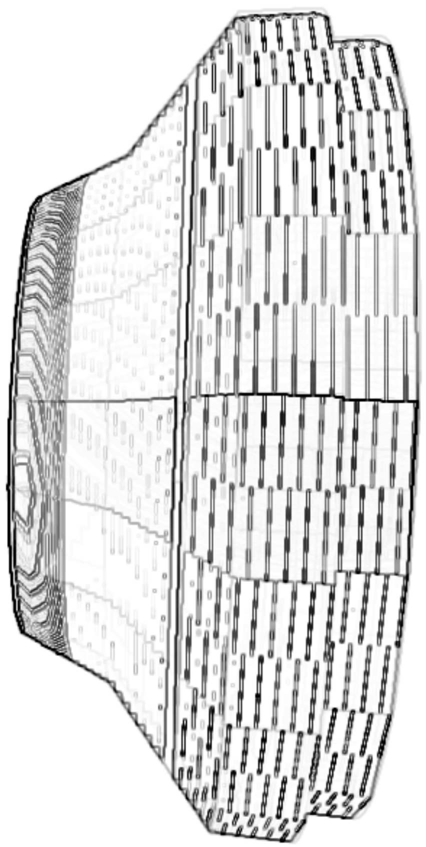
HGC Reconstruction & Performance

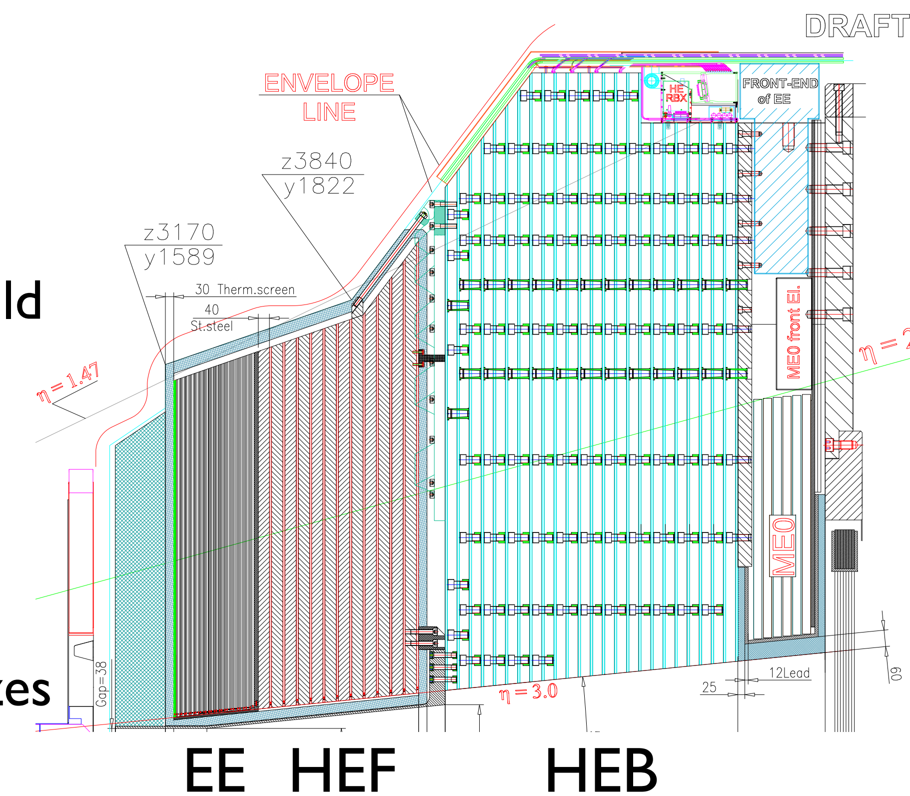
Lindsey Gray (FNAL)
on behalf of the HGCAL Software Group

1 March, 2015
USCMS Upgrade Meeting @ CalTech

The CMS High Granularity Calorimeter

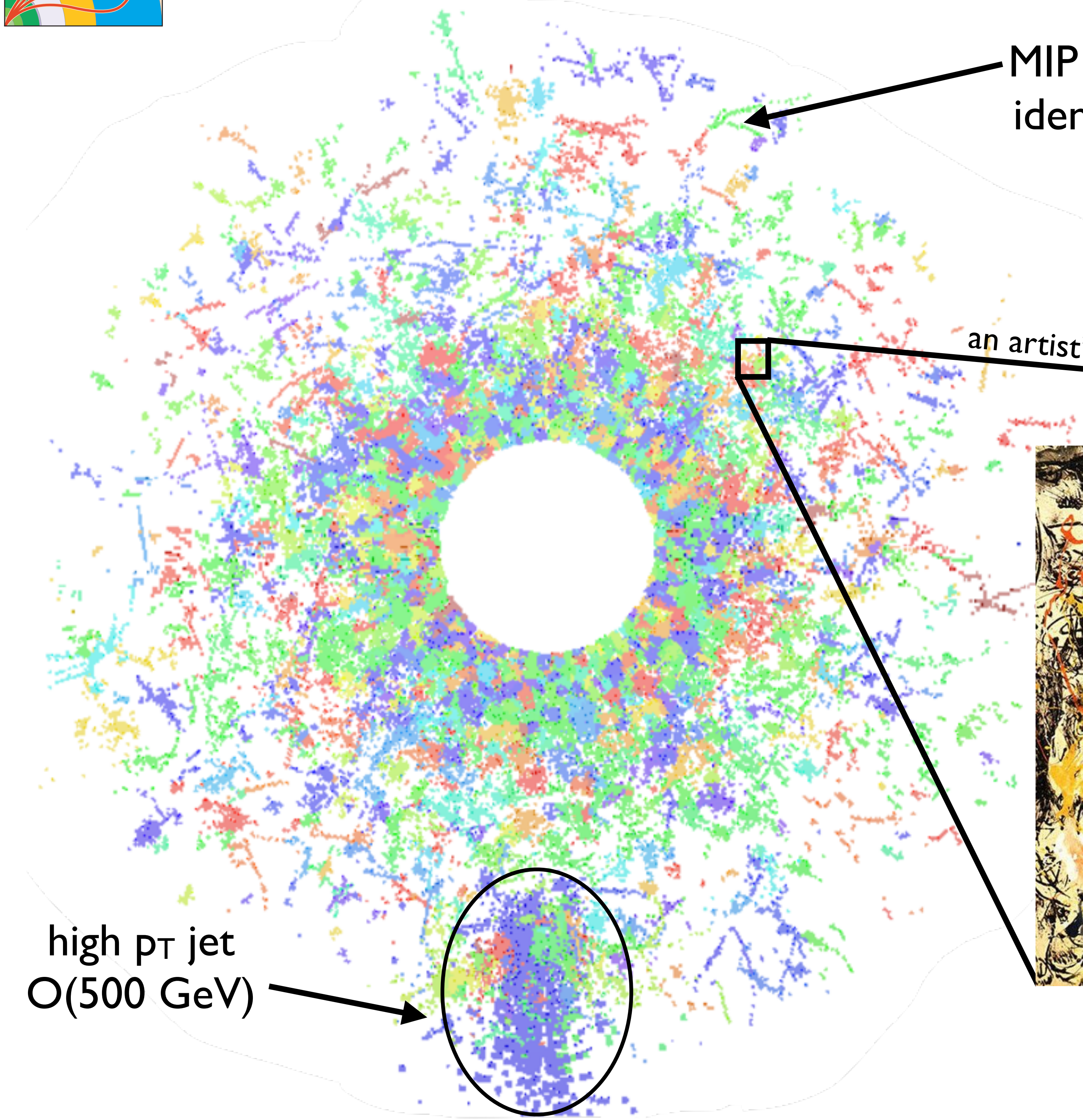


- Here I will outline only what is in our present studies
- Using HGC "V5" geometry
  - Representative of a detector we would actually build
    - ~$26X_0$ in EE, 1/3.5/6 λ_0 in EE/HEF/HEB
- Geometry validated using muon gun sample and checking expected position of hits
  - Hits found to be aligned for all layers within cell sizes
  - V5 Geometry validated for physics



Sub-detector	# layers	Structure	Cell size [mm^2]	Cell thickness [mm]
HGCEE	1	0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3.0mm Cu + 1mm Pb	10x10	0.2
	5	1.75mm W + 0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3mmCu + 1.0mm Pb + 3.0mm Cu + 0.3mm Si + 1.2mm FR4 + 2.0mm Air + 0.5mm Cu	10x10	0.2
	5	2.8mm W + 0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3mm Cu + 2.1mm Pb + 3mm Cu + 0.3mm Si + 1.2mm FR4 + 2mm Air + 0.5mm Cu	10x10	0.2
	4	4.2mm W + 0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3mm Cu + 4.4mm Pb + 3mm Cu + 0.3mm Si + 1.2mm FR4 + 2mm Air + 0.5mm Cu	10x10	0.2
	1	4.2mmW + 0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3.0mm Cu + 1mm Pb	10x10	0.2
HGCEF		15mm Stainless steel		
	12	40mm Brass + 0.5mm Cu + 2mm Air + 1.2mm FR4 + 0.3mm Si + 3.0mm Cu + 1mm Pb	10x10	0.3
		2mm Al + 2mm Foam + 2mm Al (thermal shielding) + 65mm Air gap		
HGCEB	12	34.5mm Brass + 9.0mm Scintillator or 9.0mm Brass	20x20mm	43.5

Imaging Showers with the HGCal



MIP tracks and clusters clearly identifiable by eye throughout most of detector.

