TDR Progress
Calorimeter Reconstruction

2024 ePIC S&C CERN Workshop

Derek Anderson (ISU)

April 23rd, 2024

Summary | TDR Readiness (1/3)

○ Calorimeter reconstruction started 2024 in strong state
  – Workflow completed in 2023
  – Since been in use by users

○ However, several requests were made at January CM
  – **Right:** summary of identified calorimeter software needs / wants from CM discussion

∴ **2 questions to answer:**
  1) Which ones are must-haves?
  2) Are the must-haves fulfilled?

Identified Data Model Needs

– Improved truth-cluster connections

Identified Reconstruction Needs/Wants

– Clustering implemented in all systems
– Cluster splitting/merging
– ML Integration
– Digitization noise, noise-masking and system-specific digitization model implementations
– Better neutral identification
– Easier access to janadot output

Identified Simulation Needs/Wants

– Enhanced realism in BEMC implementation and implementation of end-of-sector box material
– Dedicated studies of HGCROC vs. waveform digitizer in BEMC
– Physics-driven performance studies for nHCal
– Update ZDC default to SiPM-on-tile
– Enhanced realism in pECal implementation

Summary | TDR Readiness (2/3)

o Amount of requests and scope of some go well beyond available workforce and what can be accomplished in a few months

∴ **"Must Have:"** *has* to happen in EICrecon or have something changed in EICrecon to happen

☞ Some things can be addressed by standalone analyses or accounted-for post-hoc

Must Haves:
– Improved truth-cluster connections
– Clustering in all systems
– Cluster splitting/merging
– ML integration
– Noise-masking, channel-by-channel gain/noise setting
– Easier access to janadot output

(More details in following slides)

Identified Data Model Needs

– **Improved truth-Cluster connections**

Identified Reconstruction Needs/Wants

– **Clustering implemented in all systems**
– **Cluster splitting/merging**
– **ML Integration**
– **Digitization noise, noise-masking** and system-specific digitization model implementations
– Better neutral identification
– **Easier access to janadot output**

Summary | TDR Readiness (3/3)

Category	Request	Status
Must Haves	Improved truth-cluster connections	In progress
	Clustering in all systems	Complete
	Cluster splitting/merging	In progress
	ML integration	Complete
	Noise-masking, channel-by-channel gain/noise setting	Complete
	Easier access to Janadot output	To do
Not Must Haves	Better neutral identification	To do
	System-specific digitization models	To do

Summary | EICrecon Issues

Request	Corresponding issues
Improved truth-cluster connections	EICrecon#898, EICrecon#899
Clustering in all systems	EICrecon#1342
Cluster splitting/merging	EICrecon#897, EICrecon#1289
ML integration	EICrecon#1340
Noise-masking, channel-by-channel gain/noise setting	EICrecon#1337
Easier access to janadot output	EICrecon#1339
Better neutral identification	EICrecon#1341
System-specific digitization models	EICrecon#1338, EICrecon#1081, EICrecon#696

Details | To-Do Issues (1/2)

Better Neutral Identification

Issues: EICrecon#1341
PRs: n/a

Notes
- Desire was expressed during January discussion for better reconstruction of neutral particles
- This is something that can be handled in standalone analyses for time being
- Addressing this in a satisfactory manner will require particle flow

Easier Access to Janadot

Issues: EICrecon#1339
PRs: n/a

Notes
- Users expressed desire for an easier way of visualizing what algorithms are being run
- Concern with Janadot was that it's hard-to-find and too developer oriented
- ☞ Could we put in something to quickly export mermaid diagrams?

Details | To-Do (2/2) & Complete Issues (1/2)

System Specific Digitization Models

Issues: EICrecon#1338, EICrecon#1081, EICrecon#696
PRs: n/a

Notes
- Users expressed desire for digitization models (esp. wrt. to noise) more tuned to the actual hardware that'll be used
- However, this is a massive undertaking that's going to take coordination across several groups
- This is something that will need to wait until after the TDR...

Noise-Masking, etc.

