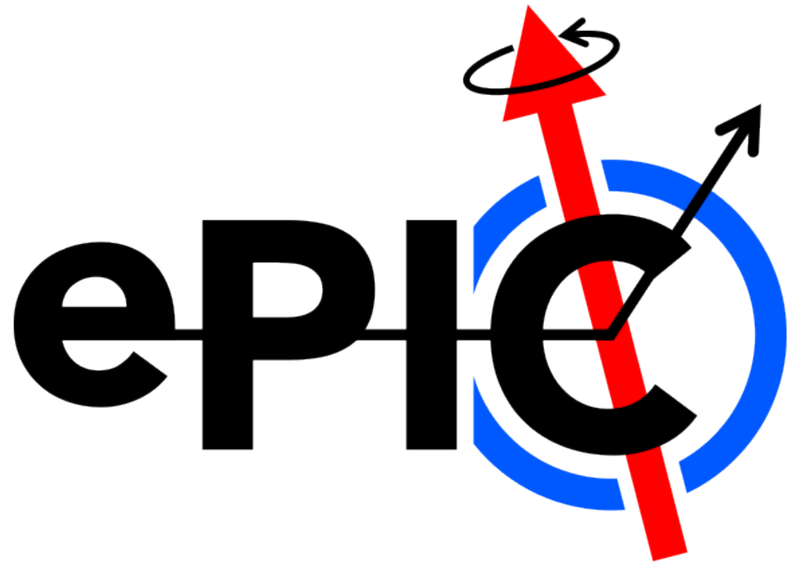
ePIC Barrel HCAL energy calibration using Machine Learning

Churamani Paudel, David Ruth

(New Mexico State University)

09/11/2026

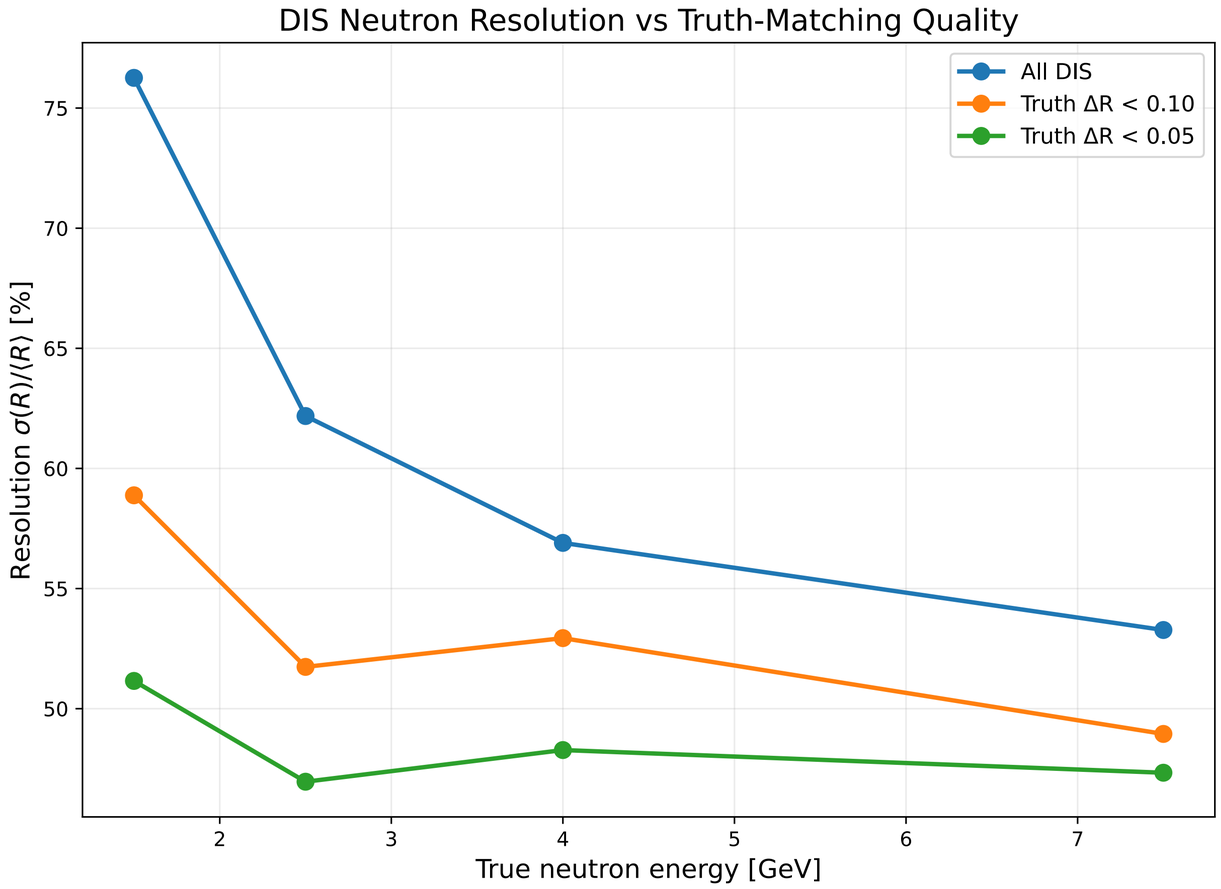


Understanding DIS neutron environment

- Neutron energy resolution in realistic DIS events remains much worse than expected.
- ML calibration improves the response, but large event-by-event fluctuations remain.
- In DIS, a neutron shower is reconstructed together with many other particles and calorimeter clusters.
- We want to identify whether poor association, shower splitting, or surrounding event activity limits the resolution.
- Goal: find reconstructed observables that can ultimately be used in real data.

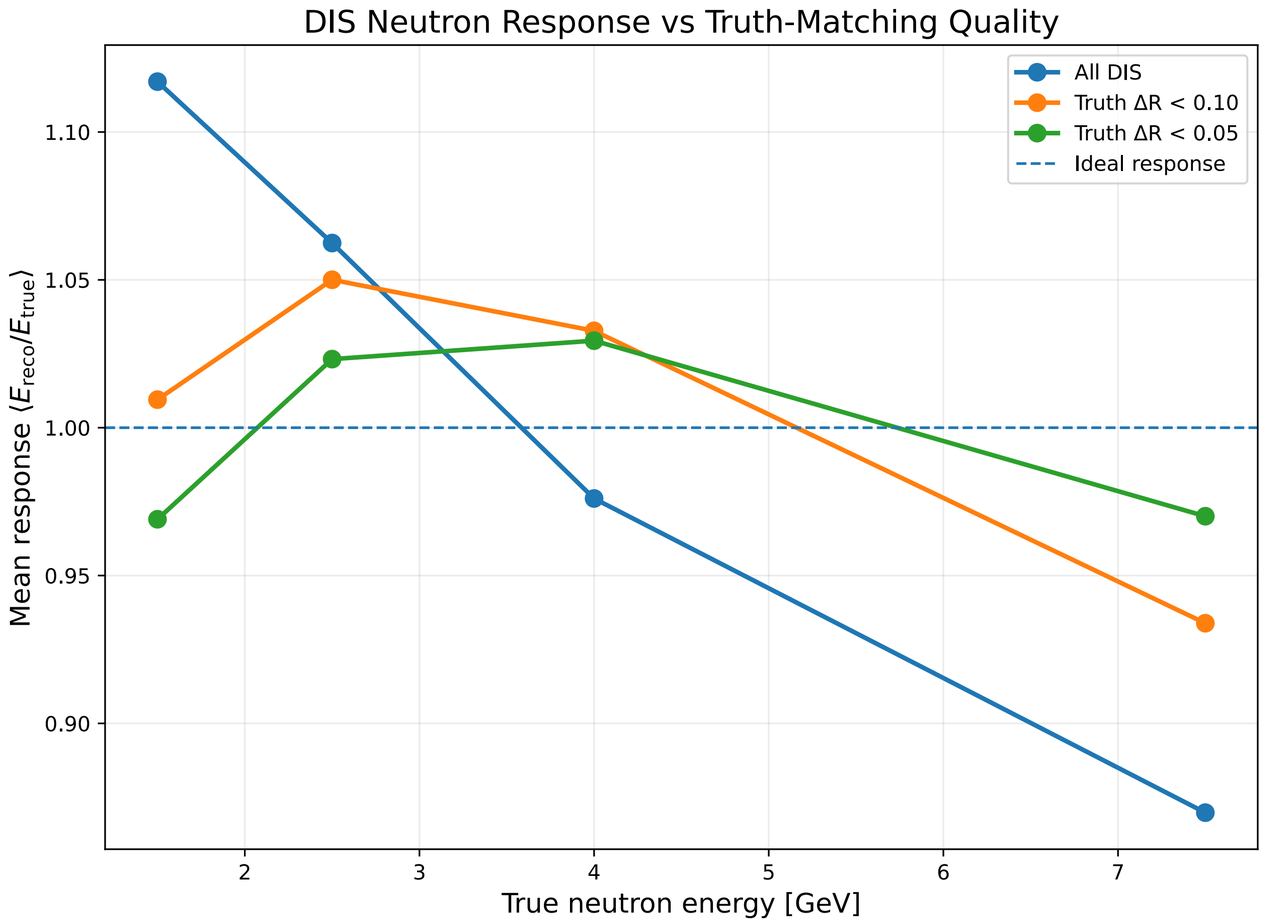
Truth Matching Strongly Affects Neutron Energy Resolution

- Tighter truth matching selects better reconstructed neutron candidates.
- Resolution improves across the full neutron-energy range.
- Mean resolution improves from about 62% to 53% for $\Delta R_{truth} < 0.10$.
- For $\Delta R_{truth} < 0.05$, the mean resolution reaches about 48%.
- Truth matching is a simulation-only diagnostic, not a real-data cut.