high p_T jet
O(500 GeV)

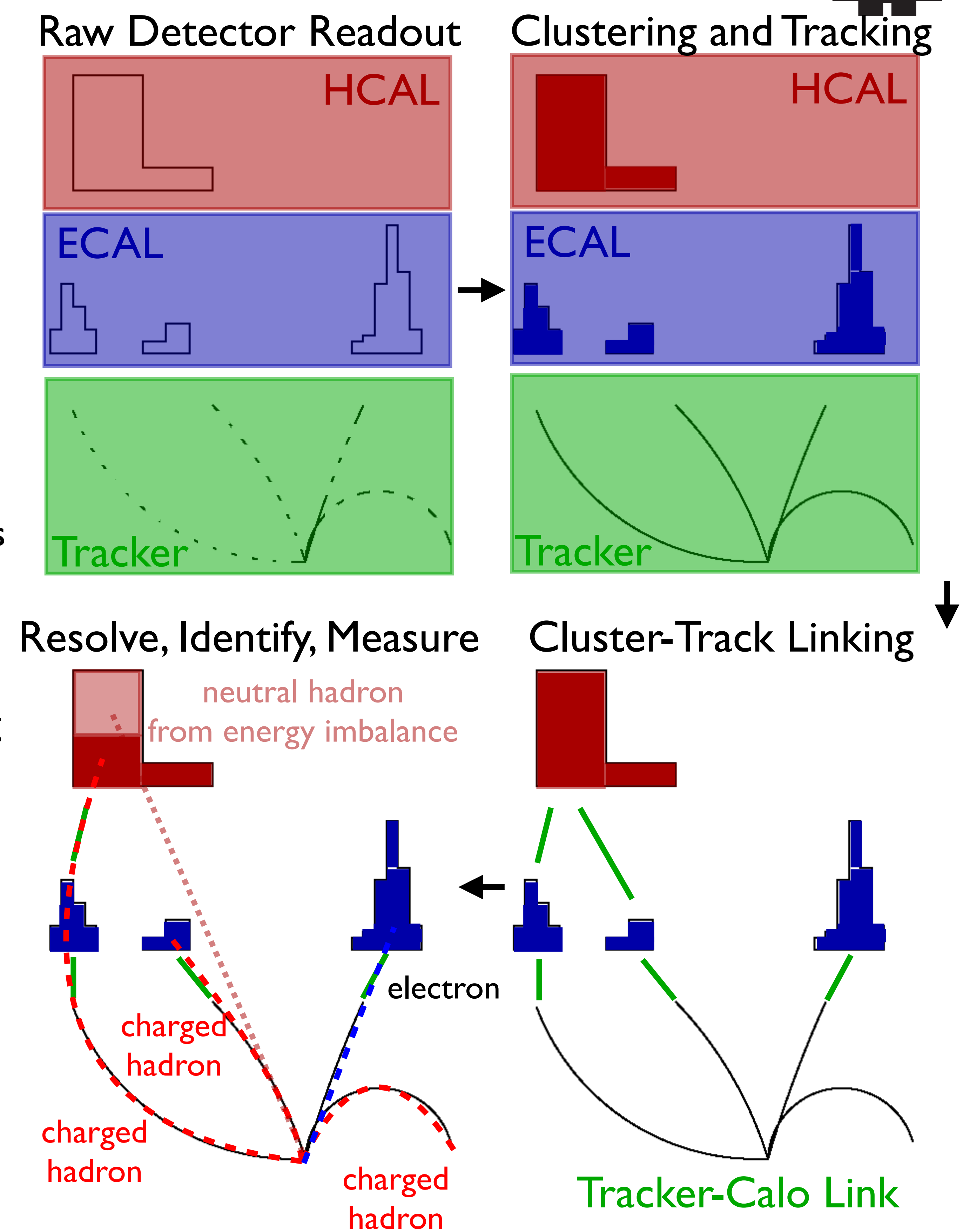
an artist's impression of the longitudinal shower footprint

J. Pollock

In this talk I will be discussing recent results with PandoraPFA and some of the road to getting there.
For results on E/γ performance and trigger please refer to previous talks a the jamborees.

What is Particle Flow?

- ◉ A reconstruction that yields an *unambiguous* list of *identified* final state particles
  - *Cluster* detector hits together in each detector
  - *Link* tracking data to calorimetric deposits
    - ~60% of particles in jets are charged hadrons
    - 30% γ, 10% neutral hadrons
    - Augment calorimeter response with tracking
  - Use of *all* detector information to *measure* and *identify* all particles in a collision
    - Optimized use of all information critical to performance
- ◉ This technique is colloquially known as “Particle Flow”

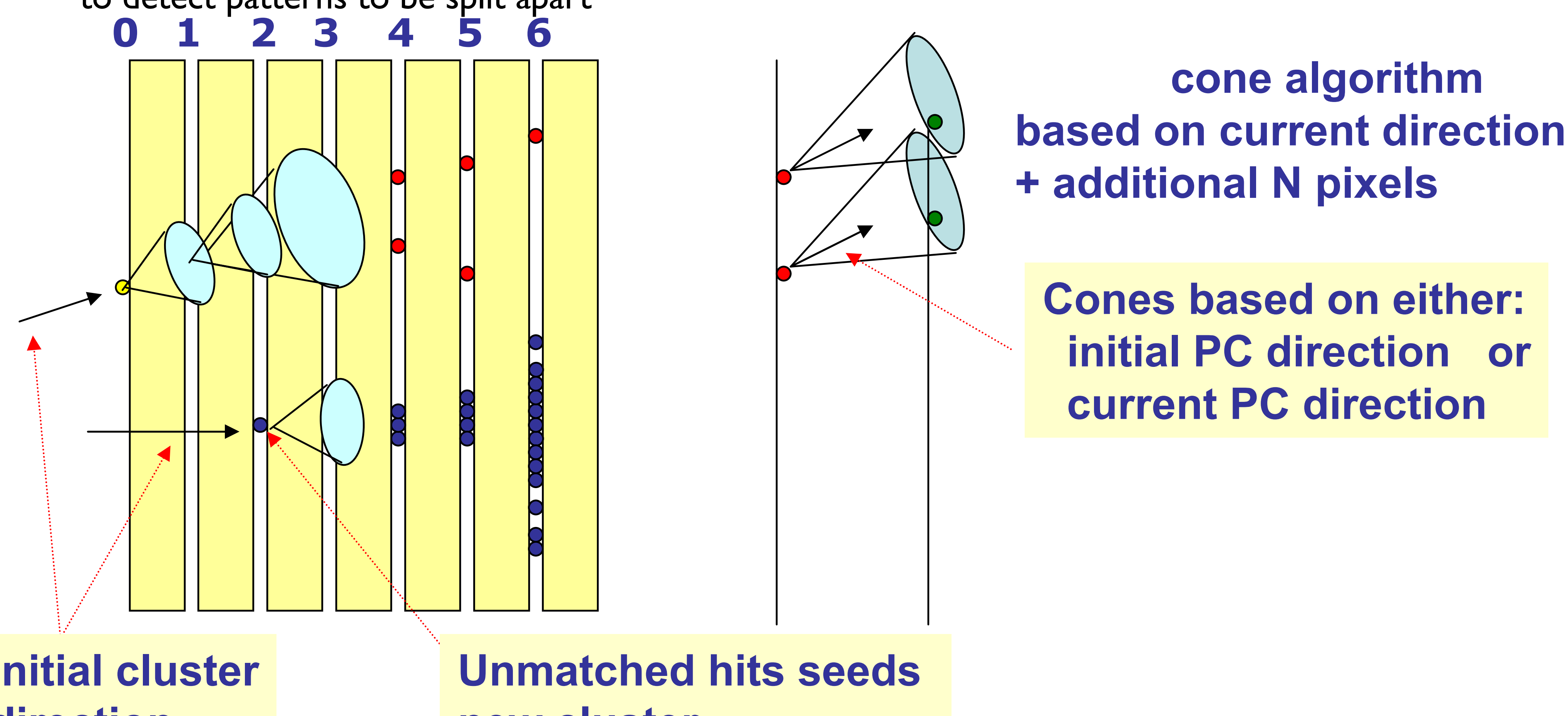


How to get to Particle Flow with HGCal

- Need to cluster together energy deposited in multiple layers
  - Energy deposits must be grouped in a way that is consistent with the formation of showers in the device
  - Any clustering must *follow* the particles propagating and showering in the detector
  - Must have a software algorithm that enables full use of the calorimeter (performance is hardware + software!)
- Have to make sure the calorimetric interpretation is stable in dense environments
  - Clusters cannot grow too large!
- Need sensible methodology for reconciling tracking information with calorimetry
  - Need to watch out for fluctuations in both calorimetry and tracking energy measurements
  - Watch out for track fakes as well!

