Status of the SiD/Iowa PFA: Diagnostics and plans

Ron Cassell, Mat Charles, Usha Mallik
and Remi Zaidan

SiD Workshop, University of Oregon, Eugene

Overview

Basic Building Blocks of the (Iowa) PFA

Input hits

γ, e and μ ID

Initial MIPs.

Remaining hits

DTree Clustering

Large Clusters

DTree Sub-clustering

MIPs, clumps,
blocks and leftovers

Track matching

Linkable
Clusters

Seeds

Scoring

Shower building

Reconstructed Particles

MC hits within 100 ns from IP are digitized

Hits belonging to photons, muons, electrons and initial MIPs are removed from the hit list for clustering algorithm

Next Use Directed Tree Clustering for classifying the remaining hits into sub-cluster types like MIPs, Clumps, Blocks and leftovers.

Energy from the leftovers is shared among MIPs, clumps and blocks.

Tracks without an initial MIP are tentatively matched to anything in the calorimeter.

Build hadron showers for one charged track at a time starting with lowest momentum.

Unused clusters are then used to build neutral hadron showers.

Outline

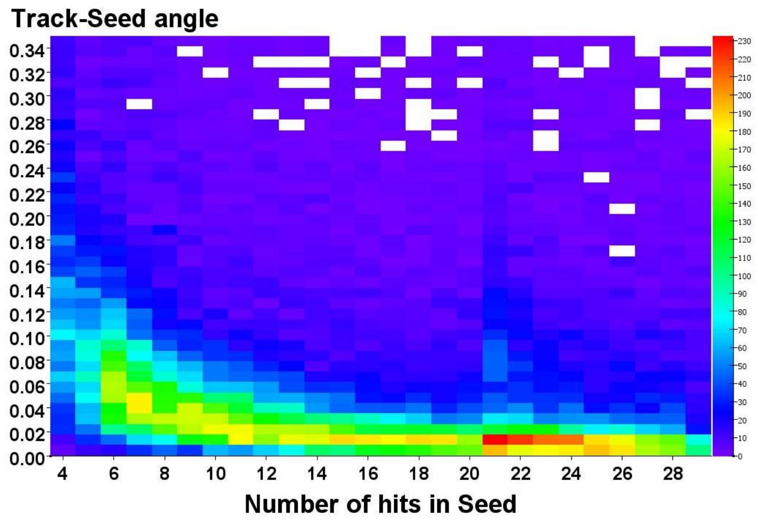
- Diagnostic tools are developed to test the PFA performance at each step of the algorithm:
  - Track-seed matching:
    - Matching quality, properties of unmatched tracks.
  - DTree sub-clusters:
    - Purities and energy contributions.
  - Link properties:
    - Variables used for scoring.
    - Scores before and after the cone algorithm.
  - Shower properties:
    - Efficiencies and purities.
    - Energy-momentum balance.
- Data samples:
  - 10,000 qq events at 500 GeV.

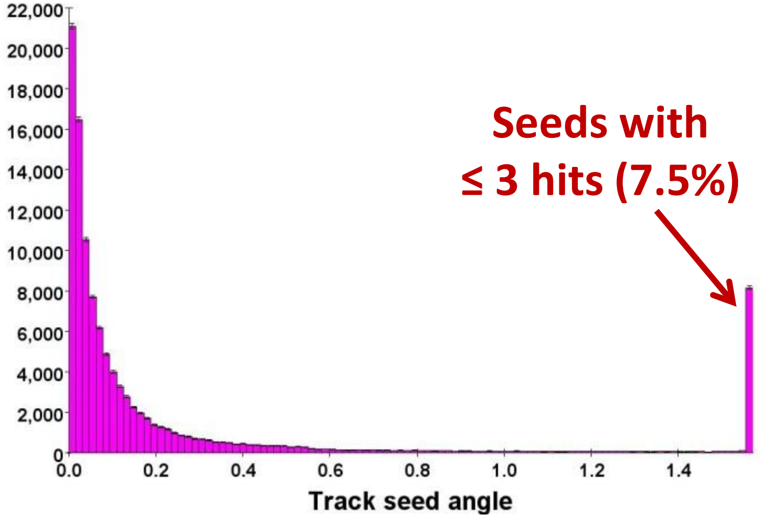
Track-Seed Matching

Track-Seed Matching

- Seed direction is used in the link scoring:
  - Direction is not defined for seeds with less than 4 hits.
  - Direction information is less accurate for seeds with less number of hits.
- Improvement in this area was tried and was satisfactory at this stage:
  - Improvement in the final PFA performance was screened by algorithm performance downstream.

- Possible improvements:
  - Use track direction instead of seed direction for seeds with few hits.





Seeds with ≤ 3 hits (7.5%)

DTree Sub-Clusters

DTree Sub-Clusters

- Possible improvements:
  - Better purity at sub-cluster level.
  - Better treatment of shared clusters.

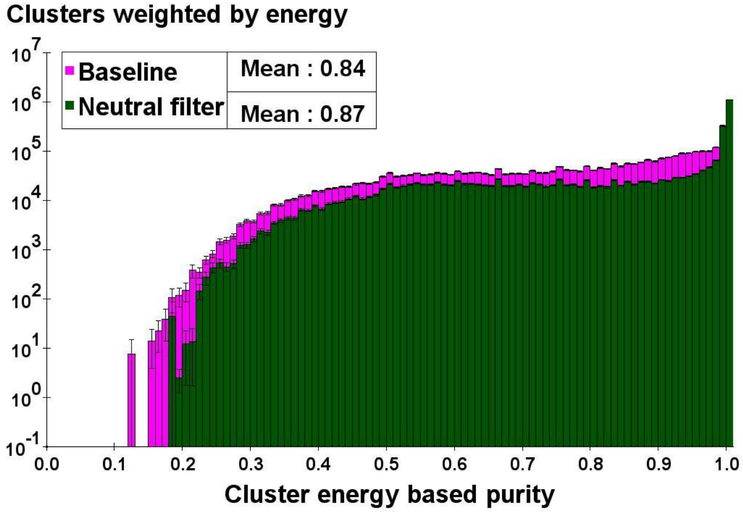
Cluster type	Baseline	Neutral filter
Mip	95%	97%
Clump	84%	88%
Block	82%	86%
Leftovers	74%	79%
Photon	89%	94%

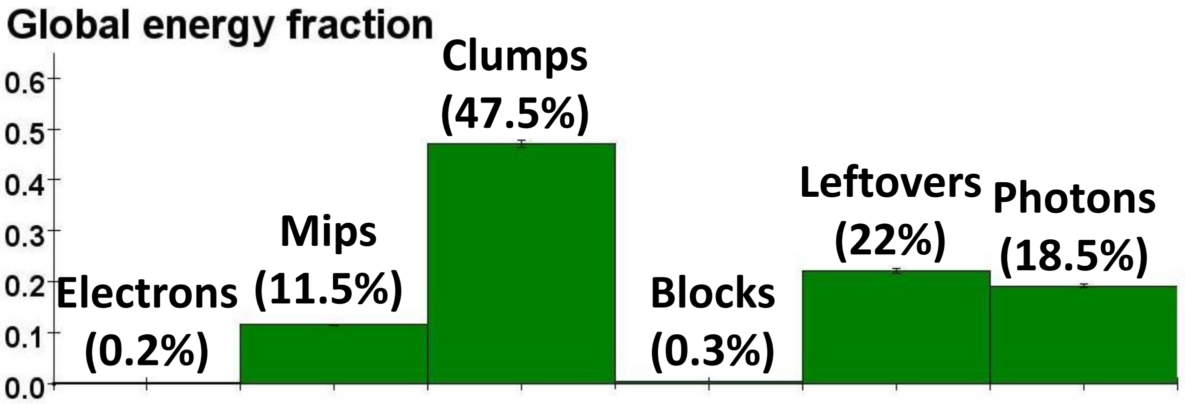
Overall purity
84-87%

Most of the energy goes to clumps:
47.5%

Expect more Photons:
26% → 18.5%

Significant amount of energy is shared:
22 %



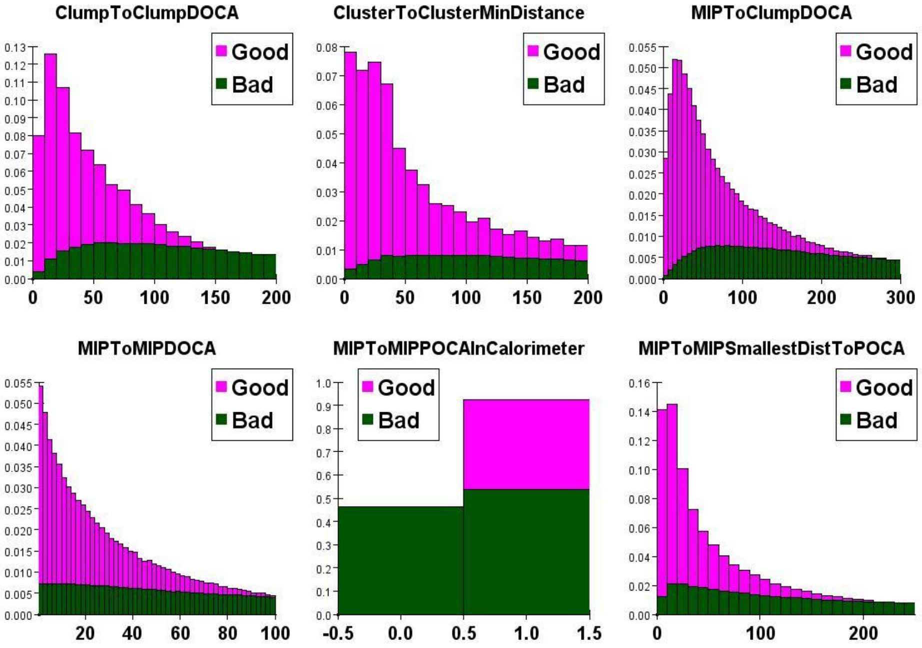


Link Scoring

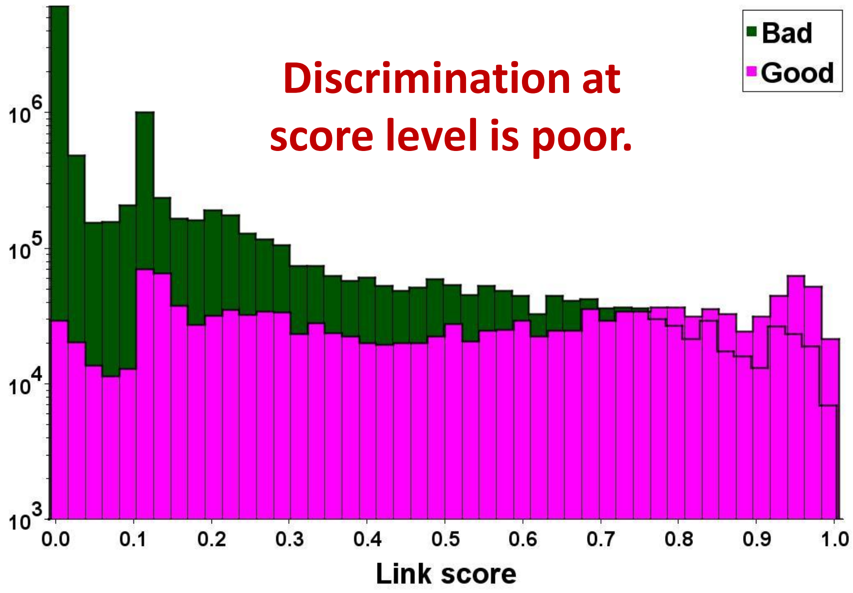
Link Scoring

- Potential improvements:
  - Better definition of a good link.
  - Better use of available information.
  - Use same clustering for training and analysis.

