Summary | TLDR

Status

☞ Fulfillment of all must-have TDR requests is possible by June

☞ **But need support!**

Needed Help/Feedback

- **We need help:** testing cluster merging algorithms (May 12th – May 26th)
- **For all DSCs:** are your clustering parameters okay? Yes or no?
- Look at PR#1349/PR#1410. Is this sufficient for noise masking?
- Look at PR#1396 and provide feedback if any

Summary | TDR Requests and Status

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	In Progress
Not Must Haves	Better neutral identification	To Do
	System-specific digitization models	To Do

Summary | EICrecon Issues, PRs, and Notes

Request	EICrecon Issues	EICrecon PRs/Notes
Improved truth-cluster connections	Issue#898, Issue#899	PR#1396, PR#1382
Clustering in all systems	Issue#1342	n/a
Cluster splitting/merging	Issue#897, Issue#1289	PR#1406
ML integration	Issue#1340, Issue#1394	PR#1358, PR#1414
Noise-masking, non-uniform gain/noise setting	Issue#1337	PR#1349, PR#1410
Easier access to janadot output	Issue#1339	In upcoming JANA2 release
Better neutral identification	Issue#1341	Standalone (but see PR#1404)
System-specific digitization models	Issue#1338, Issue#1081, Issue#696	For post TDR software freeze

Details | Not Must-Haves

[To Do] Better Neutral Identification

Issues: Issue#1341
PRs: PR#1404

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

[To Do] System Specific Digitization Models

Issues: Issue#1338, Issue#1081, Issue#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 software freeze

[In Progress] Easier Access to Janadot

Issues: Issue#1339
PRs: See upcoming JANA2 release

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
☞ Nathan working on better factory/algorithm/etc. reporting for upcoming JANA2 release

[Complete] Noise-Masking, etc.

Issues: Issue#1337
PRs: PR#1349, PR#1410

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
☞ Is this is sufficient?
☞ If not, what more do we need?