Getting a Rough Picture: Initial Clustering

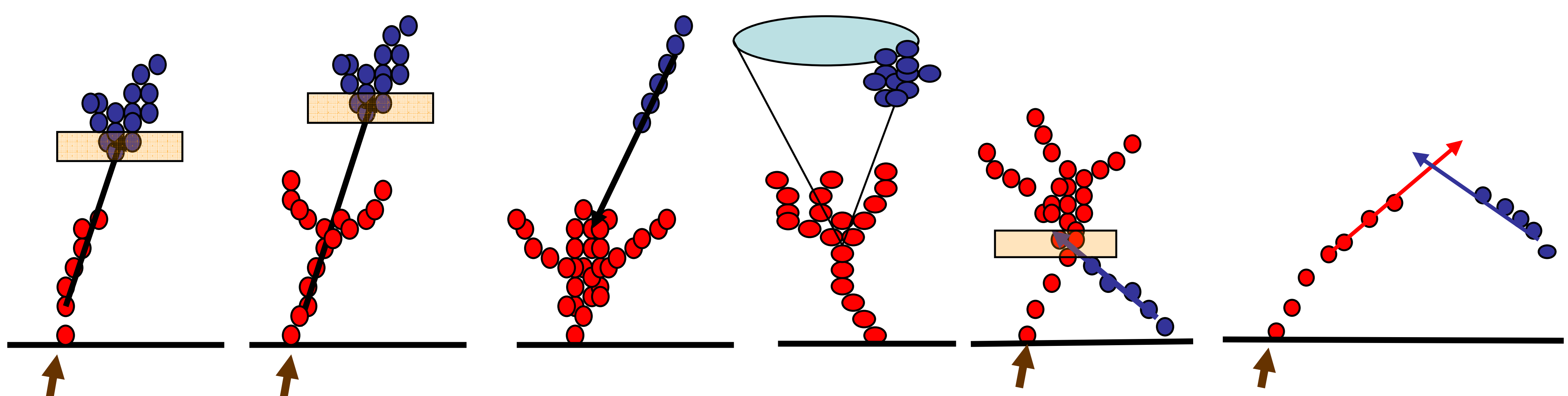
- Track seeded initial cluster positions and directions (optional)
- Loop over calorimeter hits to find nearest cluster
  - First stepping back N (default = 3) layers looking in a narrow cone ($\theta/2$ = ~8 degrees)
  - If no previous layer match, look in narrow region (1-2 pads) for cluster in same layer
  - If no match *at all* seed a new cluster with expected direction given by pointing back to IP
- By design this will fragment clusters apart, but gives first reasonable clustering to start
  - Use other algorithms to focus on putting the event back together (will talk about the main ones)
  - With this device it is easier and more efficient to detect patterns that you should merge together than to detect patterns to be split apart



From M.Thompson

Refining the Event: Topological Associations

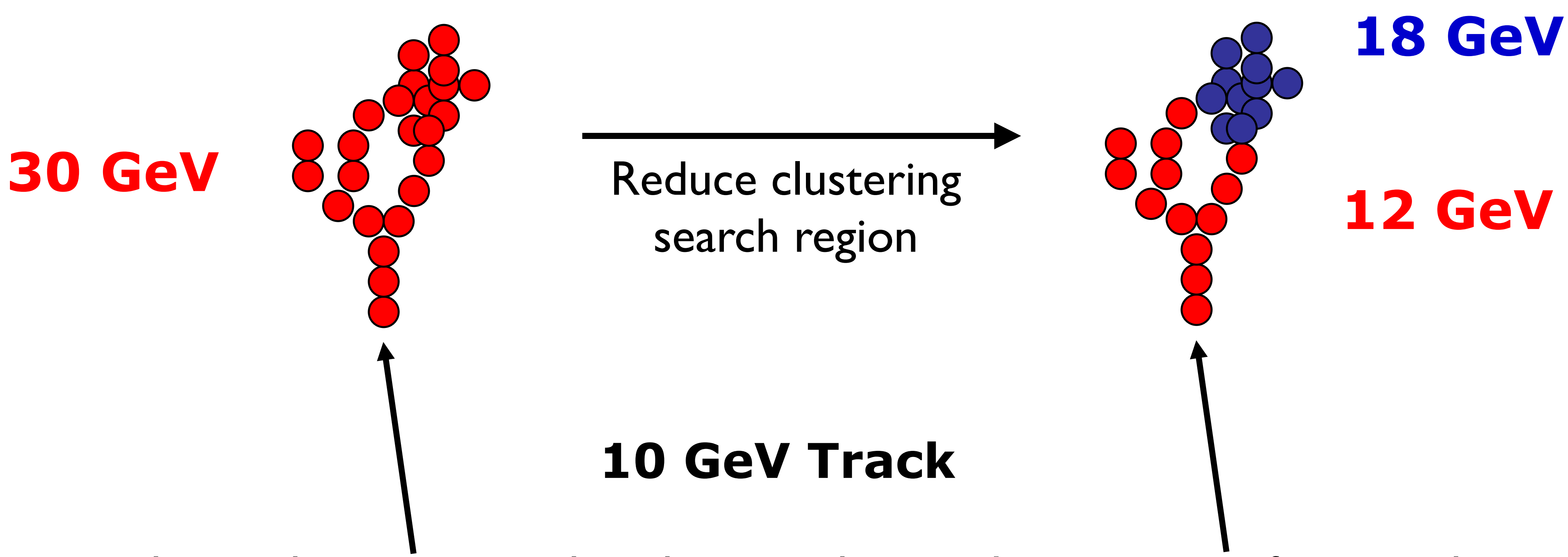
- Use the longitudinal granularity and tracking capabilities of the HGCal to gather fragmented clusters together
  - Since MIP-like clusters will point with very high precision, most cluster-cluster associations are accurate
  - Exploit in-situ cluster direction fit used during initial clustering step
  - Few mistakes at this step thanks to longitudinal granularity
- Ensure that gross mistakes for charged energy component are not allowed by requiring merged clusters to be E/p consistent with parent tracks



Forward Pointing | Back Pointing | Forward Scattered Neutral | Back Scattered | Loopers (not so relevant for EndCap)

Refining the Event: Iterative Reclustering

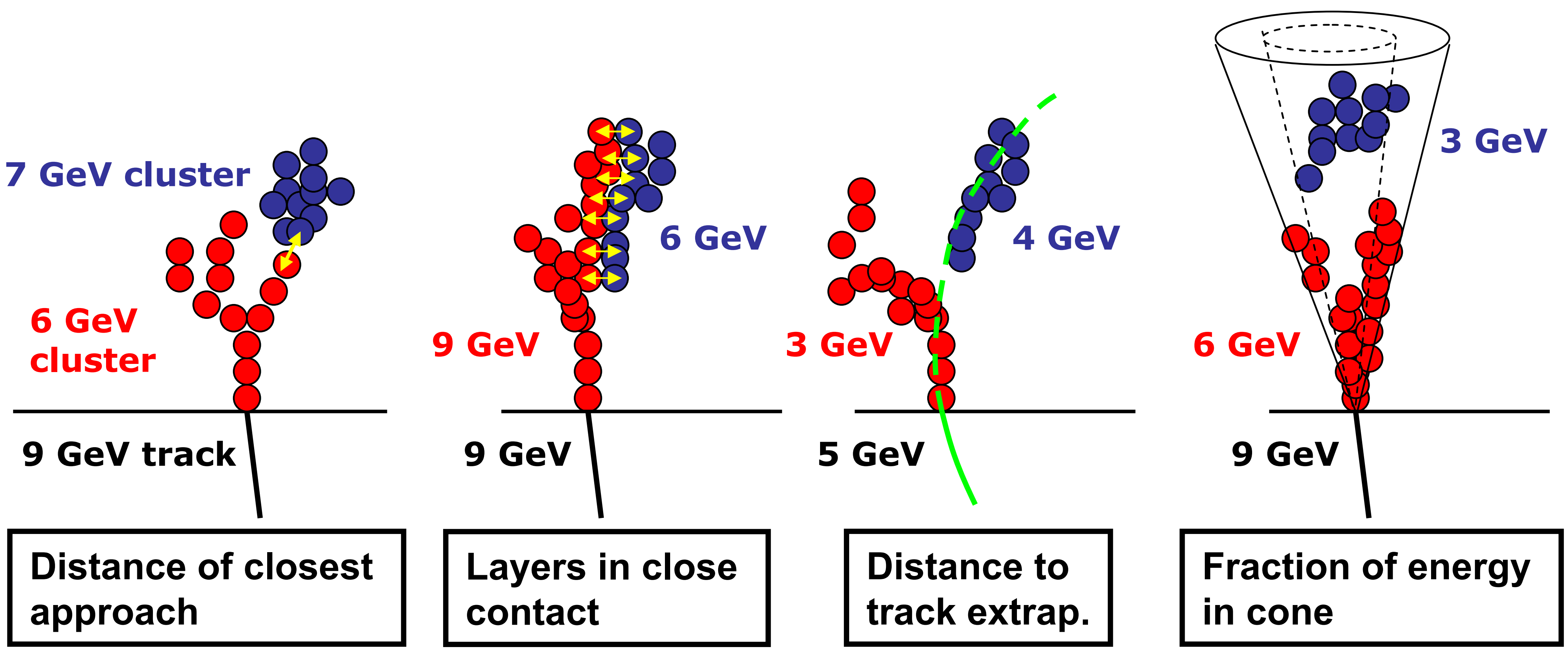
- Look at all track-cluster associations for cases where cluster contains more energy than the track
  - Typically look at 3σ deviations
    - Requires a clean set of tracks, need a priori fake rejection in CMS
  - Alter the clustering parameters, starting from coarser clustering to very narrow clustering, to attempt to break cluster into better-matching pieces
- Keep the reclustering result with the best energy balance in the local charged component
  - This is sensitive to both upwards and downward fluctuations in the cluster energy gathering efficiency (you can make a cluster bigger if track energy is much too large)



Compare to the track momentum but do not subtract cluster energy from track energy.
(This fails at large track momentum and you throw away information)
Get the best calorimeter-defined clustering with respect to input track information.

