ePIC Barrel HCAL energy calibration using Machine Learning

Churamani Paudel, David Ruth

(New Mexico State University)

09/18/2026

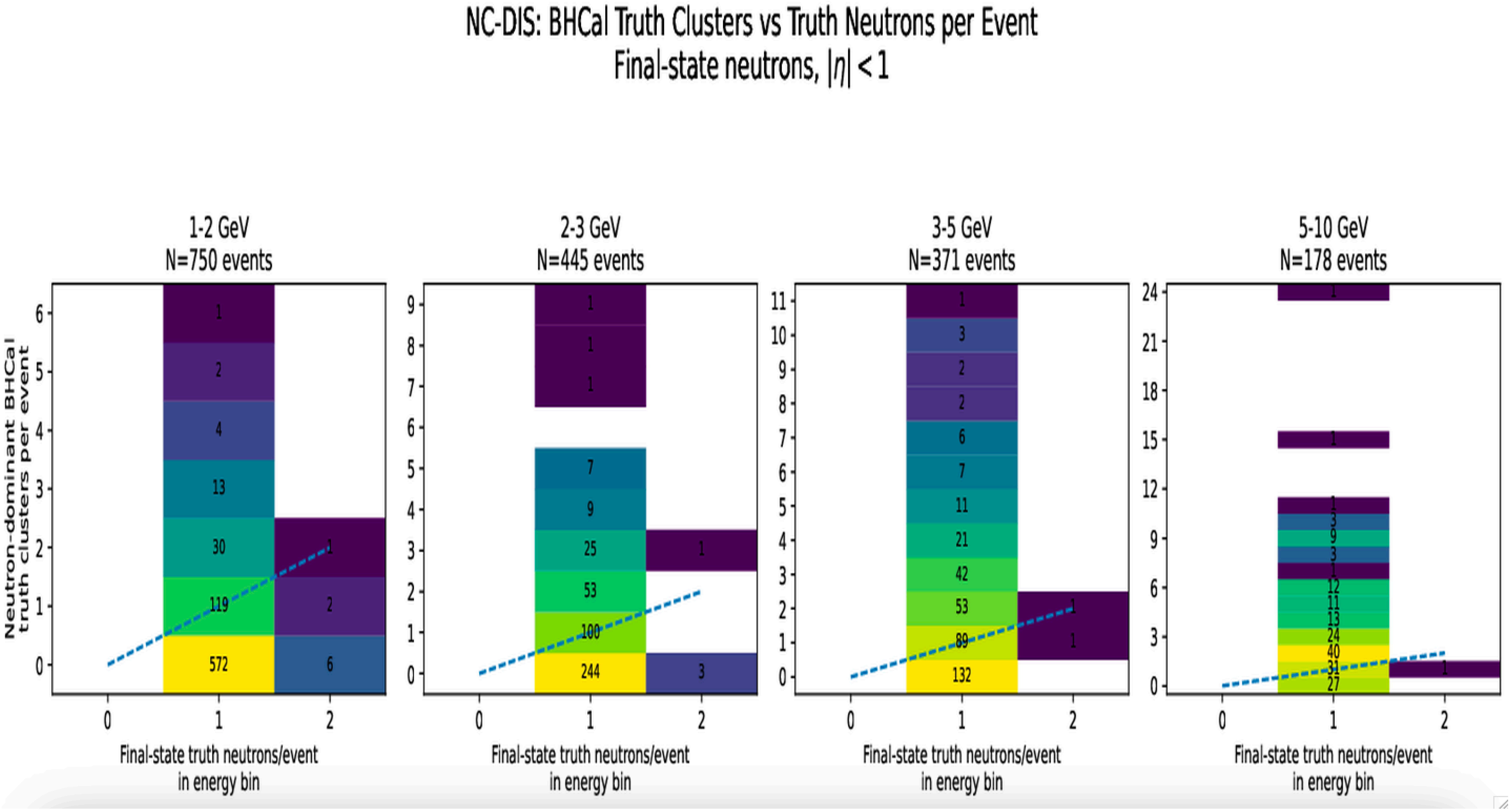
Truth Clusters vs Reconstructed Clusters

- Truth clusters use Monte Carlo information to identify the simulated particle responsible for the energy deposits.
- Reconstructed clusters use only detector hits and the geometrical clustering algorithm.
- Truth clustering provides a simulation-based reference for the shower origin.
- Reconstructed clustering can divide one neutron shower into several separate clusters.
- One neutron does not necessarily correspond to exactly one truth cluster because a hadronic shower can contain several contributing particle tracks.

Goal: compare reconstructed cluster multiplicity with the truth reference to identify reconstruction-driven fragmentation.