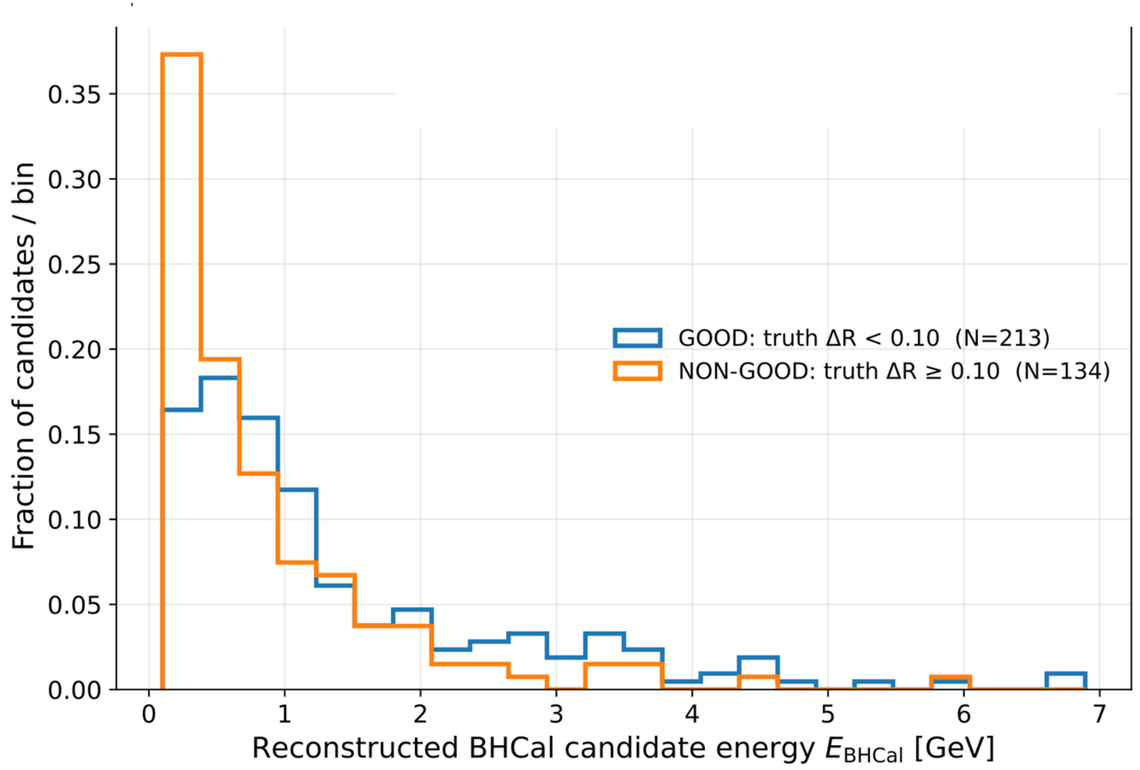
Truth Matching Also Improves the Neutron Energy Response

- Better truth matching also improves the reconstructed energy response.
- Tighter ΔR selections move Epred/Etrue closer to unity.
- Both response bias and resolution improve simultaneously.
- Candidate association is therefore important for energy calibration.
- Goal: identify a real-data-usable proxy for truth-match quality.



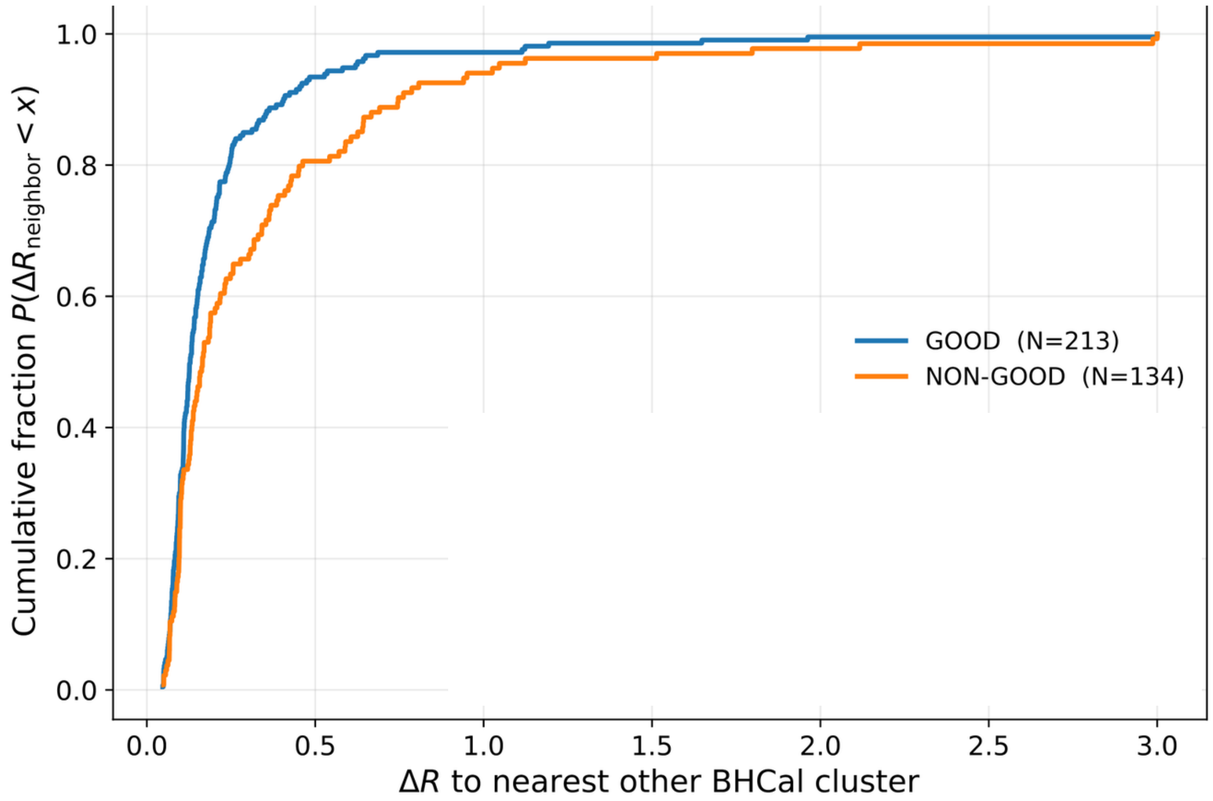
Energy Alone Only Partially Separates Candidate Quality

- Independent sample
- GOOD and NON-GOOD energy distributions overlap strongly.
- Candidate energy contains some discrimination, but is not sufficient.
- A direct energy cut would also bias an energy-calibration analysis.



