Status of the SiD/Iowa Particle Flow Algorithm

Remi Zaidan, Garabed Halladjian,
Mat Charles[1], Ron Cassell[2],
&
Usha Mallik

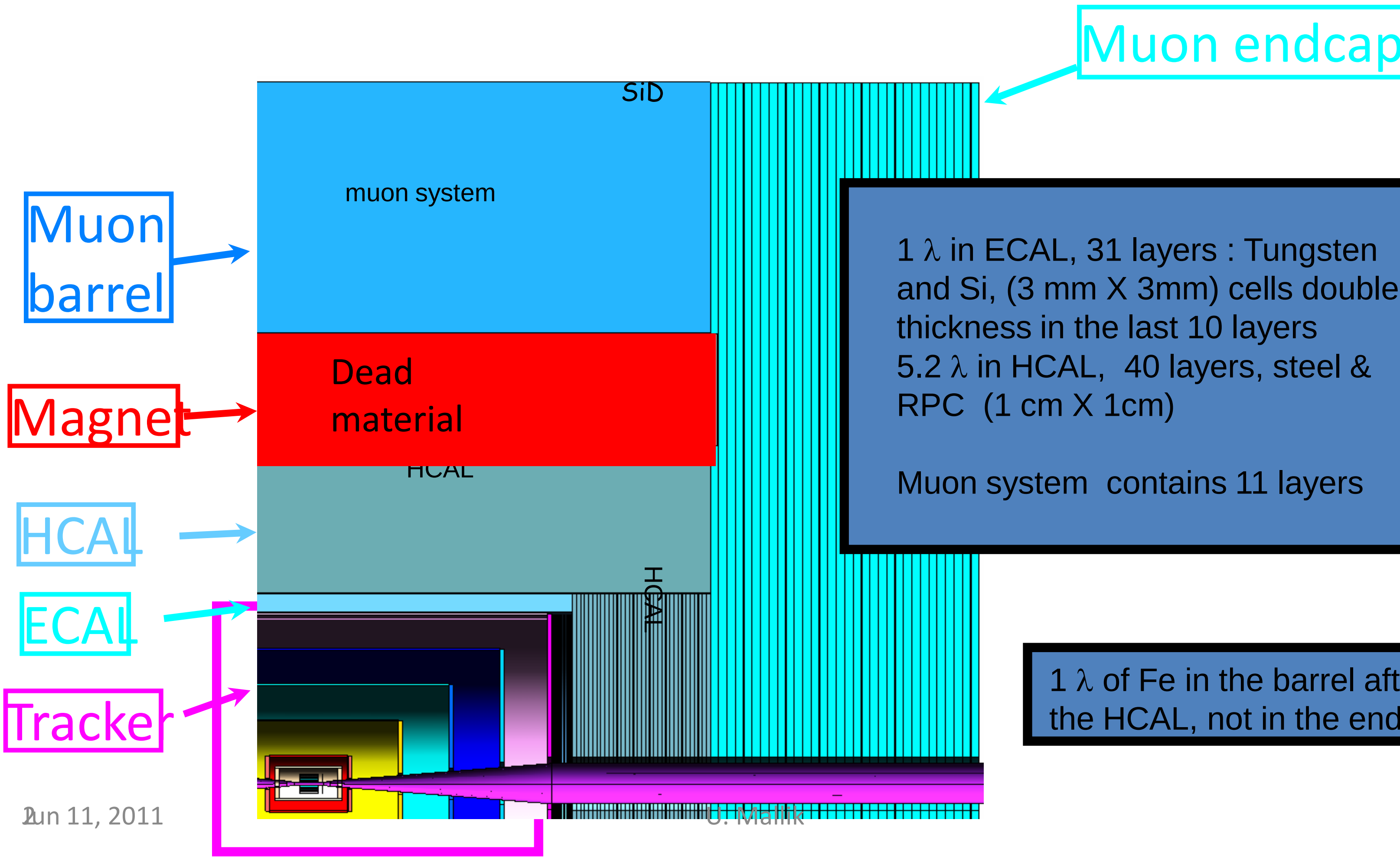
THE UNIVERSITY OF IOWA

1: Oxford University, 2: SLAC National Lab

The Collider and the Detector (SiD02)

