Cluster Finding Comparisons

Ron Cassell

SLAC

Clustering Studies

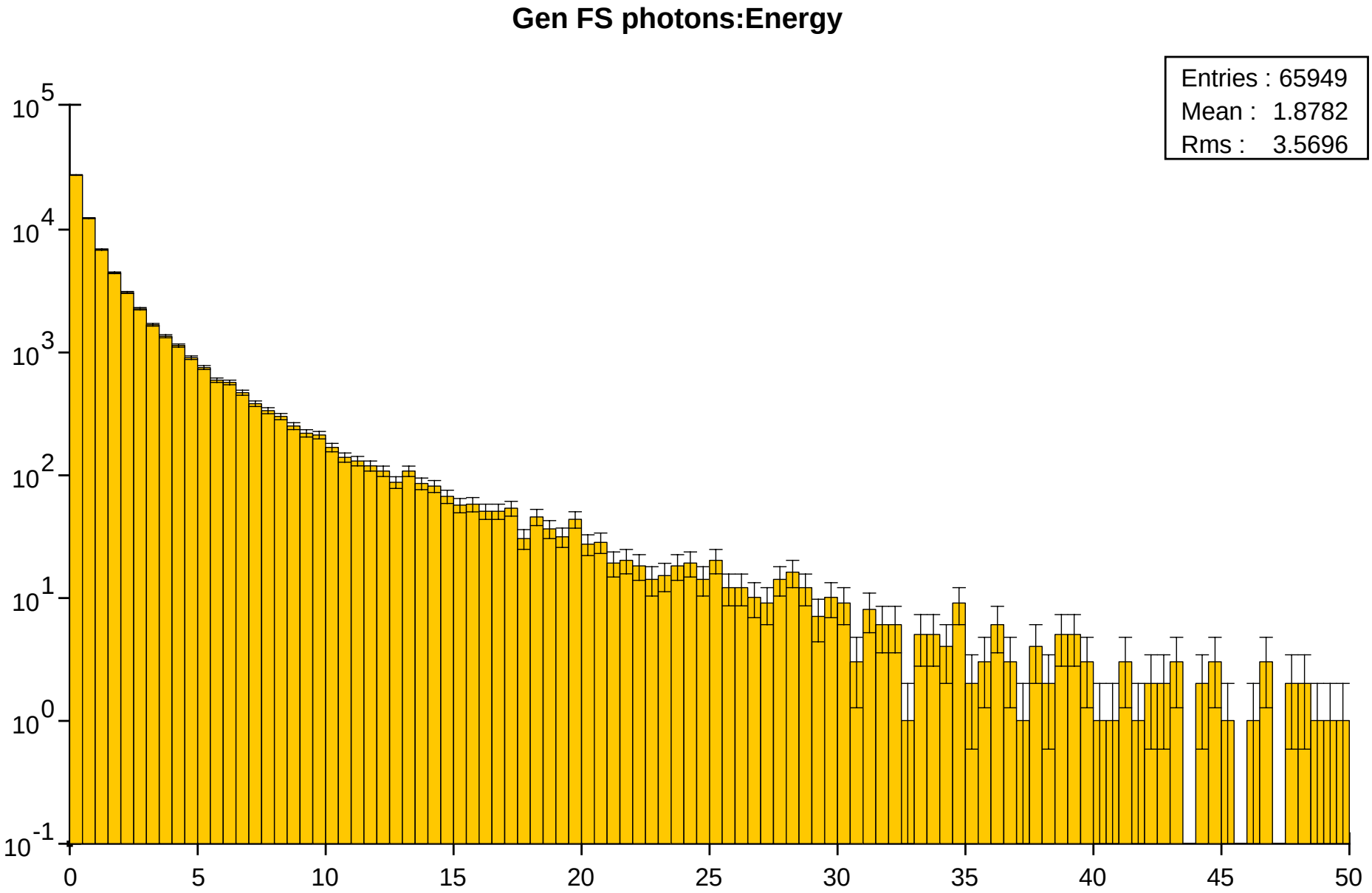
- This report studies clustering in the EM calorimeter, using SLIC simulated ttbar events at 500 GeV in the acme0605 detector
- The focus is on finding and identifying photons. Hadronic studies more difficult, not done yet.
- Attempt is to characterize performance of different clustering algorithms, and identify problem areas in photon finding

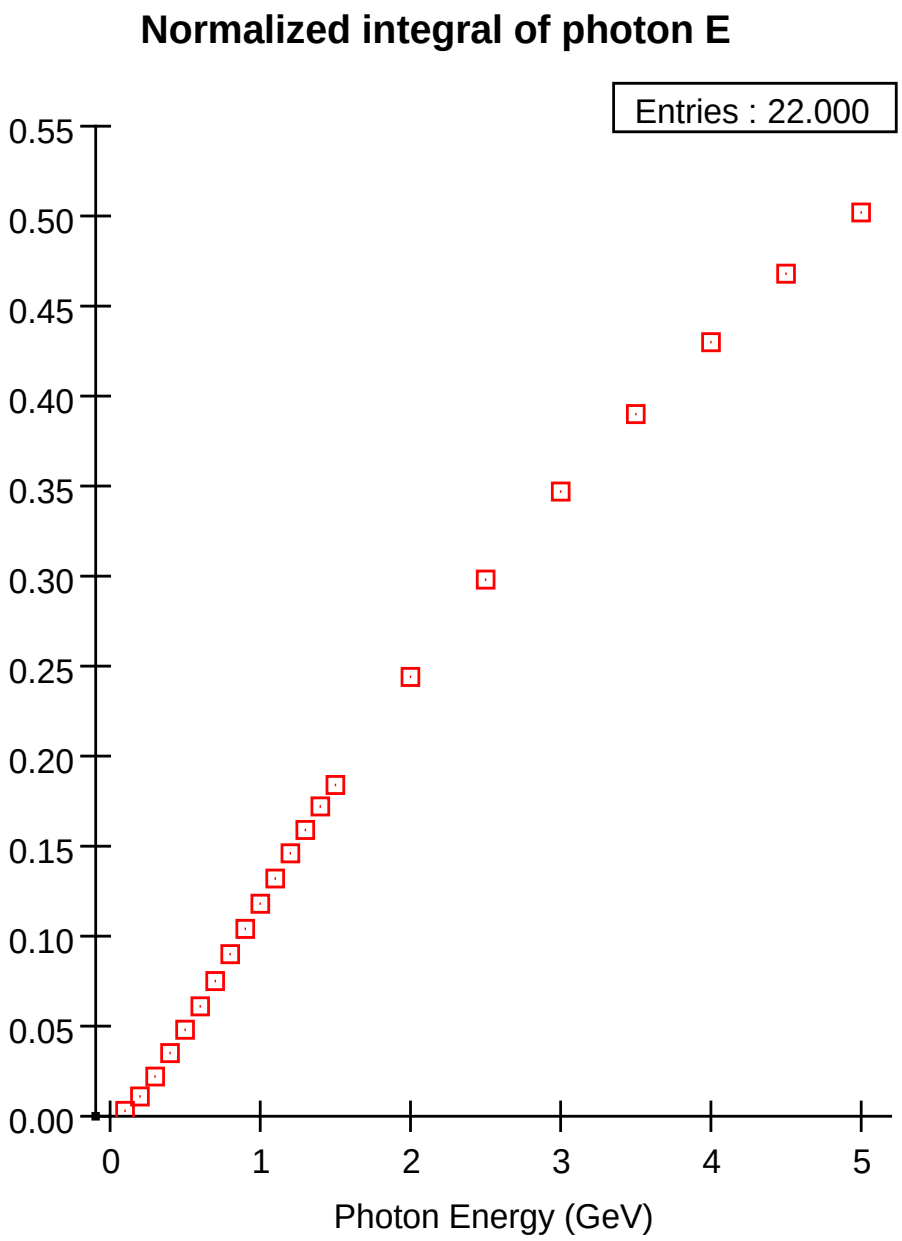
Clusterers available in org.lcsim

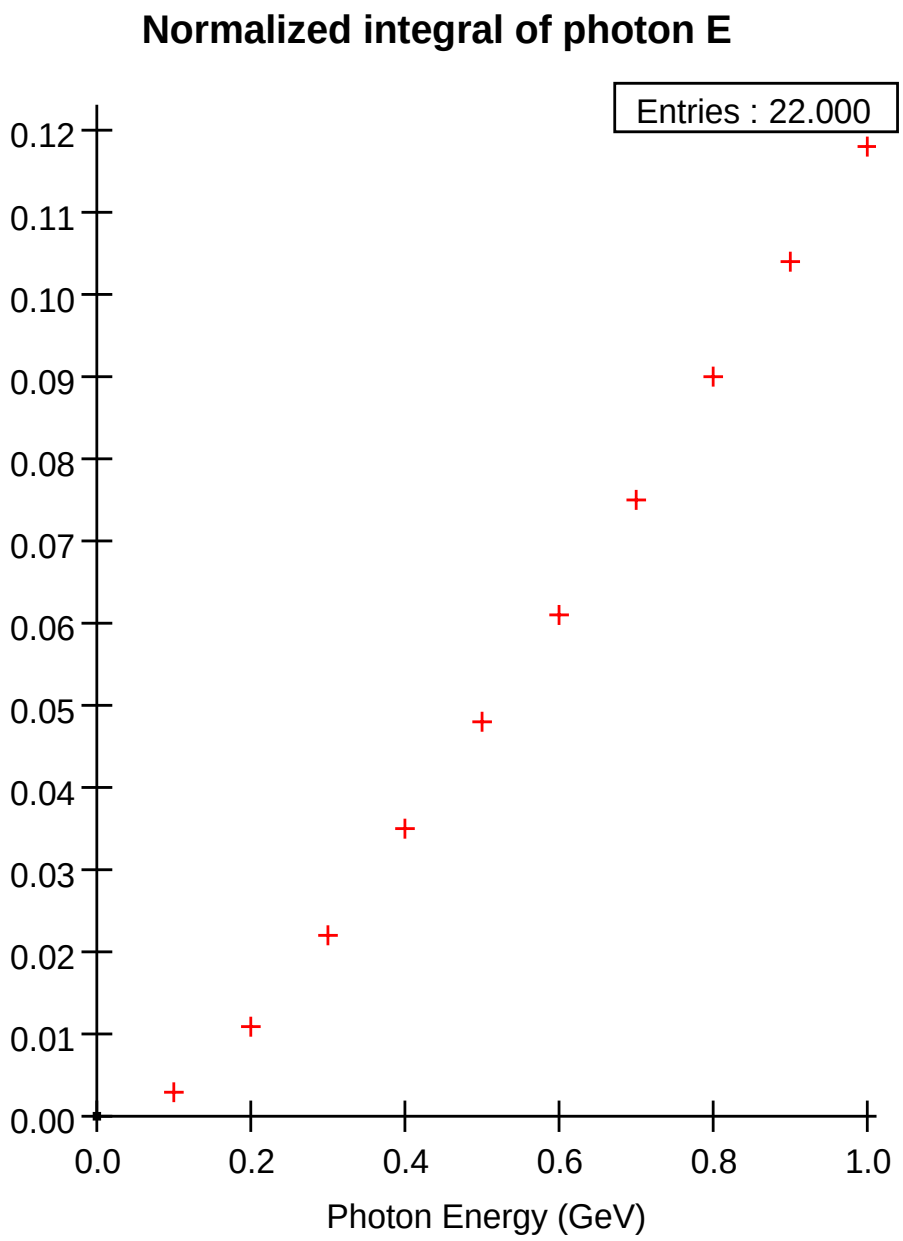
- Cheater
- NearestNeighbor – combines adjacent cells into clusters. Definition of adjacent can be set independently in 3 directions: u,v,layer
- MST – Minimum spanning tree. Default metric is distance, so should be very close to NN with appropriate parameters. Only studied with default parameters in this report
- Fixed Cone – should be fast but takes more than an order of magnitude more cpu than other clusterers. Problem being studied, but excluded from this report
- Directed Tree – uses local hit densities to do clustering. Still contains a memory leak, so # of events limited

Photons in ttbar events at 500 GeV

- The amount of information available to separate photons from hadrons decreases with energy
- Some initial cluster energy and #hits cut will be necessary
- Look at photon energy distributions to see what is being cut