Follow-Up Truth Cluster vs no. of neutrons/event

- The x-axis is the number of final-state truth neutrons in each energy bin.
- The y-axis is the number of neutron-dominant BHCal truth clusters in the same event.
- Most NC-DIS events contain one selected neutron in a given energy bin.
- One neutron does not always correspond to exactly one truth cluster.
- The truth-cluster multiplicity develops a broader tail as neutron energy increases.



BHCal Cell Spacing and the Choice of eta-phi Clustering Range

$$\Delta\phi_{\text{cell}} = \frac{2\pi}{N_\phi} = \frac{2\pi}{320} = 0.0196 \text{ rad}$$

$$3\,\Delta\phi_{\text{cell}} = 3\left(\frac{2\pi}{320}\right) = 0.0589 \simeq 0.060$$

$$\Delta\phi_{\max} = 0.021,\, 0.040,\, 0.060,\, 0.080,\, 0.100$$

$$\eta = \operatorname{asinh}\left(\frac{z}{r_T}\right), \qquad r_T = \sqrt{x^2 + y^2}$$

$$\Delta\eta_{\text{cell}} \simeq 0.091 - 0.092$$

$$\Delta\eta_{\max} = 0.095,\, 0.185$$

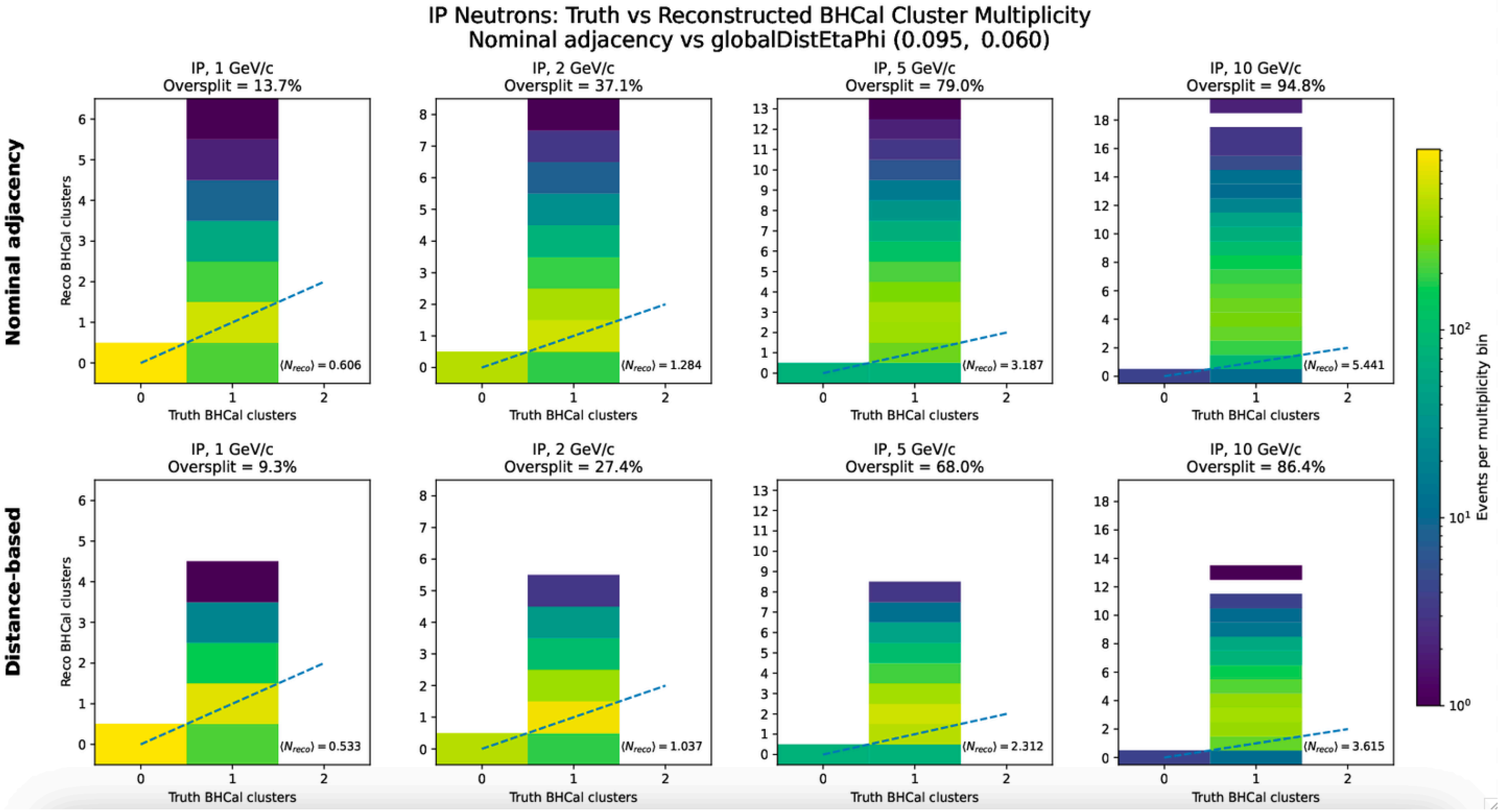
$$0.095 \approx 1\,\Delta\eta_{\text{cell}}$$