Developed for high energy electron-positron linear collider

Nominal $\sqrt{s}$ = 500 GeV , upgrade to 1 TeV

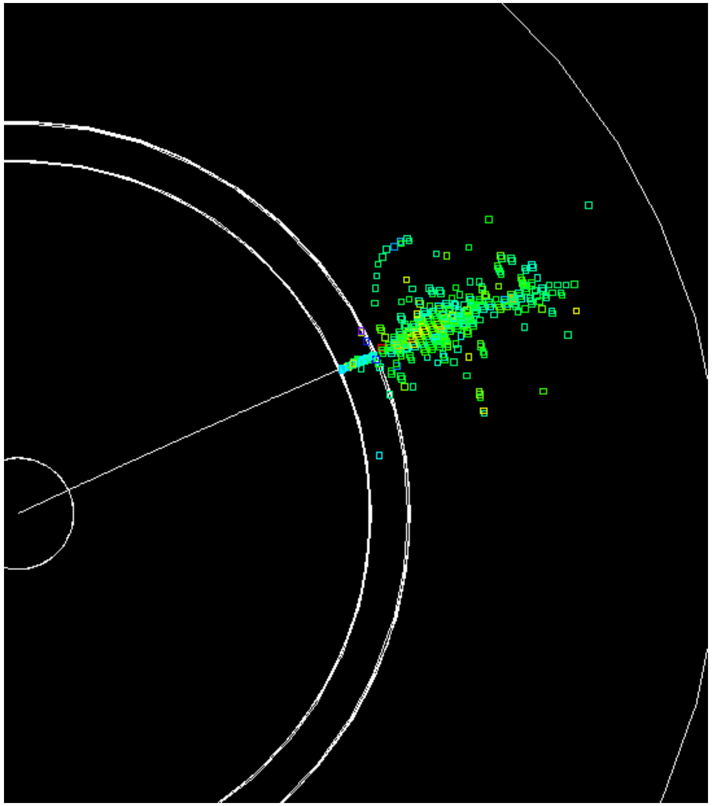


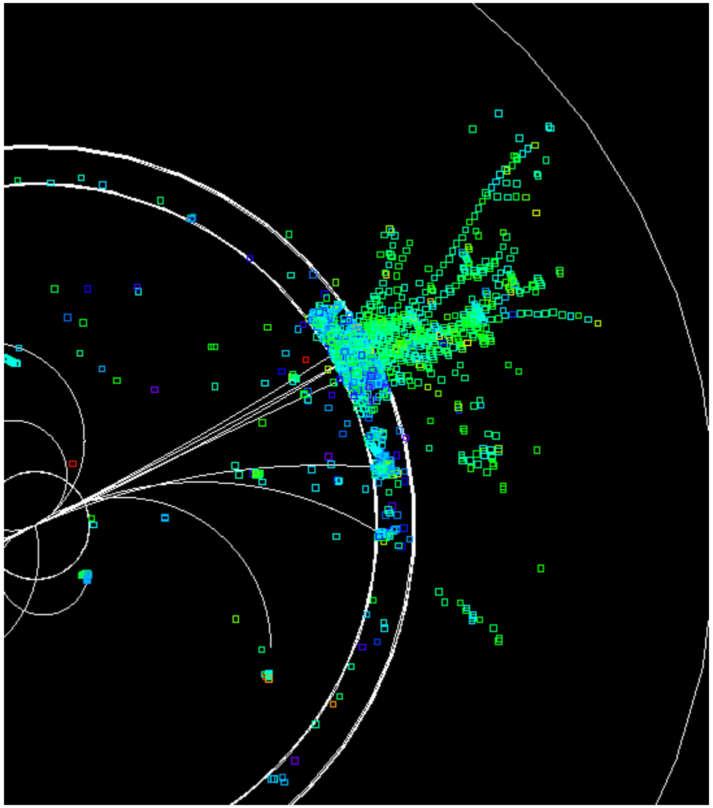
1 λ in ECAL, 31 layers : Tungsten and Si, (3 mm X 3mm) cells double thickness in the last 10 layers
5.2 λ in HCAL, 40 layers, steel & RPC (1 cm X 1cm)

Muon system contains 11 layers

1 λ of Fe in the barrel after the HCAL, not in the endcap

A (half) event display

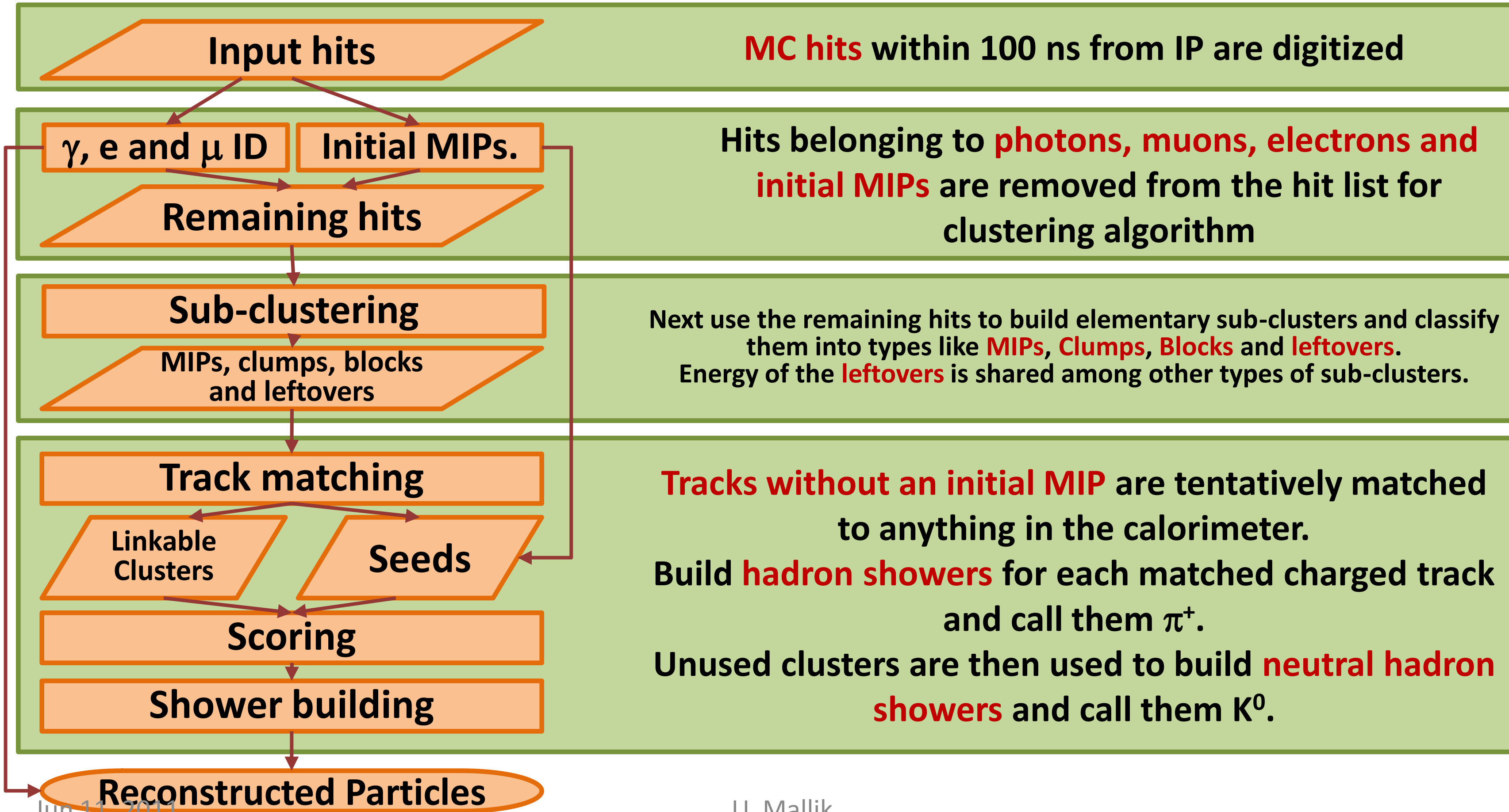




A Particle Flow Algorithm (PFA) provides improved resolution which translates into superior jet-jet reconstructed mass resolution, such as the difference between the W and the Z mass, necessary to reach the physics goals of such a collider