Refining the Event: Fragment Removal

- Final step of the clustering before particle flow
- Previous clustering steps naturally seed "fragments"
  - Split-off clusters on periphery of larger ones
  - A cause of double counting or "confusion" if that cluster belongs to a charged object (as energy usually taken from track)
- Look for residual topological associations
  - Clusters with shared boundaries or containment within projection of cluster envelope
  - Clusters along track propagation in calorimeter



Lindsey Gray, FNAL

How to get to Particle Flow with HGCal

- Need to cluster together energy deposited in multiple layers
  - Energy deposits must be grouped in a way that is consistent with the formation of showers in the device
  - Any clustering must *follow* the particles propagating and showering in the detector

✓ Well documented in ILD studies

- Have to make sure the calorimetric interpretation is stable in dense environments
  - Clusters cannot grow too large!

✓ Well documented in ILD studies

- Need sensible methodology for reconciling tracking information with calorimetry
  - Need to watch out for fluctuations in both calorimetry and tracking energy measurements
  - Watch out for track fakes as well! (req. optimization @ CMS)

~✓ Well documented in ILD studies

OK, That's Great: Can it work in 140PU?

- CLIC pileup scenarios are a much more forgiving environment than 140PU @ LHC
  - Occupancies are at least ~3x less
    - HGCal endcap has finer granularity then ILD design
  - Much more data to process and associate in 140PU at a proton machine
  - Still, original algorithms took 2 minutes per event at CLIC pileup
    - This might be bad…
- Indeed, using PandoraPFA out of the box in 140PU takes about 1 hour per event
  - Not going in the HLT as it stands…
- However, there exist tools and ways of thinking about this problem from computational geometry and graph theory that can mitigate or entirely remove the underlying performance bottlenecks
  - Will give an overview of this
  - This is critical to getting sensible result in 140PU
    - Cutting away Calo Hits and Tracks makes it harder to do particle flow in busy environments… avoid this as much as possible!

Computational Geometry & Graph Theory

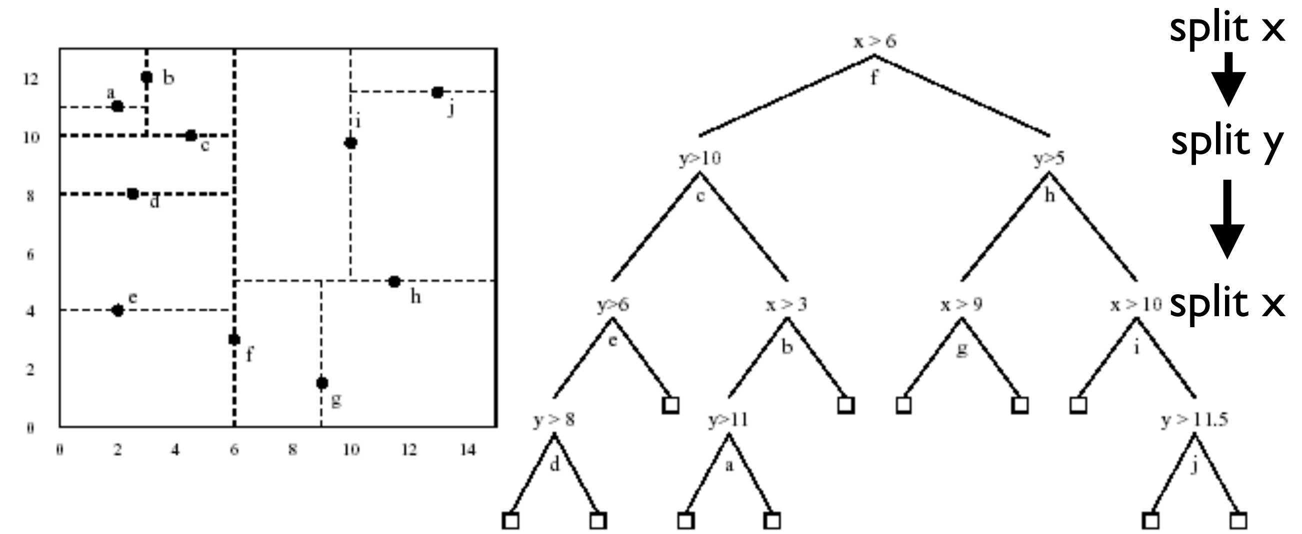
Computational Geometry

- Study of the algorithms that associate points in space
  - Nearest neighbour searches (core of any clustering)
  - Hull finding (get set of outermost points)
- Results from computational geometry speed up these operations by orders of magnitude in typical cases
  - N^2 -> N*log(N) to search neighbors of hits in a region

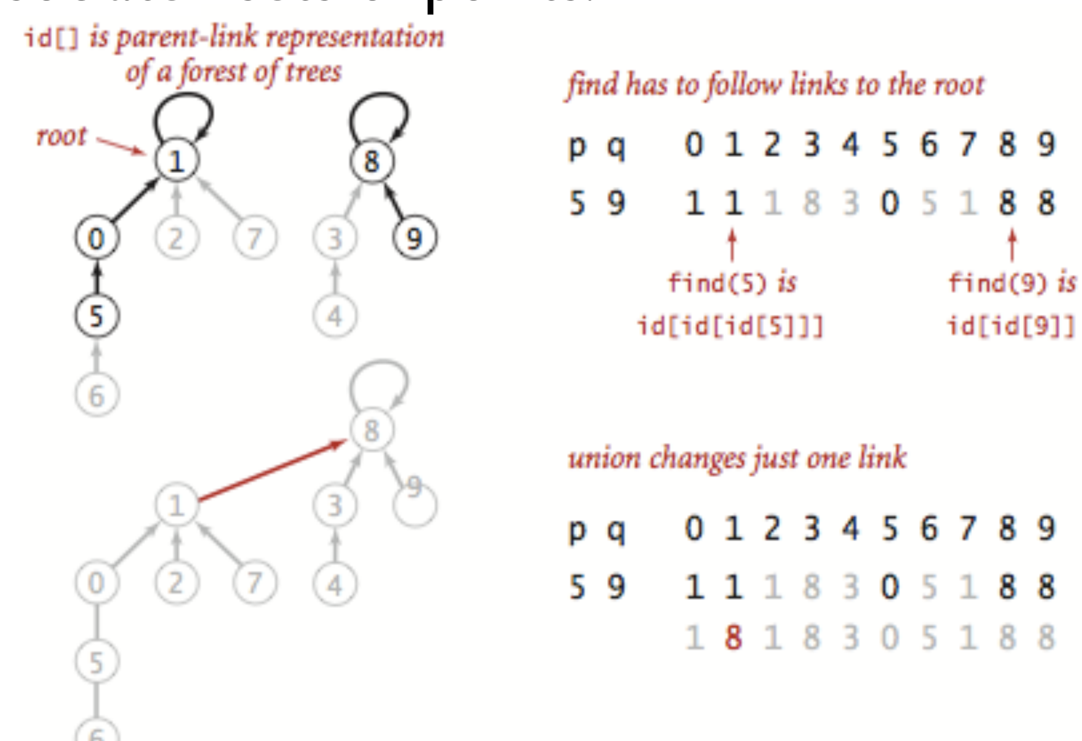
Graph Theory

- Provide efficient way to manage associated sets of points
  - Represent hits as a disconnected graph to start (a hit is a vertex of the graph)
  - Associate hits = building edges in the graph between vertices
- No need to search over and over again for the association you are adding a hit to
  - N^2 -> N*iterlog(N) (almost linear)

A kd-tree in 2 dimensions:



QuickUnion efficiently represents associated sets of points:



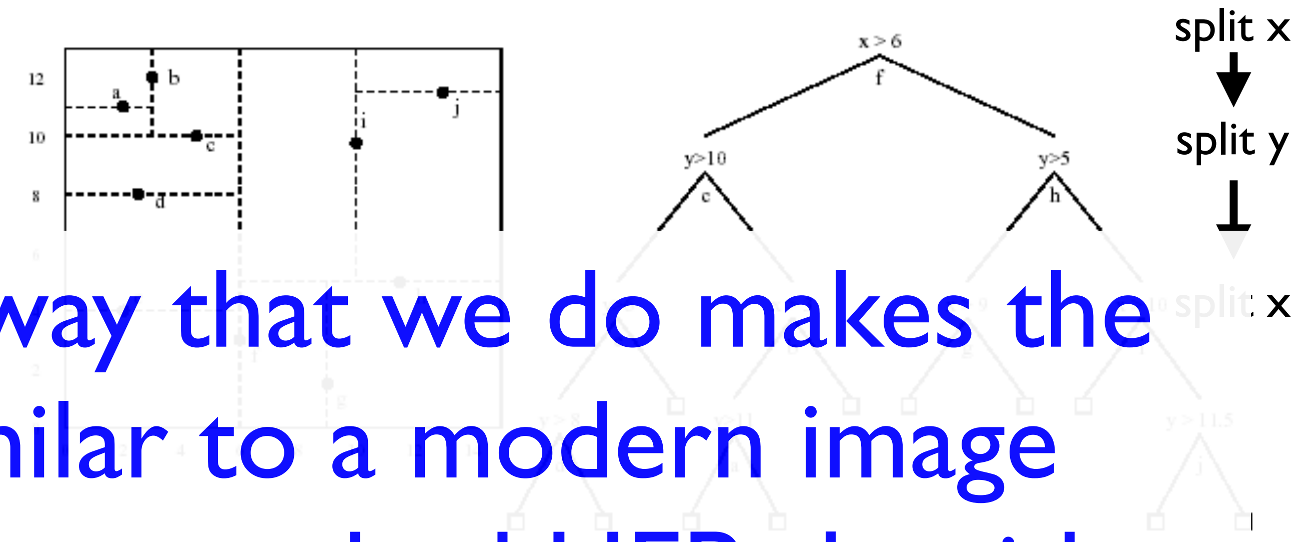
Quick-union overview

Computational Geometry & Graph Theory

◉ Computational Geometry

- Study of the algorithms that associate points in space
  - Nearest neighbour searches (core of any clustering)
- operations by orders of magnitude in typical cases
  - N^2 -> N*log(N) to search neighbors of hits in a region