A "good" link is a link where both clusters have dominant energy contribution from the same MC particle

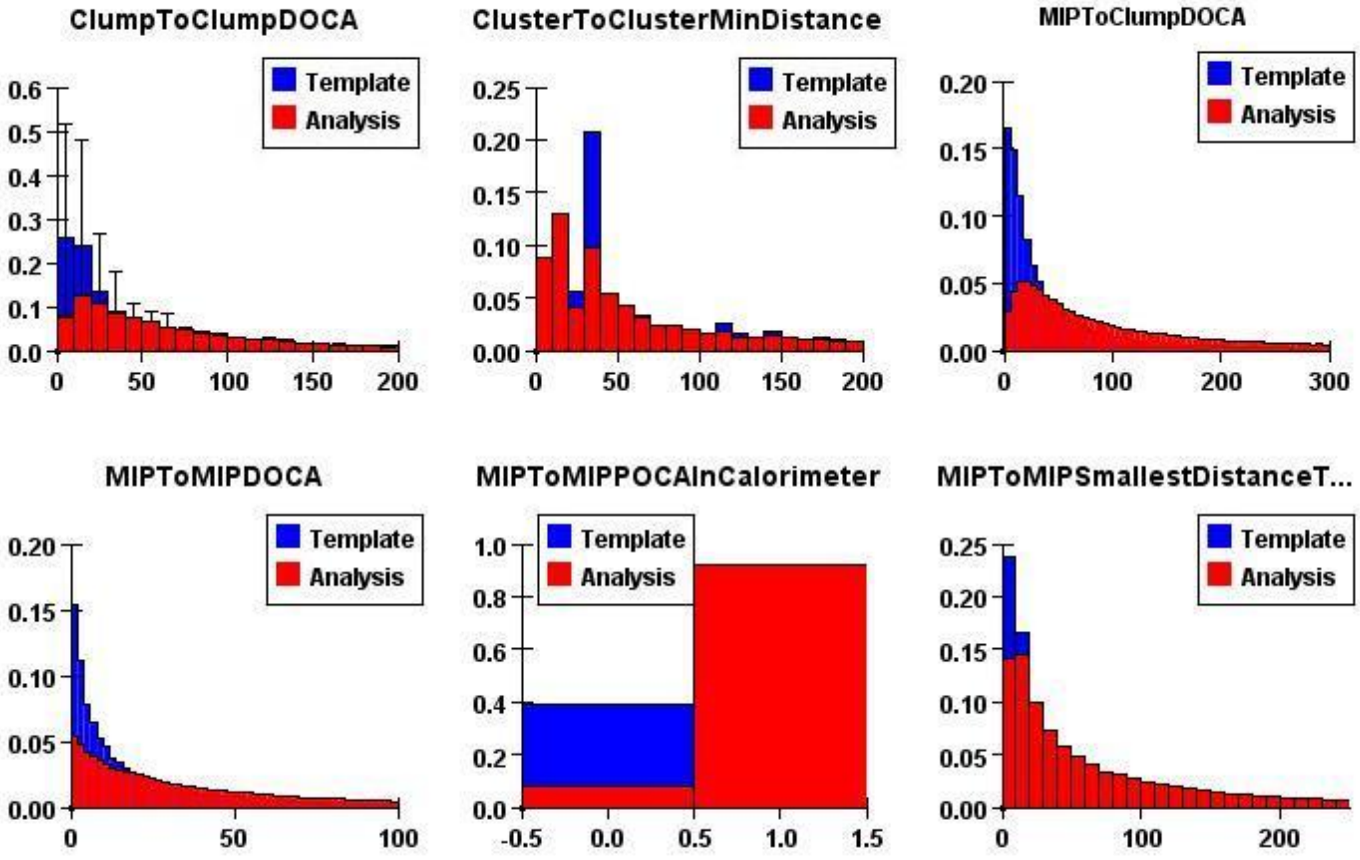


Discrimination at variable level is good.



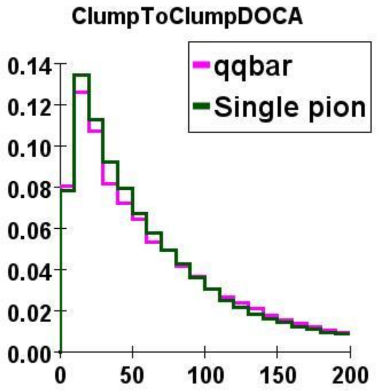
Discrimination at score level is poor.

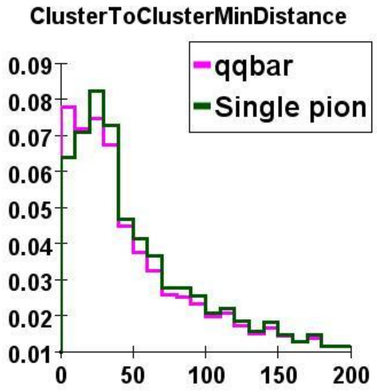
Link Scoring

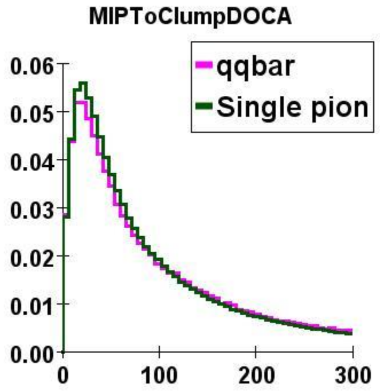


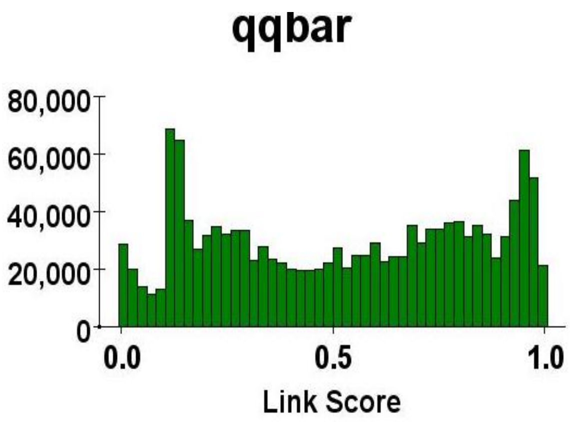
A different clustering algorithm is used when training and when evaluating the likelihoods.

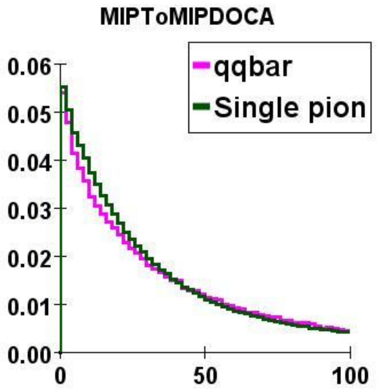
Link Scoring: Comparison to single pions.

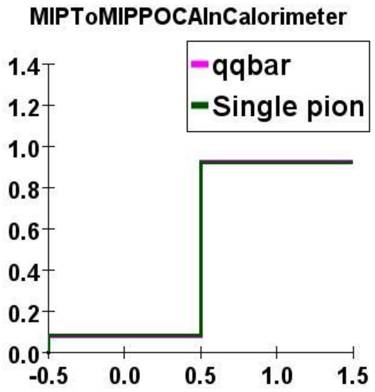


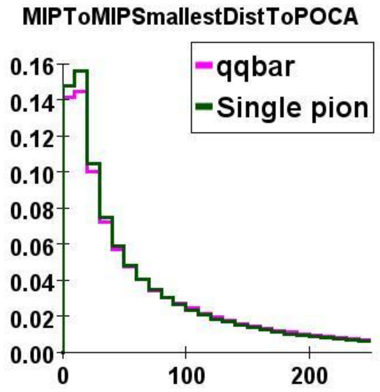


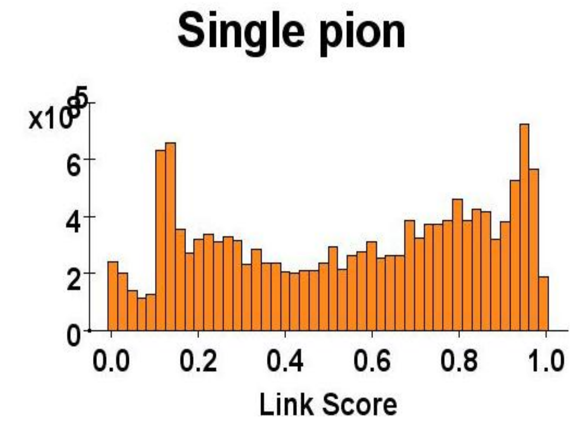










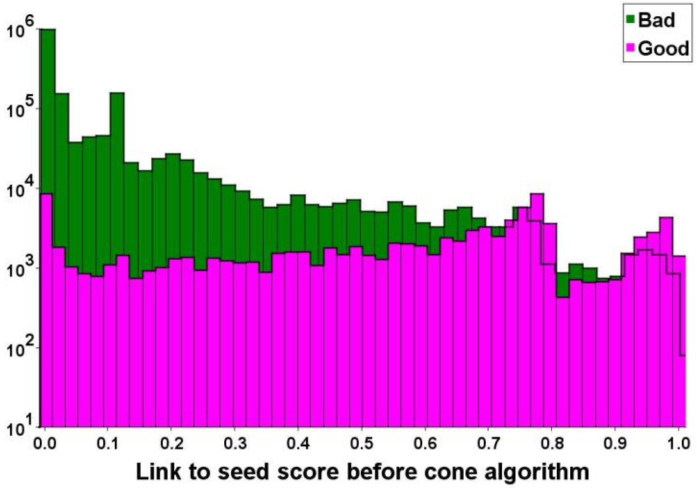


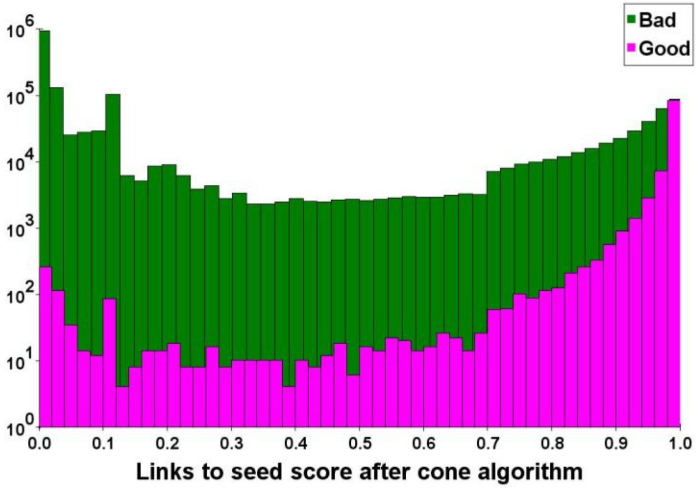
Very similar distributions in general.

