Double Majority Rule
Benchmark development for tracking with backgrounds

Jeetendra Gupta (University of Illinois Chicago)
Barak Schmookler (University of Houston)
Austin Baty (University of Illinois Chicago)

Joint ePIC/EICUG Meeting, GlasGow
July 16, 2026

UIC group's work is supported by US DOE Office of Science

Analysis Flow

Set 1 : Reconstructable Particle

(every final-state, charged MC particle that left at least 4 pristine hits in the tracker)

- CentralTrackingRawHitAssociations → getSimHit() → getParticle()
- Keep status==1 and charge ≠ 0 particle
- getRawHit() → hit key (joins to TrackerHitCollection hits; only computed for particles that passed the cut above)
- Apply quality==0 for pristine hits
- Keep particle with ≥4 pristine hits

Set 2 : CKF Track

(every reconstructed CKF track that has at least 4 (or 5) measurements)

- CentralCKFTracks/CentralCKFTruthSeededTracks
- Keep tracks with ≥4 (or ≥5) measurements
- getMeasurements() → getHits()
- hit.getObjectID() → matching key (same key type as getRawHit() in Set 1)

Double Majority Rule

Do one-to-one matching by applying double majority rule (hit efficiency > 50%, hit purity > 50%)

Note: Using March campaign files → root://dtn-eic.jlab.org//volatile/eic/EPIC/RECO/26.03.0/epic_craterlake/Bkg_Exact1S_2us/GoldCt/10um/DIS/NC/10x100/minQ2=1/pythia8NCDIS_10x100_minQ2=1_beamEffects_xAngle=-0.025_hiDiv_1.0000.eicrecon.edm4eic.root

Set 1 Reconstructable particles

For example, I take one event and see how many particles are there, how many of them have at least 4 quality==0 hits (pristine hits)

Rec hits on Tracker

Event 9955 | Collection: SiEndcapTrackerRecHits | rechits: 2157
Event 9955 | Collection: SiBarrelVertexRecHits | rechits: 1469
Event 9955 | Collection: SiBarrelTrackerRecHits | rechits: 591
Event 9955 | Collection: MPGDBarrelRecHits | rechits: 97
Event 9955 | Collection: OuterMPGDBarrelRecHits | rechits: 28
……

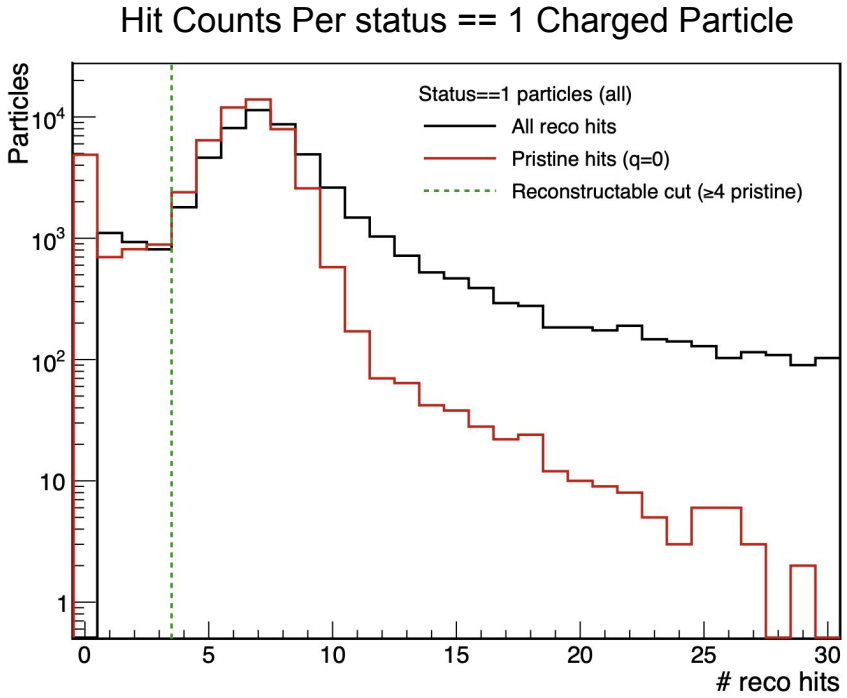
Final State Particles

Event 9955 has 5 **status==1** charged particles with tracker hits
--> **Reco-hit reconstructable** = 5
--> **Pristine-hit reconstructable** = 4

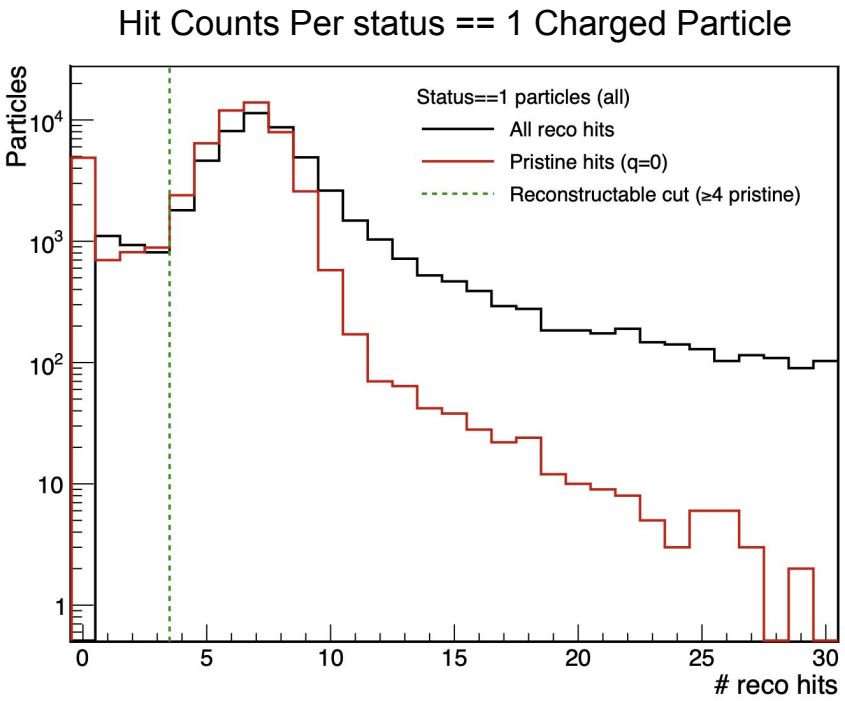
For example:
Particle #14 has 13 reco hits and 6 pristine hits (quality==0)
Particle #16 has 24 reco hits and 3 pristine hits (quality==0)

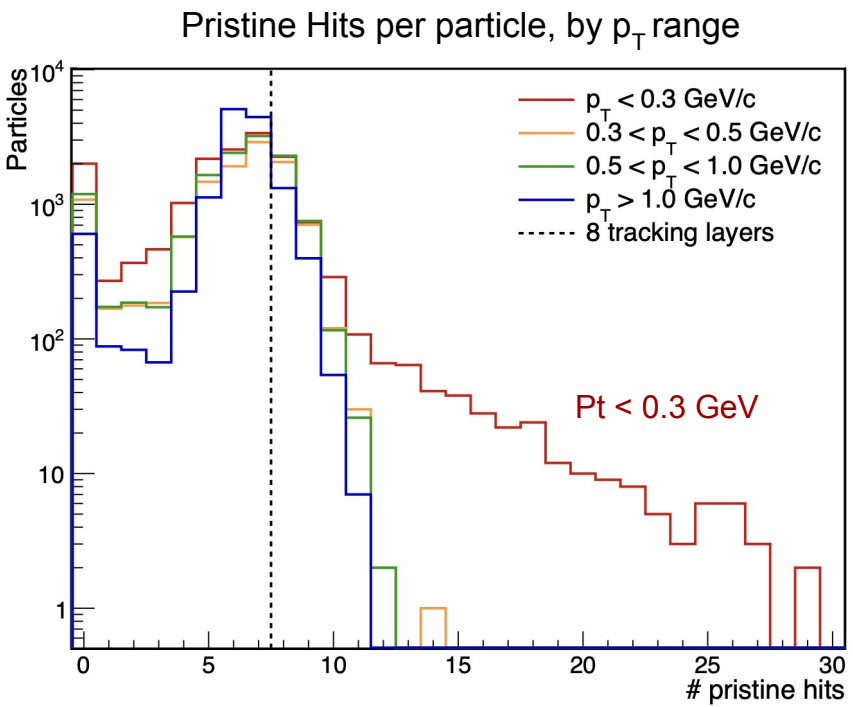
A lot of background hits but only a few final state charged particles

