MC particle to hit association definition and implications for track reconstruction metrics

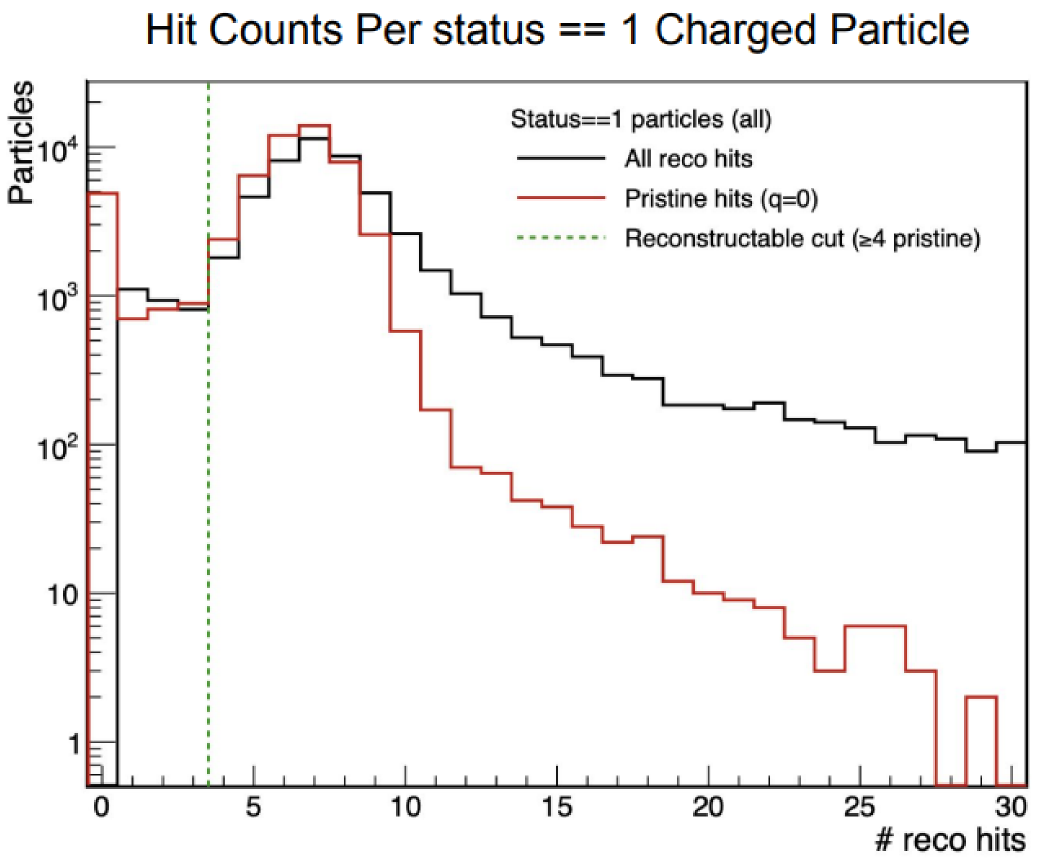
Barak Schmookler & Jeetendra Gupta

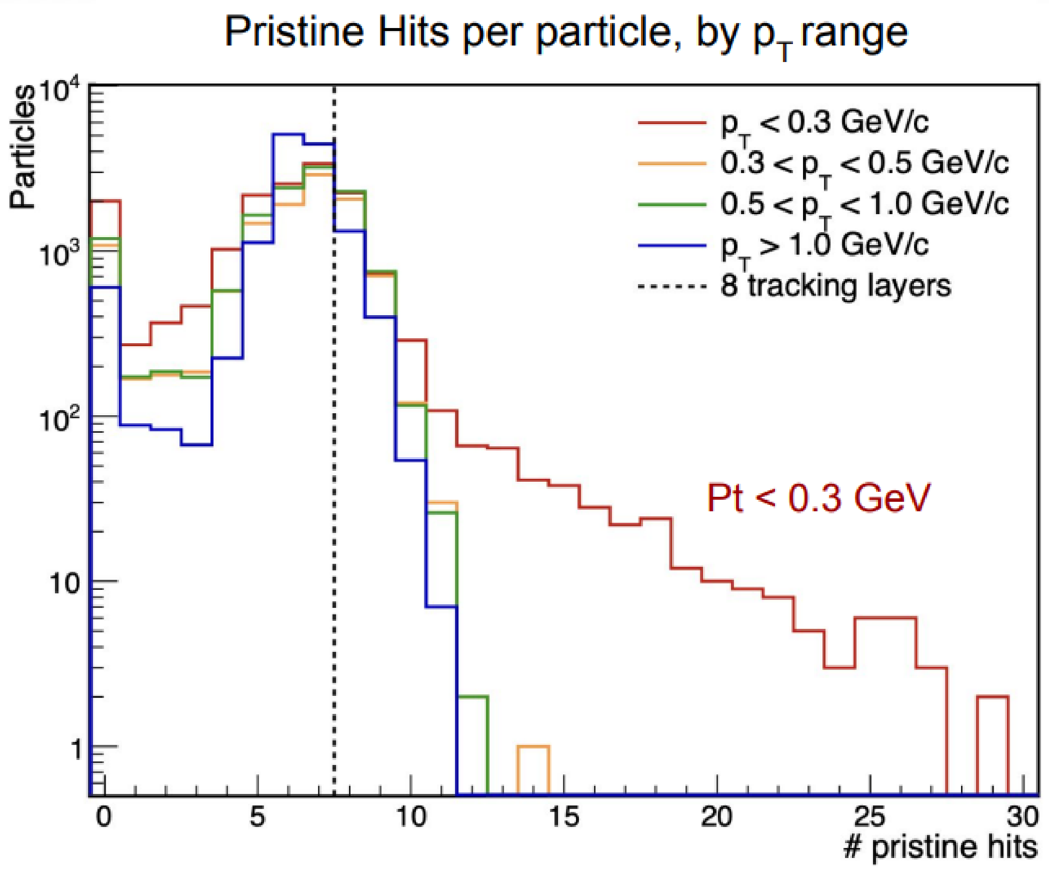
Follow up to presentations on 06/11/26 and 07/16/26

Motivation

- When studying track reconstruction performance – tracking efficiency, tracking purity, track resolutions – we need a method to match MC particles to reconstructed tracks.
- If using tracking detector hit-based matching, we can use some combination of the metrics below to define the matching criterion:
  - track hit efficiency – what fraction of an MC particle's hits are used in a given track fit
  - and track hit purity – what fraction of a track's hits originate from a given MC particle
- In DD4hep, there are some complications to defining whether a hit is associated to a specific MC particle, and this can have an impact on the quoted tracking performances.

Tracking detector hits associated to MC particles





The tracking detector hits linked to a final-state, signal particle (i.e., status == 1 particle) have a quality flag.

- Quality = 0 means that the hit 'came' from that particle. We call these hits "pristine" in the above plots.
