PFA Status at Iowa

Mat Charles
The University of Iowa

Foreword

(The non-algorithmic bit)

- Code is in CVS.
  - Currently unstable & living in contrib `org.lcsim.contrib.uiowa.ExampleRunAndWriteOutPFA`
  - Several other people running it and looking at output -- thanks for being the guinea pigs!
  - Plan to release a "stable" version in medium-term that people can work on, even if performance is not terrific.
  - Uses PFA template/framework

Algorithm overview

Most PFAs work in roughly the same way:

- Find photons & identify them
- Extrapolate tracks to the calorimeter
- Find charged clusters & associate them to tracks
- Call the big leftover pieces neutral hadrons
- Calibration for EM and hadronic showers

Take all that as read. What are the distinctive features of this PFA?

Internal structure of hadronic showers

Hadronic showers have 3+1 basic components:

- A dense clump of hits (EM core)
- MIP-like trails/tentacles (secondary charged particles)
- Small, displaced fragments/satellites (secondary neutrals)
- If charged, probably a MIP trail coming in

So we start by looking for these components:

- MIP-like clusters: (semi)isolated hits in subsequent layers
- Clumps: clusters where the local hit density is high

... and then we'll combine them into shower skeletons

Building shower skeletons

- Start with generous "envelope" clustering to reduce combinatorics. (3cm MST)
  - Problem: What about fragments that are far away?
- Look at pairs of components & compute a likelihood based on geometric/topological variables:
  - MIP-MIP
  - MIP-clump
  - ... but not clump-clump (too little information)
- If likelihood is above threshold, accept the link.
- Use links to build the components into skeletons