What is in a PFA

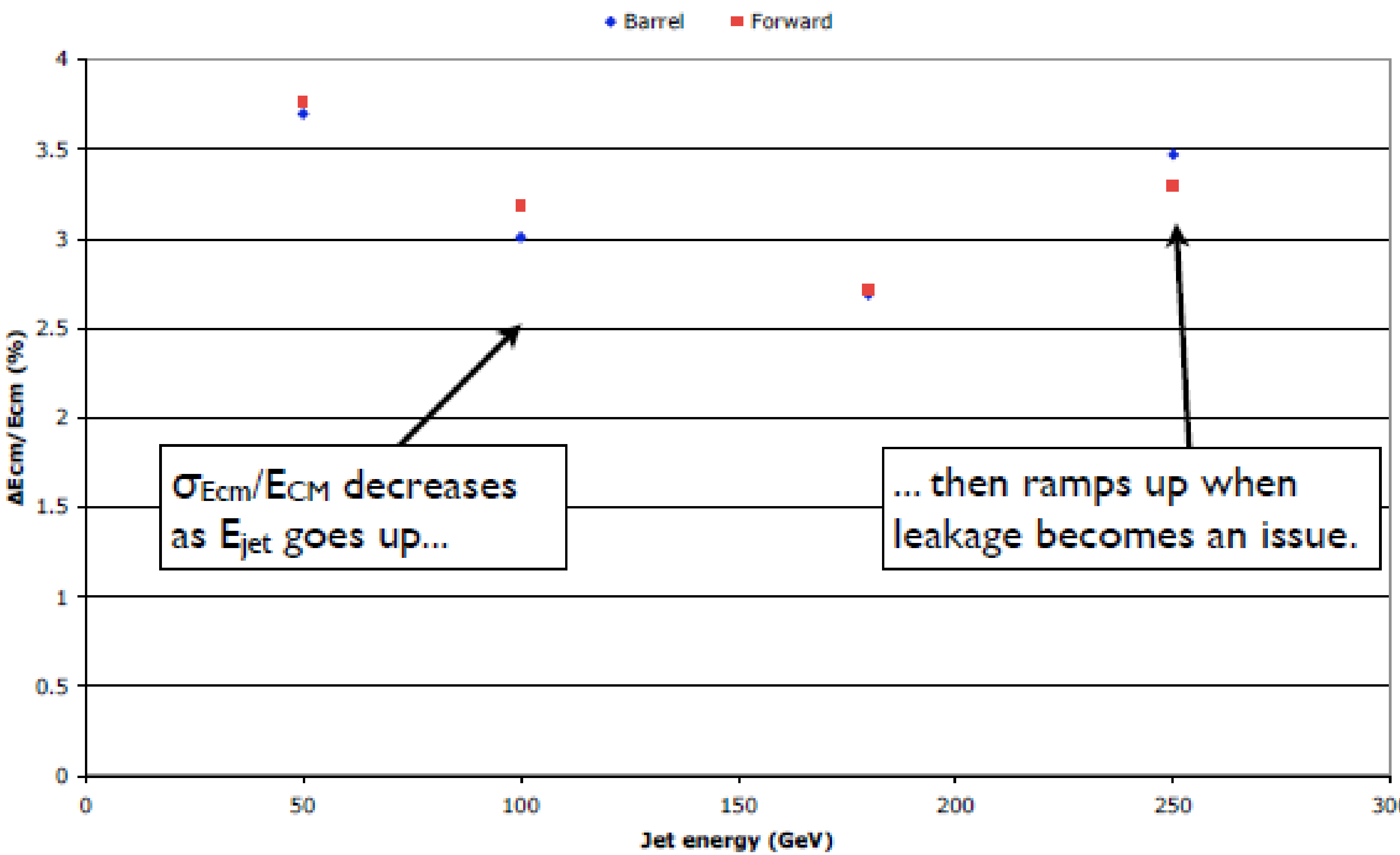
- Reconstruct the complete shower from each charged hadron track
- Replace the shower energy by momentum measured by the tracker
- The leftover is neutral hadrons, which contributes ~10% to each event
- Result: much improved jet resolution, and jet-jet reconstruction mass resolution



Performance at LOI

with SiD02 geometry

- Last significant resolution result used for the LOI
  - In 2009 for validation of the SiD concept (for all physics)



At higher energies resolution worsened partly due to leakage but also due to algorithm

Phoenix from the ashes

- Previous algorithm optimized at lower CM energies, adequate, but not optimal
- **Confusion increases significantly at higher energies with overlaps for 500 GeV and higher collider energies**
- **Difficult to use incremental improvement**

  Overall PFA is a combination of many interrelated/interdependent algorithmic pieces, simply improving a complex piece does not help because of algorithms executed downstream
- **A complete rebuilding effort used keeping a similar philosophy**

Still use (SiD02 for exact comparison, but SiD03 compatible)

Developed a set of tools to quality-control each piece of algorithm independently such as measuring purity and efficiency of a cluster or a shower

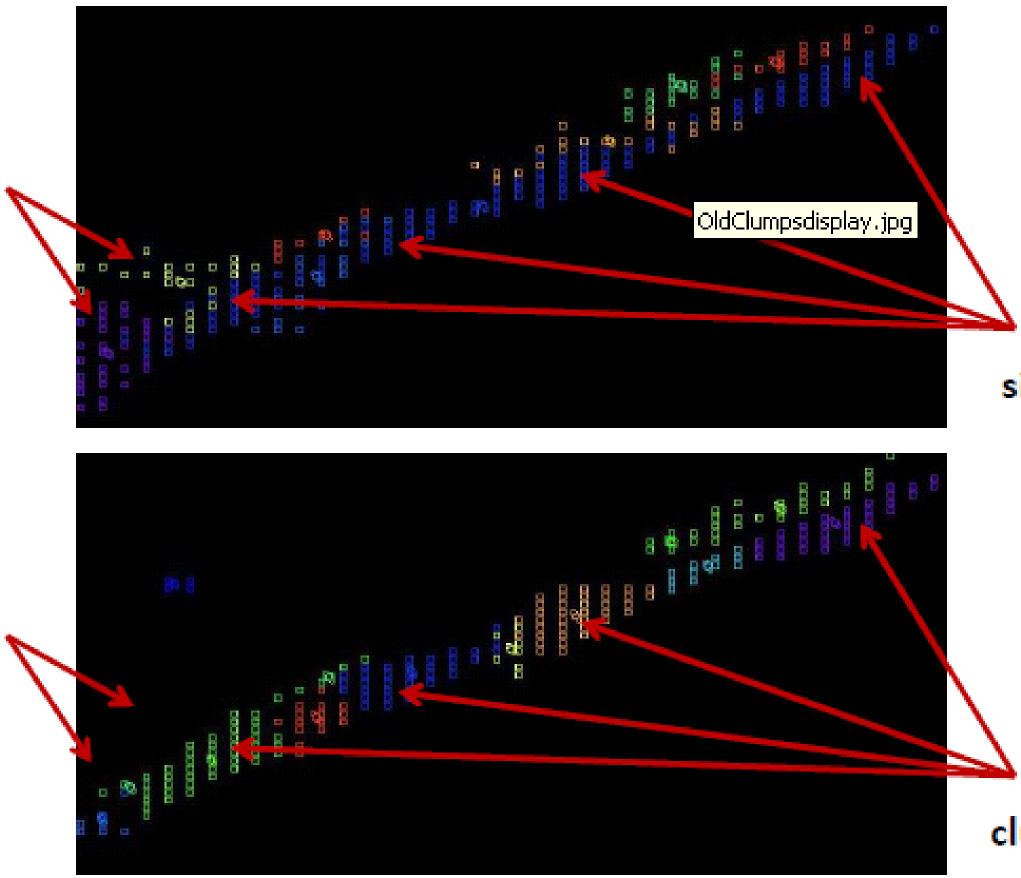
Changes in the algorithm

- Redefining sub-clusters and photons
- Connecting sub-clusters
- Shower Building