- Quality ≠ 0 means that the hit came from a secondary particle, but that secondary particle had too small an energy to be saved to the MC Particle list.

Example from event with single charged pion at η = 3.6

```
TrackerEndcapHits = (vector<edm4hep::SimTrackerHitData>*)0x5610ef8915d0
TrackerEndcapHits.cellID = 1586815173047565 1919, 15973986210049709135, 16019867786599748687, 16021556662229816399
TrackerEndcapHits.eDep = 0.000011, 0.000035, 0.000030, 0.000050
TrackerEndcapHits.time = 3.173552, 4.008810, 4.960220, 4.961260
TrackerEndcapHits.pathLength = 0.040061, 0.040061, 0.068549, 0.055542
TrackerEndcapHits.quality = 0, 0, 1073741824, 1073741824
TrackerEndcapHits.position.x = -39.6739, -49.8942, -55.8866, -56.0502
TrackerEndcapHits.position.y = -33.7965, -43.0422, -41.0262, -41.0045
TrackerEndcapHits.position.z = 949.865, 1199.87, 1199.87, 1199.87
TrackerEndcapHits.momentum.x = -0.410268, -0.406529, -0.000413, -0.000124
TrackerEndcapHits.momentum.y = -0.367710, -0.370419, 0.000147, 0.000112
TrackerEndcapHits.momentum.z = 9.980886, 9.980414, 0.000381, -0.000342
_TrackerEndcapHits_particle = (vector<podio::ObjectID>*)0x5610f6f7e8c0
_TrackerEndcapHits_particle.index = 0, 0, 0, 0
_TrackerEndcapHits_particle.collectionID = 2714477136, 2714477136, 2714477136, 2714477136
```