Set 1 Reconstructable particles



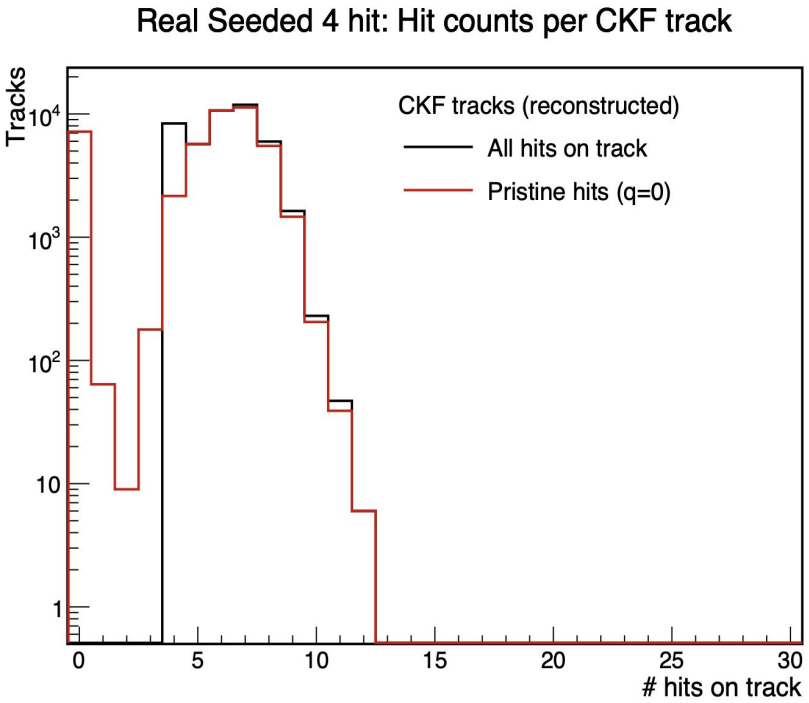
Set 1 Reconstructable particles

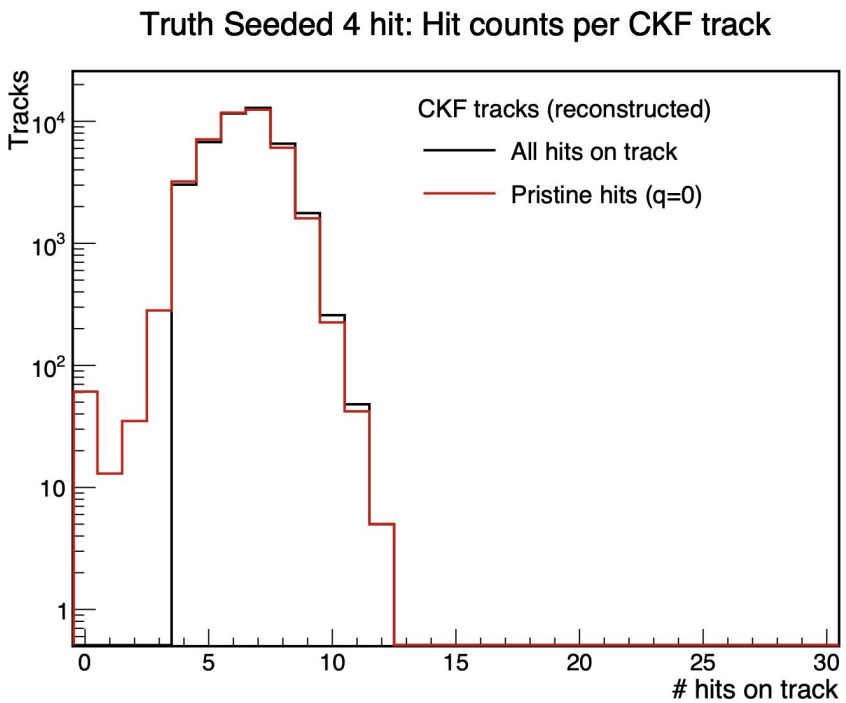




Pristine hits > 14 are coming from low pt looper

Set 2 CKF Track





No Loopers in CKF Tracks in both Real Seeded and Truth Seeded tracks
CKF Tracks' hits are almost all pristine hits

Double Majority Rule

Defintions:

c(t,p) = number of (pristine) hits shared between track ‘t’ and particle ‘p’
c(:,p) = total (pristine) hits (quality==0) belonging to particle ‘p’
c(t,:) = total (pristine) hits on the track (signal + background + noise)

(We are using pristine hits for calculations)

Hit efficiency = c(t,p) / c(:,p)
Hit purity = c(t,p) / c(t.:)

Track Efficiency = $N_{\text{Matched Particle,Track}}$ / $N_{\text{Reconstructable Particle}}$

(Set 1)

Track Purity = $N_{\text{Matched Particle,Track}}$ / $N_{\text{all Set 2 Tracks}}$

Double Majority Rule: Hit Efficiency > 50% , Hit Purity > 50%

Double Majority Rule

Config: CKF tracks require ≥4 hits

10k Events:

Real Seeded ≥4 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 44480
Double-majority matches: 36927
Track Efficiency: 79.6665%
Track Purity: 83.0193% (Fake Rate: 16.9807%)
Pure tracks: 36926 Mixed tracks: 1 Pure fraction: 99.9973%

Truth Seeded ≥4 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 42894
Double-majority matches: 41948
Track Efficiency: 90.4988%
Track Purity: 97.7946% (Fake Rate: 2.20544%)
Pure tracks: 41947 Mixed tracks: 1 Pure fraction: 99.9976%

Pure tracks → hit purity > 0.99, Mixed tracks → hit purity ≤ 0.99 for matched tracks

Truth seeded track for ≥ 4 hit have 10% higher **track efficiency** and **track purity**

Double Majority Rule

Config: CKF tracks require ≥4 hits

Real Seeded ≥4 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 44480
Double-majority matches: 36927
Track Efficiency: 79.6665%
Track Purity: 83.0193% (Fake Rate: 16.9807%)
Pure tracks: 36926 Mixed tracks: 1 Pure fraction: 99.9973%

Truth Seeded ≥4 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 42894
Double-majority matches: 41948
Track Efficiency: 90.4988%
Track Purity: 97.7946% (Fake Rate: 2.20544%)
Pure tracks: 41947 Mixed tracks: 1 Pure fraction: 99.9976%

Config: CKF tracks require ≥ 5 hits

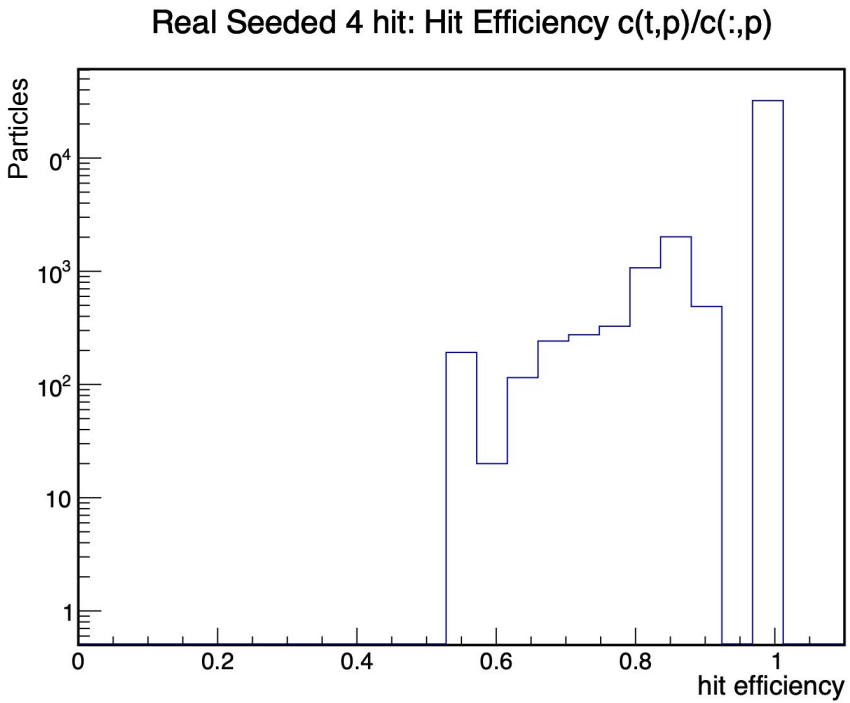
Real Seeded ≥5 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 36106
Double-majority matches: 35242
Track Efficiency: 76.0312%
Track Purity: 97.607% (Fake Rate: 2.39295%)
Pure tracks: 35241 Mixed tracks: 1 Pure fraction: 99.9972%

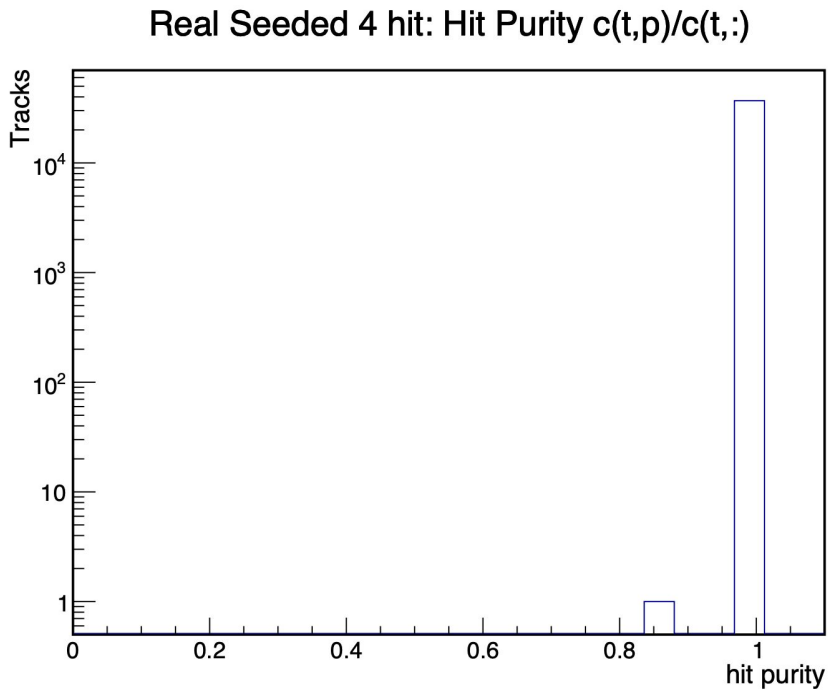
Truth Seeded ≥5 hit
Set 1 (reconstructable): 46352
Set 2 (track candidates): 39865
Double-majority matches: 39487
Track Efficiency: 85.1894%
Track Purity: 99.0518% (Fake Rate: 0.9482%)
Pure tracks: 39486 Mixed tracks: 1 Pure fraction: 99.9975%

Pure tracks → hit purity > 0.99, Mixed tracks → hit purity ≤ 0.99 for matched tracks

Truth seeded track for ≥ 4 hit have 10% higher track efficiency and track purity
Track purity increases by ~15% for ≥ 5 hit track