THE UNIVERSITY OF IOWA

The changes I

Sub-cluster categories: Clumps, Blocks, leftovers, redefined

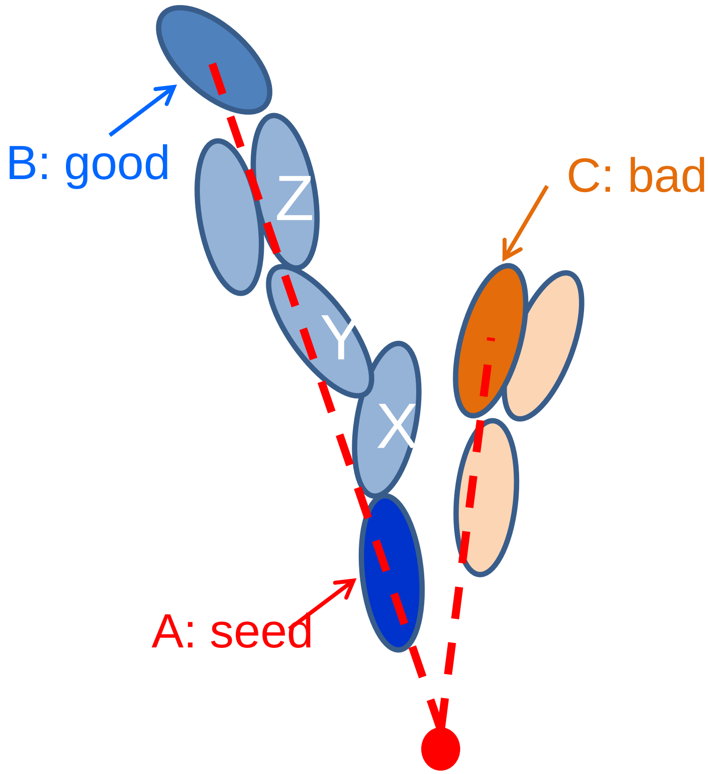


Old:
single clump

New:
clump broken

The changes II

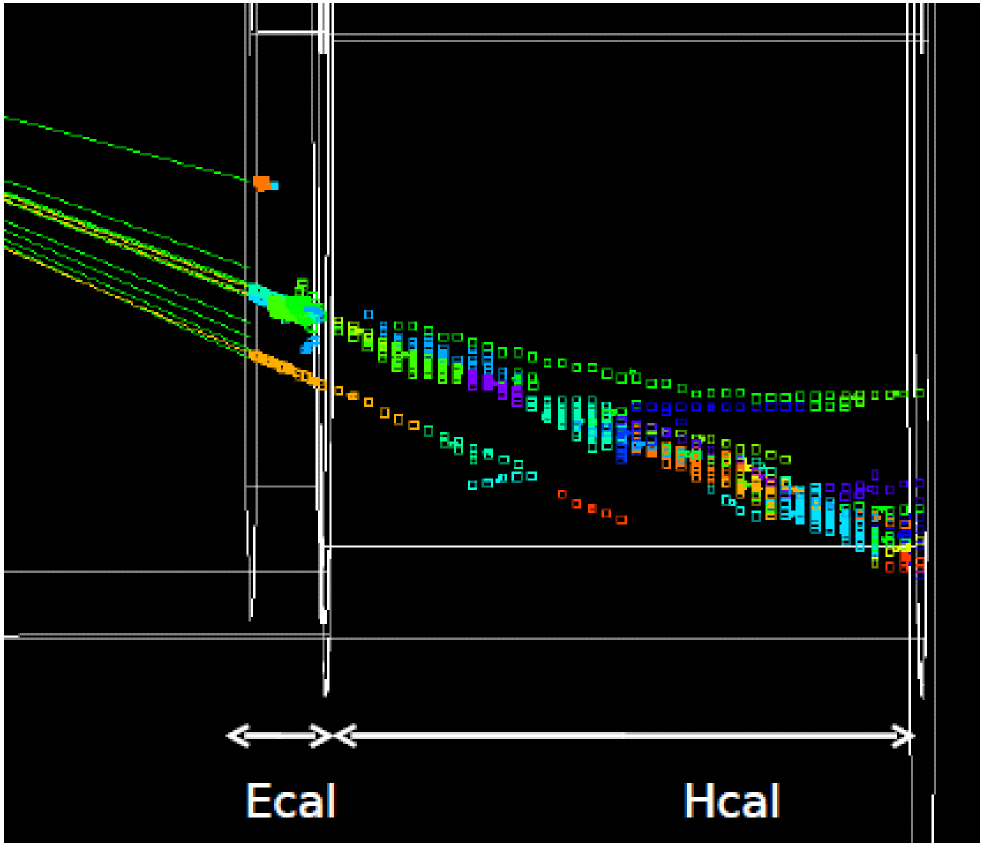
Connecting sub-clusters by interpolation



The University of Iowa

The Changes: III

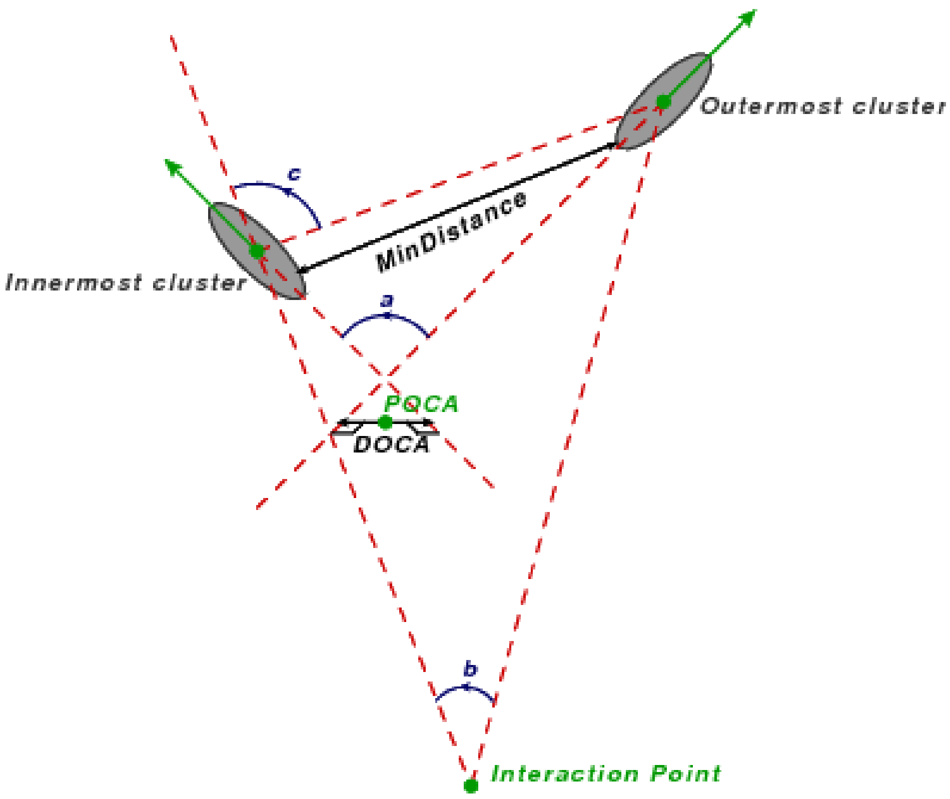
Scales in ECAL and HCAL very different, use different likelihood probability density functions (PDF)