Example from event with single charged pion at $\eta = 3.6$

```
TrackerEndcapHits = (vector<edm4hep::SimTrackerHitData>*)0x5610ef8915d0
TrackerEndcapHits.cellID = 15868151730475651919, 15973986210049709135, 16019867786599748687, 16021556662229816399
TrackerEndcapHits.eDep = 0.000011, 0.000035, 0.000030, 0.000050
TrackerEndcapHits.time = 3.173552, 4.008810, 4.960220, 4.961260
TrackerEndcapHits.pathLength = 0.040061, 0.040061, 0.068549, 0.055542
TrackerEndcapHits.quality = 0, 0, 1073741824, 1073741824
TrackerEndcapHits.position.x = -39.6739, -49.8942, -55.8866, -56.0502
TrackerEndcapHits.position.y = -33.7965, -43.0422, -41.0262, -41.0045
TrackerEndcapHits.position.z = 949.865, 1199.87, 1199.87, 1199.87
TrackerEndcapHits.momentum.x = -0.410268, -0.406529, -0.000413, -0.000124
TrackerEndcapHits.momentum.y = -0.367710, -0.370419, 0.000147, 0.000112
TrackerEndcapHits.momentum.z = 9.980886, 9.980414, 0.000381, -0.000342
 TrackerEndcapHits_particle = (vector<podio::ObjectID>*)0x5610f6f7e8c0
_TrackerEndcapHits_particle.index = 0, 0, 0, 0
_TrackerEndcapHits_particle.collectionID = 2714477136, 2714477136, 2714477136, 2714477136
```

- There are four hits linked to the generated (10 GeV/c) charged pion.

Example from event with single charged pion at η = 3.6

```
TrackerEndcapHits = (vector<edm4hep::SimTrackerHitData>*)0x5610ef8915d0
TrackerEndcapHits.cellID = 1586815173047565l919, 15973986210049709135, 16019867786599748687, 16021556662229816399
TrackerEndcapHits.eDep = 0.000011, 0.000035, 0.000030, 0.000050
TrackerEndcapHits.time = 3.173552, 4.008810, 4.960220, 4.961260
TrackerEndcapHits.pathLength = 0.040061, 0.040061, 0.068549, 0.055542
TrackerEndcapHits.quality = 0, 0, 1073741824, 1073741824
TrackerEndcapHits.position.x = -39.6739, -49.8942, -55.8866, -56.0502
TrackerEndcapHits.position.y = -33.7965, -43.0422, -41.0262, -41.0045
TrackerEndcapHits.position.z = 949.865, 1199.87, 1199.87, 1199.87
TrackerEndcapHits.momentum.x = -0.410268, -0.406529, -0.000413, -0.000124
TrackerEndcapHits.momentum.y = -0.367710, -0.370419, 0.000147, 0.000112
TrackerEndcapHits.momentum.z = 9.980886, 9.980414, 0.000381, -0.000342
_TrackerEndcapHits_particle = (vector<podio::ObjectID>*)0x5610f6f7e8c0
_TrackerEndcapHits_particle.index = 0, 0, 0, 0
_TrackerEndcapHits_particle.collectionID = 2714477136, 2714477136, 2714477136, 2714477136
```

- There are four hits linked to the generated (10 GeV/c) charged pion.
- However, only two of the hits directly come from the charged pion. The other two come from low-energy secondaries.

Example from event with single charged pion at η = 3.6

```
TrackerEndcapHits = (vector<edm4hep::SimTrackerHitData>*)0x5610ef8915d0
TrackerEndcapHits.cellID = 15868151730475651919, 15973986210049709135, 16019867786599748687, 16021556662229816399
TrackerEndcapHits.eDep = 0.000011, 0.000035, 0.000030, 0.000050
TrackerEndcapHits.time = 3.173552, 4.008810, 4.960220, 4.961260
TrackerEndcapHits.pathLength = 0.040061, 0.040061, 0.068549, 0.055542
TrackerEndcapHits.quality = 0, 0, 1073741824, 1073741824
TrackerEndcapHits.position.x = -39.6739, -49.8942, -55.8866, -56.0502
TrackerEndcapHits.position.y = -33.7965, -43.0422, -41.0262, -41.0045
TrackerEndcapHits.position.z = 949.865, 1199.87, 1199.87, 1199.87
TrackerEndcapHits.momentum.x = -0.410268, -0.406529, -0.000413, -0.000124
TrackerEndcapHits.momentum.y = -0.367710, -0.370419, 0.000147, 0.000112
TrackerEndcapHits.momentum.z = 9.980886, 9.980414, 0.000381, -0.000342
_TrackerEndcapHits_particle = (vector<podio::ObjectID>*)0x5610f6f7e8c0
_TrackerEndcapHits_particle.index = 0, 0, 0, 0
_TrackerEndcapHits_particle.collectionID = 2714477136, 2714477136, 2714477136, 2714477136
```

- There are four hits linked to the generated (10 GeV/c) charged pion.
- However, only two of the hits directly come from the charged pion. The other two come from low-energy secondaries.
- In this case, the hits coming from the low-energy secondaries are on the same layer as one of the direct hits, but a few millimeters away.