Hit Efficiency, Hit Purity for matched ≥ 4 hit track





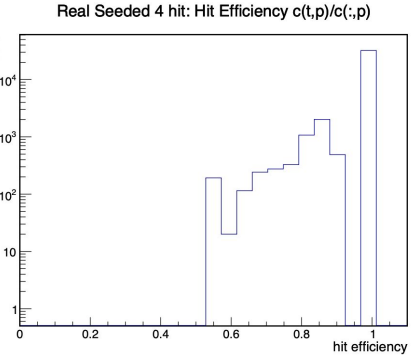
Hit Efficiency ~1, Hit Purity = 1

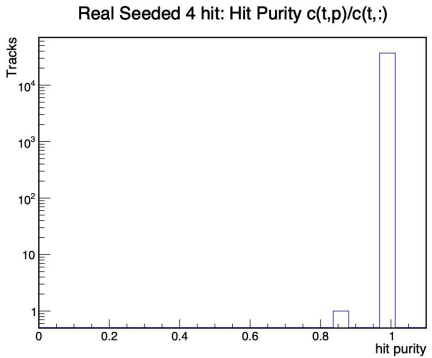
≥ 4 hit track

Efficiency

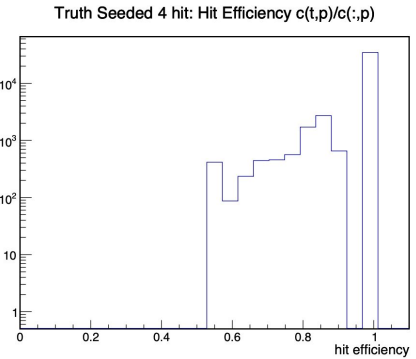
Purity

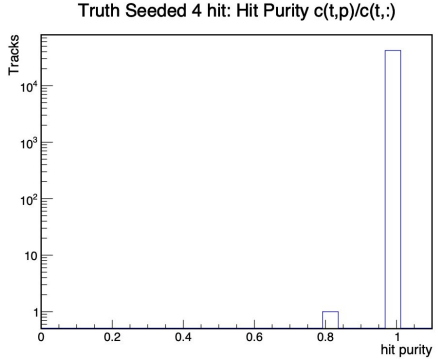
Real Seeded





Truth Seeded

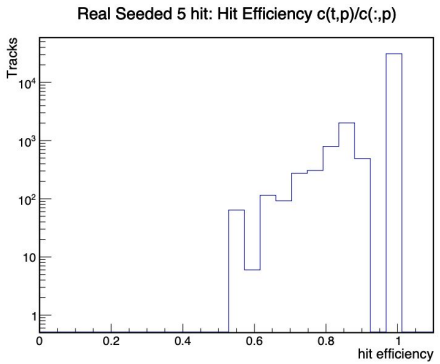


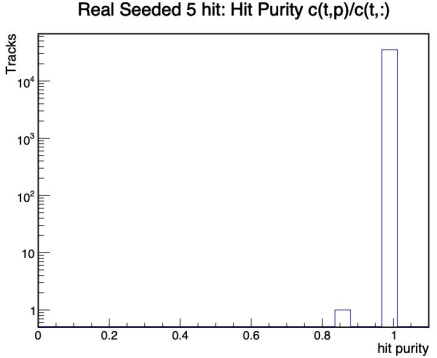


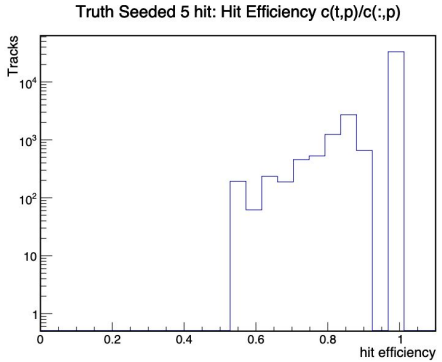
≥ 5 hit track

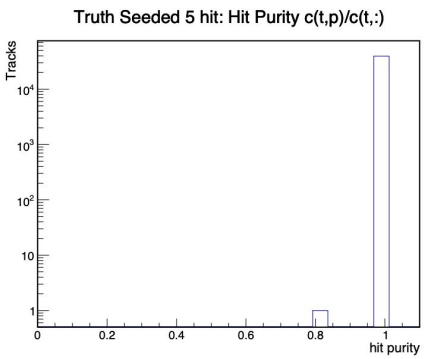
Efficiency

Purity





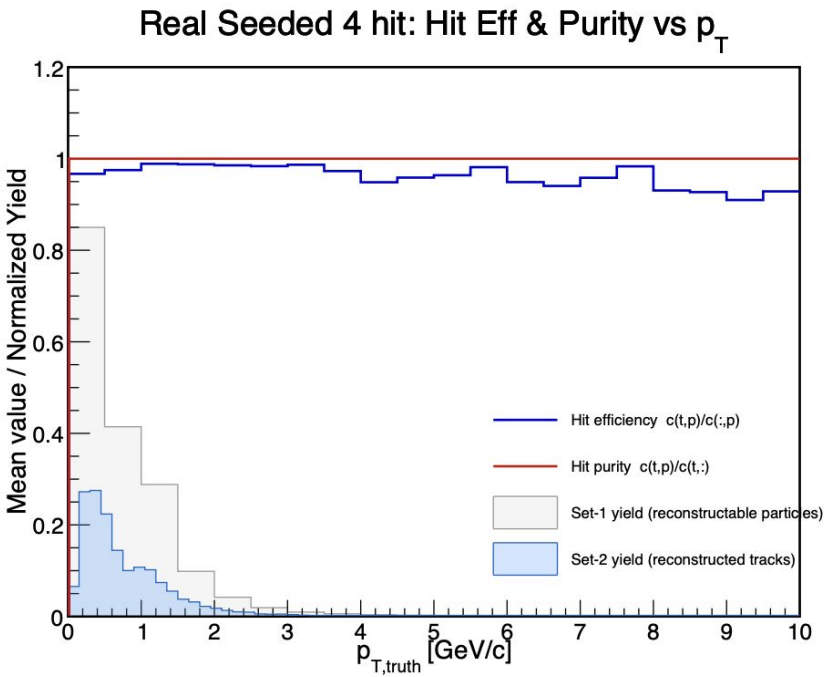


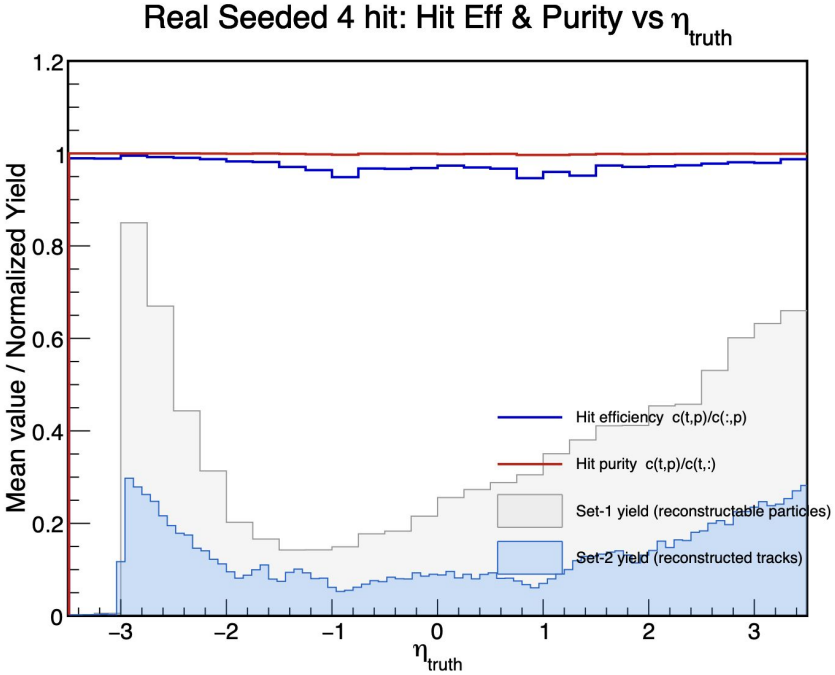


Real seeded and Truth Seeded have similar distribution

≥ 5 hit track: Hit Efficiency is mostly ~1, Hit Purity is also mostly ~1

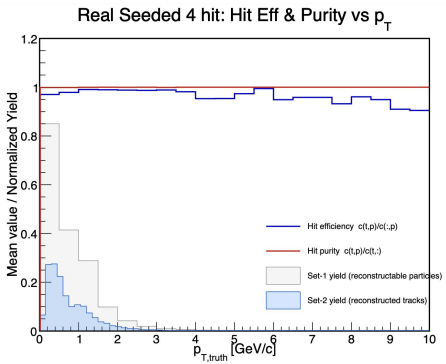
Hit Efficiency, Hit Purity vs pt, η for matched ≥ 4 hit track

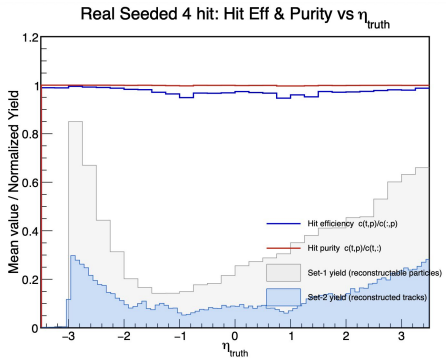


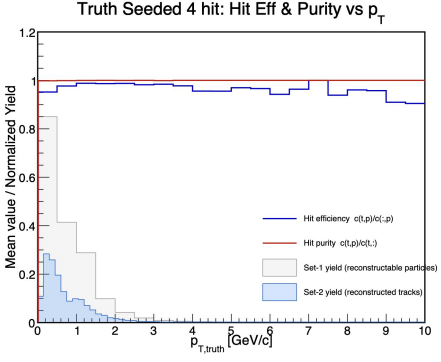


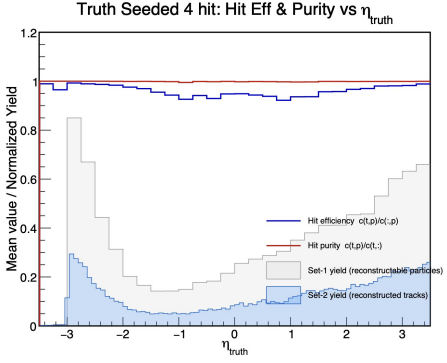
Track purity is ~1 for every pt and η
Hit efficiency decreases for higher pt, and for mid rapidity region

