

Electron ID and Holistic Reconstruction: Workfest Summary

Shujie Li, Tyler Kutz, Dmitry Kalinkin,
Daniel Brandenburg, Derek Anderson
ePIC Collaboration Meeting
07.27.2024





Holistic Reco

- Identify what we need for *holistic* reconstruction
- Composed of:
 - > Overview session summarizing where reconstruction currently is
 - > Open discussion session to discuss where and how to get to holistic reconstruction
- Will also help identify development priorities for remainder of 2024 and into 2025



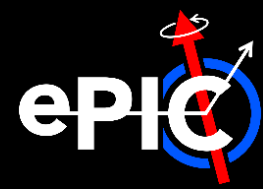
Electron ID

- Catalyze work on next-steps for the electron-finder
- Small working session for focused discussion and co-working



PID Software

- Report software status
- Talk on PID experience and analysis at BELLE-II



Workfest Agenda [~30 participants!]

- 1) [8 – 9:45 am] **Morning 1:**
Intro and Context
 - Flash updates from different *categories* of reconstruction
 - Frames forward-focused open discussion
- 2) [9:45 – 10 am] **Coffee break**
- 3) [10 am – noon] **Morning 2:**
Open Discussion
 - What do we need for truly *holistic* reconstruction
 - What are the steps to get there?
- 4) [noon – 1 pm] **Lunch**
- 5) [1 – 1:30 pm] **Afternoon 1:**
The BELLE-II PID Experience
- 6) [1:30 – 2 pm] **Afternoon 2:**
Overflow Discussion
 - Accommodate discussion from morning or BELLE-II
- 7) [2 – 3 pm] **Afternoon 2:**
Electron-Finder Working Session I
 - Focused work session on the electron-finder
- 8) [3 – 3:15 pm] **Coffee Break**
- 9) [3:15 – 4:15 pm] **Afternoon 3:**
Electron-Finder Working Session II
- 10) [4:15 – 4:30 pm] **Afternoon 4:**
Closeout discussion



- S&C focus during 1st half of 2024 on priorities for (pre-) TDR identified at [January CM](#)
 - ⇒ See earlier [S&C report on Saturday \[8:30 am\]](#) for more details
- Workfest part of ongoing discussion for identifying priorities for remainder of year and beyond
 - Previous discussions:
 - > [07.09.2024 reco meeting](#)
 - > [07.17.2024 S&C meeting](#)
- **Focus for workfest discussion:** what do we need for truly *holistic* reconstruction?
 - ⇒ **Holistic:** something synthesizing information from multiple, distinct sources

Identified priority tasks in 2023

- Primary vertexing
- Electron finder
- Low-Q2 tagger
- Particle flow

All either implemented or in progress

Identified reconstruction priorities from CERN Workshop 2024

- Modularity of algorithms
- Use holistic information from various detector components in algorithms
 - > Integration of FF/FB systems
- Web-based event display

Holistic Reconstruction | holistic reconstruction?



Digitization

Canonical Reco

Early Synthesis

Late Synthesis

Example algorithms:

- Timeframe unfolder
- Unfolded event assessor
- Digitization

Example algorithms:

- Tracking reco.
- Calorimeter reco.
- PID reco.
- FF/FB reco.

Example algorithms:

- Vertexing (primary and secondary)
- Electron identification
- Early stages of particle flow

Example algorithms:

- DIS electron selector
- Kinematic calculation
- Jet reconstruction
- Final stages of particle flow

Holistic Reconstruction | holistic reconstruction!



Digitization

Example algorithms:

- Timeframe unfolder
- Unfolded event assessor
- Digitization

Canonical Reco

Example algorithms:

- Tracking reco.
- Calorimeter reco.
- PID reco.
- FF/FB reco.

