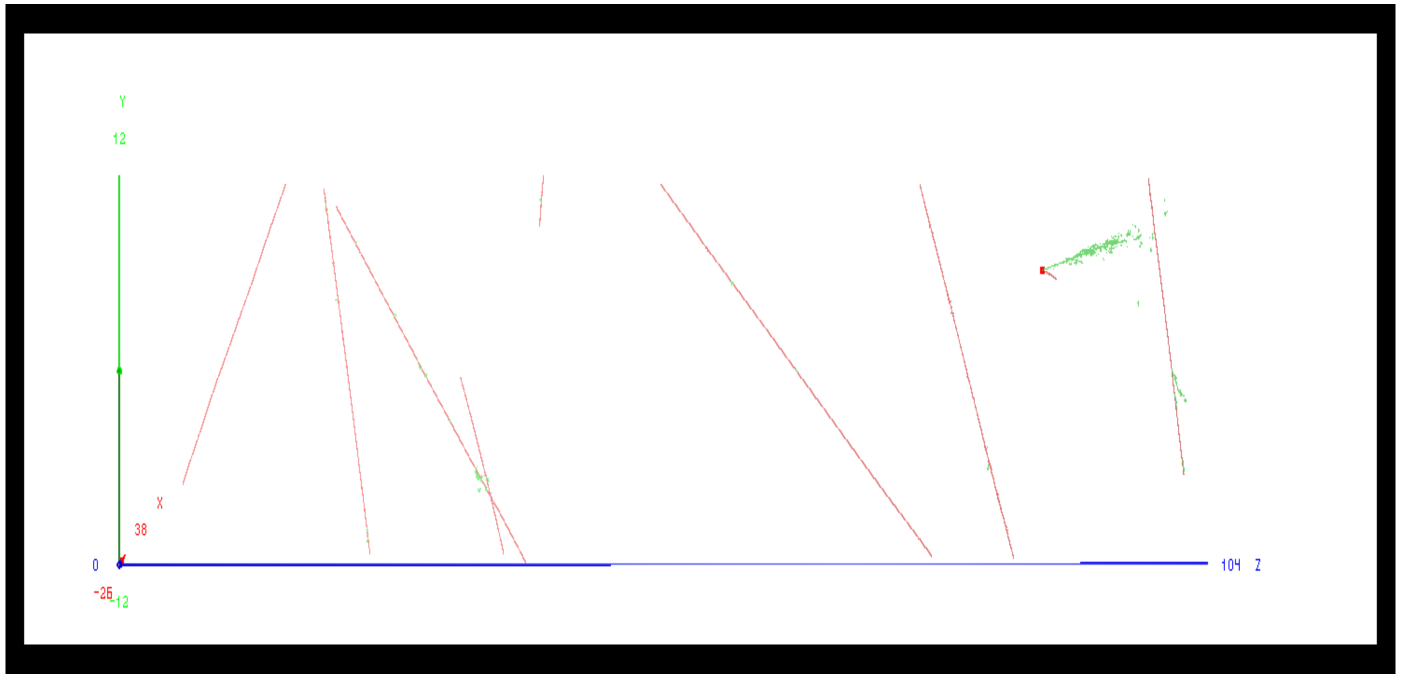


Picture: M. Soderberg

Cambridge: Pandora MicroBooNE



Challenge in MicroBooNE (surface detector): cosmic rays



2-steps chain to remove cosmic tracks:

- pandoraCosmic pass: optimized for cosmics, strongly track-oriented
- cosmic ray removal
- pandoraNu pass: optimised for neutrino reconstruction, identify neutrino interaction vertex

Cambridge: Pandora MicroBooNE

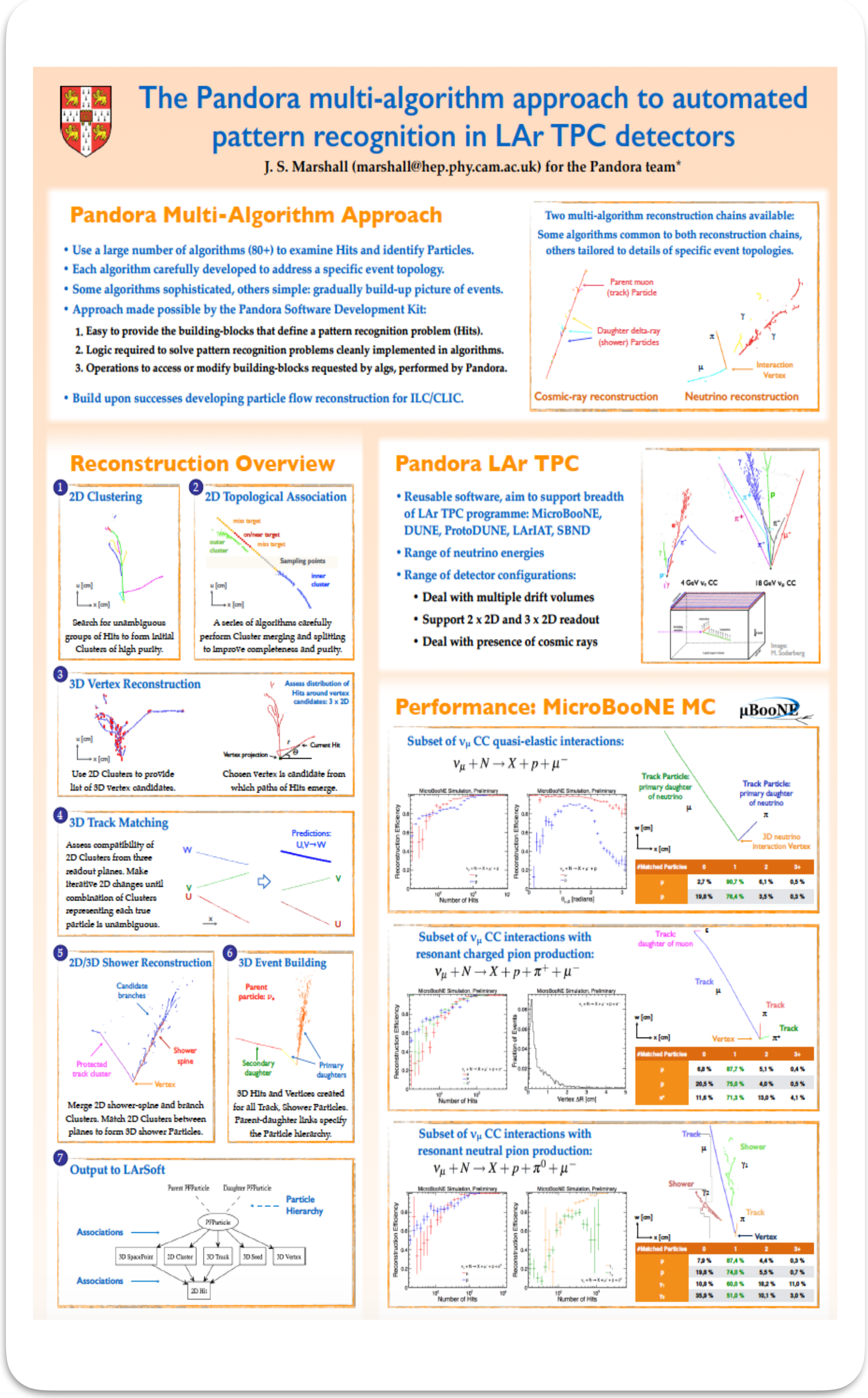
Pandora has been used for the official MicroBooNE summer analyses presented at Neutrino 2016

Public note available here

MicroBooNE Public Notes Page

Back to the Publications Page

- 7/4/16 MICROBOONE-NOTE-1019-PUB
  Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber
- 7/4/16 MICROBOONE-NOTE-1017-PUB
  A Method to Extract the Charge Distribution Arriving at the TPC Wire Planes in MicroBooNE
- 7/4/16 MICROBOONE-NOTE-1016-PUB
  Noise Characterization and Filtering in the MicroBooNE TPC
- 7/4/16 MICROBOONE-NOTE-1015-PUB
  The Pandora multi-algorithm approach to automated pattern recognition in LAr TPC detectors



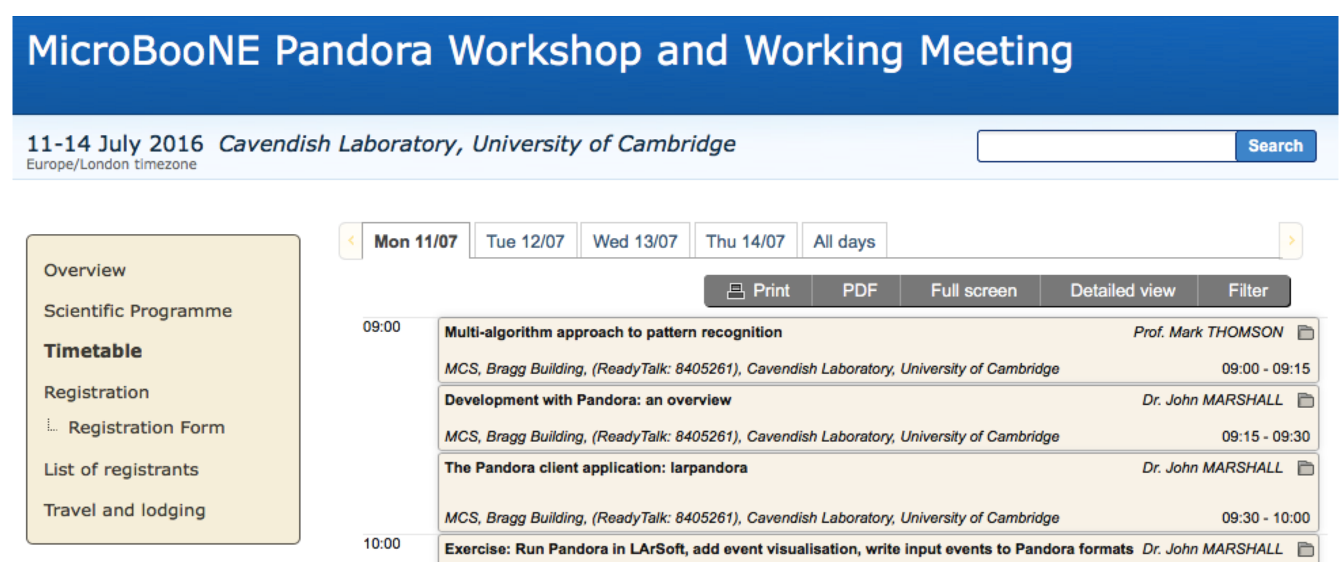
MicroBooNE paper in preparation! (with the Editorial Board now)

Cambridge: Pandora MicroBooNE

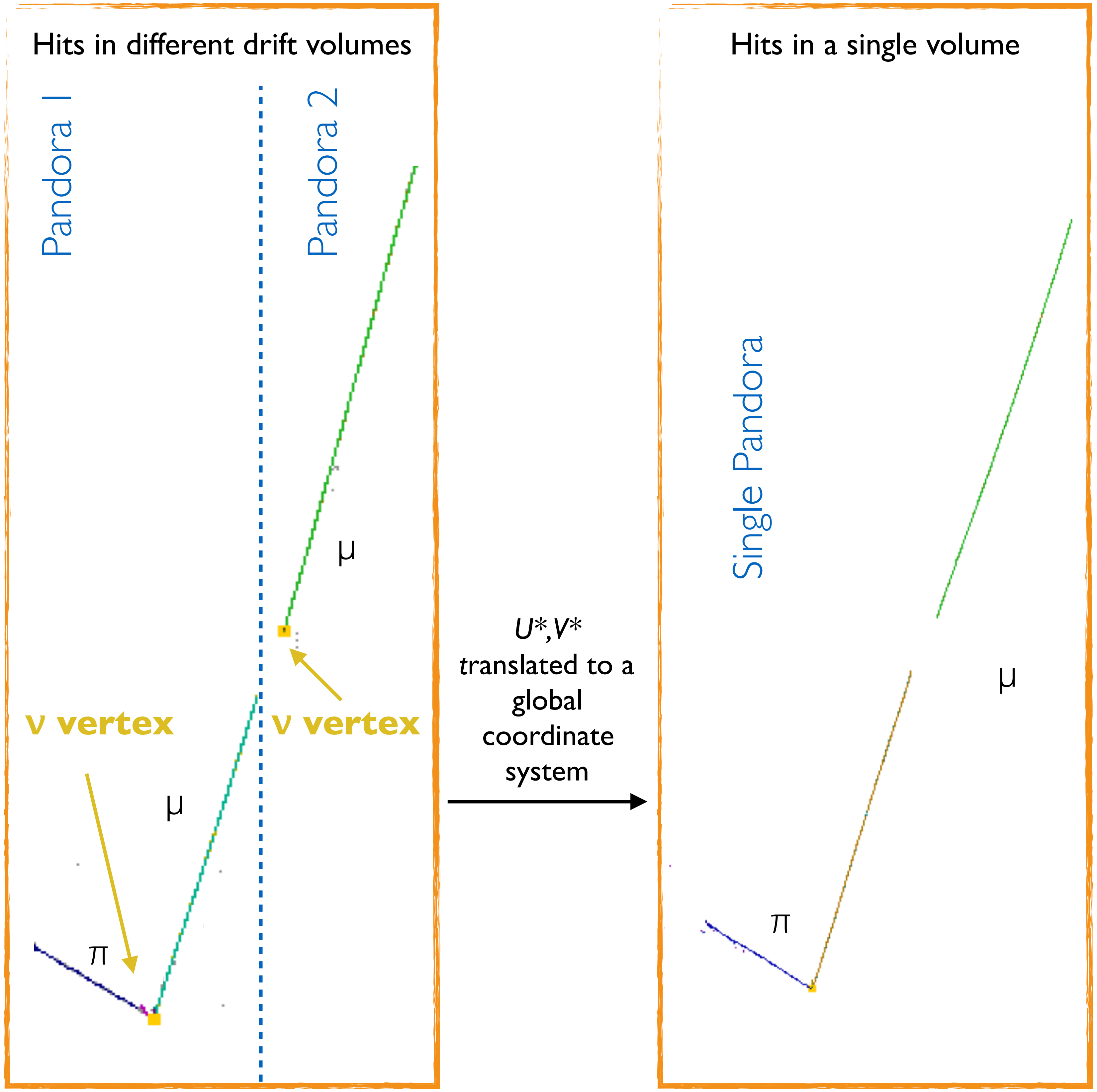
We also hosted the first Pandora Workshop and Working Meeting this summer in Cambridge, focused on MicroBooNE

- **Comprehensive description of all algorithms**
- **Exercises**
  1. Run Pandora in LArSoft
  2. Create and configure a new algorithm
  3. Add algorithm implementation for cluster creation
  4. Add algorithm implementation for cluster merging
  5. Particle creation, and more….

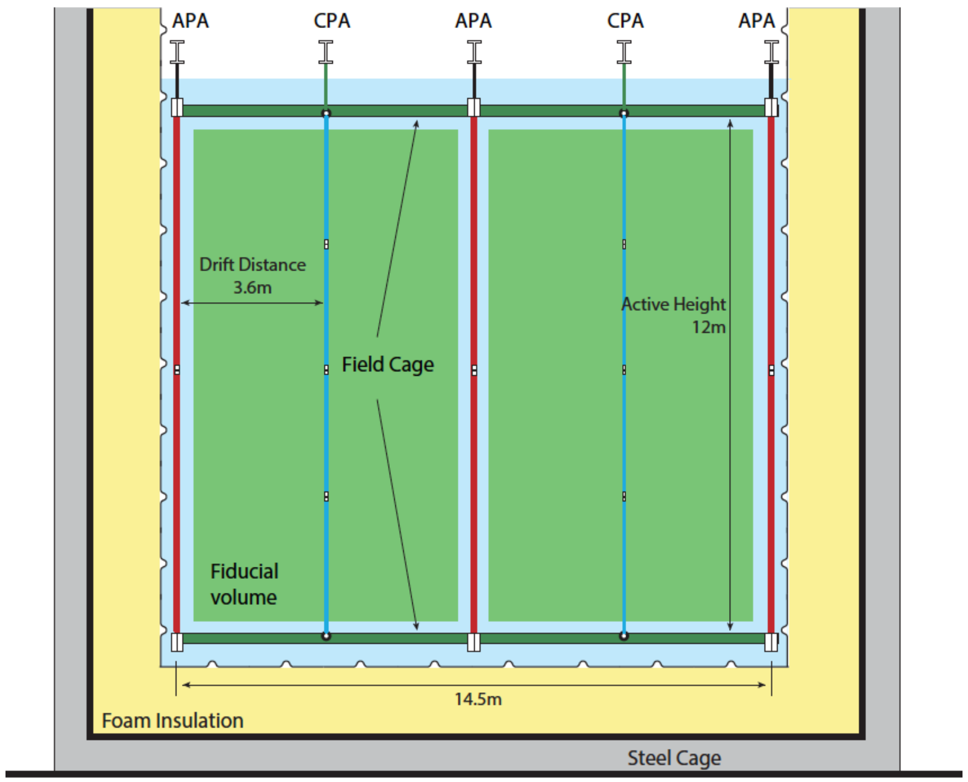
Pandora workshop material available from Indico or GitHub



Cambridge: Pandora DUNE



DUNE single phase 10kton module



Previously Pandora had to handle multiple "drift volumes" in the DUNE context. Now moved to use a single instance of Pandora, so a single neutrino interaction reconstructed (single neutrino vertex) and no need for stitching of PFOs from different volumes afterwards

Cambridge: Pandora DUNE

Interaction type | Final state particles

A selection of exclusive final states

CCQE NEvents Correct [%]	μ 9339 81,0	μ+p 12603 72,8	NCQE p 2378 73,8
CCRES NEvents Correct [%]	μ 792 82,3	μ+p 2452 71,0	NCRES p 1090 78,8
CCRES NEvents Correct [%]	$\mu+2\pi^C$ 167 44,3	$\mu+\pi^C+p$ 7814 58,4	NCRES π^0 2492 33,1
CCRES NEvents Correct [%]	$\mu+\pi^0$ 1362 36,6	$\mu+\pi^0+p$ 3373 30,3	NCRES π^0+p 1795 33,5
CCDIS NEvents Correct [%]	μ 397 83,4	μ+p 1054 62,8	NCDIS π^C 1287 40,5
CCDIS NEvents Correct [%]	$\mu+\pi^C$ 9484 50,8	$\mu+\pi^C+p$ 3946 42,3	NCDIS π^C+p 2773 37,3
CCDIS NEvents Correct [%]	$\mu+2\pi^C$ 2148 35,5	$\mu+2\pi^C+p$ 3214 27,9	NCDIS π^0 1358 31,0
CCDIS NEvents Correct [%]	$\mu+\pi^0$ 828 31,3	$\mu+\pi^0+p$ 4882 22,8	NCDIS π^0+p 915 25,7

Correct fraction

nEvents, for context

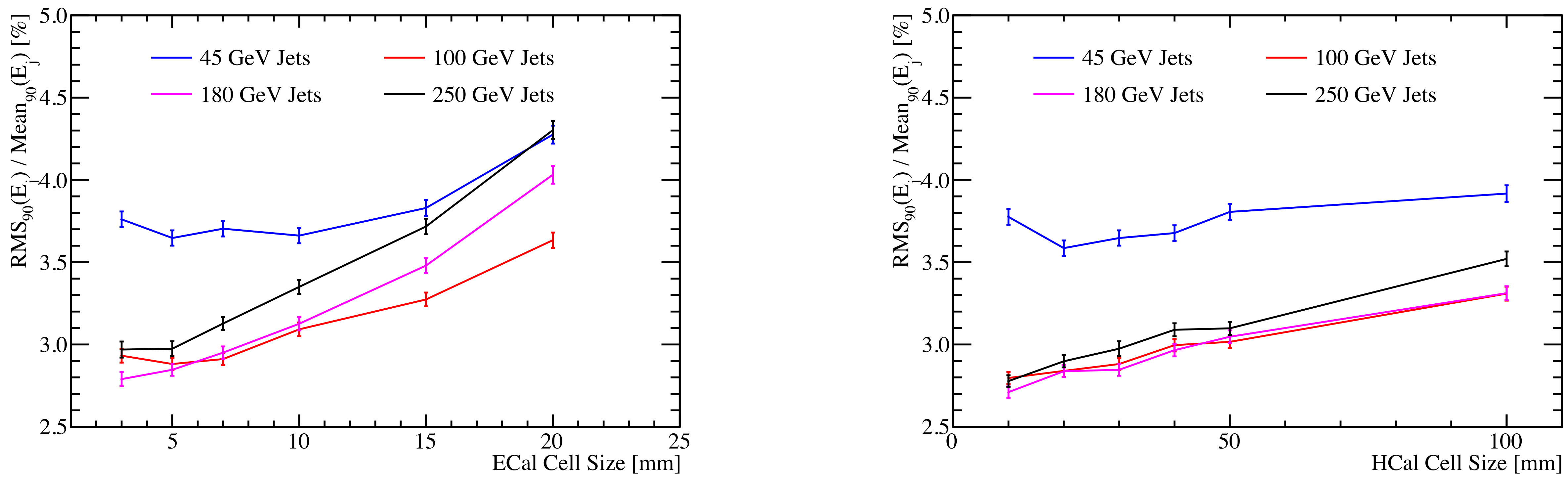
Very strict metrics: demanding perfection

Algorithm development and performance evaluation has been mainly in a MicroBooNE context.

Started tuning for DUNE, first flavour of performance metrics

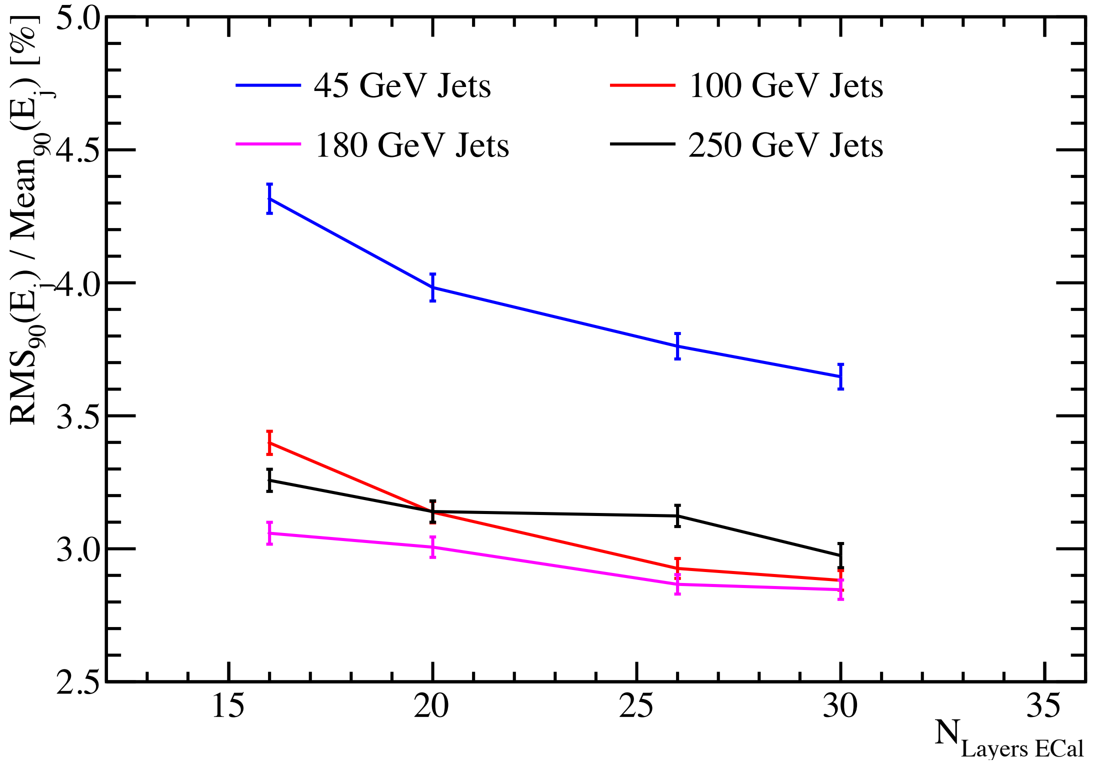
- Final results for study optimising the calorimeters for future linear collider completed
- Full calibration procedure was applied to each detector model to ensure optimal performance was achieved

S.Green

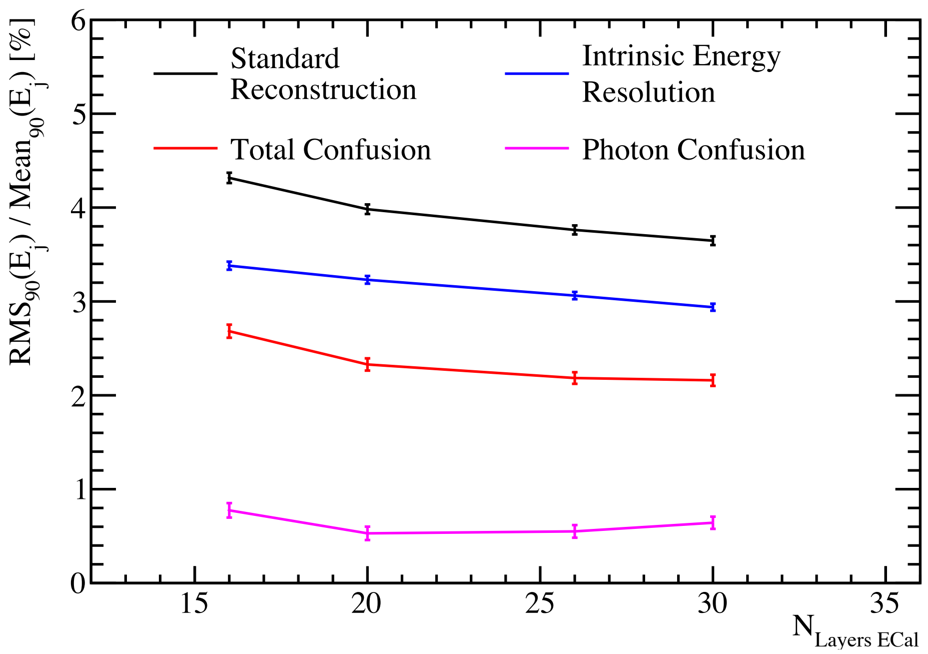


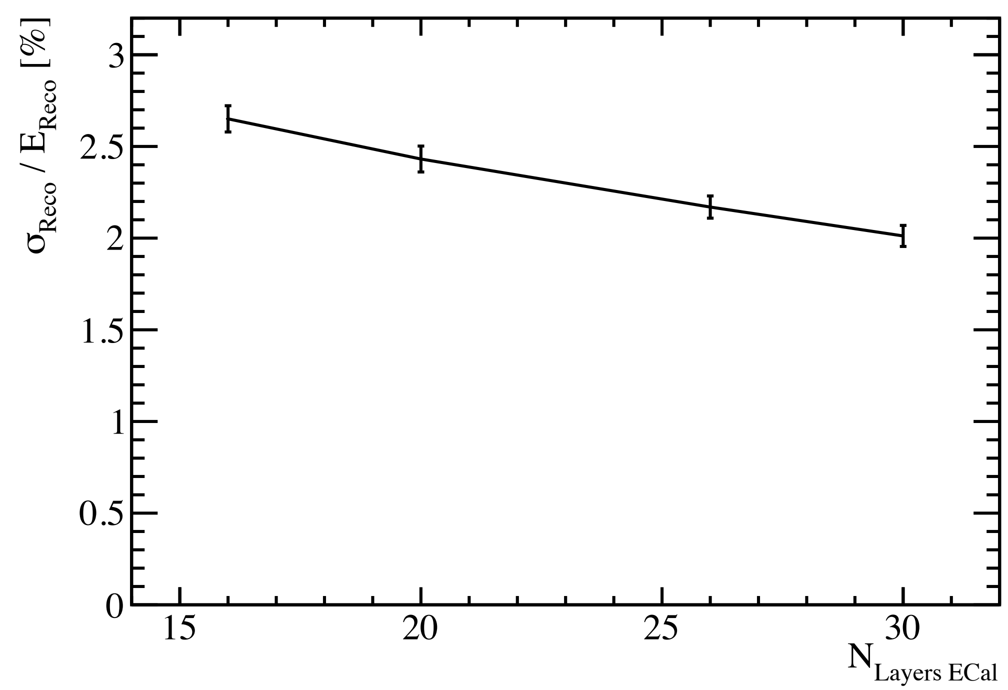
Cambridge: Pandora LC

- Performance trends understood using decomposition of jet energy resolution into different contributions as well as single particle studies.



S.Green

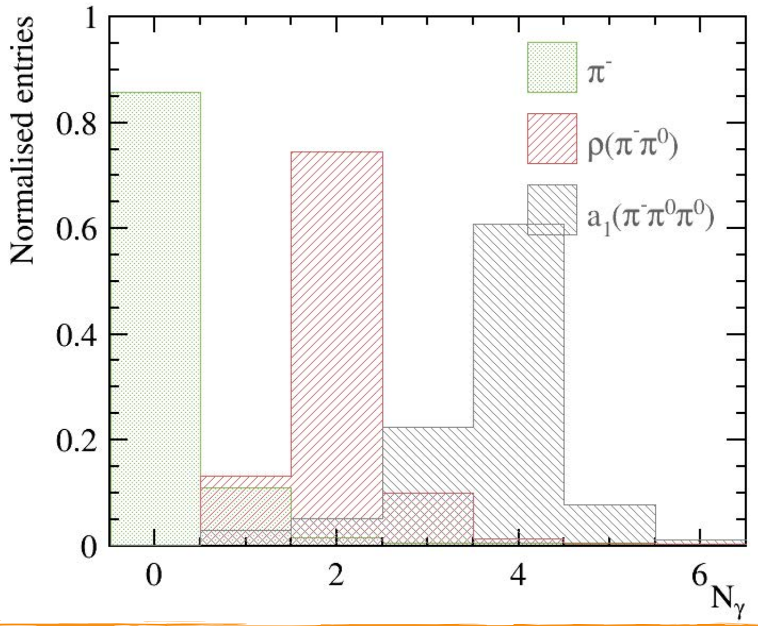


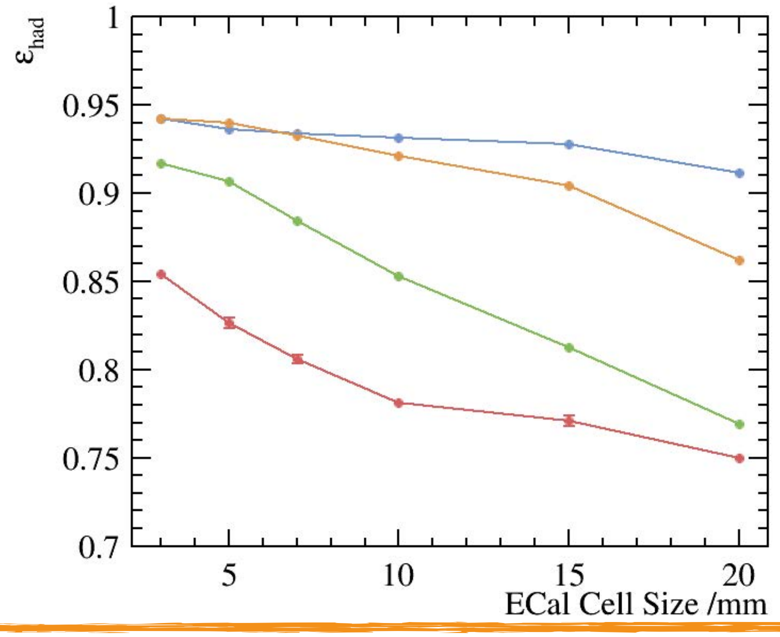


Cambridge: Pandora LC

- Tau final state separation studies. Test of detector performance, important to reconstruct photons correctly.
- Use e+e-→ $\tau^+\tau^-$, no ISR. MVA trained on 29 variables.
- For ILD, tables show selection efficiency for $\sqrt{s}$ = 100GeV. (Number < 0.25% not shown)

Reco ↓ True →	$e^- \bar{\nu}_e$	$\mu^- \bar{\nu}_\mu$	π^-	$\rho(\pi^-\pi^0)$	$a_1(\pi^-\pi^0\pi^0)$	$a_1(\pi^-\pi^-\pi^+)$	$\pi^-\pi^-\pi^+\pi^0$
$e^- \bar{\nu}_e$	**99.7**	-	0.9	0.6	0.4	-	-
$\mu^- \bar{\nu}_\mu$	-	**99.5**	0.6	-	-	-	-
π^-	-	0.3	**94.0**	0.8	-	0.4	-
$\rho(\pi^-\pi^0)$	-	-	3.4	**93.6**	9.5	0.6	2.3
$a_1(\pi^-\pi^0\pi^0)$	-	-	-	4.5	**89.7**	-	0.6
$a_1(\pi^-\pi^-\pi^+)$	-	-	0.9	-	-	**96.8**	6.4
$\pi^-\pi^-\pi^+\pi^0$	-	-	-	0.3	-	2.0	**90.6**





B. Xu

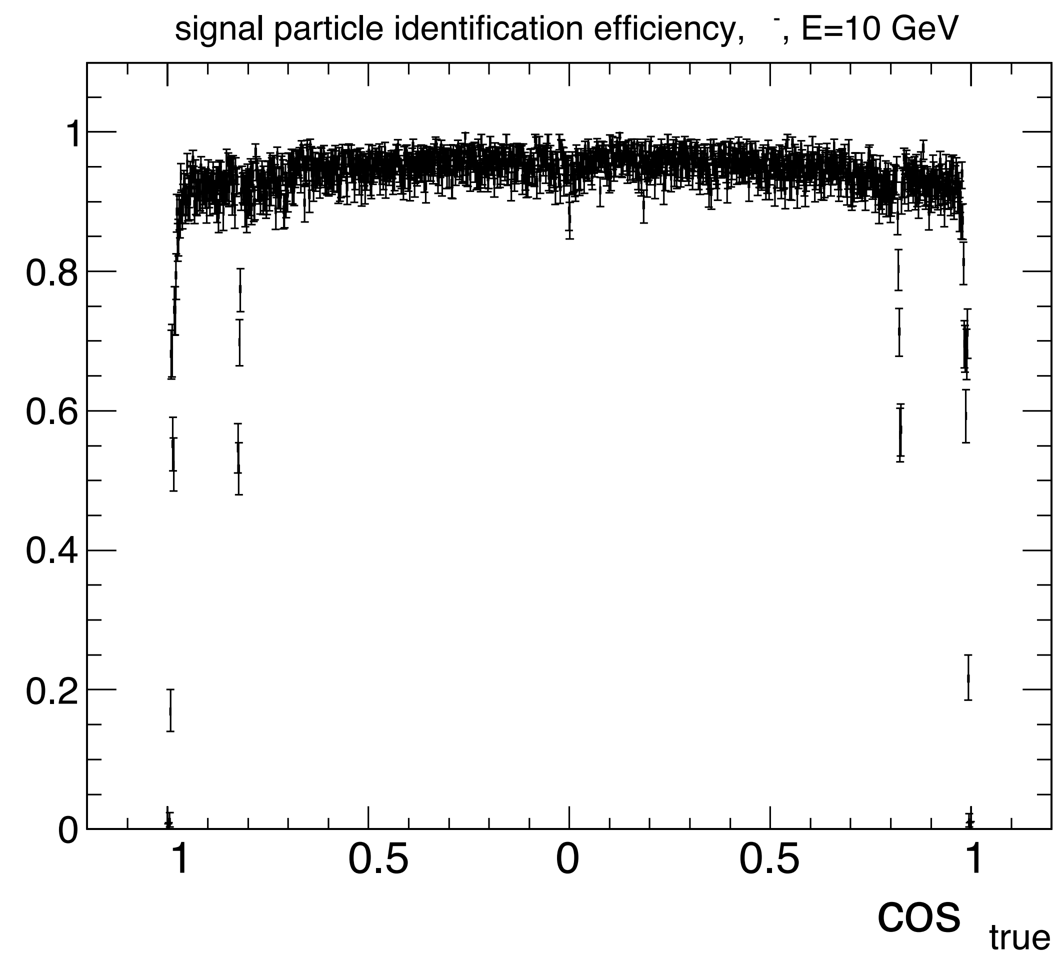
The tau lepton hadronic decay correct classification probability. Blue, orange, green and red lines represent $\sqrt{s}$ = 100, 200, 500 and 1000 GeV.