Link Scoring:
the Cone Algorithm

- The cone algorithm computes a score between any cluster and the seeds:
  - Based on the opening angle between the cluster and the seed .
  - The link score is only modified if the new score is larger than the old score.
- The algorithm is pretty aggressive:
  - Brings a lot of background into the signal region.
  - Causes the shower building to start far in the calorimeter.

- Potential improvements:
  - Replace the cone algorithm by a more sophisticated algorithm.
  - Apply the cone algorithm at a second pass after a first "shower-skeleton" reconstruction.





Shower building

Shower building

The University of Iowa

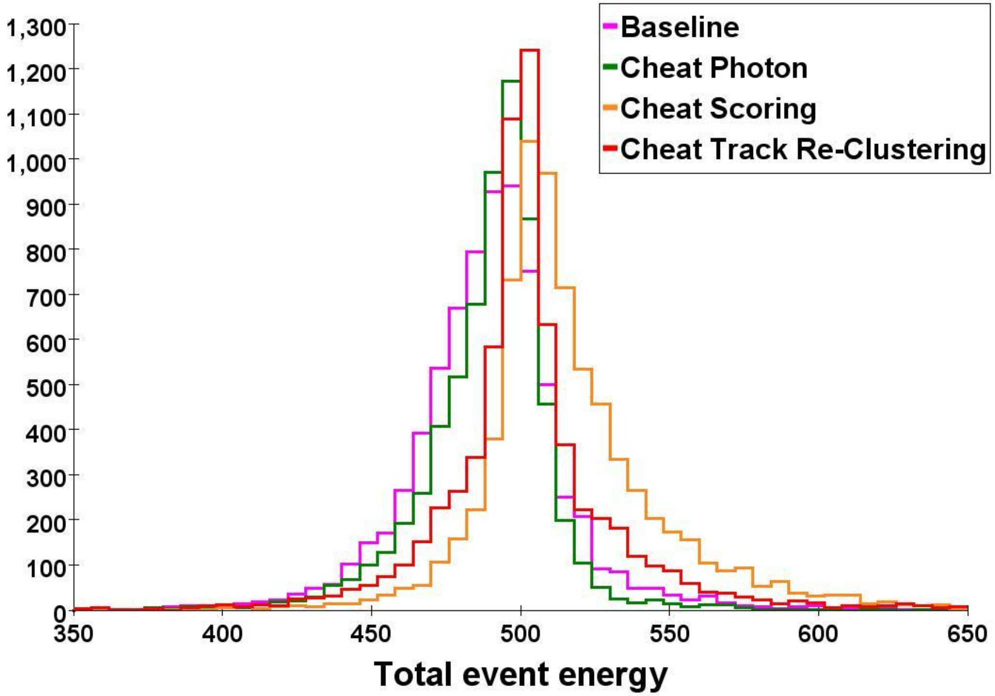
A - Baseline: Algorithm is run out of the box.

B - Cheat Photon Finding: Photons are reconstructed based on MC truth.

C - Cheat Scoring: Link scores are set to 0 or 1 based on MC truth.

D - Cheat track Re-clustering: Clusters are assigned to tracks based on MC truth.

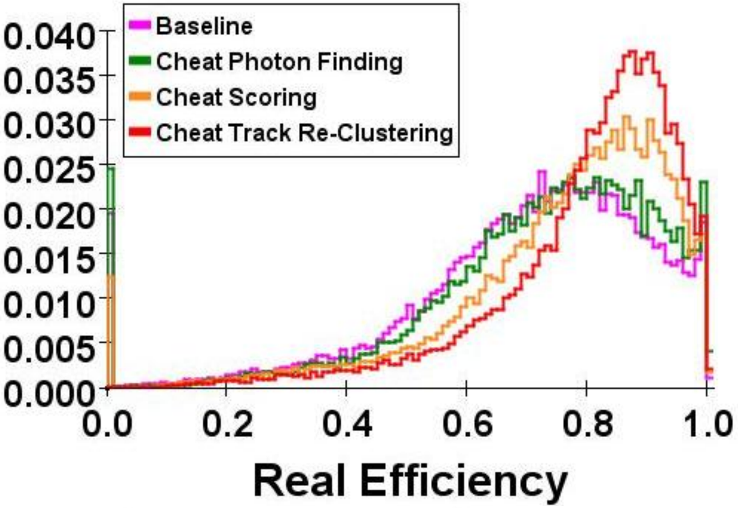
PFA	RMS90
Baseline	3.4%
Cheat Photon Finding	2.8%
Cheat Scoring	3.9%
Cheat Track Re-Clustering	3.6%

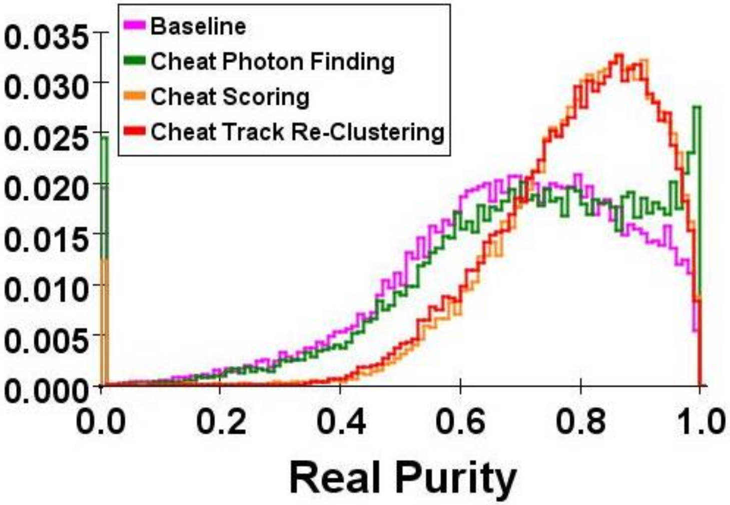


Photon veto

- Identified photons that overlap with initial MIPs and muons are vetoed as photons:
  - Their hits may be used for neutral or hadron showers.
- Photon veto is removing real photon:
  - Intrinsic photon efficiency an purity are at the level of 90%.
  - Observed photon efficiency: ~70%

- Possible improvements:
  - Work to have better purity at the photon level.
  - Change the vetoing algorithm.

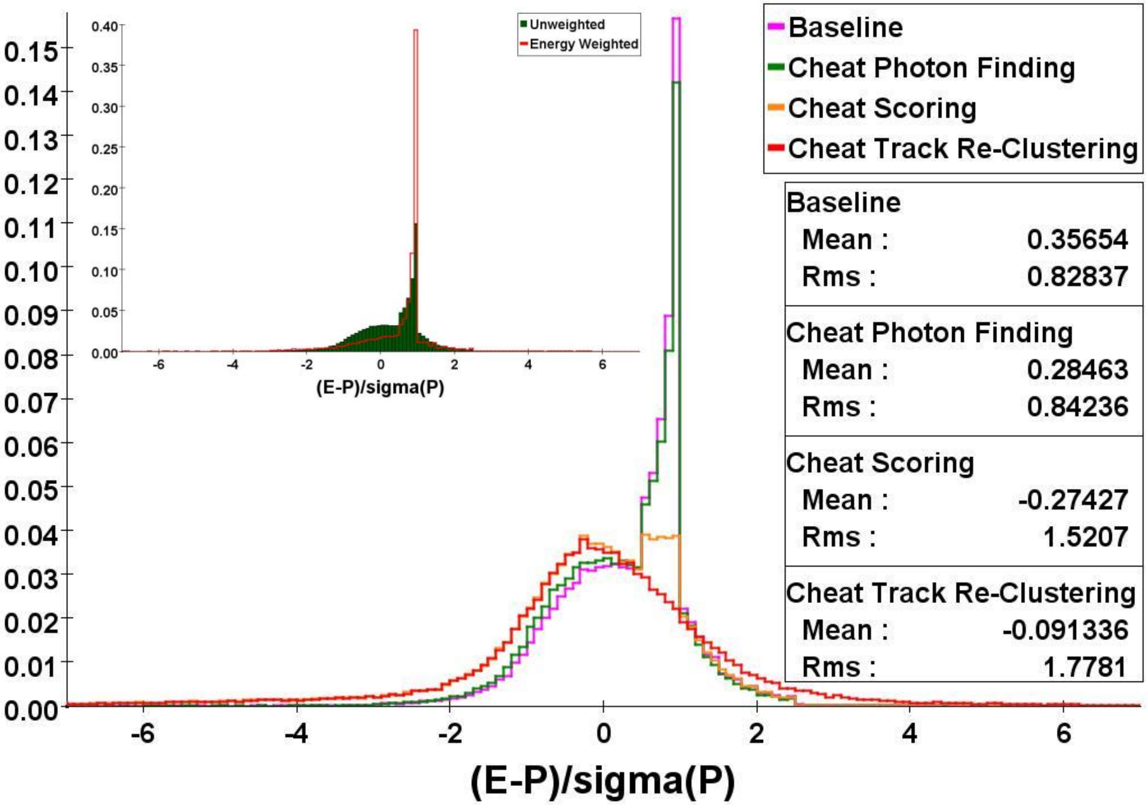




Shower Normalized energy residual to track momentum

- Shower building keeps on propagating until stopped by an energy-momentum constraint:
  - E < P+σ(P) at first iteration.
  - Spike at E-P=1σ(P) appears.
  - Tracks in the spike are mostly high momentum tracks.
- Perfect track-cluster assignment fixes the central value but the RMS is sill too high:
  - Affected by purities at cub-cluster level..

- Possible improvements:
  - Remove E/P constraint from the algorithm.