Early Synthesis

Example algorithms:

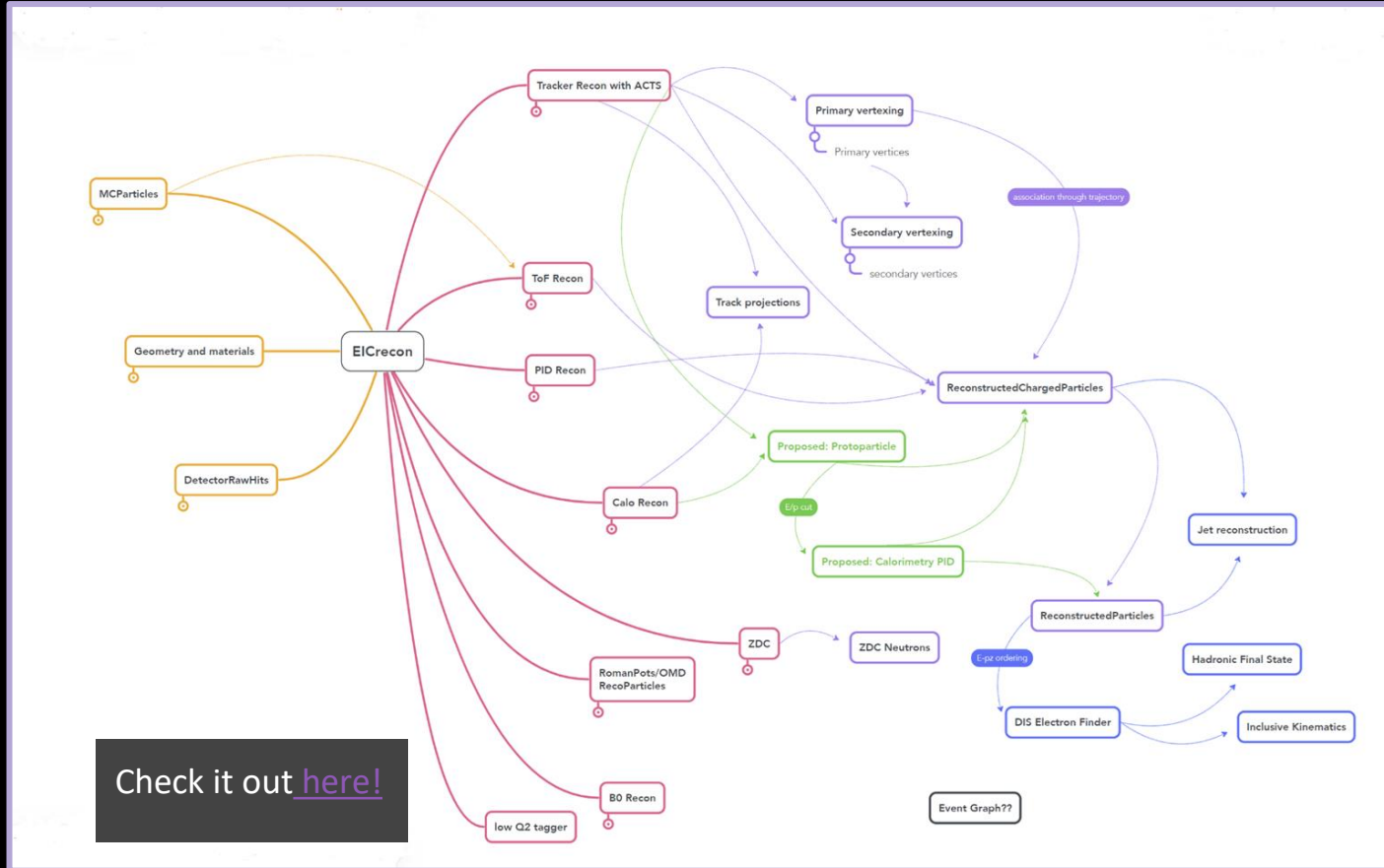
- Vertexing (primary and secondary)
- Electron identification
- Early stages of particle flow