CERN: DDMarlinPandora

In the last year, the CERN group continued development of DDMarlinPandora interfacing DD4hep geometry and PandoraPFA reconstruction for high granularity calorimeters:

- More use of DD4hep Geometry information
- Replacing previous assumptions about the order of polygons
- updating track selection criteria for pure silicon tracking
- bug fixes

Work continues to solve remaining issues and improve reconstruction efficiencies, particularly the track cluster matching at gaps between barrel and endcap.

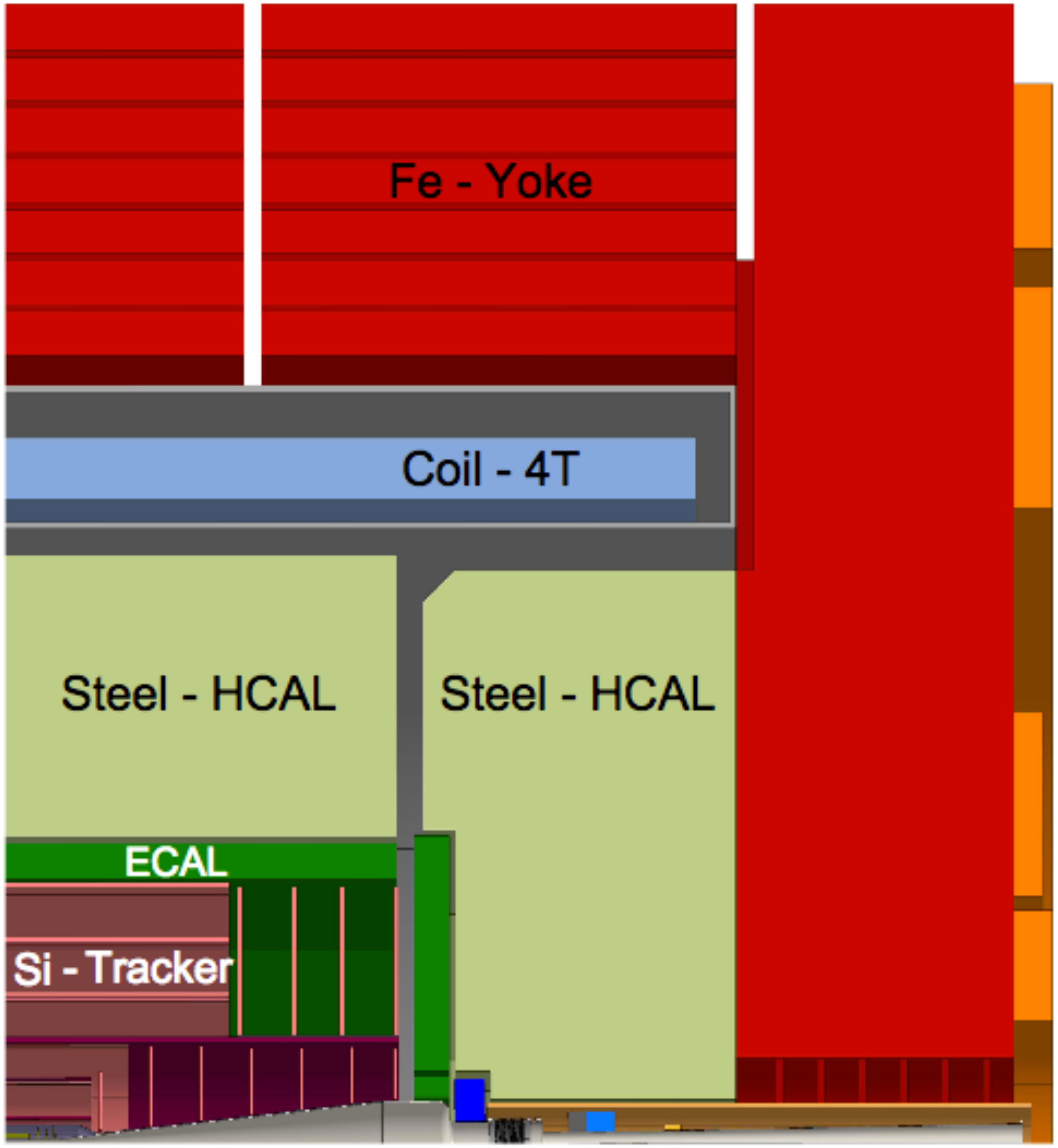


CERN: CLIC detector model

The new CLIC detector model now used to run simulation and reconstruction

New optimised model CLICdet for new benchmark studies: 4 Tesla solenoid field, all steel scintillator HCAL, silicon-tungsten ECAL, pure silicon tracking

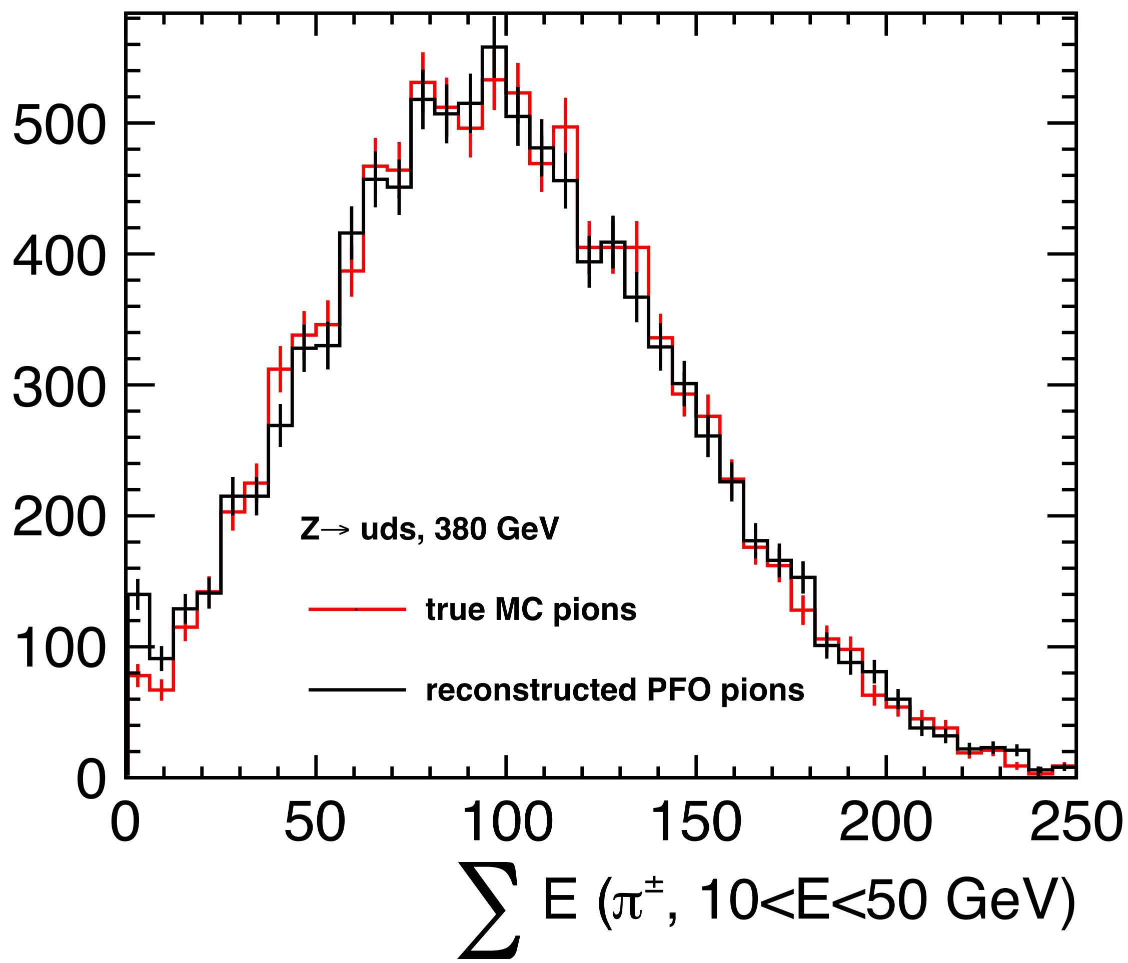
More information here:
https://indico.cern.ch/event/577810/contributions/2487776/

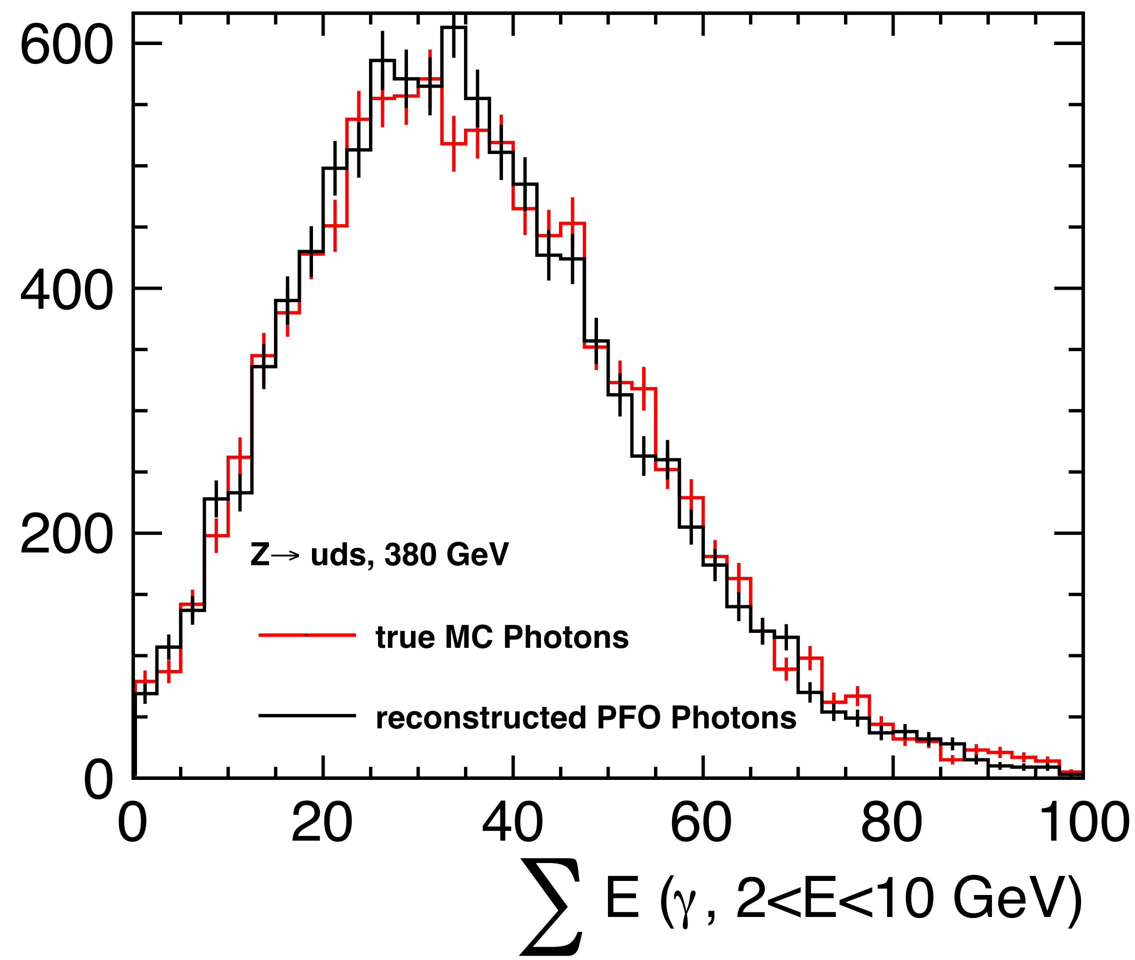


CERN: Example performance plot

- Reconstruction performance improved during the last year
- Reconstruction calibrated with standard PandoraPFA calibration method

Example plots: Good agreement between MCTruth and reconstructed energy for the sum of energy of pions/photons in jet events





WP3: Work on ARBOR

Optimisation of High Granularity PFA
IHEP
CNRS-LLR, CNRS-IPNL

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 654168.

PFA efforts on ARBOR (IHEP, IPNL, LLR)

- Continuous cooperation between LLR (V. Boudry H. Videau, J.C. Brient), IHEP (M. Ruan, D. Yu, L. Liao) and IPNL (B. Li, R. Été) on ARBOR integration
  - Rémi Été's PhD defended March 2017.
    - Includes description of ArborPFA and integration in PandoraPFA on ILD cases
- March 2017, 20–24th:
  - Mini-Workshop (1 day) on PFA at LLR: CALICE & ILD, CEPC, CMS, ATLAS
  - Latest results have been presented at the CALICE meeting:
    - https://agenda.linearcollider.org/event/7454/timetable/#20170324.detailed
    - M. Ruan on ARBOR, LICH & CEPC optim studies
    - K. Shpak Separation of two showers one of which is EM using Pandora, Garlic and Arbor
    - Lepton Identification (LICH) Paper submitted to EPJ (feb 2017). https://arxiv.org/abs/1701.07542

Physics > Instrumentation and Detectors

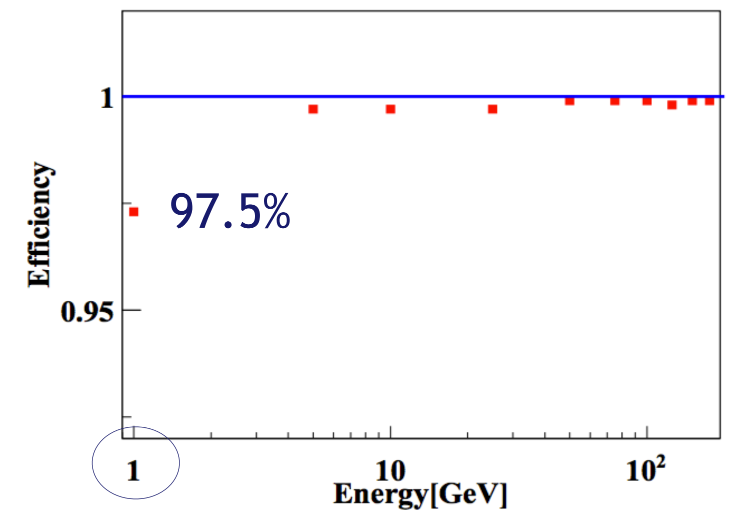
Lepton identification at particle flow oriented detector for the future e^+e^- Higgs factories

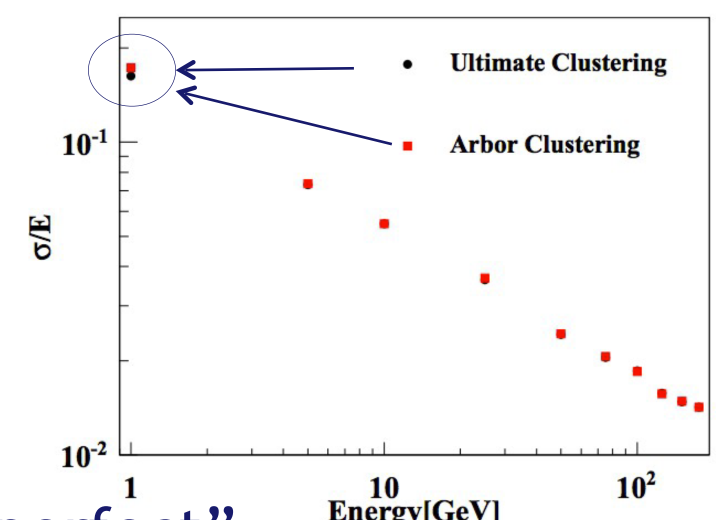
Dan Yu, Manqi Ruan, Vincent Boudry, Henri Videau

(Submitted on 26 Jan 2017 (v1), last revised 17 Feb 2017 (this version, v2))

CEPC-det optimisation studies

- CEPC : 1M Higgs + 10–100G Z
  - Detector ~ILD-o2_v5 mod. : SiW ECAL + (S)DHCAL
    - e.g. no Power-pulsing & passive cooling → lower granularity
- Used as a test stand for ARBOR and optimisation on PFA objects
  - Energy resolution: γ’s, leptons, Jets
  - Performance: H→ γγ, H recoil, H→WW, ZZ
  - Cracks, dynamics, inhomogeneity on photons
- Example of Performance Plots
  - Clustering of photons: ideal vs realistic
  - Longitudinal & lateral segmentations, W thickness
  - Separation for H→γγ, Z→ττ





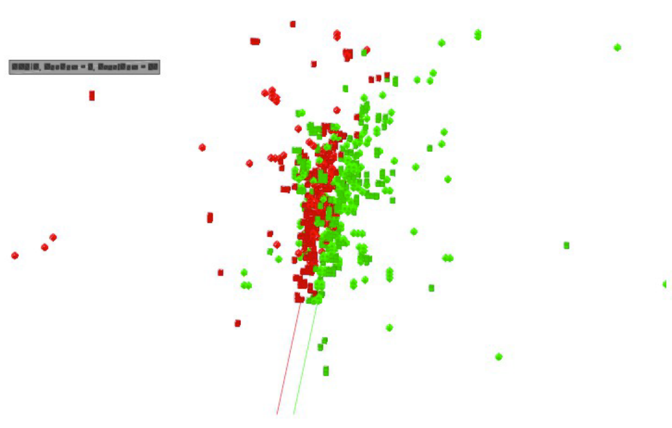
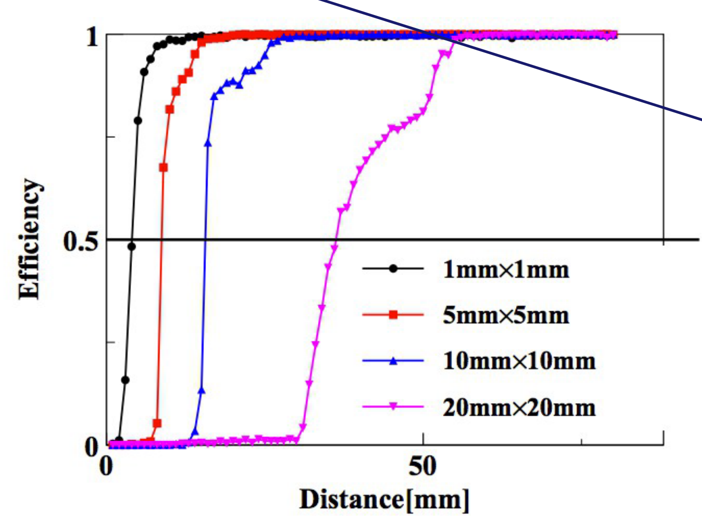


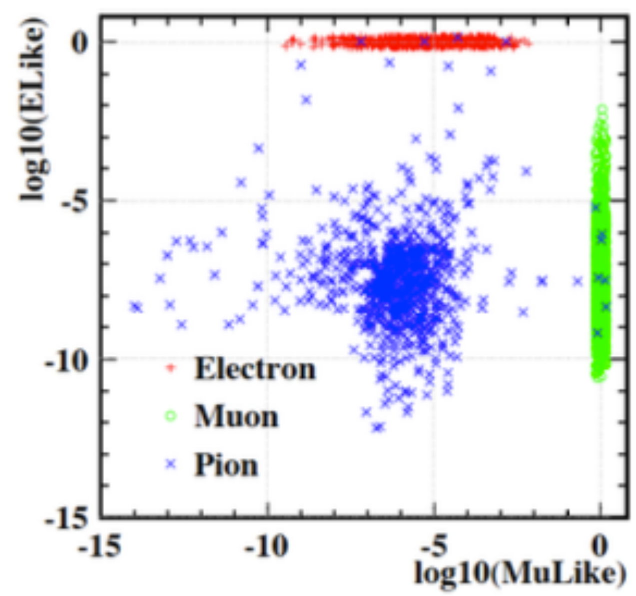
Figure 11. Event display of reconstructed di-photon.

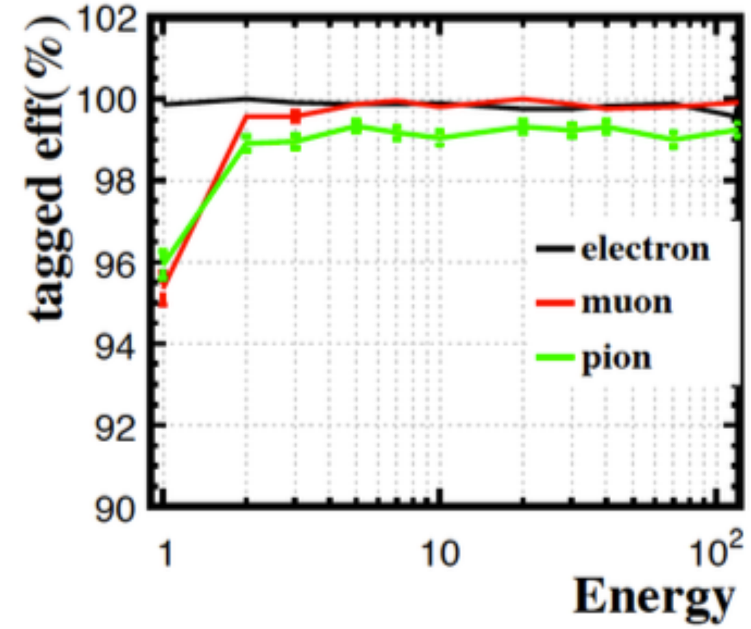


“Close to perfect”

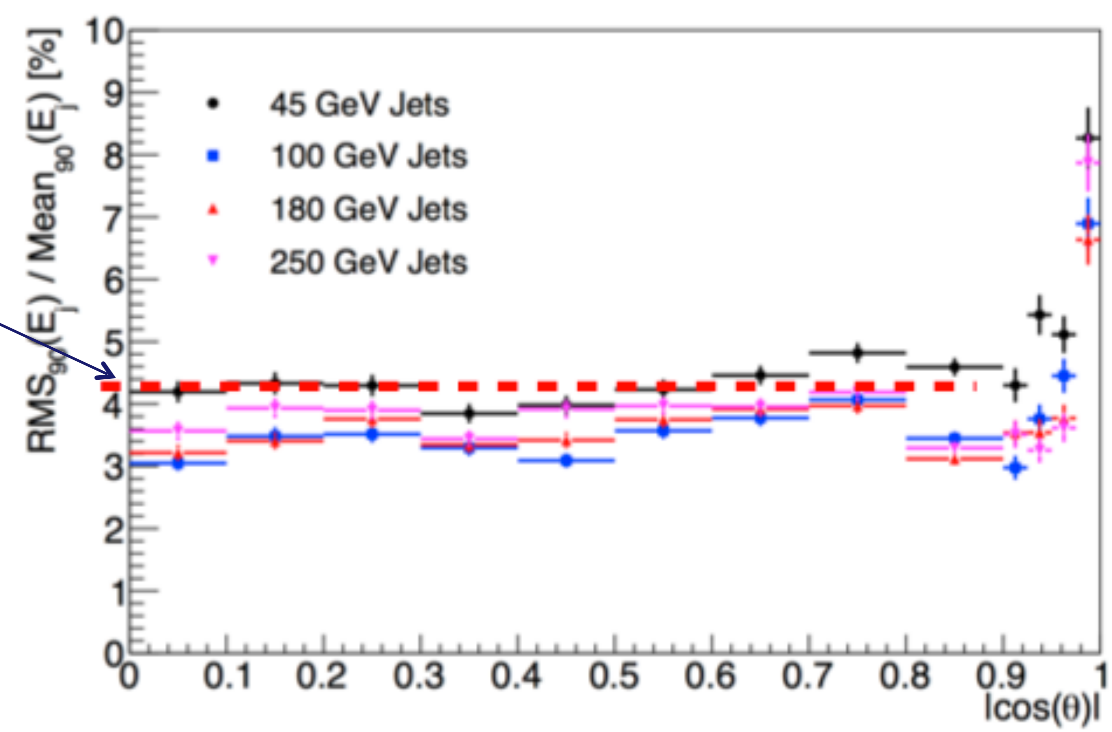
Lepton ID for Calorimeter with High granularity (LICH)

- BDT on 24 input variables
- Optim granularity:
  (HCAL,ECAL) × (#layers, cell size)
  - No Significant effect for
    E > 2 GeV charged Particles
- Effect on JER on ννH: 3% improv't.
  - Photon direction reconstruction
    → ISR reduction (28% improv't)
  - With ν vetoed, RMS_{90}^{rel} = 3.8% / $\sqrt{2}_{jets}$ → 4.3%
    - E(Jets) ⊂ 30–100 GeV
    - Closing to PandoraPFA using SDHCAL
      (PhD A. Steen, 205LYO10230)



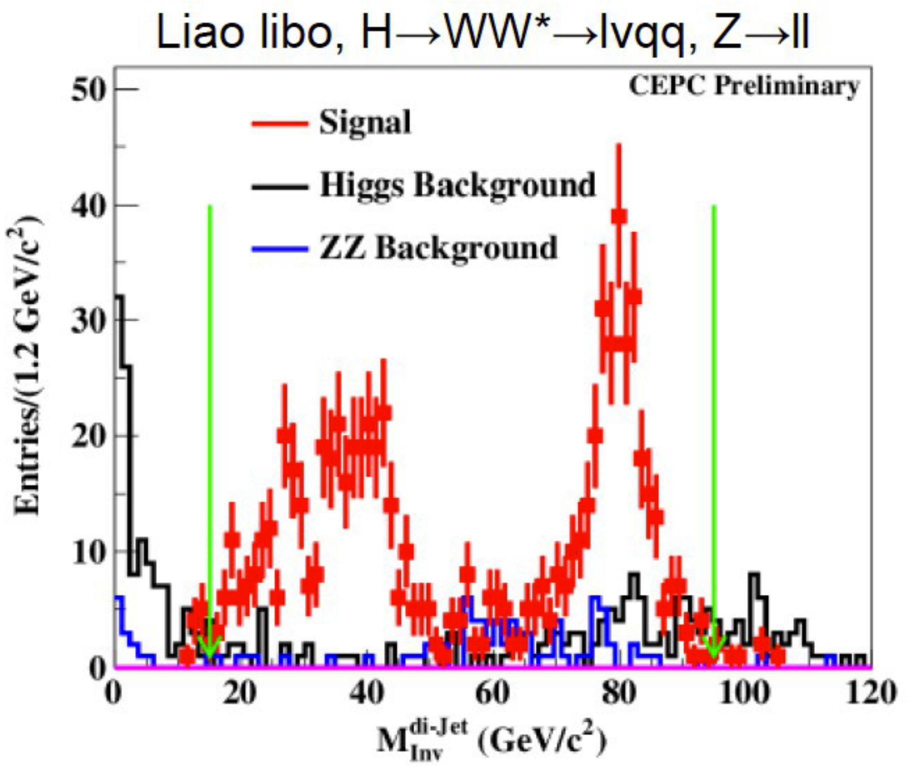


BDT method using 4 classes of 24 input discrimination variables.



- H recoil on ℓℓH events
- Br(H→WW)
  - H→WW,ZZ ⇒ Higgs width & perfect test

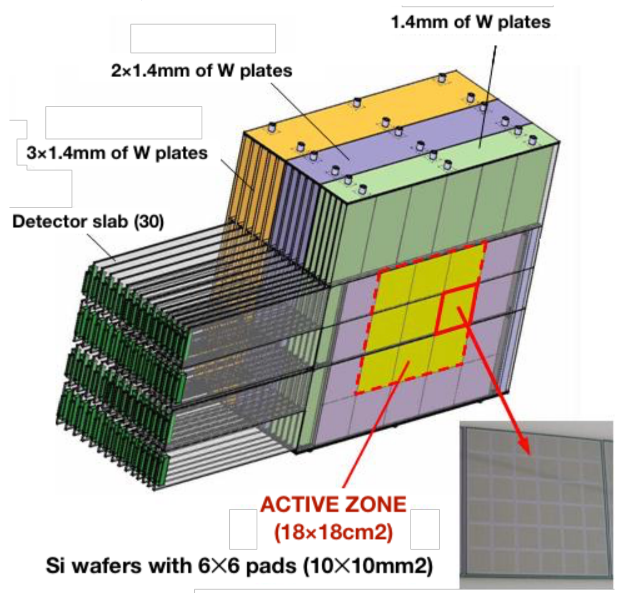
    bed for detector/reconstruction performance...