Conclusion

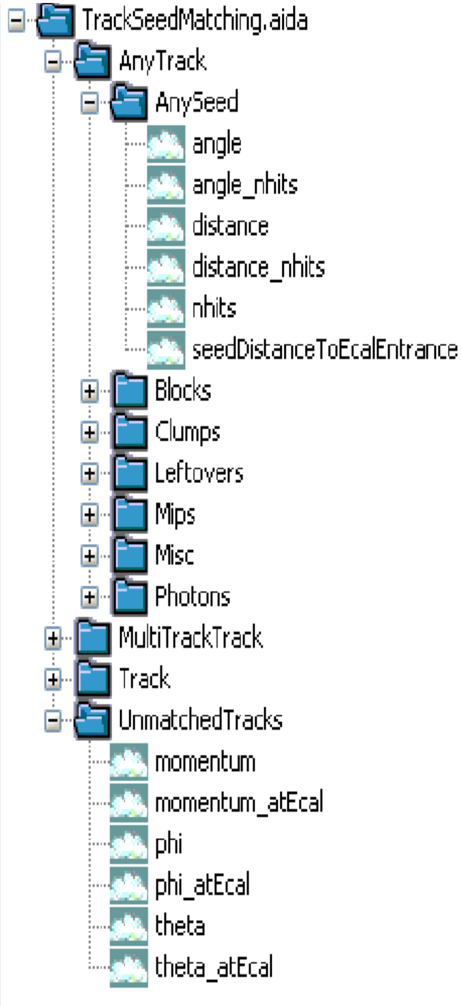
- Areas to work on:
  - **Photon finding:**
    - Need for better photon purity and efficiency.
    - Photon veto: is it doing the correct thing?
  - **Sub-clustering:**
    - Need better purity for Clumps.
    - A lot of energy goes into shared clusters.
  - **Linking:**
    - Better definition of what is a “good” link: may be too ambitious!?
    - Use same re-clustering for training and evaluating.
    - Many information are still used empirically in penalty factors.
  - **Shower building algorithm:**
    - E/P constraint should not be used in a PFA.
    - The cone algorithm is aggressive and hides algorithm problems.

Garabed Halladjian has recently joined the PFA effort.

Backup

Track-Seed Matching: Definitions

- Tracks are extrapolated to the innermost layer with hits from the seed:
  - Angle is computed between the seed direction from energy tensor calculation and the tangent to the extrapolated track.
  - Distance is computed between the track interception point and the closest hits in the cluster on the same layer.
- Seed distance to Ecal entrance is the depth of the innermost layer with hits from the seed.
- Plots are made per seed type and also separating simple from multiple tracks.
- Multiple tracks:
  - Angle and distance to seeds are computed using extrapolation results averaged on sub-tracks.
  - Angle is the maximum angle between "sub-tracks" at the extrapolation layer.
  - Distance is the maximum distance between interception points.
- For unmatched tracks:
  - Momentum, theta and phi are plotted for all unmatched tracks and for those that reach the Ecal.



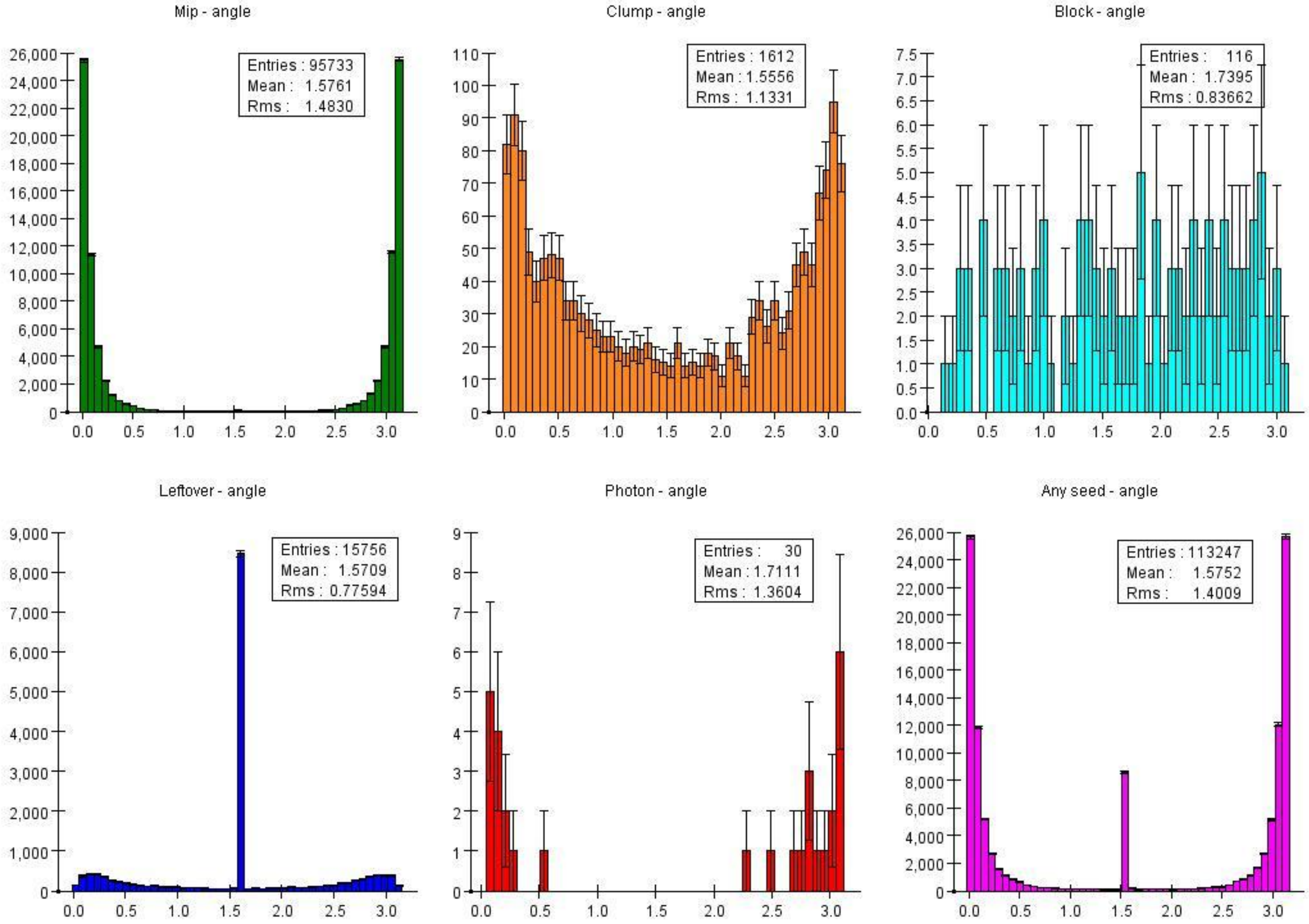
- TrackSeedMatching.aida
  - AnyTrack
    - AnySeed
      - angle
      - angle_nhits
      - distance
      - distance_nhits
      - nhits
      - seedDistanceToEcalEntrance
    - Blocks
    - Clumps
    - Leftovers
    - Mips
    - Misc
    - Photons
  - MultiTrackTrack
  - Track
  - UnmatchedTracks
    - momentum
    - momentum_atEcal
    - phi
    - phi_atEcal
    - theta
    - theta_atEcal