Late Synthesis

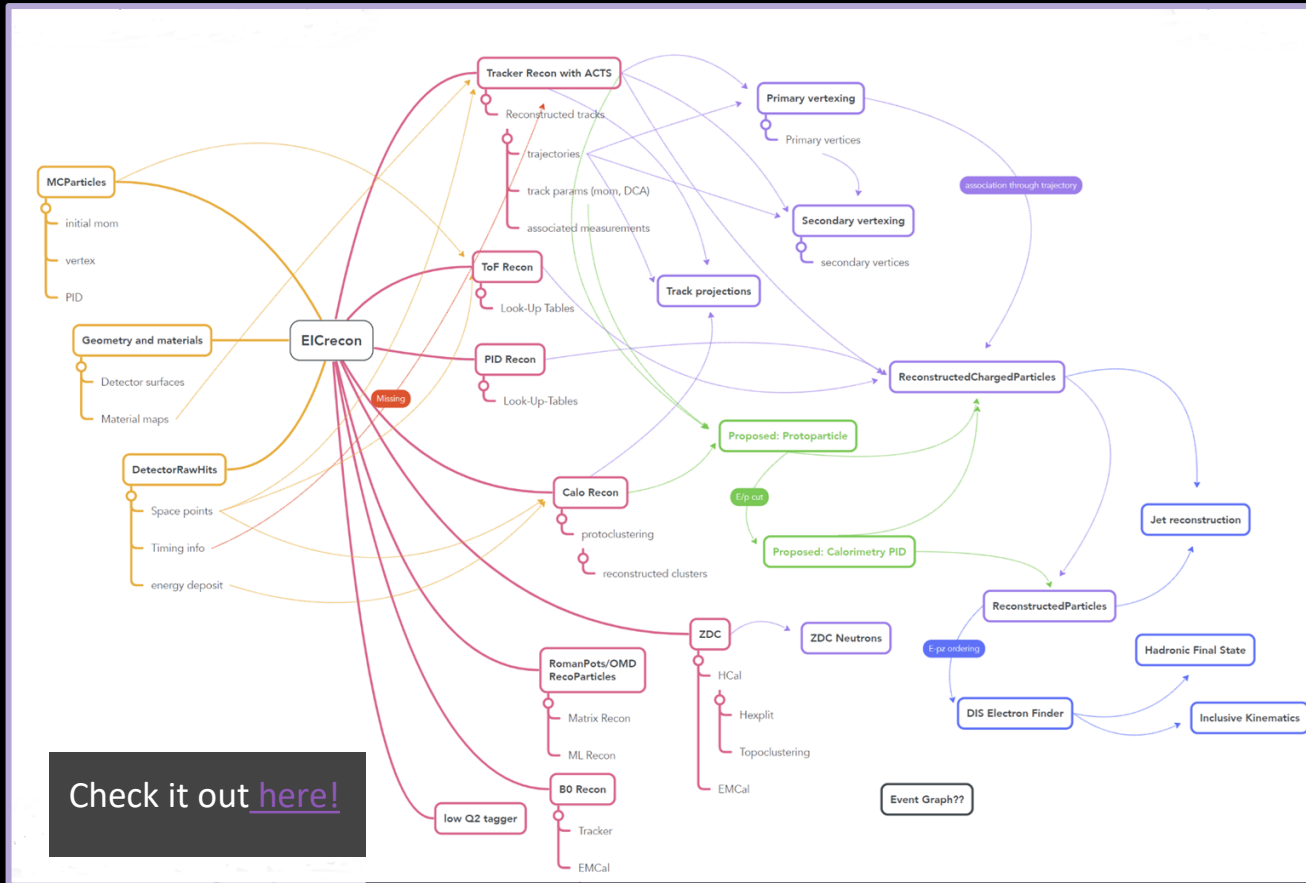
Example algorithms:

- DIS electron selector
- Kinematic calculation
- Jet reconstruction
- Final stages of particle flow

Holistic Reconstruction | mapping out EICrecon



Holistic Reconstruction | mapping out ElCrecon



Reco Status | Calorimeter

- Overall Status**
 - Basic workflow in place
 - What's in place?**
 - All basic components (e.g. simple clustering)
 - What's missing?**
 - More system-specific reconstruction routines
- Workflow Inputs/Outputs:**
 - Input:** *Hits, e.g. `EcalEndcapNHits` (`edm4hep::SimCalorimeterHit`)
 - Output:** *Clusters, e.g. `EcalEndcapNClusters` (`edm4ec::Cluster`)
- Near Term Goals:**
 - Track-based re-clustering
 - Proper truth-reconstructed cluster associations
- Longer Term Goals:**
 - System-specific digitization routines
 - Proper digitized-simulated hit associations