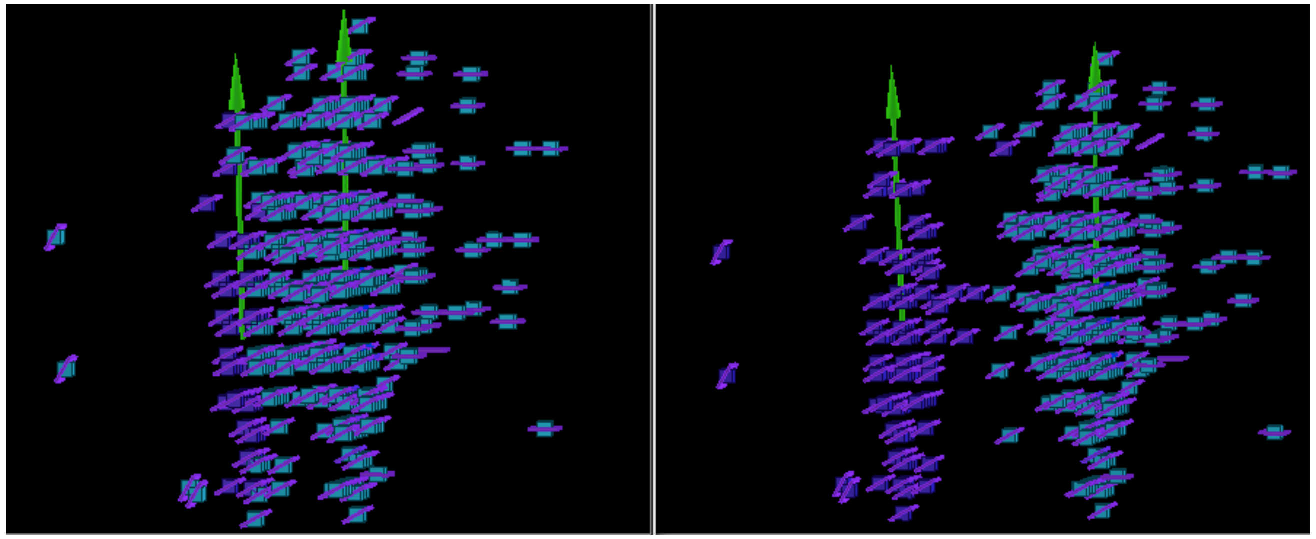


- Conclusion:
  - ARBOR in heavy phase of testing with ≠ detector configuration many different channels
    - Improved re-connections
  - Added LICH lepton in the code (v3.3)
  - Tests continue with re-written part → Arbor V3.4
    - Coherent results

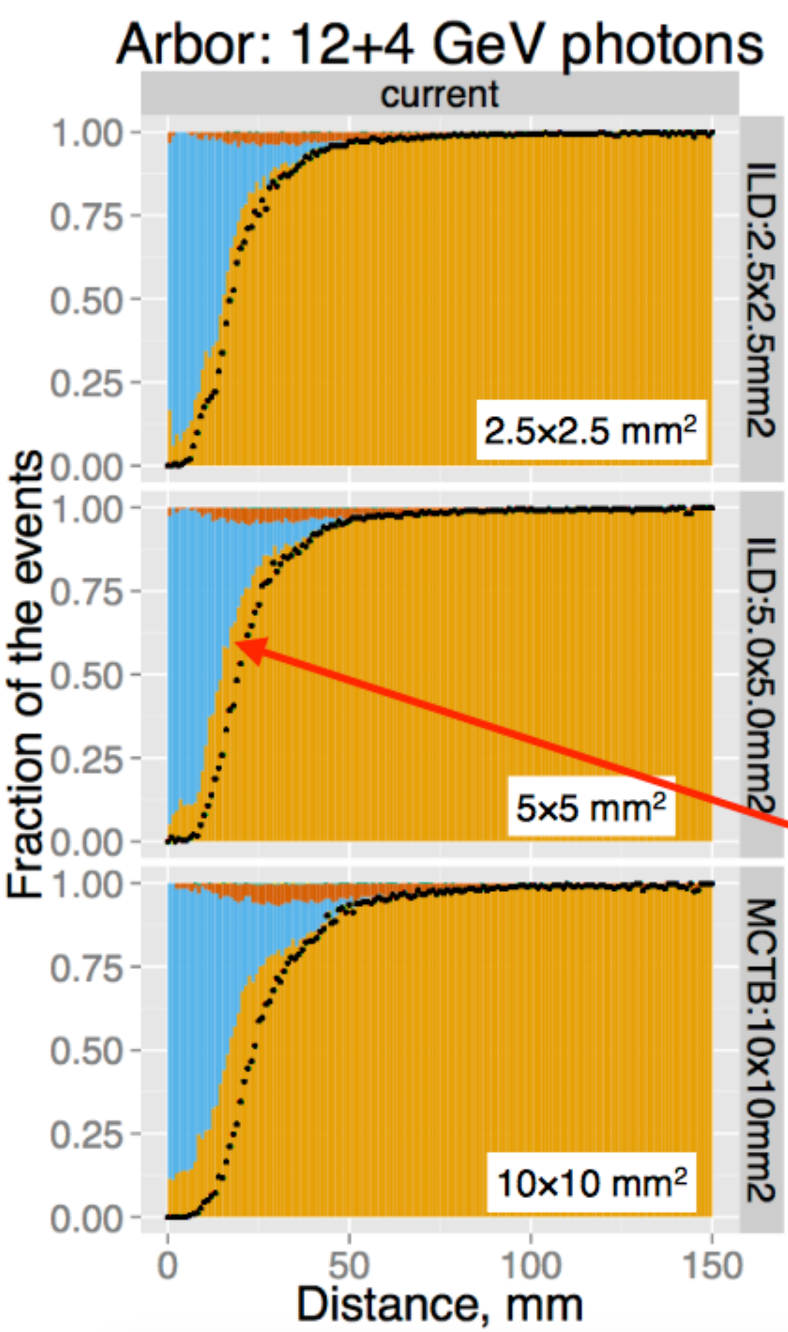
Separation of EM shower's Kostiantyn Shpak (LLR)

- Confusion studies on Beam test data & simulation (CALICE & ILD) reconstruction of simplest 2 shower events (e+−e+; γ−γ; e+,γ−π+) obtained by **event mixing**
  - Pandora old v00-14 & September'16 v02-04 (better photon reconstruction); not used for SDHCAL
  - Garlic old v2.11 & new v3.0.3; only ECAL
  - Arbor March'15; not used for AHCAL
- Data:
  - CALICE: SiW ECAL physics prototype *standalone* (FNAL'11) + AHCAL prototype (CERN'07)
  - ILD with 2.5x2.5mm^2 or 5x5mm^2 ECAL pixels and with AHCAL (3x3cm^2) or SDHCAL (1x1cm^2)
  - 4-32 GeV particles
- No final official CALICE results yet (in very final approval line)
  - Analysis Note to be released soon
  - Sim only here.





γ–γ: classification of inefficiencies



- are preliminary data

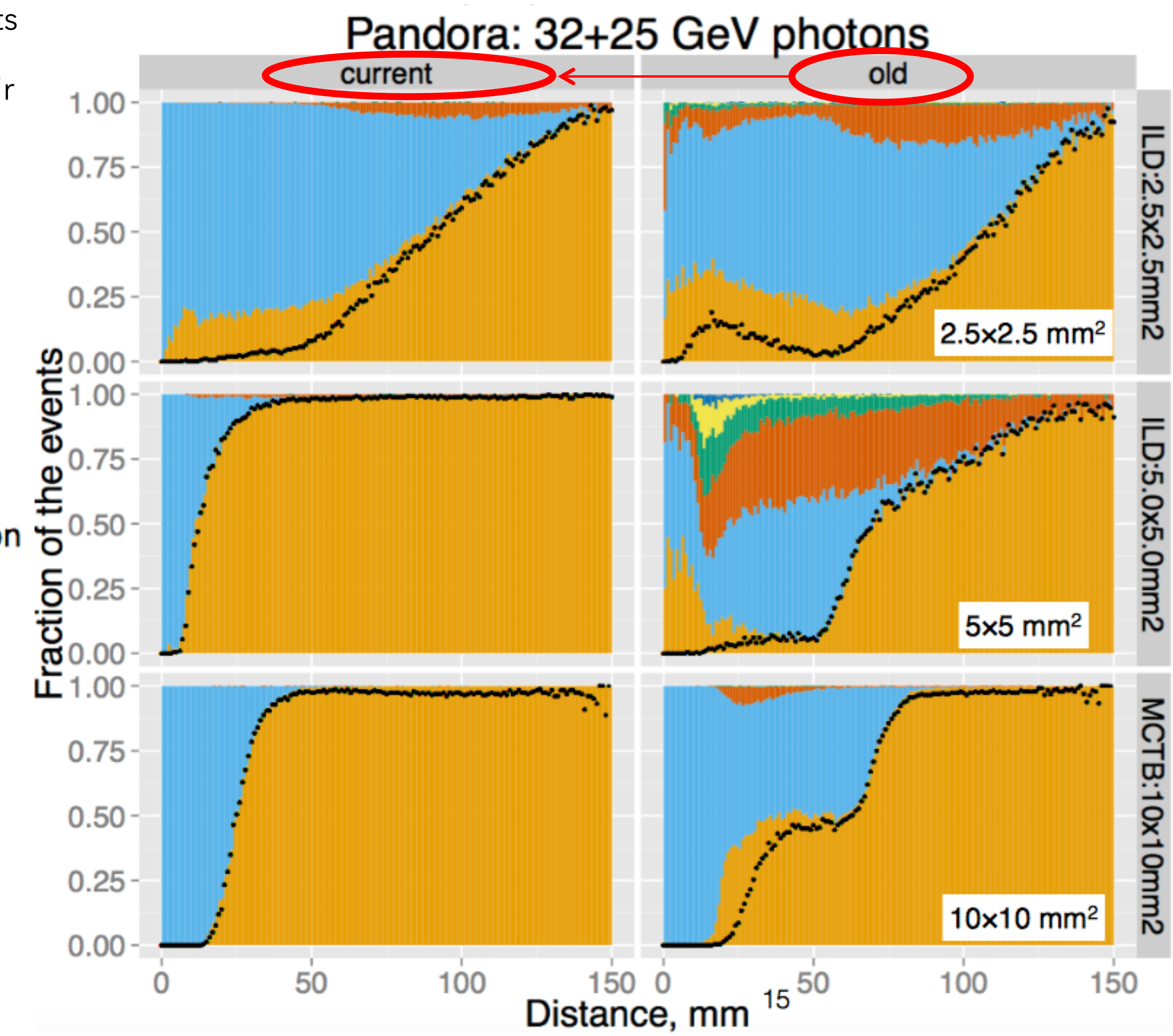
Colored bands: fraction of events where 1, 2, 3, . . . neutral clusters reconstructed, regardless of their energy and position.

Black dots: γ–γ reconstruction efficiency

N Clusters
- 2
- 1
- 3
- 4
- 5

(1, 3, 4, 5) } inefficient reconstruction

orange band events above black dots failed PID or E(±20%) or X,Y(±5 mm)



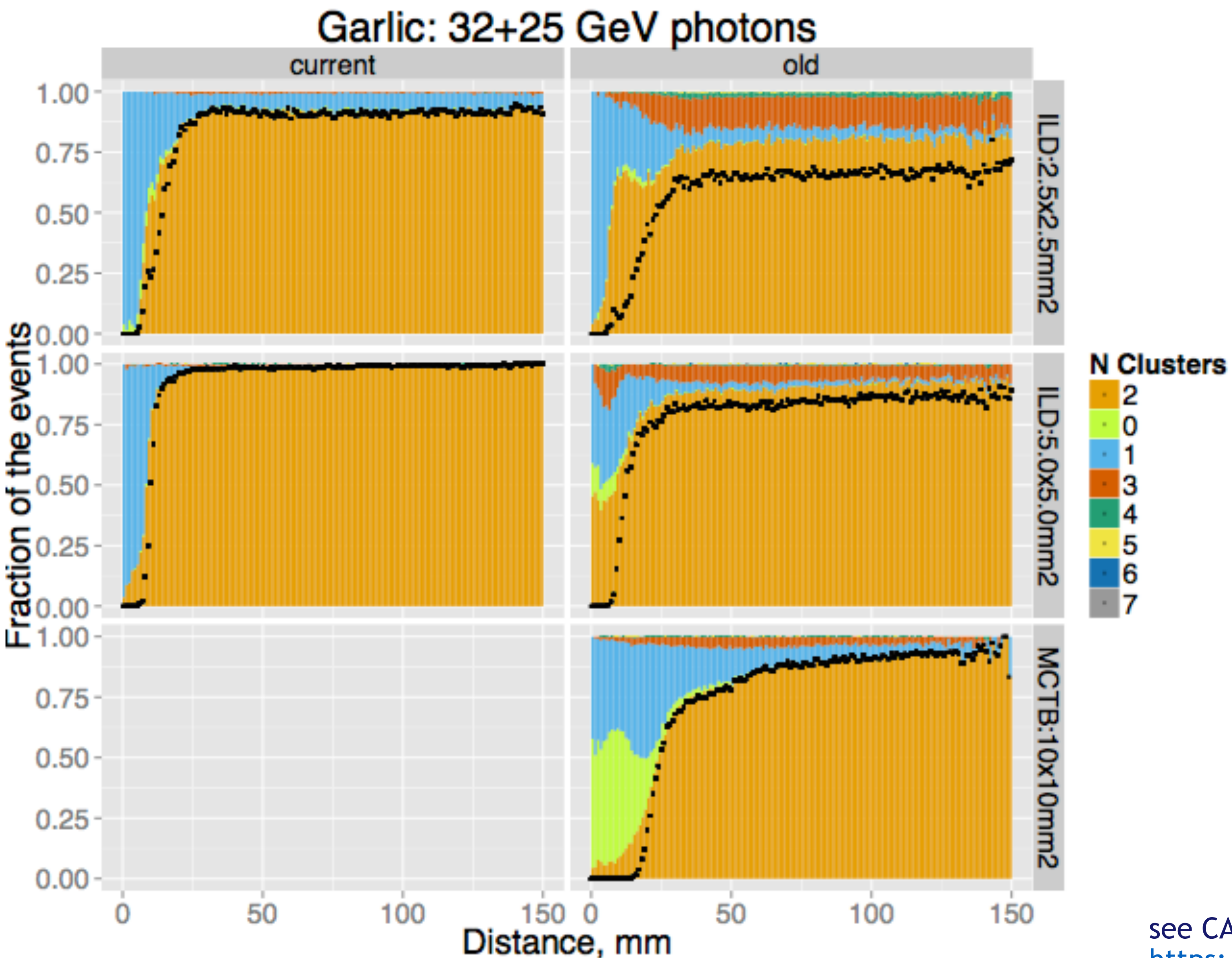
12

15

- Example of comparisons. Distance = between shower barycenters

Kostiantyn Shpak (LLR)

γ–γ: classification of inefficiencies



- Good agreement MC/Data
  - Has helped improving algorithms (bugs found and corrected)

Complementary algorithms:
  - PandoraPFA better at High E
  - Garlic is better for γ-γ separation
  - Arbor better for low E photons
    - Requires add'l tuning for Garlic & Pandora with 2.5mm cells

Also studied π – γ «jets» in ILD & CALICE
  - π – γ separation
  - RMS(ΔE)

see CALICE meeting 2017 slides https://agenda.linearcollider.org/event/7454/contributions/38730/attachments/31388/47213/CALICEmeeting2017.pdf

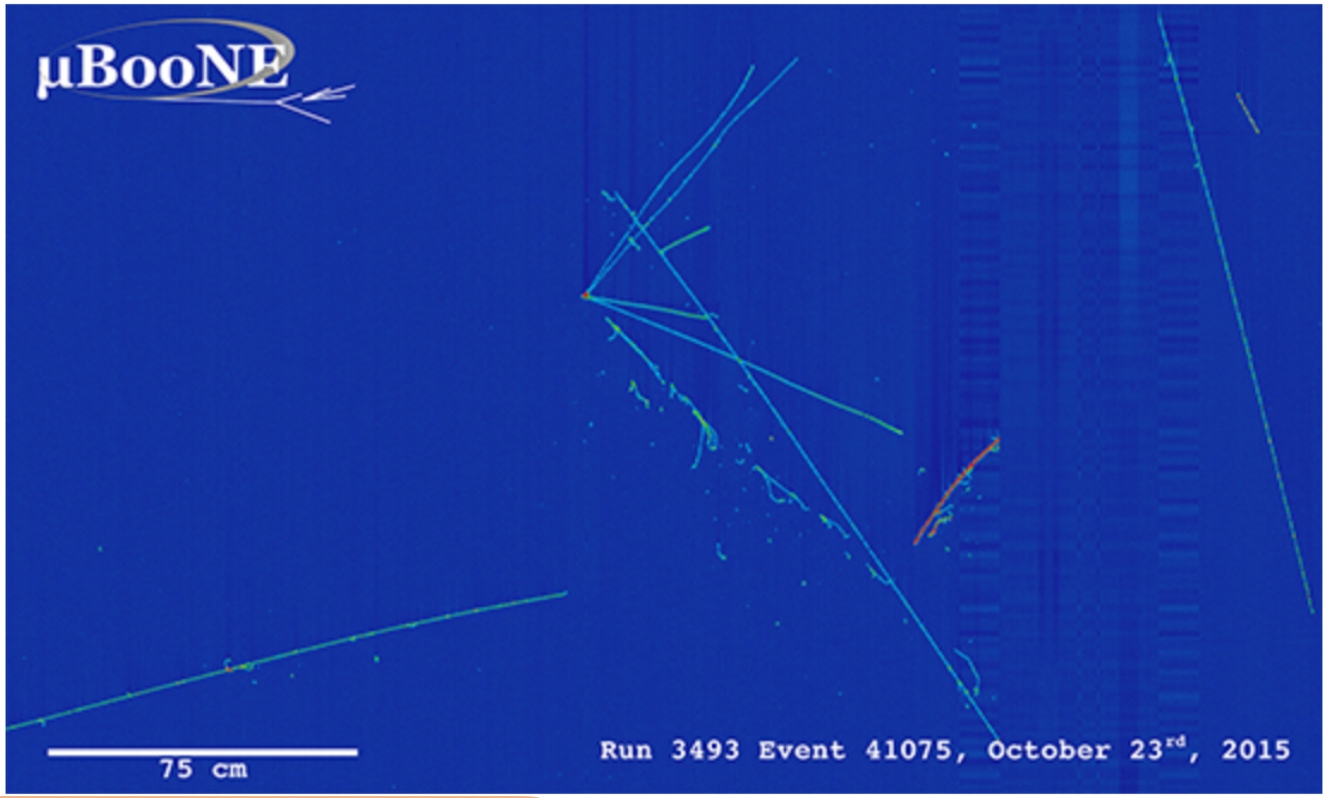
Kostiantyn Shpak (LLR)

Thanks!

n track
News for the Detector Community

Pandora: opening the box for neutrinos

Barbara Warmbein (DESY), 13/02/2017



Pattern recognition rules in particle physics. When particles collide, many things happen at the

And we keep working on it!

Pandora Documentation

http://arxiv.org/abs/1506.05348

or EPJC.75.439 or PandoraPFA github page

The Pandora Software Development Kit for Pattern Recognition

J. S. Marshall[a,*], M. A. Thomson[a]

[a] *Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom*

Abstract

The development of automated solutions to pattern recognition problems is important in many areas of scientific research and human endeavour. This paper describes the implementation of the Pandora Software Development Kit, which aids the process of designing, implementing and running pattern recognition algorithms. The Pandora Application Programming Interfaces ensure simple specification of the building-blocks defining a pattern recognition problem. The logic required to solve the problem is implemented in algorithms. The algorithms request operations to create or modify data structures and the operations are performed by the Pandora framework. This design promotes an approach using many decoupled algorithms, each addressing specific topologies. Details of algorithms addressing two pattern recognition problems in High Energy Physics are presented: reconstruction of events at a high-energy e^+e^- linear collider and reconstruction of cosmic ray or neutrino events in a liquid argon time projection chamber.

Keywords: Software Development Kit, Pattern recognition, High Energy Physics

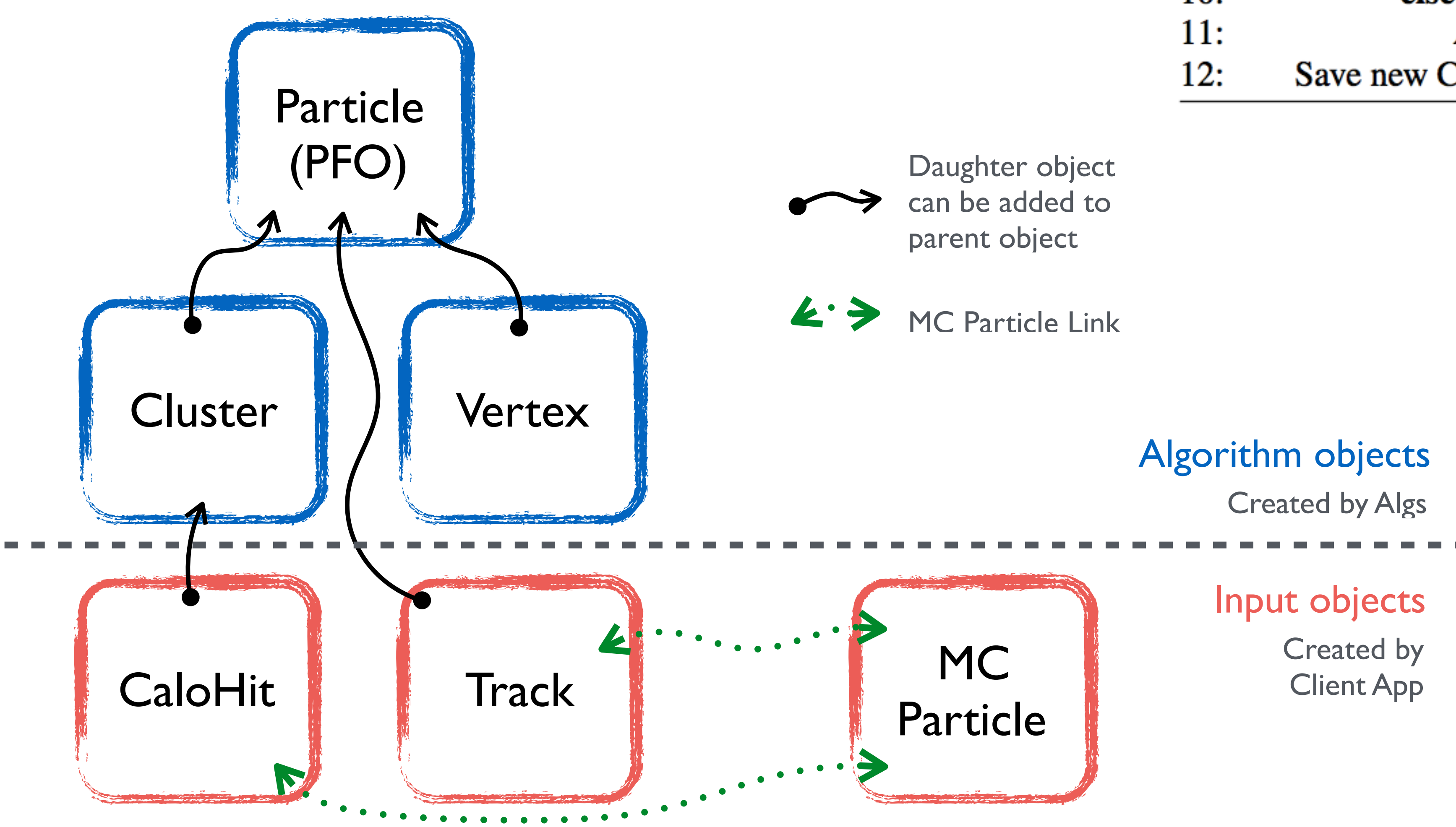
Pandora Algorithms

- Pandora algorithms provide the step-by-step instructions for finding patterns in the provided data.
- They use the APIs to access objects and to request the Pandora Managers to make new objects or modify existing objects.