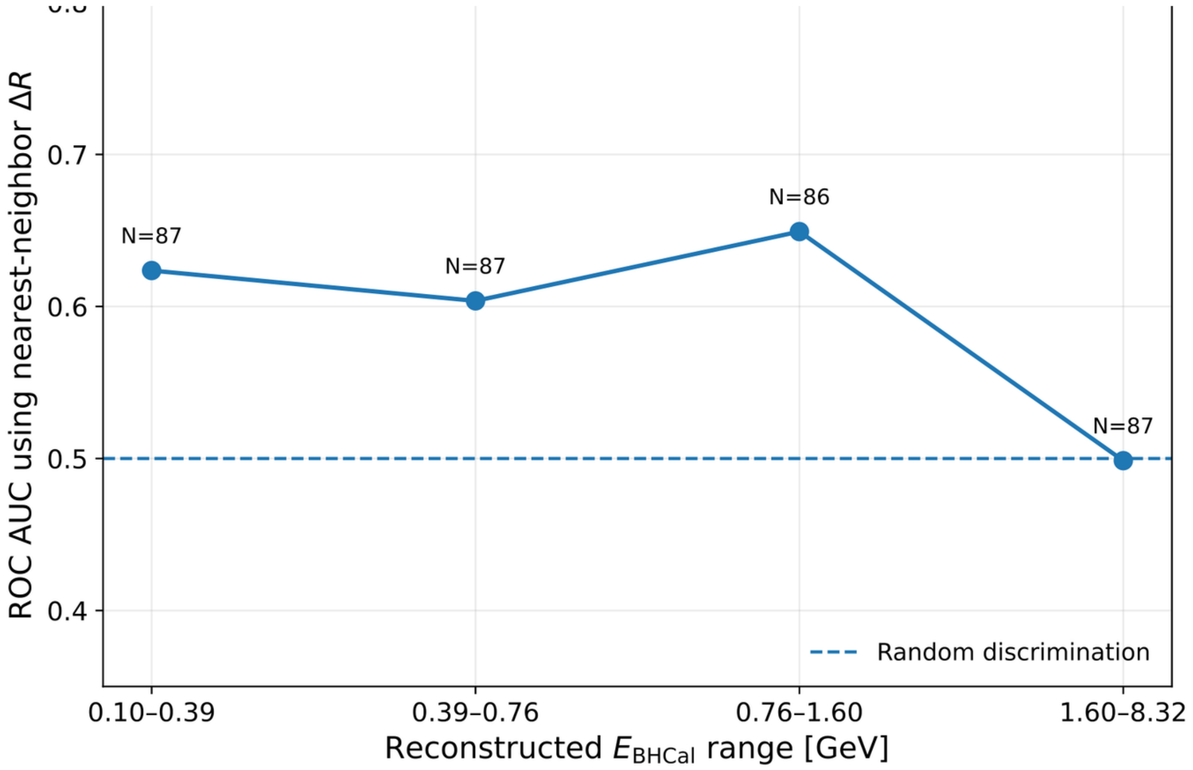
Nearby BHCal Structure Carries Candidate-Quality Information

- Observable: ΔR to the nearest other BHCal cluster.
- GOOD candidates preferentially have a closer neighboring BHCal cluster.
- The effect is reproduced in the independent 80-file sample.
- Better neutron candidates are not necessarily isolated.
- Nearby activity may reflect fragmented hadronic shower development.



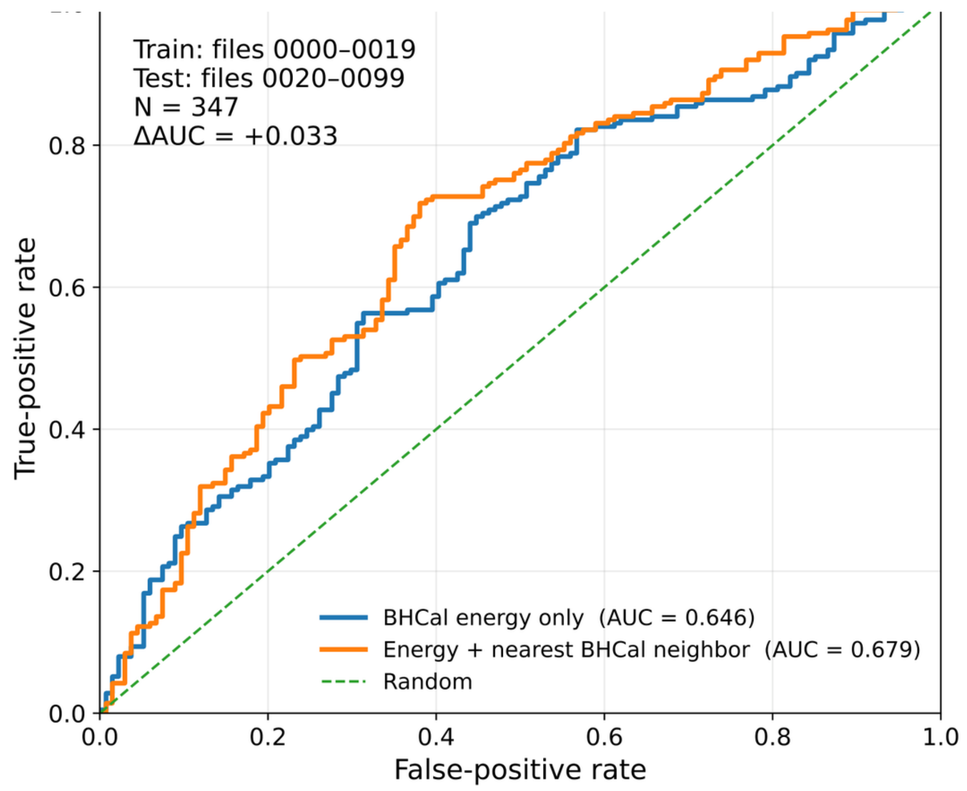
Neighbor Distance Retains Information Within Energy

- Neighbor-distance discrimination was tested within reconstructed-energy bins.
- AUC ≈ 0.60–0.65 over most of the low/intermediate-energy range.
- The effect is therefore not explained entirely by candidate energy.
- Discrimination becomes weak in the highest-energy bin.
- Neighbor distance provides complementary, energy-dependent information.



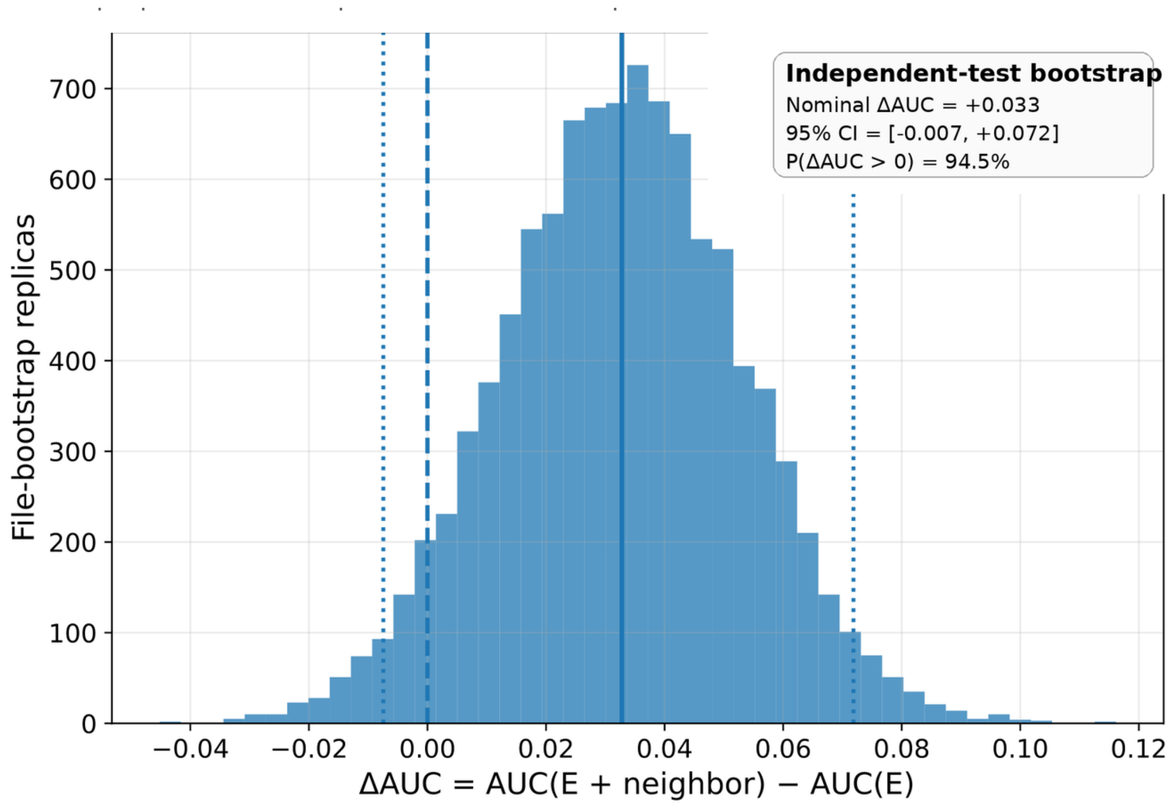
Nearest BHCal Neighbor Improves Strict Independent Validation

- Train: files 0000–0019.
- Test: completely independent files 0020–0099.
- Independent test sample candidates: N = 347 candidates.
- Energy only: AUC = 0.646.
- Energy + nearest BHCal neighbor: AUC = 0.679.
- ΔAUC = +0.033.