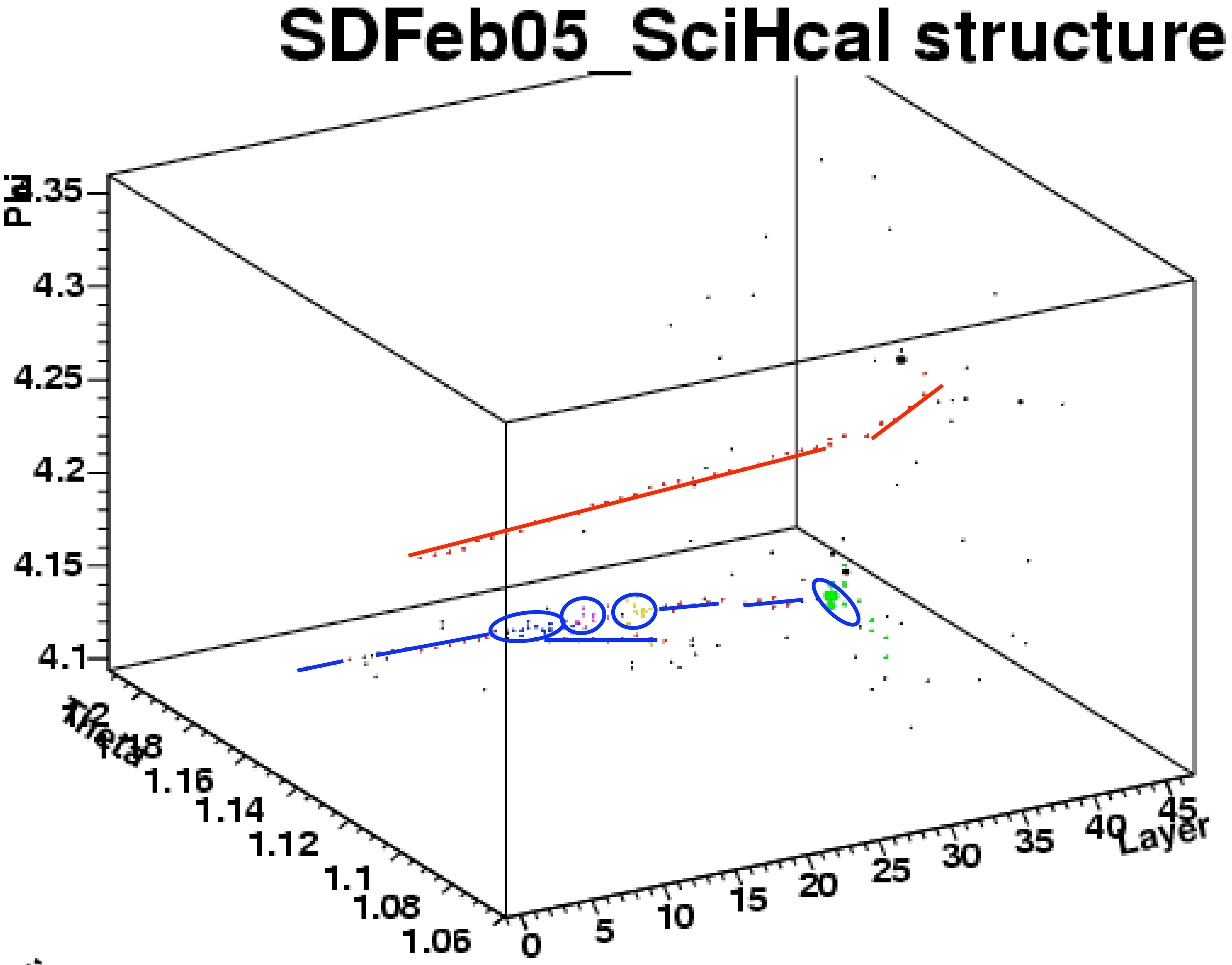
Likelihood variables

- MIP-MIP links:
  - Distance of closest approach (DOCA)
  - Whether point of closest approach (POCA) is in calorimeter
  - Smallest distance from a cluster hit to POCA
- MIP-clump links:
  - DOCA (track to cluster center of energy)
  - Smallest 3D distance from MIP hit to clump hit

The likelihood distributions are "trained" on simulated events from same detector, then stored in conditions database. Retuning on a different detector just means one person running a batch job once.

An old example

This $K_s \rightarrow \pi^+\pi^-$ event was made on a very old detector but illustrates the approach:



MIP-like segments are approximated as lines.

MIP-clump pointing is done with clump center of energy.