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
```

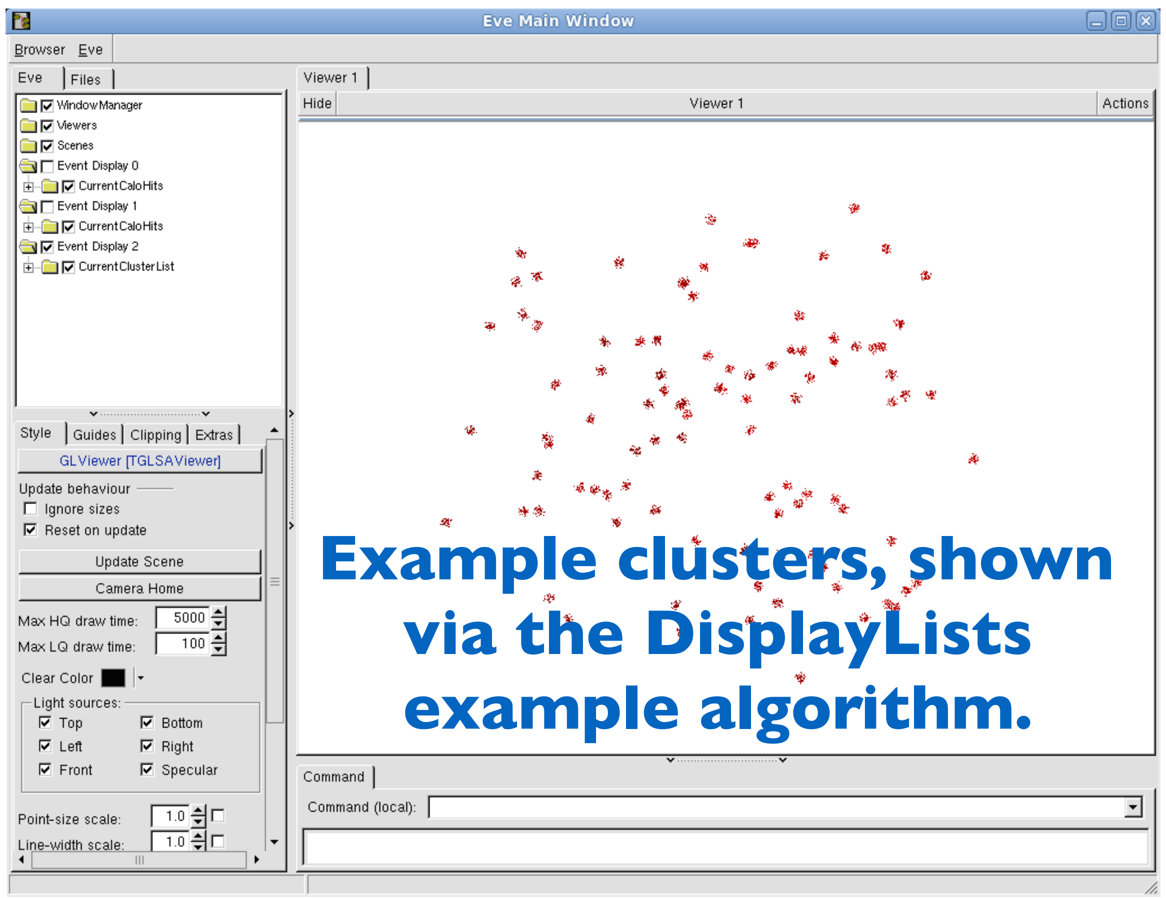
Simplified algorithm implementation



Generic, reusable Event Data Model

Pandora Learning Library

- Pandora algorithms create and/or modify clusters, vertices and PFOs. Their decisions (the algorithm logic) whether to proceed with operations can be complex and use-case specific.
- The aim of the Pandora ExampleContent library and test application is to demonstrate the key Pandora functionality in a very simple testing and learning environment.
- The ExampleContent library is structured in exactly the same manner as the LCContent and LArContent libraries, currently in use for Linear Collider and LAr TPC reconstruction.

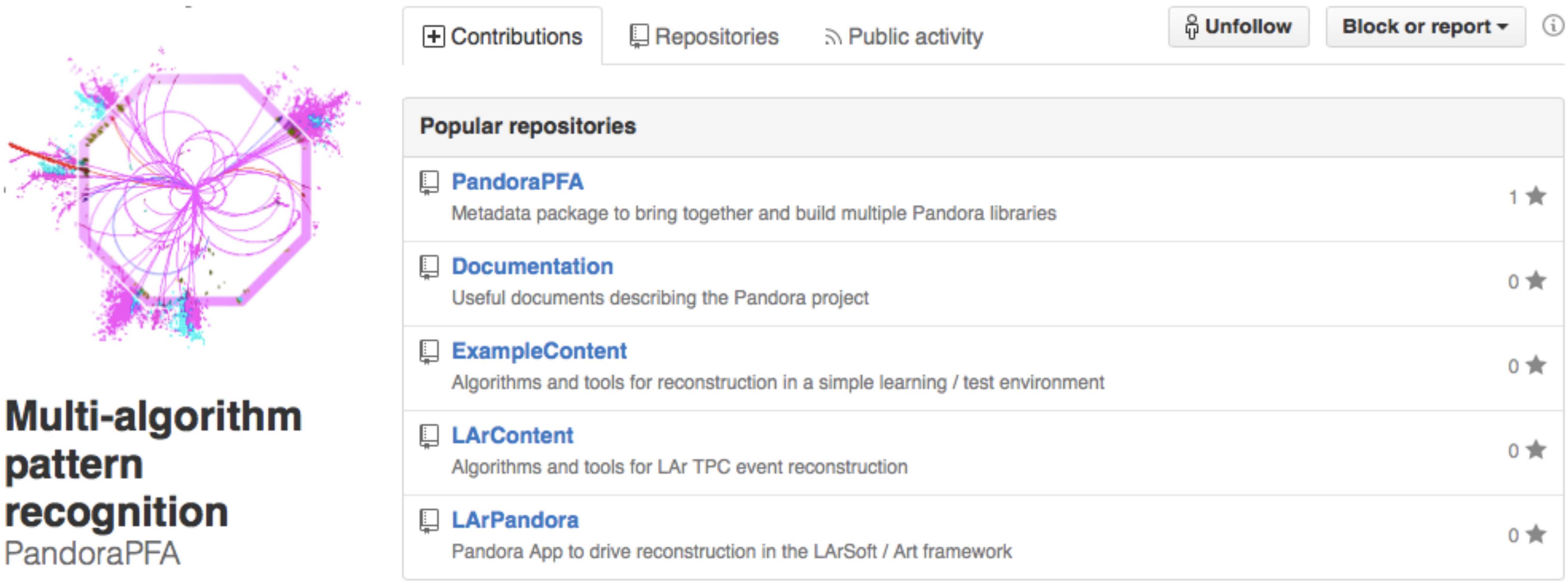


Example clusters, shown via the DisplayLists example algorithm.

- The library consists of example Algorithms, AlgorithmTools, Plugins and Helper functions:
  - Example list access and display
  - Example Cluster, Vertex and PFO creation
  - Cluster manipulation, including merging, deletion, fragmentation and reclustering
  - Creating and saving new lists of objects
  - Using Algorithm Tools and Plugins
  - Writing a tree using PandoraMonitoring.

https://github.com/PandoraPFA/Documentation

Pandora Distribution



Pandora now uses git to provide a distributed version control system.

Distributed approach is perfect for balancing development on multiple Pandora projects.

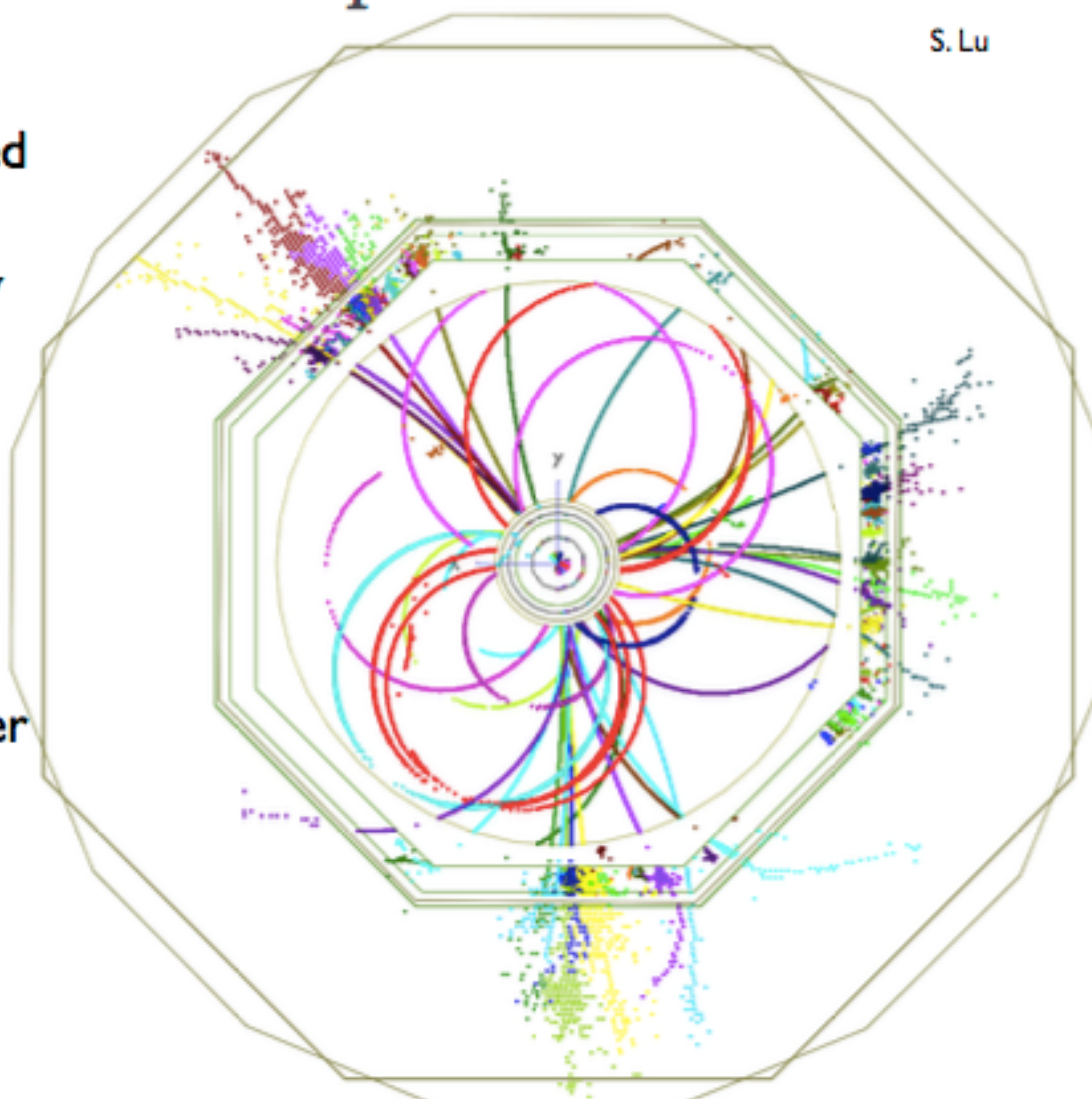
Source code is officially distributed from the github remote repository:

https://github.com/PandoraPFA

CERN: ILC/CLIC Geometry Migration

Event Simulated, Reconstructed and Visualized Fully with DD4hep

- **ILD_o1_v05** model implemented in **DD4hep**
- $Z \rightarrow uds$ event at $\sqrt{s} = 500$ GeV simulated in **DDSim**
- Tracks reconstructed using **DDSurfaces**
- PFOs from **DDMarlinPandora** using the **DDRec** data structures
- Event display from the **CED** viewer interfaced with **DD4hep**
  - Also uses **DDRec** and **DDSurfaces**



Significant progress using Pandora LC reconstruction with DD4HEP

Only needed to migrate geometry usage in the MarlinPandora client app…

…not throughout the 70+ pattern recognition algorithms

CERN: DDMarlinPandora

- Calibration for PFA is one of the tasks that must be performed by the Pandora user/client app.
- CERN have updated the semi-automated procedure from the LCPandoraAnalysis package, originally provided by S. Green, to make it more robust and generically applicable.
- A brief description and tutorial is available: https://twiki.cern.ch/twiki/bin/view/CLIC/DD4hepClicDetectorCaloCalibration

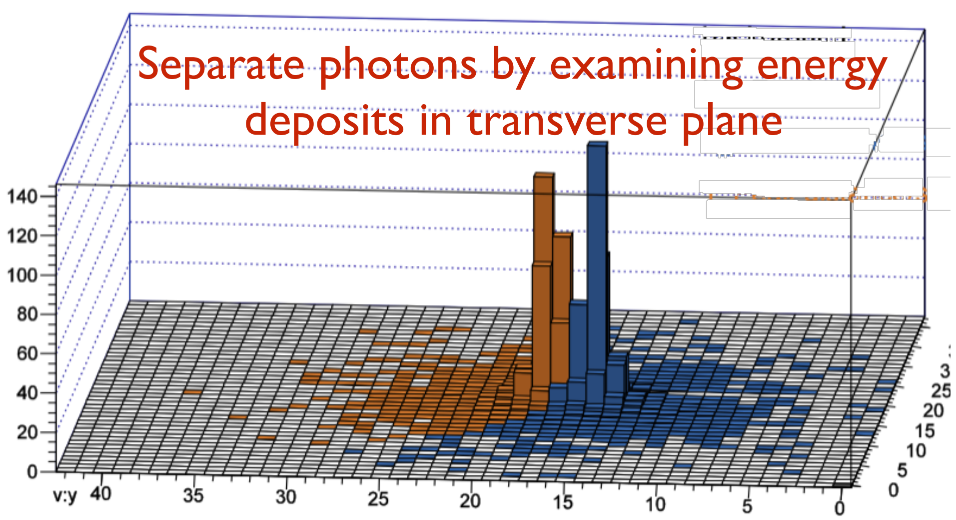
- Efforts to understand the quality cuts applied to inner detector tracks in MarlinPandora (strongly tied to the TPC-based tracker in ILD) and produce similar cuts for new CLIC model*
- Important decision: want to extract charged particle four-momentum from tracker measurements, but need to know when can trust track and its projection to the calorimeter.
- Using Pandora debugging tools, including Pandora Visual Monitoring toolkit, which allows algorithms to display e.g. relevant tracks/clusters at user-chosen points during reconstruction.

***CLIC model has only Silicon Tracking**

Cambridge: Pandora LC

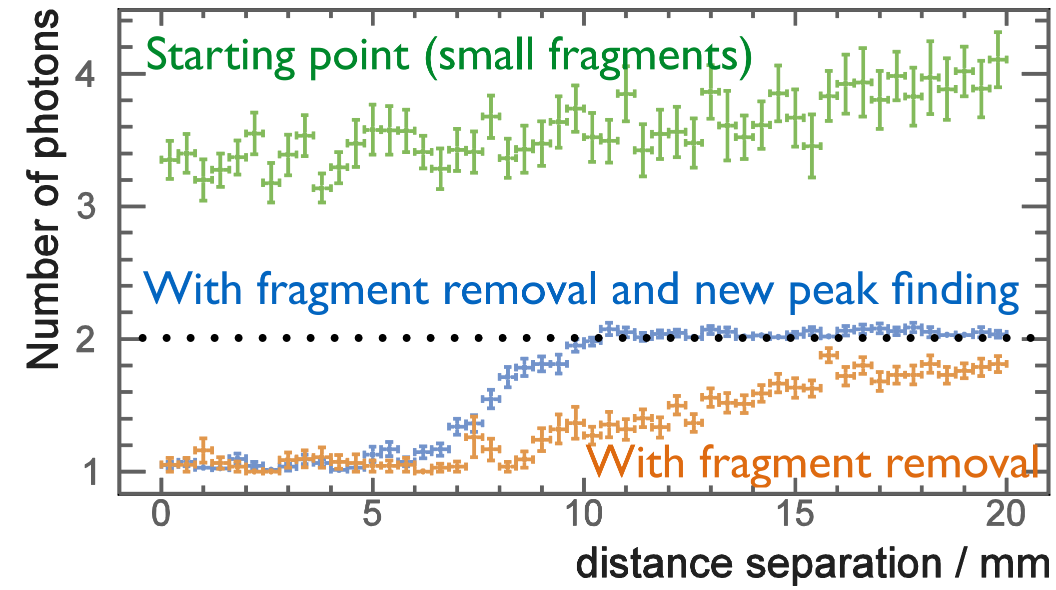
AIDA 2020

Improvements to reconstruction (fragment removal and separation of nearby photons) just released. See presentations in upcoming CLICdp analysis and ILD software/analysis meetings:

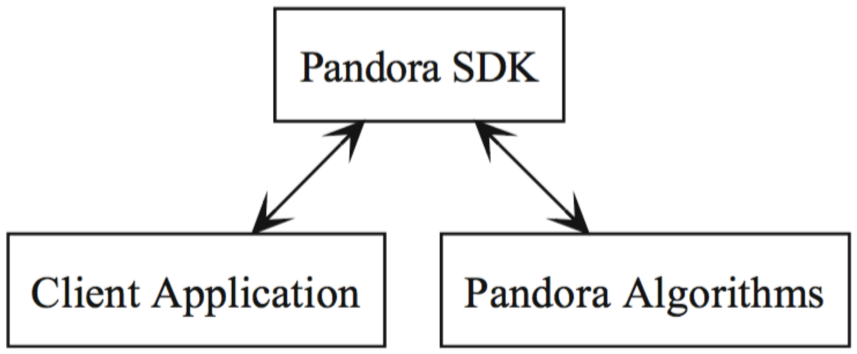


B. Xu

E.g. Separation of nearby high energy (500GeV) photons