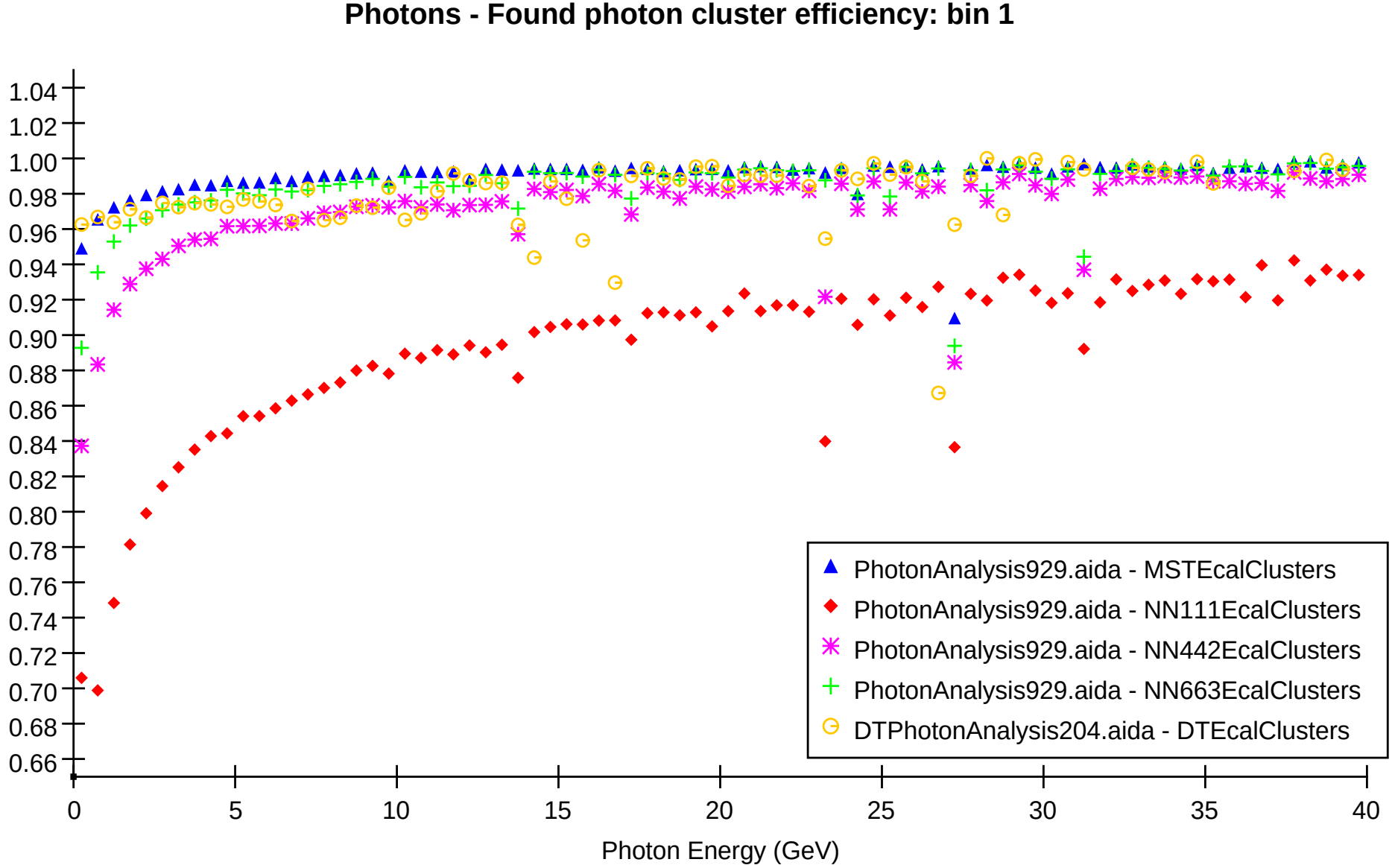

Gen FS photons:Energy
Entries : 65949
Mean : 1.8782
Rms : 3.5696
10^5
10^4
10^3
10^2
10^1
10^0
10^-1
0
5
10
15
20
25
30
35
40
45
50

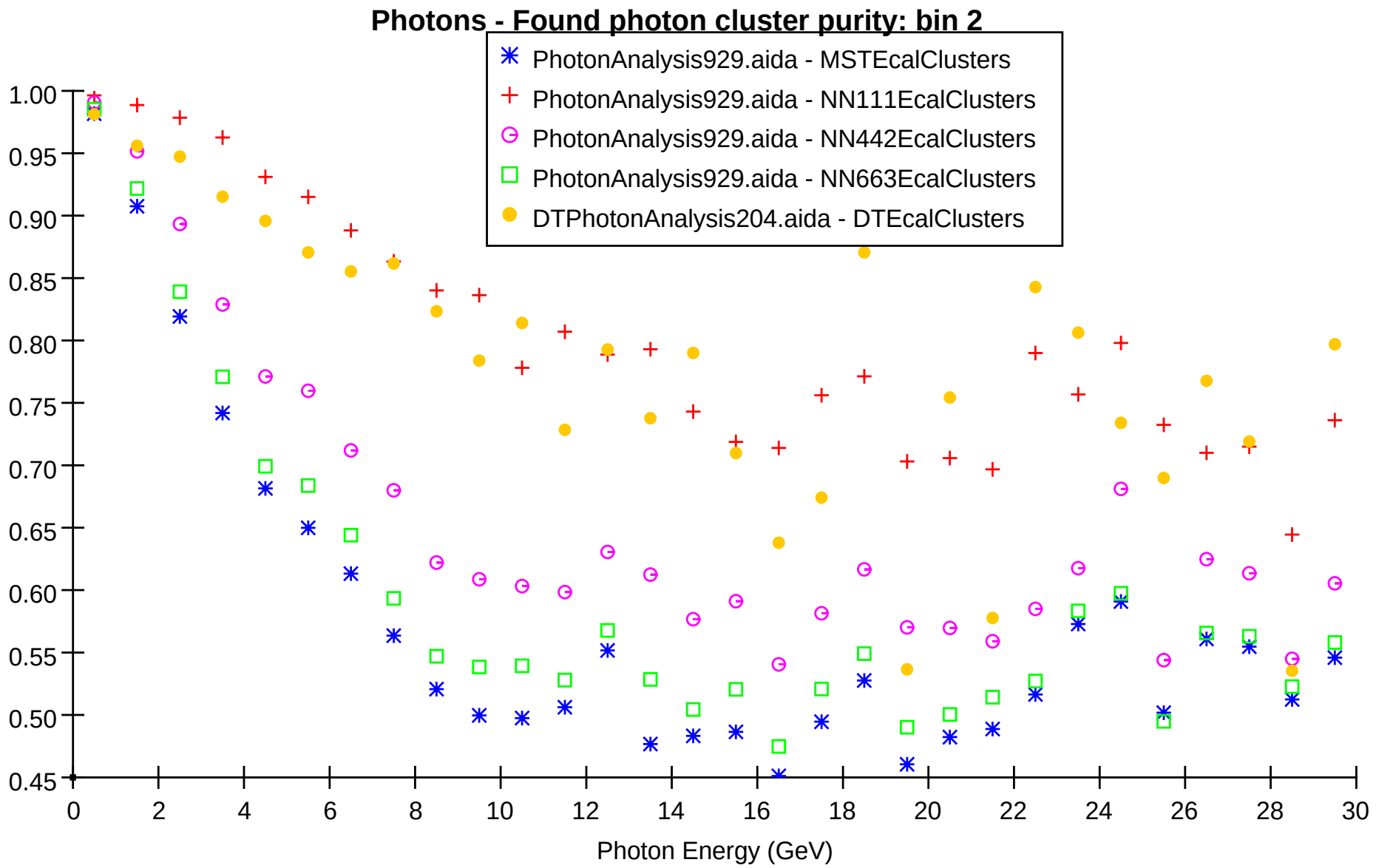

Normalized integral of photon E
Entries : 22.000
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
Photon Energy (GeV)


Normalized integral of photon E
Entries : 22.000
0.12
0.11
0.10
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.2
0.4
0.6
0.8
1.0
Photon Energy (GeV)

Analysis definitions

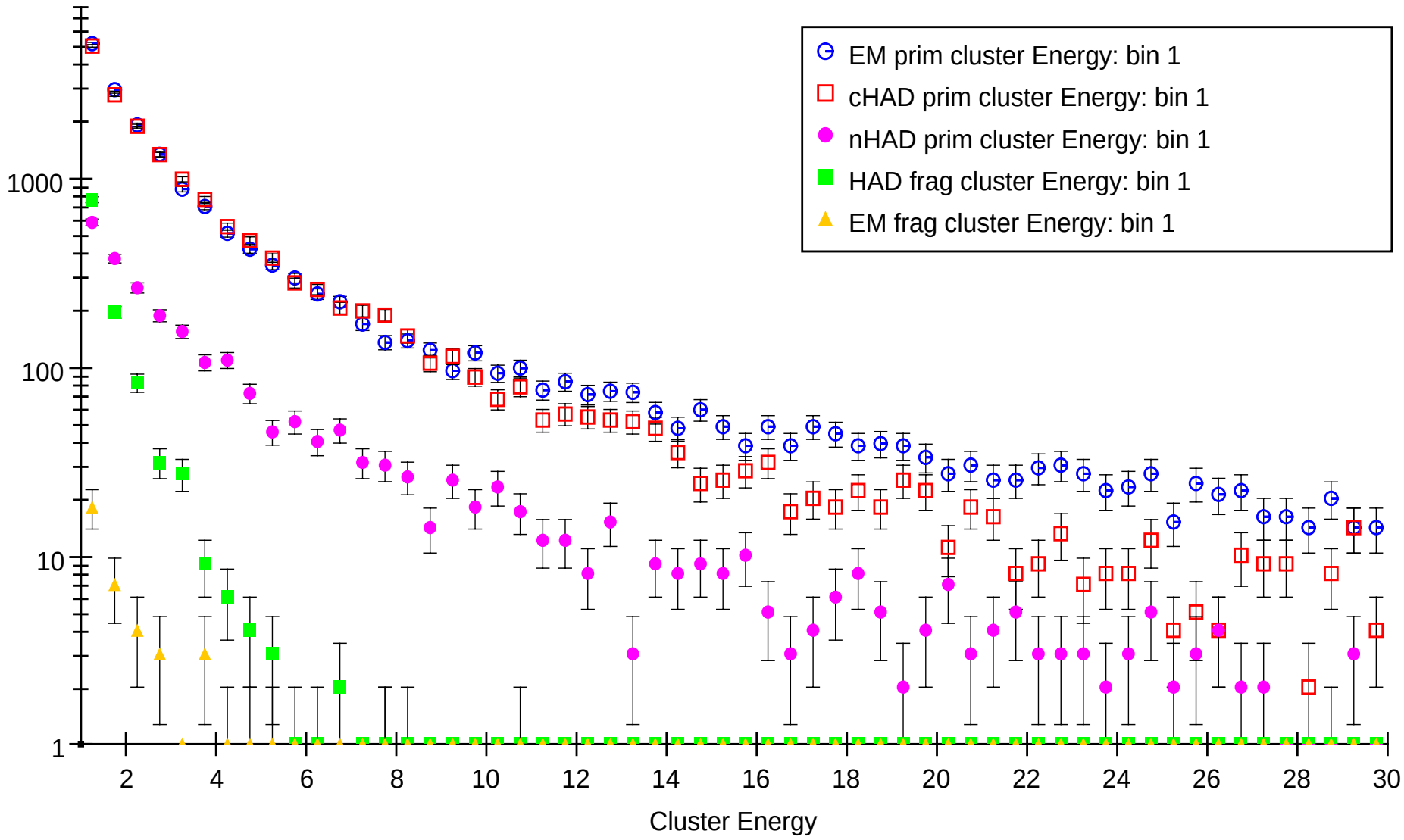
- For each cluster, the particle association is determined by the maximum energy contribution to that cluster.
- Each particle may contribute to multiple clusters. The max cluster for each particle is simply the cluster to which that particle contributes the most energy.
- For each cluster, if it is the max cluster for the particle associated with it, it is a primary. Otherwise a fragment.
- Each cluster is then put in 1 of 5 categories:
  EM primary, neutral Hadron primary, charged Hadron primary, EM fragment, or Hadron fragment
- For primaries: Cluster efficiency = (Particle contribution to cluster)/(Total energy deposited by particle in calorimeter). Cluster purity = (Particle contribution to cluster)/(cluster energy)