Derek A., ISU

Reco Status | Tracking

Workflow Inputs/Outputs:

- Input:** *ReHits from silicon tracker, `MPGD, TOF, BO`
- Output:** `CentralCKF(Seeded)Trajectories(Unfiltered)` Tracks
- `TrackParameters`

Seed finder: Can find multiple triplets from one track
 CKF allow track candidates split in track finding from one initial guess
 => use ambiguity solver to filter duplicates

Red: work in progress

- Overall Status**
 - Basic workflow in place
 - What's in place?**
 - tracking finding/fitting with space points
 - realistic seeding and ambiguity resolution
 - What's missing?**
 - Charge sharing and hits clustering
 - Timing info
- Near Term Goals:**
 - Understand tracking behaviors with hit-track and seed-track associations
 - Optimize algorithm performance
- Longer Term Goals:**
 - Tracking with time frame
 - Noise and clustering

*See Barak's talk in general meeting for more info

Shujie Li, Berkeley Lab

Tracking/Vertexing Workflow

Overall Status:

- Basic workflow in place
- What's in place?**
 - All basic components
- What's missing?**
 - edm4ec: Vertex associated Particles not filled

Workflow Inputs/Outputs:

- Inputs:** `CentralCKF(Seeded)ActsTrajectories`
- Outputs:** `CentralTrackVertices (edm4ec::Vertex)`

Near Term Goals:

- primary-vertexing benchmark for TDR
- fill in missing associated Particles in output

Long Term Goals:

- algorithm/parameter tuning for different classes of events
- MC/generated vertices and associations
- secondary vertexing

More info available at https://indico.bnl.gov/event/20727/contributions/93551/attachments/56958/94919/20240725_ePIC_Vtx.pdf

ePIC Collaboration Meeting, 07/24-07/28, 2024, Lehigh University X. Dong/LBNL

Reco Status | Secondary Vertex Finding

- Overall Status**
 - Basic workflow in place based on ACTS
 - What's in place?**
 - SecondaryVertexFinder class available
 - FullBilIorFilter and DCA cut working
 - What's missing?**
 - Import of primary vertices for track rejection
- Workflow Inputs/Outputs:**
 - Input: Tracks and primary Vertex**
 - `ActsExamples:Trajectories` (`edm4ec::VertexCollection`)
 - Output: Secondary vertices** (`edm4ec::VertexCollection`)
- Near Term Goals:**
 - Improvement via primary vertex import
 - Full pythia event validation
- Longer Term Goals:**
 - Comparison to KFParticle secondary vertex finder

Nico S., ORNL

Reco Status | ToF

- Overall Status**
 - Digitization (Yellow in right chart)
 - in progress
 - What's in place?
 - In progress but EICRecon w/o Digitization looks working
 - What's missing?
 - EICRecon w/ Digitization
- Workflow Inputs/Outputs:**
 - Input: ToF Raw hits + Raw Hits other than ToF
 - Output: ToF RecHits
- Mid Term Goals (for preTDR):**
 - Complete Digitization
 - Optimize material Budget
 - Performance study under beam backgrounds

Kentaro K., Shinshu U.

IRTv-2

Chandra C., INFN

- A second version of IRT has been developed during backward RICH review.
- A sophisticated chi-square based PID algorithm is used to handle much more complicated event topology.
- Tested thoroughly in a Standalone code.
- Kolja has made an effort to import the stand-alone code to EICRecon. Does not work out-of-the box. Requires, testing and step-by-step debugging.
- Alexander and Chandra plan to work on it. Will be used in both RICHes. Standalone code plots:

Reco Status | Roman pots/OMD

- Overall Status**
 - Full reco with static transfer matrix exists and works. (**note**: special cases need to be considered, e.g. light nuclei).
 - ML algorithm exists, integration with EICrecon in-progress.
- Workflow**
 - Input(s): ForwardRomanPotsRecoHits
 - Output(s): ForwardRomanPotsRecoParticles (similar for Off-Momentum Detectors)
- Near-Term goals**
 - Get ML algorithm fully-integrated.
 - Fix a few dangling issues for nuclei, and sorting of hits.
- Long-Term Goals**
 - Replace static matrix code with dynamic (polynomial) matrix code.