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
```

[Complete] Noise-Masking, etc.

Issues: Issue#1337
PRs: PR#1349, PR#1410

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
☞ Is this is sufficient?
☞ If not, what more do we need?

Details | Must-Haves (3/4)

[Complete] Clustering in all Systems

Issues: Issue#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
☞ Are current parameters okay for clustering in your system?

[Complete] ML Integration Example

Issues: Issue#1340, Issue#1394
PRs: PR#1358, PR#1414

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. Issue#1394)

[In Progress] Truth-Cluster Connections

Issues: Issue#898, Issue#899
PRs: PR#1396, PR#1382

Notes
– Longstanding question: are MC-Cluster associations working (Yes) and sufficient (TBD)
– Association logic update in progress in PR#1396

Follow-Up:
– Working on explainer of + snippet for cluster relations, associations, contributions, and what's in EICrecon output

[In Progress] Cluster Merging

Issues: Issue#897, Issue#1289
PRs: PR#1406

Notes
– Splitting functionality in place for a while, but no functionality for merging in EICrecon, though...
– Working on track-based merging algorithm in PR#1406
  › Peter exploring similar "calo-only" idea

Follow-Up:
– Testing PR#1406 is something we will need help with!!

Backup

Backup | TDR Requests (1/2)

○ 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

Backup | TDR Requests (2/2)

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**
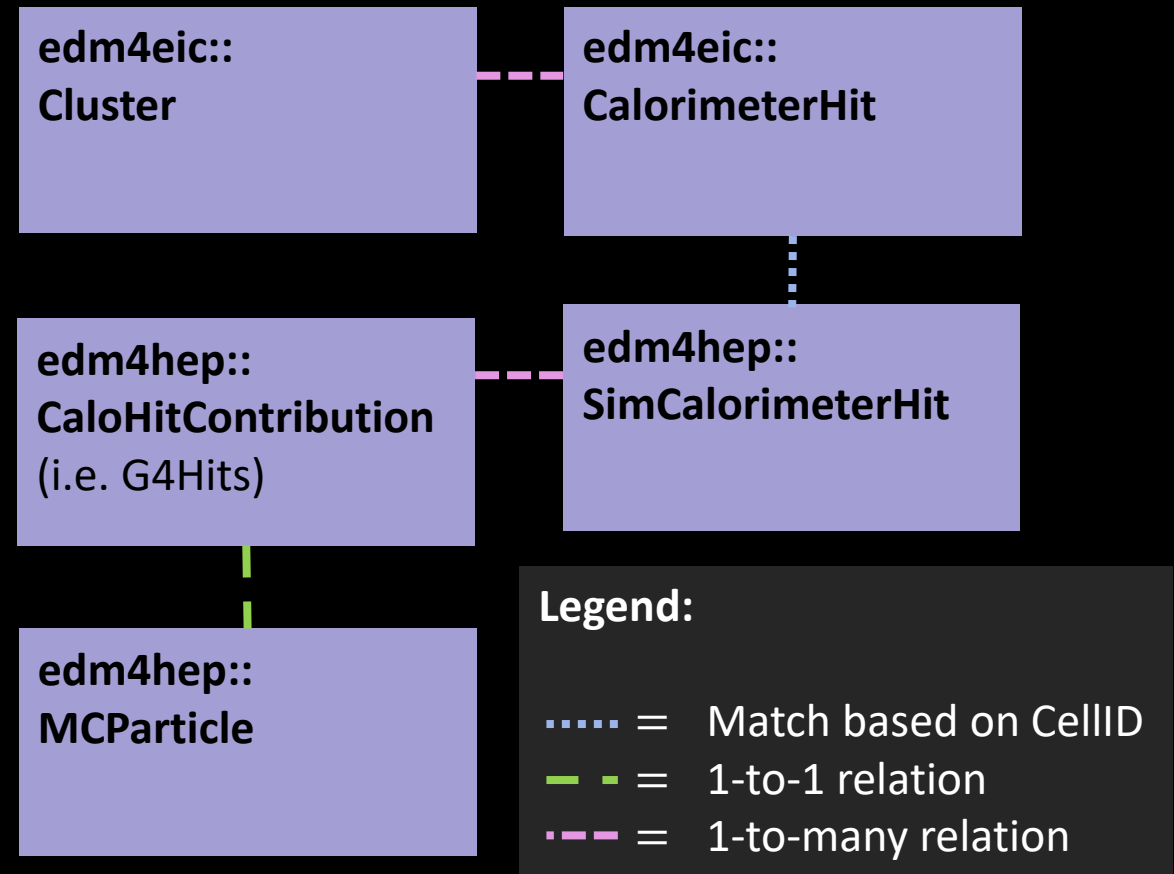
[In Progress] Truth-Cluster Connections

Issues: Issue#898, Issue#899
PRs: PR#1396, PR#1382

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