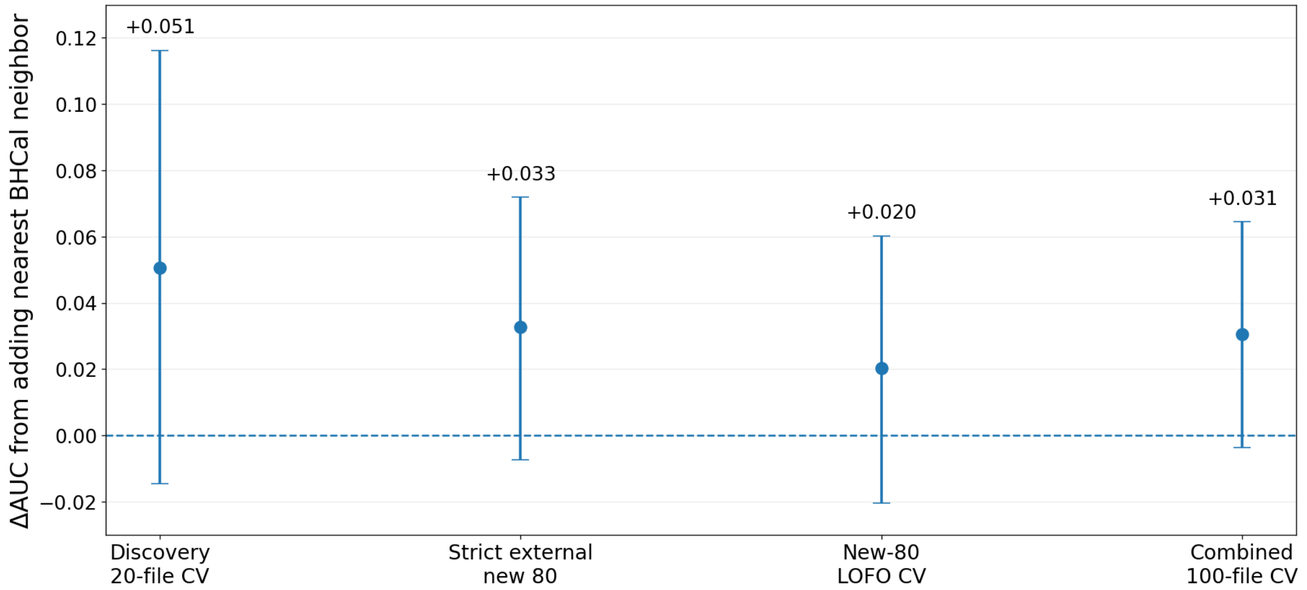
Independent Validation Favors a Positive Neighbor-Distance Gain

- 10,000 paired file-bootstrap replicas.
- Nominal $\Delta AUC = +0.033$.
- 95% interval: approximately -0.007 to $+0.072$.
- $P(\Delta AUC > 0) = 94.5\%$.
- Positive trend is reproducible, but not yet significant at the 95% level.

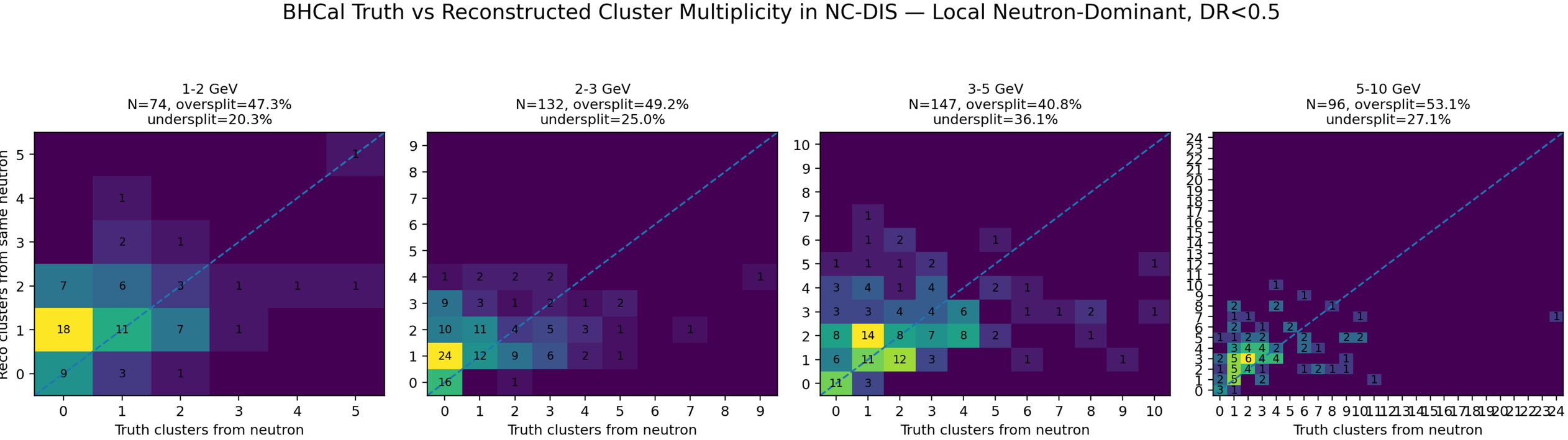


Neighbor-Distance Gain Persists Across Independent Evaluations

- Discovery 20-file CV: ΔAUC = +0.051.
- Strict independent new-80 test: ΔAUC = +0.033.
- New-80 LOFO CV: ΔAUC = +0.020.
- Combined 100-file CV: ΔAUC = +0.031.
- All central values remain positive.
- More statistics are needed before defining a final selection.

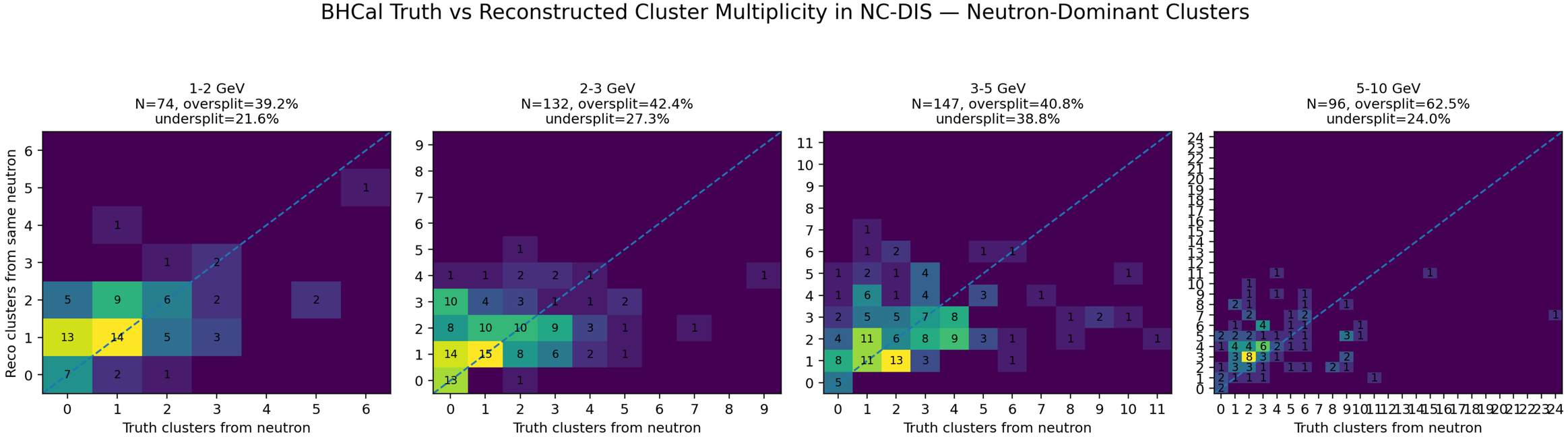


Local Neutron-Dominant Clusters Show Persistent Oversplitting



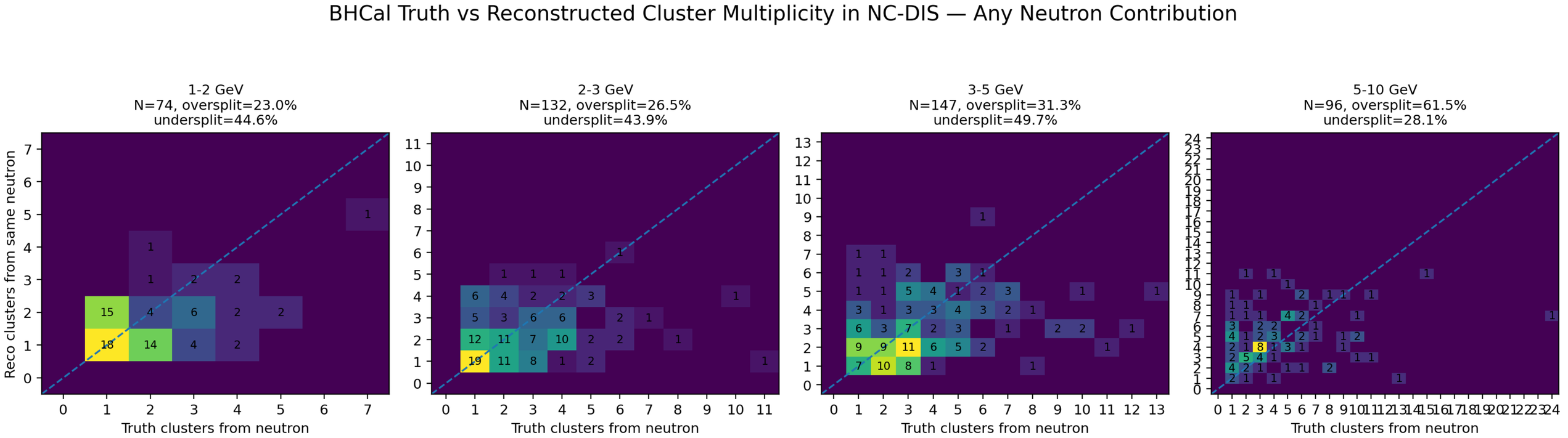
- Require neutron dominance and $\Delta R < 0.5$.
- Focuses on the local neutron shower.
- Closest DIS analogue to the single-neutron study.
- Oversplitting remains about 40–53% across 1–10 GeV.
- Local shower fragmentation is common in DIS.