Issues: EICrecon#1337
PRs: EICrecon#1349

Notes
- However, more tractable items did come up in follow-up conversations: ability to implement dead maps + more control over gain/noise
- PR#1349 addresses this nicely: gains for each channel can be adjusted w/ adjacency matrix-style expression

Follow-Up
☞ Confirm with DSCs that this is sufficient
☞ If not, identify anything w/ quick turnaround

Examples

```
−PEEMC:EcalEndcapNRecHits:samplingFraction='[row,
column](){ static map<std::tuple<double, double>,
double> sf; auto index = std::make_tuple(row,
column); if (!sf.count(index)) {
std::default_random_engine gen(row + 100 * column);
std::normal_distribution R{1.0, 0.1}; sf[index] =
R(gen); } return sf[index]; }()'
```

```
.sampFrac = "(eta == phi) ? 0.0 : 0.033"  //
unalive all channels on a diagonal, otherwise use
default
```

Noise-Masking, etc.

Issues: EICrecon#1337
PRs: EICrecon#1349

Notes
- However, more tractable items did come up in follow-up conversations: ability to implement dead maps + more control over gain/noise
- PR#1349 addresses this nicely: gains for each channel can be adjusted w/ adjacency matrix-style expression

Follow-Up
☞ Confirm with DSCs that this is sufficient
☞ If not, identify anything w/ quick turnaround

Details | Complete Issues (2/2)

Clustering in all Systems

Issues: EICrecon#1342
PRs: n/a

Notes
- In January, nHCal DSC specifically was unsure if clustering parameters were ideal
- DSC liaison confirmed at recent Calo CCWG meeting that parameters are indeed okay

Follow-Up
☞ Double-check with other DSCs to make sure no other subsystem slipped through crack

ML Integration

Issues: EICrecon#1340
PRs: EICrecon#1358

Notes
- In January, users expressed desire for an example of how to use Ort (or likewise) in an algorithm
- PR#1358 accomplishes this nicely
- And TMVA::SOFIE is now available in the shell (containers#9)

Follow-Up
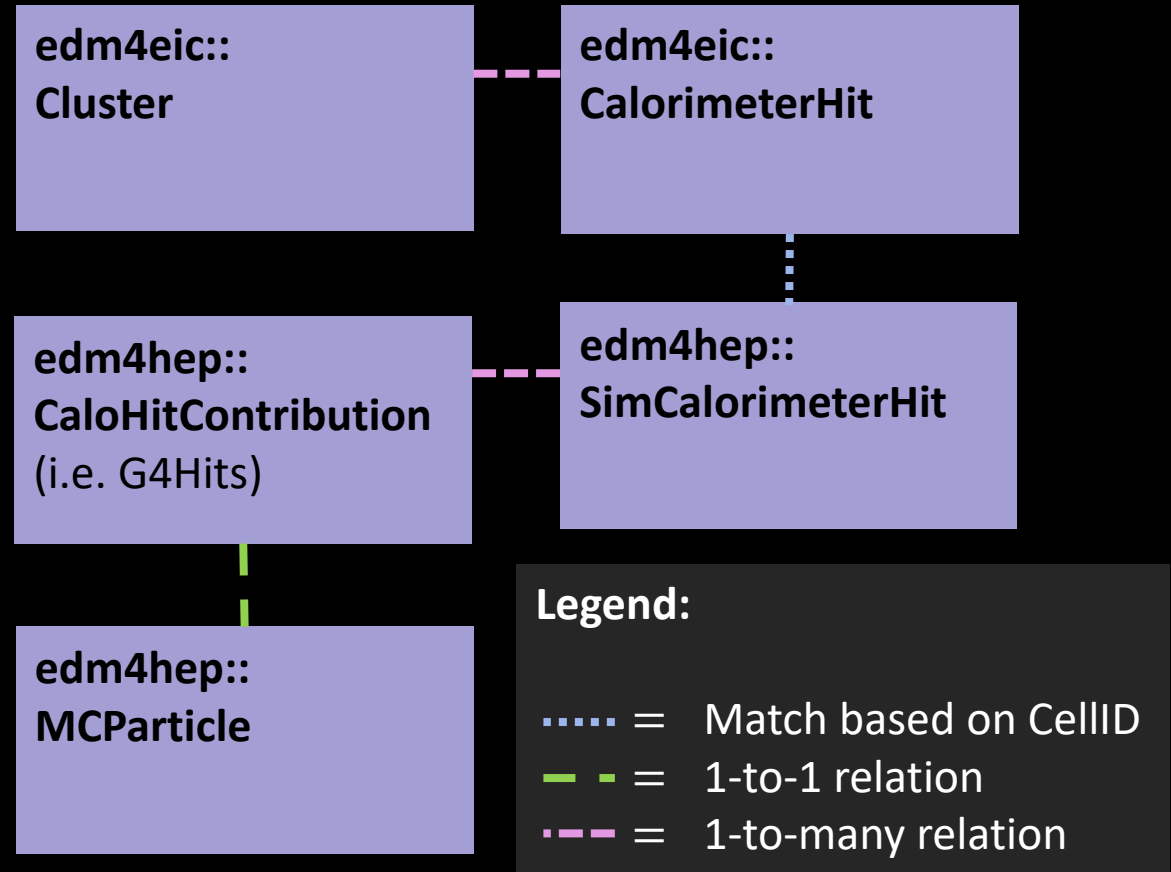☞ Small issues with stability in enabling ONNX (c.f. EICrecon#1394)