Transverse distance from the beam axis

I chose the eta and phi scan values to follow the natural BHCal cell spacing and to test how far the clustering should reach before it becomes too broader. In eta, I compared 0.095 and 0.185, which correspond approximately to one and two BHCal eta-cell spacings. Extending the eta reach gave only a modest improvement, so I kept the more conservative value of 0.095. In phi, the BHCal cell spacing is about 2 pi divided by 320, which is about 0.0196 radians. I therefore scanned 0.021, 0.040, 0.060, 0.080, and 0.100, corresponding roughly to one through five phi-cell spacings. Increasing the phi reach produced a much larger reduction in reconstructed multiplicity and oversplitting than increasing the eta reach. I therefore chose 0.060, about three phi-cell spacings, as a conservative value that recovered a large fraction of the fragmented shower while avoiding the more broader four- and five-cell connections. The final frozen choice was eta equals 0.095 and phi equals 0.060, corresponding approximately to **one eta-cell spacing** and three phi-cell spacings.

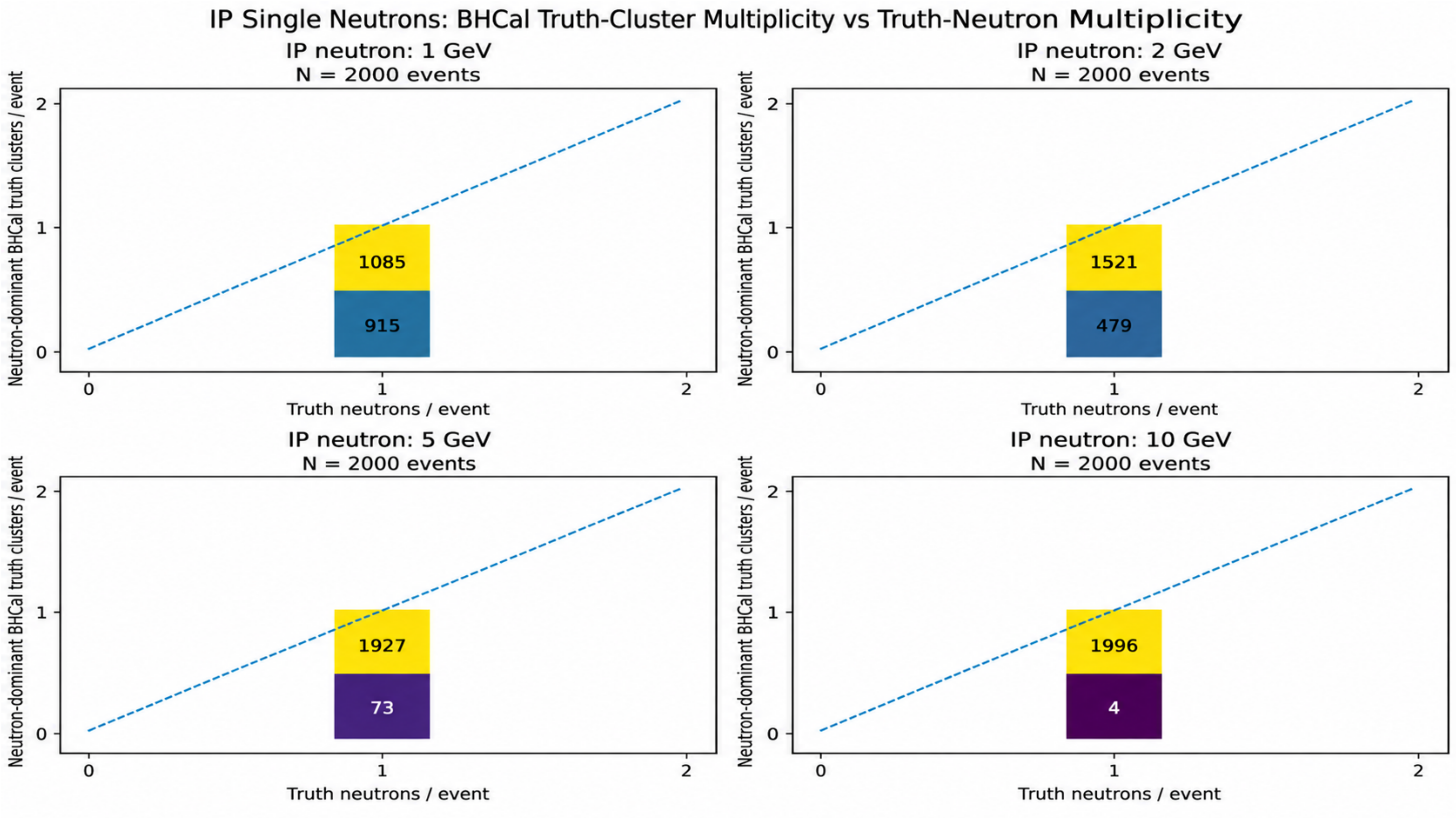
IP Neutrons: Truth vs Reconstructed Multiplicity