Neutron-Dominant Cluster Multiplicity



- Count only clusters where the neutron is the dominant contributor.
- Removes small neutron contributions to mixed clusters.
- Oversplitting remains substantial at all energies.
- About 39–42% oversplit below 5 GeV.
- Oversplitting reaches 62.5% at 5–10 GeV.

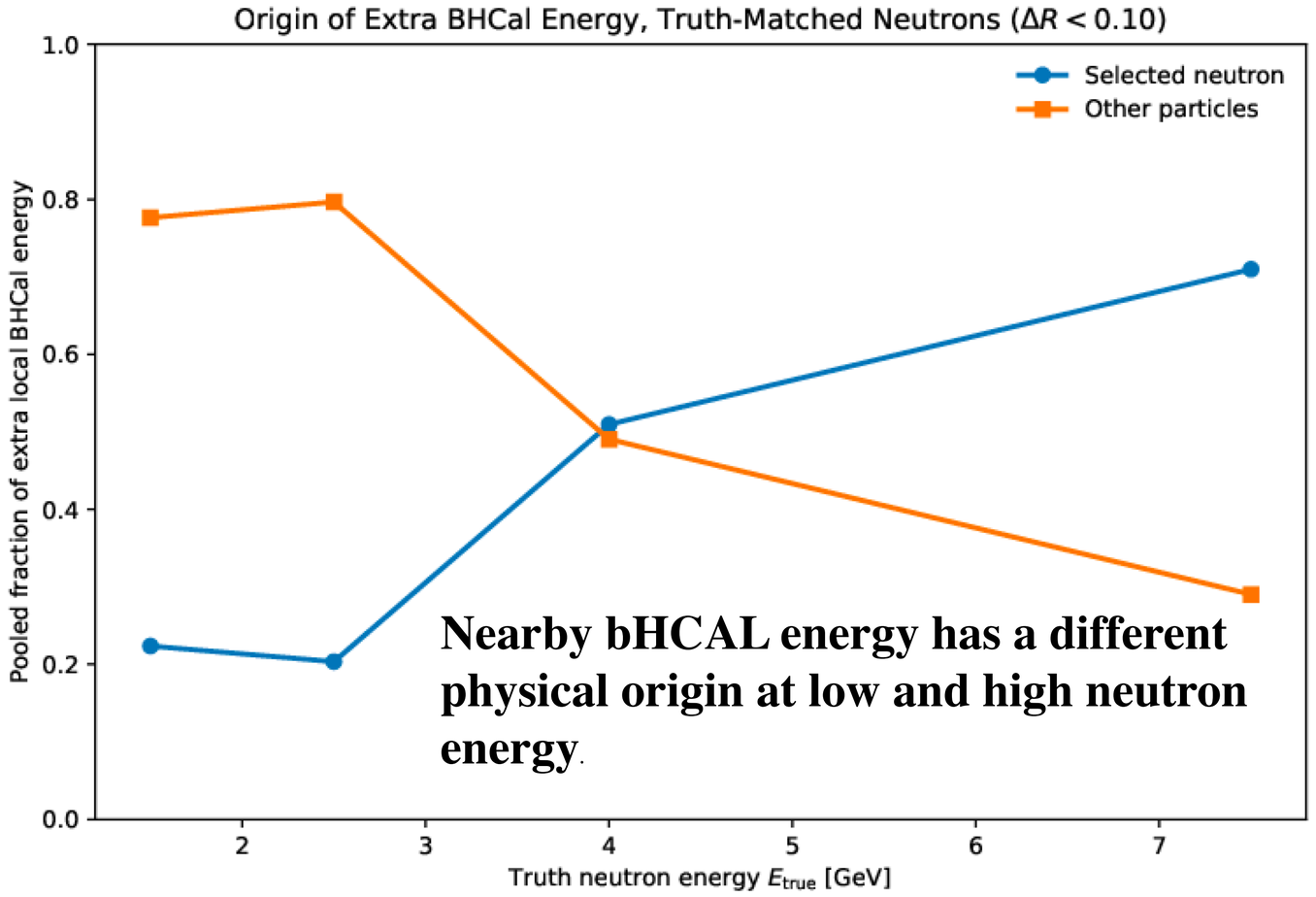
Any Neutron Contribution Gives a More Complex Picture



- Count every truth/reco cluster receiving any contribution from the neutron.
- Low-energy bins show substantial undersplitting.
- 1–5 GeV: undersplitting is about **44–50%**.
- At 5–10 GeV, oversplitting becomes dominant: **61.5%**.
- Small neutron contributions to mixed DIS clusters complicate multiplicity counting.

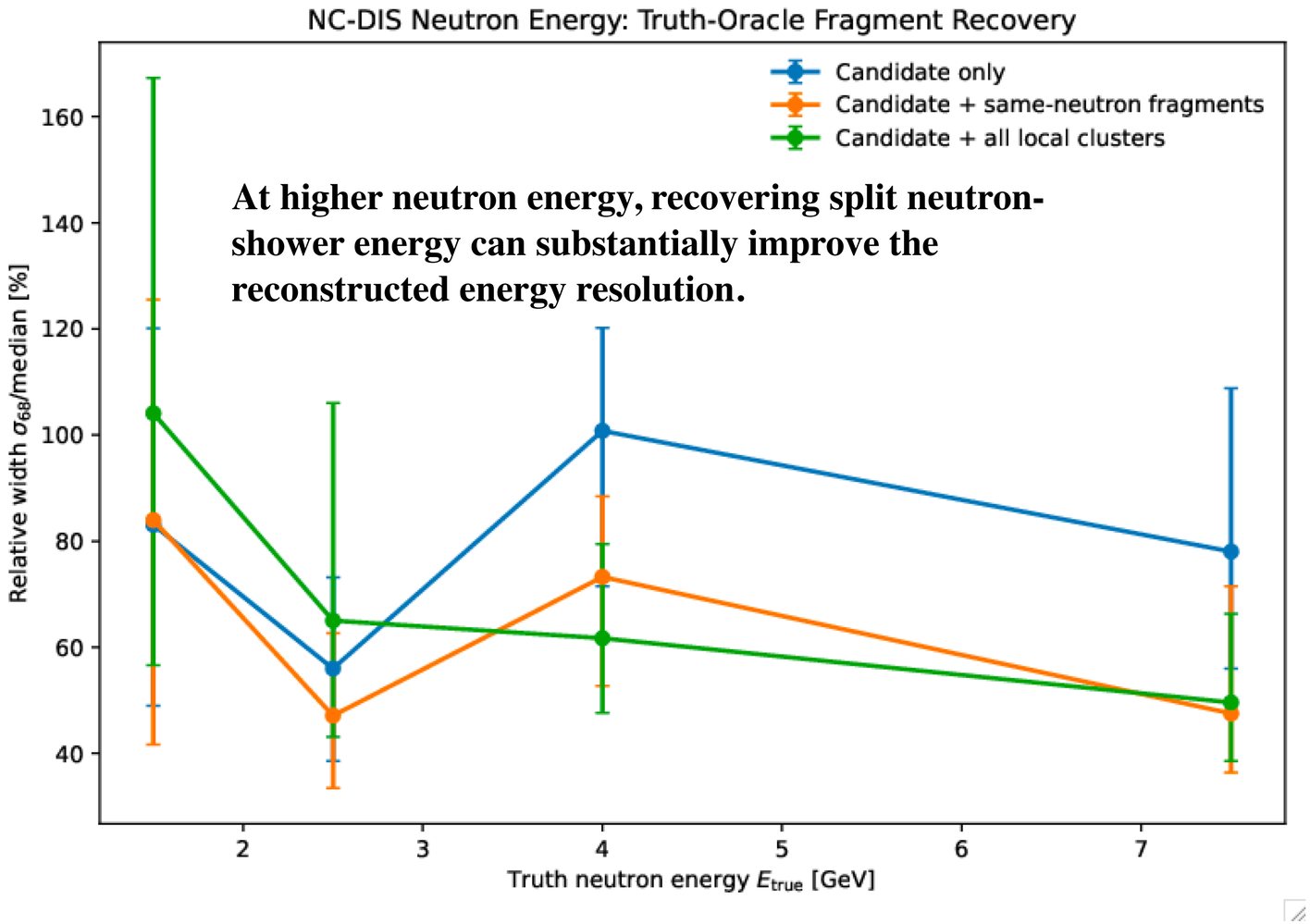
Origin of Extra Local bHCAL Energy

- Use well truth-matched neutron candidates with DeltaR < 0.10.
- At 1–3 GeV, about 80% of the extra nearby bHCAL energy comes from other DIS particles.
- At 3–5 GeV, the extra energy is roughly equally shared between the neutron shower and other particles.
- At 5–10 GeV, about 71% of the extra local energy comes from the same neutron.
- This shows a transition from low-energy contamination to high-energy neutron-shower fragmentation