A kd-tree in 2 dimensions:



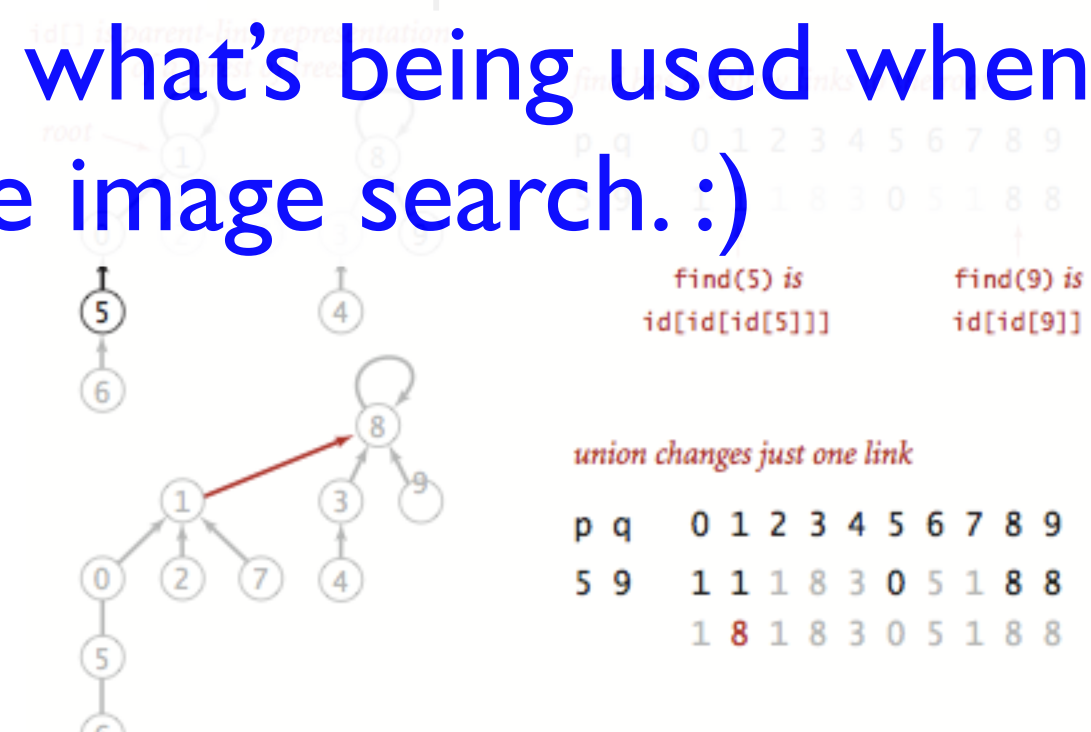
Using these algorithms the way that we do makes the HGCal software more similar to a modern image classification algorithm than any standard HEP algorithm.

◉ Graph Theory

QuickUnion efficiency represents associated sets of points:

These algorithmic concepts are what's being used when you do a Google reverse image search. :)

- points
  - Represent hits as a disconnected graph to start (a hit is a vertex of the graph)
  - Associate hits = building edges in the graph between vertices
- No need to search over and over again for the association you are adding a hit to
  - N^2 -> N*iterlog(N) (almost linear)

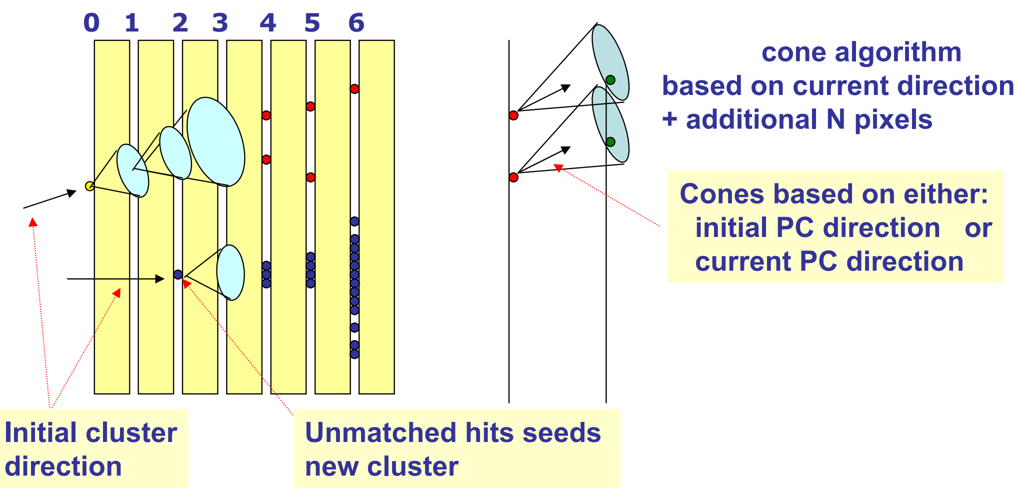


Quick-union overview

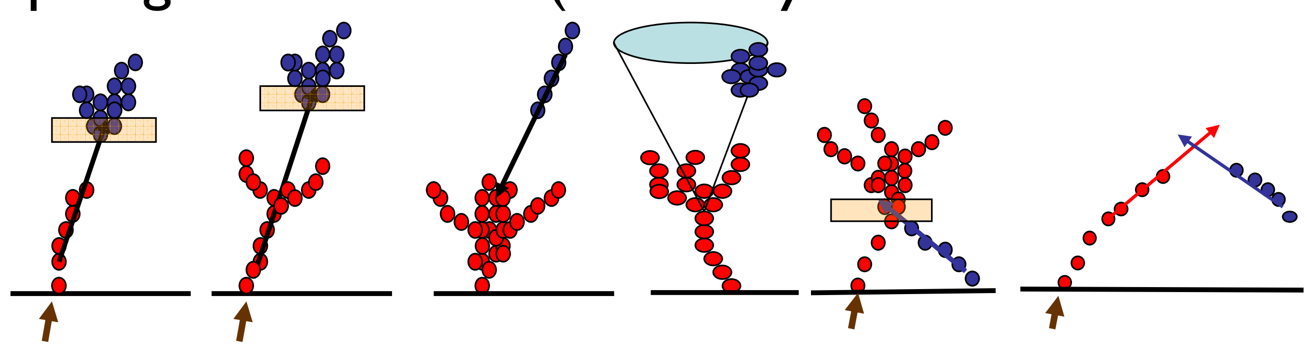
What do these algorithms enable at HL-LHC?

- These algorithms can clearly bring large gains
  - In less pressured times I have demonstrated factors of 100x reduction in algorithm timing in 140PU
    - for ARBOR or our home brew clustering
- Typically requires calm thought to get the best gains
  - We had about 1.5 weeks to overhaul much of pandora
    - Gains are there but can do better!
  - PandoraPFA steps must be rethought at a low level to take full advantage of these algorithms
- 1 hour/ev -> 10 minutes/ev
- These techniques enable physics reconstruction algorithms that exploit the full potential of HGCal in 140PU
  - New algorithms validated and show no physics performance degradation for great gains in speed in LHC 140PU

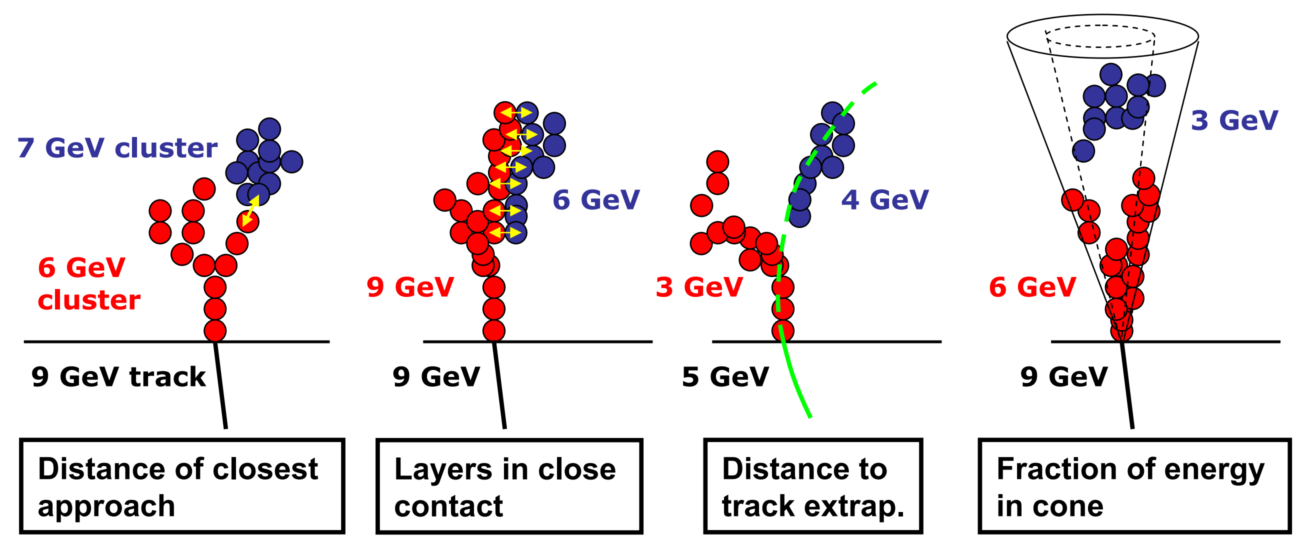
Cone clustering: 10x-20x (good!)