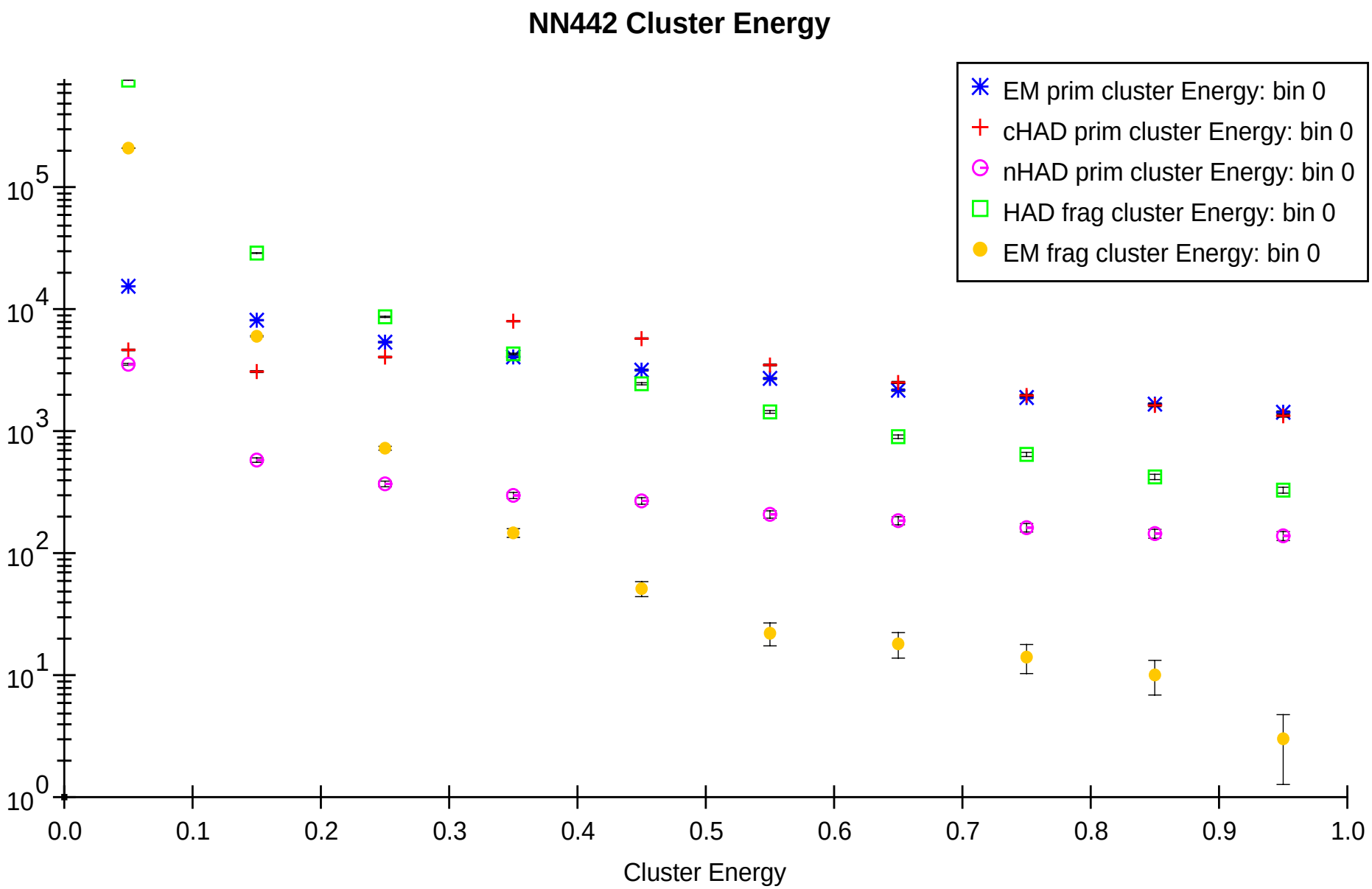

Photons - Found photon cluster efficiency: bin 1
1.04
1.02
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0.66
0
5
10
15
20
25
30
35
40
Photon Energy (GeV)
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters

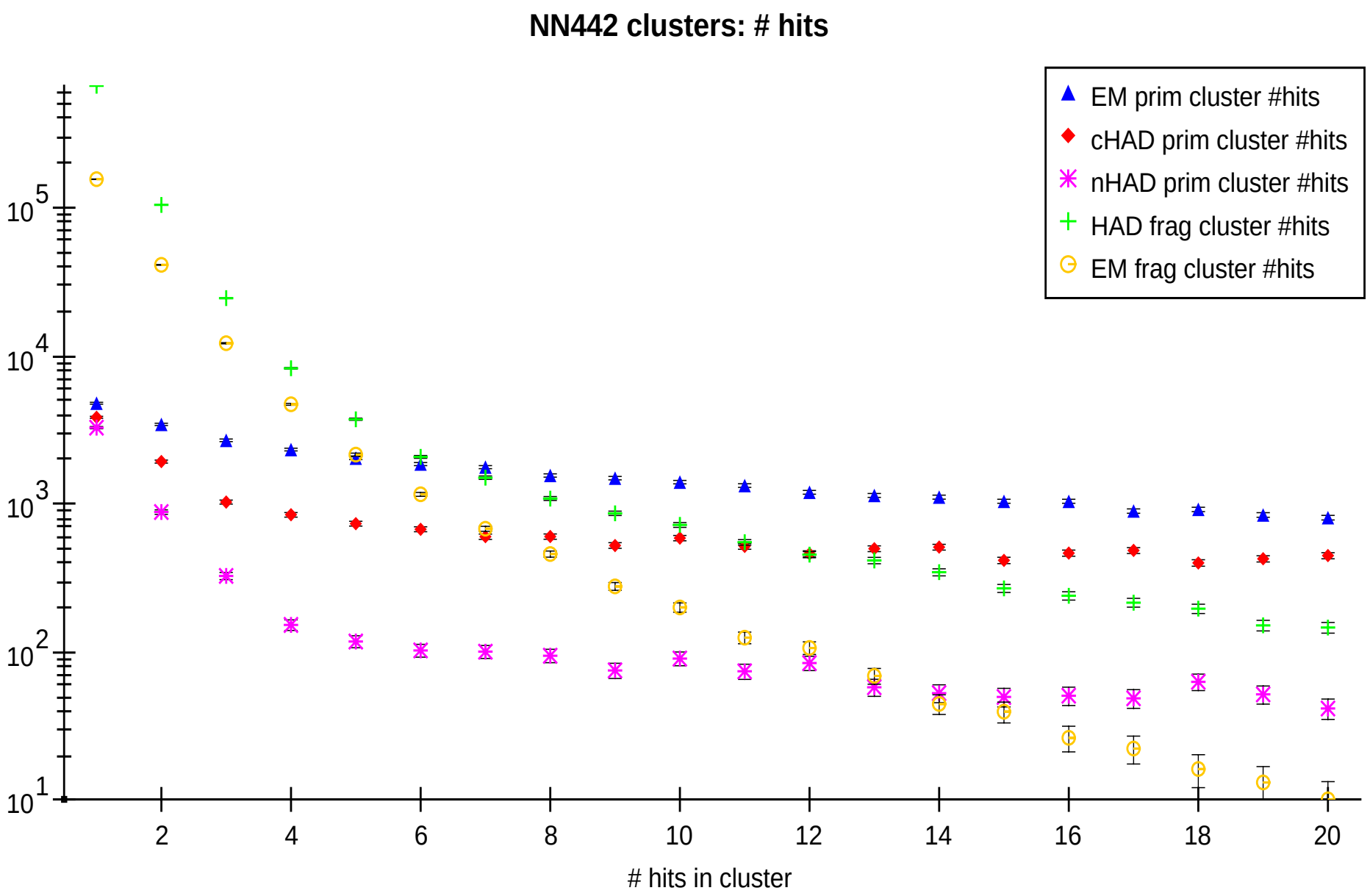

Photons - Found photon cluster purity: bin 2
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Photon Energy (GeV)

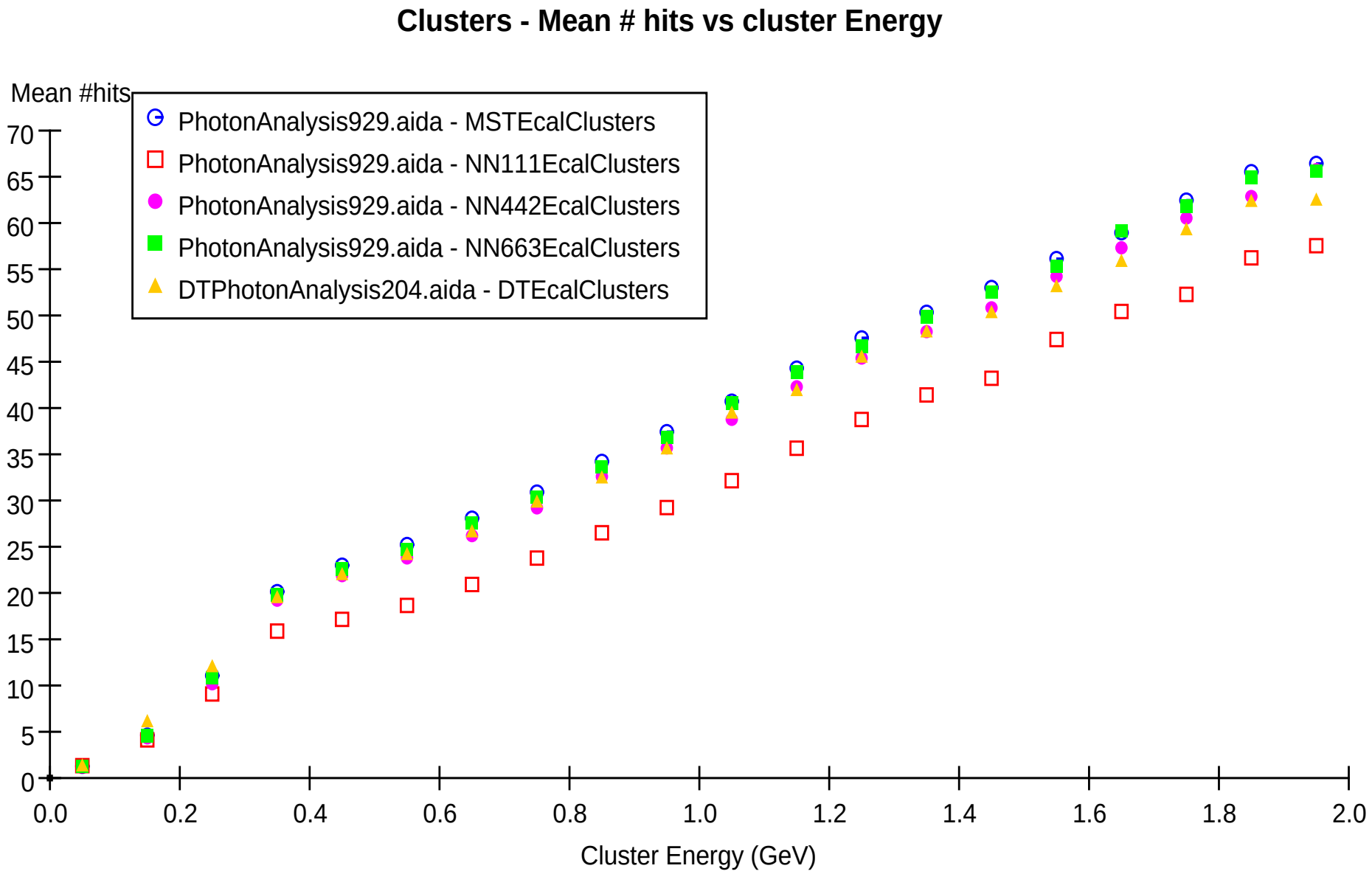
Cluster Identification

- Problem of separating photon clusters from other clusters varies with energy
- Above 5 GeV, about equal number of photon clusters and charged hadron clusters, ~ 10% neutral hadron clusters, and virtually no fragments
- Below 1 GeV, hadronic fragments are significant, and dominate below .3 GeV


NN442 cluster Energy
EM prim cluster Energy: bin 1
cHAD prim cluster Energy: bin 1
nHAD prim cluster Energy: bin 1
HAD frag cluster Energy: bin 1
EM frag cluster Energy: bin 1
1000
100
10
1
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
Cluster Energy