Truth-Guided Recovery of Same-Neutron Fragments

- Compare the selected candidate with recovery of nearby same-neutron fragments.
- At 1–2 GeV, recovering same-neutron fragments gives essentially no resolution improvement.
- At 2–3 GeV, the improvement is modest and not yet statistically conclusive.
- At 3–5 GeV, the relative width improves from about 101% to 73%.
- At 5–10 GeV, the relative width improves from about 78% to 48%.
- The improvement above 3 GeV shows that shower fragmentation is a potentially recoverable source of poor DIS neutron resolution.



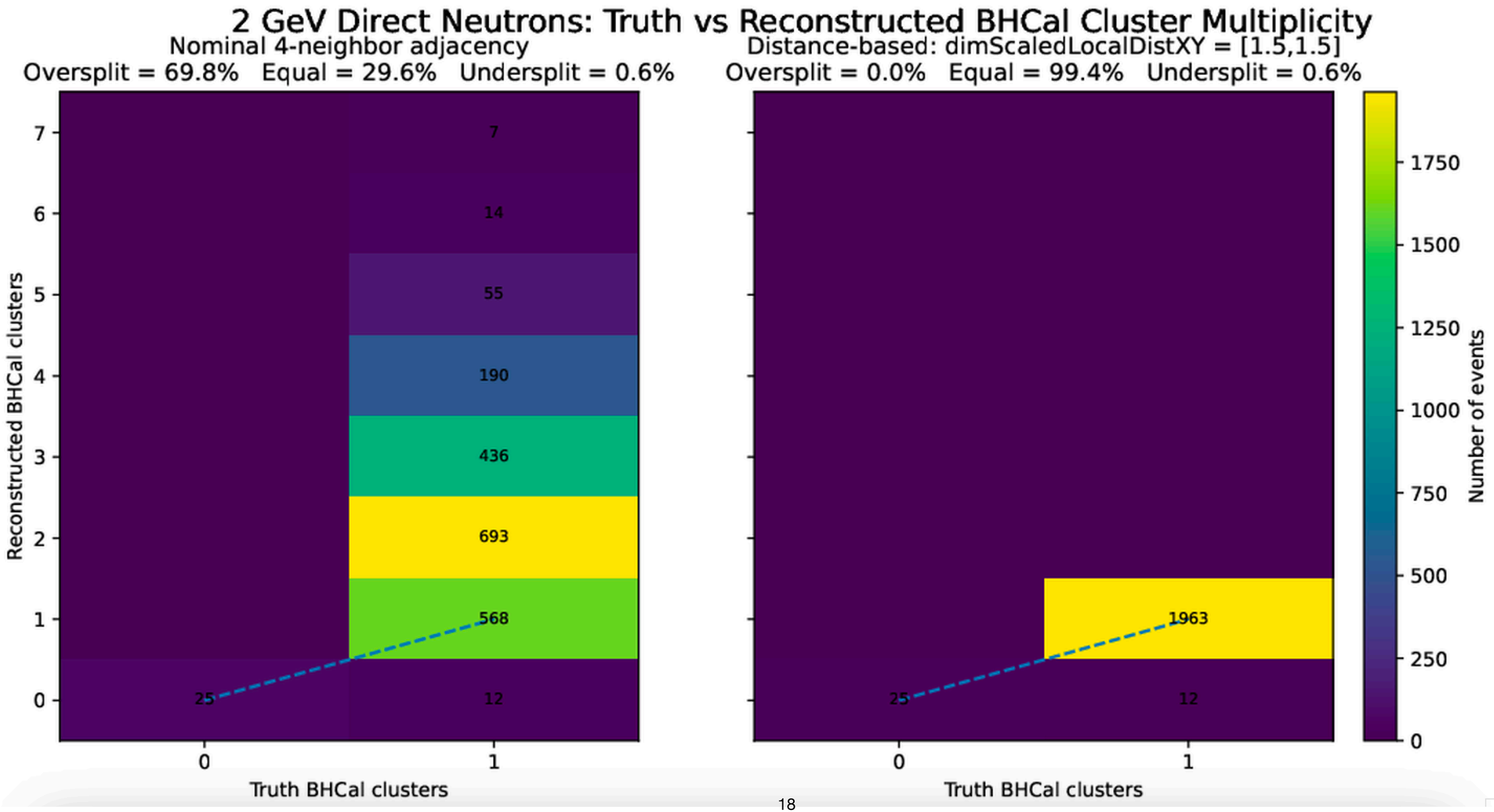
BHCal Clustering: Adjacency Matrix vs Distance-Based Neighboring

- **Adjacency-matrix clustering**
  - Neighboring cells are predefined from the BHCal cell indices/geometry.
  - Start from an active cell and add directly adjacent active cells to the cluster.
  - Continue through connected neighboring cells until no more adjacent hits remain.
  - If one neutron shower produces two blobs separated by empty cells, they are reconstructed as separate clusters.
  - This can naturally lead to cluster splitting for irregular hadronic showers.
- **Distance-based (dR) clustering**
  - Neighboring cells are defined using their spatial separation rather than only direct cell adjacency.
  - Two active cells are connected if their separation is smaller than a chosen threshold, for example dR < R(neighbor).
  - The cluster grows recursively by adding all nearby connected hit cells.
  - A suitable R(neighbor) can reconnect disconnected pieces of the same neutron shower.
  - R(neighbor) must be tuned carefully: **too small leaves shower splitting, while too large can merge unrelated DIS activity.**