- Each IP sample contains a single generated neutron propagated through the full detector.
- The top row uses nominal 4-neighbor adjacency; the bottom row uses frozen globalDistEtaPhi.
- Oversplitting is defined by reconstructed multiplicity greater than truth-cluster multiplicity.
- Nominal adjacency oversplitting increases strongly with neutron momentum.
- globalDistEtaPhi reduces oversplitting at every tested momentum without IP retuning.



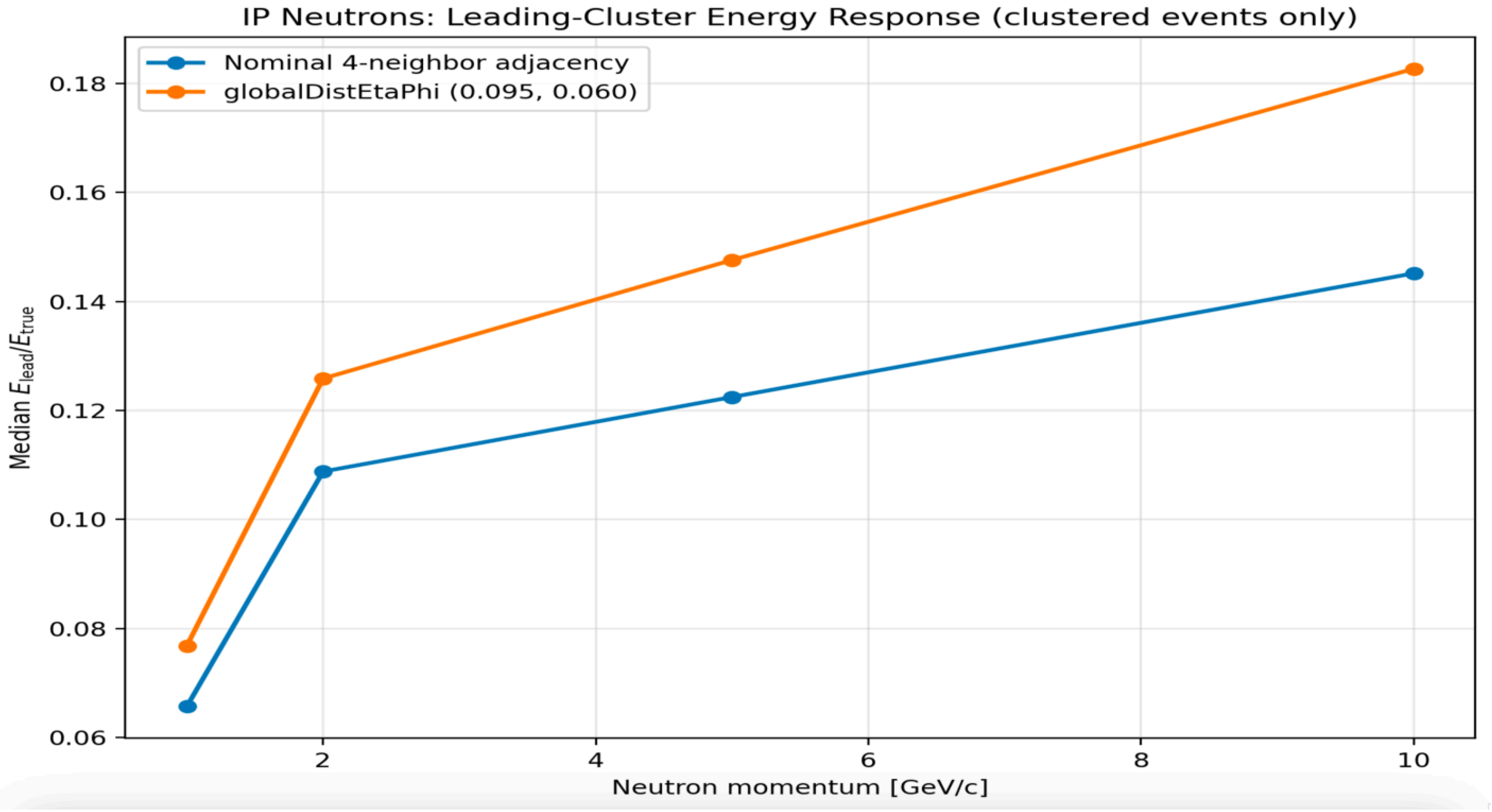
IP Neutrons: Truth Clusters vs Truth Neutrons

- There is exactly one generated truth neutron per simulated IP event.
- At 1 GeV, many events have no neutron-dominant BHCal truth cluster.
- The fraction with one truth cluster increases strongly with neutron energy.
- By 5–10 GeV, almost every event contains one neutron-dominant truth cluster.
- Large reconstructed multiplicities at high energy therefore cannot be explained by large truth multiplicity alone.