The Changes: IV

Proper Maximum Likelihood, with additional variables and correlations

Train to determine "score", basically linkage probability between 0 and 1

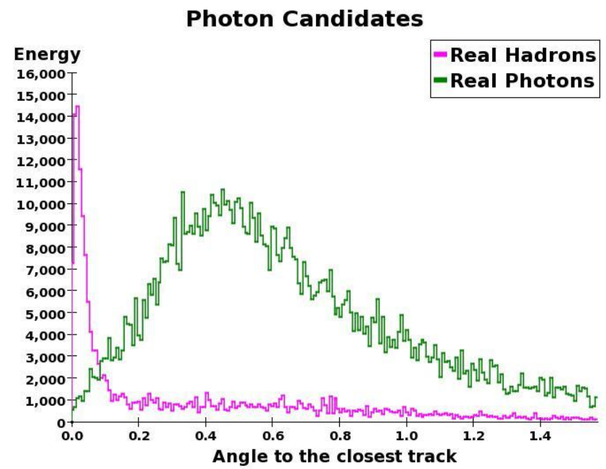


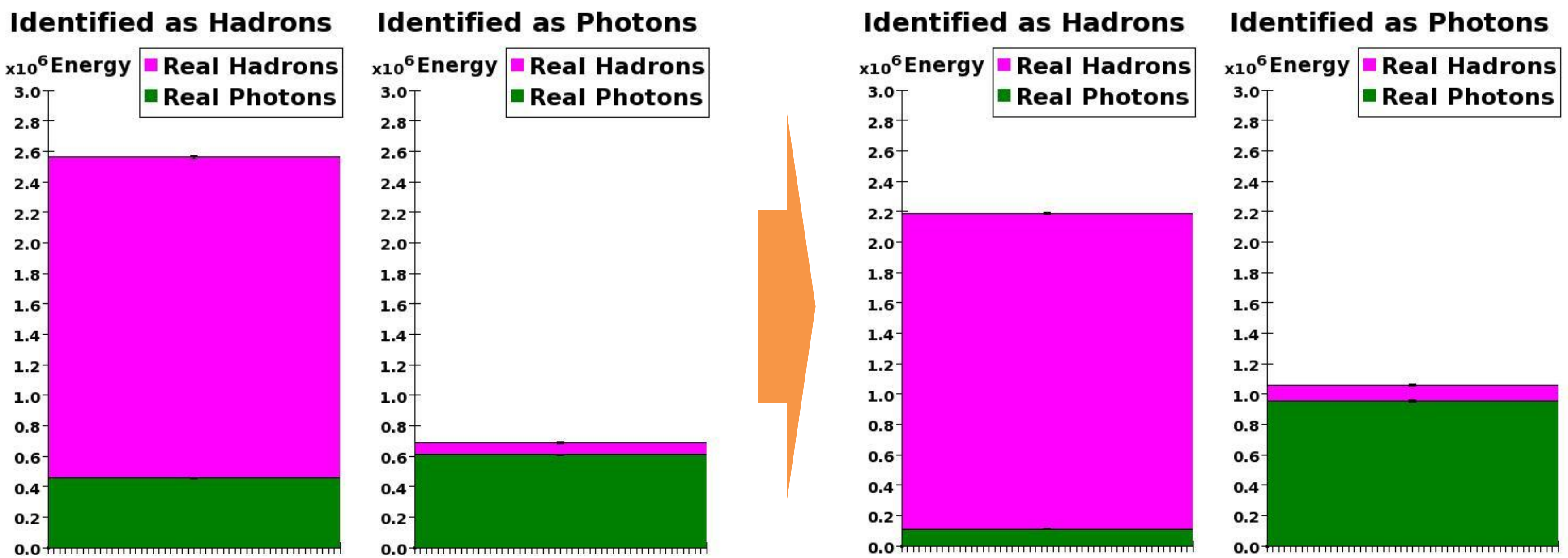
Information used in Likelihood

The University of Iowa

Improving photon efficiency

- A photon `recovered' if there is no reconstructed track within of 0.1 rad
- Some photon/hadron confusion remains
- Now limited by Photon ID
- Need to quantify the confusion effect in the PFA performance





Changes in Shower Building

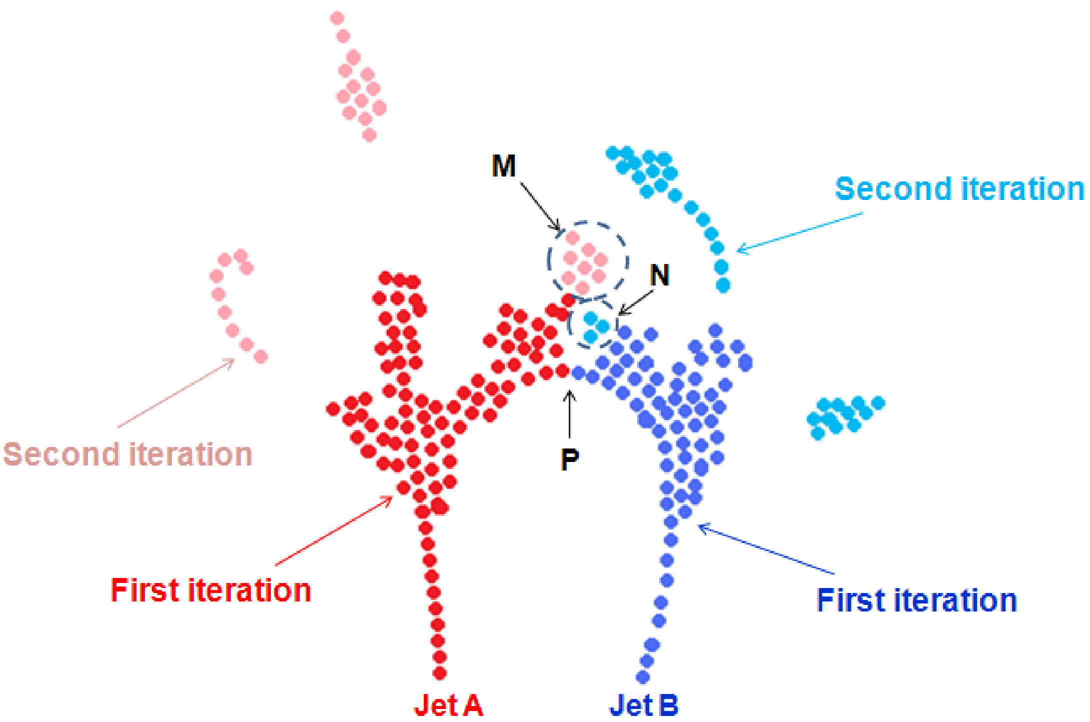
Reconstruct the skeleton of the shower with high purity in a first pass