Improved Truth-Cluster Associations

Issues: EICrecon#898, EICrecon#899
PRs: EICrecon#1382 [+1 more to come]

Context
– The biggest concern expressed by users (esp. from the HCal.s)
  › Are the truth-cluster associations working?
  › **Or more precisely:** is there enough information in the output for users to do what they need?
– The ability to understand how energy is flowing in the detector will be critical for more advanced studies

edm4eic:: Cluster

edm4eic:: CalorimeterHit

edm4hep:: CaloHitContribution (i.e. G4Hits)

edm4hep:: SimCalorimeterHit

edm4hep:: MCParticle

Legend:

- ····· = Match based on CellID
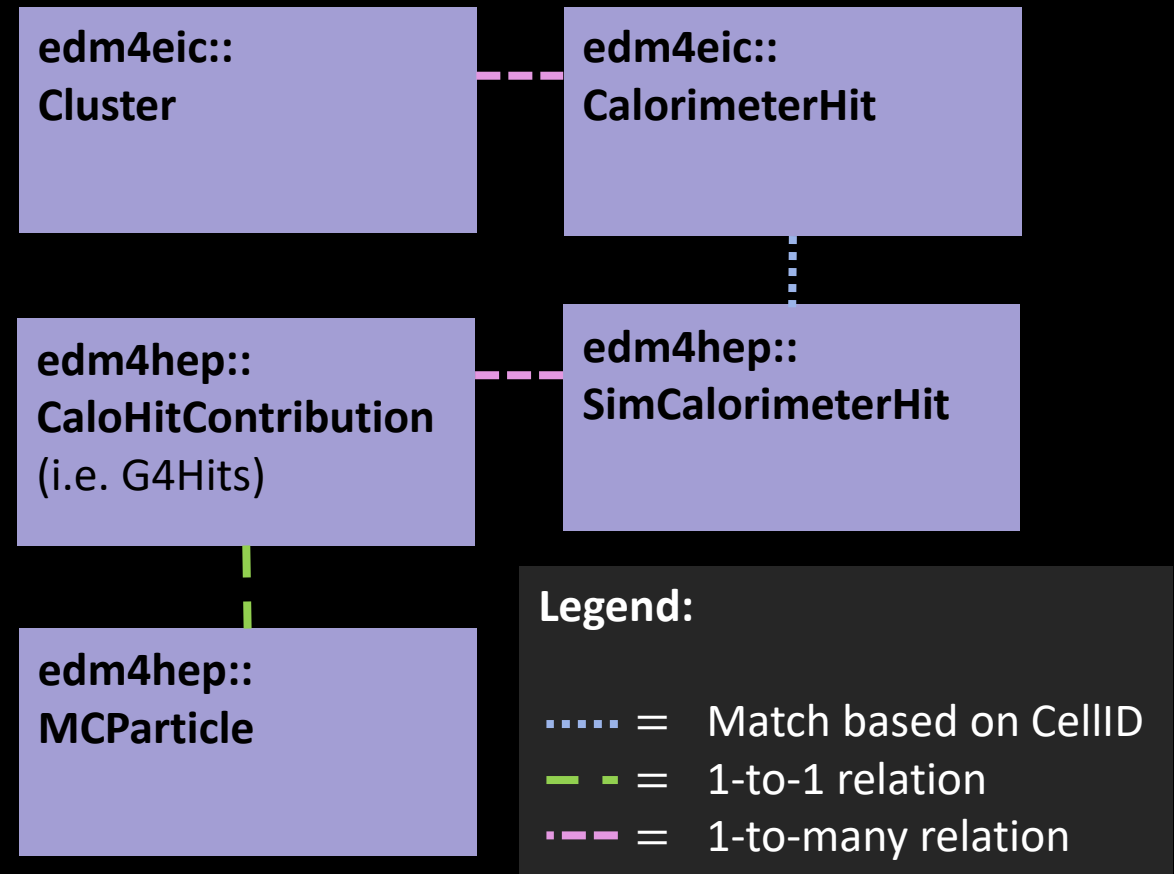- – – = 1-to-1 relation
- ·–·– = 1-to-many relation