Track-Seed Matching

Matching efficiency
to a mip:
84.5%

Fraction of seeds
with <4 hits:
7.5%

Type	rate
Mips	84.5%
Leftovers	14%
Other	1.5%

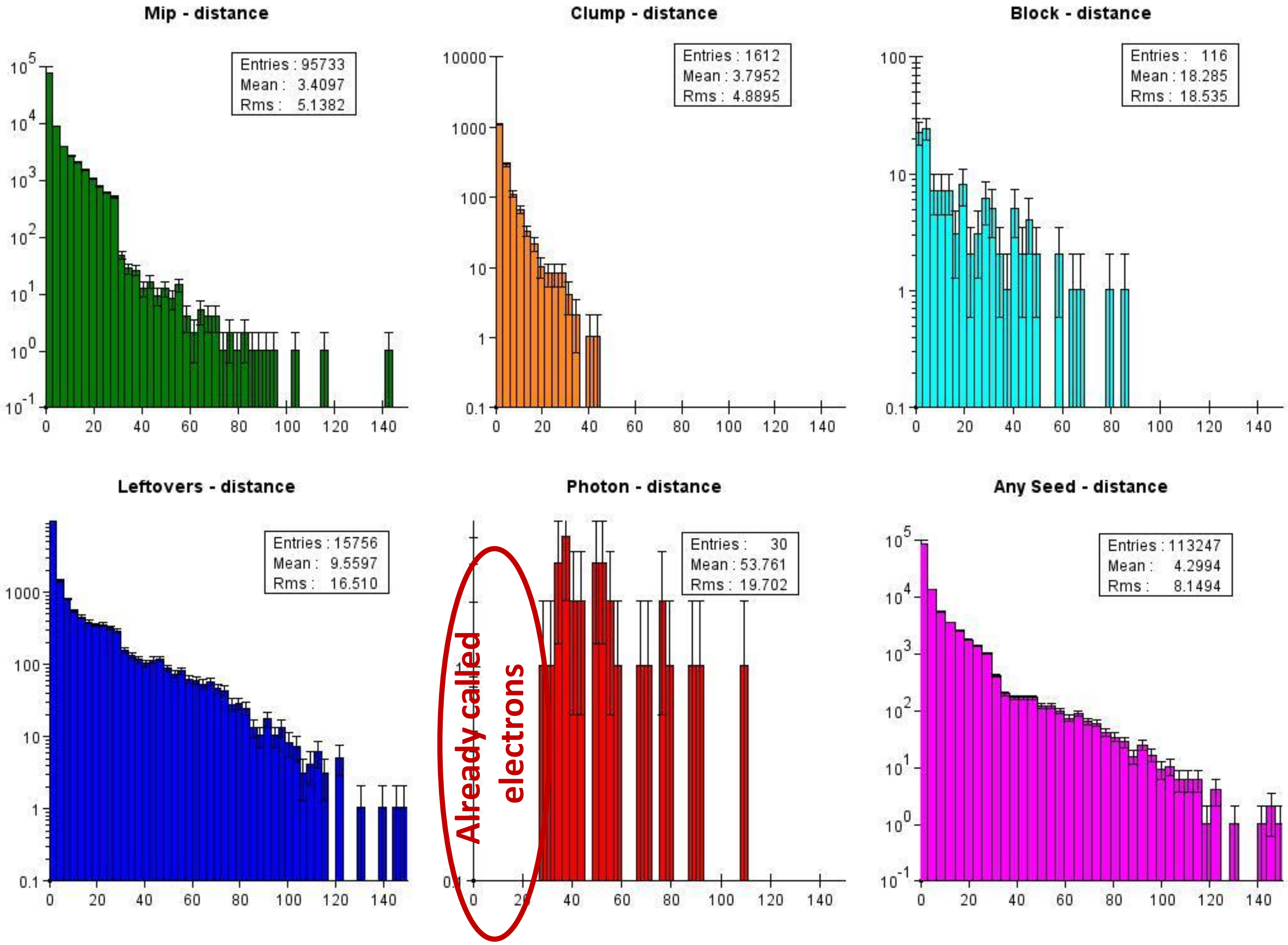


Track-Seed Matching

Matching efficiency
to a mip:
84.5%

Fraction of seeds
with <4 hits:
7.5%

Type	rate
Mips	84.5%
Leftovers	14%
Other	1.5%



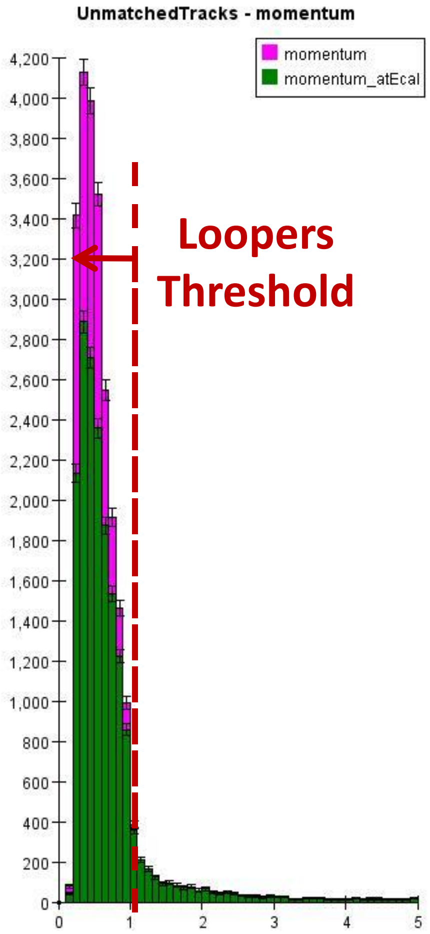
Already called electrons

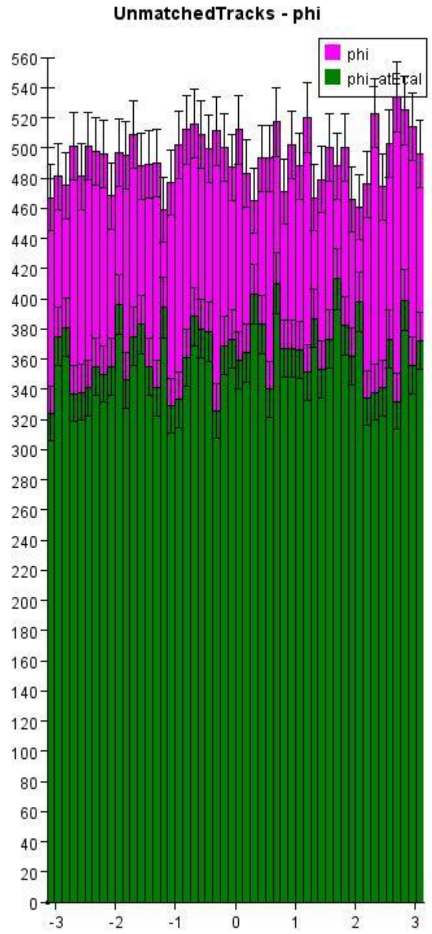
Track-Seed Matching: Unmatched tracks

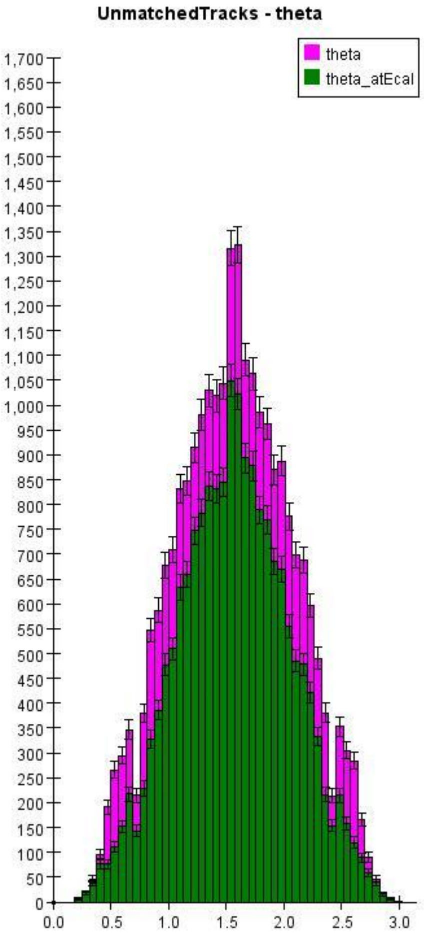
Fraction of unmatched tracks: **18%**

Fraction of unmatched tracks with p<1 GeV: **90%**

Fraction of unmatched tracks that reach Ecal: **75%**







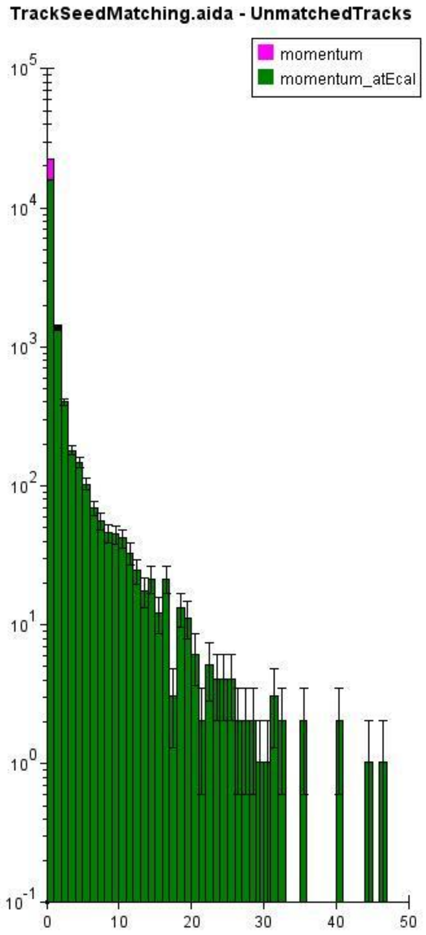
Track-Seed Matching: Unmatched tracks

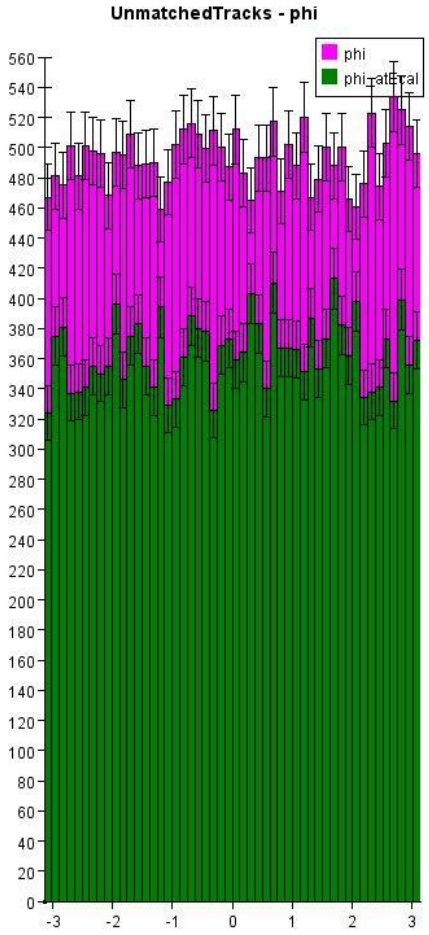
Fraction of unmatched tracks: **18%**

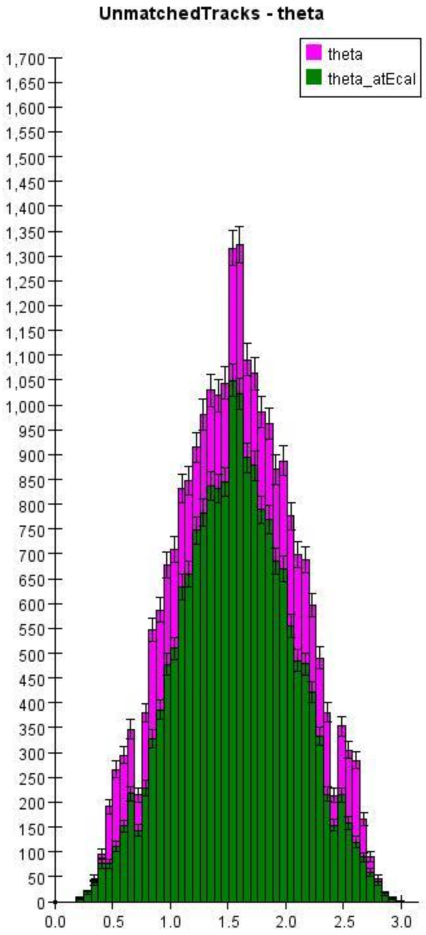
Fraction of unmatched tracks with p<1 GeV: **90%**