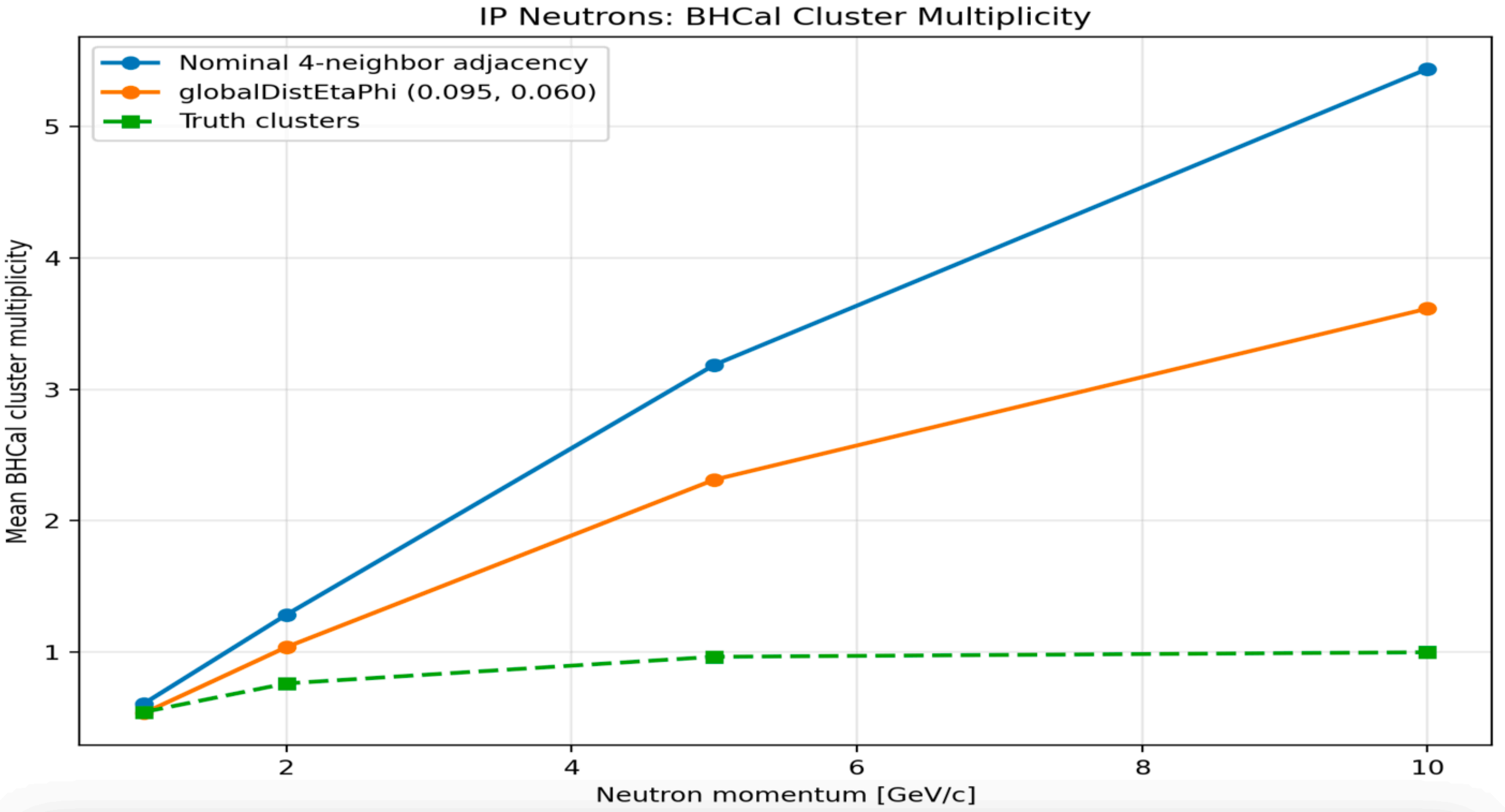
IP neutrons, Leading Clusters Energy Response

- The response is defined as the median leading-cluster energy divided by the true neutron energy.
- This figure uses only events with at least one reconstructed BHCal cluster.
- globalDistEtaPhi gives a larger leading-cluster response than nominal adjacency.
- The difference becomes larger as neutron momentum increases.
- This indicates that the new clustering reconnects more shower energy into the leading cluster.