Only outgoing direction of linkage between sub-clusters allowed

All tracks are treated democratically, not momentum ordered

Multiple assignments allowed, to be resolved in the second pass

A simple search for neutral hadron showers



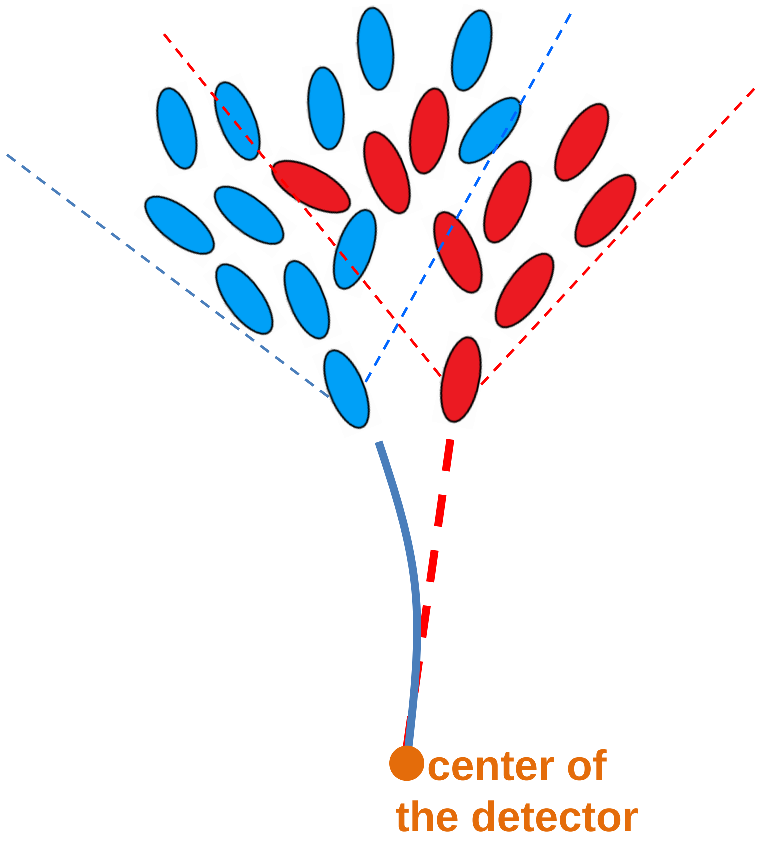
Build a robust shower core with high purity, reasonable efficiency

Leave ambiguities for later

Leave back-scattering for later

A simple check for neutral hadron showers

Being aware of the presence of a nearby neutral shower helps in charged shower building/resolving.



THE UNIVERSITY OF IOWA

Taking stock of confusion matrix

	True photons	True charged hadrons	True neutral hadrons
Identified as photon	30.7%	0.5%	2.9%
Identified as hadron, assigned to charged shower	1.0%	30.0%	2.3%
Identified as hadron, not yet linked to a shower	2.4%	20.0%	10.1%

Photon efficiency and purity both 90%.

Green = correct decision
Red = wrong decision, causes confusion
Black = photon vs neutral hadron mis-ID (not so serious)
Blue = charged clusters not yet attached to a track (needs second pass)

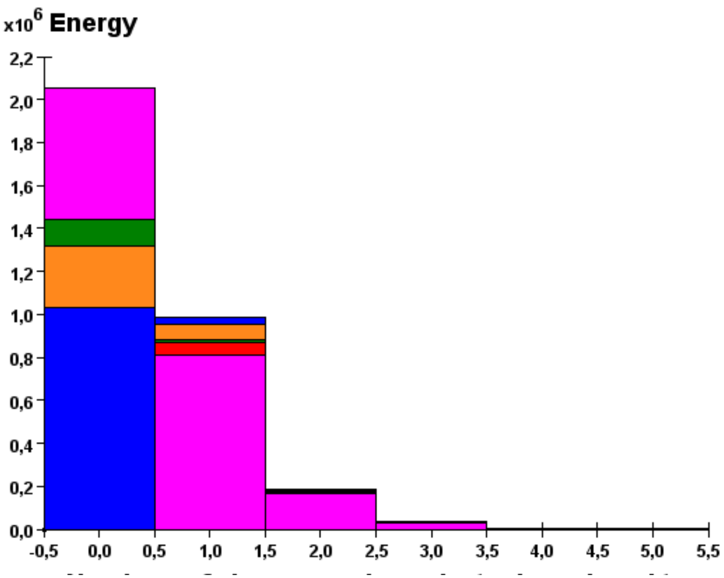
Quality control: confusion

Assigned to neutral particles (bin 0) but in reality are from:

- **Photons**
- **Neutral hadrons**
- **Charged hadrons with no reconstructed**
- **Charged hadrons**

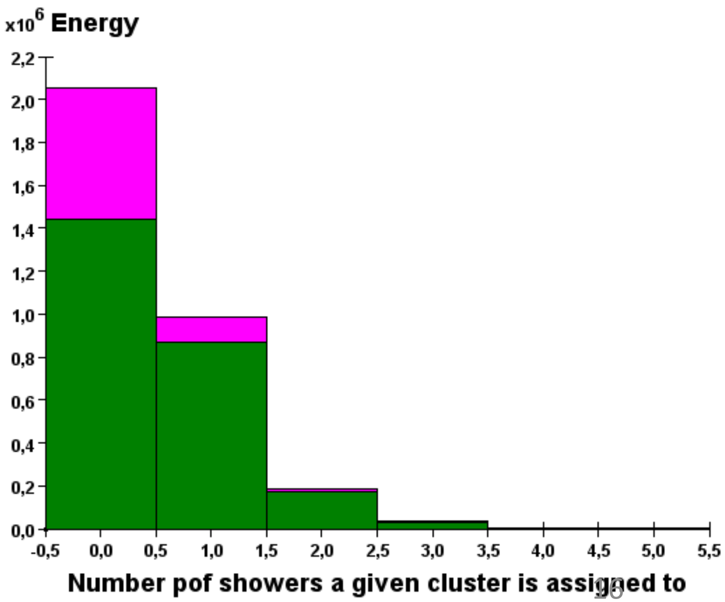
Assigned to charged particles (bin ≥ 1) but in reality are from:

- **Photons**
- **Neutral hadrons**
- **Charged hadrons with no reconstructed**
- **A different charged hadron than all the ones assigned to**
- **One of the charged hadrons assigned to**