≥ 4 hit track

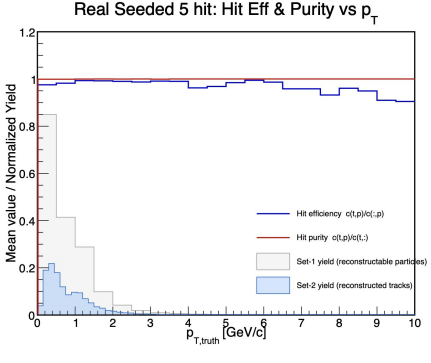


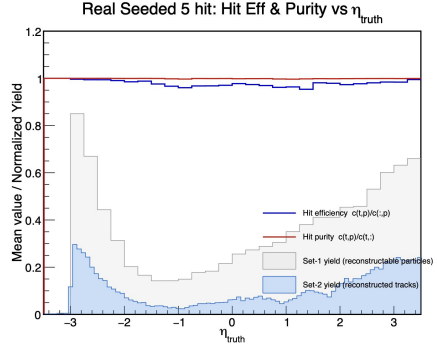


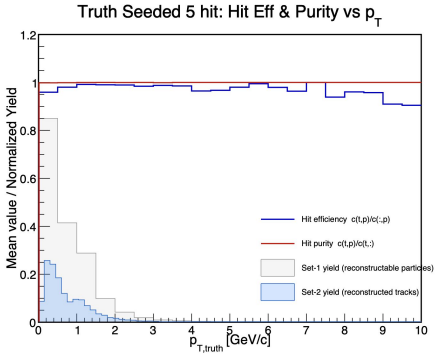


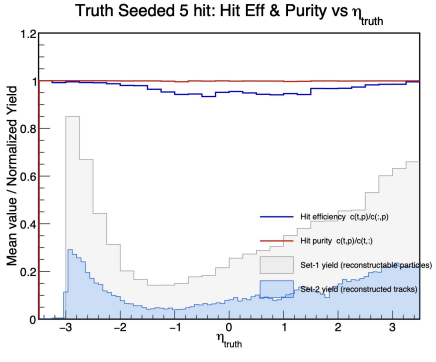


≥ 5 hit track



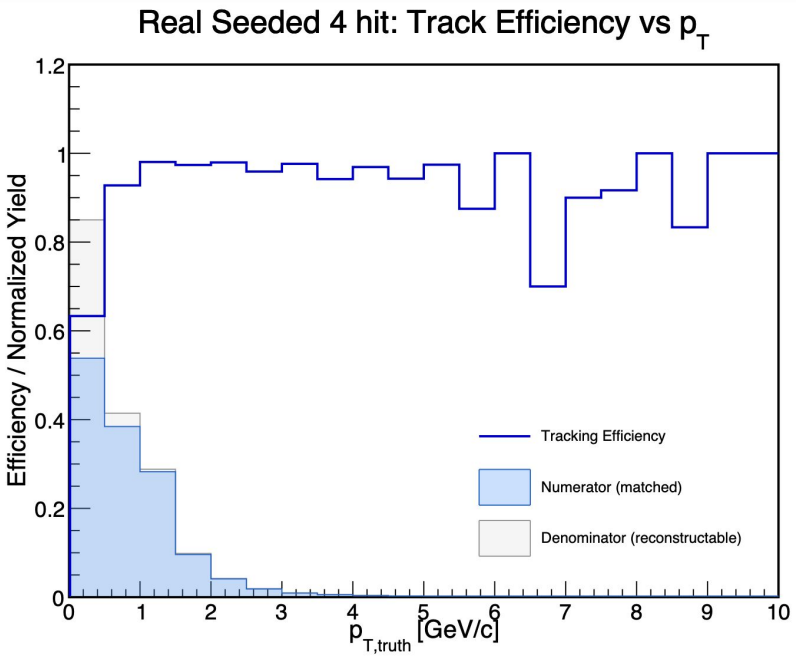


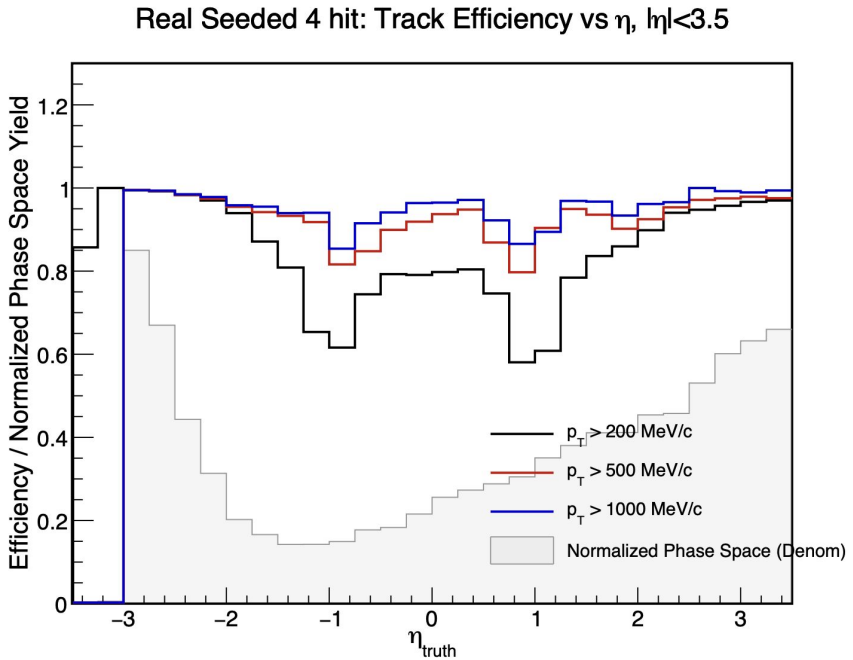




Similar observation for ≥ 5 hit track as well as for truth seeded tracks

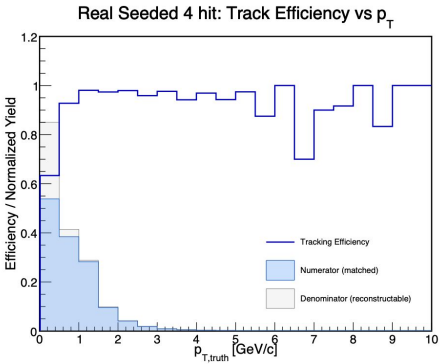
Track Efficiency vs pt and η for matched ≥ 4 hit track

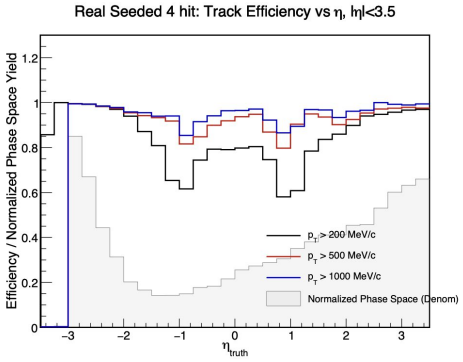


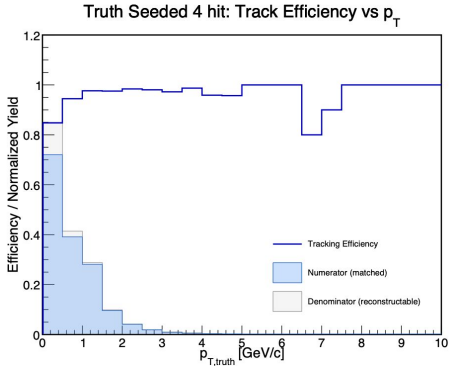


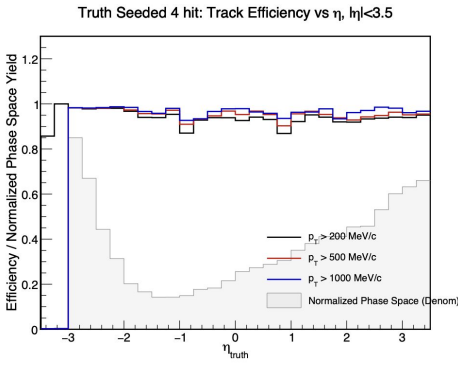
Track Efficiency fluctuates for high pt region (maybe bc of less statistics)
Track Efficiency drops at η = ±1 for all pt