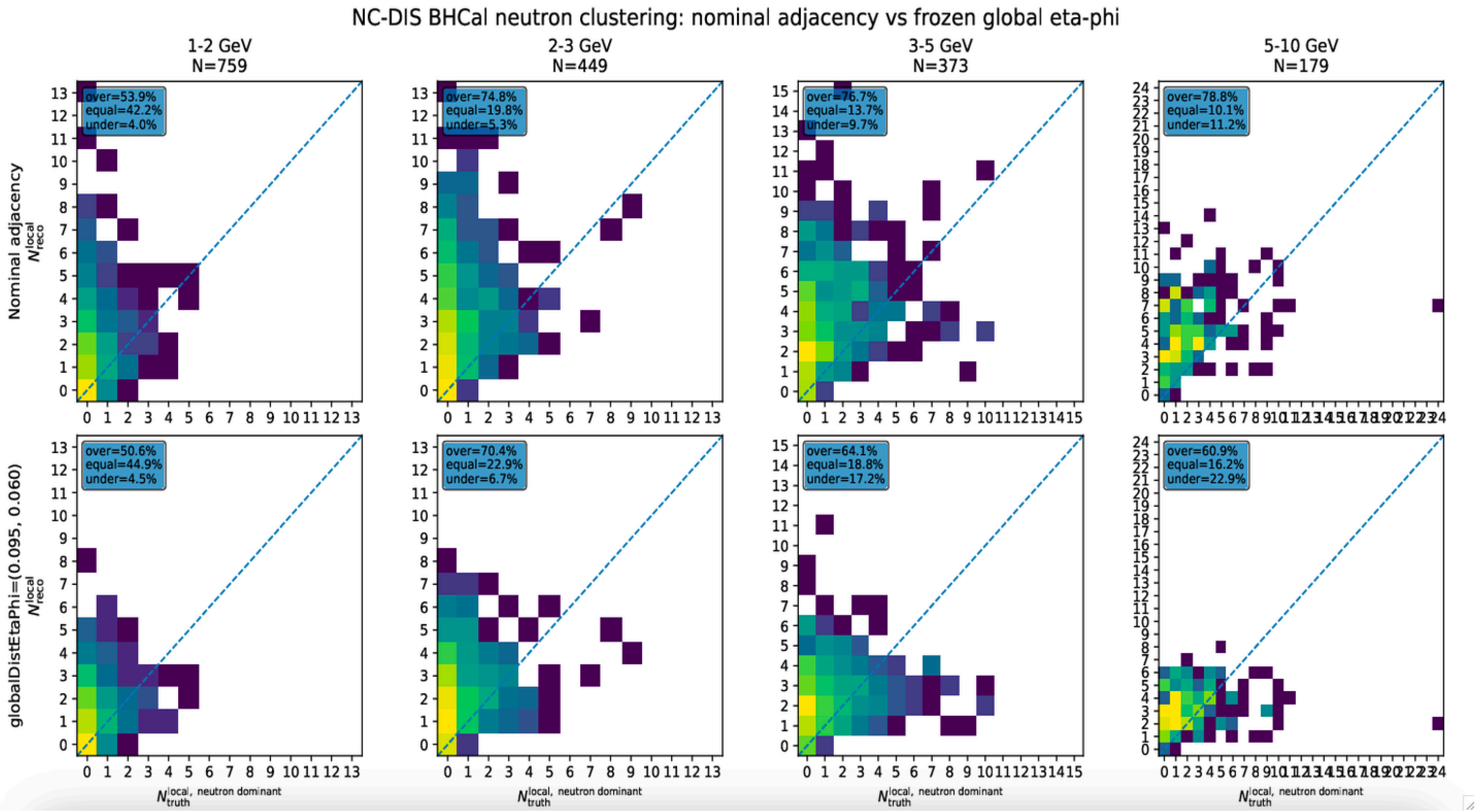
IP Neutrons: BHCal Cluster Multiplicity

- The mean truth-cluster multiplicity approaches one as neutron momentum increases.
- Nominal reconstructed multiplicity grows much faster than the truth multiplicity.
- The difference is especially large at 5 and 10 GeV.
- globalDistEtaPhi substantially reduces the reconstructed cluster multiplicity.
- A remaining gap at high energy shows that fragmentation is reduced but not completely removed.