Improved Truth-Cluster Associations

Issues: EICrecon#898, EICrecon#899
PRs: EICrecon#1382 [+1 more to come]

So what's in the EICrecon output?
– Confirmed (as of this past weekend) that **all** collections to the right are saved to EICrecon output
  › So users can go from **clusters to G4Hits to MCParticles** using EICrecon output
∴ All of the information is there...
  › But can we make life easier for users w/ a judicious choice of association b/n Clusters & MCParticles?

edm4eic:: Cluster

edm4eic:: CalorimeterHit

edm4hep:: CaloHitContribution (i.e. G4Hits)

edm4hep:: SimCalorimeterHit

edm4hep:: MCParticle

Legend:

····· = Match based on CellID

– – = 1-to-1 relation

·– – = 1-to-many relation

Improved Truth-Cluster Associations
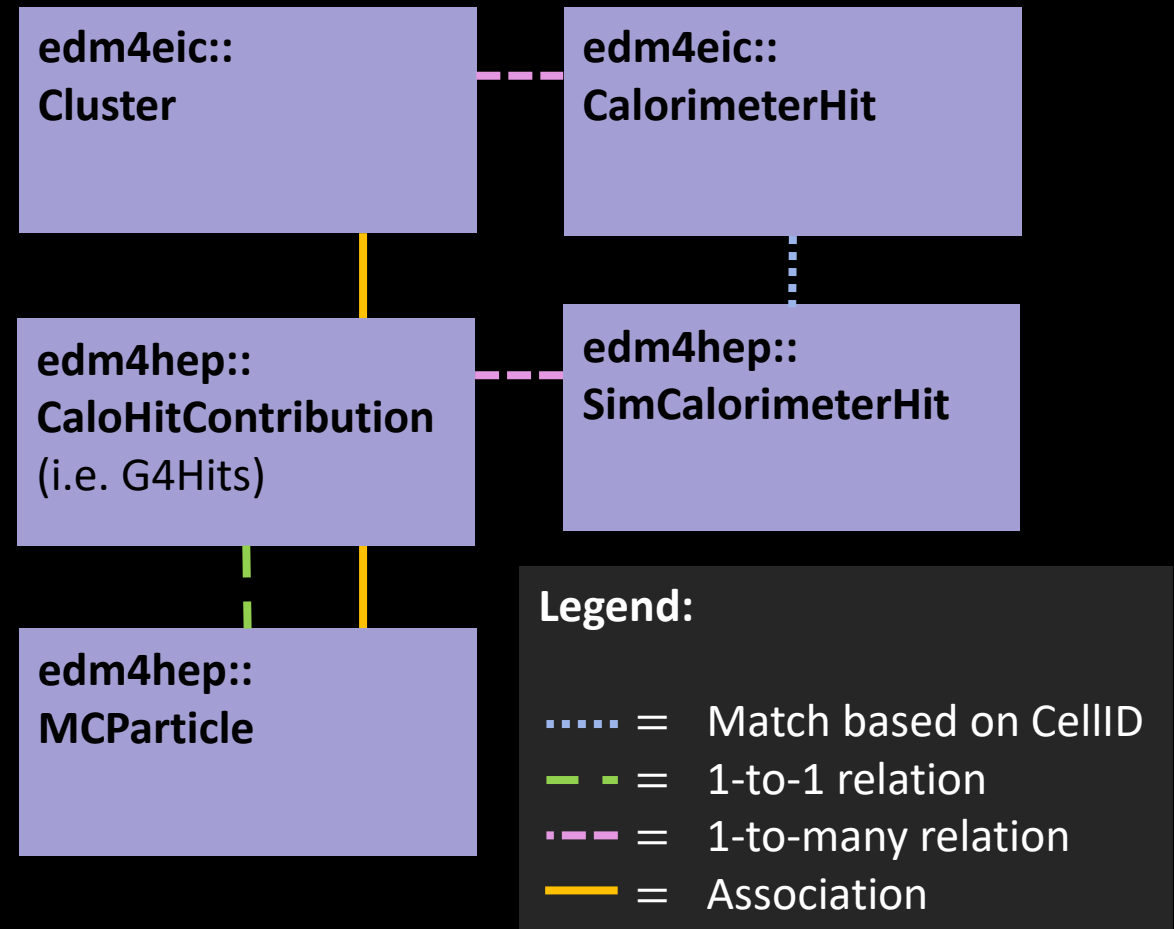
Issues: EICrecon#898, EICrecon#899
PRs: EICrecon#1382 [+1 more to come]