≥ 4 hit track

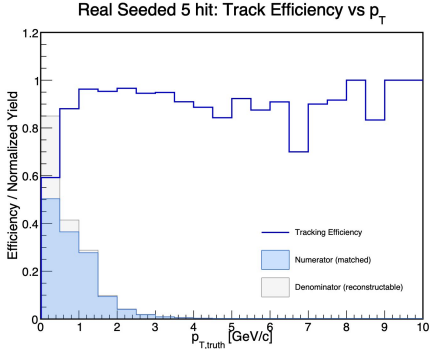


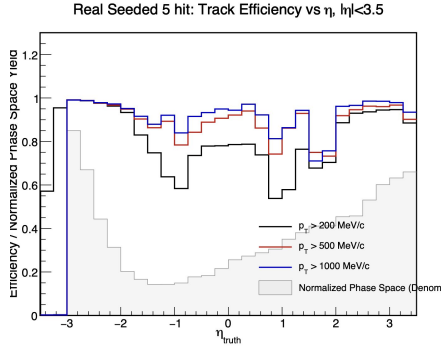


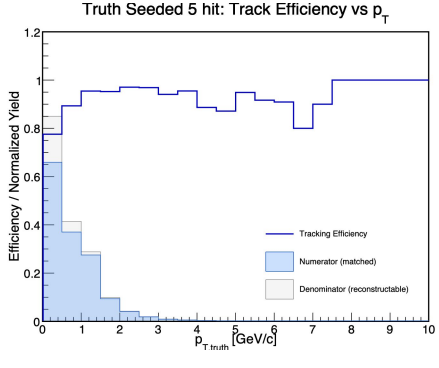


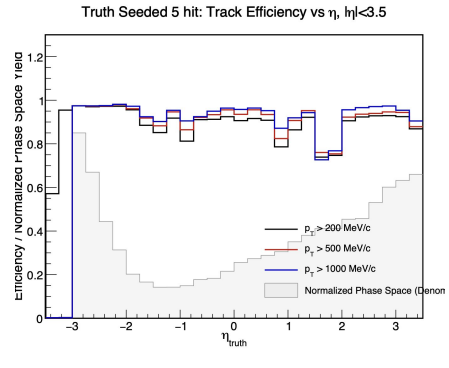


≥ 5 hit track





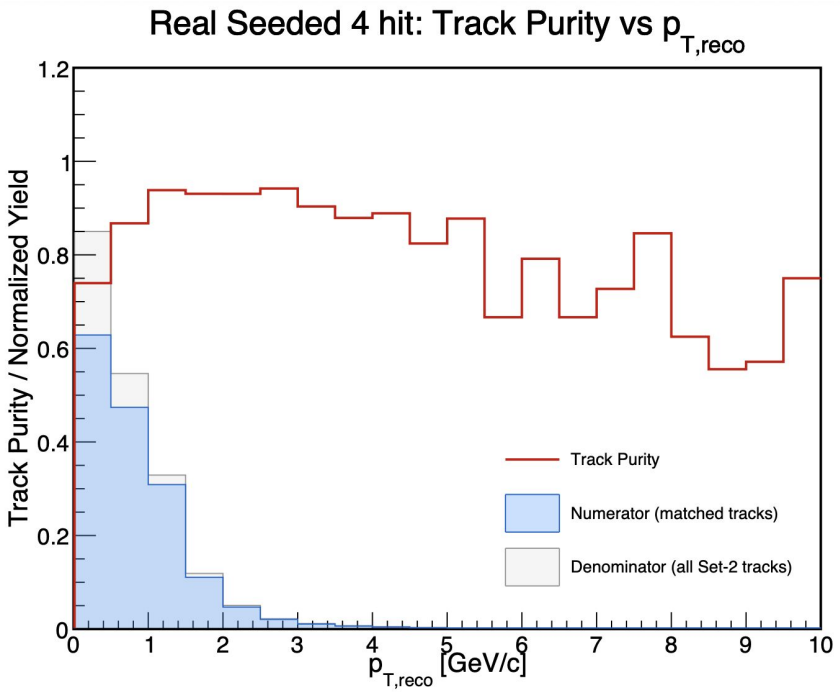


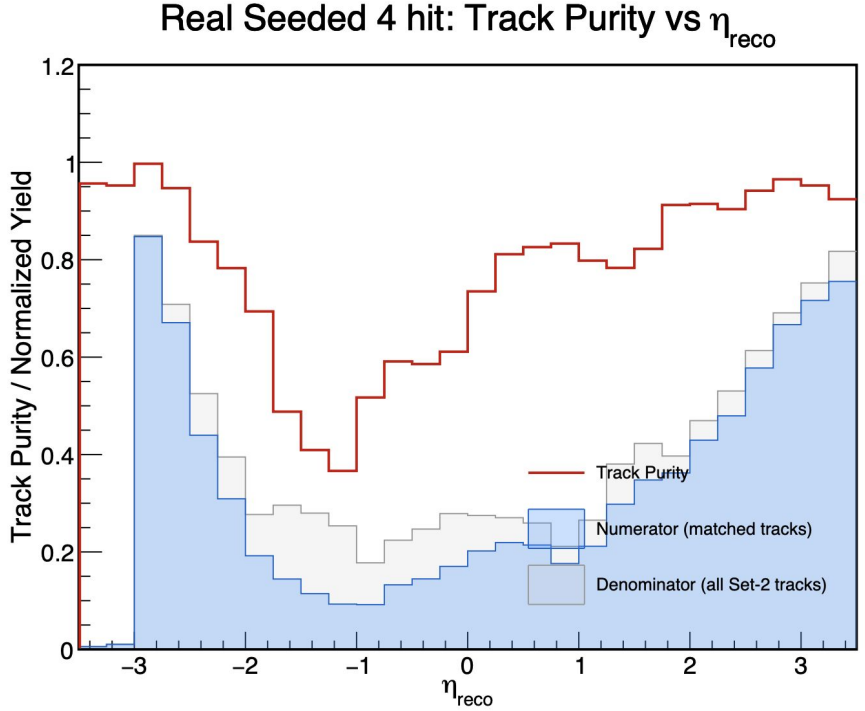


≥ 5 hit track shows similar effects

Truth seeded tracks also shows some fluctuations for all pt and η

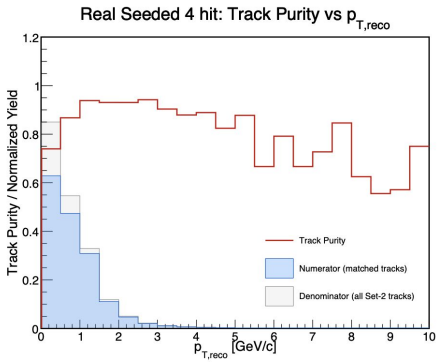
Track purity vs pt and η for matched ≥ 4 hit track

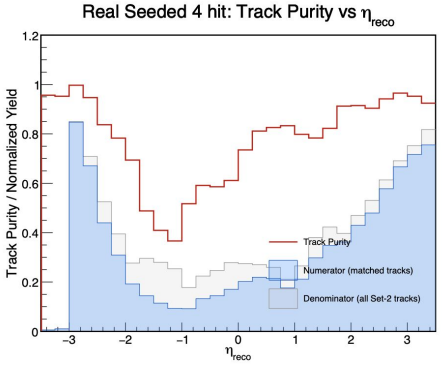


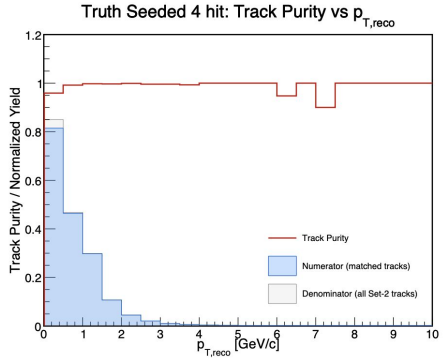


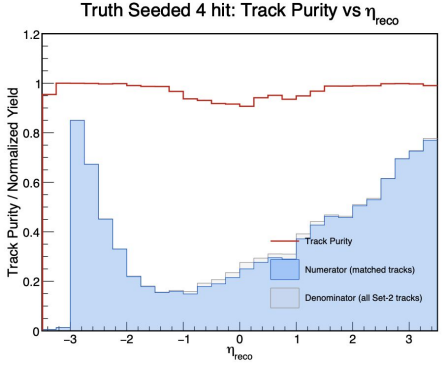
Track purity is below 1 for all pt and decreases at pt increases
Track purity drops in negative eta region