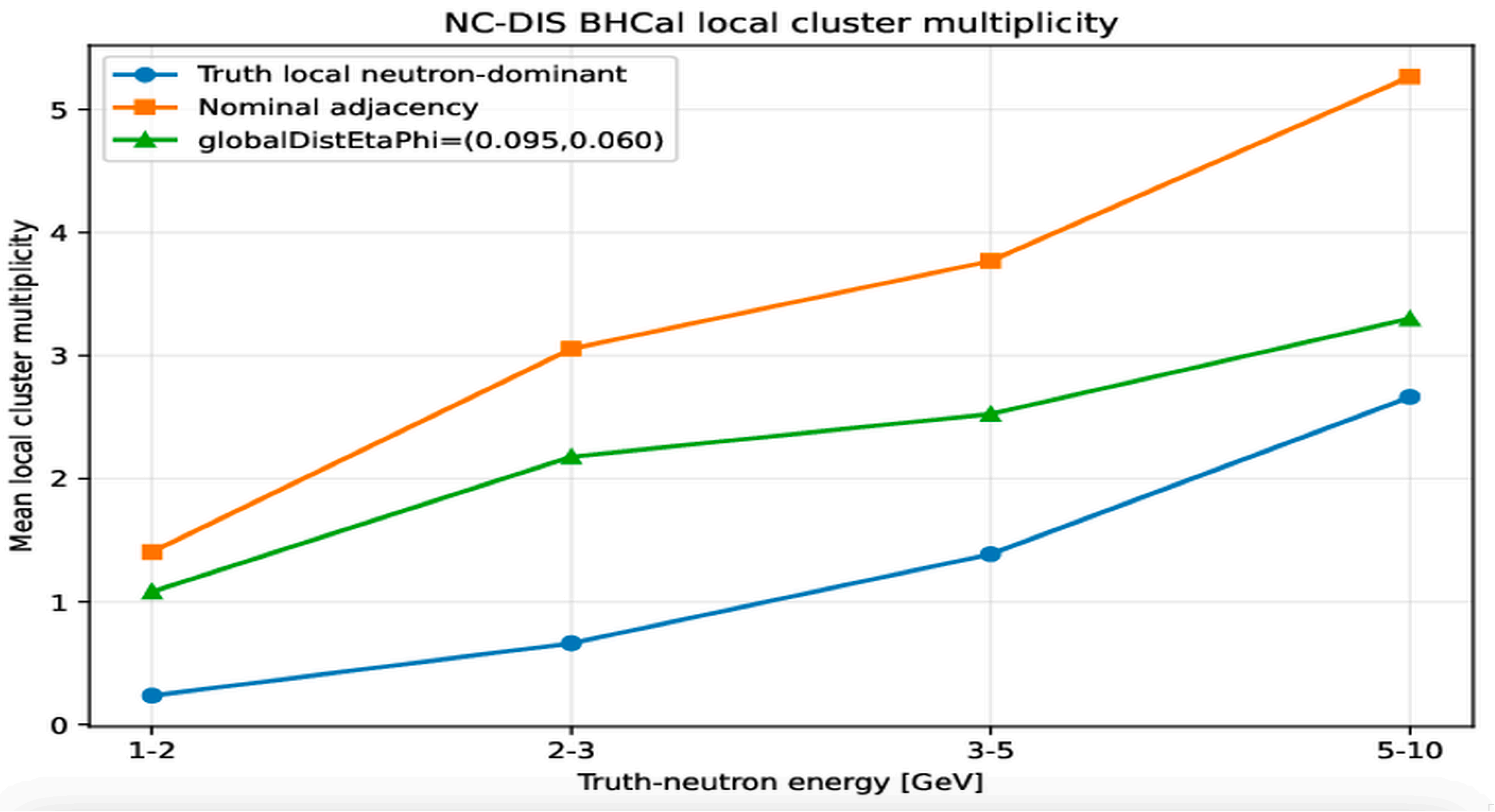
DIS Neutrons: Adjacency vs globalDistEtaPhi

- The comparison uses the same fixed truth-neutron population for both clustering methods.
- The x-axis is local neutron-dominant truth-cluster multiplicity.
- The y-axis is local reconstructed BHCal cluster multiplicity around the neutron direction.
- Nominal adjacency produces substantial excess reconstructed multiplicity in all energy bins.
- Frozen globalDistEtaPhi shifts the distributions closer to the truth reference, especially above 3 GeV.