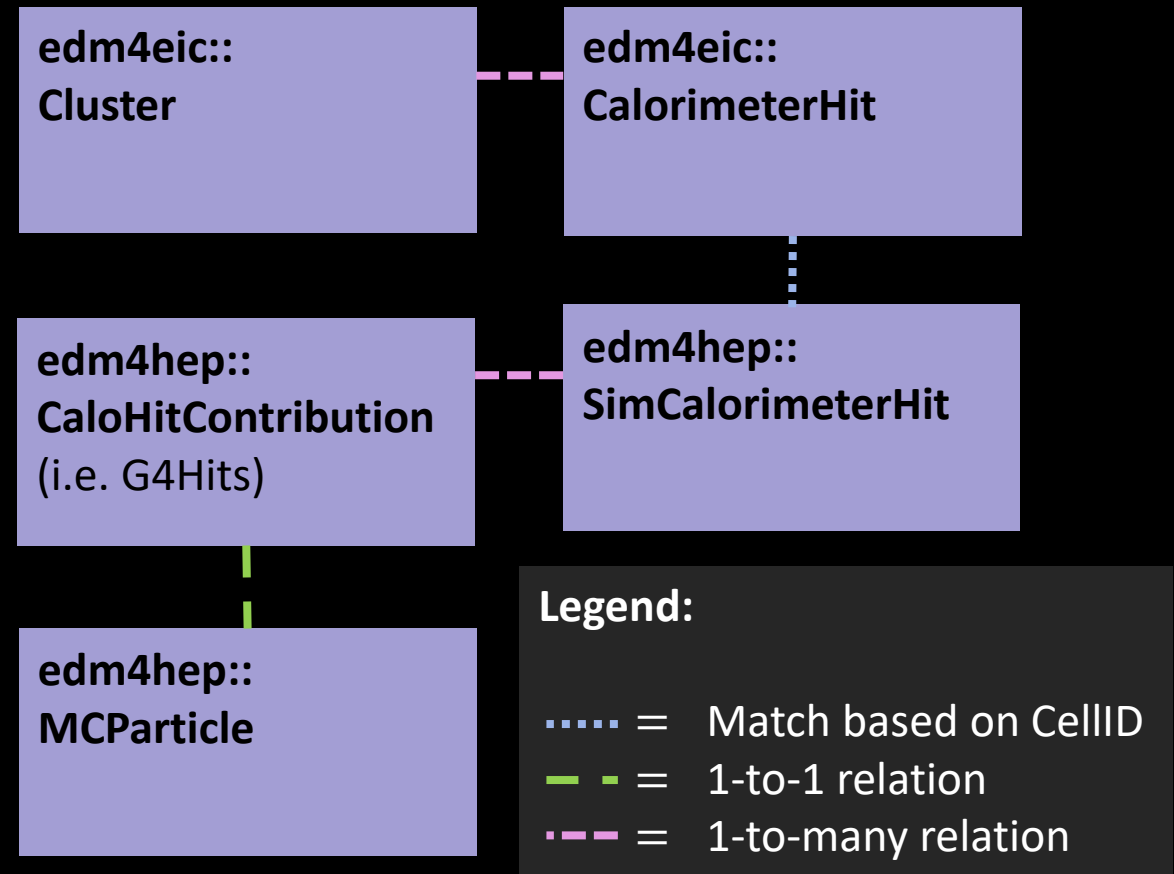– – = 1-to-1 relation
·– – = 1-to-many relation

[In Progress] Truth-Cluster Connections

Issues: Issue#898, Issue#899
PRs: PR#1396, PR#1382

So what's in the EICrecon output?
– Confirmed (as of this 04.23.2024) 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

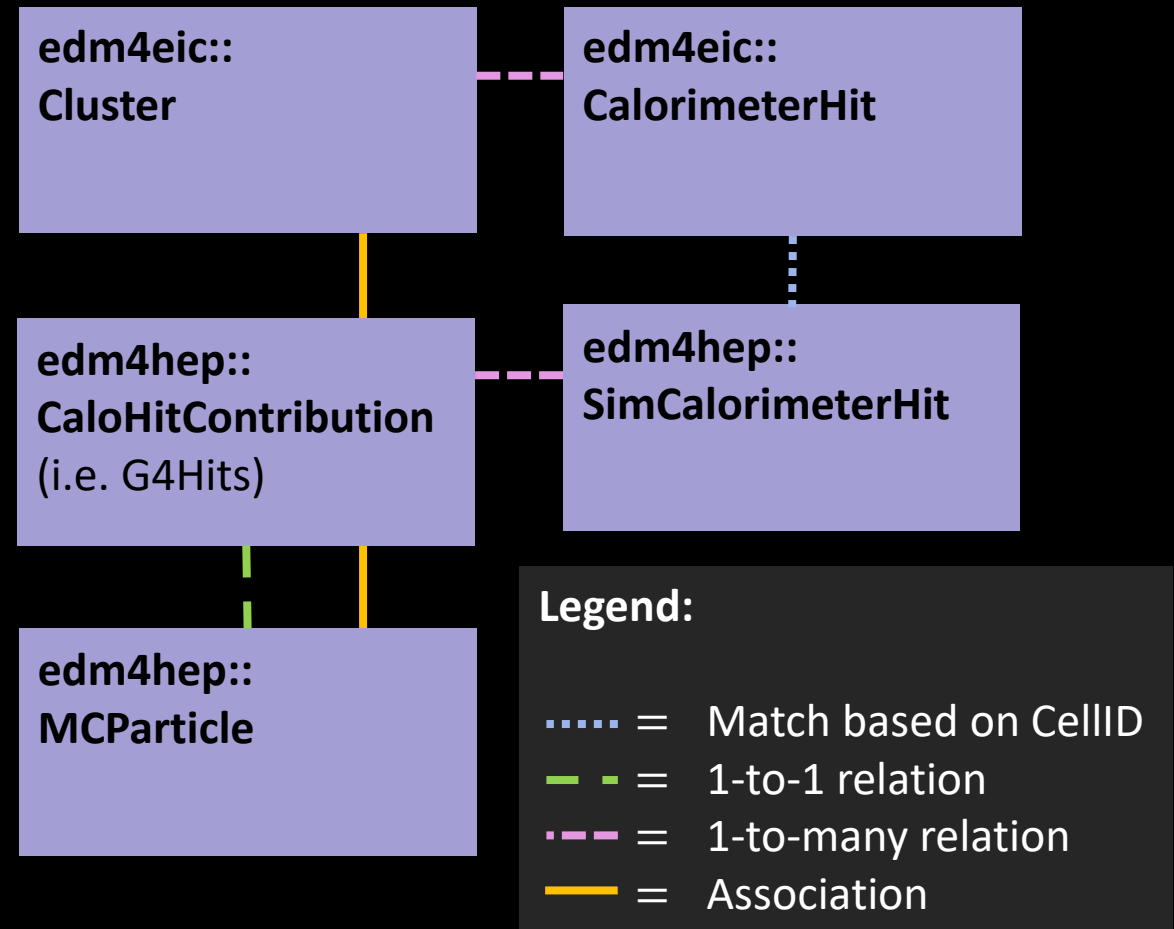**[In Progress] Truth-Cluster Connections**

Issues: Issue#898, Issue#899
PRs: PR#1396, PR#1382

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

[In Progress] Truth-Cluster Connections

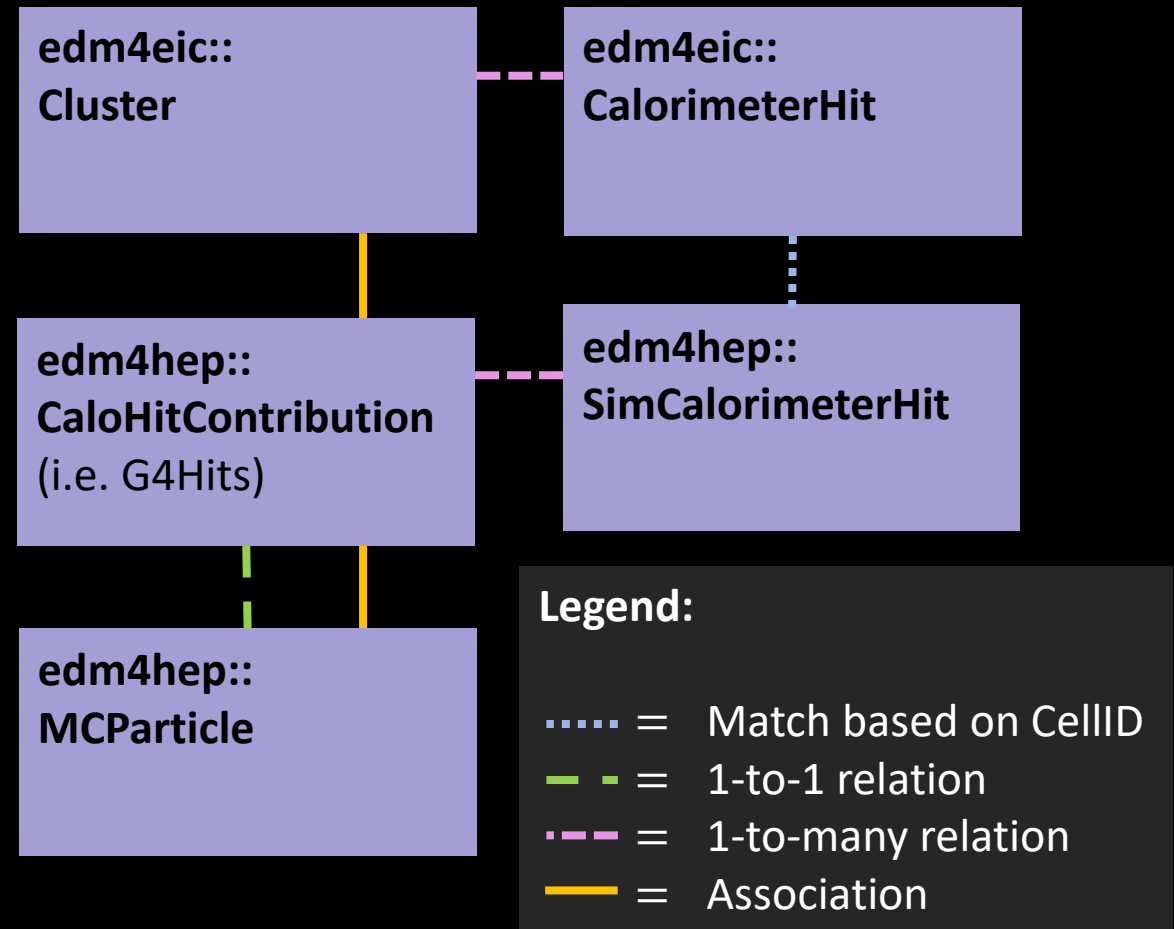**Issues:** Issue#898, Issue#899
PRs: PR#1396, PR#1382

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::          edm4eic::
Cluster            CalorimeterHit

edm4hep::          edm4hep::
CaloHitContribution   SimCalorimeterHit
(i.e. G4Hits)

edm4hep::
MCParticle
```