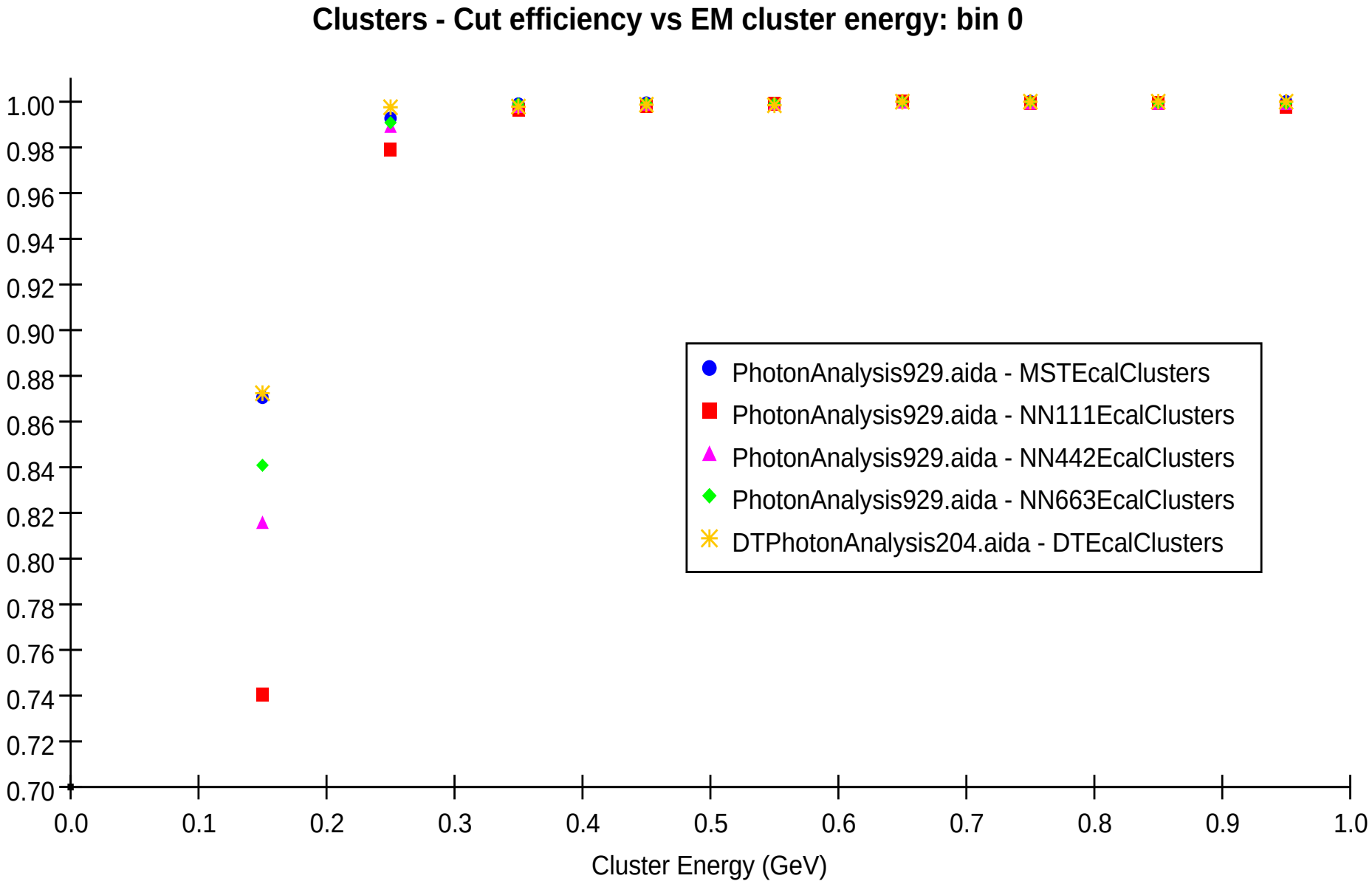

NN442 Cluster Energy
EM prim cluster Energy: bin 0
cHAD prim cluster Energy: bin 0
nHAD prim cluster Energy: bin 0
HAD frag cluster Energy: bin 0
EM frag cluster Energy: bin 0
10^5
10^4
10^3
10^2
10^1
10^0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Cluster Energy


NN442 clusters: # hits
EM prim cluster #hits
cHAD prim cluster #hits
nHAD prim cluster #hits
HAD frag cluster #hits
EM frag cluster #hits
10^5
10^4
10^3
10^2
10^1
2
4
6
8
10
12
14
16
18
20
hits in cluster


Clusters - Mean # hits vs cluster Energy
Mean #hits
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
Cluster Energy (GeV)

Loss accounting and cuts

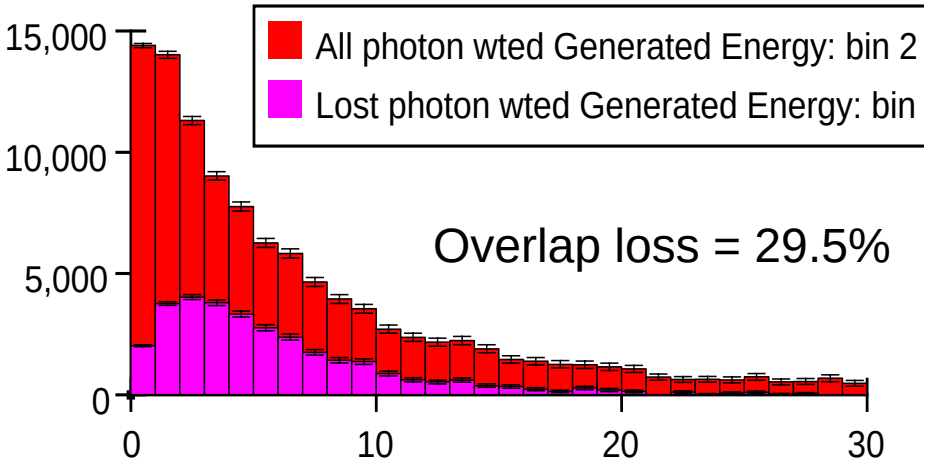
- 1.67% of photon energy lost to “invisible” photons(beam pipe, conversions beyond the EM calorimeter, low energy with no EM hits)
- Require #hits > 5, and E > .1 GeV
- Leads to additional .5-.7% loss


Clusters - Cut efficiency vs EM cluster energy: bin 0
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0.74
0.72
0.70
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Cluster Energy (GeV)
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters

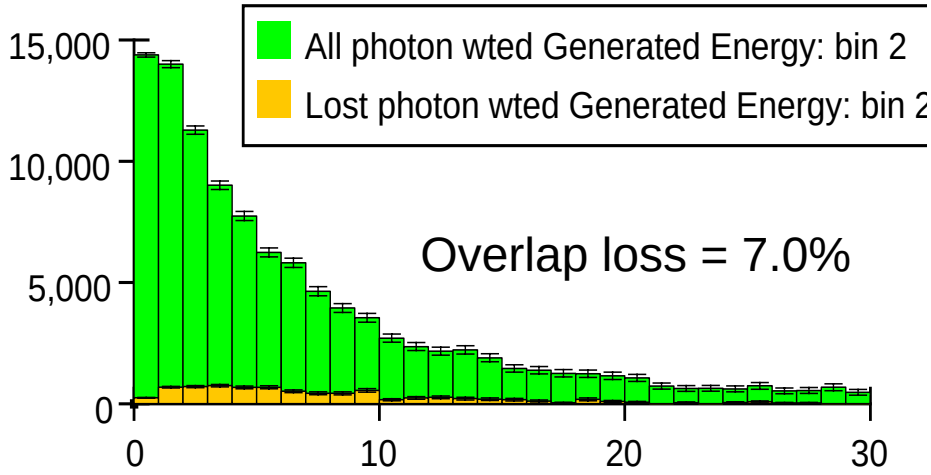
More losses

- If a particle is not the max contributor to its max cluster, declare the particle lost to overlap.
- Look at overlap losses

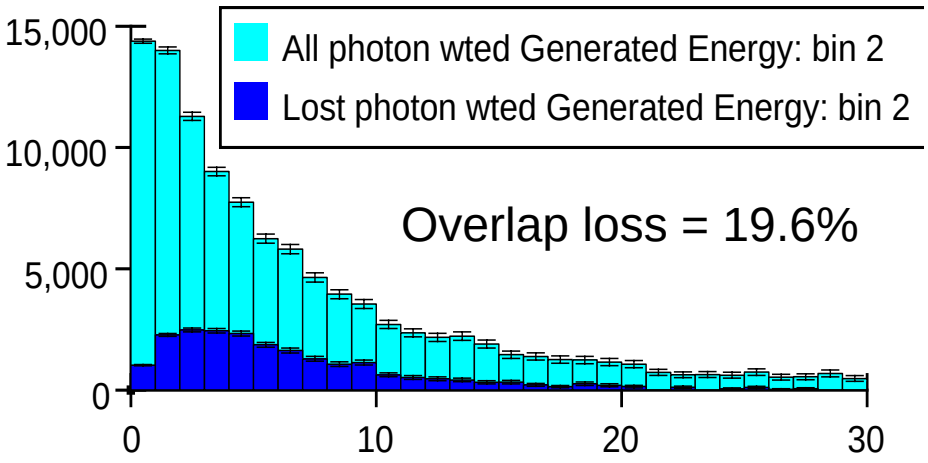
PhotonAnalysis929.aida - MSTEcalClusters - Photons



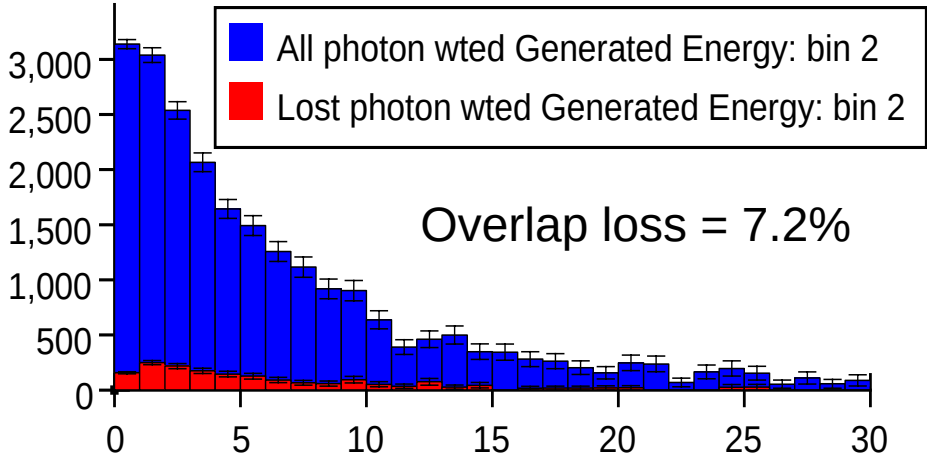
PhotonAnalysis929.aida - NN111EcalClusters - Photons



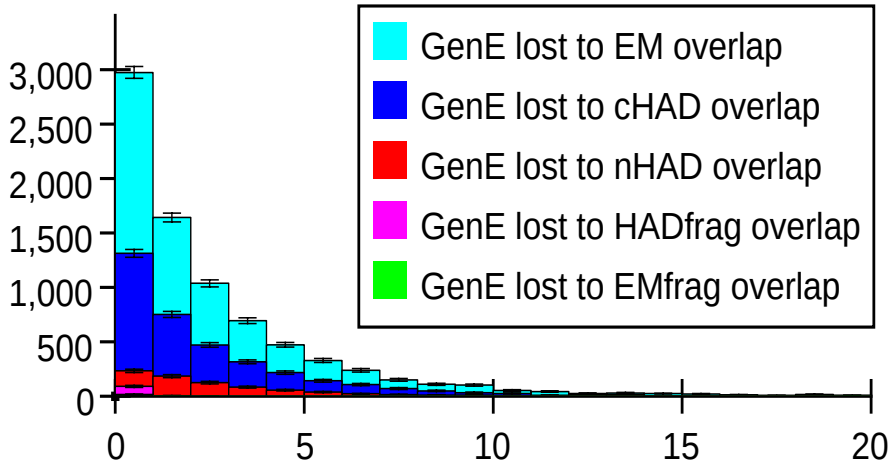
PhotonAnalysis929.aida - NN442EcalClusters - Photons



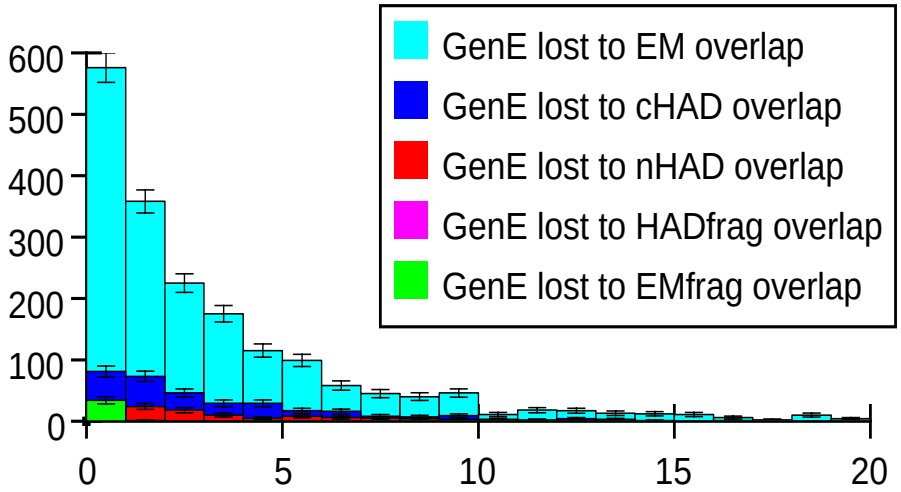
DTPhotonAnalysis204.aida - DTEcalClusters - Photons



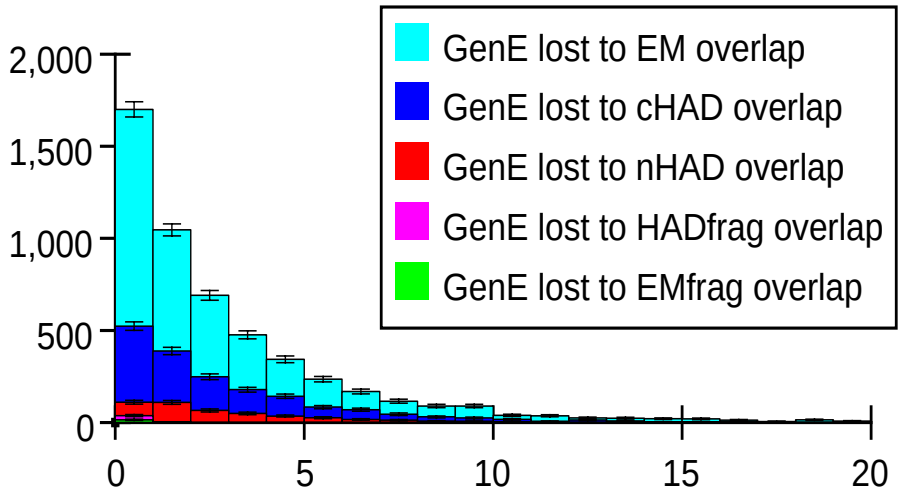
PhotonAnalysis929.aida - MSTEcalClusters - ...