Topological Assc.: ~3x (have only fixed bad instances)

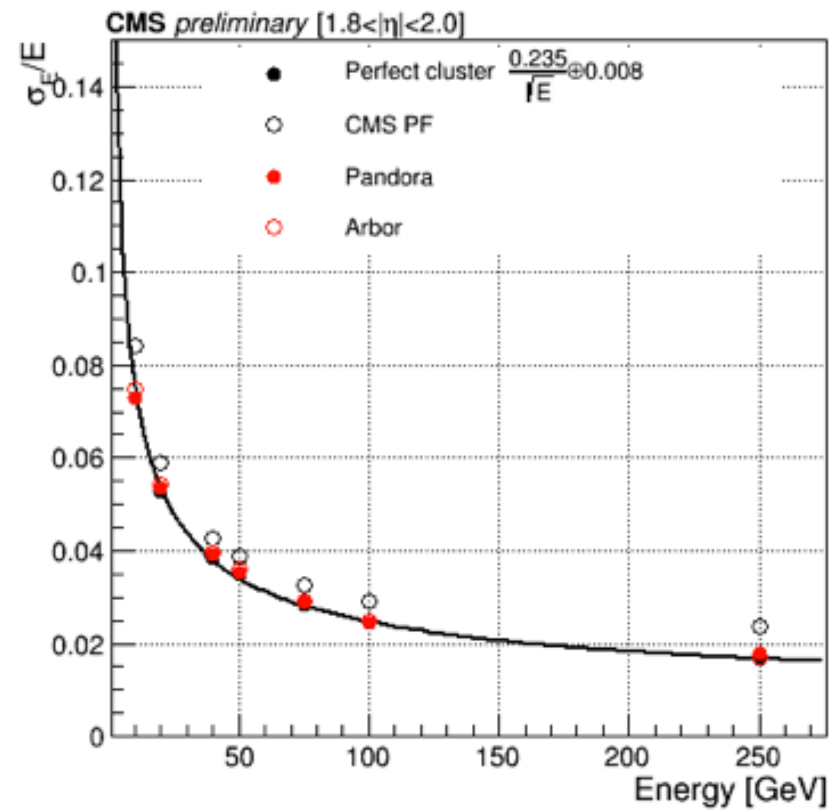


Frag. Removal: 2x (can definitely do better)

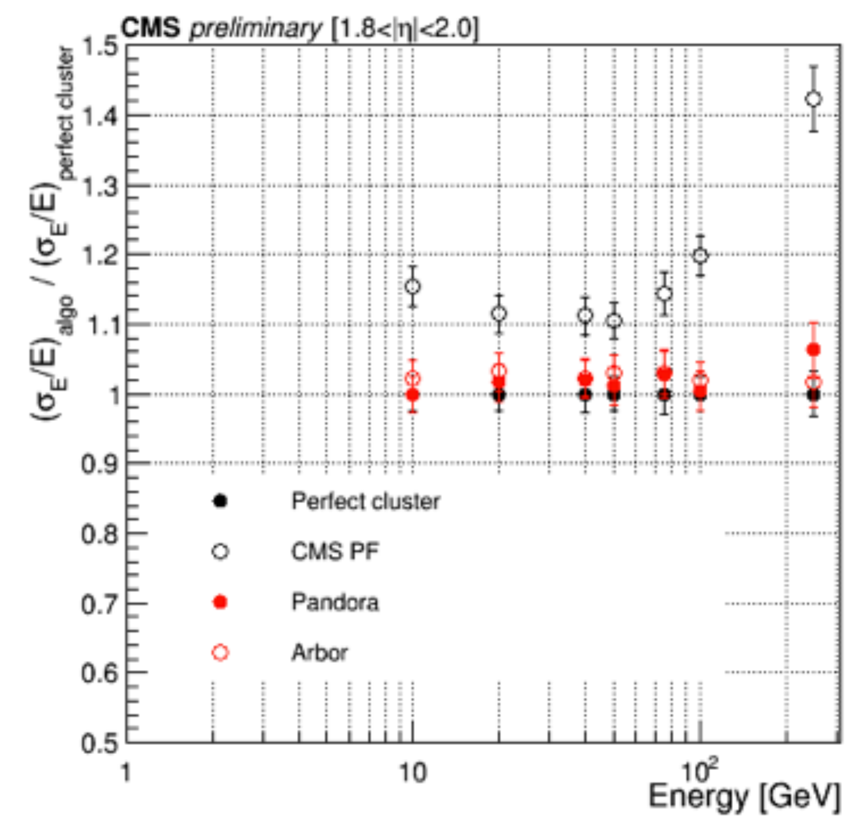


Performance: γ Calibration

- Calorimeter calibration is fundamentally important to doing particle flow in HGCal
  - Pattern recognition is bootstrapped off of energy measurement!
    - In reclustering, fragment removal
  - Particle IDs also use energy containment and expected longitudinal profiles
    - Measuring the correct energy is deeply important!
- PandoraPFA performs very well in gather energy
  - Achieves results near "perfect clustering"
    - Matching RecHits to SimHits

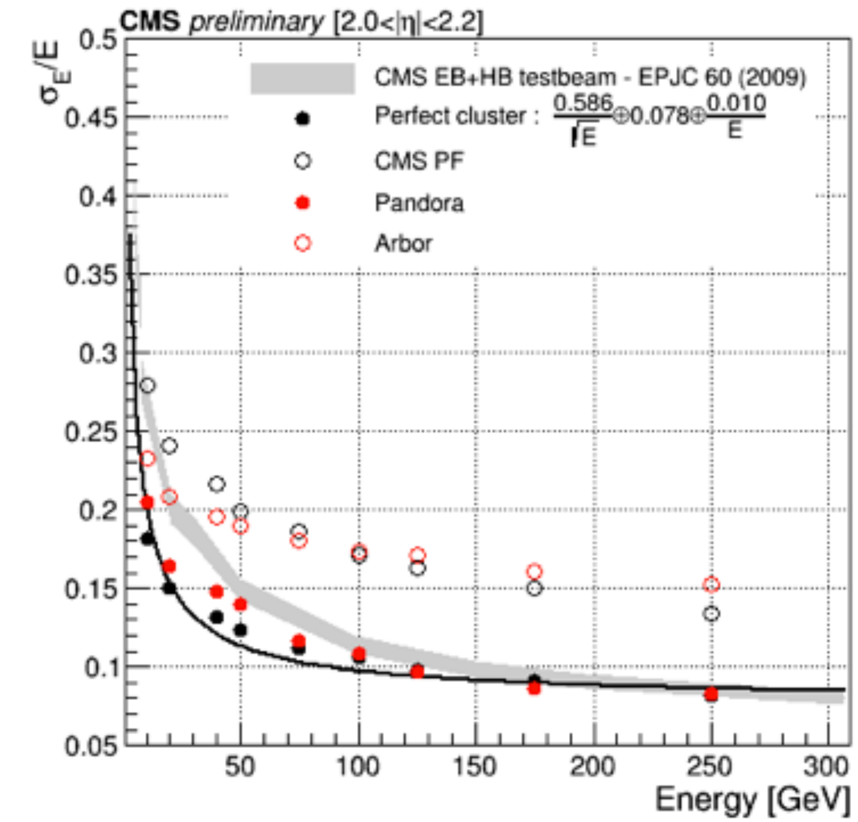


Energy reconstructed using weights-per-layer: $E_{\rm rec} = \sum_{i=1}^{30} X_{0,i} \cdot \frac{E_i}{\rm MIP}$

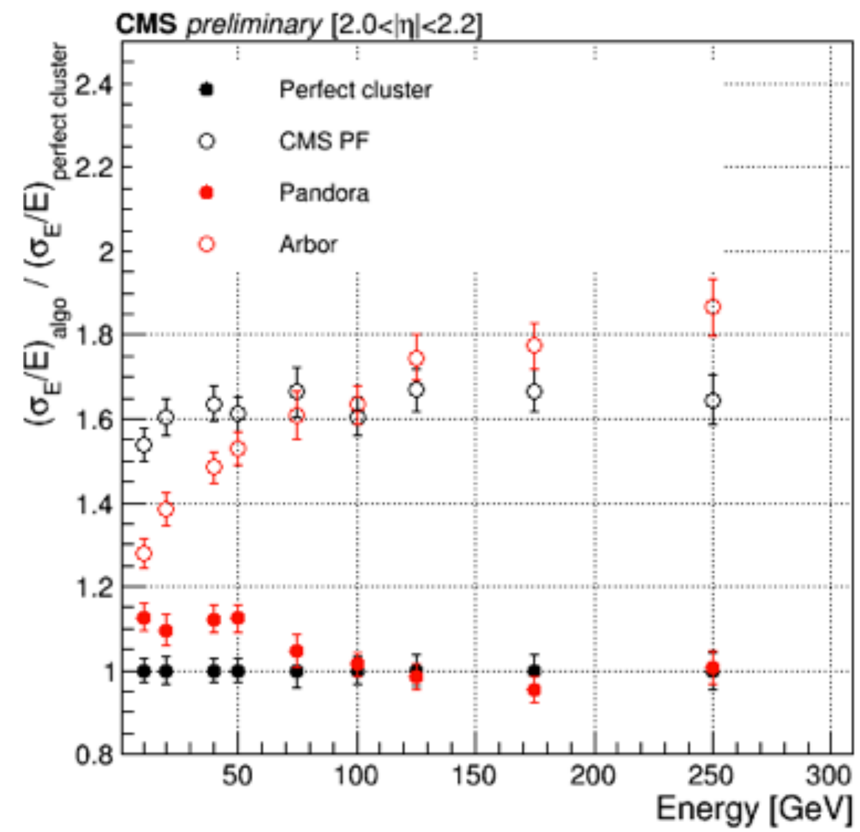


Performance: Hadron Calibration

- Again, Pandora clustering achieves good performance
  - On par with perfect clustering (grouping Calo Hits by sim-hit match)
  - Before any additional corrections ~60% stochastic term
    - Investigating source of residual constant term
    - Can improve the resolution further by exploiting fine-granularity sampling of the shower to tag EM-rich showers
- With established good performance we can apply software compensation techniques to improve this performance
  - Next slide