Alex J., BNL

Reco Status | B0 tracker

- Overall Status**
 - Full reco has been tested, but with recent changes, correct output for ACTS tracking a bit unclear.
 - B0 field map needs to be put into a PR and merged (see below).
- Workflow**
 - Input(s): B0TrackerRecoHits
 - Output(s): ReconstructedChargedParticles***
- Near-Term goals**
 - Make PR for the B0 field map (on a private branch), and get it merged.
- Long-Term Goals**
 - Include charge sharing digitization and change segmentation to correct value (currently set to a value to provide expected resolutions).

Alex J., BNL

Electron ID + reconstruction status

- Overall Status
 - Basic workflow in place
 - **What's in place?**
 - › All basic components
 - **What's missing?**
 - › Fully reconstruction-based matching/associations
 - › eID parameters (E/p, E-pz) in output
 - › Calorimeter energy for electron reconstruction
- Workflow Inputs/Outputs:
 - **Input: Reconstructed particles, calorimeter clusters, cluster-particle associations (truth info)**
 - › (edm4eic::ReconstructedParticle)
 - › (edm4eic::Cluster)
 - › (edm4eic::MCRecoClusterParticleAssociation)
 - **Output: DIS electron candidates (subset of reconstructed particles)**
 - › (edm4eic::ReconstructedParticle)

Tyler K., MIT

Reco Status | Particle Flow

- Overall Status
 - Work started, but on hold...
 - **What's in place?**
 - › Truth-based track-EMCal matching (i.e. *MatchClusters* algorithm)
 - **What's missing?**
 - › Non-truth based track-cluster matching
 - › Integration of HCal info
- PFAlpha Inputs/Outputs:
 - **Input: Track projections, Clusters,**
 - › (edm4eic::TrackSegment)
 - › (edm4eic::Cluster)
 - **Output: Reconstructed particles**
 - › (edm4eic::ReconstructedParticle)
- Near Term Goals:
 - Finish implementation of PFAlpha
- Longer Term Goals:
 - Refinement of PFAlpha using PID
 - Explore alternative algorithms vs. eta

Derek A., ISU

- Strongly recommend checking out both slides and live notes!
 - All are on [indico page](#)

- **Summary:** (almost) all of our canonical reconstruction is in place
 - 👉 Reconstruction moving from truth to enhanced realism

- **Example:** Reconstructed Charged Particles
 - Default will use KF seeds by September
 - > Now: realistic seeded quantifies flagged by “seeded” in name
 - Now use LUTs instead of truth PID