Fraction of unmatched tracks that reach Ecal: **75%**

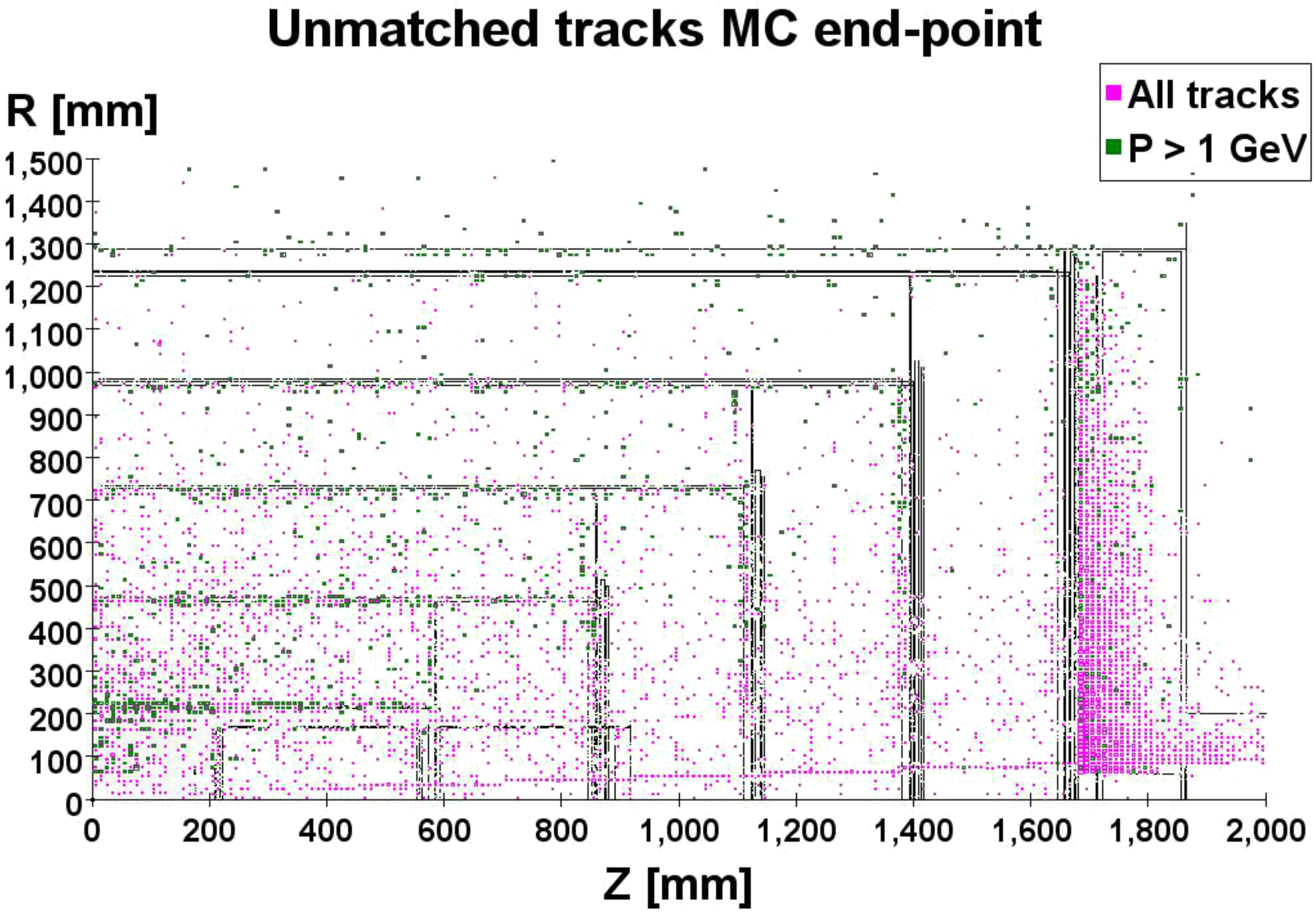
Momentum tail up to 50 GeV







Track-Seed Matching

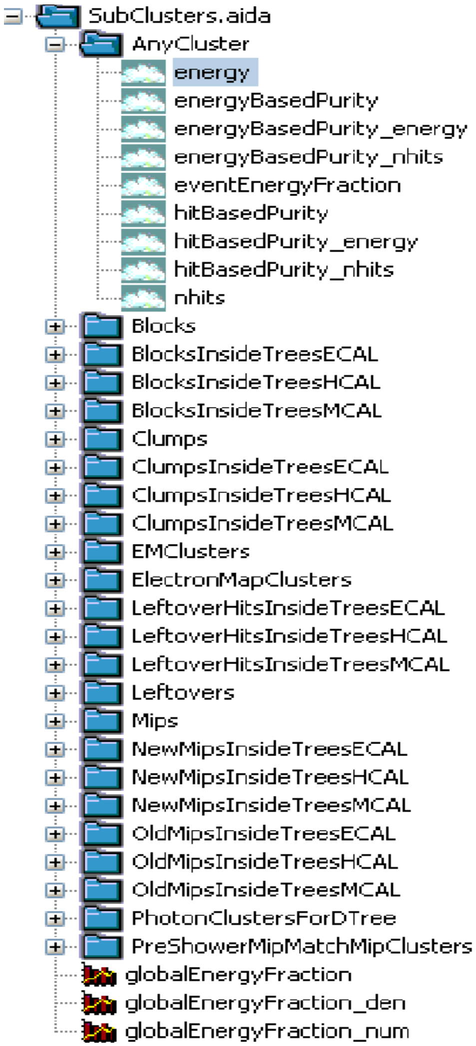


DTree Sub-Clusters: Definitions

- Plots are defined per cluster type.
- Energy fraction ignores energy from non-listed cluster types (missing muons):
  - The total energy is computed by summing up cluster energies from listed types.
- Defined a “per-event” energy fraction and a “global” energy fraction.
- Two definitions for purity:
  - Hit based purity

    # hits from dominant particle / # hits in cluster
  - Energy based purity:

    Energy from dominant particle / cluster energy



- SubClusters.aida
  - AnyCluster
    - energy
    - energyBasedPurity
    - energyBasedPurity_energy
    - energyBasedPurity_nhits
    - eventEnergyFraction
    - hitBasedPurity
    - hitBasedPurity_energy
    - hitBasedPurity_nhits
    - nhits
  - Blocks
  - BlocksInsideTreesECAL
  - BlocksInsideTreesHCAL
  - BlocksInsideTreesMCAL
  - Clumps
  - ClumpsInsideTreesECAL
  - ClumpsInsideTreesHCAL
  - ClumpsInsideTreesMCAL
  - EMClusters
  - ElectronMapClusters
  - LeftoverHitsInsideTreesECAL
  - LeftoverHitsInsideTreesHCAL
  - LeftoverHitsInsideTreesMCAL
  - Leftovers
  - Mips
  - NewMipsInsideTreesECAL
  - NewMipsInsideTreesHCAL
  - NewMipsInsideTreesMCAL
  - OldMipsInsideTreesECAL
  - OldMipsInsideTreesHCAL
  - OldMipsInsideTreesMCAL
  - PhotonClustersForDTree
  - PreShowerMipMatchMipClusters
  - globalEnergyFraction
  - globalEnergyFraction_den
  - globalEnergyFraction_num

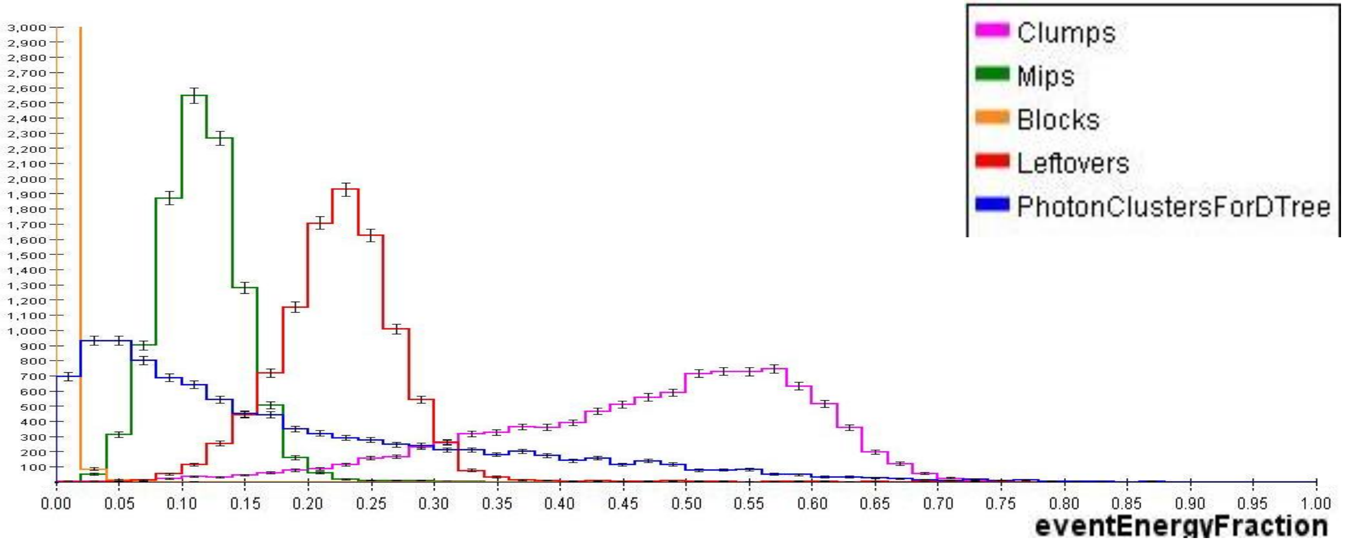
DTree Sub-Clusters: Energy contributions

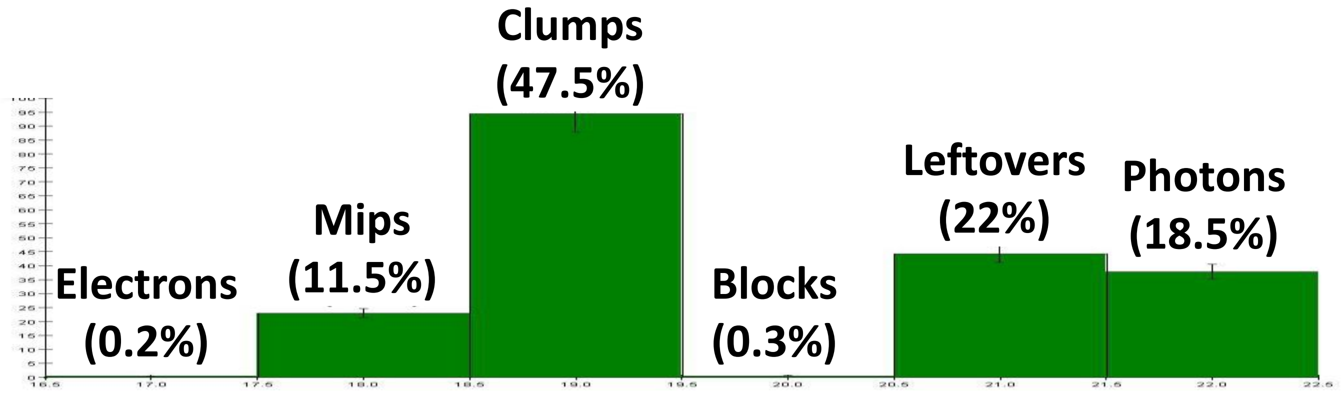
Most of the energy goes to clumps:
47.5%