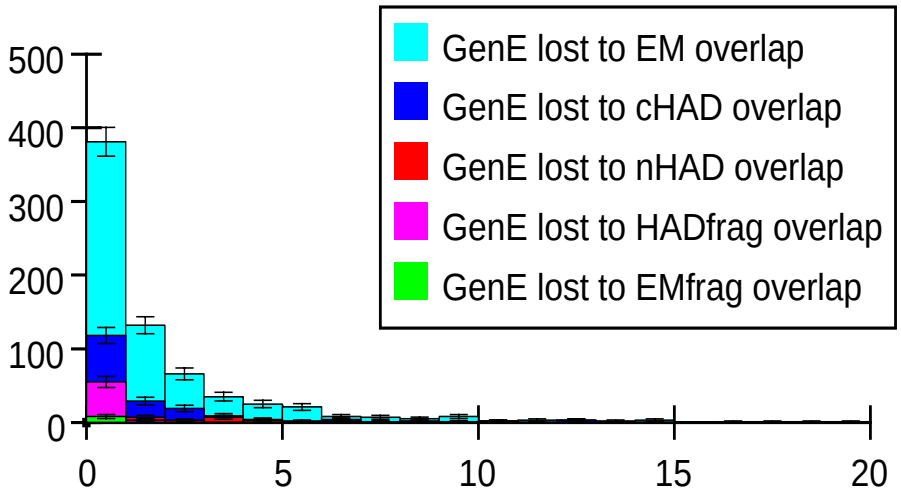
PhotonAnalysis929.aida - NN111EcalClusters - Photons



PhotonAnalysis929.aida - NN442EcalClusters - Photons

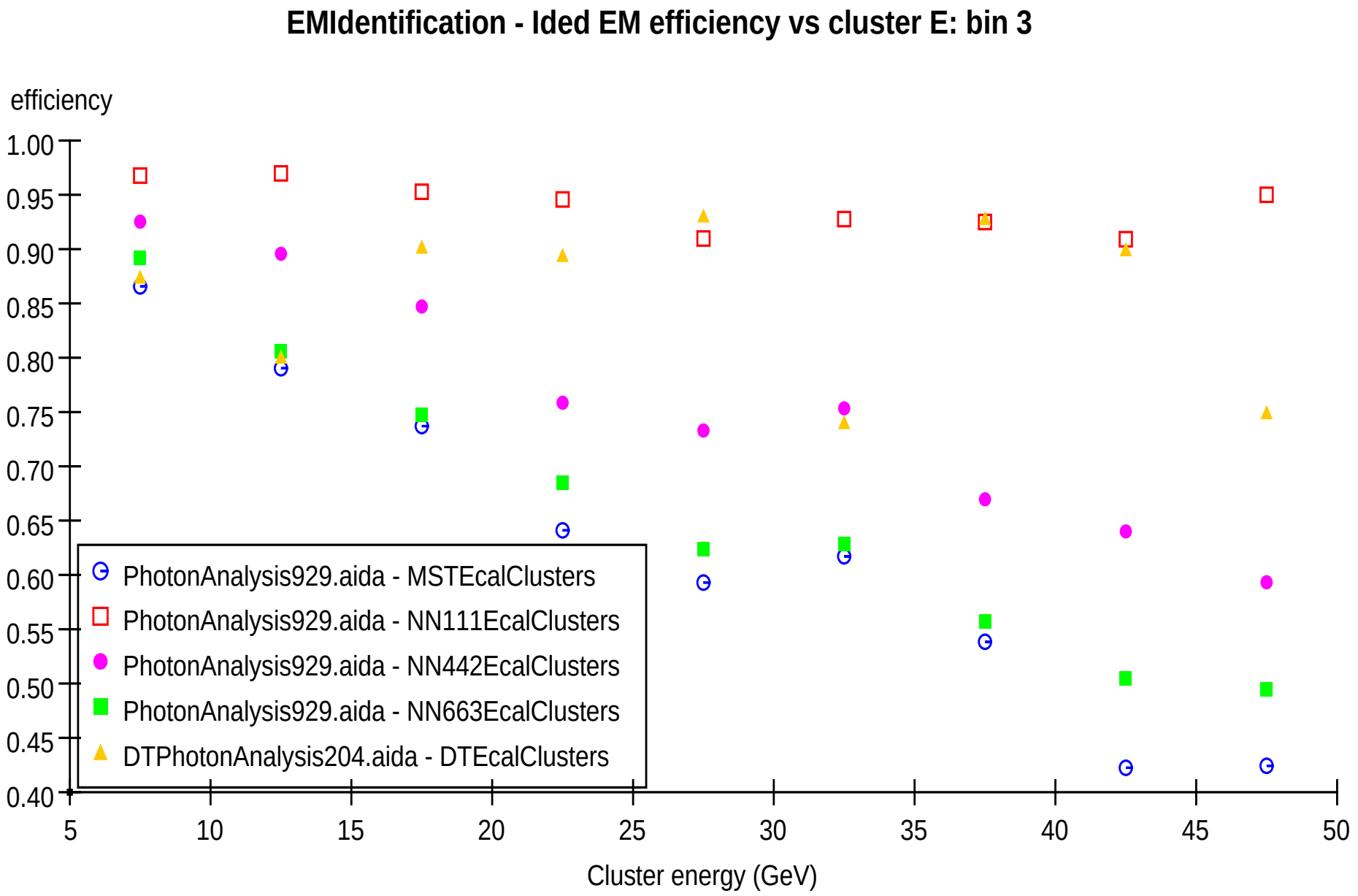


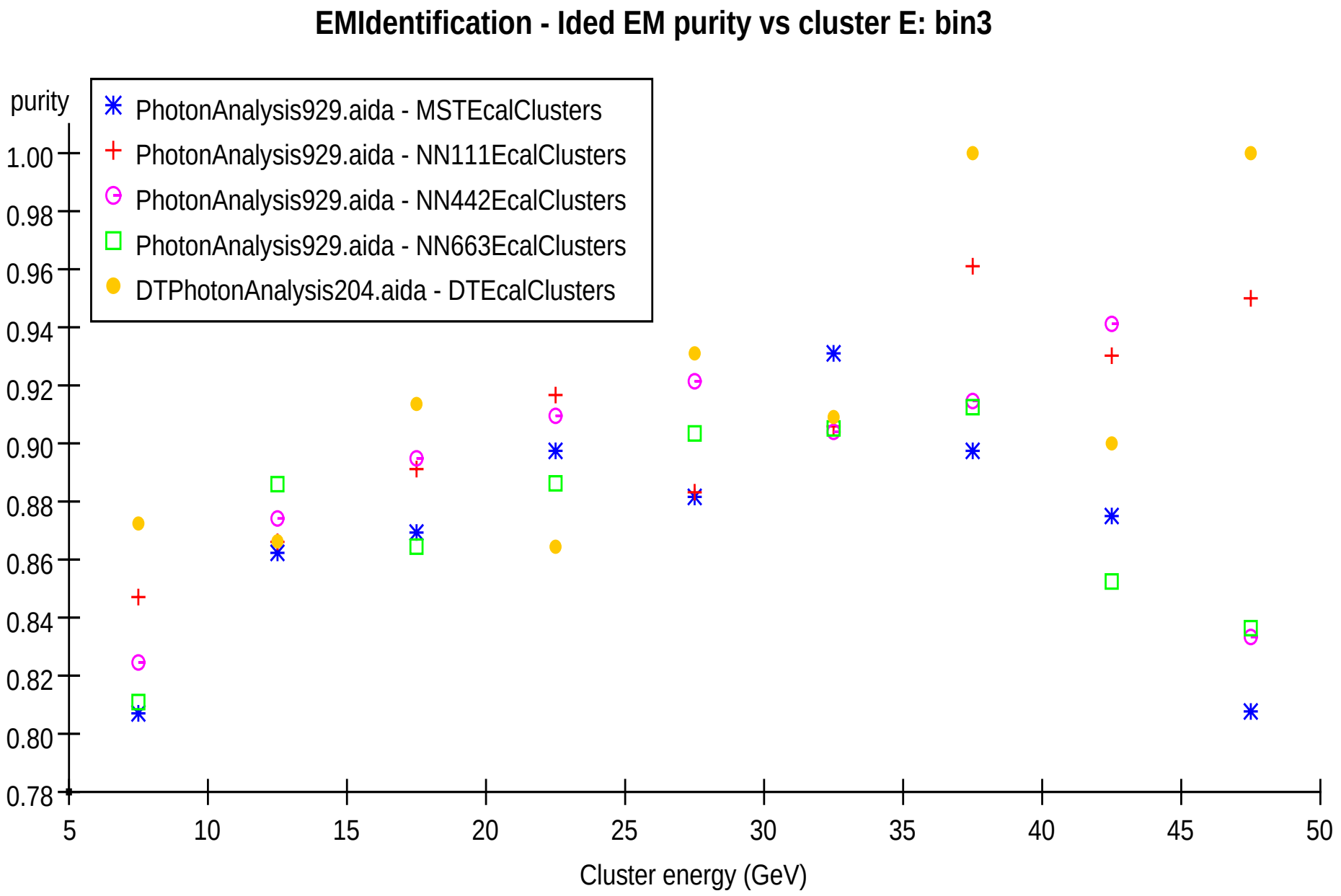
DTPhotonAnalysis204.aida - DTEcalClusters - Photons