Belle II



Barrel cherenkov: TOP

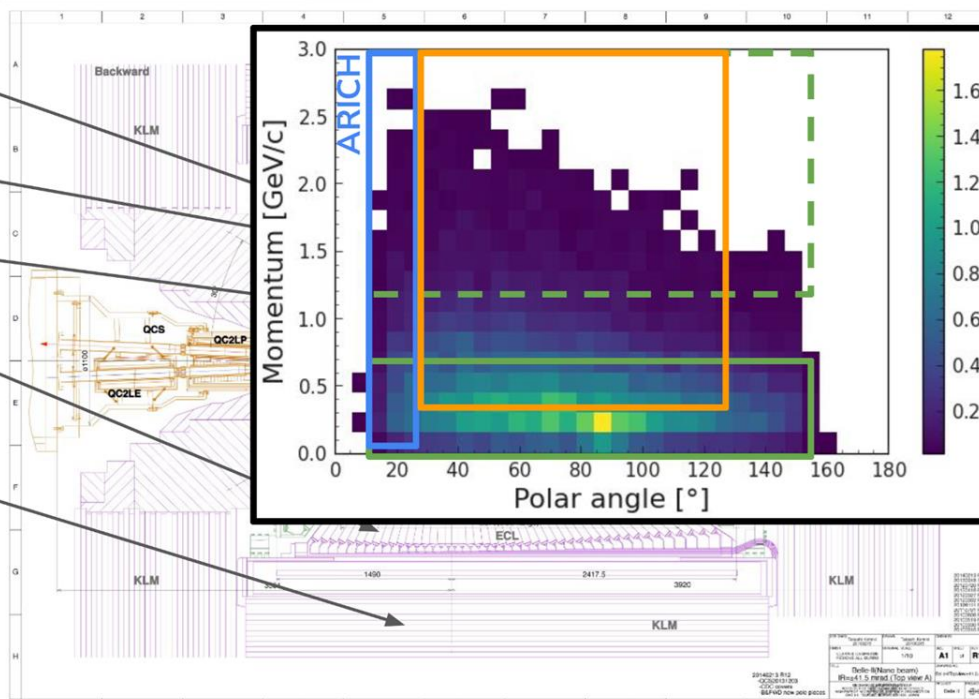
Endcap Cherenkov: ARICH

dEdx: SVD + CDC

Shower shape: ECL

Pen. depth: KLM

Umberto T., INFN



Check out recording and slides [here!](#)

Combining information



Each sub-detector provides a likelihood value for 6 possible PID hypotheses:

- electron, muon, pion, kaon, proton, deuteron
- The likelihood values are calculated comparing the observed signal with the expectation for each particle hypothesis (based in MC, data template, or analytic models)
- If particle is out-of-acceptance, $\text{LogL} = 0$ for all hypotheses

$\mathcal{L}_\alpha^d = \mathcal{L}^d(\mathbf{x}|\alpha)$ Likelihood for hypothesis α from detector d that observed \mathbf{x} hits

$\mathcal{L}(\mathbf{x}|i) = \exp\left(\sum_{d \in D} \log \mathcal{L}^d(\mathbf{x}|i)\right)$ Likelihood for hypothesis α from all detectors

Umberto T., INFN

$$P(A_i|\mathbf{x}) = \frac{P(\mathbf{x}|A_i) \cdot P(A_i)}{\sum_j P(\mathbf{x}|A_j)P(A_j)} \Rightarrow P(i|\mathbf{x}) = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j} \quad \text{PID probability}$$

5

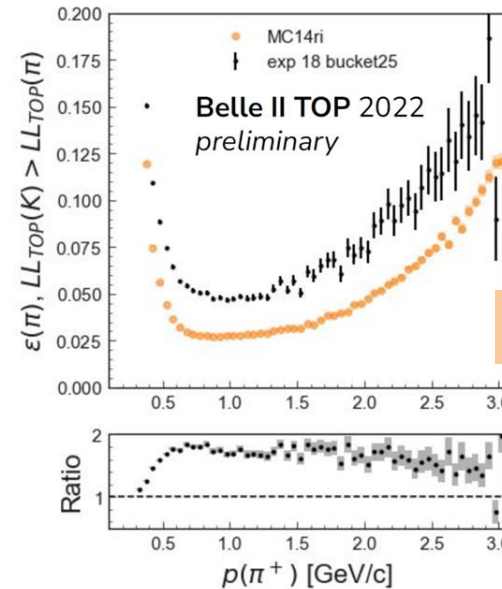
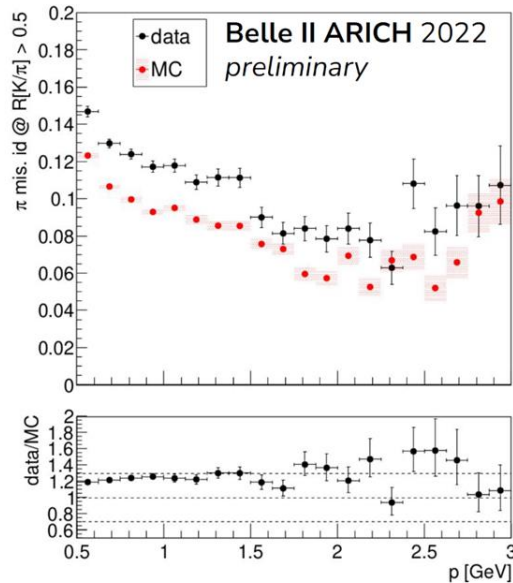
Check out recording and slides [here!](#)

Expectations VS reality



Performance observed in data still don't match with (optimistic) MC

- Many lessons learned so far!