Other important features

Matching tracks to clusters

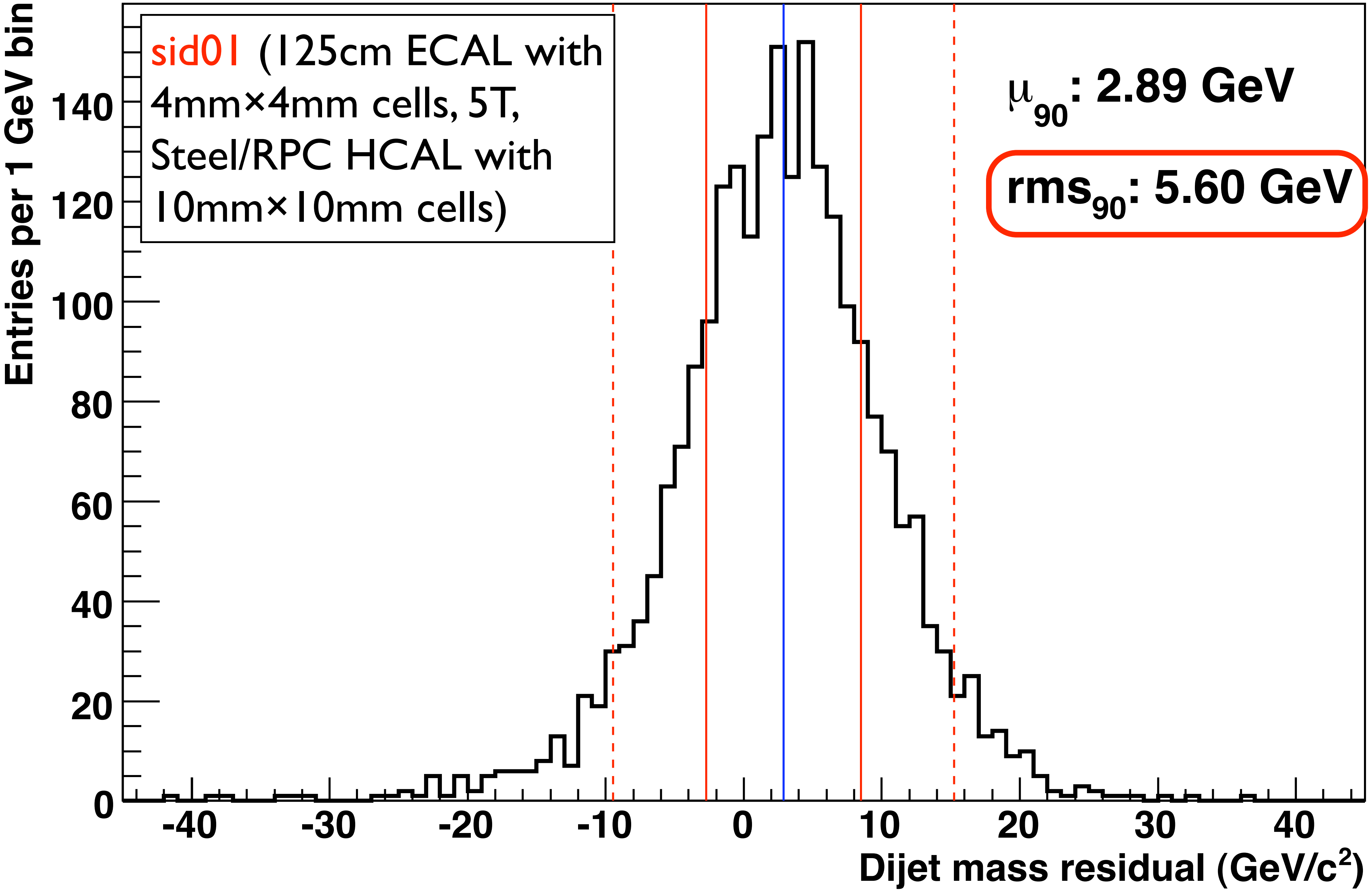
- Cheating, though trying to be somewhat realistic about it
- Use Ron Cassell's list of reconstructible final-state tracks
- Estimate ECAL entry point with local helix extrapolation from outermost tracker hits
- Match to MIP stub preferentially (distance < 3cm, dot-product of directions > 0.85)
- Otherwise match to any nearby cluster (distance < 3cm)

Sanity checks:

- If >1 track attached to skeleton, tighten L cuts until fixed.
- Require E/p cut (3σ), otherwise throw out track & treat cluster as neutral. (Trade confusion for resolution.)

Current performance

Looking at $e^+e^- \rightarrow$ Z(qq) Z(vv) @ 500 GeV for q=uds and computing $m_{reco} - m_{true}(Z)$, i.e. residuals:



Needs to be 3-4 GeV for physics...

How to do better?

Tried a number of things:

- Better track extrapolation -- helped with perfect CAL pattern recognition, but not for mine;
- Using better photon clustering (Ron Cassell) with real photon ID (H-matrix by Graham/Steve/Norman/Ron)
- Cheating on track-cluster matching -- actually made the overall resolution worse

... what? How can cheating make the resolution worse? Both Ron & I saw this with independent implementations, so it's a real effect. What's going on?

What’s going on?

		Efficiency	Purity
Real track matching	Charged	58%	94%
	Neutral had	78%	30%
	Photons	78%	81%
Cheat track matching	Charged	84%	81%
	Neutral had	34%	39%
	Photons	78%	81%

What's going on?