Current BHCAL Clustering

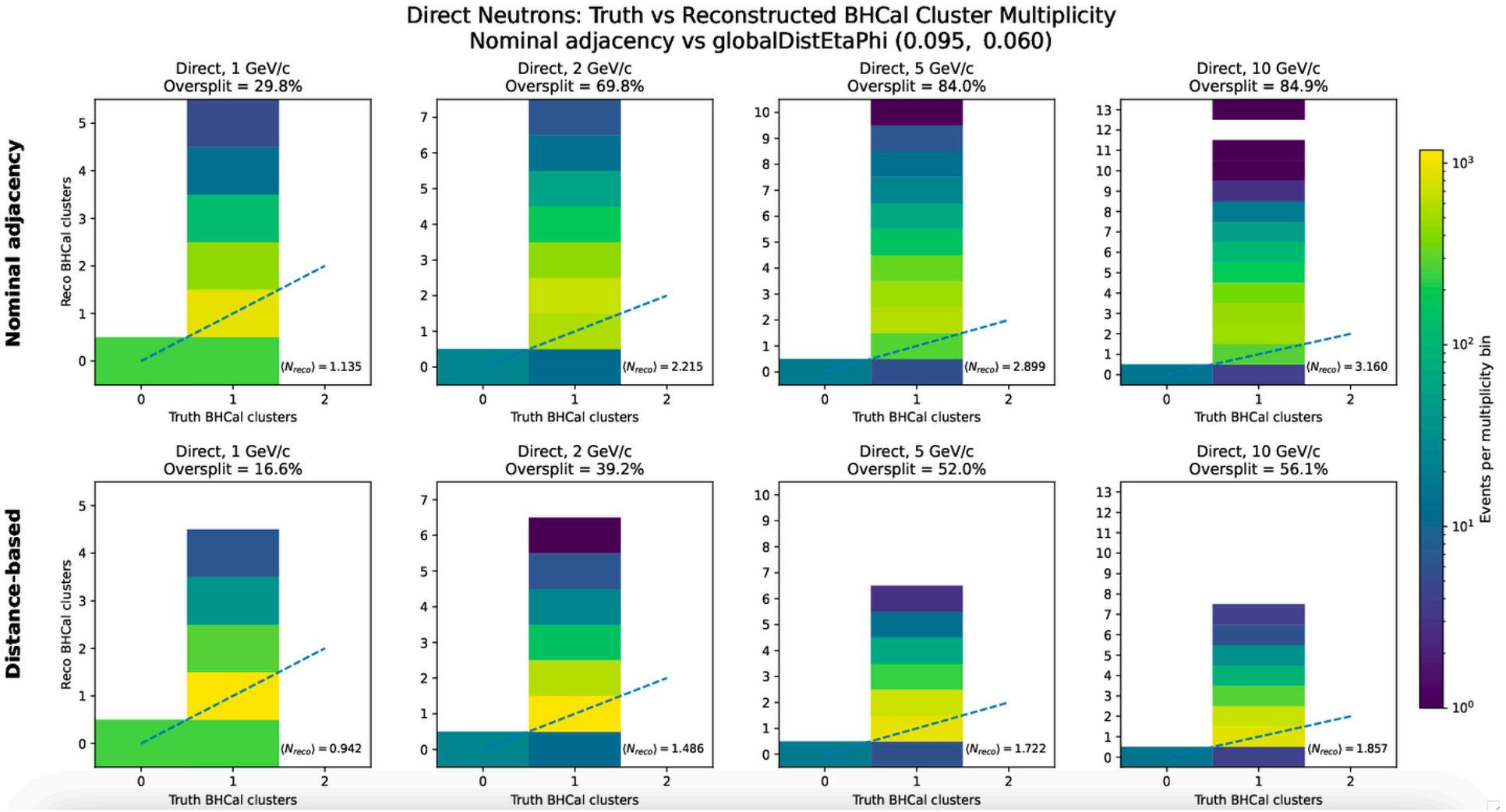
- Current BHCal Island Clustering uses an adjacency-matrix definition of neighboring cells.
- Hadronic neutron showers can produce disconnected energy blobs, which can be reconstructed as separate clusters.
- This explains why one neutron can appear as multiple reconstructed BHCal clusters.
- Derek suggested testing a distance-based neighbor definition, with a tunable distance or dR parameter.
- The goal is to recover neutron shower fragments without merging unrelated DIS activity.

Truth vs Reconstructed BHCal Cluster Multiplicity



Geometry-Based Clustering Reduces BHCal Neutron Shower Fragmentation

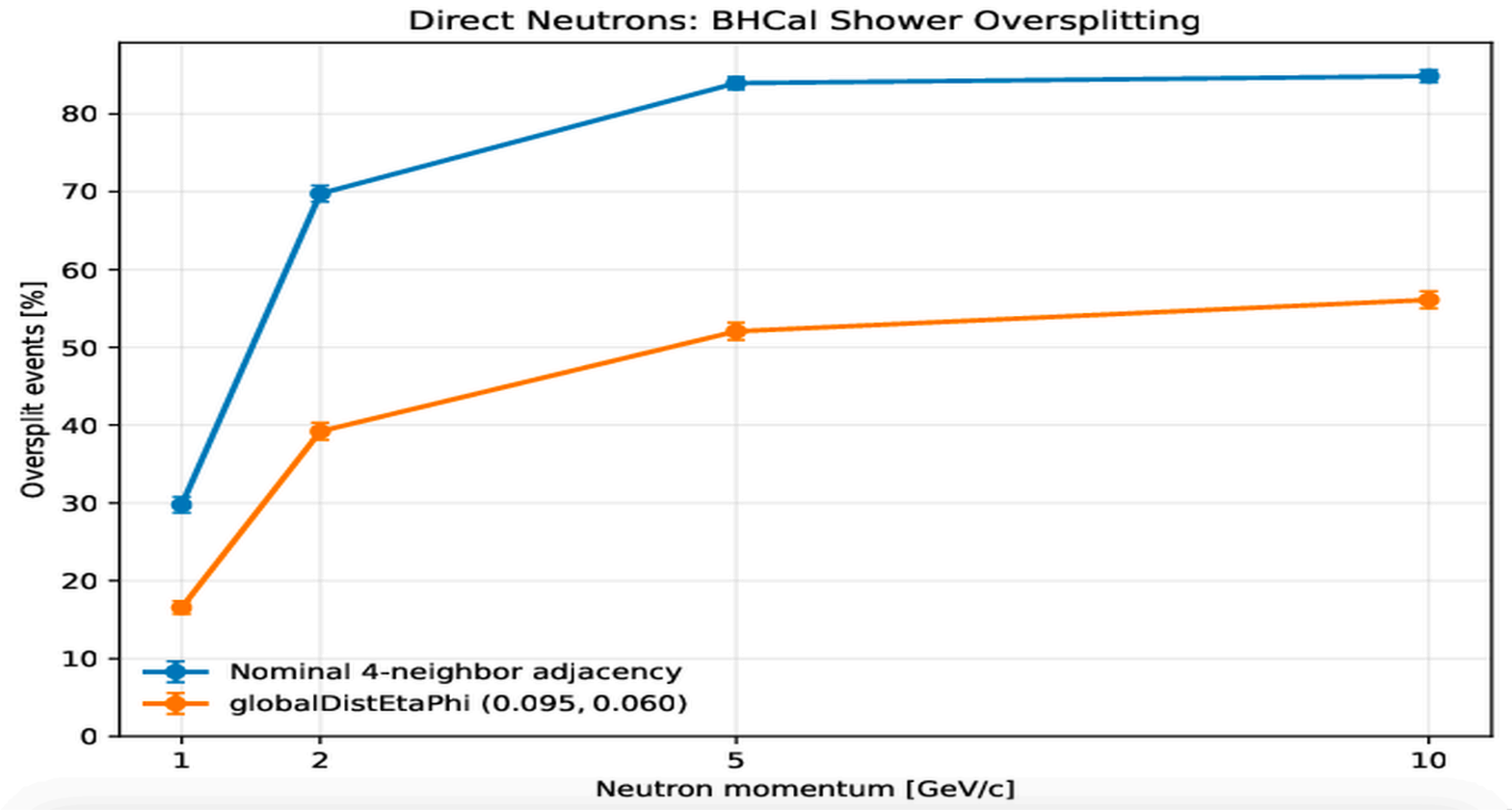
- Nominal adjacency strongly oversplits direct-neutron showers, especially above 2 GeV/c.
- The frozen global eta–phi neighbor criterion reduces oversplitting at all four momenta.
- The reduction is largest at 2–10 GeV/c, by about 29–32 percentage points.
- The new clustering also recovers more energy into the leading neutron cluster and improves its resolution.



Using the frozen global eta–phi neighbor criterion reduces neutron-shower oversplitting across the full 1–10 GeV/c direct-neutron sample. The largest absolute reductions are about 29–32 percentage points in the 2–10 GeV/c range.