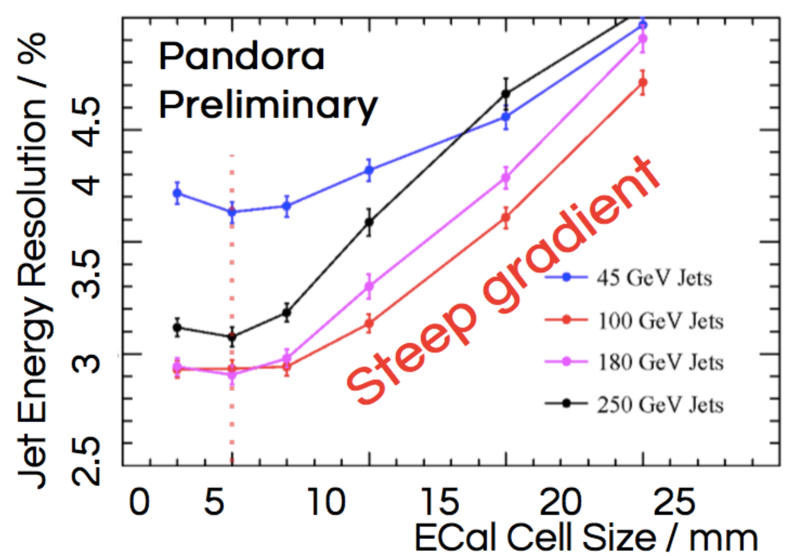
Currently supported:

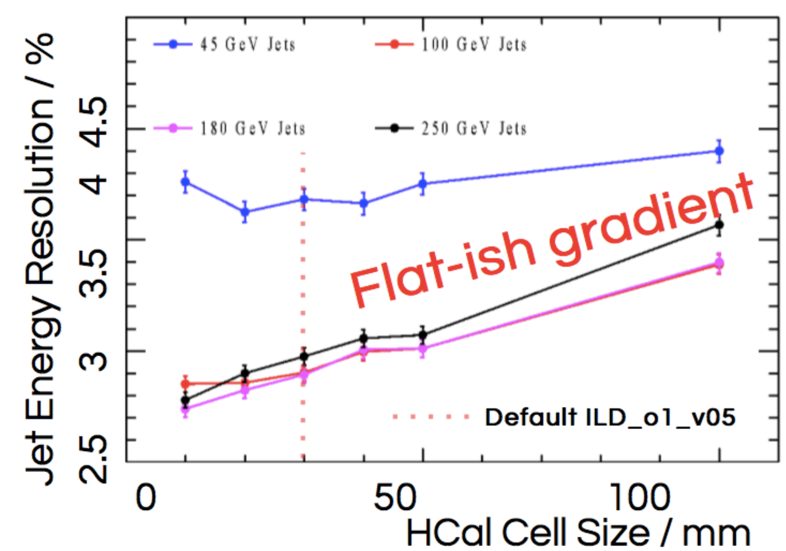


MarlinPandora
DDMarlinPandora
CMSPandoraTranslator
[SlicPandora]

LCPandoraContent

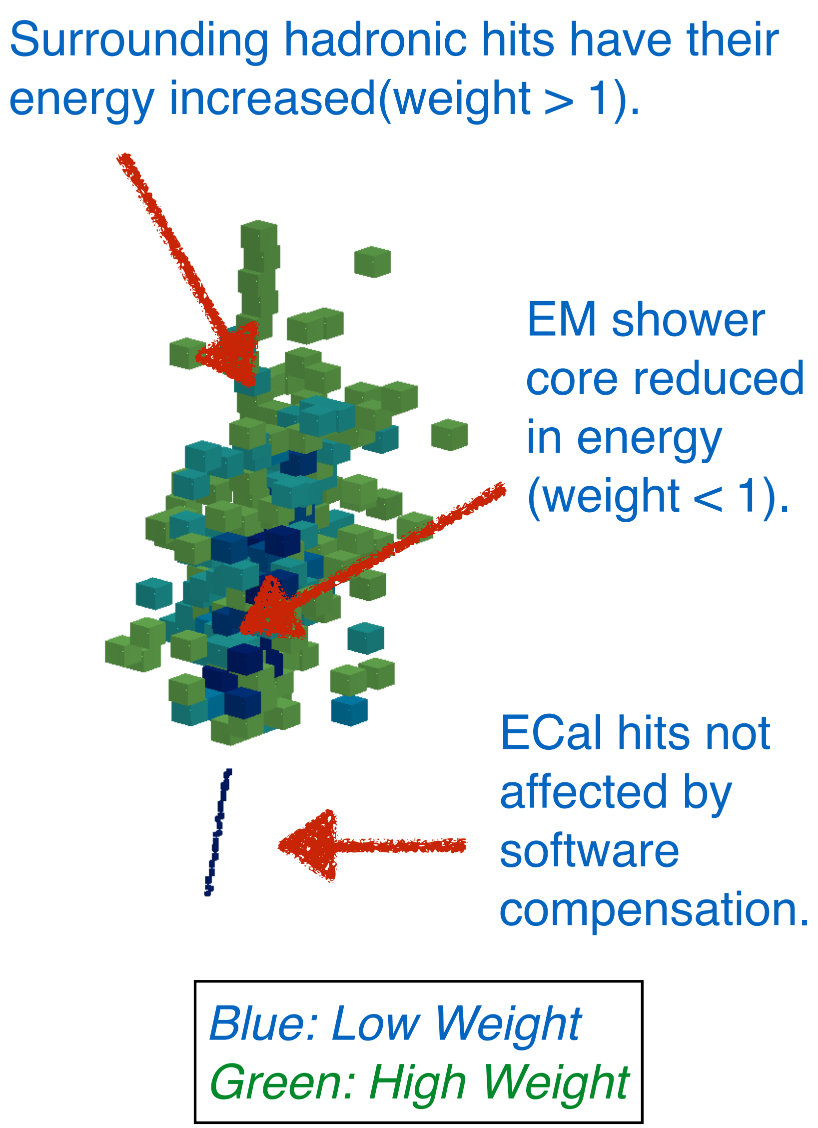
Detector optimisation studies (Cambridge/DESY):





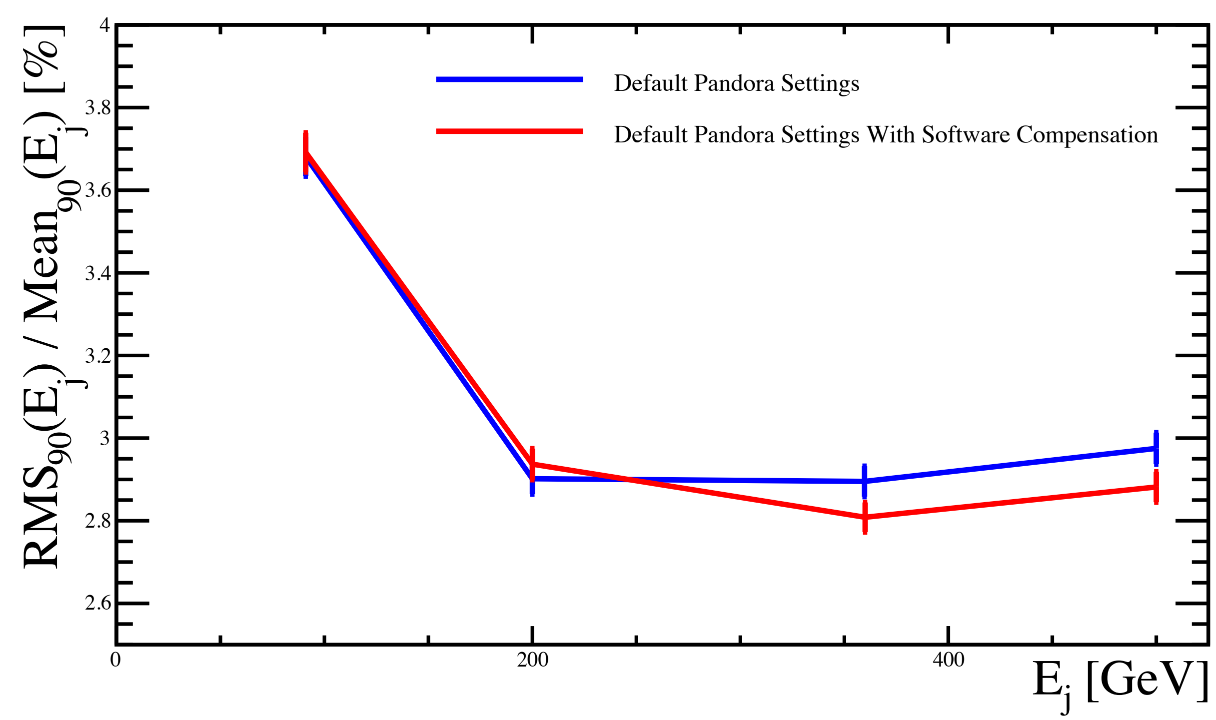
S. Green

Cambridge: Pandora LC

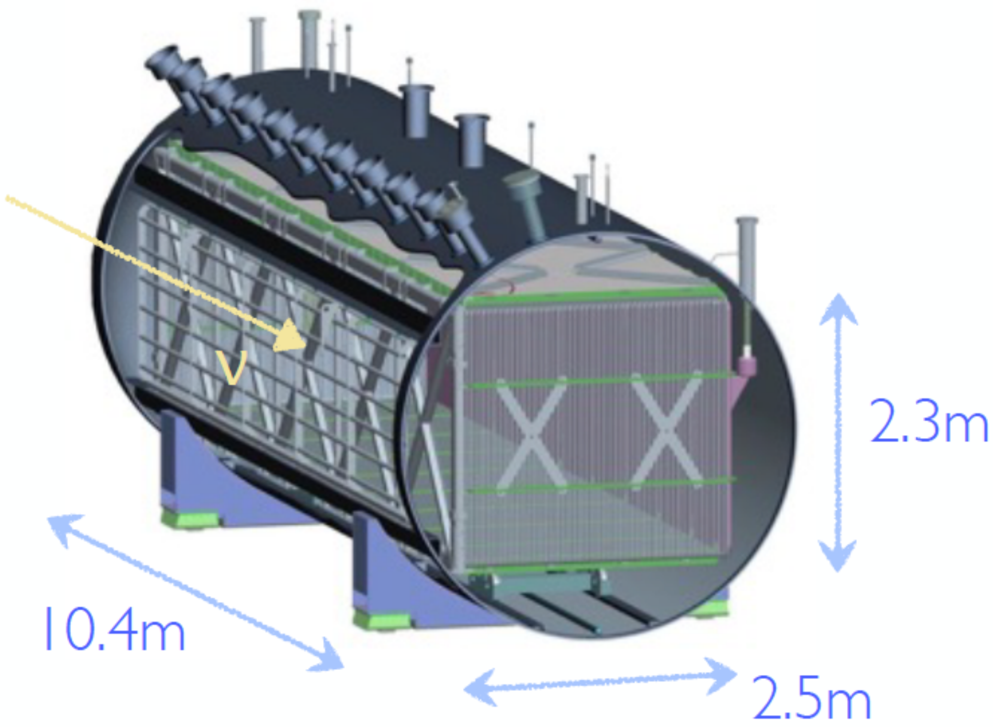


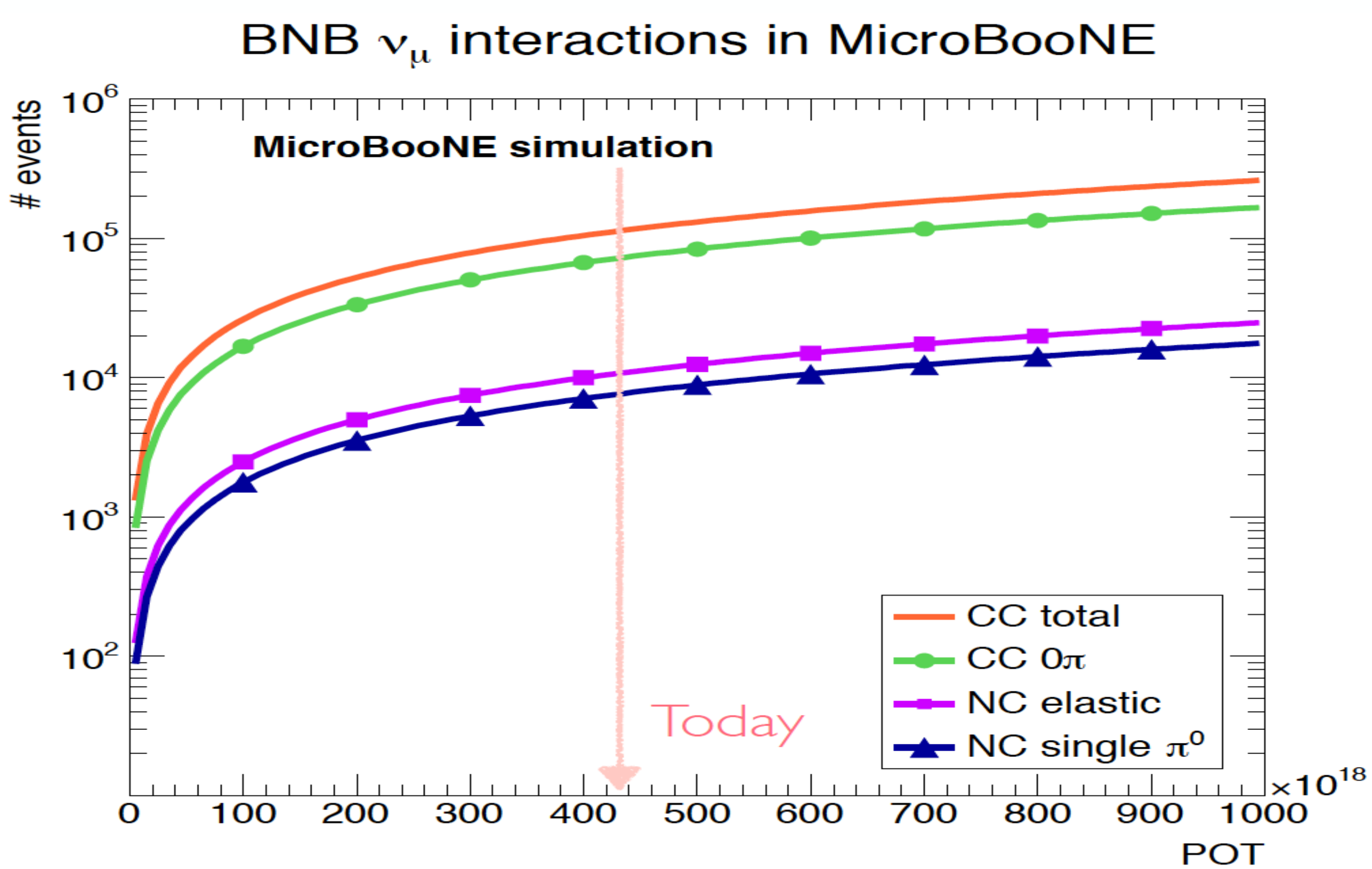
Event display with colour indicating weight applied in software compensation. Cluster from 45 GeV jet.

- **Goal:** Improve the energy estimators for hadronic clusters via a reweighting technique based on hit energy density.
- Compensate for "invisible" energy component of hadronic showers such as low energy neutrons and nuclear binding energy losses.



Jet energy resolution as a function of jet energy with and without software compensation. Improvement seen when using software compensation for high energy jets.





MicroBooNE in numbers

- 89 tons of LAr active volume
- 470m from BNB target
- 3 planes of ~3000 wires each
- 3 mm wire pitch
- 32(+4) PMTs
- Installed in summer 2015
- Data taking since fall 2015
- First year of data taking:
  - 3.6 E20 POT
  - ~100k neutrino events

Recent detector paper (accepted in JINST) available here: https://arxiv.org/pdf/1612.05824v2.pdf

MicroBooNE TPC

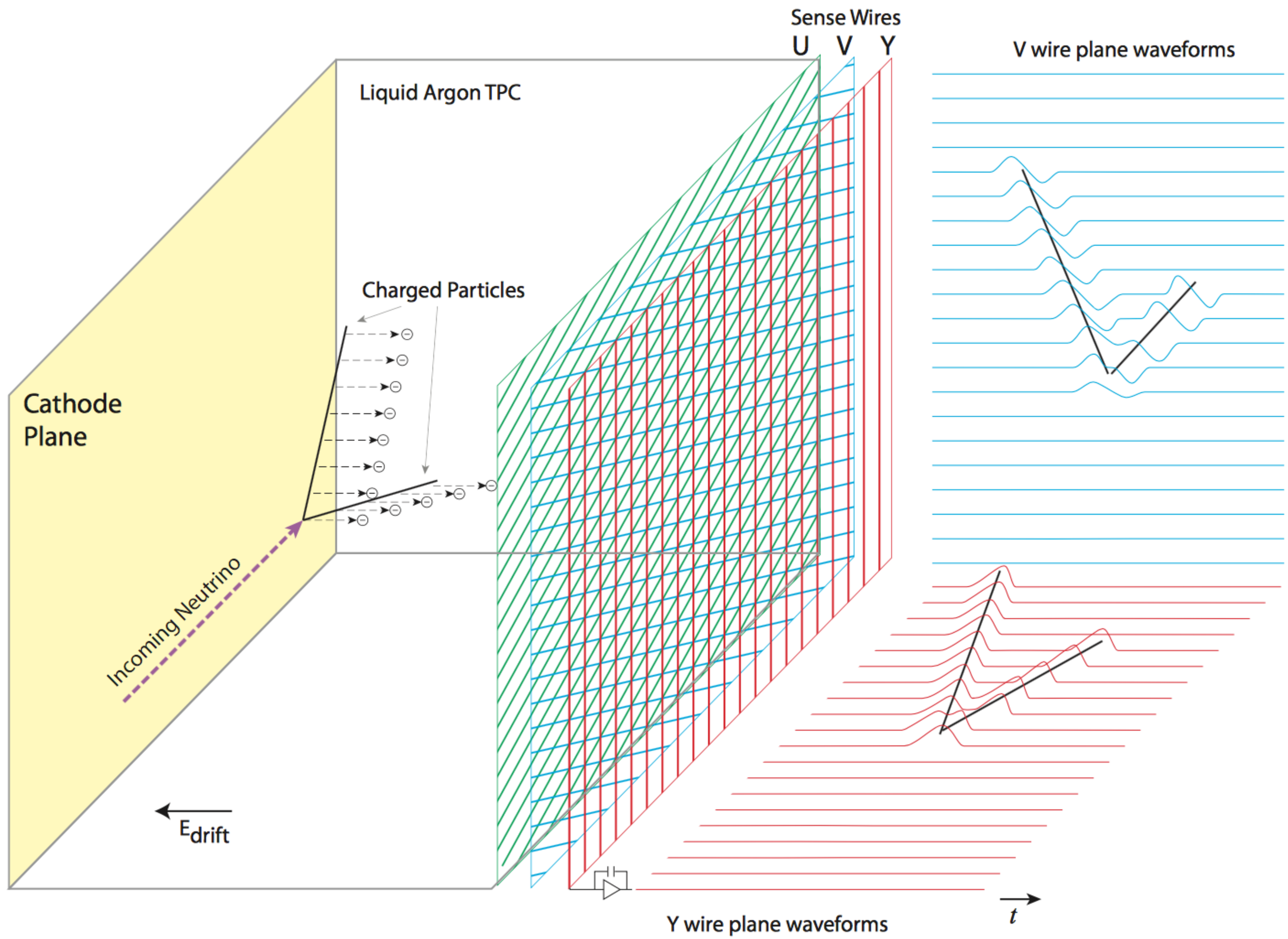
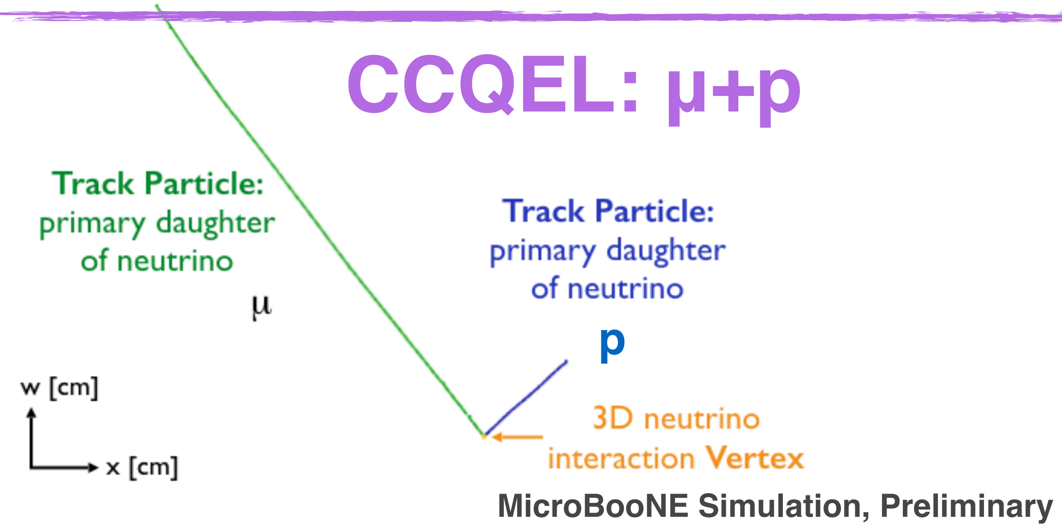


Figure 2. Operational principle of the MicroBooNE LArTPC.

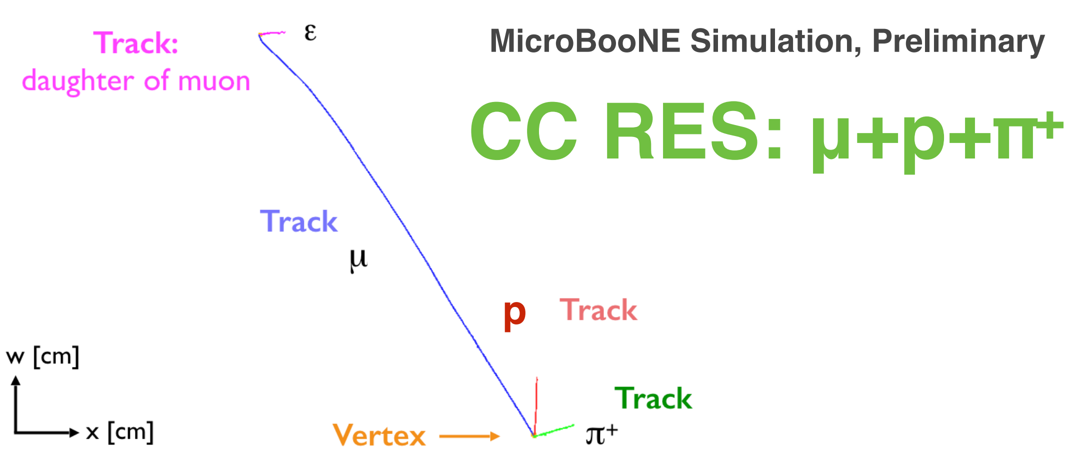
Cambridge: Pandora MicroBooNE

CCQEL: μ+p



#Matched Particles	0	1	2	3+
μ	2.7 %	90.7 %	6.1 %	0.5 %
p	19.8 %	76.4 %	3.5 %	0.3 %

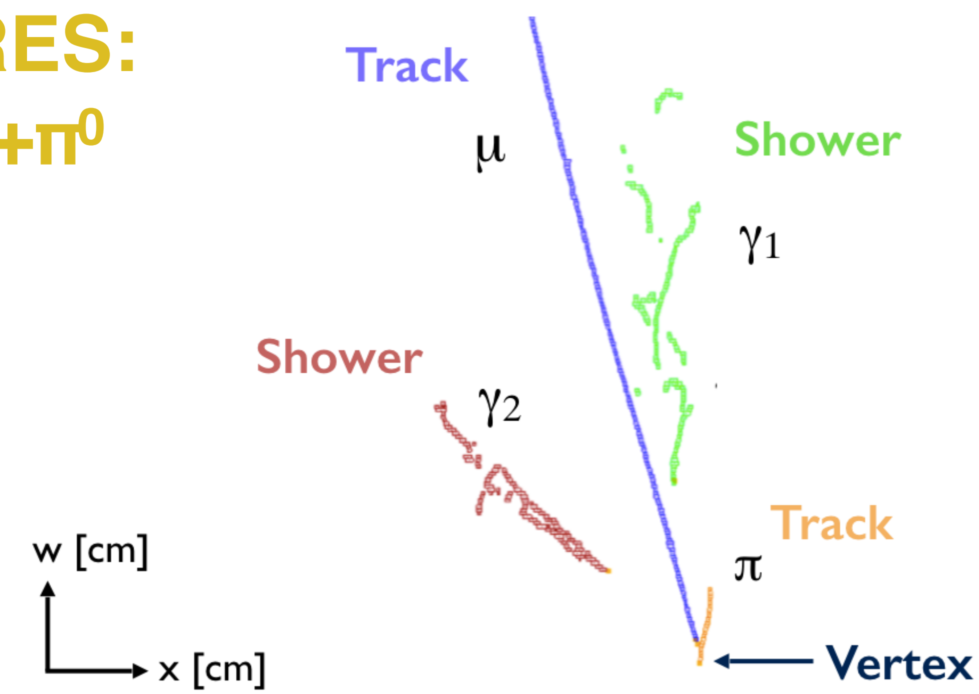
CC RES: μ+p+π⁺



#Matched Particles	0	1	2	3+
μ	6.8 %	87.7 %	5.1 %	0.4 %
p	20.5 %	75.0 %	4.0 %	0.5 %
π^+	11.6 %	71.3 %	13.0 %	4.1 %

Examples of matching true-reconstructed particles for different interaction types (Results as of summer 2016)

CC RES: μ+p+π⁰



MicroBooNE Simulation, Preliminary

#Matched Particles	0	1	2	3+
μ	7.9 %	87.4 %	4.4 %	0.3 %
p	19.8 %	74.0 %	5.5 %	0.7 %
γ_1	10.8 %	60.0 %	18.2 %	11.0 %
γ_2	35.9 %	51.0 %	10.1 %	3.0 %

Improved performance since summer, and in other final state channels will be publicly available soon!

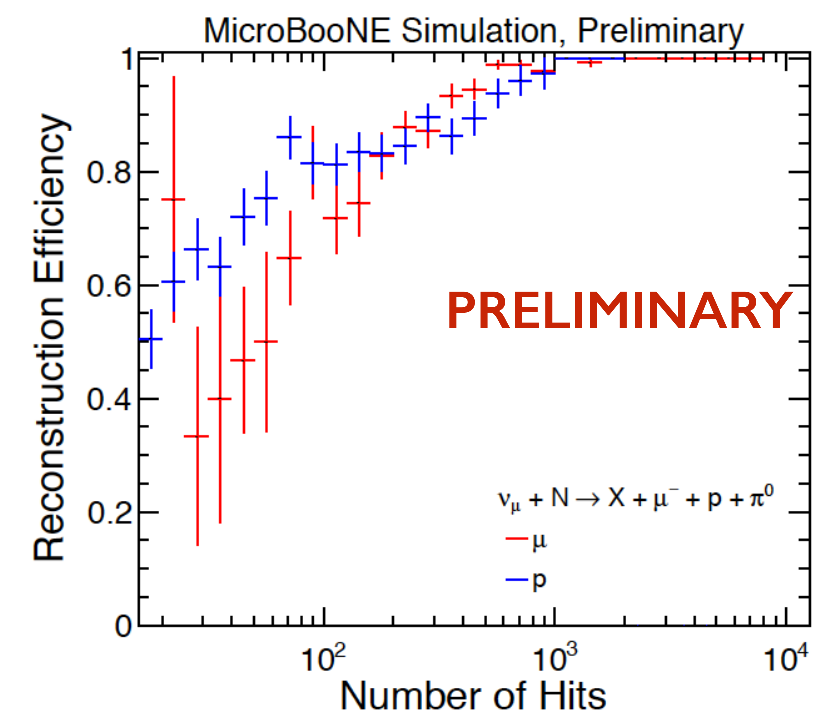
Cambridge: Pandora MicroBooNE

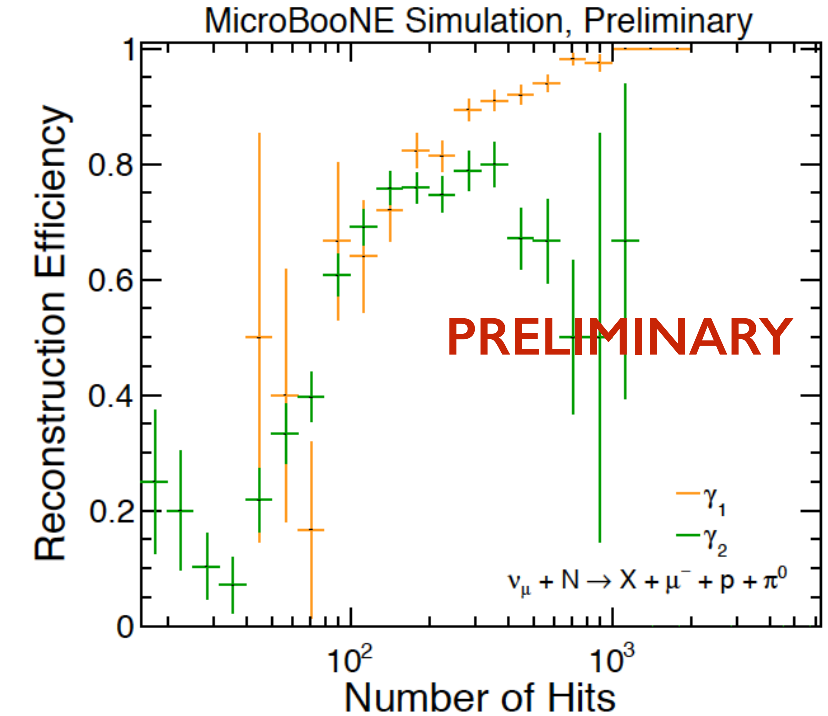
Reconstruction efficiency per particle in different interaction types

Also in the public note and backup slides showing:
- Efficiency vs #reconstructed hits
- Completeness, purity, and vertex resolution

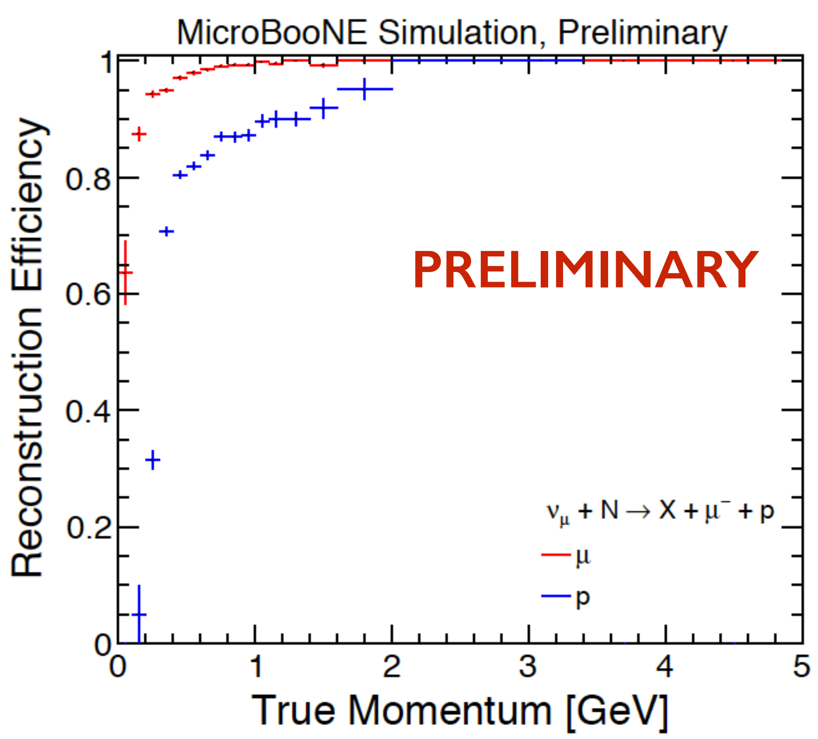
Updated plots available soon!

CC RES: μ+p+π⁰

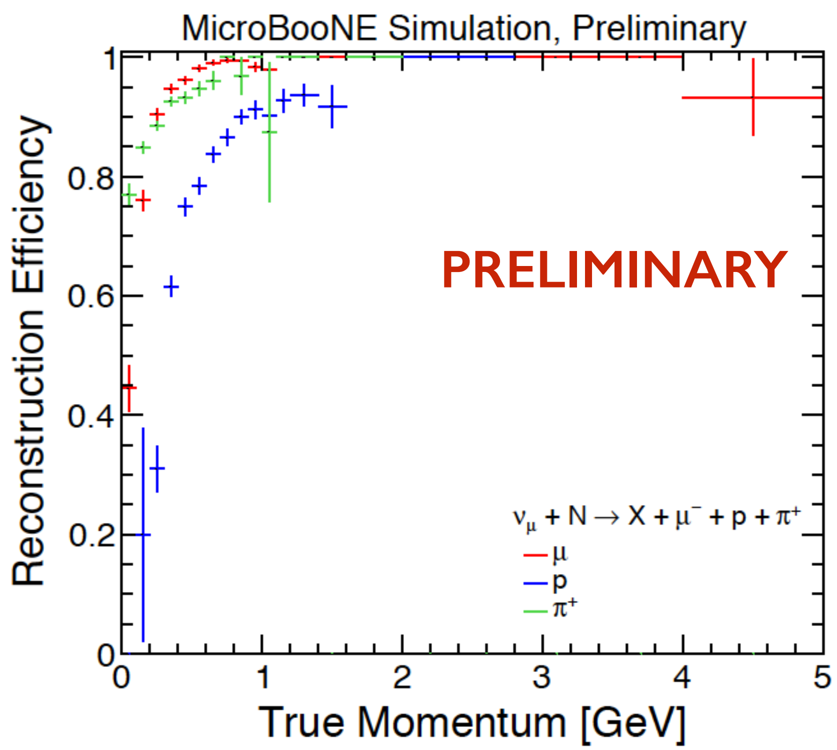




CCQEL: μ+p



CC RES: μ+p+π⁺

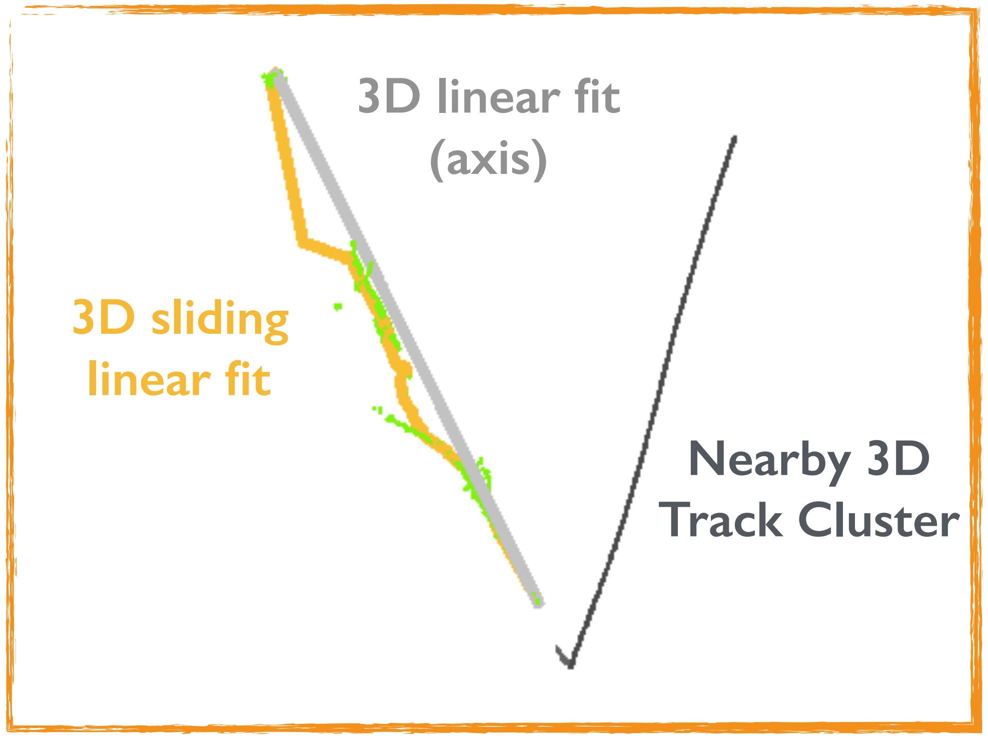


Cambridge: Pandora MicroBooNE

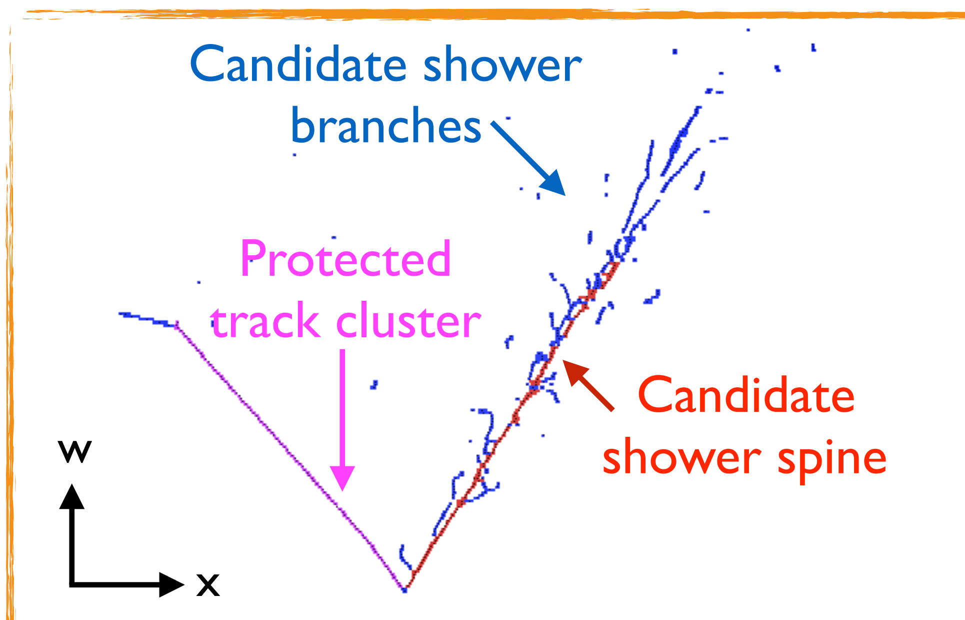
AIDA 2020

Some of the algorithm improvements since summer

Shower Completeness:
3D Sliding Cone Fits

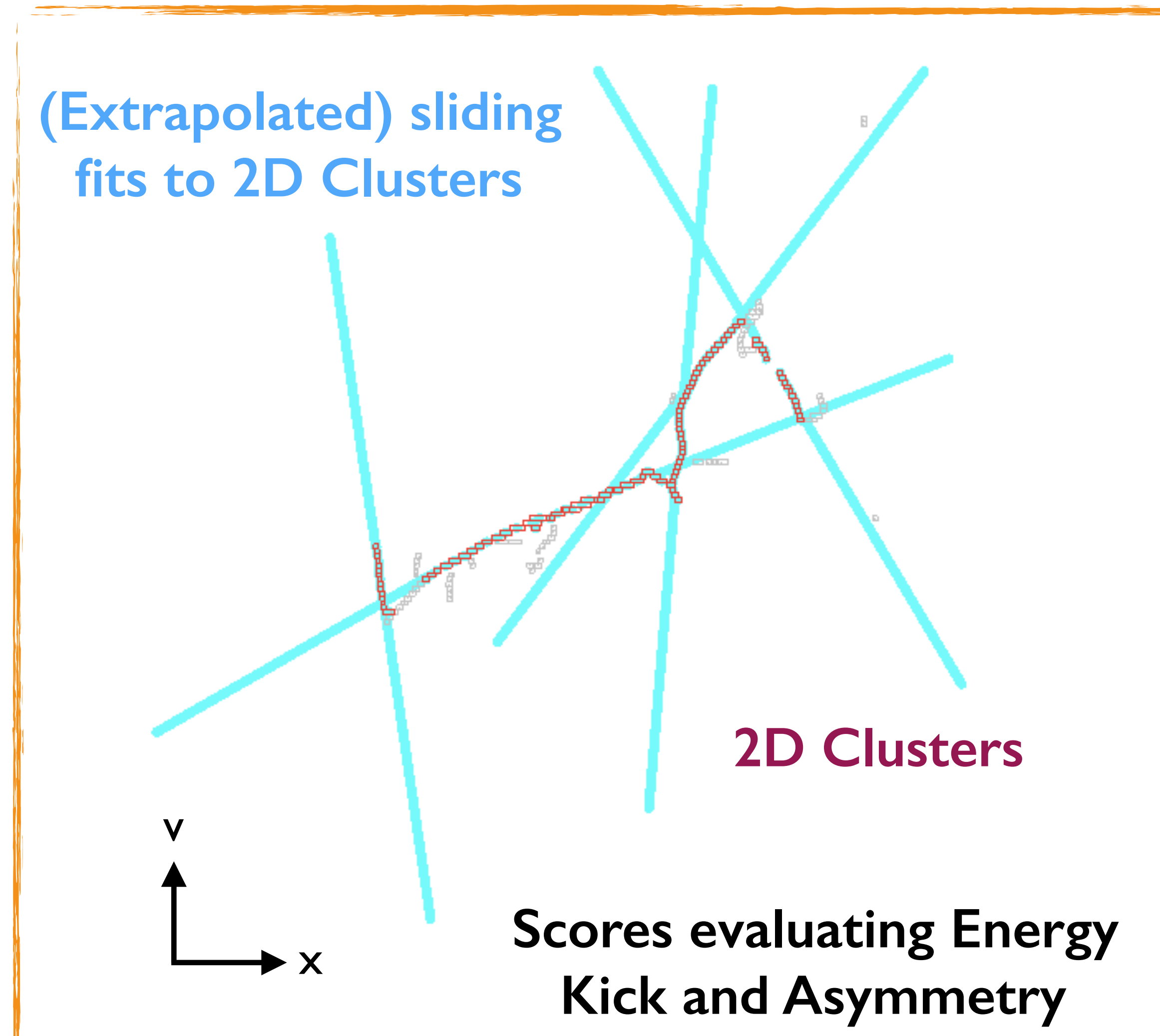


Track/Shower ID:
Tension track/shower reco chain



Separation based on broad topological properties for the moment

Vertex Improvements:
Additional vertex candidates + improved vertex scoring

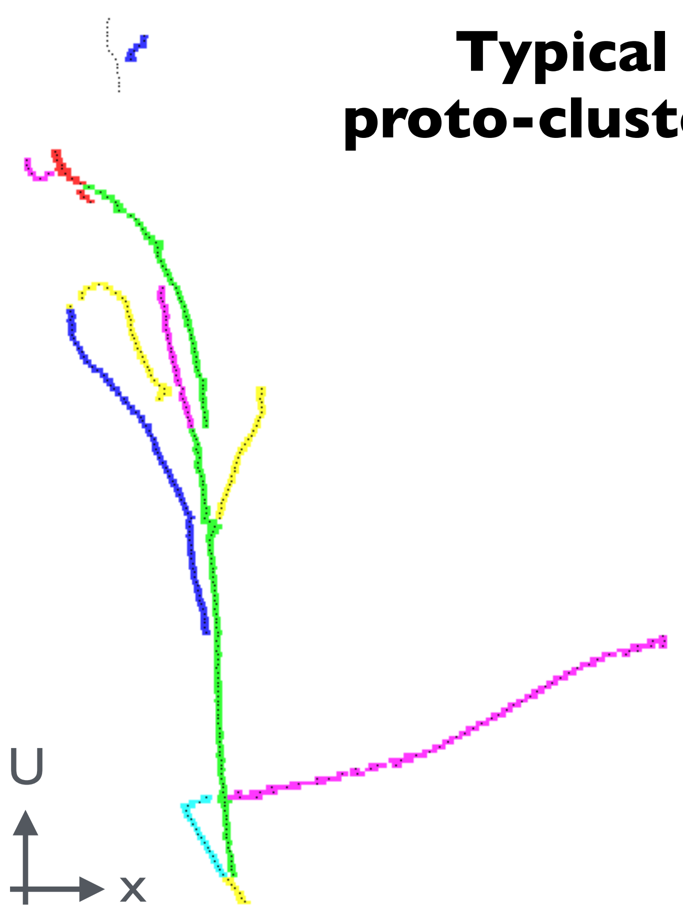


Pandora LAr TPC Example

Typical example: Begin to solve the LAr TPC pattern recognition problem...

- Start with cautious, track-oriented 2D clustering. Important to avoid mistakes at this early stage.
- Use series of topological-association algorithms to carefully merge/split the 2D proto-clusters.

Typical 2D proto-clusters



Snippet from PandoraSettings XML file:

```
<algorithm type = "LArClusteringParent">
    <algorithm type = "LArTrackClusterCreation" description = "ClusterFormation"/>
    <InputCaloHitListName>CaloHitListW</InputCaloHitListName>
    <ClusterListName>ClustersW</ClusterListName>
    <ReplaceCurrentCaloHitList>false</ReplaceCurrentCaloHitList>
    <ReplaceCurrentClusterList>true</ReplaceCurrentClusterList>
</algorithm>

<algorithm type = "LArLayerSplitting"/>
<algorithm type = "LArLongitudinalAssociation"/>
<algorithm type = "LArTransverseAssociation"/>
<algorithm type = "LArLongitudinalExtension"/>
<algorithm type = "LArTransverseExtension"/>
<algorithm type = "LArOvershootSplitting"/>
<algorithm type = "LArBranchSplitting"/>
<algorithm type = "LArKinkSplitting"/>
```

Cluster Creation

Cluster Merging/Splitting

Pandora LAr TPC Example

The Pandora LAr TPC algorithms cluster 2D hits, then group 3x2D clusters into particles.
3D spacepoints are created for each particle and the full particle hierarchy is reconstructed:

5 GeV ν_e CC: Display 1/4

The reconstructed neutrino particle contains:

- Metadata: PDG code, 4-momentum, etc
- A 3D interaction vertex
- A list of daughter particles

3D neutrino interaction vertex

Pandora LAr TPC Example

The Pandora LAr TPC algorithms cluster 2D hits, then group 3x2D clusters into particles.
3D spacepoints are created for each particle and the full particle hierarchy is reconstructed:

5 GeV ν_e CC: Display 2/4

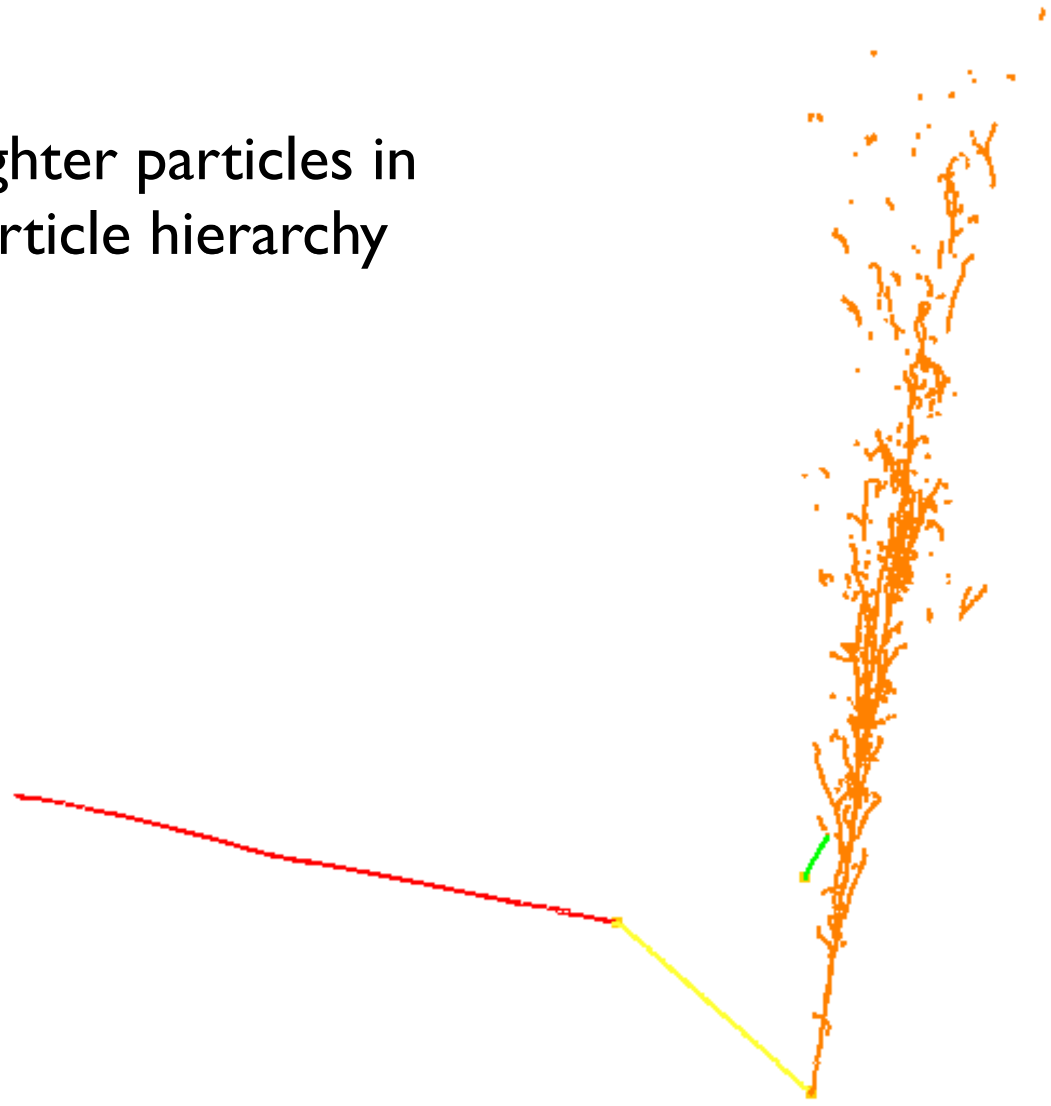
+ Primary daughter particles of the neutrino, each of which has:
  - Particle metadata
  - A list of 2D clusters and a 3D cluster
  - A 3D interaction vertex
  - A list of any further daughter particles

Pandora LAr TPC Example

The Pandora LAr TPC algorithms cluster 2D hits, then group 3x2D clusters into particles.
3D spacepoints are created for each particle and the full particle hierarchy is reconstructed:

5 GeV ν_e CC: Display 3/4

+ Complete list of daughter particles in the reconstructed particle hierarchy

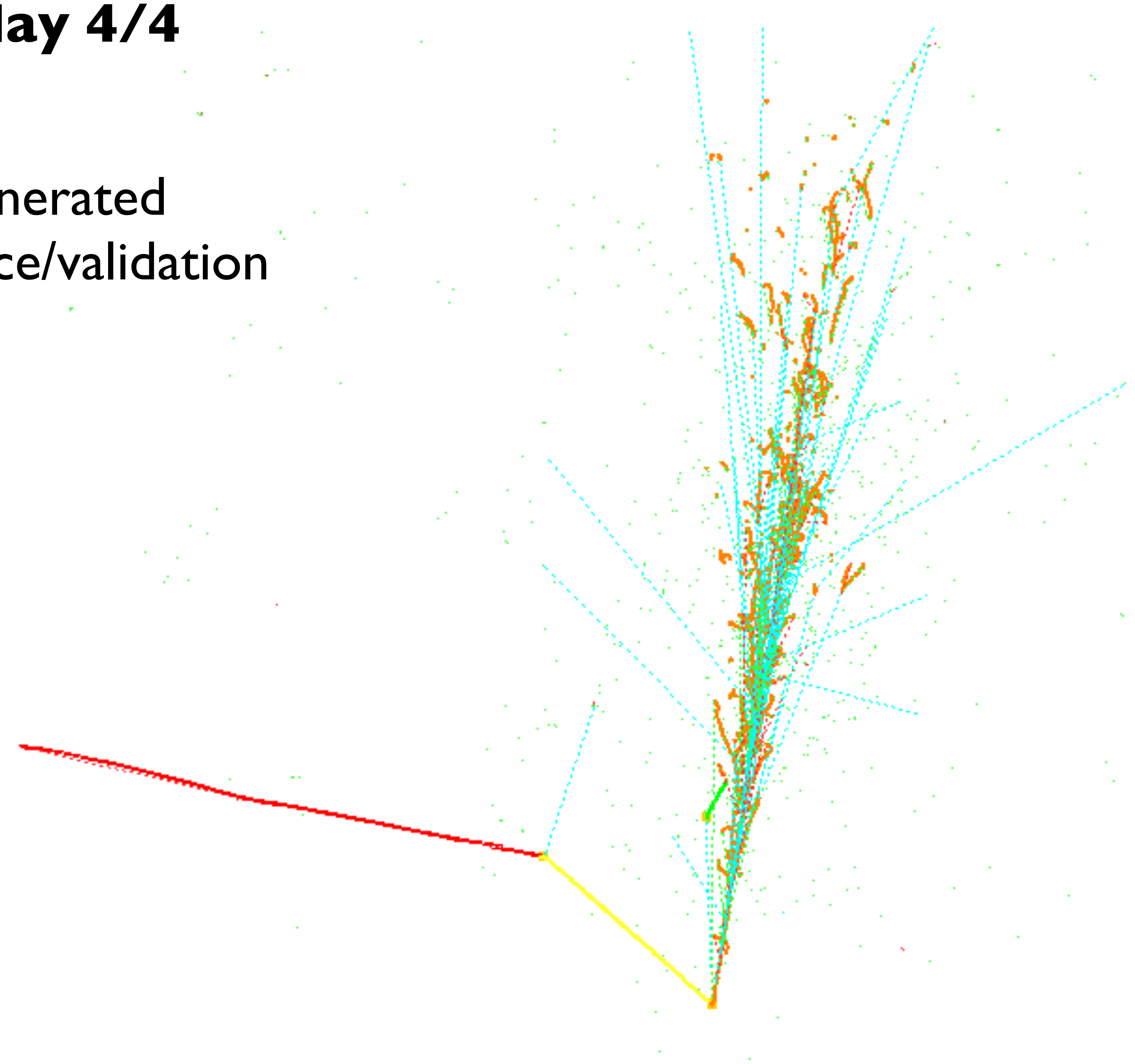


Pandora LAr TPC Example

The Pandora LAr TPC algorithms cluster 2D hits, then group 3x2D clusters into particles.
3D spacepoints are created for each particle and the full particle hierarchy is reconstructed:

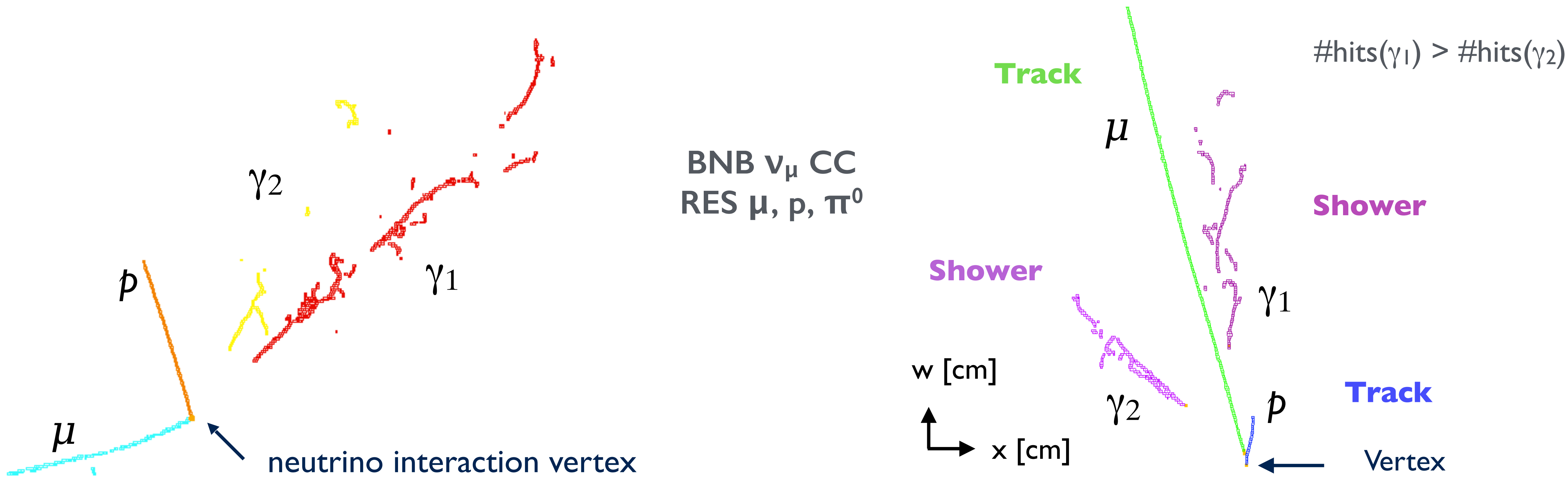
5 GeV ν_e CC: Display 4/4

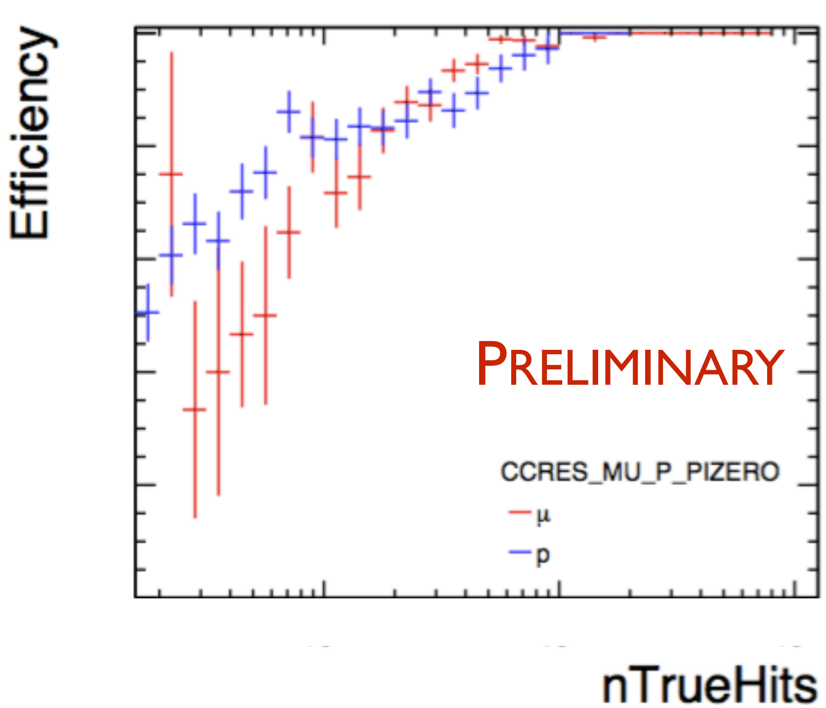
+ Overlay details of generated particles, for reference/validation

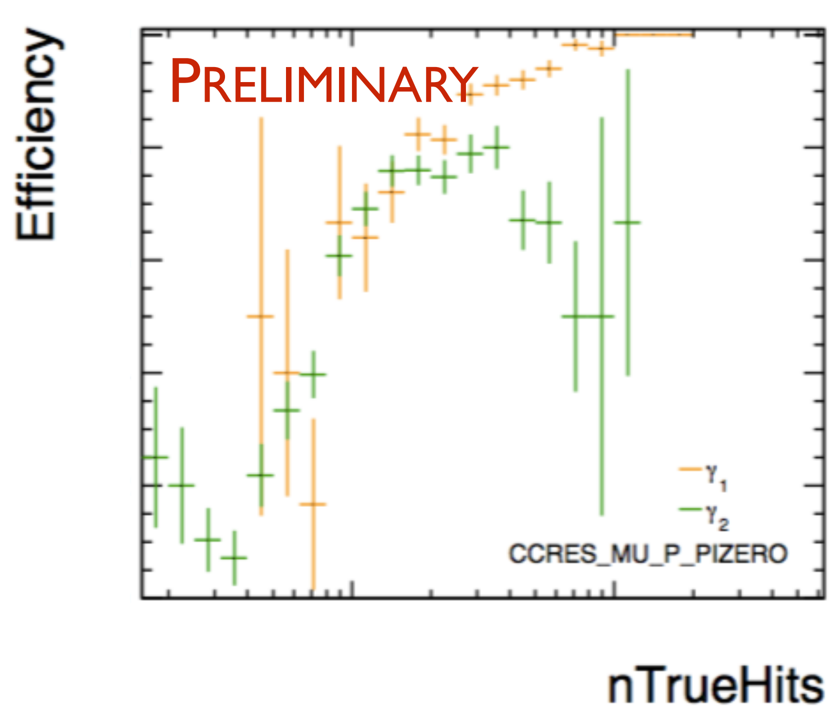


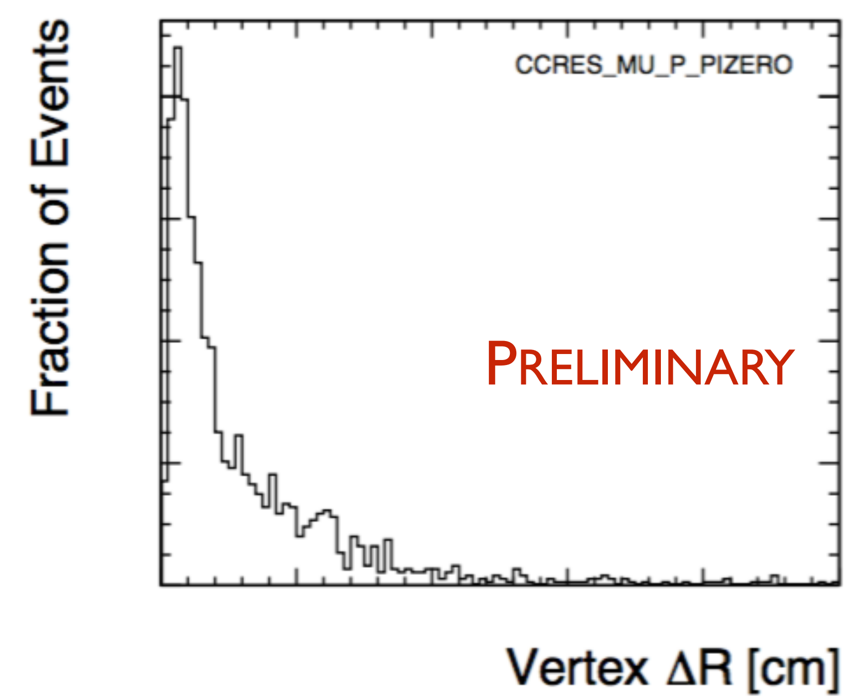
Cambridge: Pandora LAr TPC

Significant effort to assess and improve reconstruction of events with single π^0s







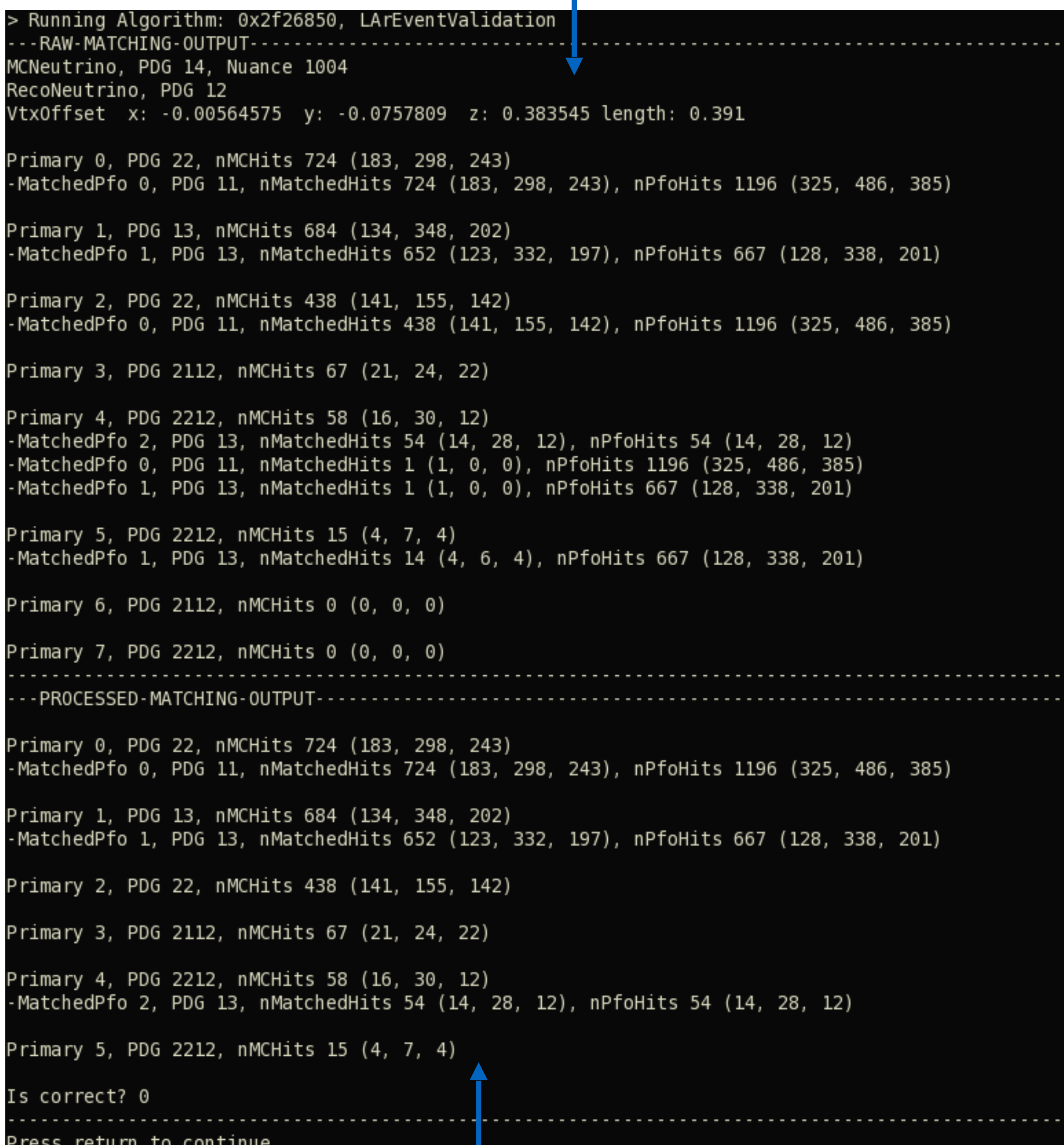


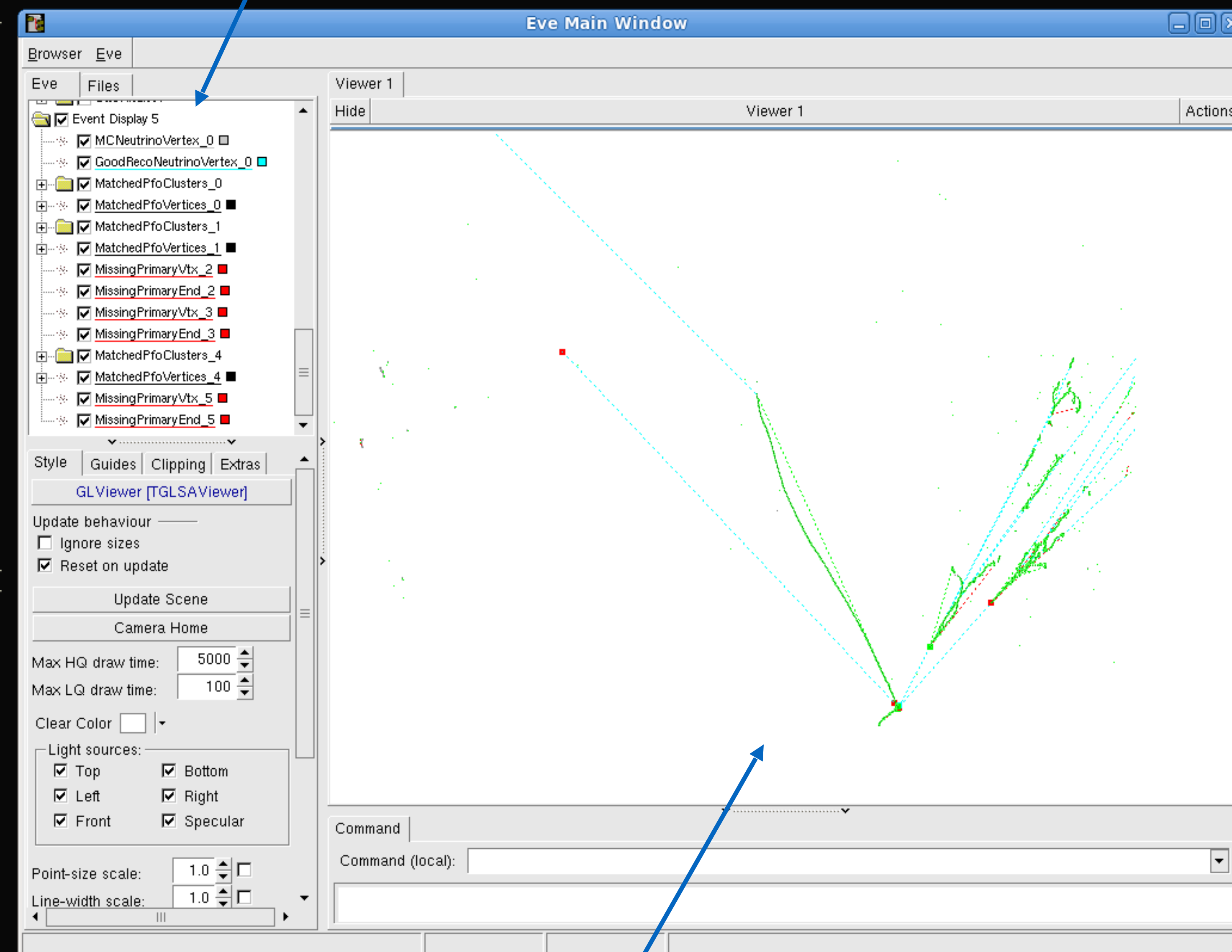
Cambridge: Pandora LAr TPC

New interactive event visualisation and validation tools to drive reconstruction development:

Comprehensive list of reco particle to MC primary matches

Clickable list of objects for each MC primary
Indication of reco quality: "angry" colours for poor matches



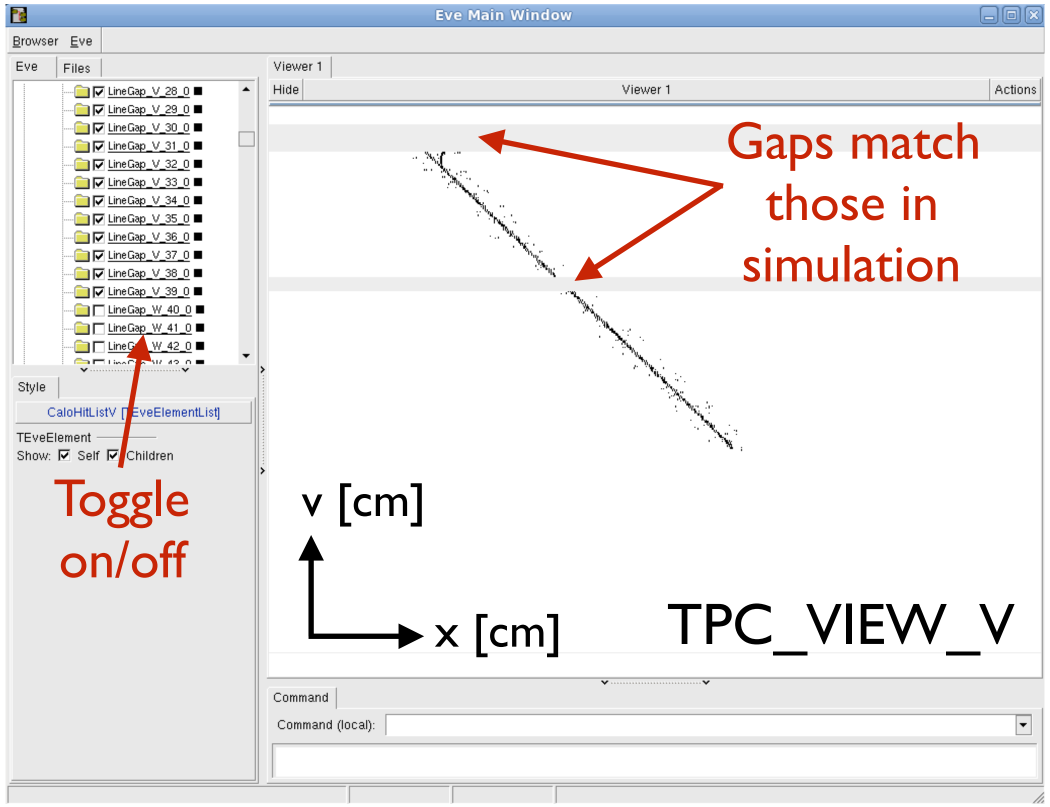


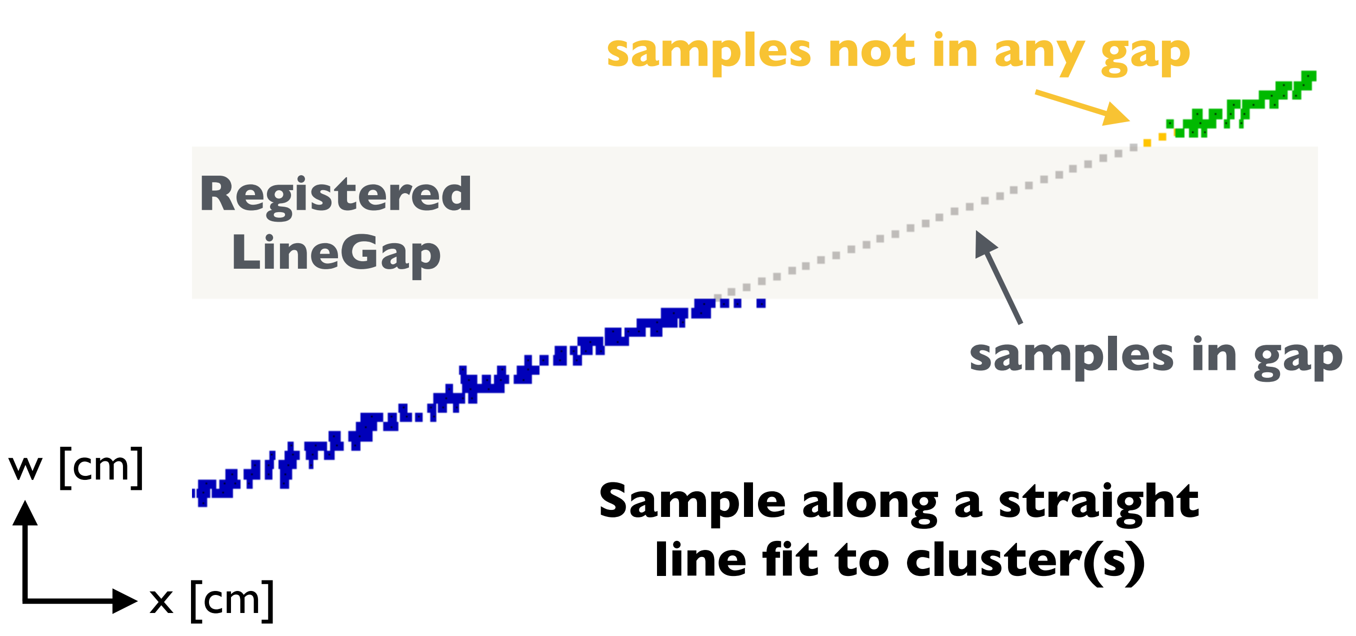
Output of our interpretative matching scheme

Matched particles appear in green
Split particles appear in red
Red markers placed at vtx/end of missing particles

Cambridge: Pandora SDK

- **Recently added support for a new kind of gap to Pandora SDK (existing Box & ConcentricGaps).**
- **LineGaps** can represent continuous sets of wires with static bad channel status in a LAr TPC.
- Gaps defined in LArSoft and information is then available for use in Pandora algs and visualisation.
- Idea is that a few algs can specifically allow e.g. 2D cluster merges across (large) gap regions.





- **E.g. Broken single muon clusters:**
  - Good 'pointing' information, but a significant cluster separation.
  - Algs can now find out whether separation is due to a known gap.
  - Can confidently merge clusters.

Cambridge: Pandora LAr TPC

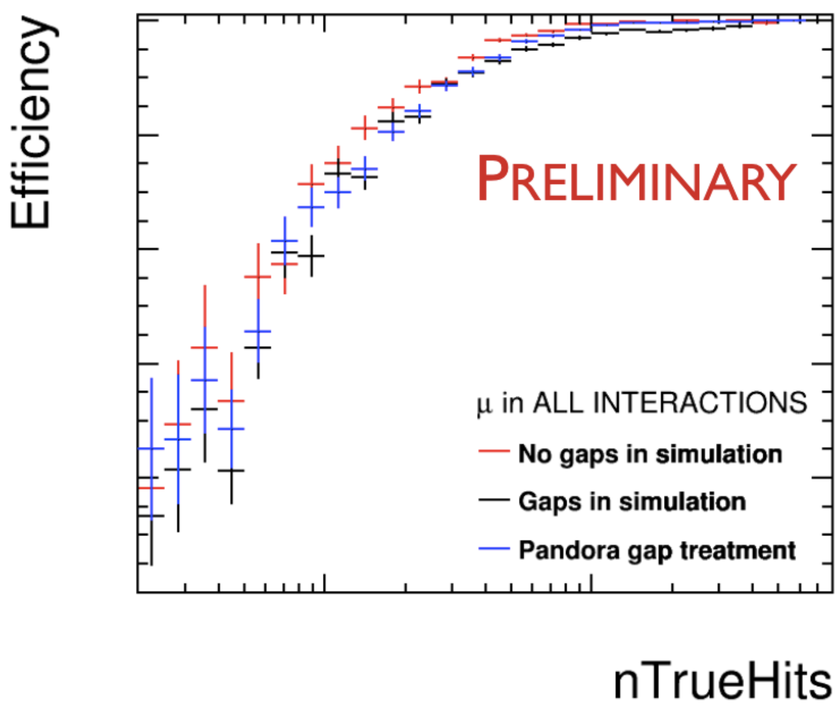
Creation

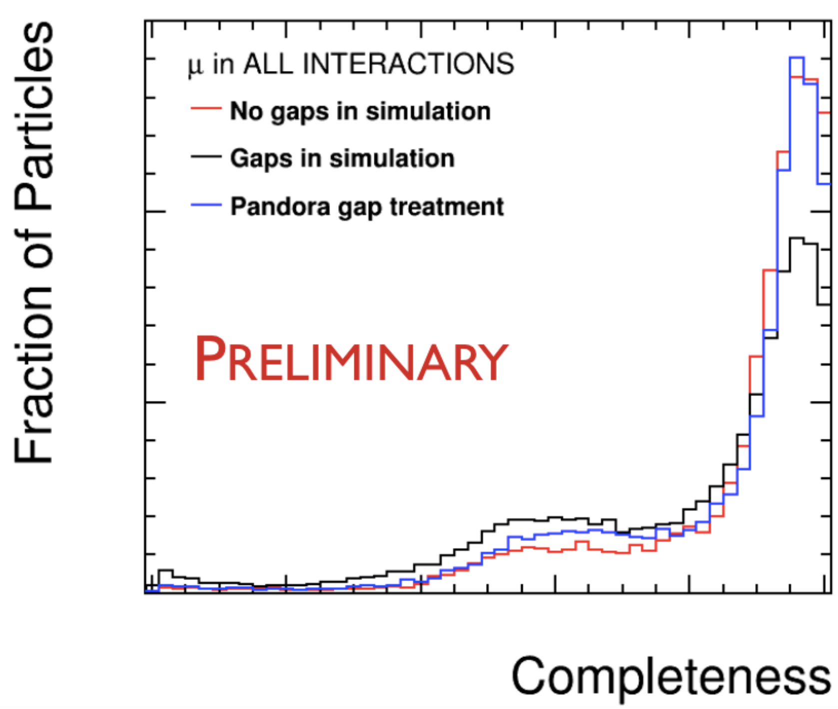
```
PandoraApi::Geometry::LineGap::Parameters parameters;
parameters.m_hitType = my_hit_type; // e.g. TPC_VIEW_U, TPC_VIEW_V, TPC_VIEW_W
parameters.m_lineStartZ = my_gap_start_position; // e.g. u, v, or w wire position
parameters.m_lineEndZ = my_gap_end_position; // e.g. u, v, or w wire position
PANDORA_THROW_RESULT_IF(STATUS_CODE_SUCCESS, !=, PandoraApi::Geometry::LineGap::Create(*pPandora, parameters));
```

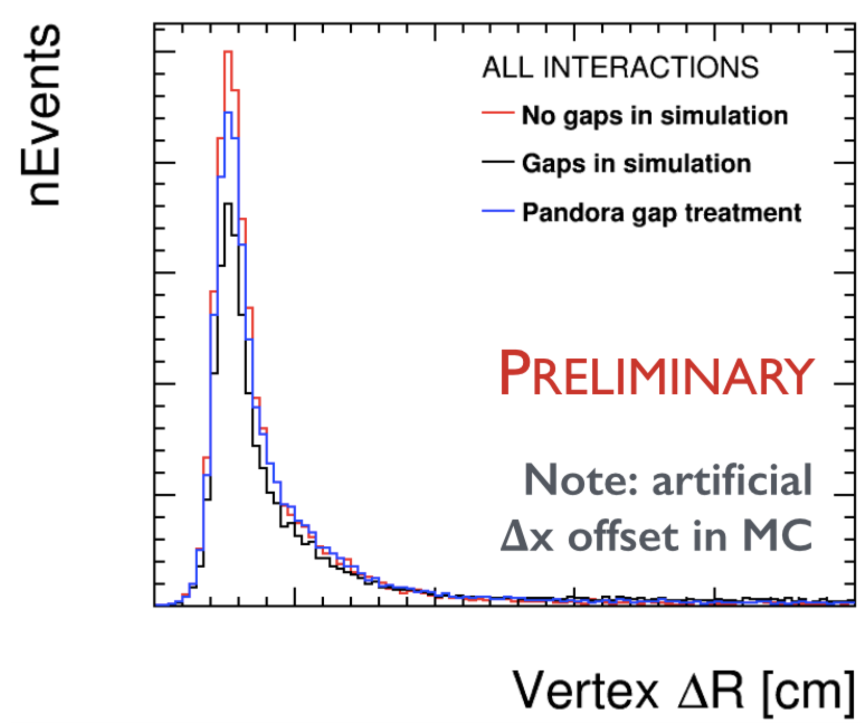
Interface

```
/**
 *  @brief  Whether a specified position lies within the gap
 *
 *  @param  positionVector the position vector
 *  @param  hitType the hit type, providing context to aid interpretation of provided position vector
 *  @param  gapTolerance tolerance allowed when declaring a point to be "in" a gap region, units mm
 *
 *  @return boolean
 */
virtual bool IsInGap(const CartesianVector &positionVector, const HitType hitType, const float gapTolerance = 0.f) const = 0;
```

Use of gap information in LAr TPC algorithms enables significant performance recovery:



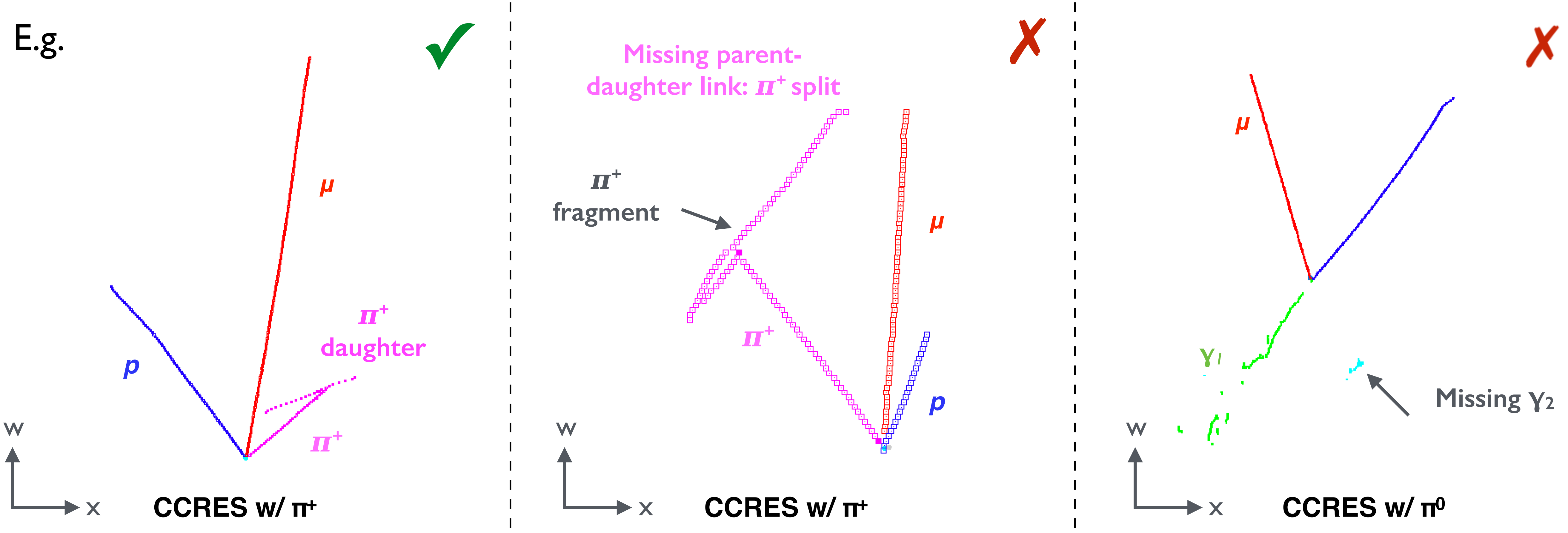




Pandora Performance Metrics

- **Will present the fraction of events deemed "correct" by Pandora performance metrics:**
  - Consider exclusive final-states where all true particles pass simple quality cuts (e.g. nHits)
  - Correct means exactly one reco primary particle is matched to each true primary particle
  - Well-defined (see uboone note and backup); ***very strict*** assessment of pattern recognition

E.g.



Pandora Very strict performance Metrics

Example of Pandora MC-reco matching output for DUNE 10kton

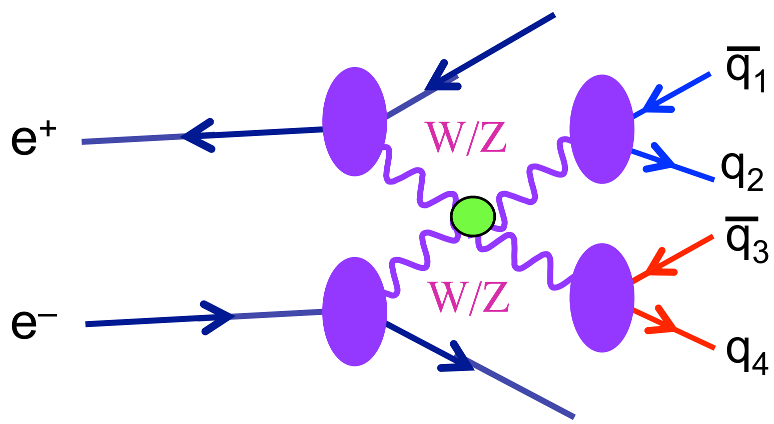
```
---PROCESSED-MATCHING-OUTPUT-------------------------------------------------------------------
MinPrimaryGoodHits 15, MinHitsForGoodView 5, MinPrimaryGoodViews 2
UseSmallPrimaries 1, MinSharedHits 5, MinCompleteness 0.1, MinPurity 0.5

Primary 0, PDG 13, nMCHits 13105 (3870, 4366, 4869), [nGood 13068 (3859, 4356, 4853)]
-MatchedPfo 0, PDG 13, nMatchedHits 11860 (3454, 3973, 4433), nPfoHits 11900 (3462, 3991, 4447)
-(Below threshold) MatchedPfo 3, PDG 11, nMatchedHits 452 (133, 153, 166), nPfoHits 452 (133, 153, 166)
-(Below threshold) MatchedPfo 4, PDG 11, nMatchedHits 435 (139, 140, 156), nPfoHits 435 (139, 140, 156)
-(Below threshold) MatchedPfo 9, PDG 11, nMatchedHits 54 (31, 11, 12), nPfoHits 54 (31, 11, 12)
-(Below threshold) MatchedPfo 12, PDG 11, nMatchedHits 37 (10, 13, 14), nPfoHits 37 (10, 13, 14)
-(Below threshold) MatchedPfo 13, PDG 11, nMatchedHits 30 (12, 12, 6), nPfoHits 30 (12, 12, 6)

Primary 1, PDG 211, nMCHits 1953 (651, 583, 719), [nGood 1940 (650, 579, 711)]
-MatchedPfo 1, PDG 13, nMatchedHits 1781 (651, 492, 638), nPfoHits 1847 (653, 539, 655)
-(Below threshold) MatchedPfo 5, PDG 11, nMatchedHits 8 (0, 8, 0), nPfoHits 8 (0, 8, 0)

Primary 2, PDG 211, nMCHits 434 (138, 131, 165), [nGood 422 (136, 126, 160)]
-MatchedPfo 2, PDG 13, nMatchedHits 375 (126, 105, 144), nPfoHits 376 (126, 106, 144)
-(Below threshold) MatchedPfo 11, PDG 11, nMatchedHits 32 (10, 9, 13), nPfoHits 36 (10, 13, 13)

Primary 3, PDG 22, nMCHits 394 (141, 128, 125), [nGood 387 (140, 126, 121)]
-MatchedPfo 10, PDG 11, nMatchedHits 80 (31, 17, 32), nPfoHits 80 (31, 17, 32)

Primary 4, PDG 22, nMCHits 185 (66, 74, 45), [nGood 178 (65, 72, 41)]

Primary 5, PDG 2212, nMCHits 36 (5, 18, 13), [nGood 25 (4, 10, 11)]

(Non target) Primary 6, PDG 2212, nMCHits 18 (9, 3, 6), [nGood 14 (7, 1, 6)]
-(Below threshold) MatchedPfo 14, PDG 11, nMatchedHits 9 (0, 3, 6), nPfoHits 26 (0, 6, 20)

(Non target) Primary 7, PDG 2212, nMCHits 11 (4, 3, 4), [nGood 9 (4, 2, 3)]
-MatchedPfo 15, PDG 11, nMatchedHits 6 (0, 3, 3), nPfoHits 6 (0, 3, 3)