Legend:

..... = Match based on CellID
- - - = 1-to-1 relation
- - - = 1-to-many relation
——— = Association

[In Progress] Cluster Merging

Issues: Issue#897, Issue#1289
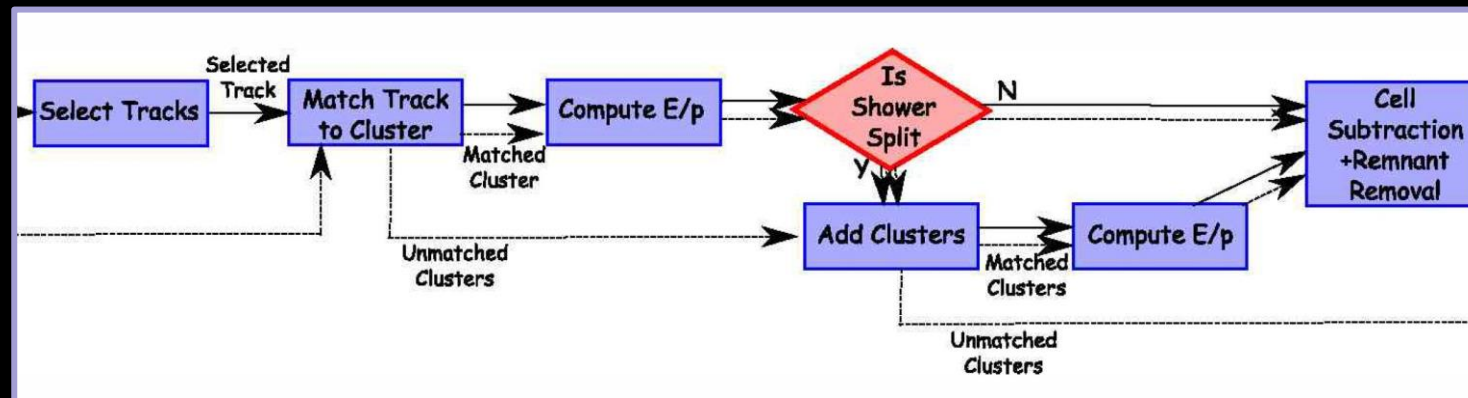**PRs:** PR#1406

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

[In Progress] Cluster Merging

Issues: Issue#897, Issue#1289
PRs: PR#1406

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)

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}

[In Progress] Cluster Merging

Issues: Issue#897, Issue#1289
PRs: PR#1406

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)

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}

Backup | Cluster Merging (4/4)

[In Progress] Cluster Merging

Issues: Issue#897, Issue#1289
PRs: PR#1406

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}

Backup | Connection & Merging Timelines

Timeline: truth-cluster associations (minimal change)

1) PR opened today
2) ~~Merged next couple days~~ PR needs some more discussion before merging

~~**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~~ **May 12th**
   - ☞ **Goal:** merged for testing **after** May campaign
3) Collecting feedback, tuning, debugging proceeds ~~May 6th~~ **May 12th – May 26th**
4) Any additional changes in by **May 31st** and merged **ahead of June campaign**

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