Umberto T., INFN

18

Check out recording and slides [here!](#)

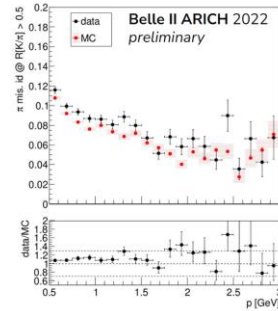
PID at BELLE-II | reality check (2/2)



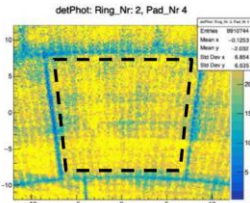
Lessons learned: ARICH tile alignment



Aerogel tile edges are responsible for most of the disagreement in ARICH



- Removing tracks extrapolated in the edges
- Improves PID (expected) reducing acceptance
 - Improves data/MC (not expected)
 - Work towards better tile alignment



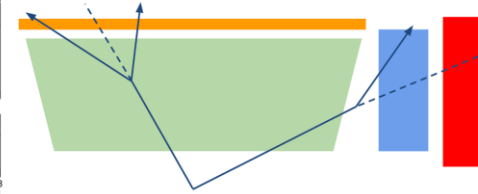
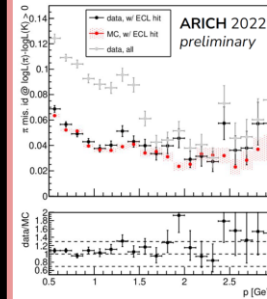
19

Mitigating material scattering



Use the Calorimeter behind ARICH and TOP to remove bad extrapolations

- Require a cluster matched with the track
- Powerful tool, but introduced correlation between subdetectors...



23

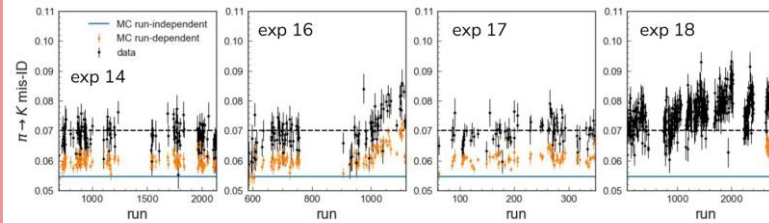
Umberto T., INFN

Lessons learned: background effects on TOP



For TOP, half of the data/MC disagreement is recovered with more realistic simulation

- Actual dead/hot channel maps from data
- Backgrounds from random triggers instead of simulation



Residual discrepancy is under investigation.

20

User's end-point



We save only the LogL values in the mDST

- ~20% of raw data are always available for extra studies

Particle identification probabilities are calculated on-fly by the analysis libraries

- Users can choose which type or probability (global, binary, ternary...)
- Users can choose which detectors are to be used

electronID, muonID, pionID, kaonID, protonID, deuteronID

pidPairChargedBDTScore(pdgCodeHyp, pdgCodeTest)

"Expert" variables

pidLogLikelihoodValueExpert(pdgCode, detectorList)

pidDeltaLogLikelihoodValueExpert(pdgCode1, pdgCode2, detectorList)

pidPairProbabilityExpert(pdgCodeHyp, pdgCodeTest, detectorList)