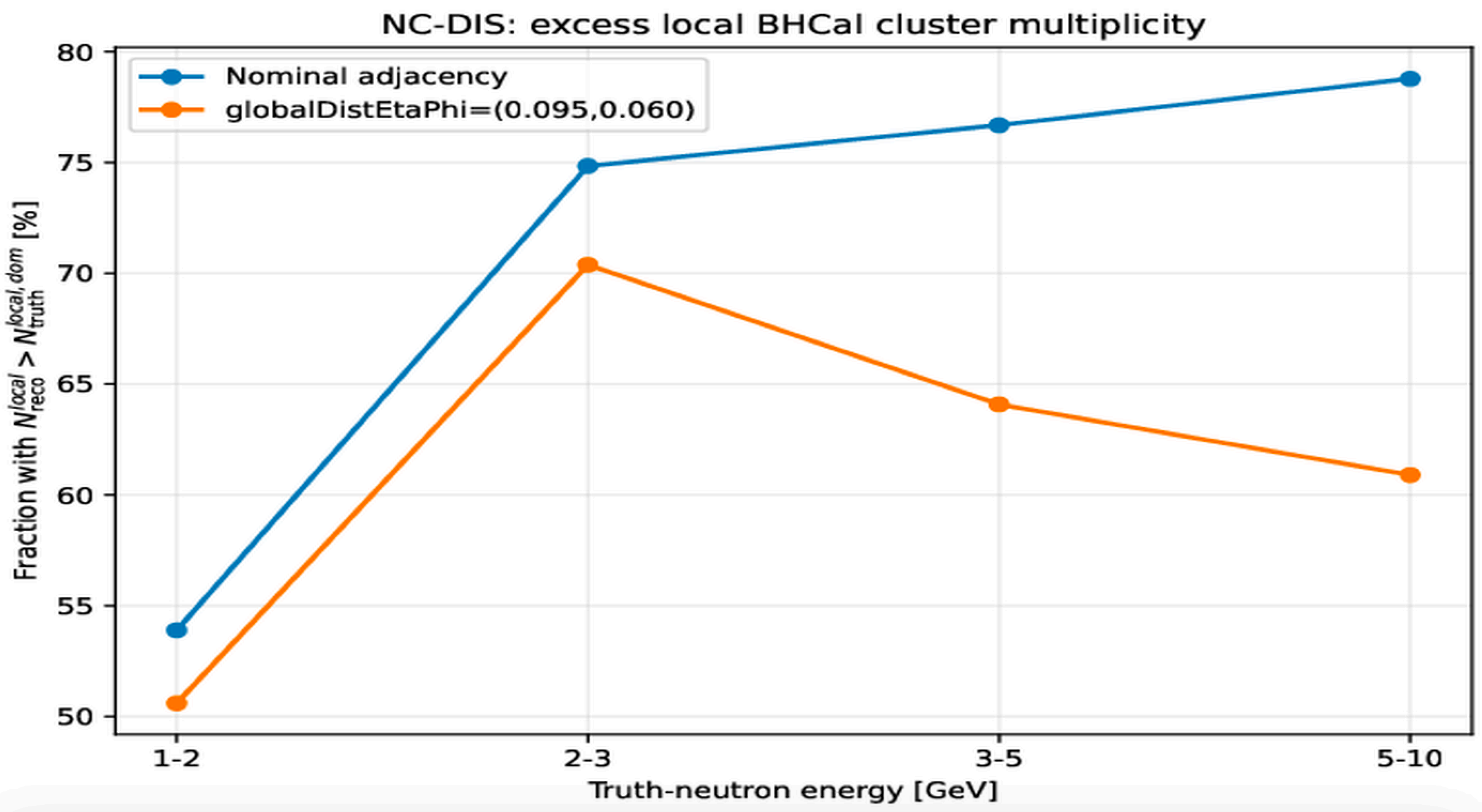
DIS Neutrons: Mean Local Intensity

- Mean truth-cluster multiplicity increases with neutron energy.
- Nominal adjacency remains well above the truth reference in every energy bin.
- globalDistEtaPhi lowers the mean reconstructed multiplicity in every bin.
- The reduction becomes increasingly important at higher neutron energy.
- The remaining excess can include both residual fragmentation and nearby DIS activity.



DIS Neutrons: Excess Local Multiplicity

- This plot shows the fraction of neutrons with reconstructed multiplicity above the truth multiplicity.
- Nominal adjacency gives a large excess-multiplicity fraction across the full energy range.
- globalDistEtaPhi lowers that fraction in every energy bin.
- The largest reduction appears in the 3–5 and 5–10 GeV regions.
- The criterion was frozen from the direct-neutron study, so this is a validation rather than a DIS optimization.



Summary

- Truth clustering is MC-based and fundamentally different from detector-level Island Clustering.
- Nominal adjacency produces increasing reconstructed fragmentation with neutron energy.
- globalDistEtaPhi clustering reduces cluster multiplicity in direct, IP, and DIS environments.
- The improvement is highly pronounced at intermediate (3-5 GeV) and high neutron energies (> 5 GeV) as compared to low energy (1-2 GeV)

Next, Clustering to ML Energy Calibration (Single Neutrons)

- Produce matched single-neutron samples using nominal **4-neighbor adjacency** and frozen `globalDistEtaPhi = (0.095, 0.060)`.
- Train the same ML energy-calibration model on both samples using identical features, architecture, and train/validation splits.
- Compare neutron energy response, bias, and resolution versus neutron momentum for the two clustering algorithms.
- Check whether reduced cluster fragmentation translates into improved ML energy resolution, especially at higher neutron energies.

Next, Towards DIS ML Validation

- Freeze the clustering parameters from the single-neutron study; no DIS retuning.
- Reconstruct the same DIS events with nominal adjacency and global η–ϕ clustering.
- Produce matched ML training/validation samples from both reconstructions.
- Compare ML resolution in the 1–2, 2–3, 3–5, and 5–10 GeV neutron-energy regions.
- Correlate ML errors with local cluster multiplicity and DIS event activity to identify the remaining resolution limitations