Is correct? 0
```

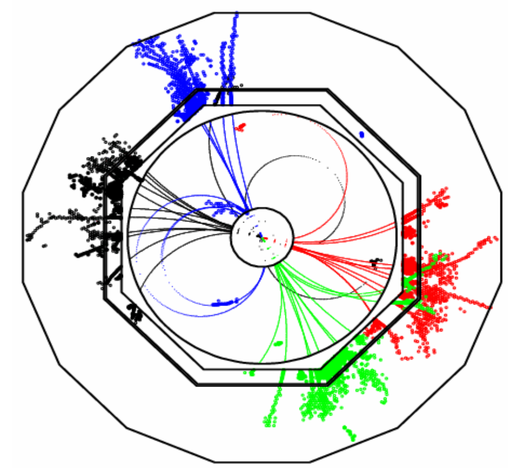
Event deemed incorrect

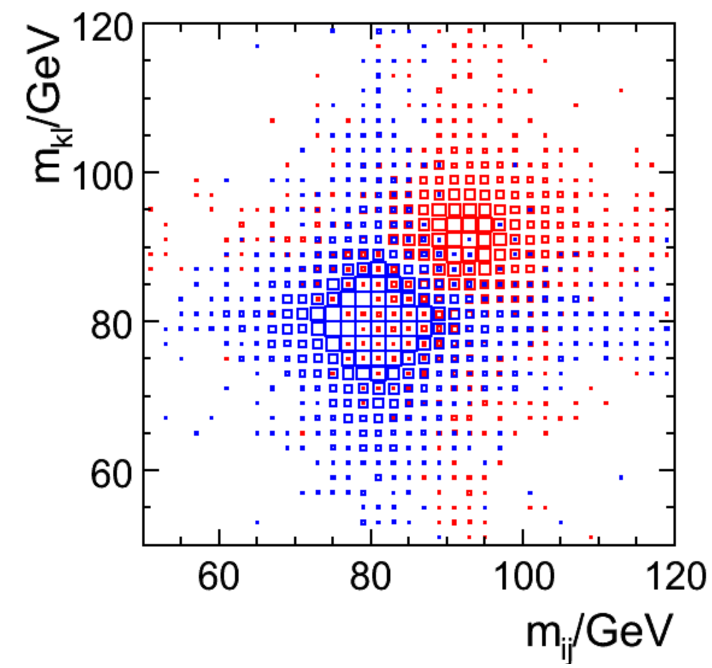
Example: correct up to here

Missing only small particles which are only a small fraction of the event

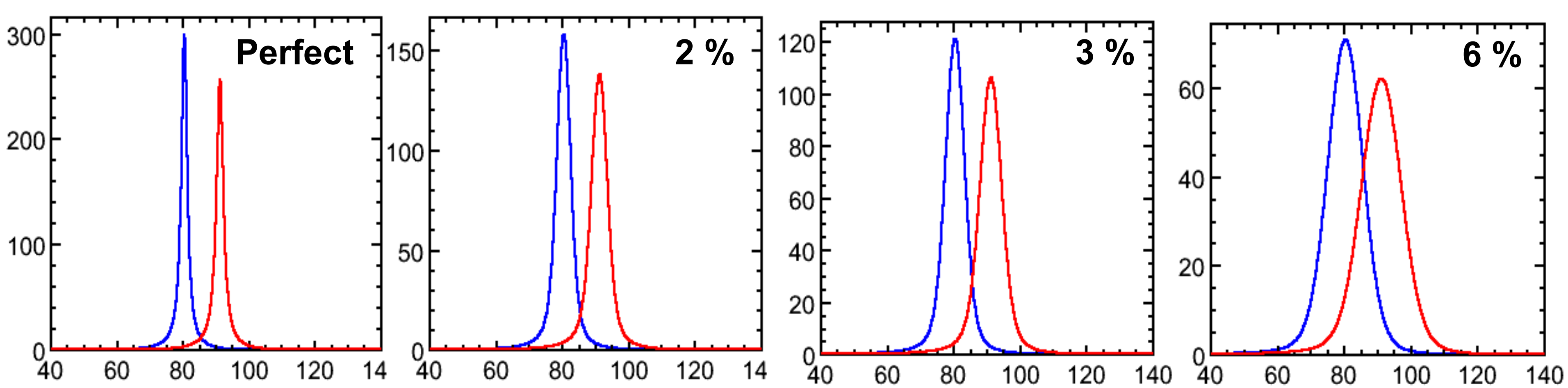
- Jet energy resolution requirements depend on physics...
- Likely to be primarily interested in di-jet mass resolution.
- Strong desire to separate W/Z hadronic decays.





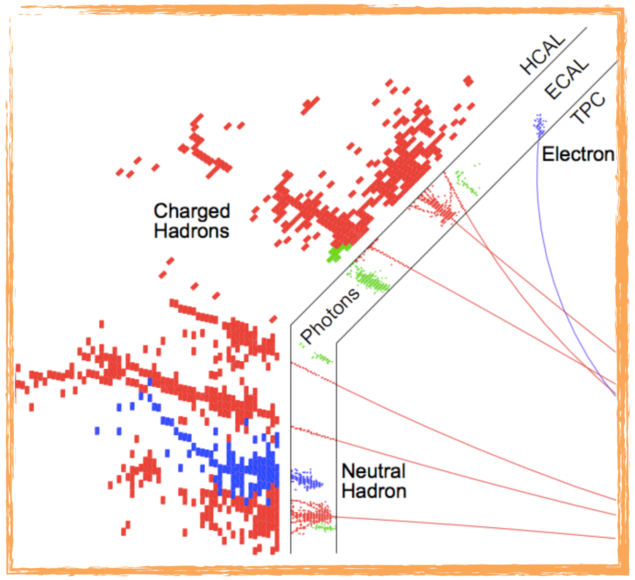


- 3-4% jet energy resolution gives decent 2.6-2.3σ W/Z separation.
- Sets a reasonable choice for LC jet energy minimal goal ~3.5%.
- For W/Z separation, not much further gain; limited by natural widths.



W/Z sep:
$(m_Z - m_W) / \sigma_m$

Jet E res.	W/Z sep
Perfect	3.1 σ
2%	2.9 σ
3%	2.6 σ
4%	2.3 σ
5%	2.0 σ
10%	1.1 σ

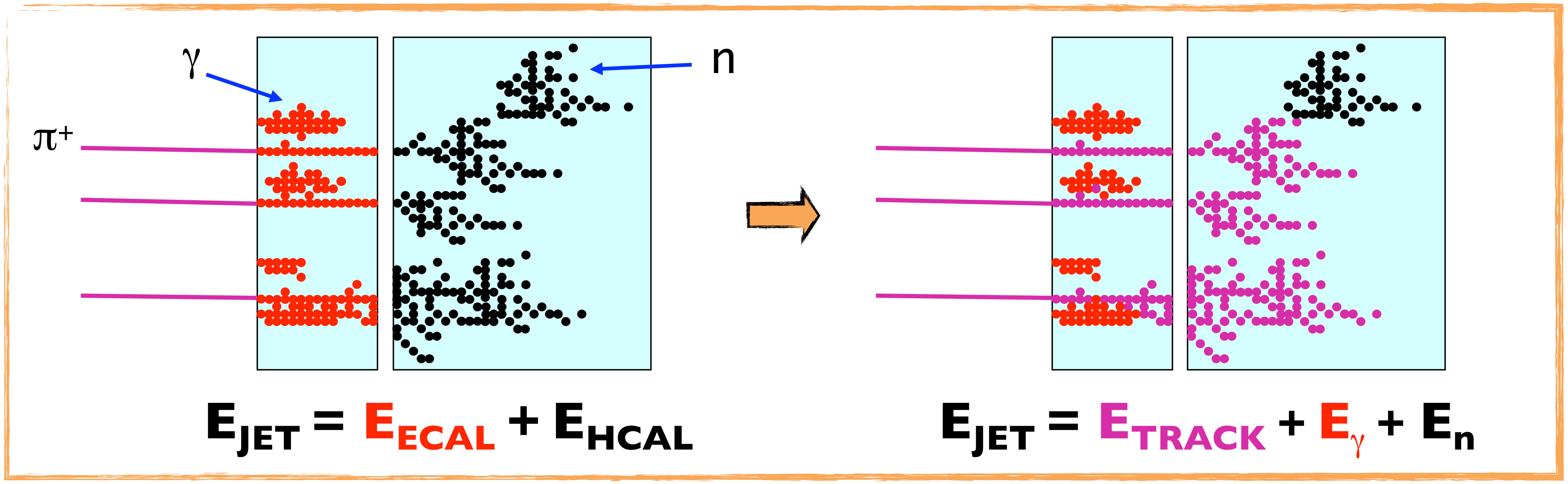


In a typical jet:

- 60 % of jet energy in charged hadrons
- 30 % in photons (mainly from $\pi^0 \rightarrow \gamma\gamma$))
- 10 % in neutral hadrons (mainly n and K_L)

Traditional calorimetric approach:

- Measure all components of jet energy in ECAL/HCAL
- Approximately 70% of energy measured in HCAL: $\sigma_E/E \approx 60\% / \sqrt{E(\mathrm{GeV})}$



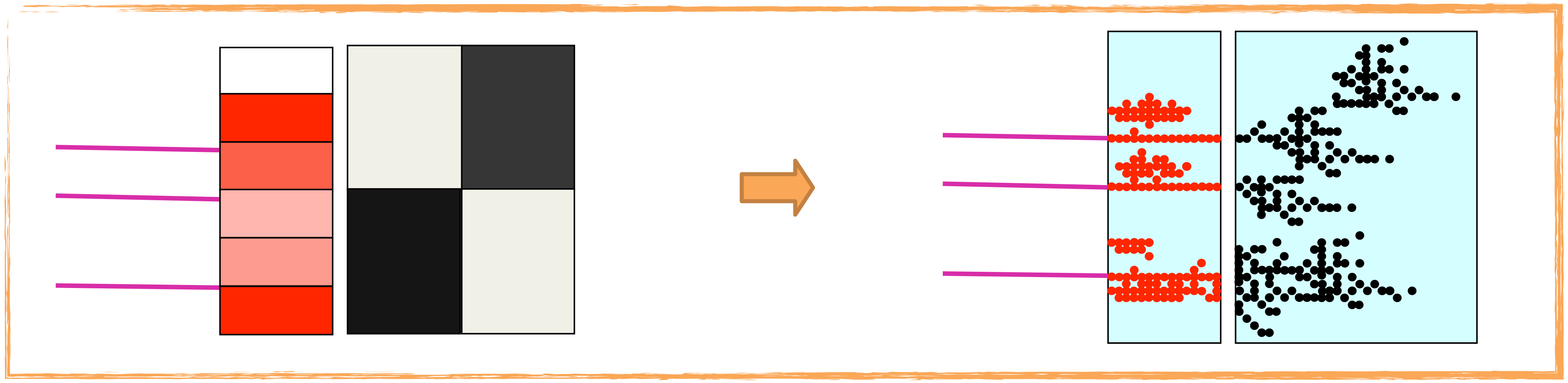
Fine granularity Particle Flow Calorimetry: reconstruct individual particles.

- Charged particle momentum measured in tracker (essentially perfectly)
- Photon energies measured in ECAL: $\sigma_E/E < 20\% / \sqrt{E(\mathrm{GeV})}$
- Only neutral hadron energies (10% of jet energy) measured in HCAL: much improved resolution.

Fine Granularity Particle Flow

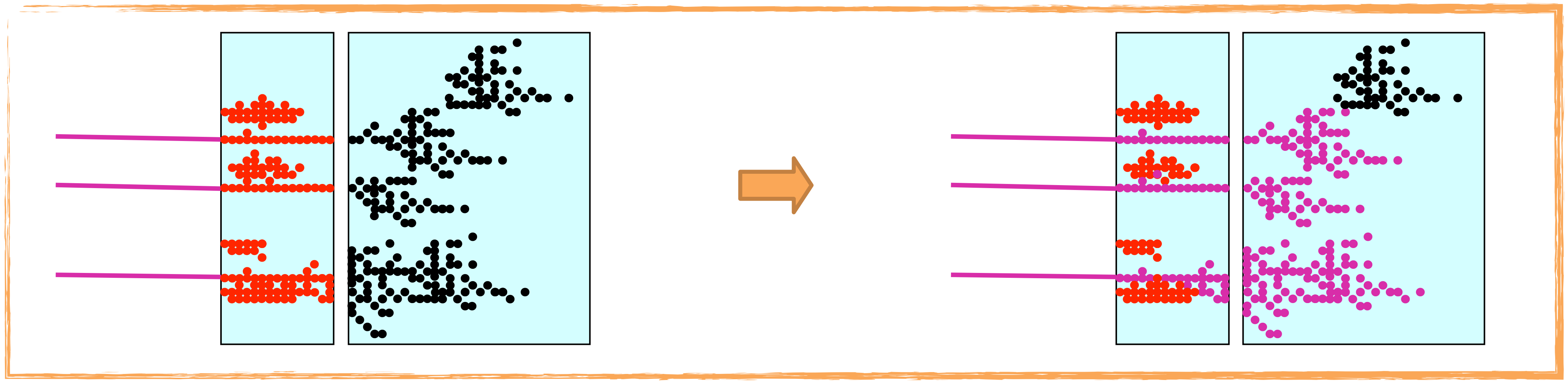
Hardware: need to be able to resolve energy deposits from different particles.

- Require highly granular detectors (as studied by CALICE).



Software: need to be able to identify energy deposits from each individual particle.

- Require sophisticated reconstruction software to deal with complex events, containing many hits.



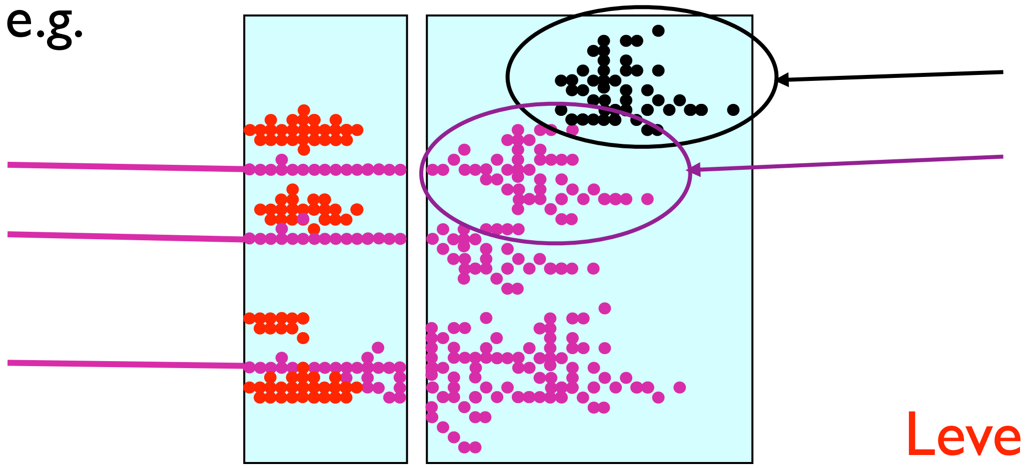
Particle Flow Calorimetry = HARDWARE + SOFTWARE

Particle Flow Algorithms

The challenge for fine granularity particle flow algorithms:

- Avoid double counting of energy from same particle
- Separate energy deposits from different particles

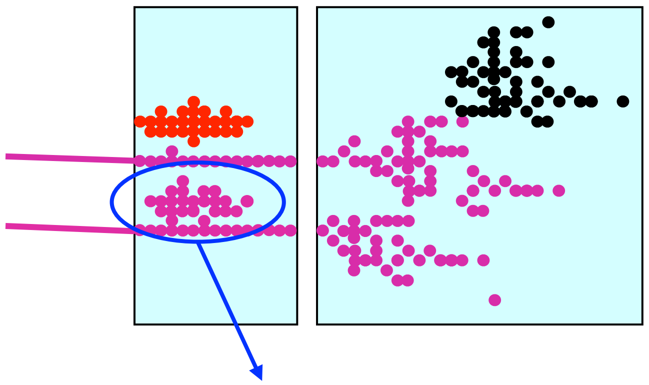
e.g.



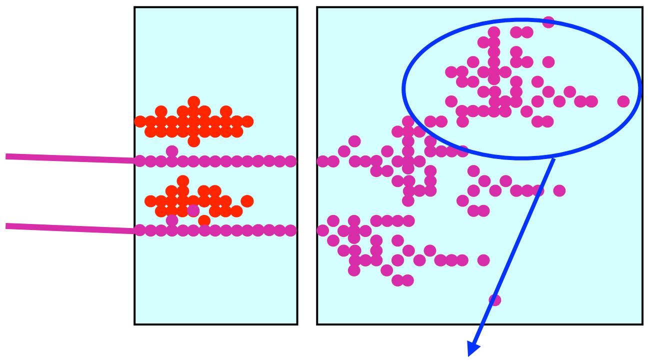
If these hits are clustered together with these, lose energy deposit from this neutral hadron (now part of track particle) and ruin energy measurement for this jet.

Level of mistakes, "confusion", determines jet energy resolution, not intrinsic calorimetric performance

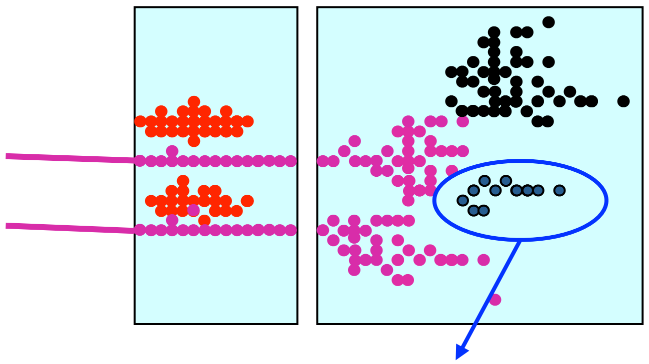
Three basic types of confusion:



Failure to resolve photons



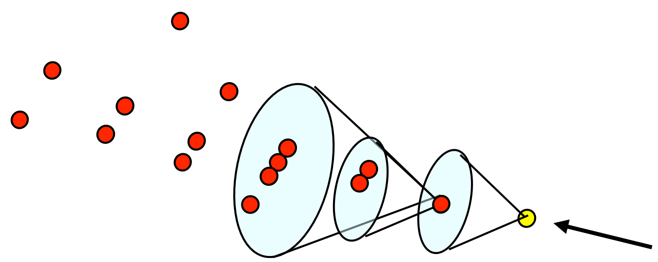
Failure to resolve neutral hadrons



Reconstruct fragments as separate neutral hadrons

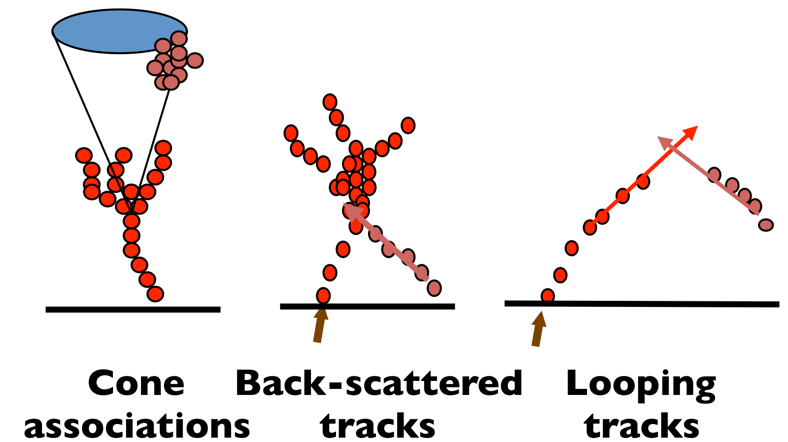
Fine-Granularity Algorithms

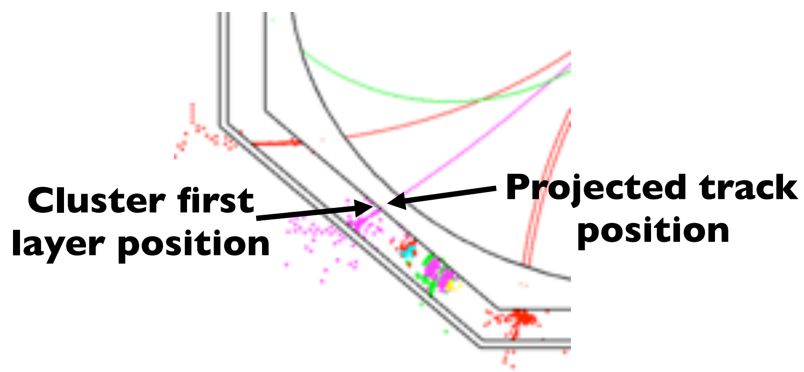
56 algorithms for fine-granularity detectors



ConeClustering Algorithm

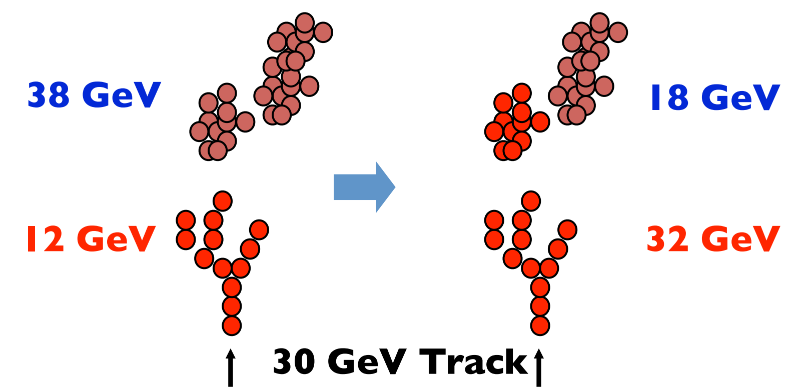
Topological Association Algorithms

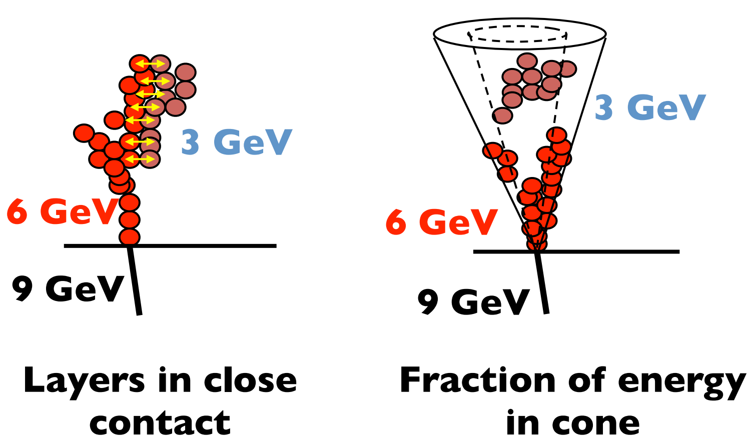




Track-Cluster Association Algorithms

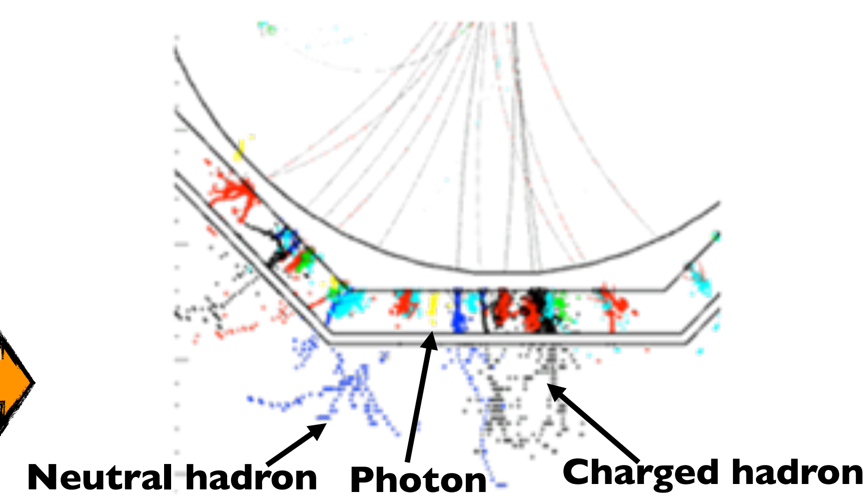
Reclustering Algorithms





Fragment Removal Algorithms

PFO Construction Algorithms



PFA Performance

- Particle Flow reconstruction inherently non-Gaussian, so resolution presented in terms of rms_{90}
  - Defined as "rms in smallest region containing 90% of events"
  - Introduced to reduce sensitivity to tails in a well defined manner

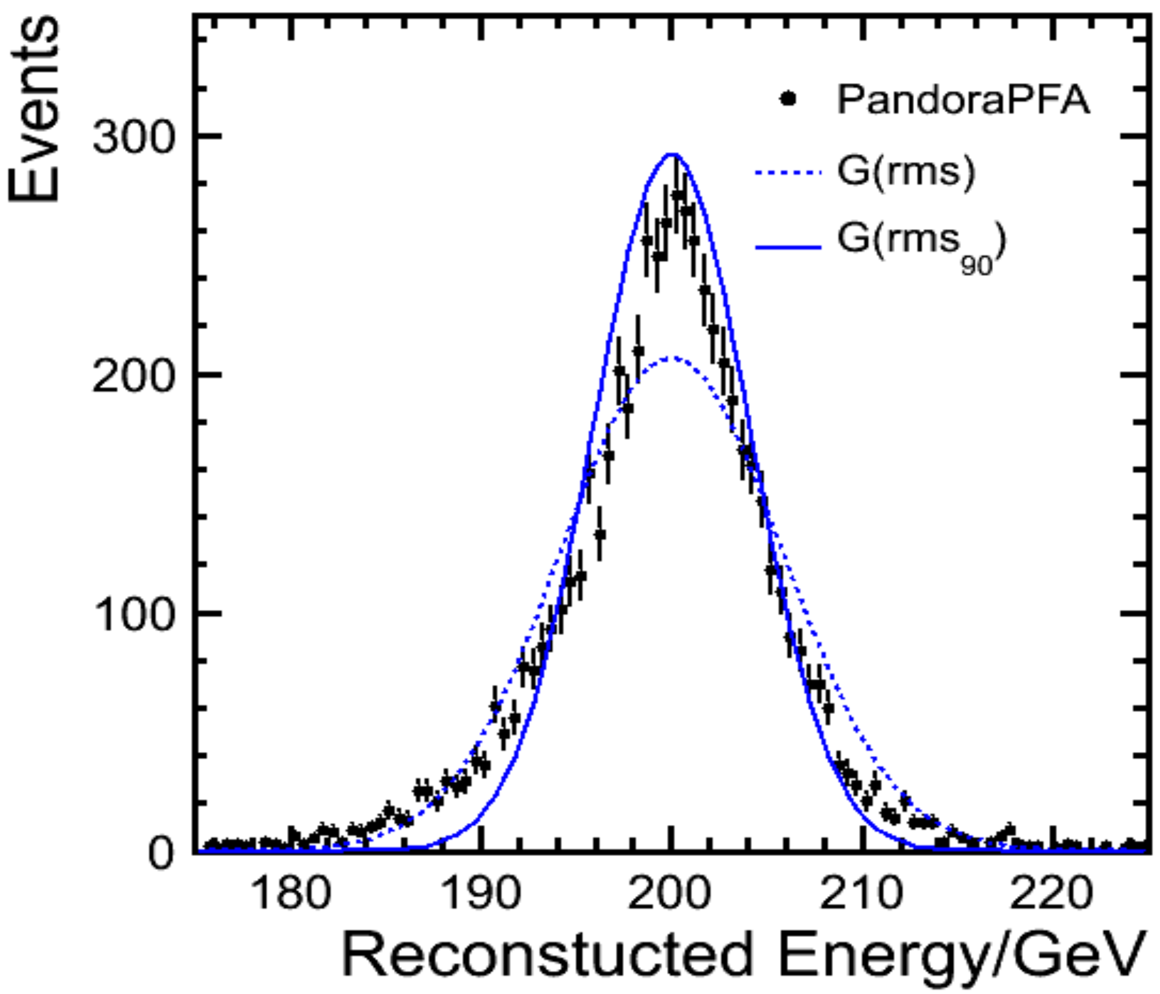
- For a true Gaussian distribution, $rms_{90} = 0.79\sigma$

- However, this can be highly misleading:
  - Distributions almost always have tails
  - Gaussian usually means fit to some region
  - $G(rms_{90})$ larger than central peak from PFA

- MC studies to determine equivalent statistical power indicate that:

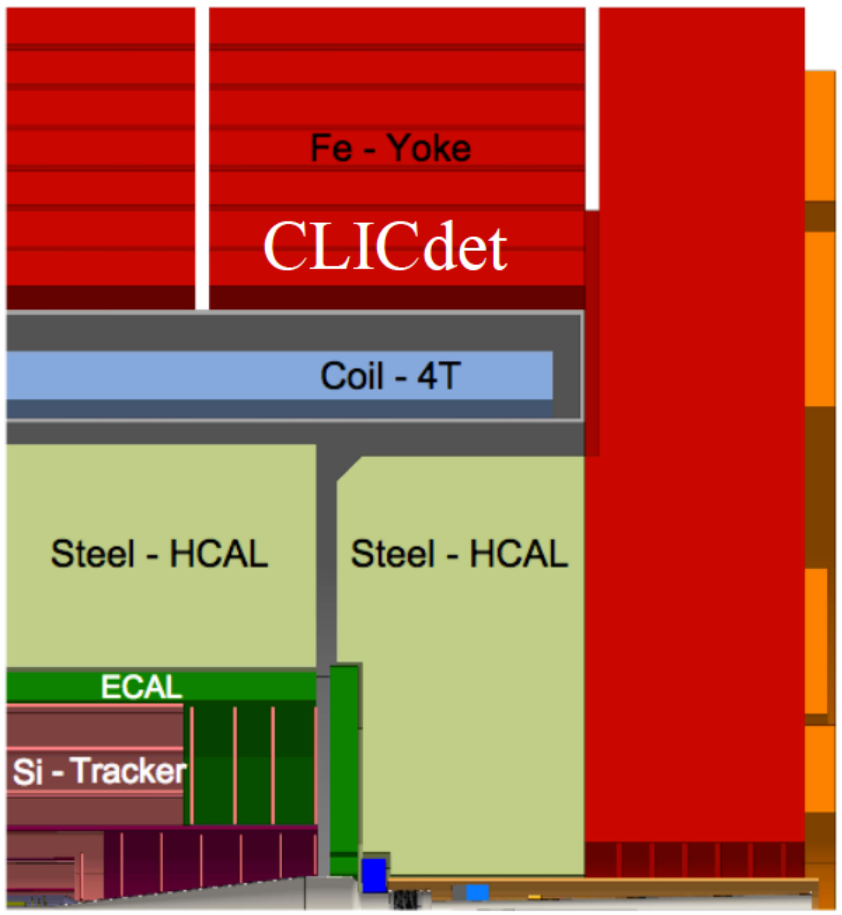
$$rms_{90} \approx 0.9\sigma_{\text{Gaus}}$$

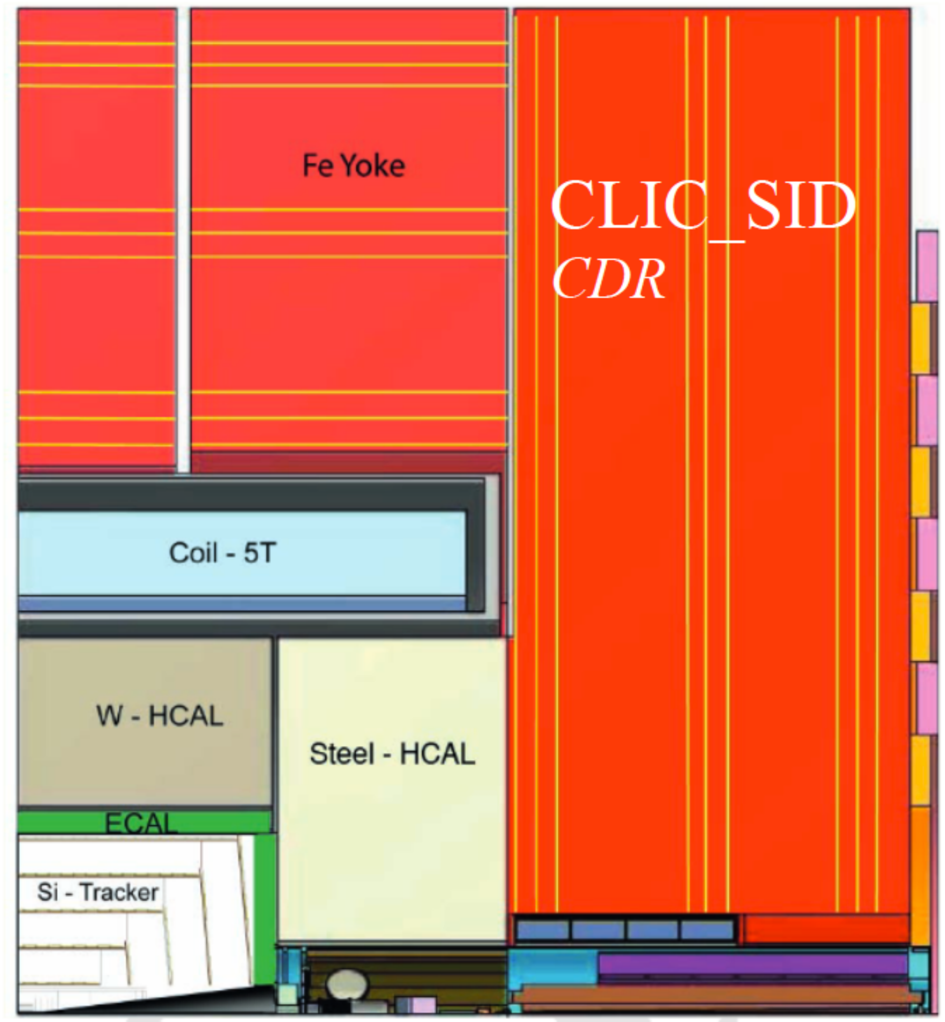