≥ 4 hit track

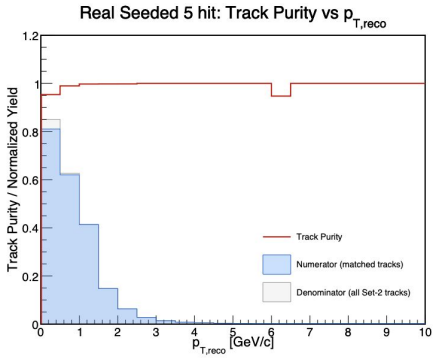


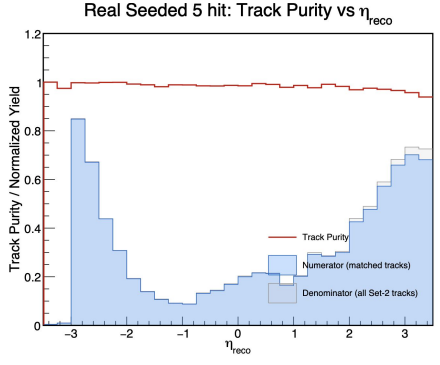


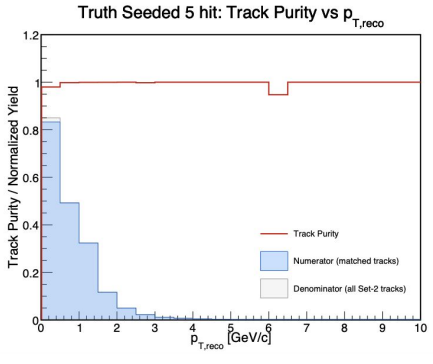


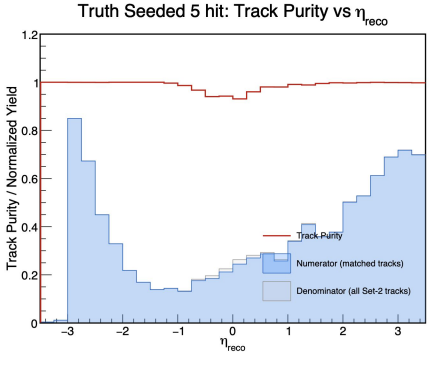


≥ 5 hit track



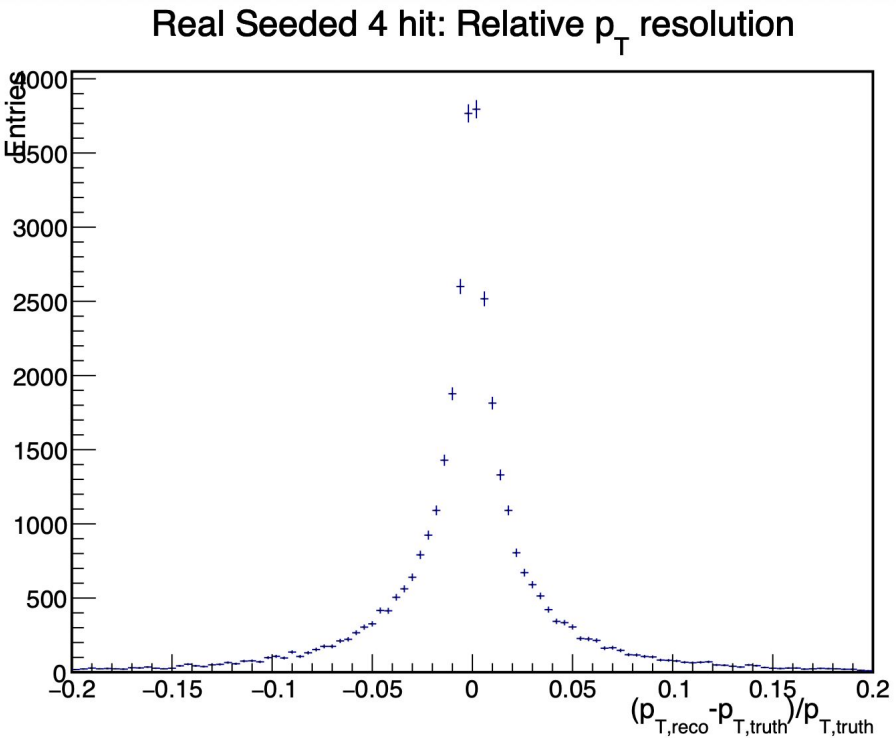


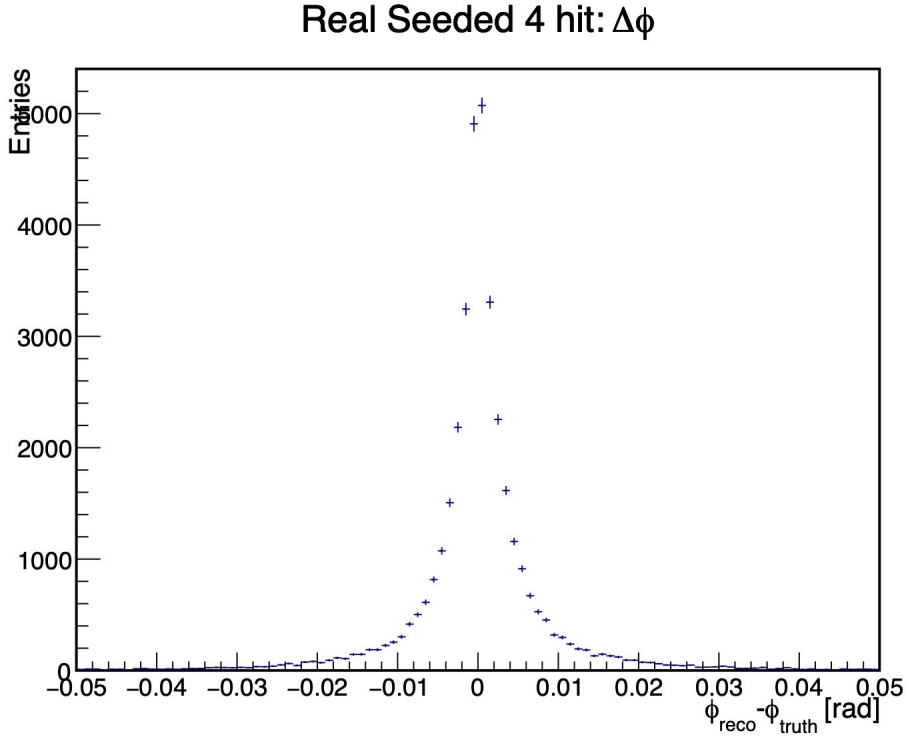




Track purity is ~1 for ≥ 5 hit track

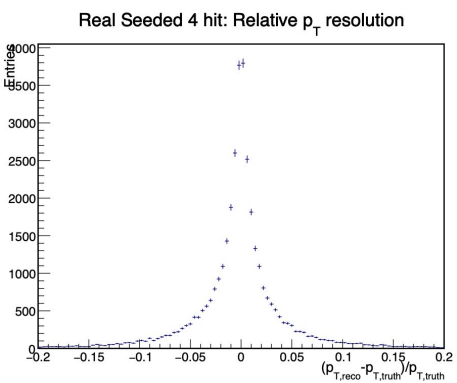
Relative pt resolution and Δϕ for matched ≥ 4 hit track

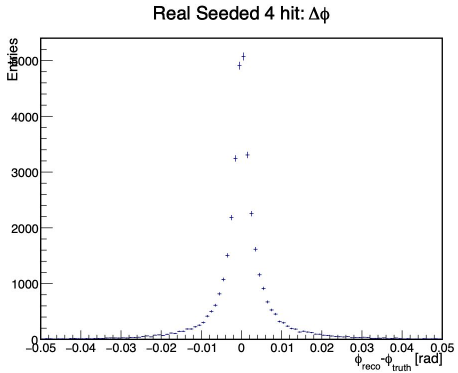


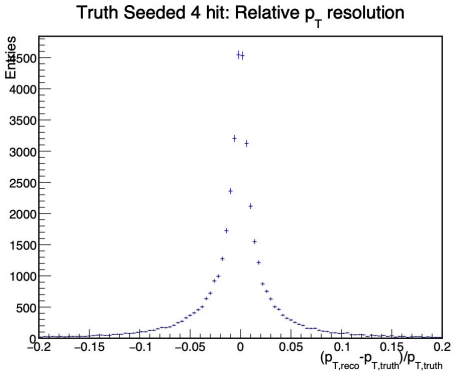


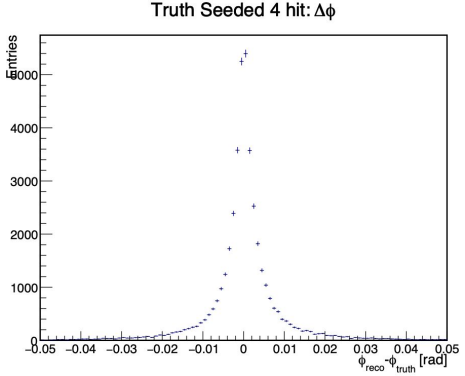
Both relative pt resolution and Δϕ is centered at 0

≥ 4 hit track

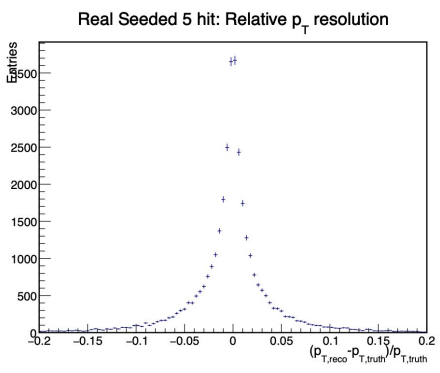


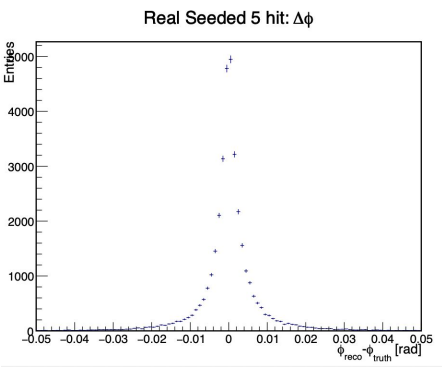


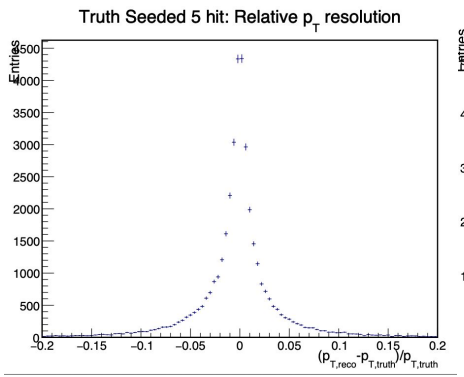


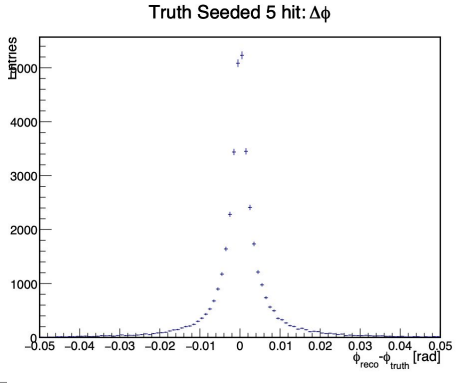


≥ 5 hit track









Similar observation for ≥ 5 hit track, as well as truth seeded tracks

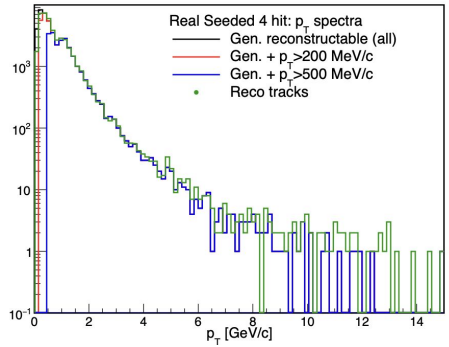
Summary & Plans Ahead

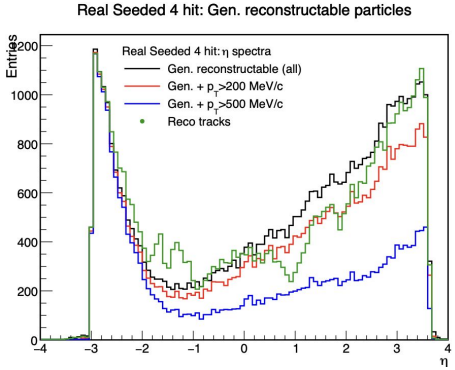
- Study of Double Majority Rule:
  - Hit efficiency, hit purity mostly ~1
  - Track efficiency drops at $\eta = \pm 1$ for all pt for ≥ 4 hit tracks
  - Track purity drops in negative η region for ≥ 4 hit tracks
  - Good resolution and observed

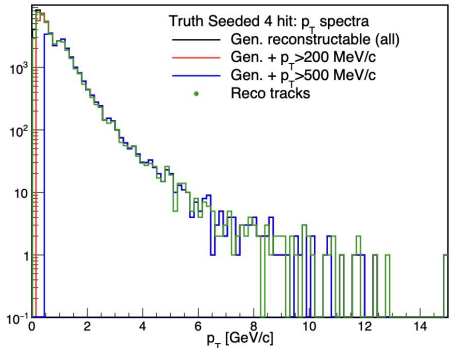
- Implementation of Double Majority Rule as a tracking benchmark that can be used for single-particle, DIS, and DIS+Background simulations

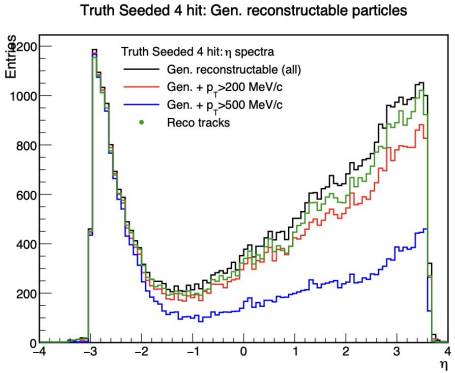
Thank you!