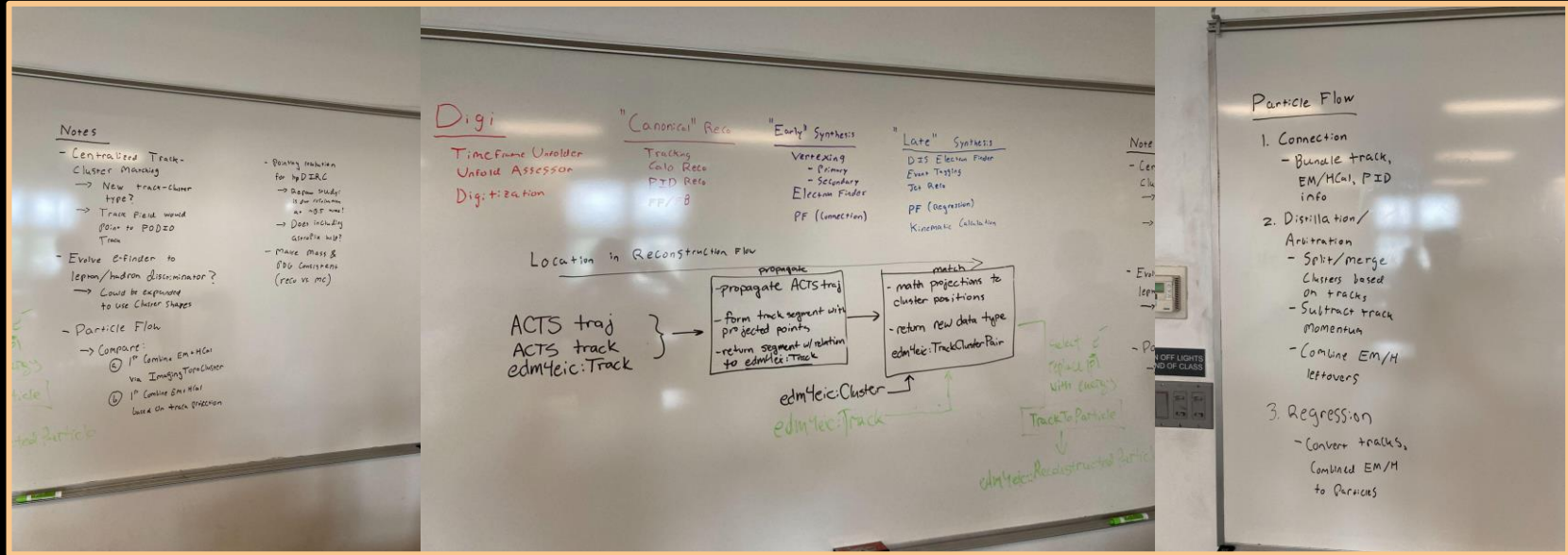
pidProbabilityExpert(pdgCodeHyp, detectorList)

Umberto T., INFN

24

Check out recording and slides [here!](#)

Discussion, Work Sessions | overview



- Open discussion spilled over into afternoon working session
 - Was extremely productive!
 - **Above:** discussion notes left on the whiteboard

- Bulk of discussion was on interplay of two early/late synthesis algorithms:
 - **Electron-Finder, and**
 - **Particle Flow**



Schematic Outline of PF: refactor current work-in-progress to modularize

1) Connection

- Bundle tracks, clusters, and PID hypotheses into proto-particles

2) Arbitration

- Split/merge clusters based on tracks
- Subtract track momentum
- Combine EM/HCal leftovers

3) Regression

- Convert proto-particles into reconstructed ones

Electron-Finder Evolution: excise truth info and refactor for synergy with other projects

1) Implement track-cluster matching algorithm

- Would produce new track-cluster data type

2) Identify hadrons using track-cluster matches

- Start with E/p (can expand later)

3) Convert identified electrons to reconstructed particles

- And pass to DIS electron selector



- **Should centralize track-cluster matching**
 - Need new data type?
 - A *TrackClusterPair* or *ProtoParticle* to bundle matched tracks and clusters
 - Track field would point to PODIO tracks (naturally)
- **Could evolve e-finder to a lepton/hadron discriminator**
 - Also could extend lepton ID to use cluster shapes
- **Compare two approaches to PF:**
 - Combine EMCal and HCal clusters via TopoClustering *before* connection + arbitration
 - Or combine EM/HCal clusters based on track projections
- **Recheck pointing resolution to hpDIRC:**
 - Repeat study to see if recent tracking changes improve resolution
 - Check if including AstroPix helps
- **Bug to fix:** make mass & PDG consistent (i.e. based on LUTs) in Reco. Charged Particles

Join in Reco WG Meetings!
Tuesdays at 3:30 pm EDT!



Thanks!