Current Truth-Cluster Assocation Logic:
1) Identify **highest energy hit** in cluster
2) Grab **1st contributing particle** of corresponding simulated hit
3) Assign that contributor as the associated particle of the cluster

Proposed Minimal Change:
- Set *highest energy contributor* as the associated MCParticle
- Will open PR today based on discussion

edm4eic:: Cluster

edm4eic:: CalorimeterHit

edm4hep:: CaloHitContribution (i.e. G4Hits)

edm4hep:: SimCalorimeterHit

edm4hep:: MCParticle

Legend:

..... = Match based on CellID

– – – = 1-to-1 relation

–·–· = 1-to-many relation

——— = Association

Details | In-Progress Issues: Associations (4/4)
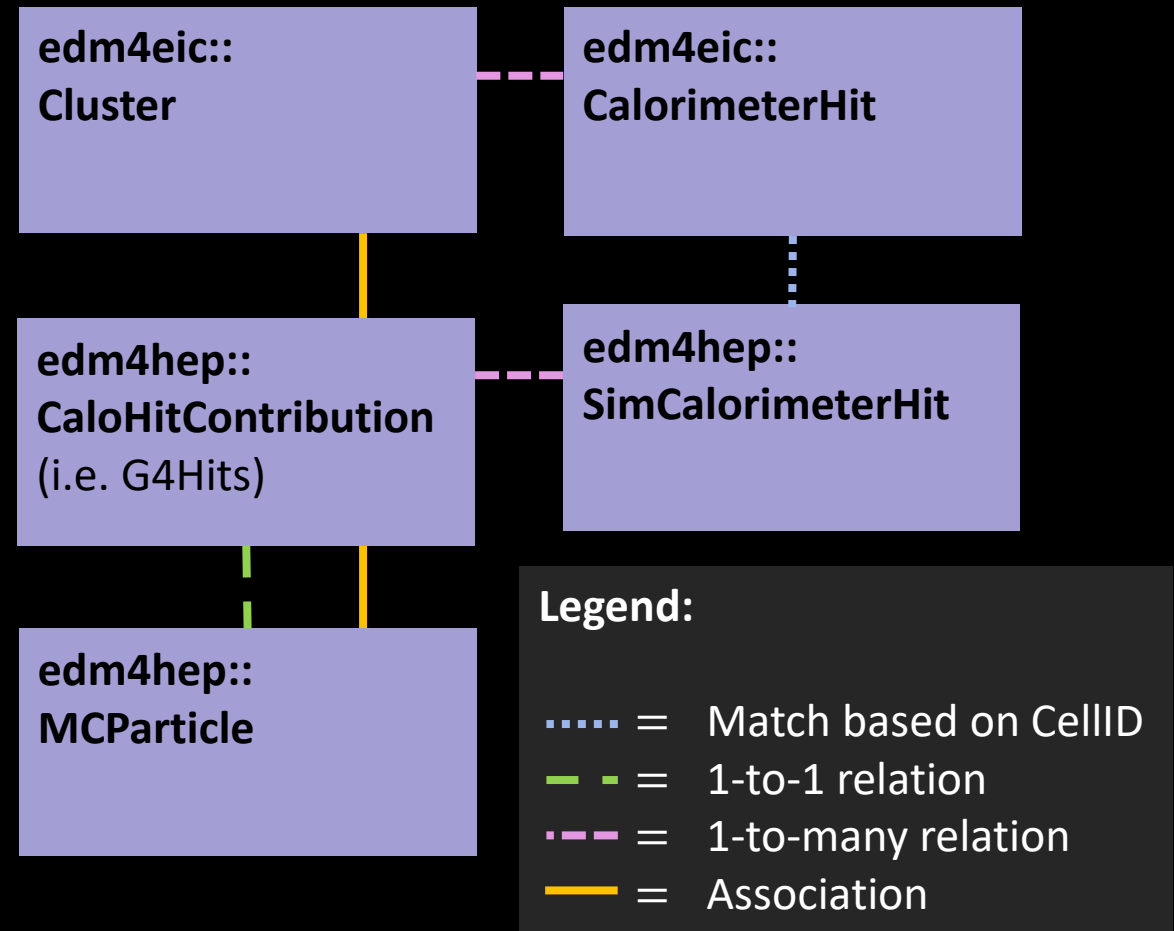
Improved Truth-Cluster Associations