pt, η for ≥ 4 hit track

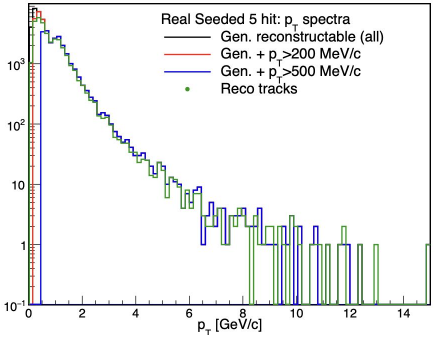


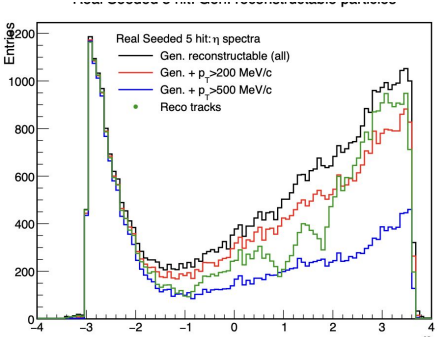


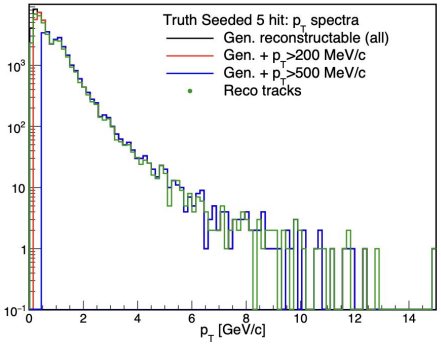


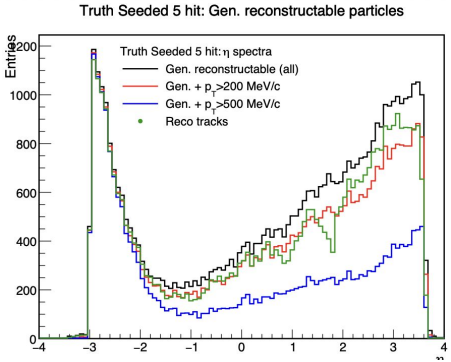


pt, η for ≥ 5 hit track

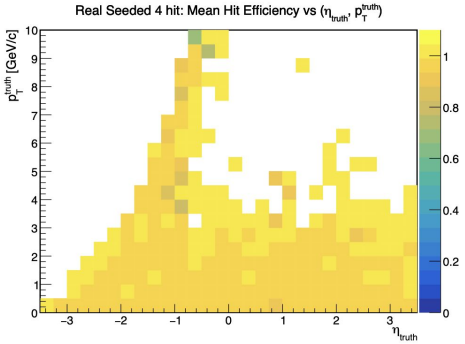


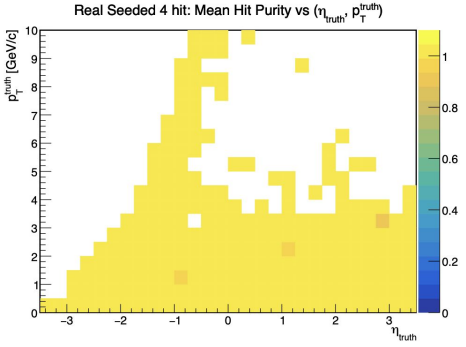


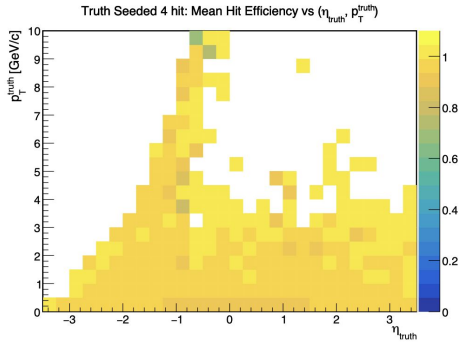


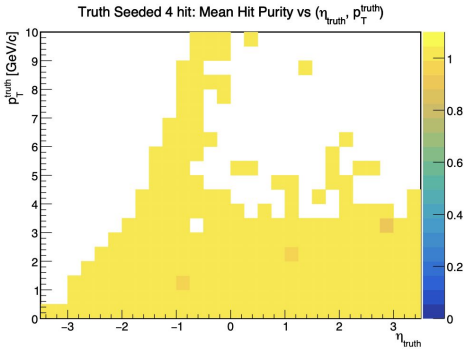


Hit efficiency, purity vs pt, η for matched ≥ 4 hit track

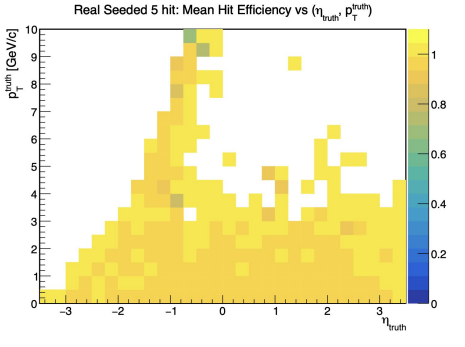


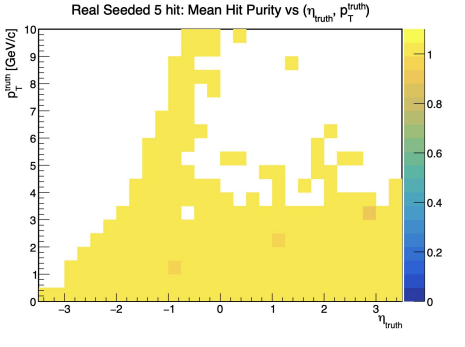


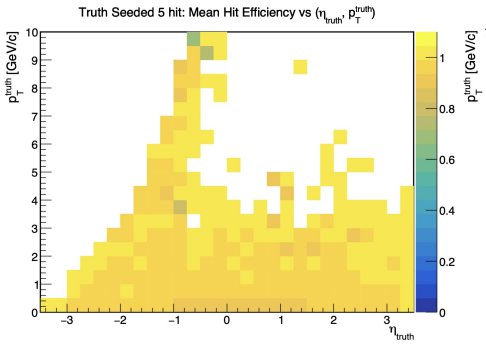


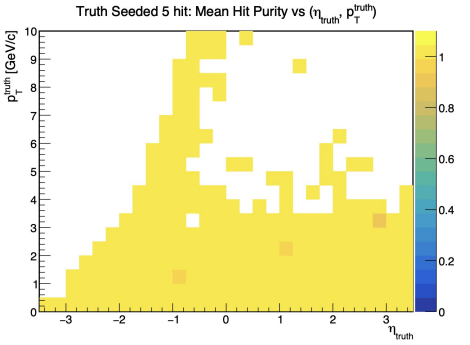


Hit efficiency, purity vs pt, η for matched ≥ 5 hit track

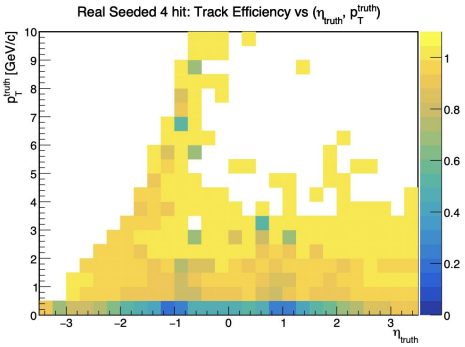


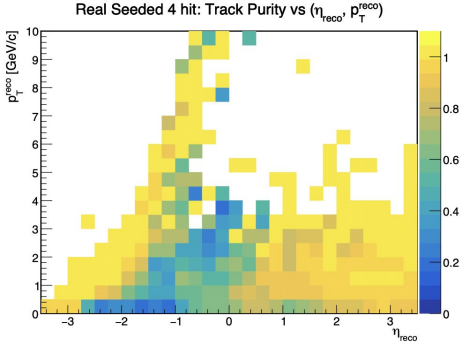


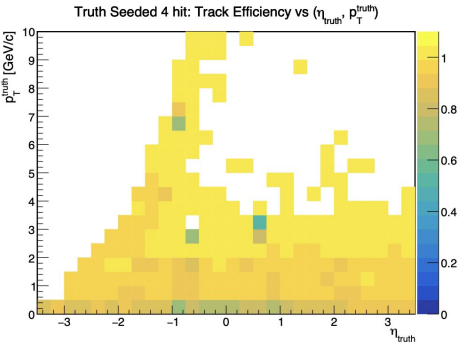


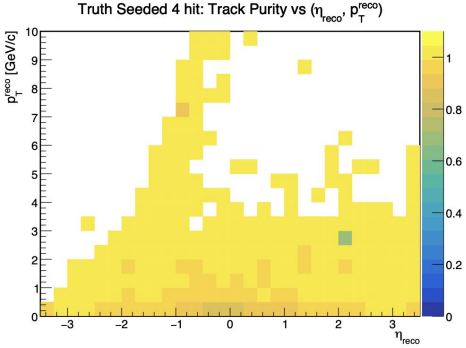


Track efficiency, purity vs pt, η for matched ≥ 4 hit track

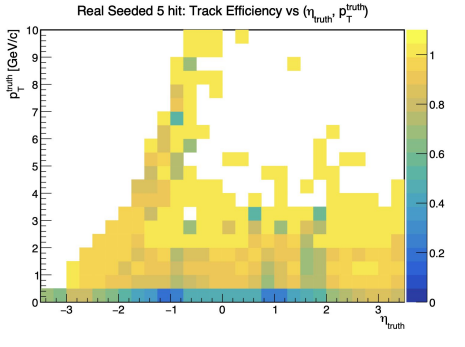


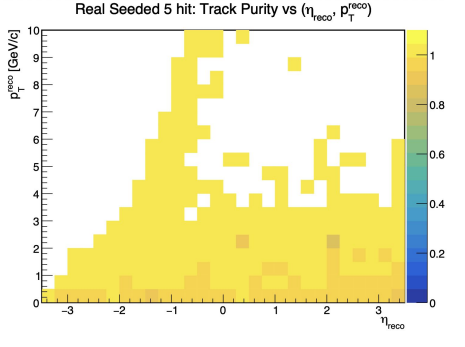


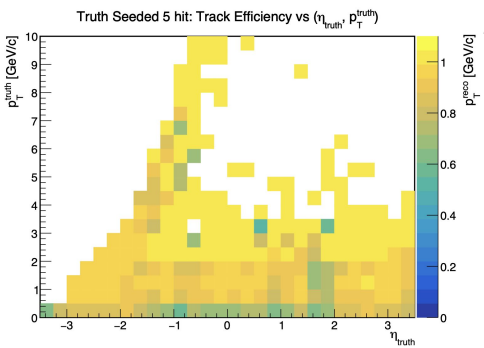


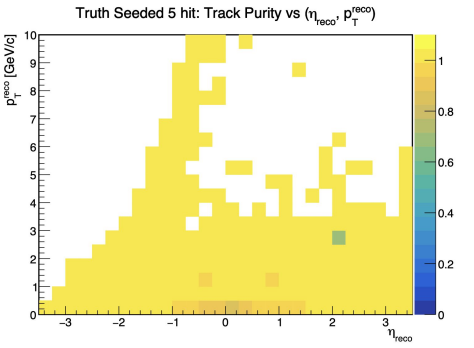


Track efficiency, purity vs pt, η for matched ≥ 5 hit track