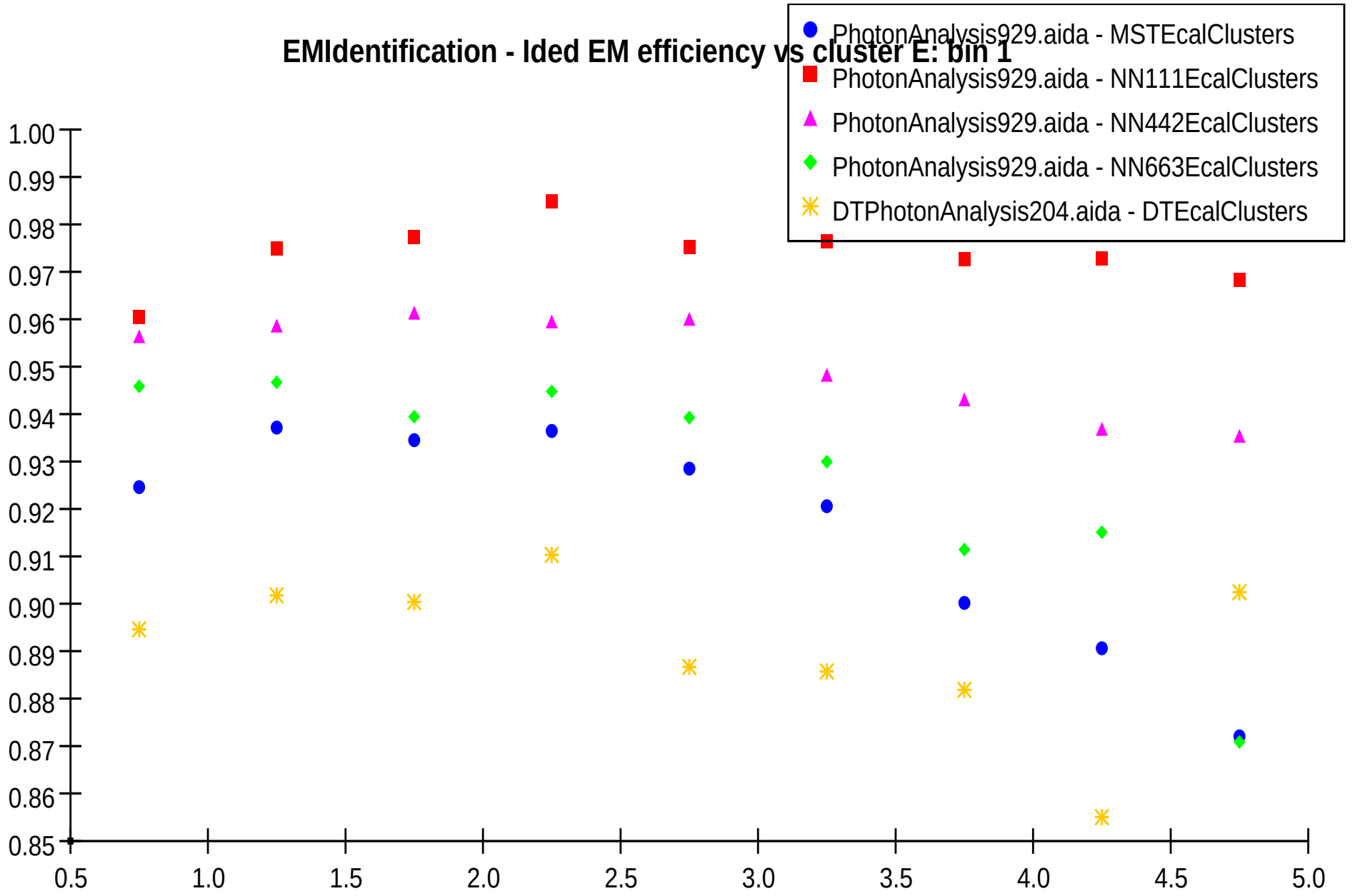


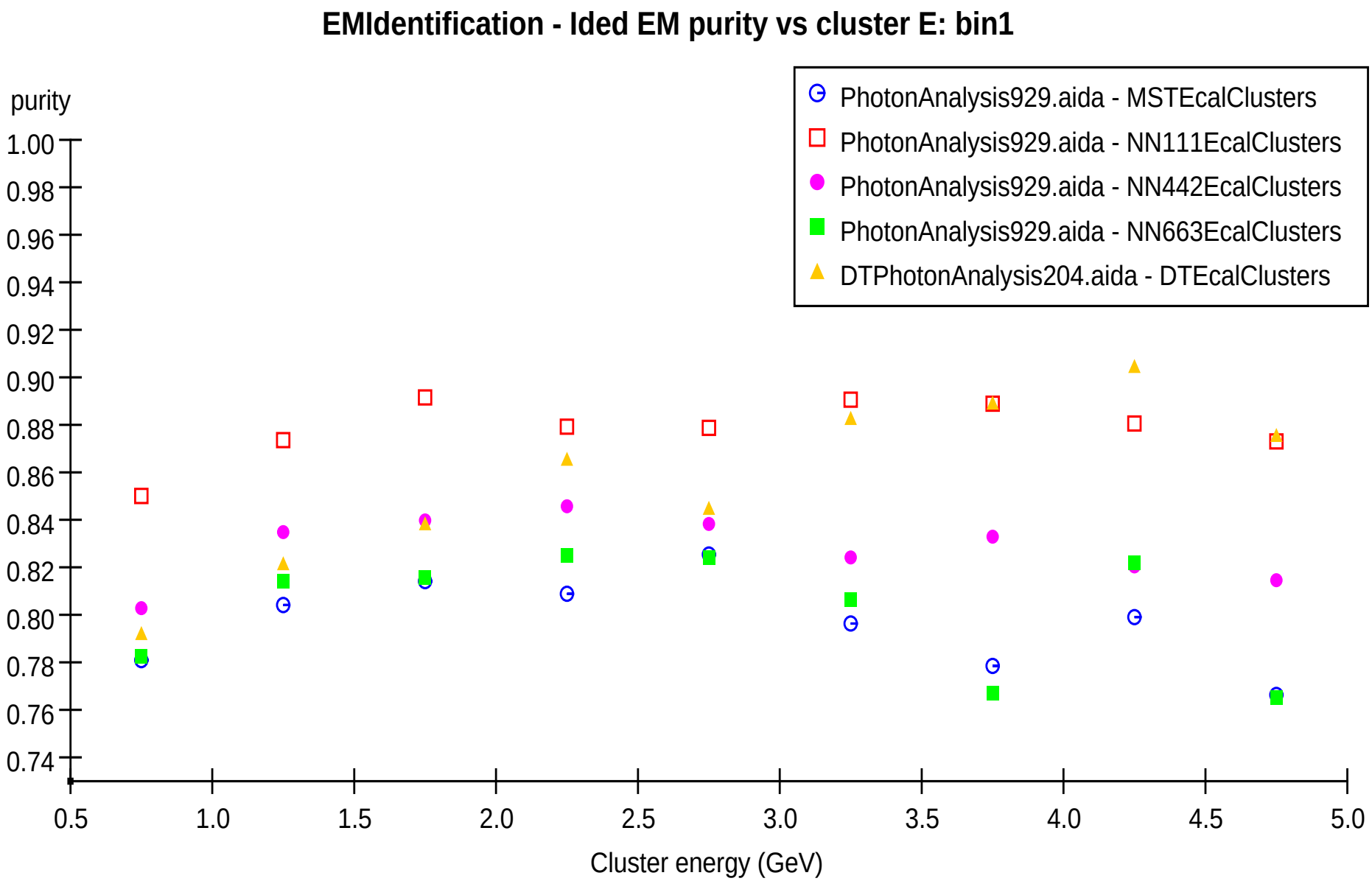
Photon ID

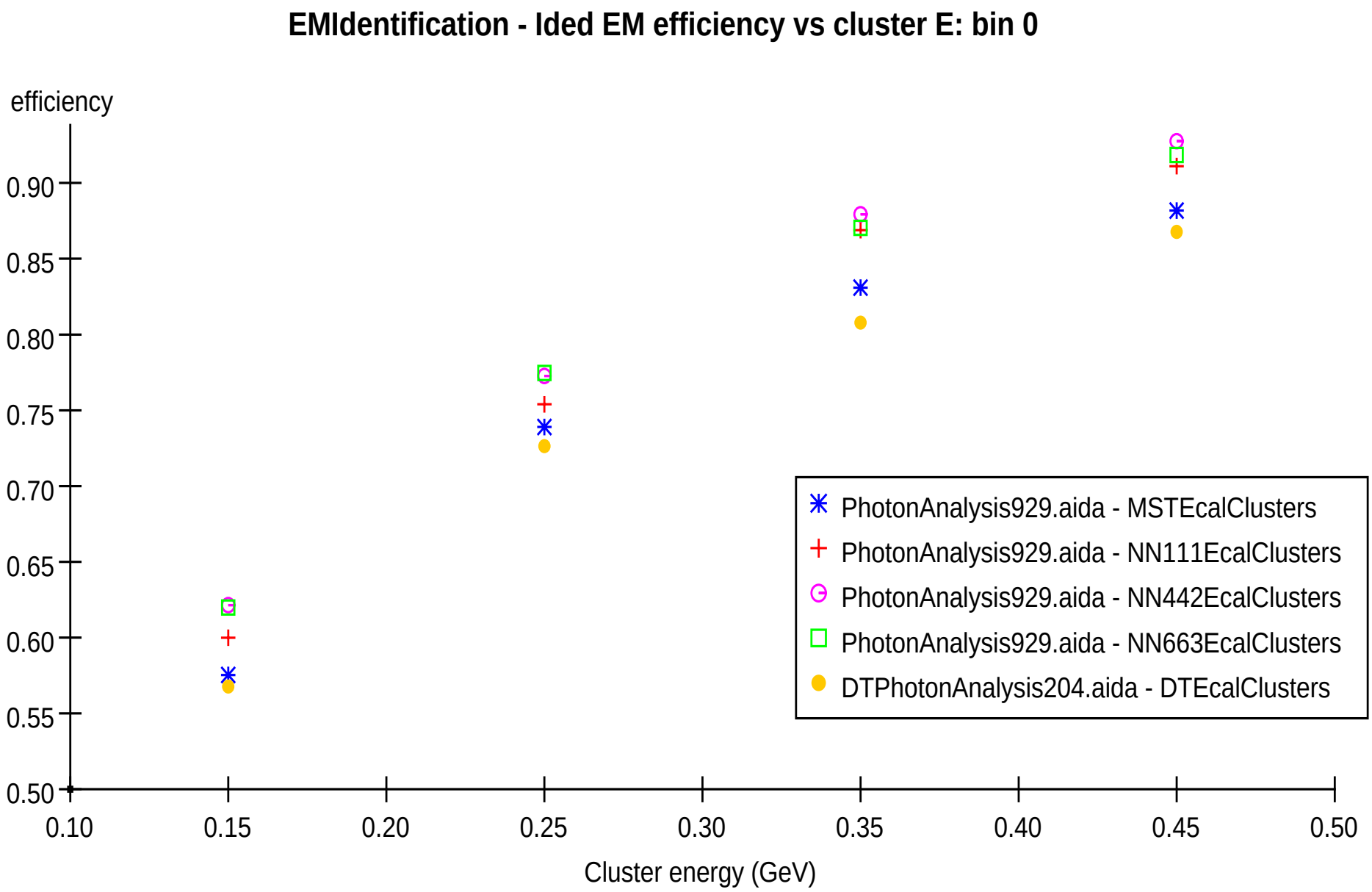
- Have shown general cluster properties assuming perfect identification.
- Add identification.
- Use HMatrix, and require first layer hit < 7, and chisq < 200.
- Use cluster categories for efficiencies and purities

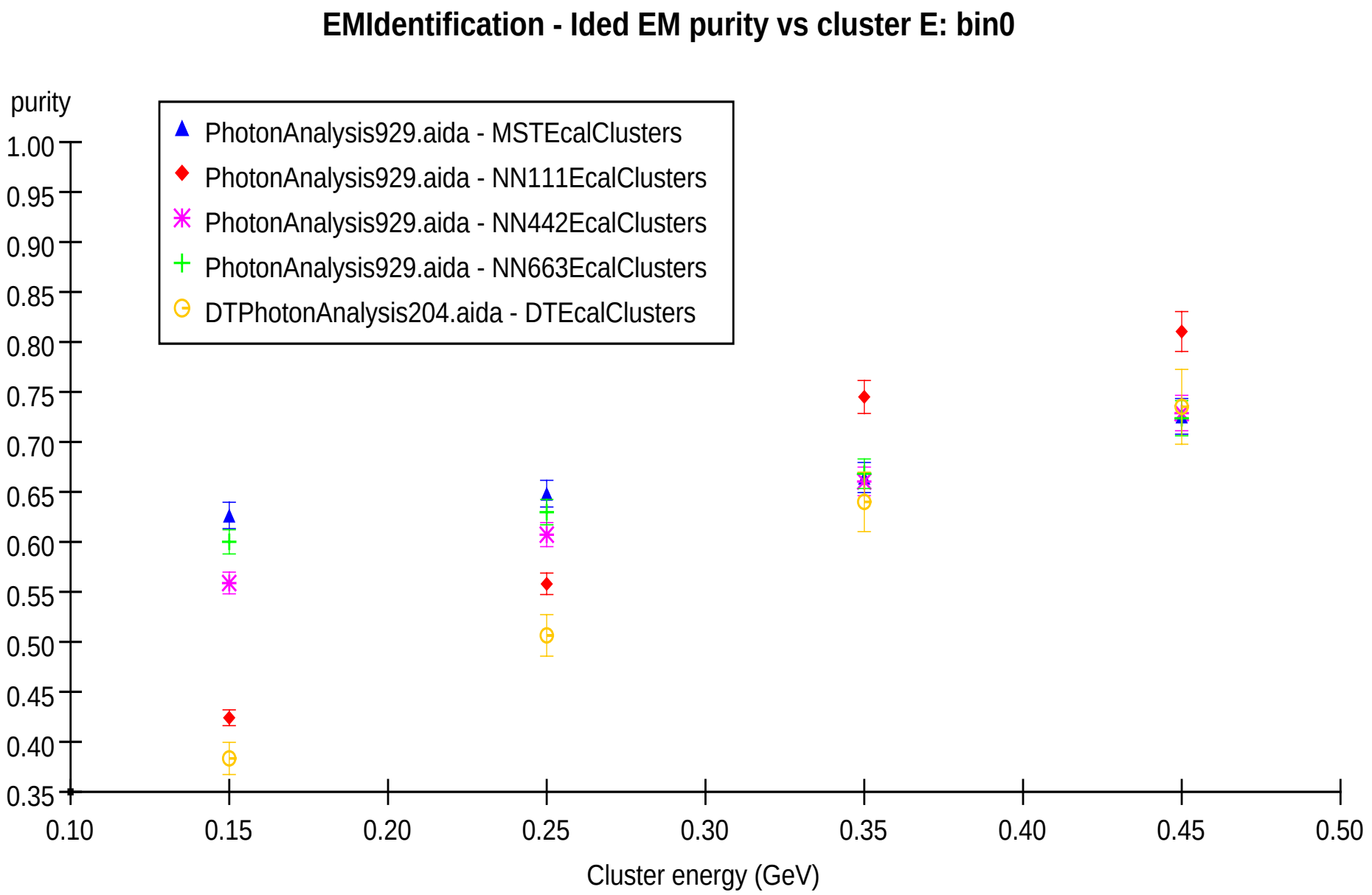

EMIdentification - Ided EM efficiency vs cluster E: bin 3
efficiency
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
5
10
15
20
25
30
35
40
45
50
Cluster energy (GeV)
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters


EMIdentification - Ided EM purity vs cluster E: bin3
purity
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
5
10
15
20
25
30
35
40
45
50
Cluster energy (GeV)


EMIdentification - Ided EM efficiency vs cluster E: bin 1
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
0.85
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

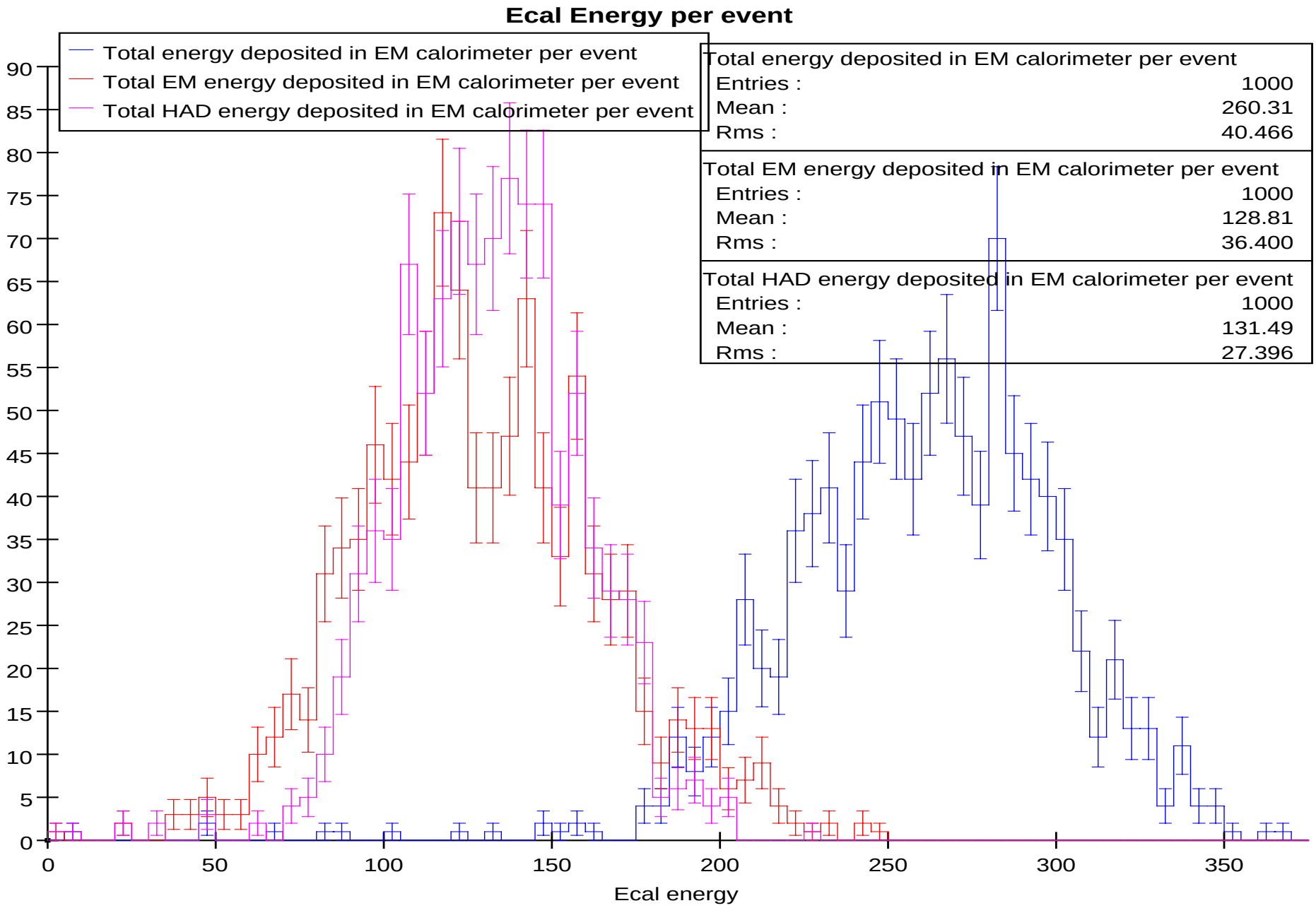

EMIdentification - Ided EM purity vs cluster E: bin1
purity
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0.78
0.76
0.74
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Cluster energy (GeV)
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters

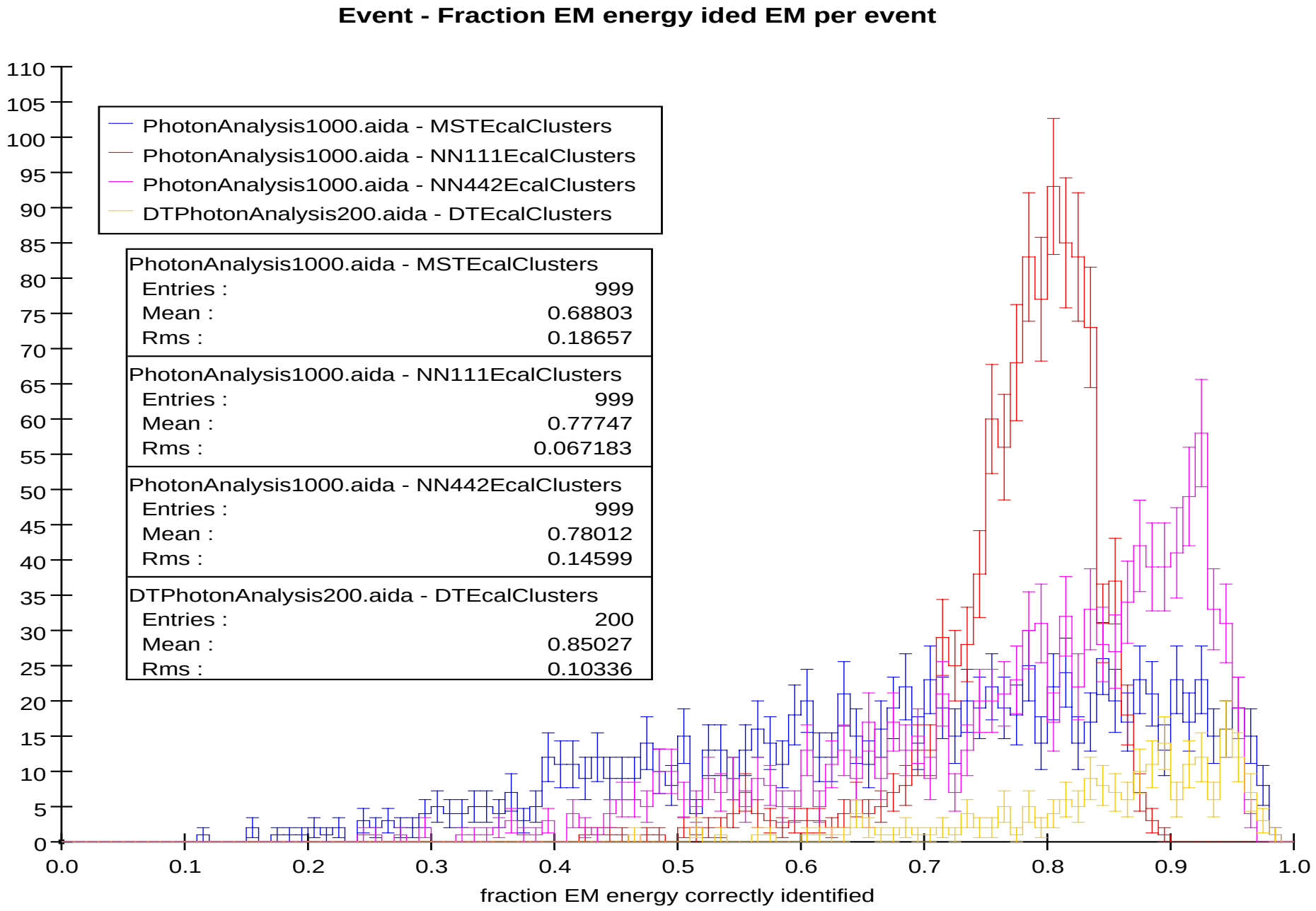

EMIdentification - Ided EM efficiency vs cluster E: bin 0
efficiency
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Cluster energy (GeV)
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters

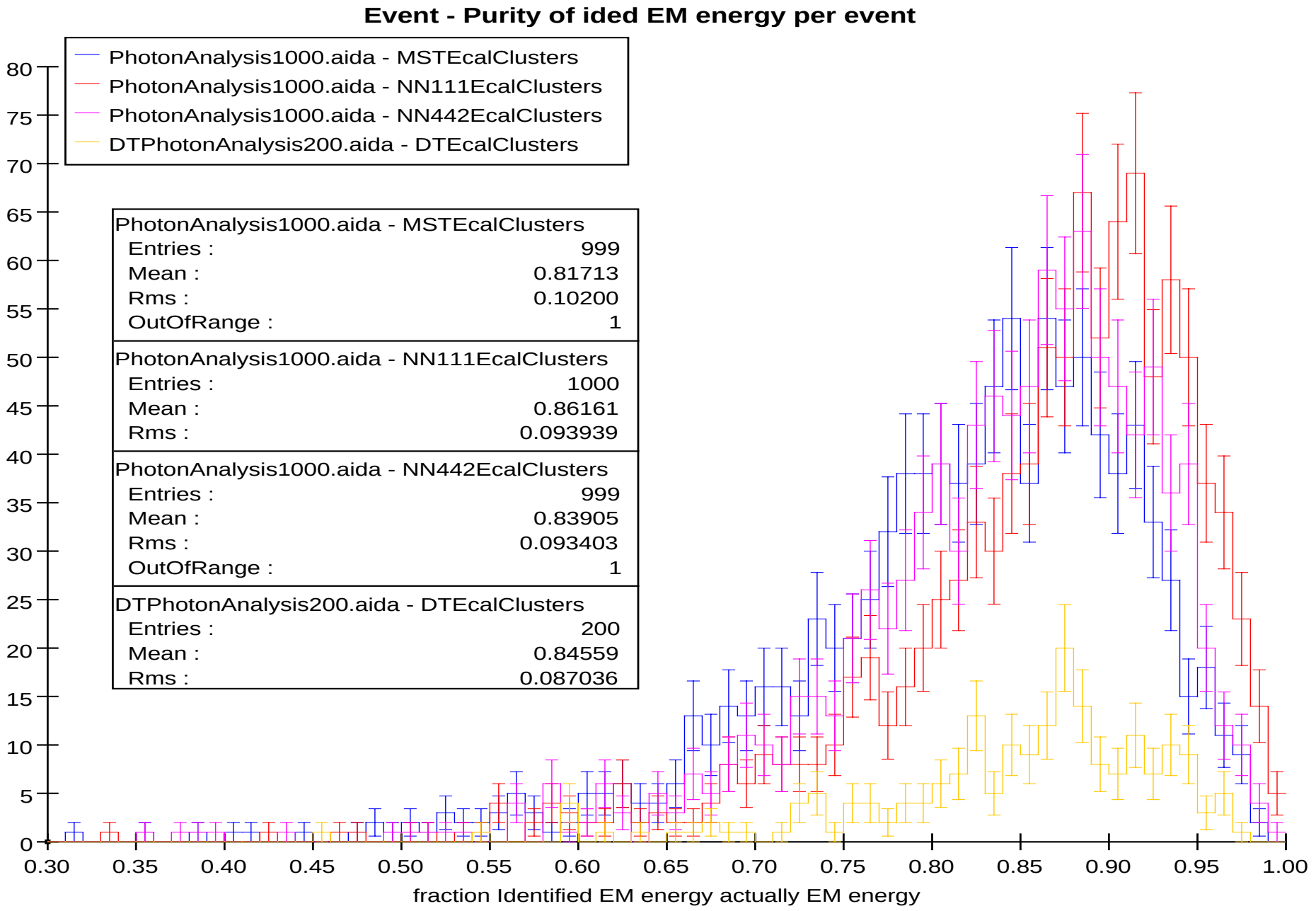

EMIdentification - Ided EM purity vs cluster E: bin0
purity
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
PhotonAnalysis929.aida - MSTEcalClusters
PhotonAnalysis929.aida - NN111EcalClusters
PhotonAnalysis929.aida - NN442EcalClusters
PhotonAnalysis929.aida - NN663EcalClusters
DTPhotonAnalysis204.aida - DTEcalClusters
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
Cluster energy (GeV)