Issues: EICrecon#898, EICrecon#899
PRs: EICrecon#1382 [+1 more to come]

Possible More Elaborate Change:
– Working on in PR#1382
– Idea: identify initiator of a shower which contributes to cluster, and associate the two
– How?
  1) Check if contributor start vertex is *outside* volume of subsystem
  2) If so, create association with weight given by eContrib / eCluster (or similar)

☞ **Thoughts?**

edm4eic:: Cluster

edm4eic:: CalorimeterHit

edm4hep:: CaloHitContribution (i.e. G4Hits)

edm4hep:: SimCalorimeterHit

edm4hep:: MCParticle

Legend:

..... = Match based on CellID

– – = 1-to-1 relation

–·– = 1-to-many relation

—— = Association

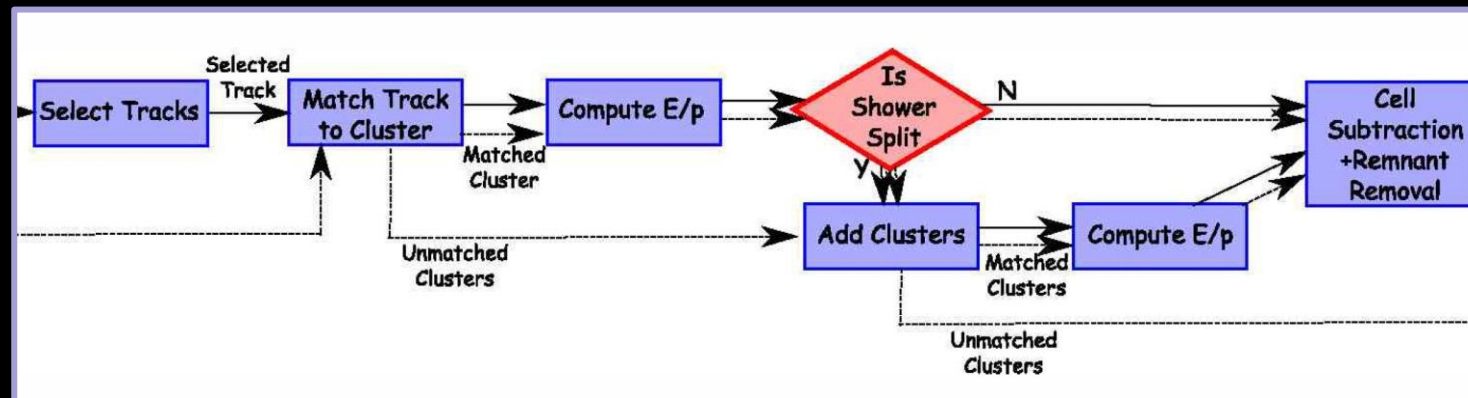## Cluster Merging

Issues: EICrecon#897, EICrecon#1289
PRs: [will open draft this week]