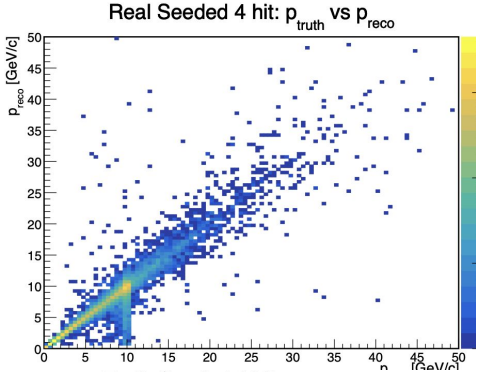


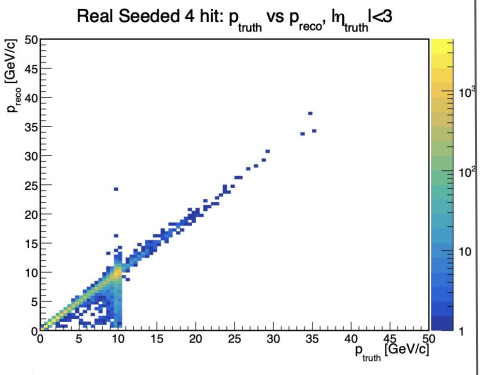


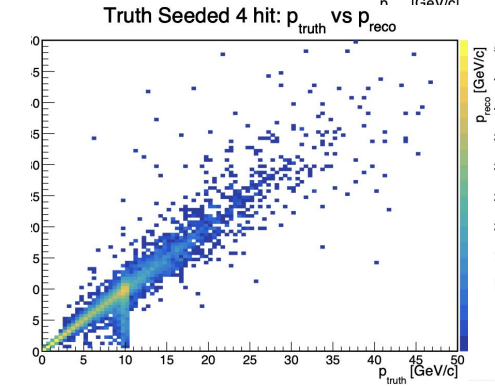


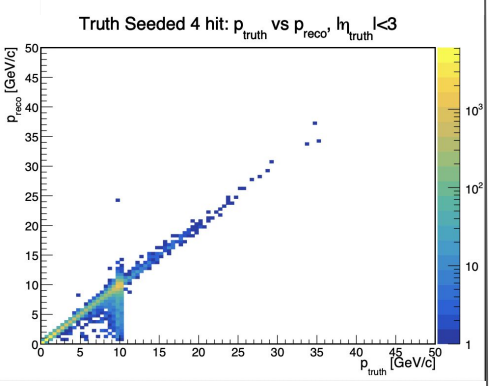


Pt (reco) vs pt (truth) for matched pairs ≥ 4 hit track

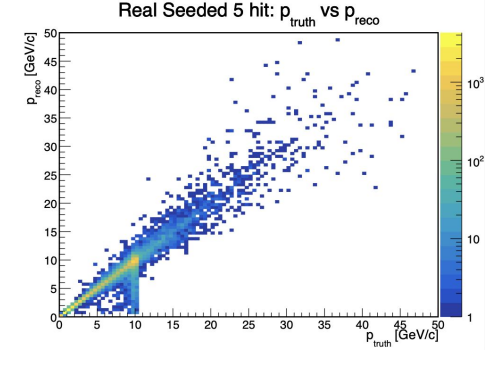


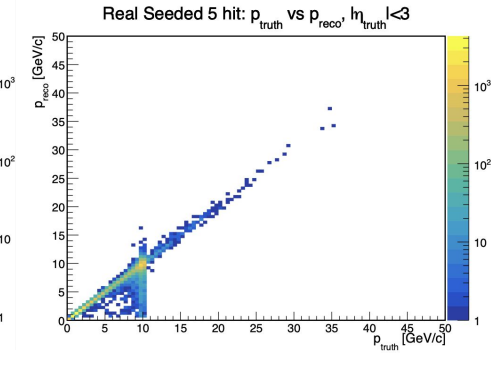


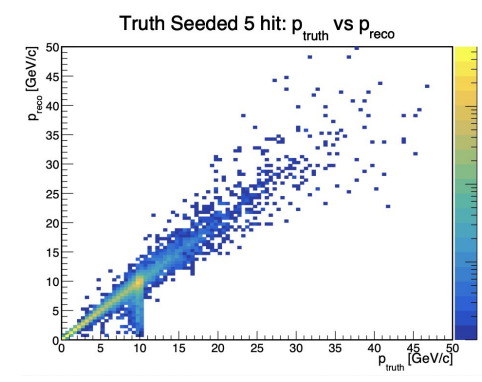


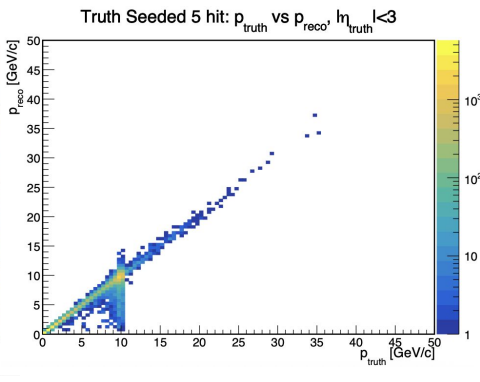


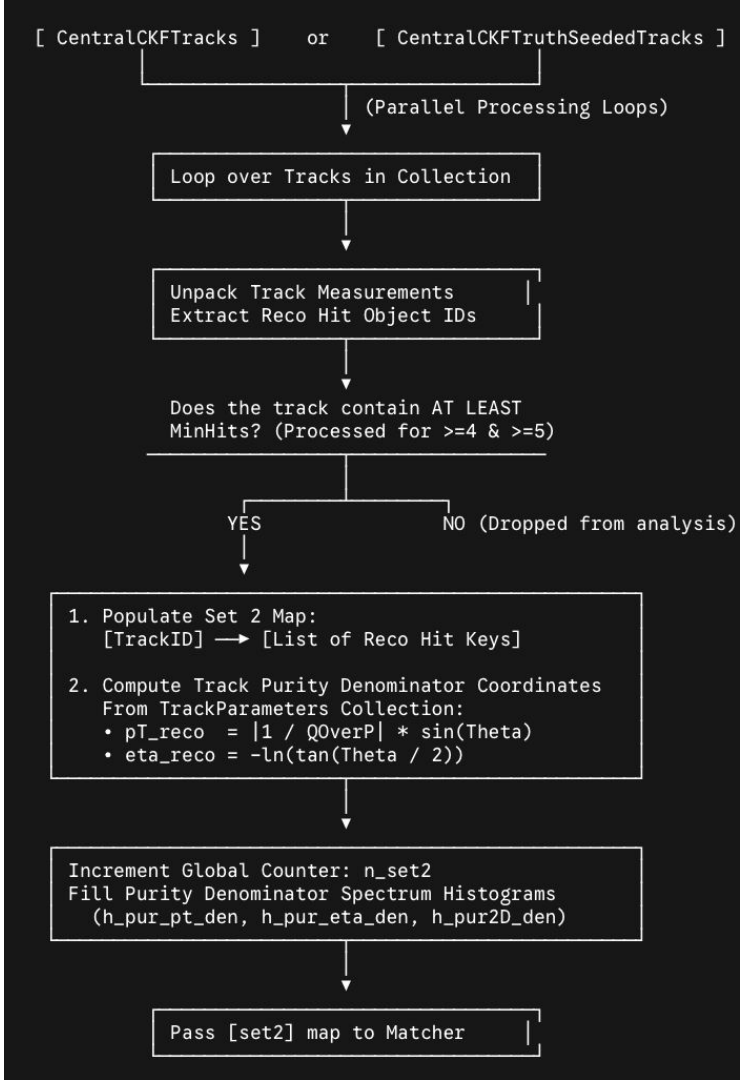
Pt (reco) vs pt (truth) for matched pairs for ≥ 5 hit track










[CentralCKFTracks] or [CentralCKFTruthSeededTracks]
(Parallel Processing Loops)
Loop over Tracks in Collection
Unpack Track Measurements
Extract Reco Hit Object IDs
Does the track contain AT LEAST
MinHits? (Processed for >=4 & >=5)
YES
NO (Dropped from analysis)
1. Populate Set 2 Map:
[TrackID] ⟶ [List of Reco Hit Keys]
2. Compute Track Purity Denominator Coordinates
From TrackParameters Collection:
• pT_reco = |1 / QOverP| * sin(Theta)
• eta_reco = -ln(tan(Theta / 2))
Increment Global Counter: n_set2
Fill Purity Denominator Spectrum Histograms
(h_pur_pt_den, h_pur_eta_den, h_pur2D_den)
Pass [set2] map to Matcher