Correct Assignment

Confusion



Next steps

Second Iteration

- Recover the unassigned charged energy

  Isolate secondary neutrals

  Isolate backscattered pieces

- Resolve overlaps between charged-neutral and charged-charged hadron

- Assign leftover hits appropriately

Expect first resolution values by end of the summer

Expectations for future

- **Understand the interplay between each piece of algorithm**

- **Tune each piece of the algorithm to get better performance at 500 GeV**

- **Tune for good performance at 1 TeV**

- **Make the algorithm work for higher energies & other detector designs**

Backups

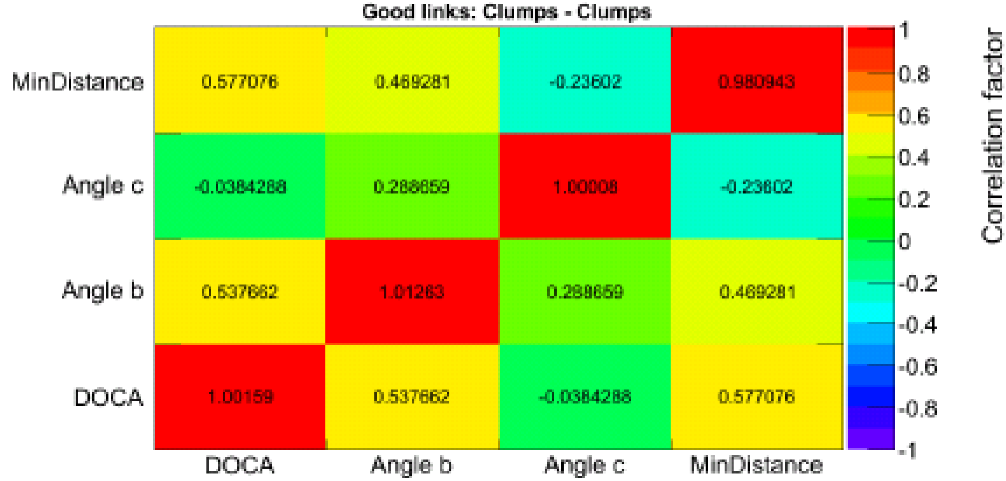
- **Correlation factor:**
  - Defined as:

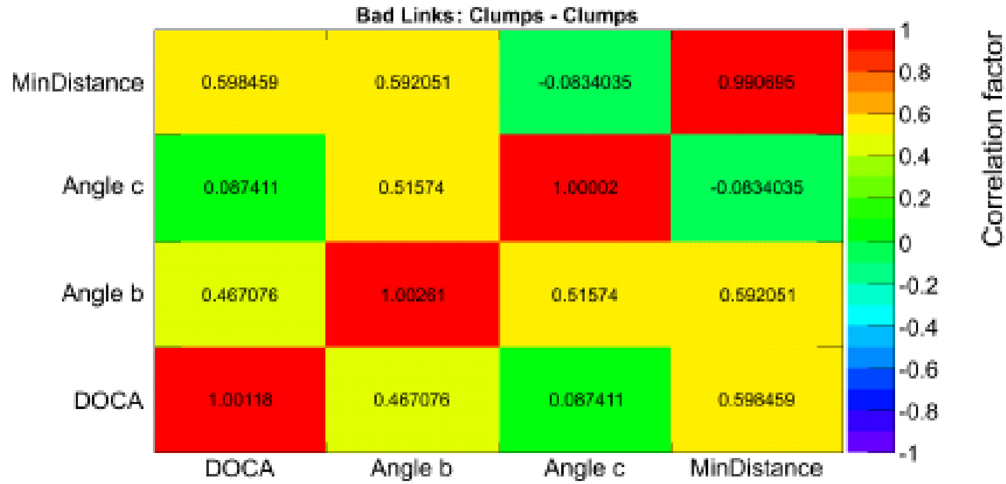
$$\rho_{x,y} = \frac{\langle (x-\bar{x})\times(y-\bar{y})\rangle}{\sqrt{\langle (x-\bar{x})^2\rangle \times \langle (y-\bar{y})^2\rangle}} = \frac{\mathrm{cov}_{x,y}}{\sigma_x \times \sigma_y}$$

  - Measures the degree of correlation:

$$-1 \le \rho_{x,y} \le 1$$

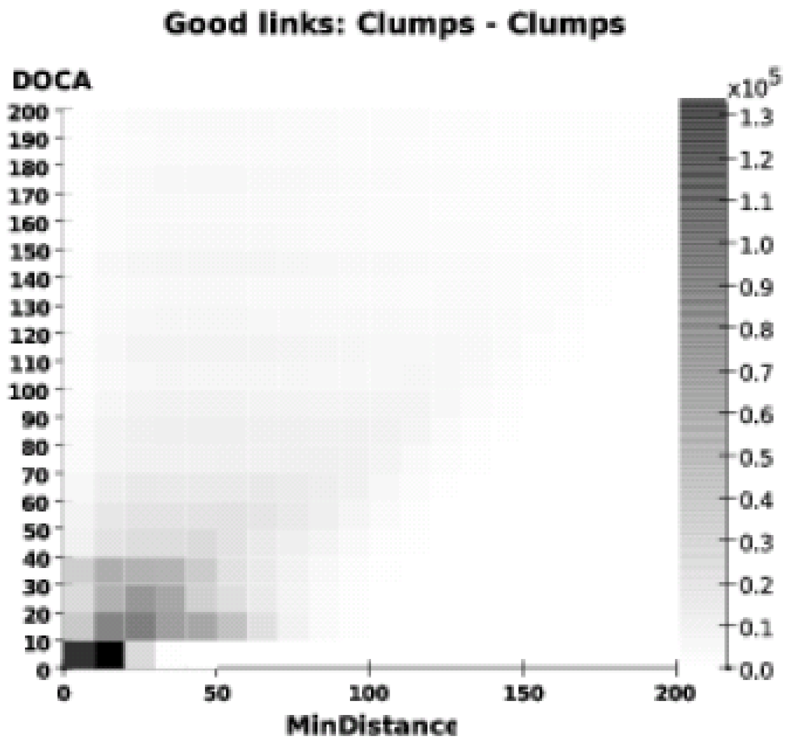
- **Likelihood function assumes independent variables:**
  - Correlations between variables may cause peaks in the likelihood distribution for background in the signal region and vice versa.