Context
- EICrecon#1289 under investigation by Akshaya
- We have cluster splitting capabilities in place…
- But we don't have any *merging* tools in place

Proposal
- Implement a track-based merging routine based on ATLAS's split recovery procedure
  - › c.f. Eur. Phys. J. C (2017) 77:466
  - › Figure illustrating routine from paper below

Cluster Merging

Issues: EICrecon#897, EICrecon#1289
PRs: [will open draft this week]

Track-Based Merging Algorithm
1) Match track projection to cluster
2) If matched,
   › Calculate significance b/n cluster energy
     & expected E_{dep}:

$$S(E_{clust}) = \frac{E_{clust} - (p_{proj} \times \langle E/p \rangle)}{\sigma(E_{dep})}$$

3) If $S < S_{cut}$, add clusters inside Δr_{add}

Note: could also make iterative…

Clusters
(output of
ClusterRecoCoG)

**Track-Cluster
Merger**

Merged Clusters
(output of
TrackClusterMerger)

Parameters: from single particle sim.s

Average E/p, $\langle E/p \rangle$
Spread of dep. energy, $\sigma(E_{dep})$
Threshold to run split-recovery, S_{cut}
Window to add clusters, Δr_{add}

<u>Cluster Merging</u>

Issues: EICrecon#897, EICrecon#1289
PRs: [will open draft this week]

Caveats
– Only works if track projections are available for a given calorimeter…

Notes
– Prepared a processor to generate histograms of splitting metrics (e.g. eClust / eSum)
  › Currently available in eic/snippets
  › Also calculates necessary parameters for algorithm
  › (Will convert into a benchmark…)

Clusters
(output of ClusterRecoCoG)

Track-Cluster Merger

Merged Clusters
(output of TrackClusterMerger)

Parameters: from single particle sim.s

Average E/p, $\langle E/p \rangle$
Spread of dep. energy, $\sigma(E_{dep})$
Threshold to run split-recovery, S_{cut}
Window to add clusters, Δr_{add}

Cluster Merging

Issues: EICrecon#897, EICrecon#1289
PRs: [will open draft this week]

Other Directions and Fallbacks
– Proposed algorithm will function as baseline
  › But with ONNX support in place and examples available…
  › Potential for quick turnaround on ML re-clustering routines?
– Worst case scenario:
  › **Is everything there for users to do standalone analyses?**

Clusters
(output of ClusterRecoCoG)

Merged Clusters
(output of TrackClusterMerger)

Track-Cluster Merger

Parameters: from single particle sim.s

Average E/p, $\langle E/p \rangle$
Spread of dep. energy, $\sigma(E_{dep})$
Threshold to run split-recovery, S_{cut}
Window to add clusters, Δr_{add}

Conclusions | Timelines

Timeline: truth-cluster associations
(minimal change)

1) PR opened today
2) Merged next couple days

Timeline: truth-cluster associations
(more elaborate change)

1) By end of week:
   › Integrate feedback,
   › complete to-do's,
   › Switch PR to open
2) Merged next week ahead of May campaign

Timeline: cluster merging

1) Draft PR for track-based merger/splitter opened by end of week
2) PR switched to open by **May 3rd**
   ☞ **Goal:** merged for testing **after** May campaign
3) Collecting feedback, tuning, debugging proceeds **May 6th – May 24th**
4) Any additional changes in by **May 31st** and merged **ahead of June campaign**

☞ **This is something we could use help with!!**

Conclusions | Questions for Discussion

– **Naïve question:** is there anything left out of the event record in MCParticles?
  › e.g. shower particles?

– Is the full (incl. CaloHitContributions) simulation output available for campaigns?
  › If so, how should we advertise it?

– Thoughts on the more elaborate association proposal?

– Thoughts on the proposed merging algorithm?

– Anything else?