Significant amount of energy is shared:
22 %

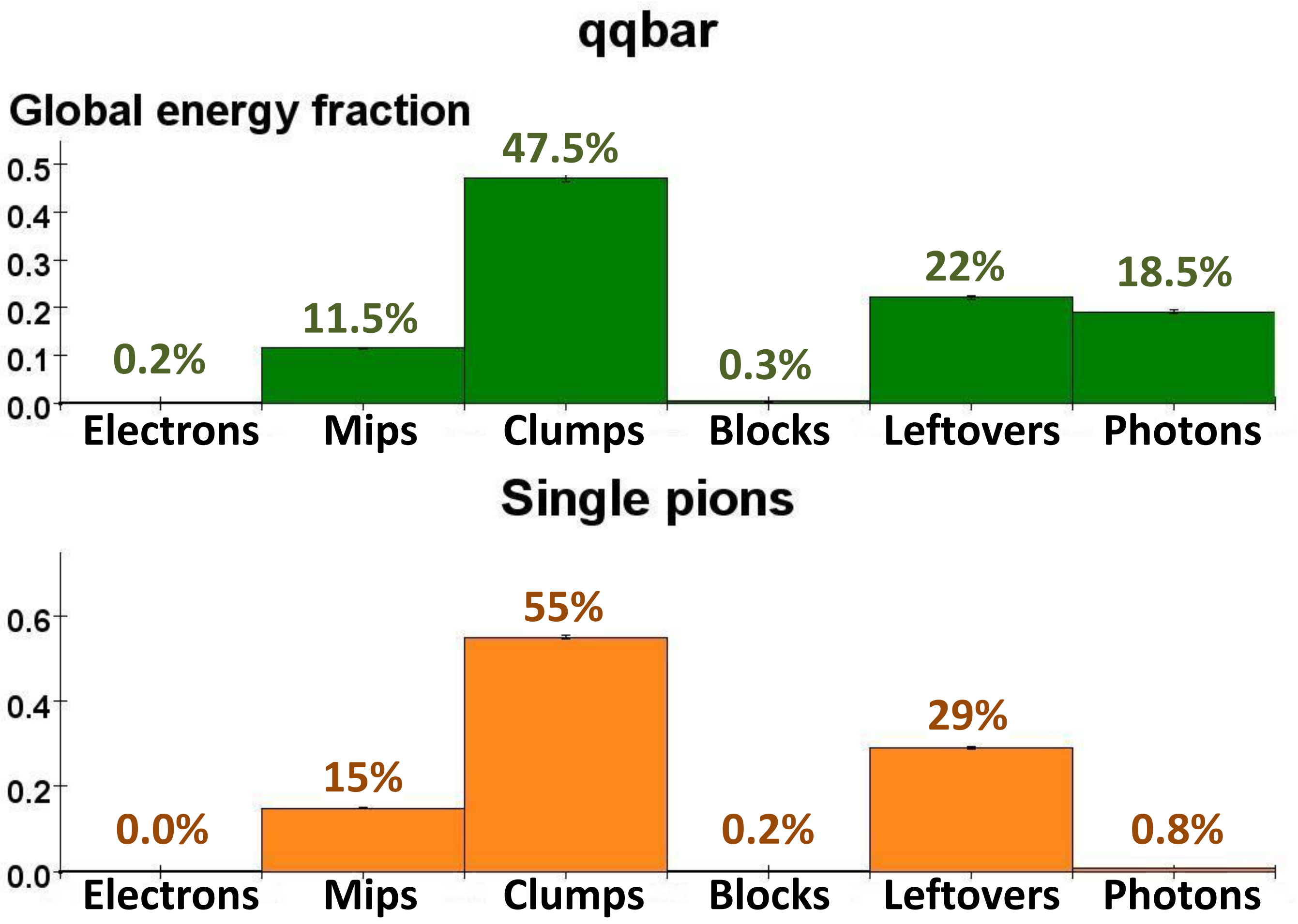
Blocks are rare:
0.3%

Photon distribution peaks at low fractions but have a large tail.





DTree Sub-Clusters: qqbar vs. Single pions

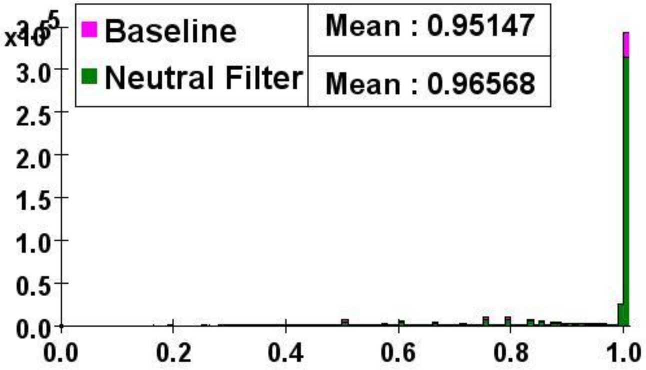


Excess in clumps in qqbar w.r.t. single pions.

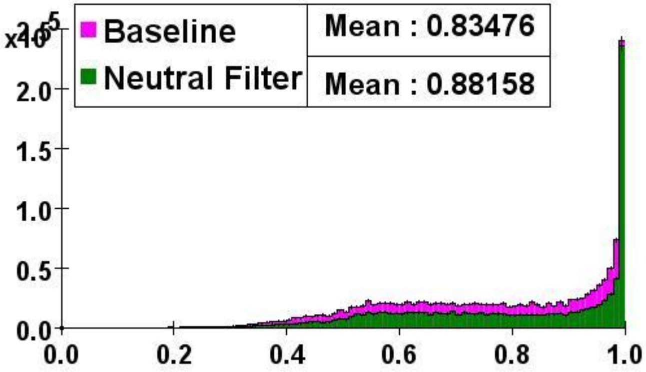
Similar distributions: no photons in single pions.