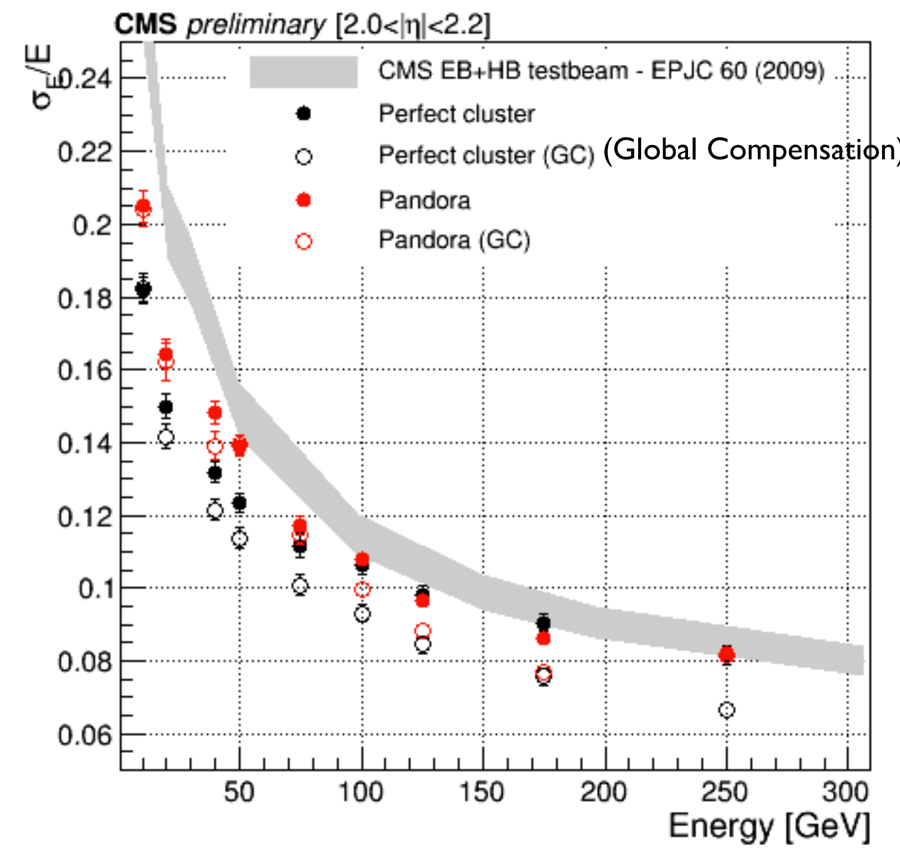


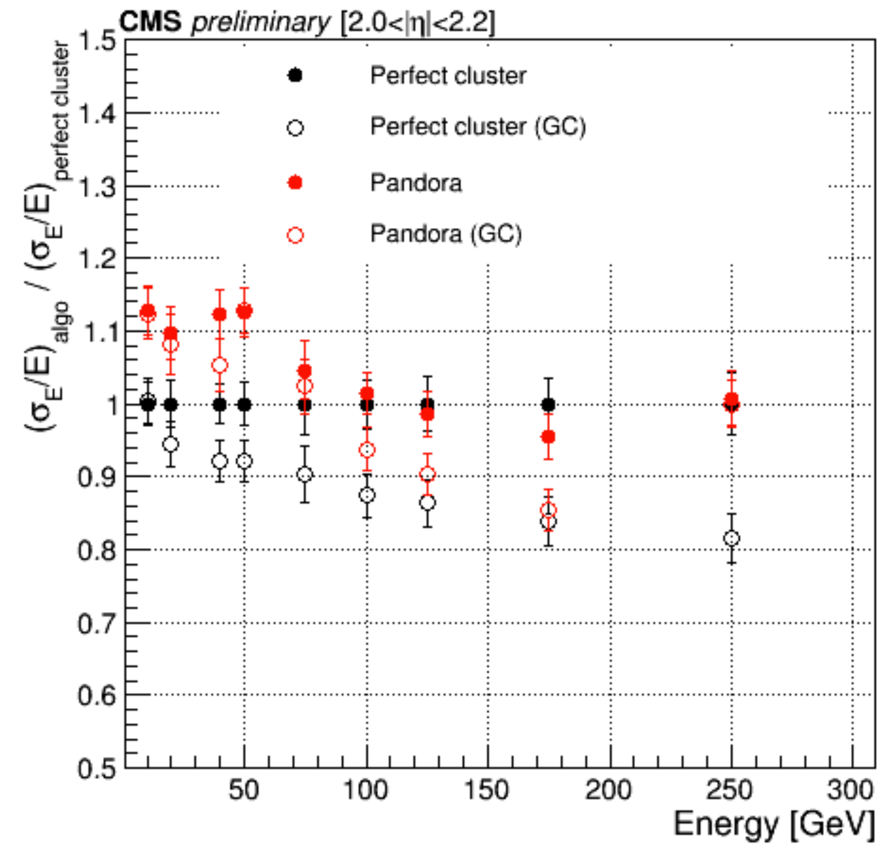
Energy reconstructed using weights-per-layer, calibrating to EM scale in each detector and then the hadron scale.



Exploiting Granularity with Software Compensation

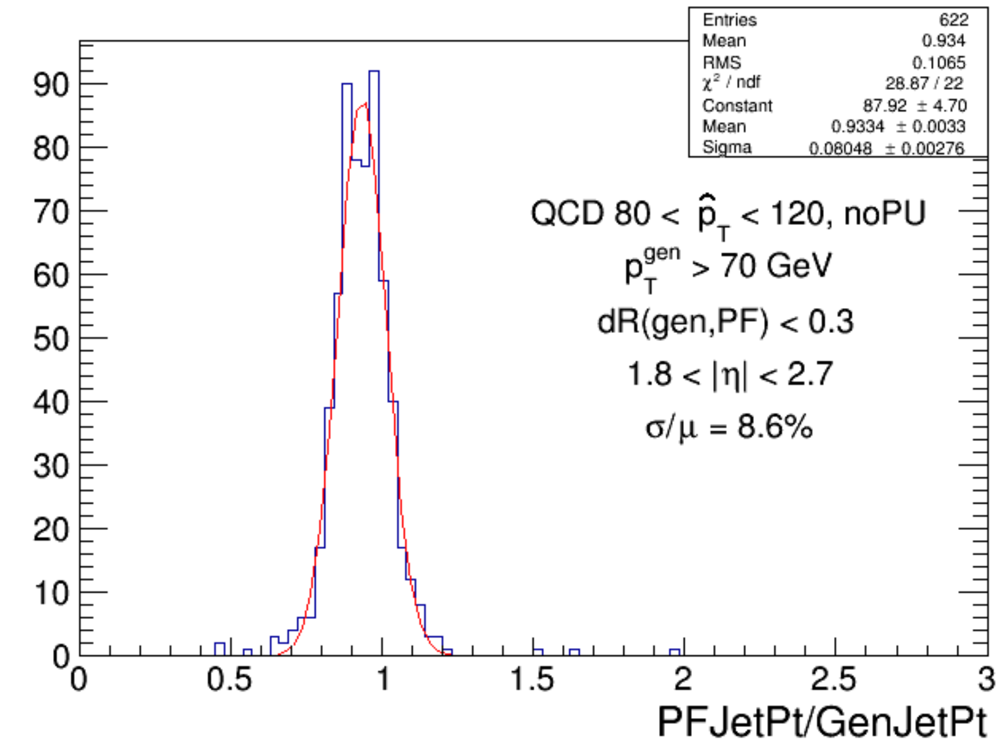
- Each hadron shower in the event fluctuates and can have more or less EM composition
  - π^0 production
  - EM cells are typically 'hotter' than the bulk of the hadron shower comprised of MIPs
  - This can result in an addition to the resolution
    - Can be corrected for with fine sampling of the shower
- Count number of hits below and above a "MIP-like" threshold
  - Correct energy by the ratio of hits that fail to hits that pass
  - Gives an educated guess as to how close the shower is to the EM scale instead of the hadronic scale