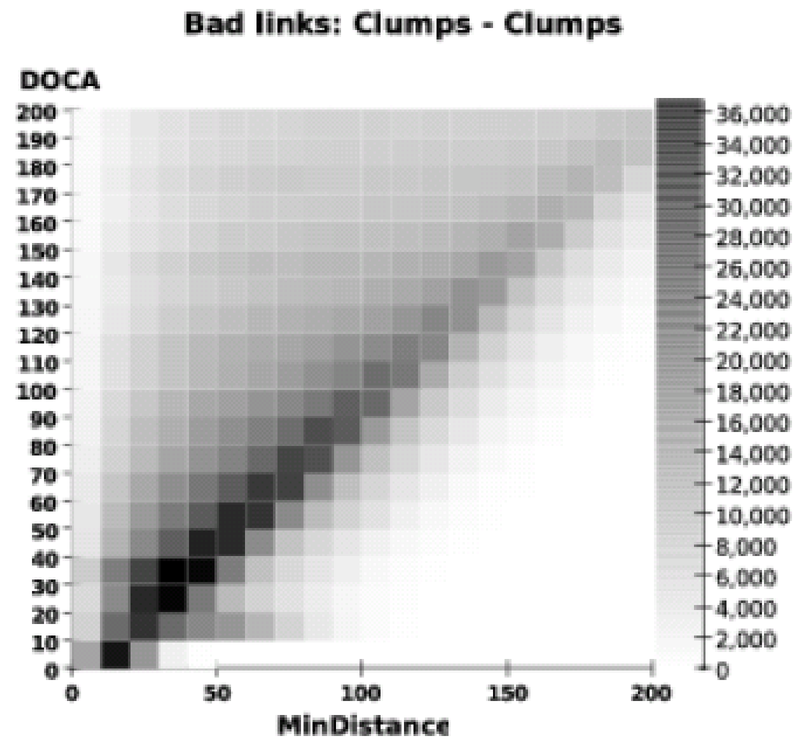




Correlations

- Using a 2-dimentional distributions for correlated variables should enhance likelihood performance with respect to using independent 1-dimentional distributions.





PFA & the Building Blocks in a nutshell

Basic Building Blocks:

Separate the EM showers and the MIPs

Classify number of showers from the calorimeter hits (Directed Tree) and categorize

For each charged track:

- Extend tracks to the beginning of the ECAL
  - Connect it with a MIP-like piece in ECAL as the seed for shower building

For all charged tracks :

- complete shower building with multiple passes
  - assign isolated hits proportionately between nearby showers

Once completed:

- Neutral hadron showers constructed from the leftover clusters

Finally: Four vectors (E,P) are written out for the event

Categorizing: DirectedTree Clustering

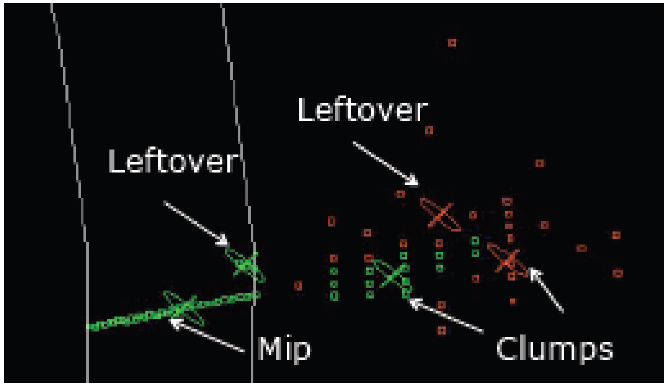
Ecal Digi Hits(Barrel, Endcap)			Hcal Digi Hits (Barrel, Endcap)		
Photon, Electron, Muon	DTree cluster (Ecal Barrel)	DTree cluster (Ecal Endcap)	Muon	DTree cluster (Hcal Barrel)	DTree cluster (Hcal Endcap)

DTree cluster

MIPs	A continuous sequence of single hits
Clumps	Group of hits with high density
Blocks	No structure, if(>= 20hits in ECal, >=15 hits in HCal)
Leftover	No structure, small number of hits (Share with others)

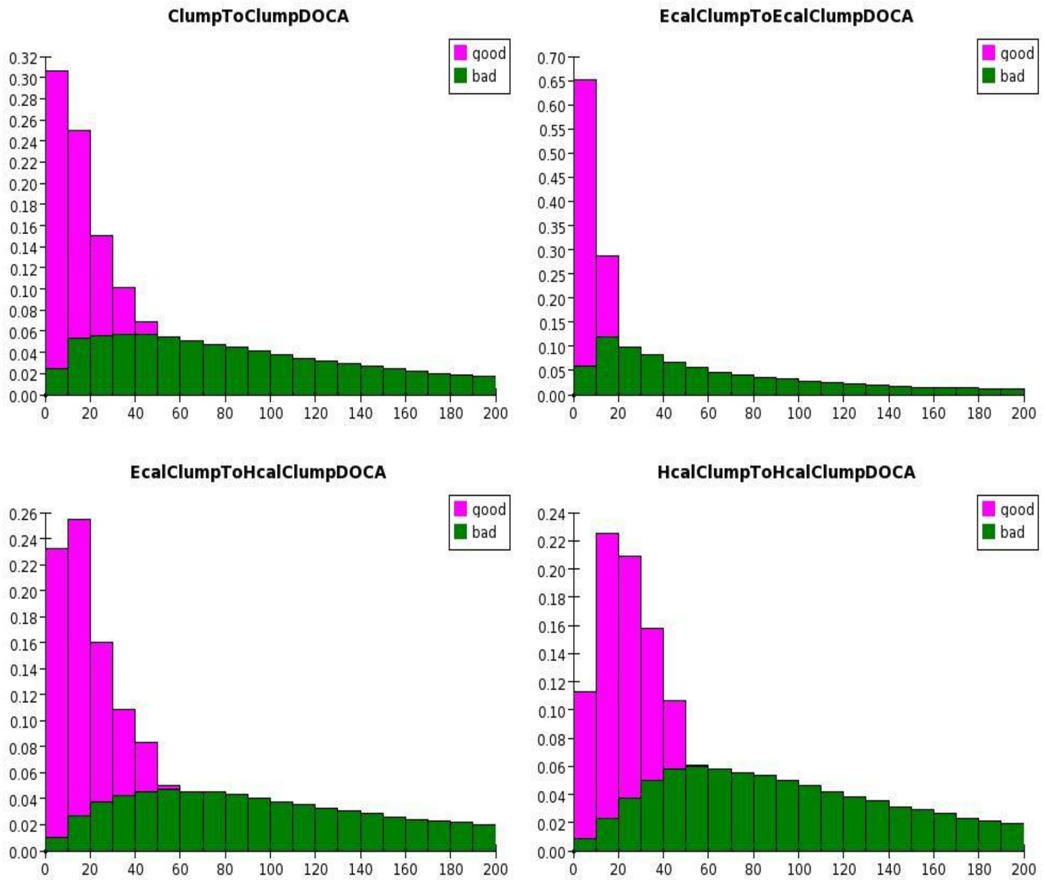
- Leakage

Some of high energy shower escapes Hcal, reaching Muon Detector. Adding the energy by using Muon Endcap as tail catcher give better resolution. (Currently not using Barrel)



Likelihood training: Ecal vs Hcal

- **A concrete example:**
  - Distance of closest approach (mm)
  - Most of the bad links between Ecal-Ecal clusters are within the good peak of the Hcal-Hcal links.
  - Mistakes made in the Ecal are most likely to propagate along with the shower building all the way through the Hcal.



Missing charged energy in the Ecal

- **We looked at the unassigned charged energy:**
  - Many of these missed clusters are situated along the extrapolation of a track and at a distance of 60-70 mm from the extrapolation of the track to the Ecal entrance.
  - This might be due to a problem at the crossing from thin to thick Ecal layers.
  - Need to investigate, but for now we decided to pick up those clusters by hand.