DTree Sub-Clusters

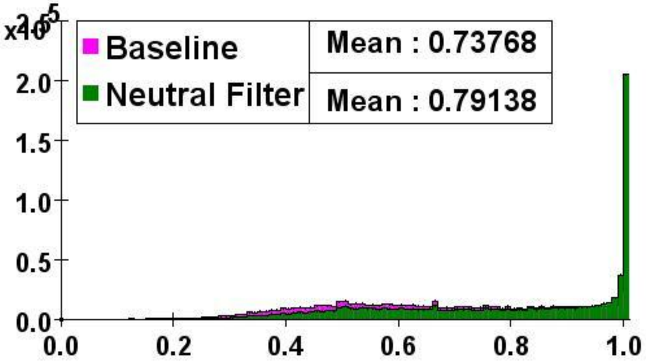
MIPs



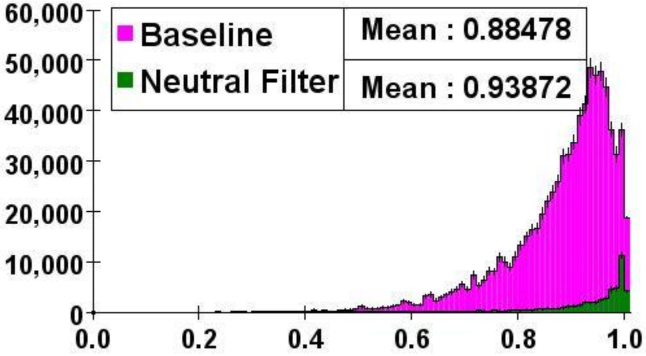
Clumps



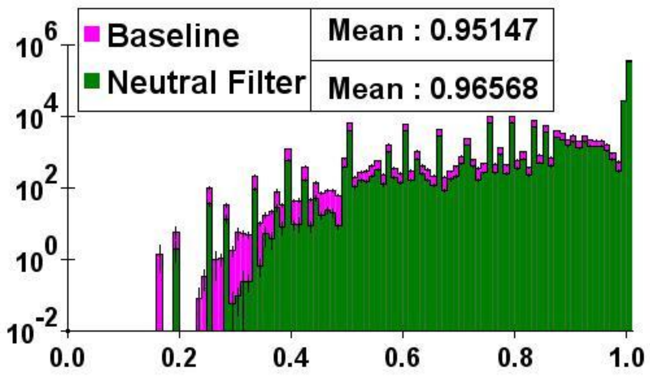
Leftovers

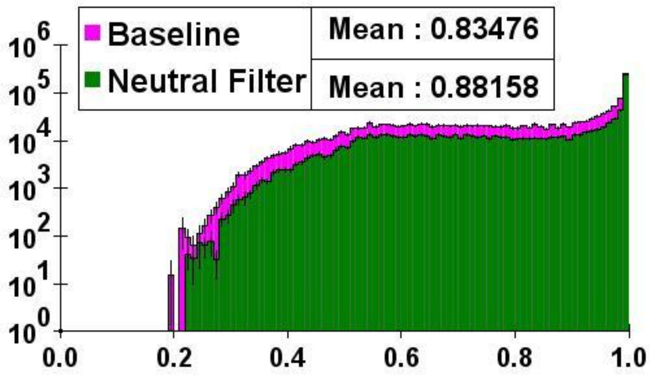


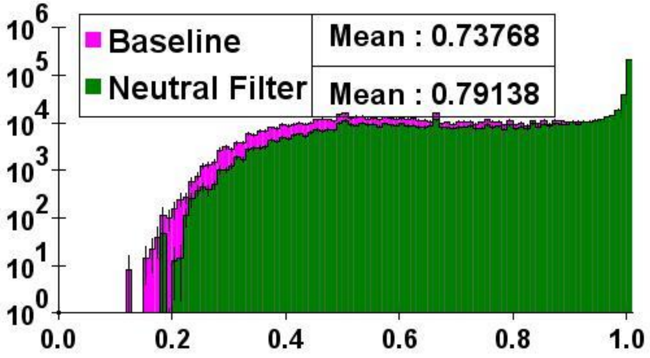
Photons

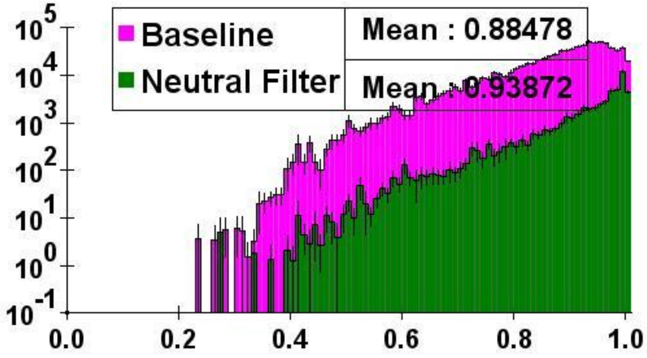


DTree Sub-Clusters


MIPs

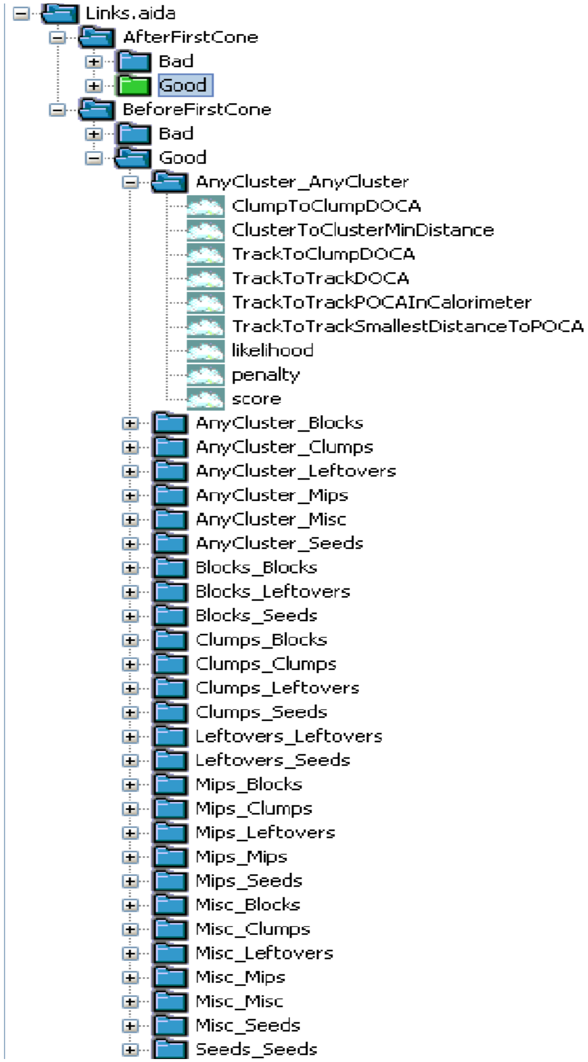

Clumps


Leftovers


Photons

Link Scoring: Definitions

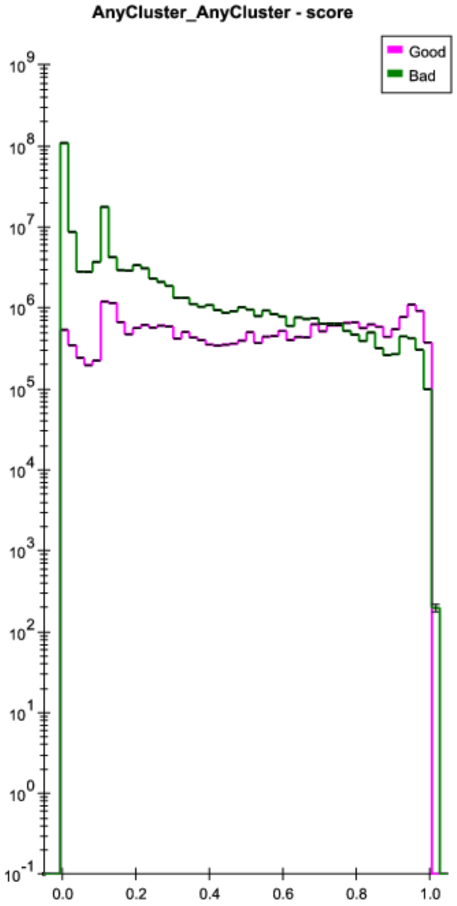
- Plots made per link type.
- A “good” link is a link where both sub-clusters have dominant energy contributions from the same MC particle.
- Made plots for variables used in likelihood:
  - To be compared with the plots in the Data-base.
- Some penalty factors are applied during scoring:
  - Penalty for belonging to separate DTree clusters: 0.8*cos(angle)
  - Penalty for proximity (not applied for mip-mip links): a/R^2
  - Other penalties depending on link type.
  - penalty = score / likelihood
- Clump-Clump likelihood is not used !?!
  - Score only computed based on angle and proximity + other types of penalties.

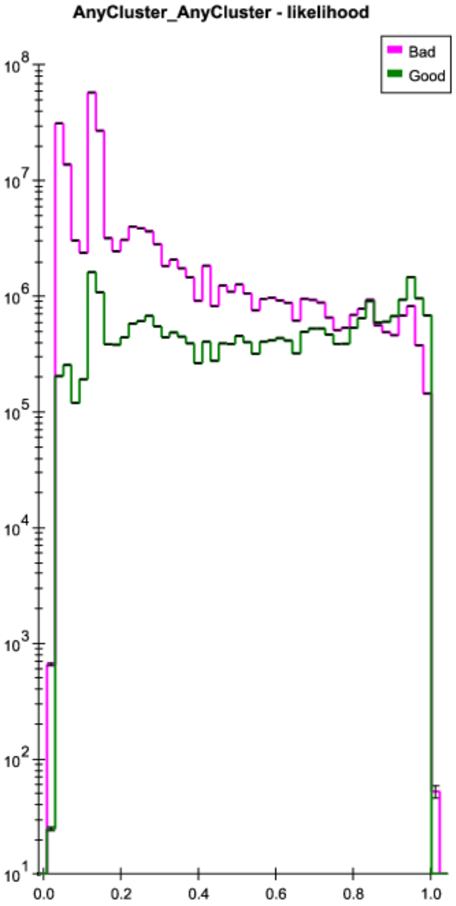


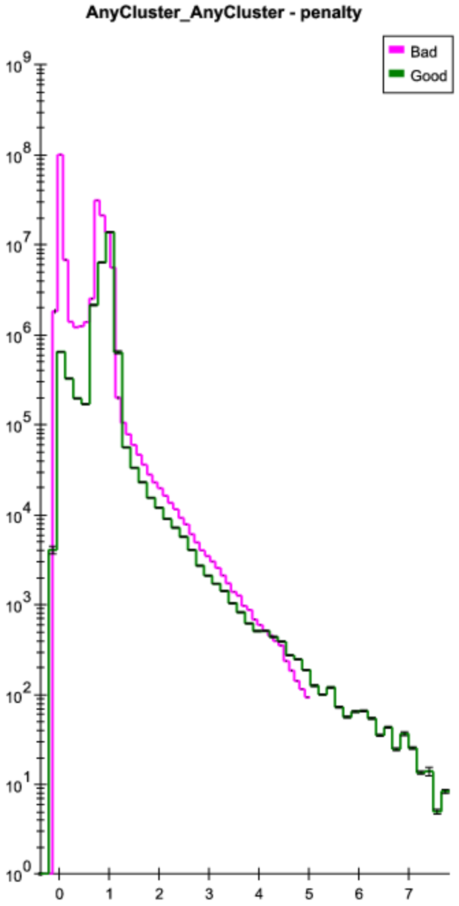
- Links.aida
  - AfterFirstCone
    - Bad
    - Good
  - BeforeFirstCone
    - Bad
    - Good
      - AnyCluster_AnyCluster
        - ClumpToClumpDOCA
        - ClusterToClusterMinDistance
        - TrackToClumpDOCA
        - TrackToTrackDOCA
        - TrackToTrackPOCAInCalorimeter
        - TrackToTrackSmallestDistanceToPOCA
        - likelihood
        - penalty
        - score
      - AnyCluster_Blocks
      - AnyCluster_Clumps
      - AnyCluster_Leftovers
      - AnyCluster_Mips
      - AnyCluster_Misc
      - AnyCluster_Seeds
      - Blocks_Blocks
      - Blocks_Leftovers
      - Blocks_Seeds
      - Clumps_Blocks
      - Clumps_Clumps
      - Clumps_Leftovers
      - Clumps_Seeds
      - Leftovers_Leftovers
      - Leftovers_Seeds
      - Mips_Blocks
      - Mips_Clumps
      - Mips_Leftovers
      - Mips_Mips
      - Mips_Seeds
      - Misc_Blocks
      - Misc_Clumps
      - Misc_Leftovers
      - Misc_Mips
      - Misc_Misc
      - Misc_Seeds
      - Seeds_Seeds

Link Scoring: Scores
All links

Scoring can be improved!

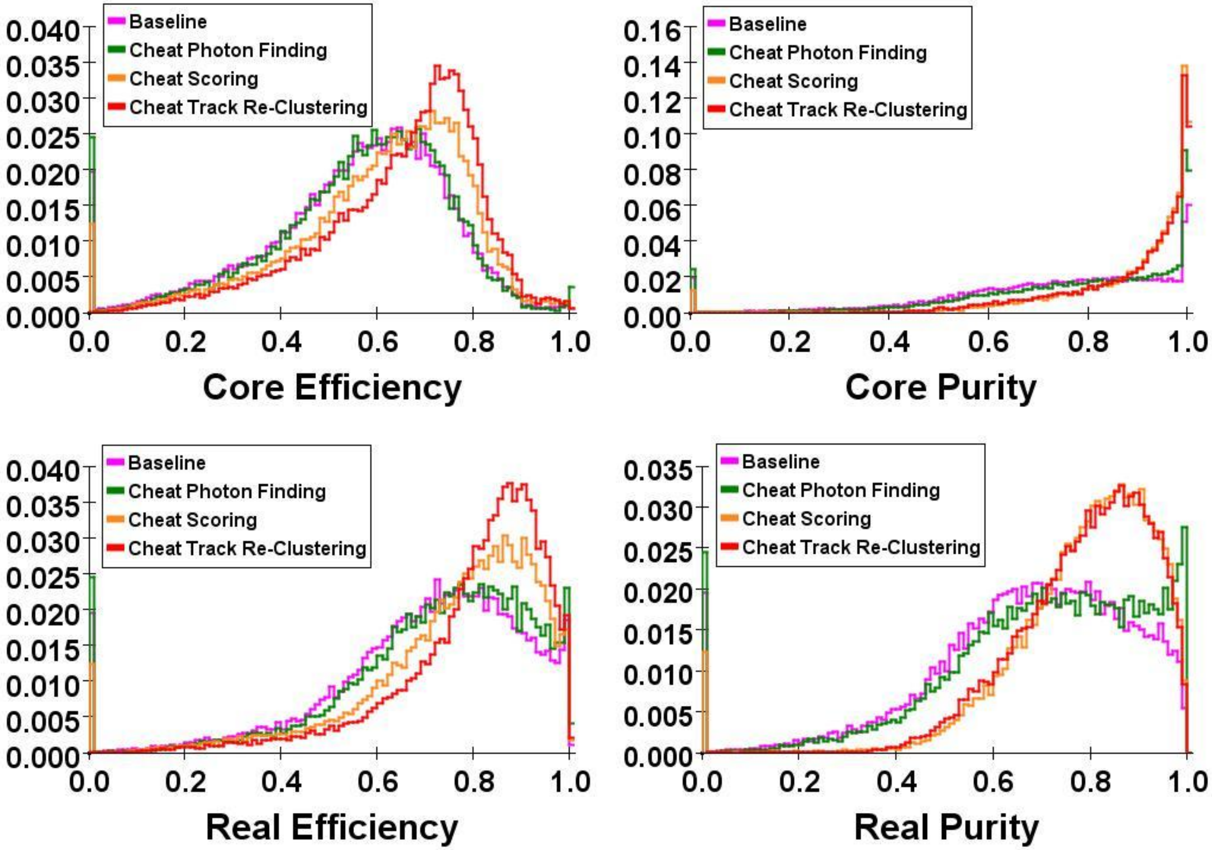






Shower Efficiency and Purity

PFA	A	B	C	D
Core (excluding shared)				
Eff	55	56	60	65
Pur	74	77	89	89
Real (including shared)				
Eff	70	72	76	80
Pur	68	70	79	79



Shower Efficiency and Purity

PFA	A	C	D
Core (excluding shared)			
Eff	61	63	66
Pur	85	92	92
Real (including shared)			
Eff	80	81	84
Pur	79	83	83