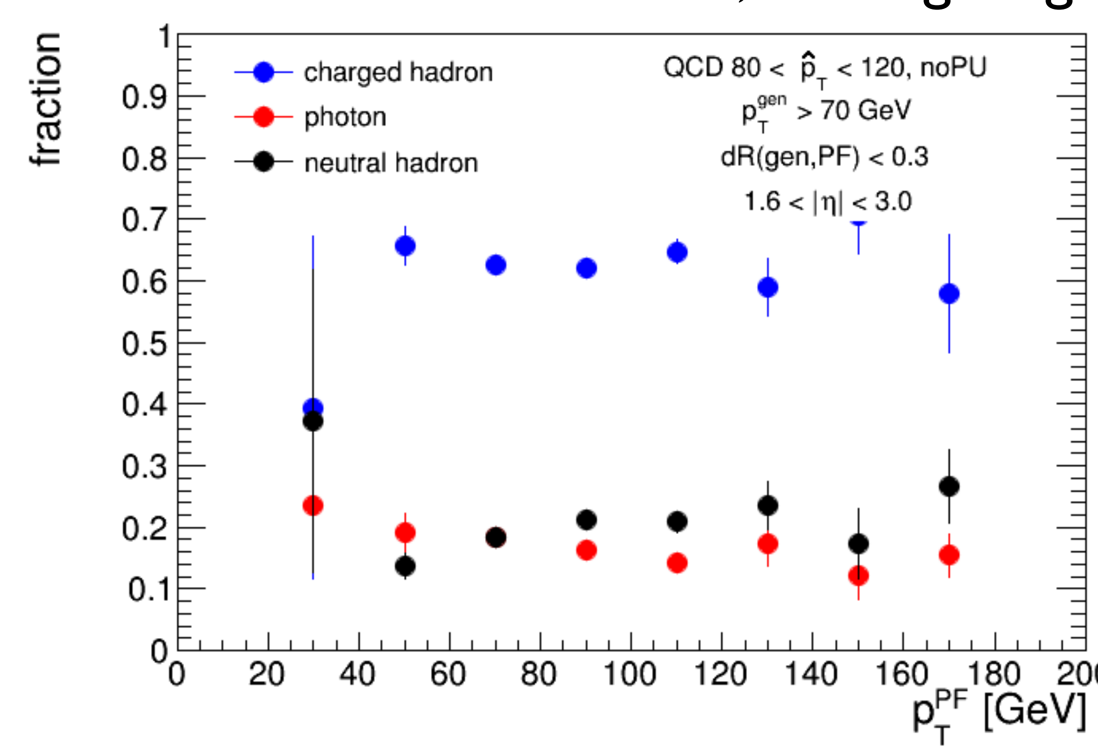


Jet Performance in 0 PU

- First results for jet energy performance look promising
    - 8.6% energy resolution when integrated with the barrel particle flow
    - 10% if you consider the endcap by itself
        - Barrel PF catches soft component!
- Some issues to chase down in the software
    - Investigating hadron calibration and photon ID, the likely culprits in poor particle ID
- No JECs are calculated or applied
    - The resolution of the integrated distribution shown can be improved by this
    - Not enough stats to make fully differential plots at present

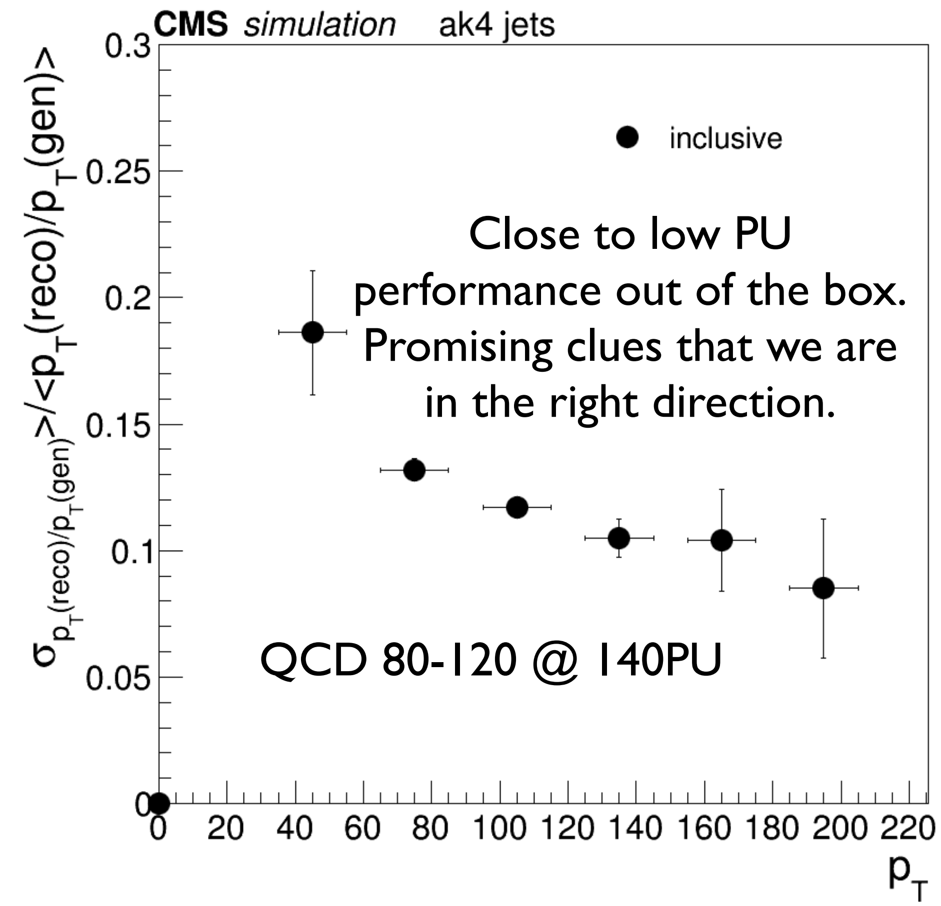


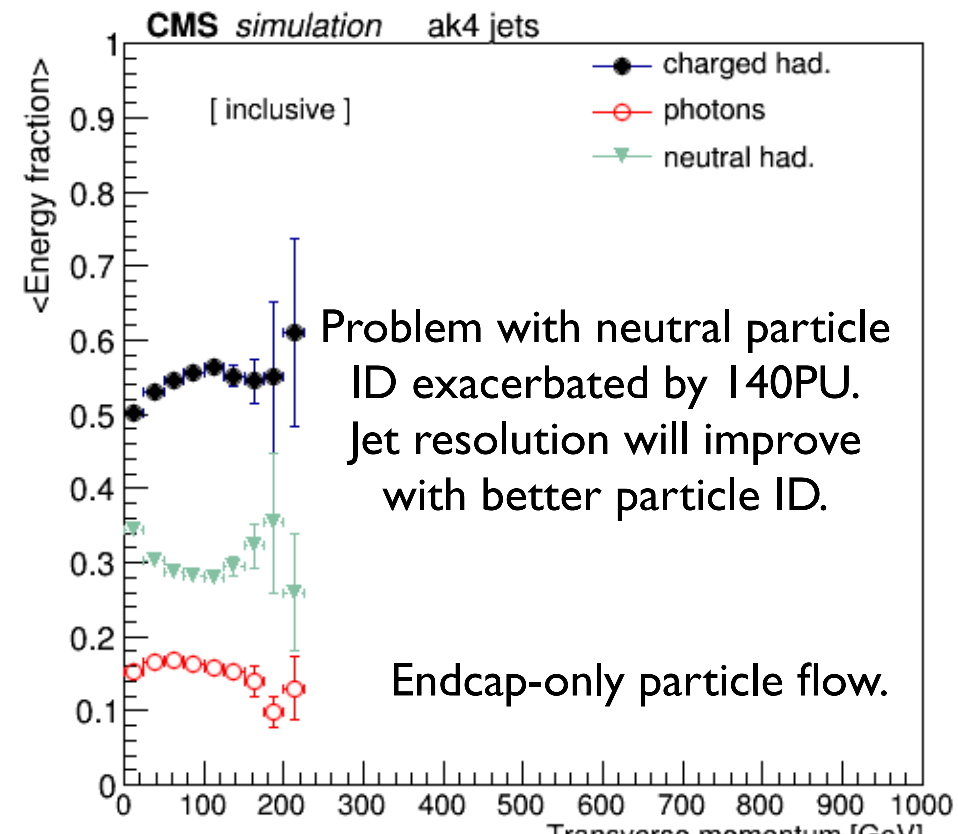
Photon fraction should be ~30%, investigating



Jet Performance in 140 PU (extra fresh)

- With the integration of PandoraPFA we can meaningfully study jets in 140PU
  - This is the first time we have been able to do this and the first time this is being shown
  - We take the present performance as a good indicator that the concepts behind the algorithms scale well to high PU
    - and of course the underlying device is robust as expected
    - Still need to understand some details of interplay of tracker at 140PU with this calorimeter
  - Not a final answer
- Please take care, these plots are very fresh
  - We are still coming to understand this!





Conclusions

- Performance of HGCal device is close to expected
  - 23.5% EM resolution, less than 1% constant term
  - 50-60% hadron resolution, 8% constant term (expect ~5%)
  - Calibration at the clustering level matches that of perfect clustering (critical for PF)
- PandoraPFA has been deployed to success
  - PandoraPFA updated to be a modern image processing algorithm, cutting edge in HEP
    - Techniques from computer science adapted for use in HEP
  - Performance of jets in zero pileup is very promising (perf
    - Event constituent reconstruction is close to expected, looking into photon identification presently
  - ~13 minute per event reconstruction time for full RECO chain is delaying 140PU results
    - but we have a good hint that things start to look good in 140PU
    - Will continue to improve algorithmic efficiency over time
  - Software quality is critically important for the performance of this device
- Higher statistics, more in depth studies of jets in pileup are waiting for the samples
  - Stay tuned!