Event efficiency/purity

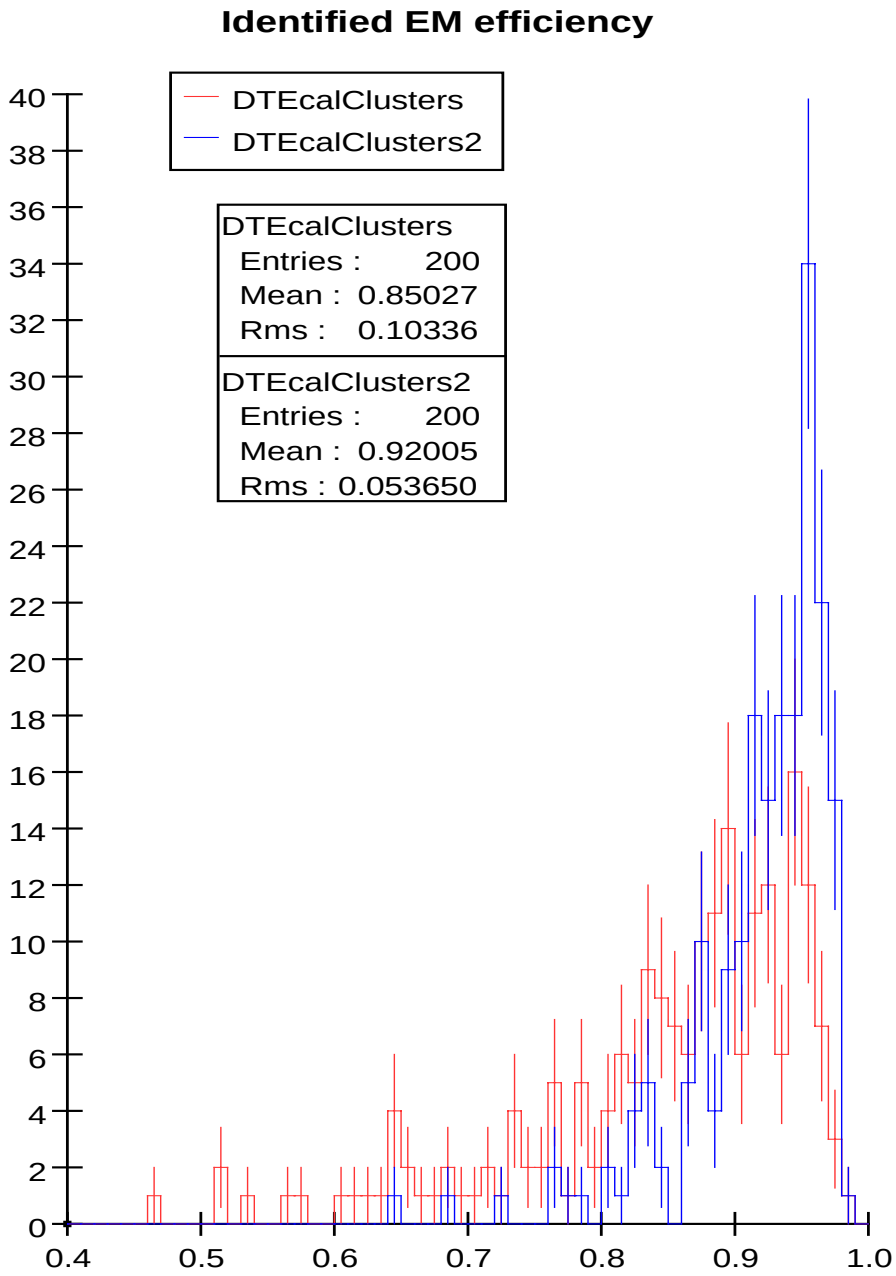
- Can also look at efficiency/purity per event (per jet would be more useful)
- Define photon event efficiency: (Total energy identified as EM from EM particles)/(Total EMcal energy from EM particles)
- Define photon event purity: (Total energy identified as EM from EM particles)/(Total energy identified as EM)

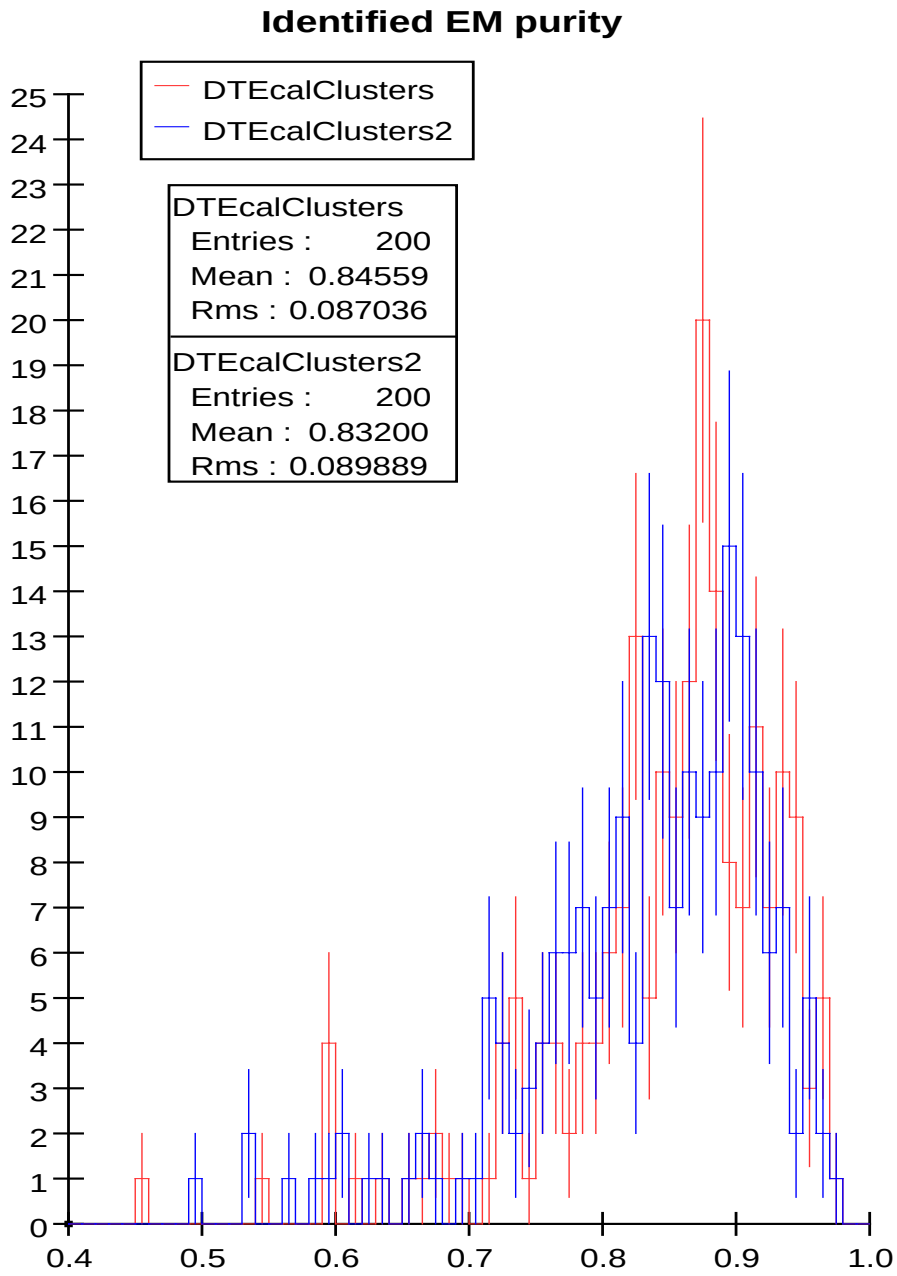

Ecal Energy per event
Total energy deposited in EM calorimeter per event
Total EM energy deposited in EM calorimeter per event
Total HAD energy deposited in EM calorimeter per event
Total energy deposited in EM calorimeter per event
Entries : 1000
Mean : 260.31
Rms : 40.466
Total EM energy deposited in EM calorimeter per event
Entries : 1000
Mean : 128.81
Rms : 36.400
Total HAD energy deposited in EM calorimeter per event
Entries : 1000
Mean : 131.49
Rms : 27.396
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
0
50
100
150
200
250
300
350
Ecal energy


Event - Fraction EM energy ided EM per event
PhotonAnalysis1000.aida - MSTEcalClusters
PhotonAnalysis1000.aida - NN111EcalClusters
PhotonAnalysis1000.aida - NN442EcalClusters
DTPhotonAnalysis200.aida - DTEcalClusters
PhotonAnalysis1000.aida - MSTEcalClusters
Entries : 999
Mean : 0.68803
Rms : 0.18657
PhotonAnalysis1000.aida - NN111EcalClusters
Entries : 999
Mean : 0.77747
Rms : 0.067183
PhotonAnalysis1000.aida - NN442EcalClusters
Entries : 999
Mean : 0.78012
Rms : 0.14599
DTPhotonAnalysis200.aida - DTEcalClusters
Entries : 200
Mean : 0.85027
Rms : 0.10336
110
105
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
fraction EM energy correctly identified


Event - Purity of ided EM energy per event
PhotonAnalysis1000.aida - MSTEcalClusters
PhotonAnalysis1000.aida - NN111EcalClusters
PhotonAnalysis1000.aida - NN442EcalClusters
DTPhotonAnalysis200.aida - DTEcalClusters
PhotonAnalysis1000.aida - MSTEcalClusters
Entries : 999
Mean : 0.81713
Rms : 0.10200
OutOfRange : 1
PhotonAnalysis1000.aida - NN111EcalClusters
Entries : 1000
Mean : 0.86161
Rms : 0.093939
PhotonAnalysis1000.aida - NN442EcalClusters
Entries : 999
Mean : 0.83905
Rms : 0.093403
OutOfRange : 1
DTPhotonAnalysis200.aida - DTEcalClusters
Entries : 200
Mean : 0.84559
Rms : 0.087036
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
fraction Identified EM energy actually EM energy

- Just for fun, use DT clustering but for IDing just use the core (NN111)


Identified EM efficiency
DTEcalClusters
DTEcalClusters2
DTEcalClusters
Entries : 200
Mean : 0.85027
Rms : 0.10336
DTEcalClusters2
Entries : 200
Mean : 0.92005
Rms : 0.053650
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40
0.4 0.5 0.6 0.7 0.8 0.9 1.0


Identified EM purity
DTEcalClusters
DTEcalClusters2
DTEcalClusters
Entries : 200
Mean : 0.84559
Rms : 0.087036
DTEcalClusters2
Entries : 200
Mean : 0.83200
Rms : 0.089889
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
0.4 0.5 0.6 0.7 0.8 0.9 1.0

Photon ID

- Comparisons between clusterers not completely fair, photon ID would need individual optimization
- Tools available for studying such an optimization

Resolution

- I haven’t done the studies, but tools exist
- Individual photon resolution: combination of intrinsic detector resolution, cluster efficiency and purity.
- Contribution to Jet energy resolution: identification efficiency and purity important

The tool

- This analysis package, specific to photons, should be in CVS by end of workshop.
- I encourage anyone doing PFA work without cheating on photons to use it, so at least you know what you are getting
- I will write a specific photon finder, and I encourage anyone else who writes one to analyze it with this package
- Usage examples will be put in the contrib area, including how to make your own plots