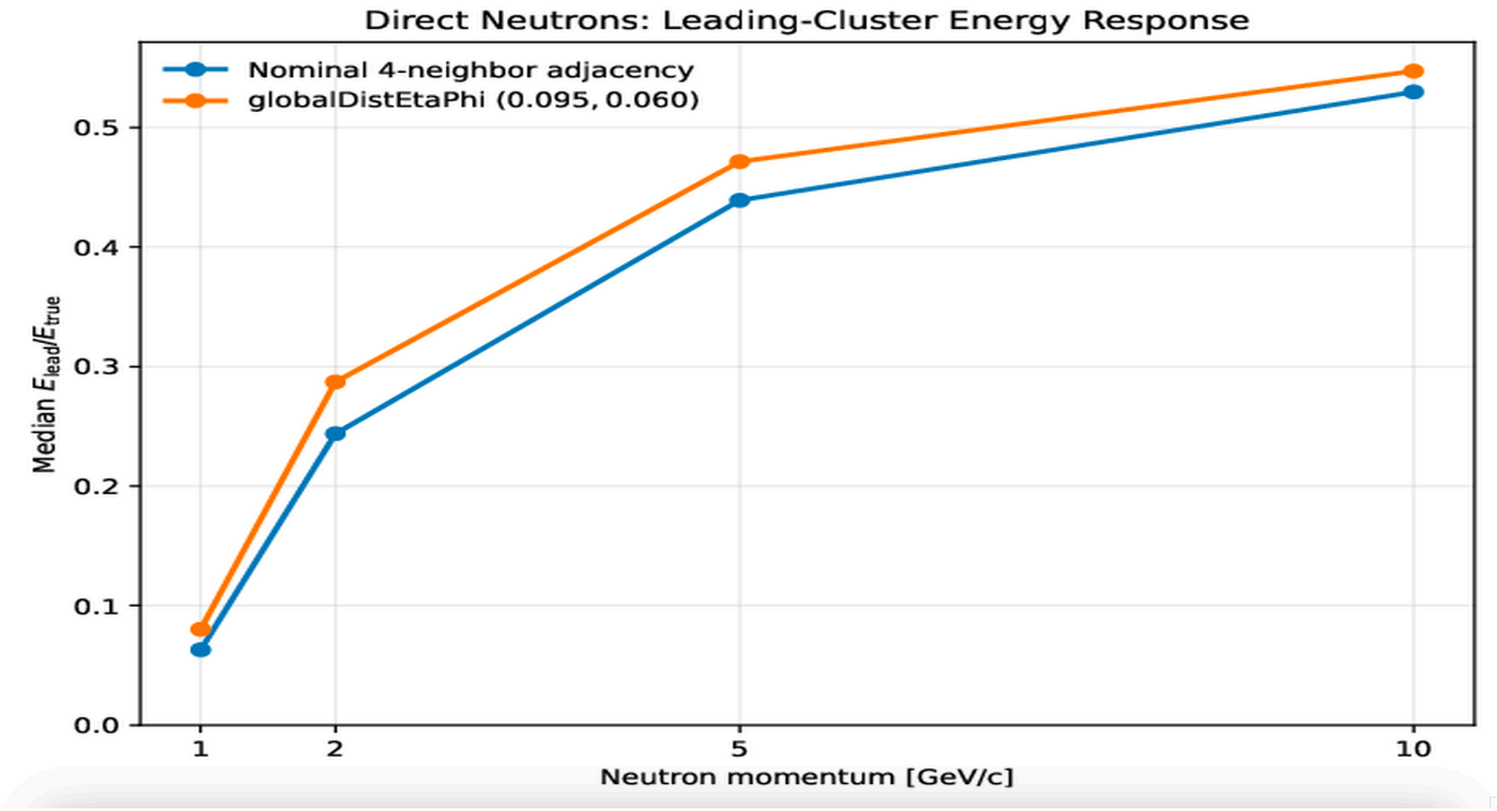
Direct Neutrons: BHCal Shower Oversplitting

- With the nominal 4-neighbor adjacency, oversplitting increases strongly with neutron momentum.
- It rises from about **30% at 1 GeV/c to about 85% at 5–10 GeV/c**.
- The frozen global eta–phi criterion reduces oversplitting at every tested momentum.
- The largest absolute reduction is about **29–32 percentage points between 2 and 10 GeV/c**.
- This shows that the current adjacency definition is an important source of reconstructed neutron-shower fragmentation



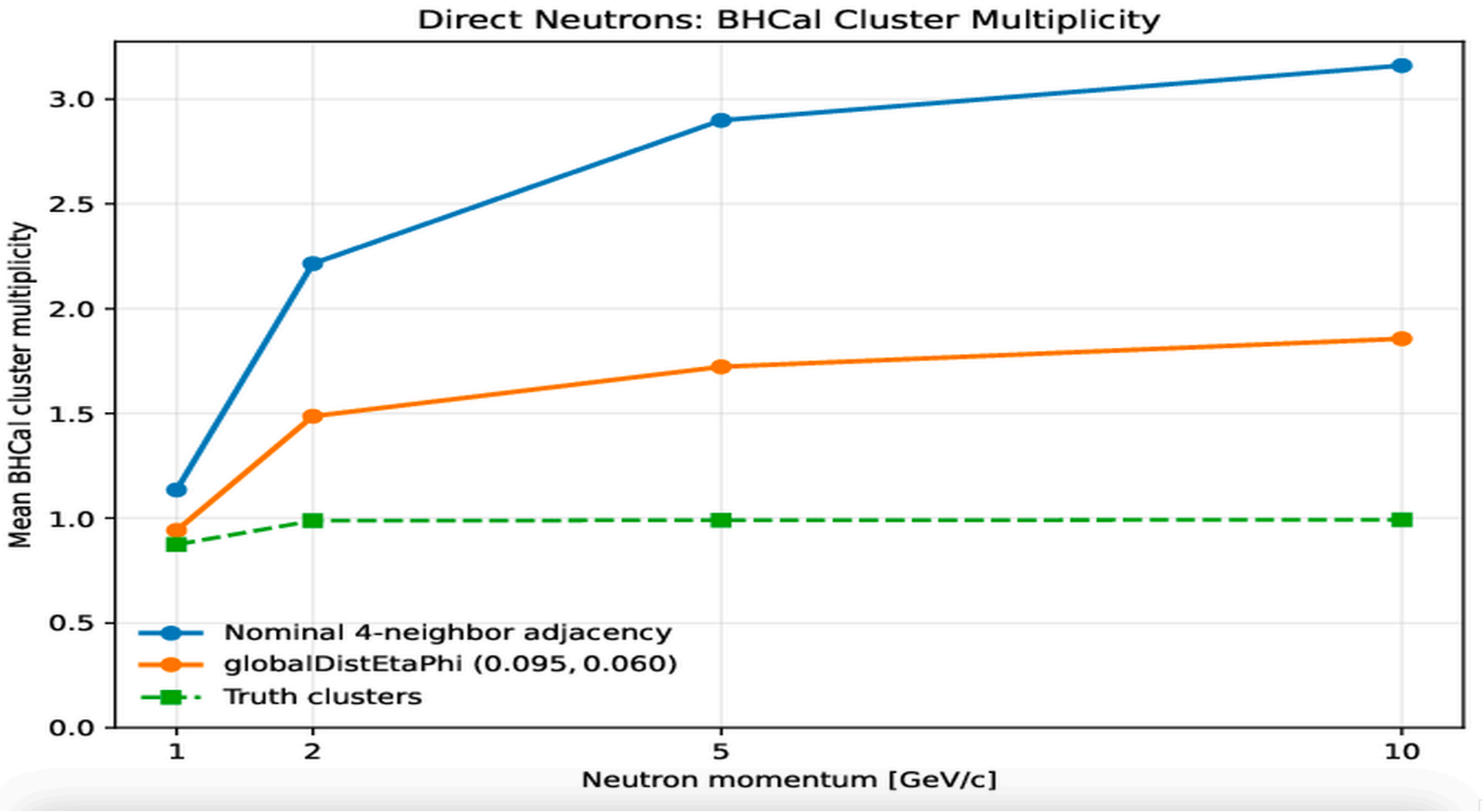
Direct Neutrons: Leading-Cluster Energy Response

- Distance-based clustering increases the fraction of the neutron energy recovered in the leading BHCal cluster.
- At 2 GeV/c, the median response increases from about **0.24 to 0.29**.
- At 5 GeV/c, it increases from about **0.44 to 0.47**, and the improvement remains visible at 10 GeV/c.
- The relative gain is largest at low and intermediate momentum, where shower fragmentation removes a larger fraction of energy from the leading cluster.
- This supports the interpretation that reconnecting separated shower fragments restores energy to the main neutron candidate.



Direct Neutrons: BHCal Cluster Multiplicity

- The truth multiplicity is close to **one cluster per neutron** over the full momentum range.
- With nominal adjacency, the reconstructed multiplicity grows from about **1.1 at 1 GeV/c to 3.2 at 10 GeV/c**.
- The frozen distance-based clustering reduces the multiplicity substantially at every momentum.
- At 10 GeV/c, the mean multiplicity decreases from **3.16 to 1.86**, moving much closer to the truth value.
- The remaining excess multiplicity shows that fragmentation is reduced substantially, but not completely eliminated.



Summary/next

- DIS studies show that nearby BHCal energy has a strongly energy-dependent origin: contamination dominates at 1–3 GeV, while same-neutron fragments become increasingly important above about 3 GeV.
- Truth-guided fragment recovery significantly improves the high-energy DIS resolution; at 5–10 GeV the relative width improves from about **78% to 48%**.
- The nominal BHCal 4-neighbor adjacency strongly oversplits isolated neutron showers, reaching about **70–85% oversplitting at 2–10 GeV/c**.
- A frozen global eta–phi neighbor criterion, **$(\Delta\eta, \Delta\phi) = (0.095, 0.060)$**, reduces oversplitting at every tested momentum.
- At 2–10 GeV/c, the oversplitting reduction is about **29–32 percentage points**, while the reconstructed cluster multiplicity moves substantially closer to truth.
- Distance-based clustering also recovers more shower energy into the leading cluster; at 2 GeV/c the leading-cluster resolution improves from **79.8% to 70.7%**.
- The summed shower response changes only weakly, supporting the interpretation that the clustering primarily **reconnects fragmented shower energy** rather than changing the underlying detector response.
- **Next:** apply the frozen clustering setting to matched **IP neutrons at 1, 2, 5, and 10 GeV/c**, then validate it in NC-DIS and quantify the balance between same-neutron fragment recovery and unrelated-particle merging.

Thank you for your attention!!!