- Now use rms_{90} as a sensible convention, but does not mean PFA produces particularly large tails.



CERN: CLIC detector model

New CLIC detector model CLICdet





New optimised model **CLICdet** for new benchmark studies:
Single detector, 4 Tesla solenoid field, all steel HCAL, smaller return Yoke, quadrupole magnet outside of detector allowing better forward HCAL coverage

Detector Validation Session,
CLIC Week 2017, March8, 2017

2

Matthias Weber
CERN

Status of ARBOR's

M. Ruan [IHEP], *B. Ma* [IHEP], *D. Yu* [IHEP/LLR], *B. Li* [LLR], *K. Shpak*[LLR], *R. Été* [IPNL]
H. Videau [LLR, *V. Balagura* [LLR], *Vincent Boudry** [LLR], *J.C. Brient* [LLR], *I. Laktineh* [IPNL]
& growing number of others…

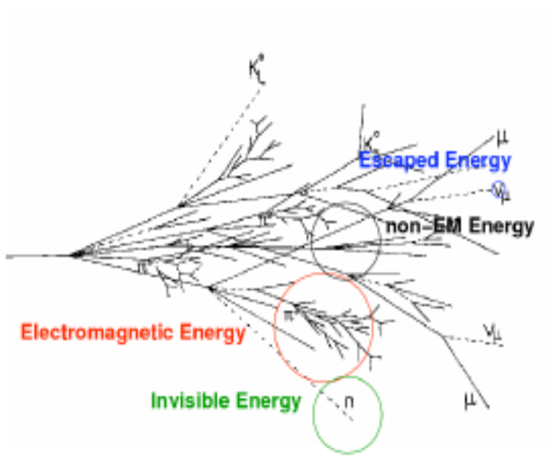
•

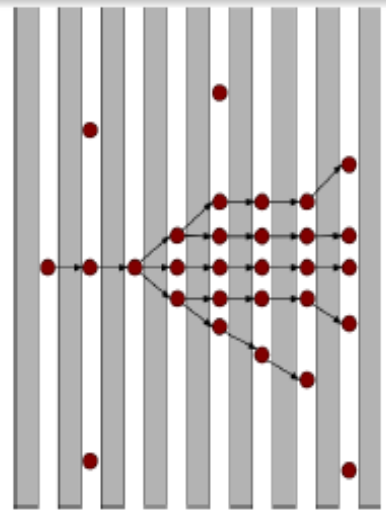
AIDA-2020 1st annual meeting
DESY
15/06/2016

ARBOR principle *by Rémi Été* [IPNL]

Principle

Particle Flow Algorithm based on hadronic shower **tree-like topology**.



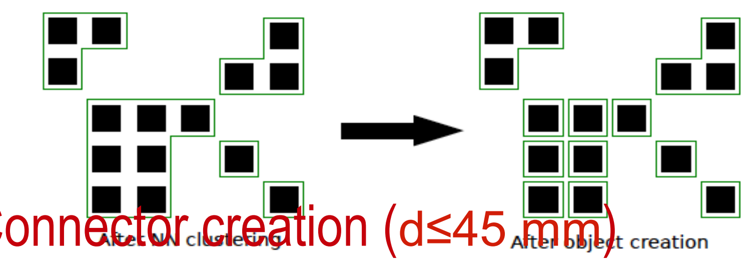


Some definitions

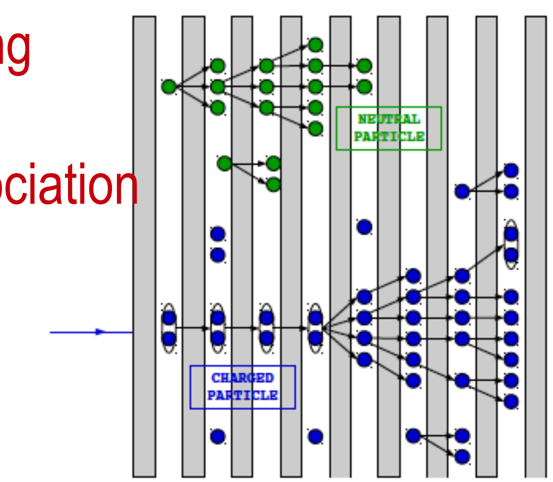
- **Object** : *Node* linked by one or many connector(s) (+ seeds and leaves)
- **Connector** : Oriented *link*. Links two objects
- **Flow direction** : Connector orientation, backward or forward
- **Tree** : Set of objects linked by connectors. For each object :
  - 0 or 1 backward connector
  - 0 or many forward connector(s)

→ Implies a unique tree structure solution (1 seed per tree)

- in-layer clustering (size ≤4)



- Connector creation (d≤45 mm)
- Connector cleaning → unique tree structure
  - Ordering using a reference direction
- Connector alignment
  - creation of secondary links between trees
- 2nd connector cleaning
- Track linking
- Neutral tree merging
- (re-clustering)
- Pointing trees association
- PFO creation



Overview of ARBOR Efforts

ARBOR "Historical Channel" (LLR + IHEP)
- from Aleph Code [H. Videau]
- Re-write in C/C++ [M. Ruan] (2010)
- ILD and CEPC optimisation studies
- esp. $\supset$ SDHCAL options

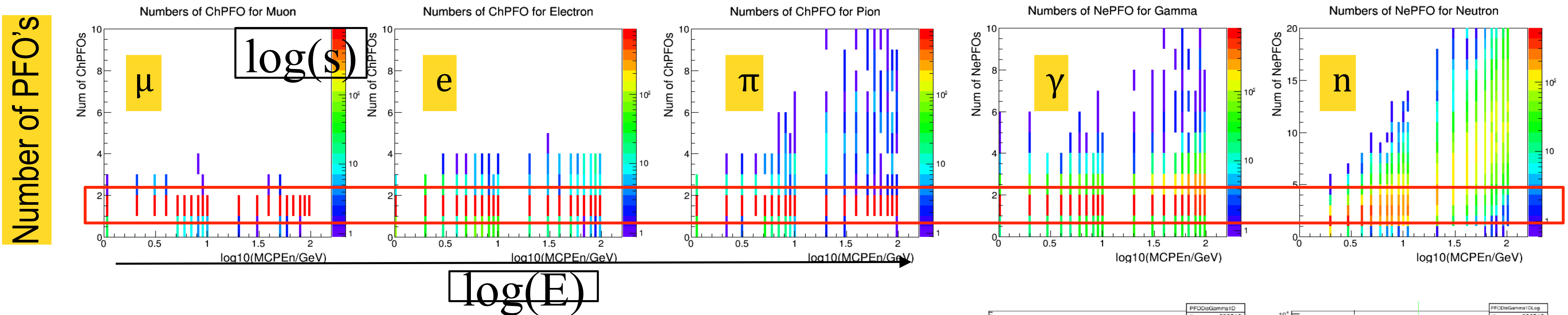
Now development towards a full package LLR+IHEP
- independent package + Marlin API
  - Git repository (IHEP)
- Implementation of p-ID
- definition of a standard set of performances

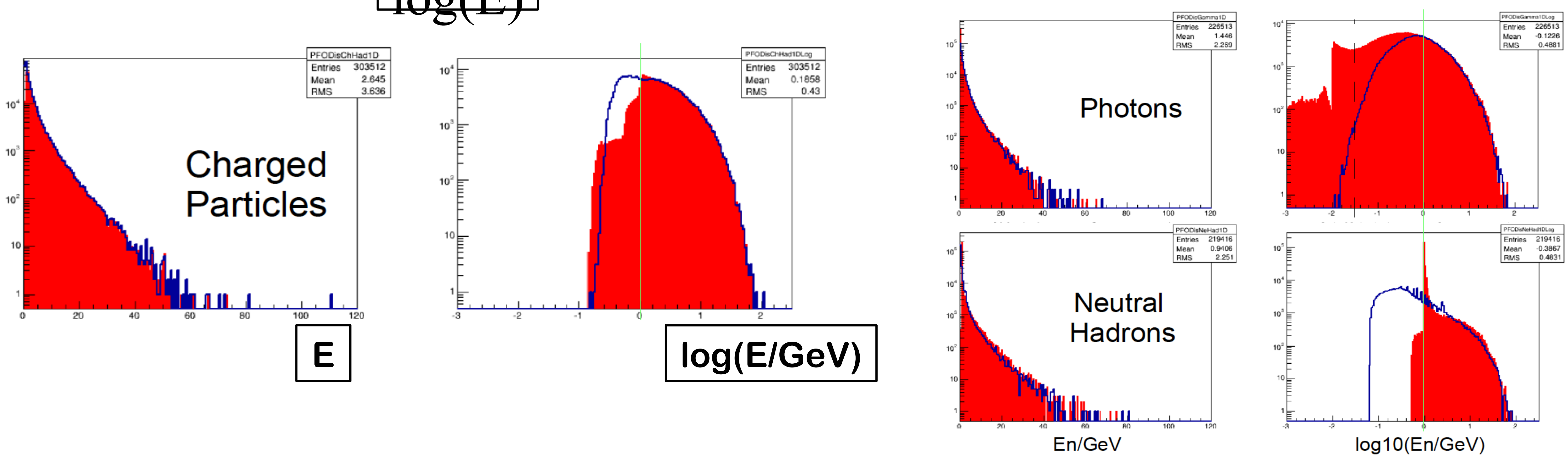
Slides from:
1st CEPC Physics Software Meeting (26-27 March 2016)

SDHCAL-ARBOR: (IPNL)
Complete re-write in C++ [R. Été] (2014)
- based on PandoraPFA framework
- used for beam test results of SDHCAL
- Toward full Detector treatment "ArborPFA"
  - Git repository (GitHub)

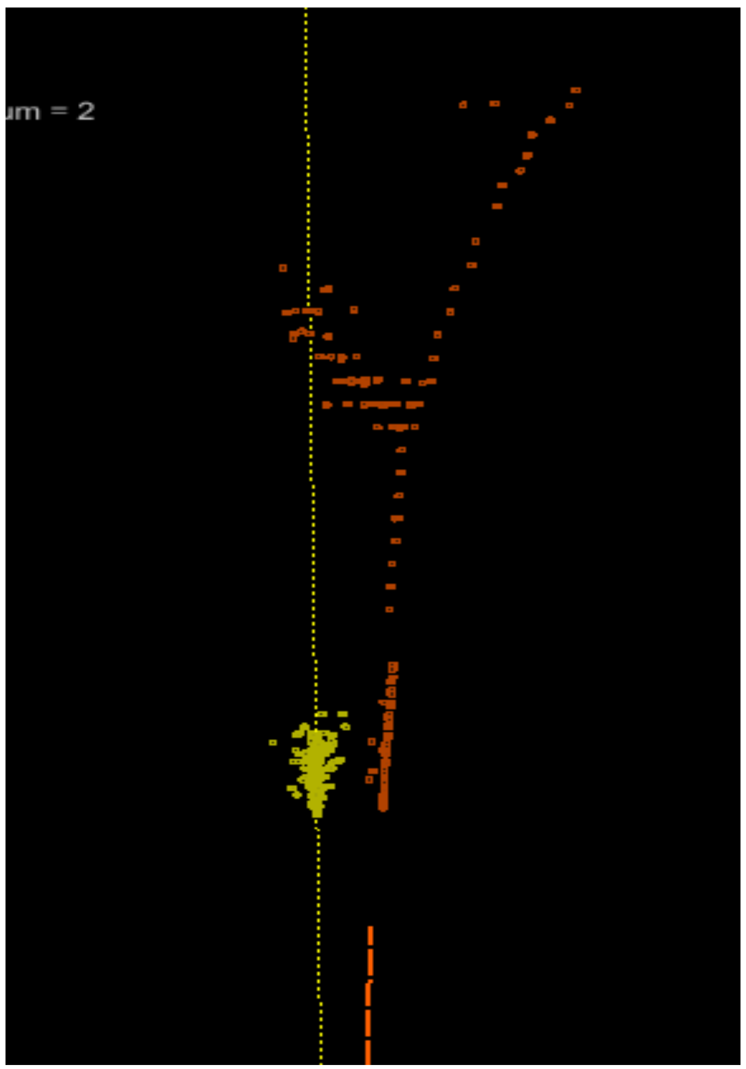
ARBOR in CMS:
- Re-write of code ↔ CMS SW framework [L. Gray]
- performances implementation of KD-tree optim

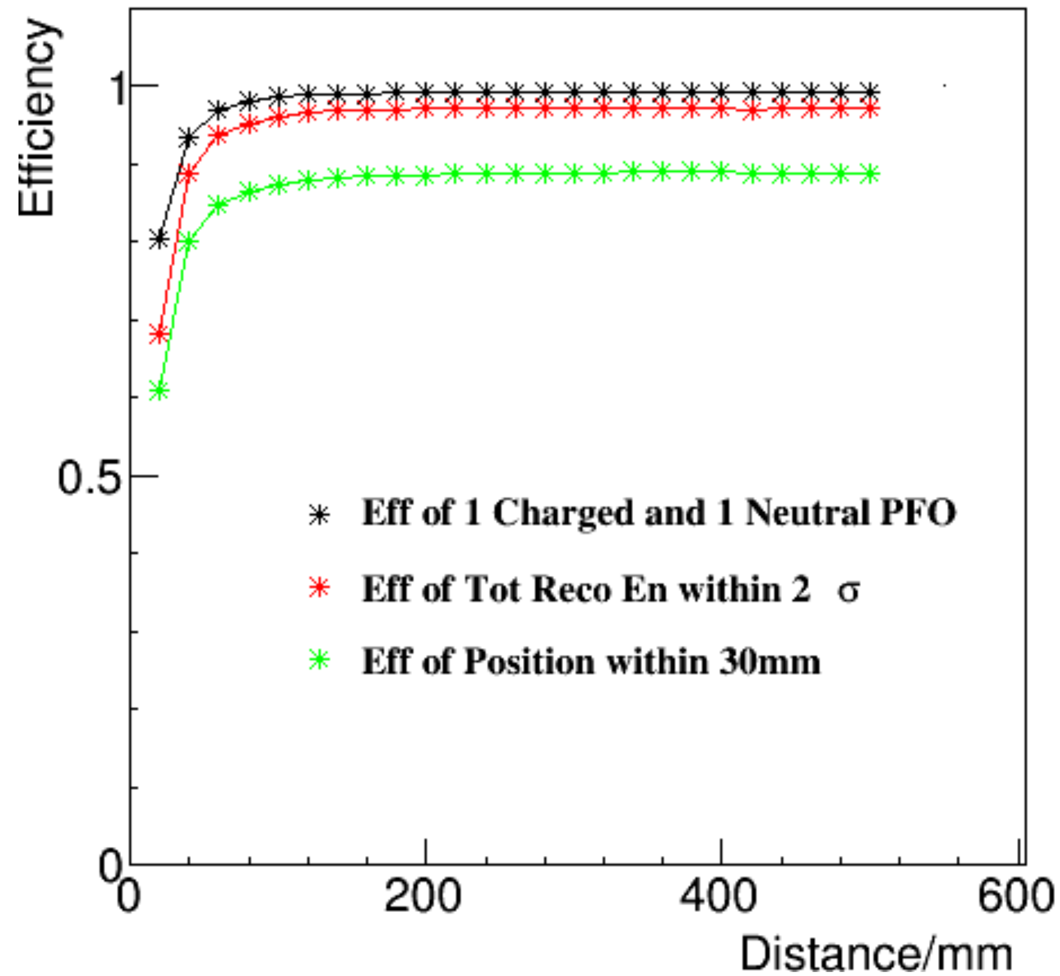
Quality plot: single particle response



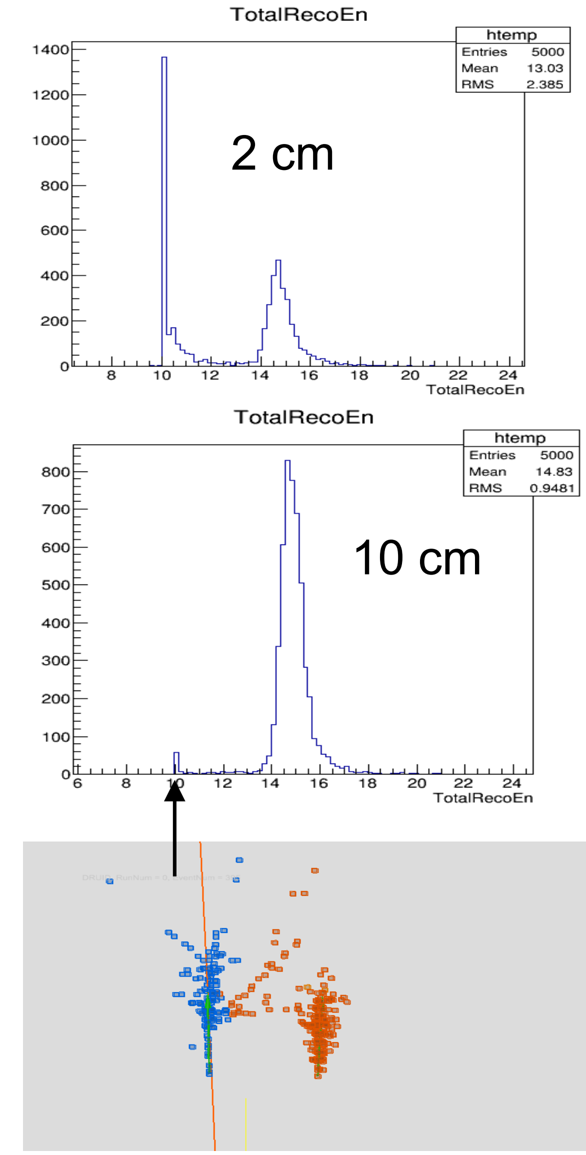


Quality plots: Separation





Tiny inefficiency: bridging effect by fragments



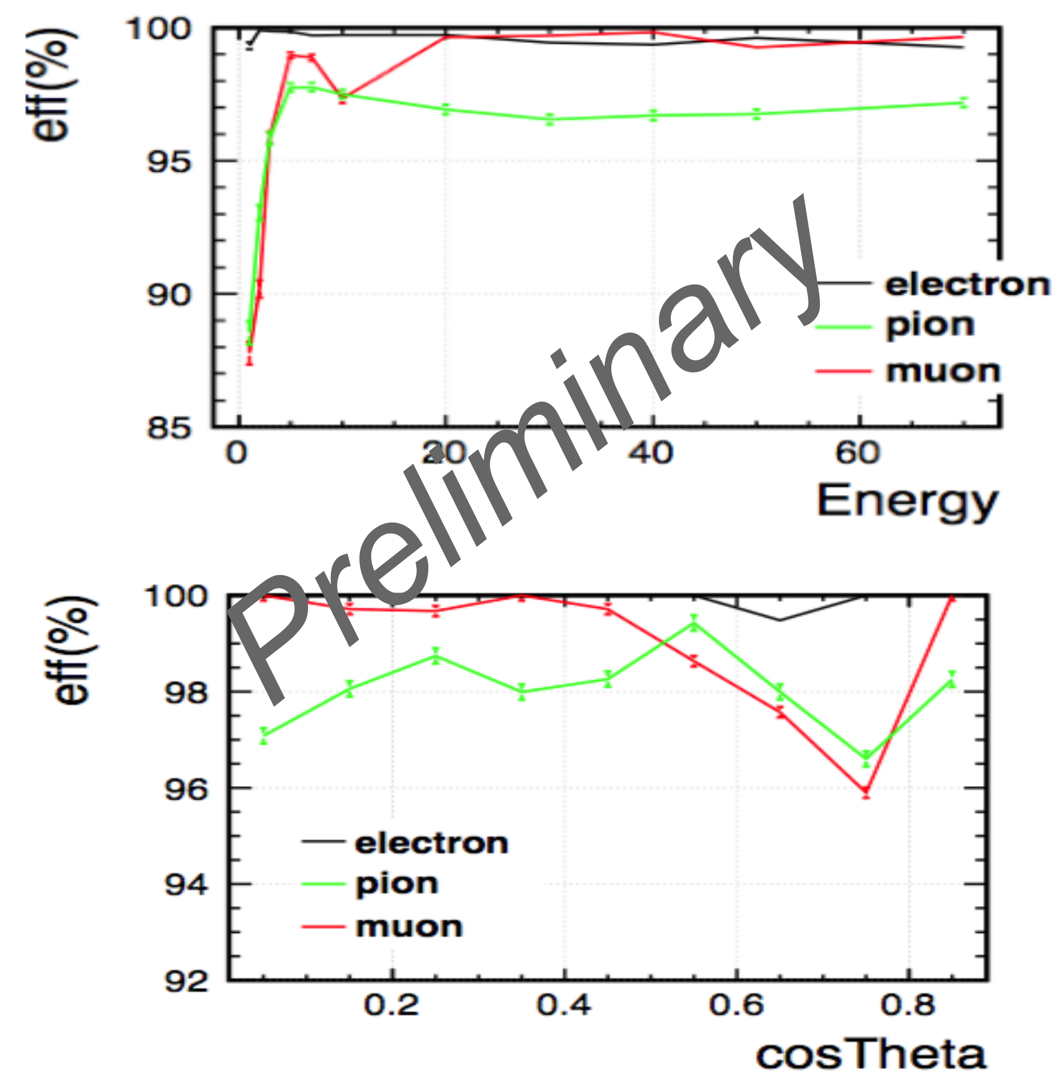
Lepton ID, next step: TMVA based [Dan YU]

TMVA including tracking, $\mathrm{d}E/\mathrm{d}x$ and calo (25 variables)

- For each charged PFO, provide a electron-likeness and muon-likeness value
- Typical working point:
  - ε > 99% for E > 20 GeV Lepton
  - ε > 97% for E > 5 GeV Pion

To be polished, encapsulated & integrated

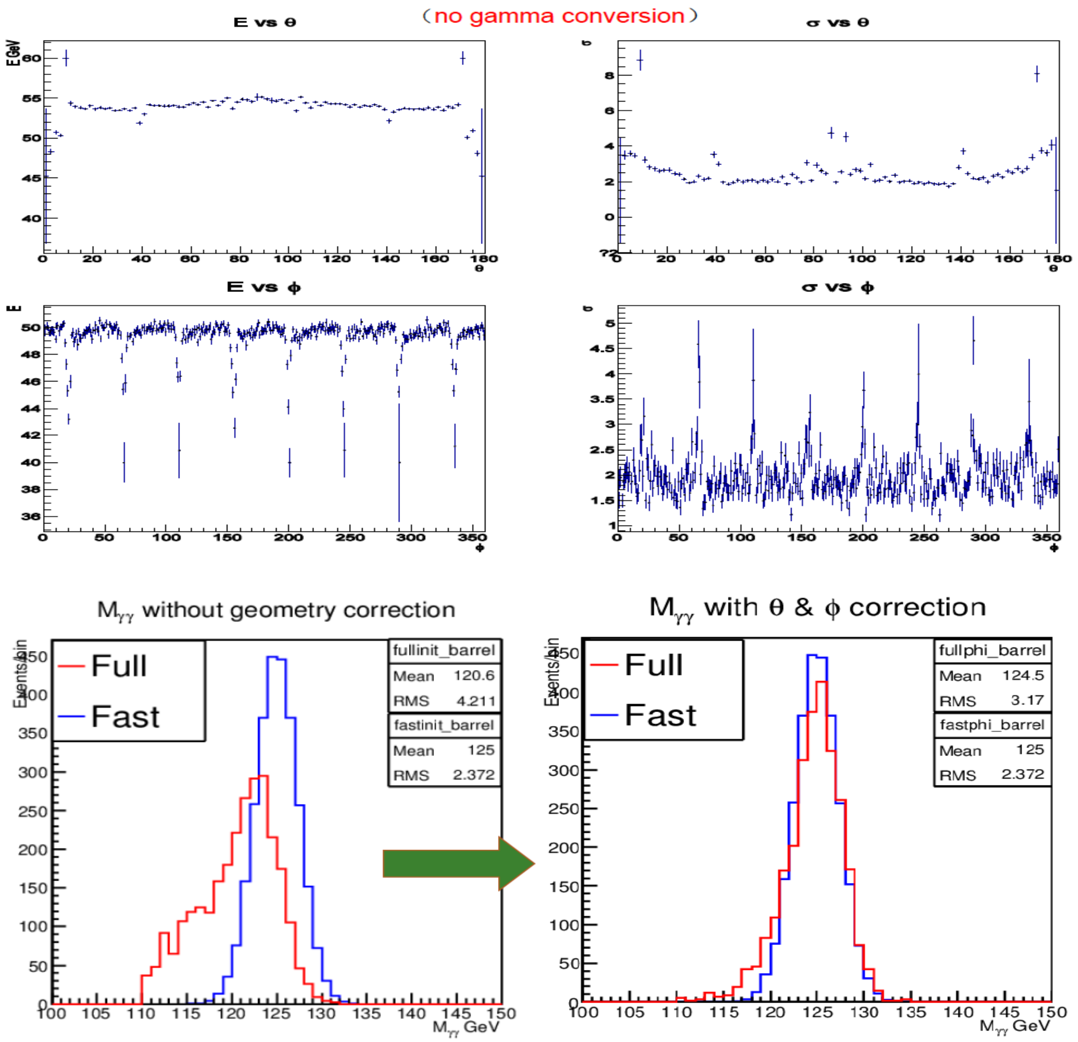
Note and Marlin module integration in preparation

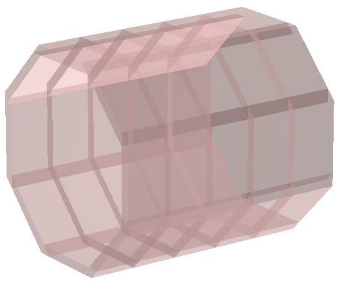


Arbor: photon reconstruction

U. of Chicago

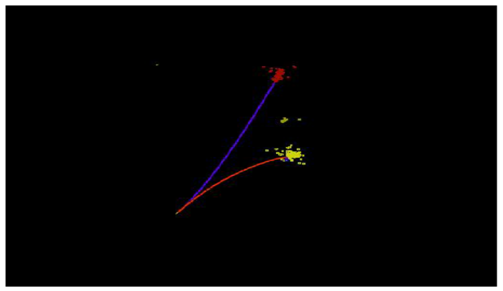
Angular Correlation of EM Shower energy response

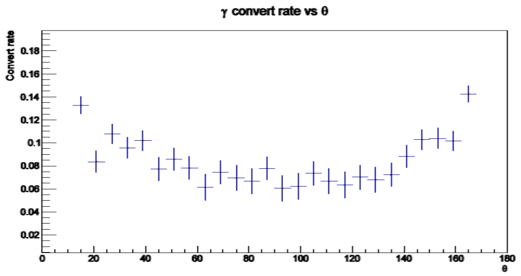


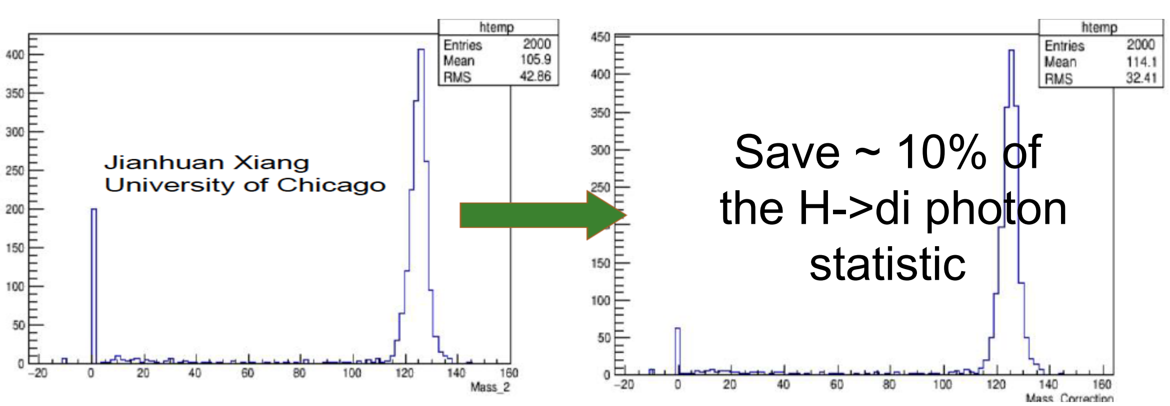


ECAL Barrel of ILD/CEPC_v1

Photon conversion & recovery



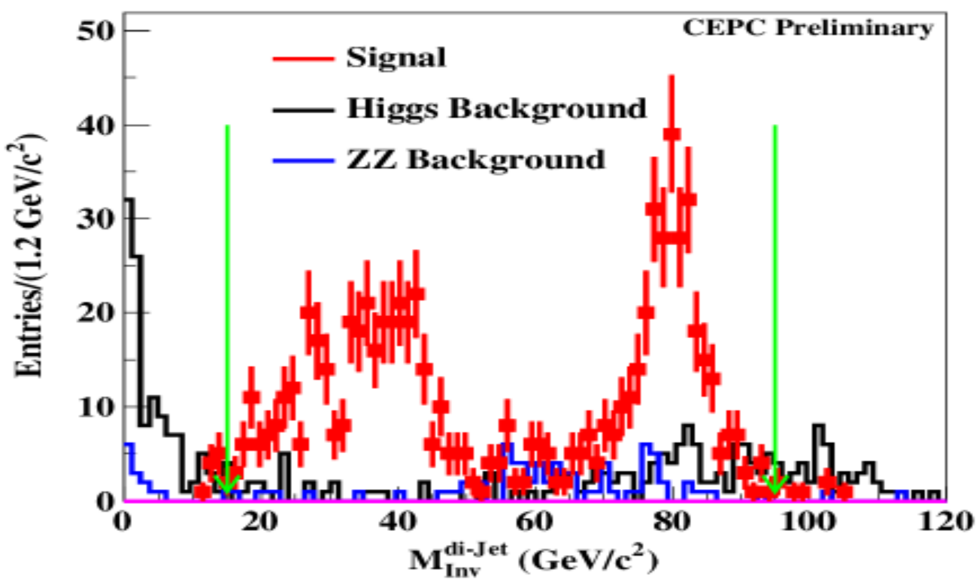


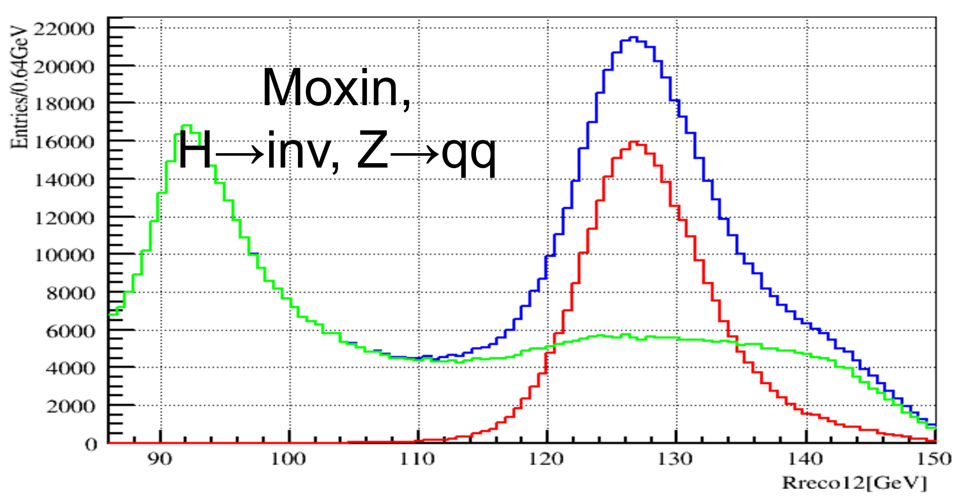


AIDA 2020

Arbor: JER MET studies

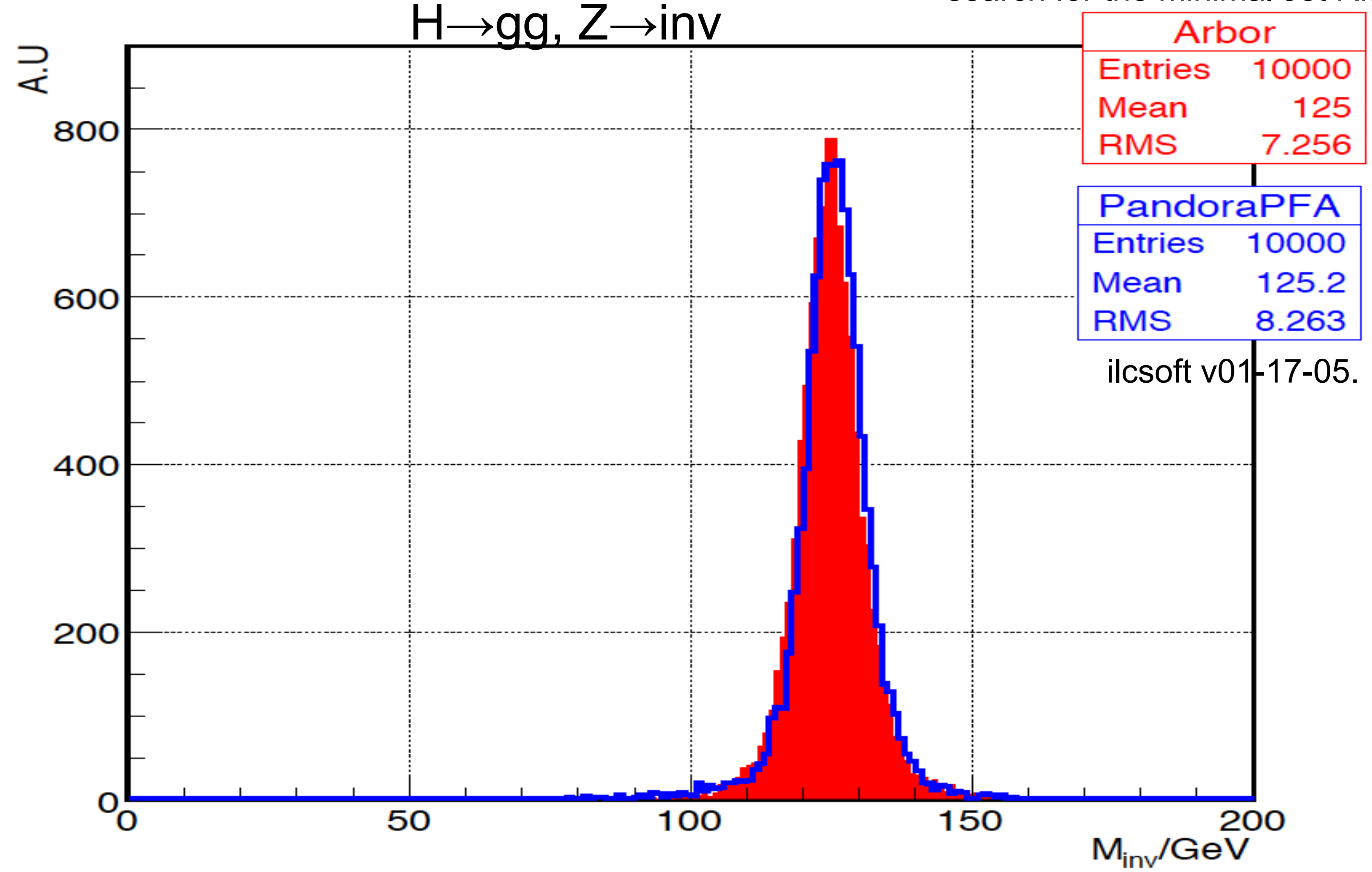
Liao libo, H→WW*→lvqq, Z→ll





unique HCAL Calibration Constant
search for the minimal Jet RMS

H→gg, Z→inv



MET: usually no ambiguity;

Jet: Highly depending on Jet clustering if #Jet > 2...