Example from event with single charged pion at η = 3.6

```
TrackerEndcapHits = (vector<edm4hep::SimTrackerHitData>*)0x5610ef8915d0
TrackerEndcapHits.cellID = 15868151730475651919, 15973986210049709135, 16019867786599748687, 16021556662229816399
TrackerEndcapHits.eDep = 0.000011, 0.000035, 0.000030, 0.000050
TrackerEndcapHits.time = 3.173552, 4.008810, 4.960220, 4.961260
TrackerEndcapHits.pathLength = 0.040061, 0.040061, 0.068549, 0.055542
TrackerEndcapHits.quality = 0, 0, 1073741824, 1073741824
TrackerEndcapHits.position.x = -39.6739, -49.8942, -55.8866, -56.0502
TrackerEndcapHits.position.y = -33.7965, -43.0422, -41.0262, -41.0045
TrackerEndcapHits.position.z = 949.865, 1199.87, 1199.87, 1199.87
TrackerEndcapHits.momentum.x = -0.410268, -0.406529, -0.000413, -0.000124
TrackerEndcapHits.momentum.y = -0.367710, -0.370419, 0.000147, 0.000112
TrackerEndcapHits.momentum.z = 9.980886, 9.980414, 0.000381, -0.000342
_TrackerEndcapHits_particle = (vector<podio::ObjectID>*)0x5610f6f7e8c0
_TrackerEndcapHits_particle.index = 0, 0, 0, 0
_TrackerEndcapHits_particle.collectionID = 2714477136, 2714477136, 2714477136, 2714477136
```

If we choose to decrease the MCParticle list storage threshold in the ddsim (npsim) steering file, then those hits would no longer be linked to the generated pion.

```
## MinimalKineticEnergy to store particles created in the tracking region
SIM.part.minimalKineticEnergy = 1.0
```

Implications for track reconstruction performance

- We studied the impact of two hit-to-particle association definitions:
    1. Definition 1: For a given MC particle, we define the hits associated to that particle to be all quality = 0 and quality ≠ 0 hits that DD4hep links to that particle.
    2. Definition 2: For a given MC particle, we define the hits associated to that particle to be only the quality = 0 hits that DD4hep links to that particle.
- In both cases, we use all hits (i.e., fit measurements) associated to a given track fit. This includes hits from signal particles, hits from background particles, hits from higher energy secondaries (that are in the MCParticle list) and hits from lower energy secondaries (that are not in the MCParticle list).
- We then apply the **double majority requirement** (> 50% hit efficiency and hit purity) to define the particle-track matching and look at the resultant tracking efficiency and tracking purity.

Implications for track reconstruction performance

Results from 10,000 events from 26.03 campaign
10x100 GeV setting with $Q^2 > 1$ GeV^2 and all beam backgrounds included

	Definition 1	Definition 2
Number of Reconstructable MC Particles (at least 4 associated tracking detector hits)	50785	46352
Number of Valid Tracks (at least 5 fit measurements on track)	36106	36106
Double-majority matches	33448	35240
Tracking efficiency	65.86%	76.03%
Tracking purity	92.64%	97.60%

Implications for track reconstruction performance

Results from 10,000 events from 26.03 campaign
10x100 GeV setting with $Q^2 > 1$ GeV^2 and all beam backgrounds included

	Definition 1	Definition 2
Number of Reconstructable MC Particles (at least 4 associated tracking detector hits)	50785	46352
Number of Valid Tracks (at least 5 fit measurements on track)	36106	36106
Double-majority matches	33448	35240
Tracking efficiency	65.86%	76.03%
Tracking purity	92.64%	97.60%

The particle to hit association definition we choose has an impact on the values we quote for track efficiency and purity.

Conclusions

- The way we define the association between the generated MC signal particles and the tracking detector hits affects the values we extract for the various tracking metrics.
- In general, it makes more sense to only count quality = 0 hits as associated to a given MC particle, since those are stable to the *minimalKineticEnergy* threshold set in the simulation steering file. But there may be exceptions we want to consider.
- We should think about what we want to do inside EICRecon (e.g., in the track->particle hit association) and in the tracking benchmarks we are writing.