This is somewhat OK (track fails E/p cut and is treated as neutral)

		Efficiency	Purity
Real track matching	Charged	58%	94%
	Neutral had	78%	30%
	Photons	78%	81%
Cheat track matching	Charged	84%	81%
	Neutral had	34%	39%
	Photons	78%	81%

What’s going on?

This is somewhat OK (track fails E/p cut and is treated as neutral)

		Efficiency	Purity
Real track matching	Charged	58%	94%
	Neutral had	78%	30%
	Photons	78%	81%
Cheat track matching	Charged	84%	81%
	Neutral had	34%	39%
	Photons	78%	81%

This is bad: a lot of neutral energy is sucked into charged clusters

What's going on?

This is somewhat OK (track fails E/p cut and is treated as neutral)

		Efficiency	Purity
Real track matching	Charged	58%	94%
	Neutral had	78%	30%
	Photons	78%	81%
Cheat track matching	Charged	84%	81%
	Neutral had	34%	39%
	Photons	78%	81%

This is also bad: charged fragments dumped into neutrals & double-counted

This is bad: a lot of neutral energy is sucked into charged clusters

What's going on?

This is somewhat OK (track fails E/p cut and is treated as neutral)

		Efficiency	Purity
Real track matching	Charged	58%	94%
	Neutral had	78%	30%
	Photons	78%	81%
Cheat track matching	Charged	84%	81%
	Neutral had	34%	39%
	Photons	78%	81%

This is also bad: charged fragments dumped into neutrals & double-counted

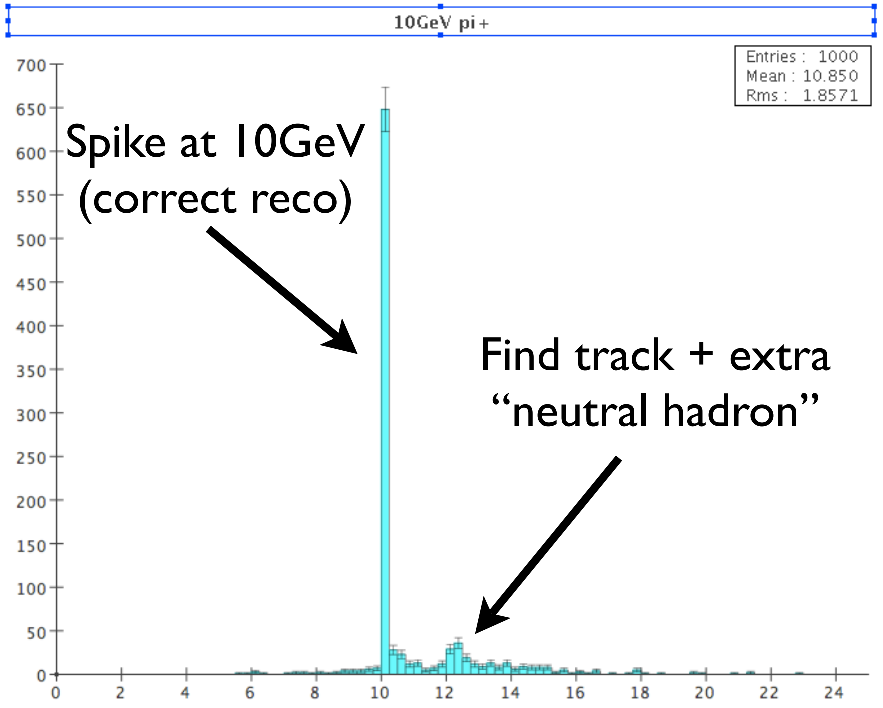
This is bad: a lot of neutral energy is sucked into charged clusters

The E/p cut is hiding [some of the] problems in clustering.

Diagnostics: single-particle events

Here are a bunch of single-particle 10GeV events

10GeV π^+



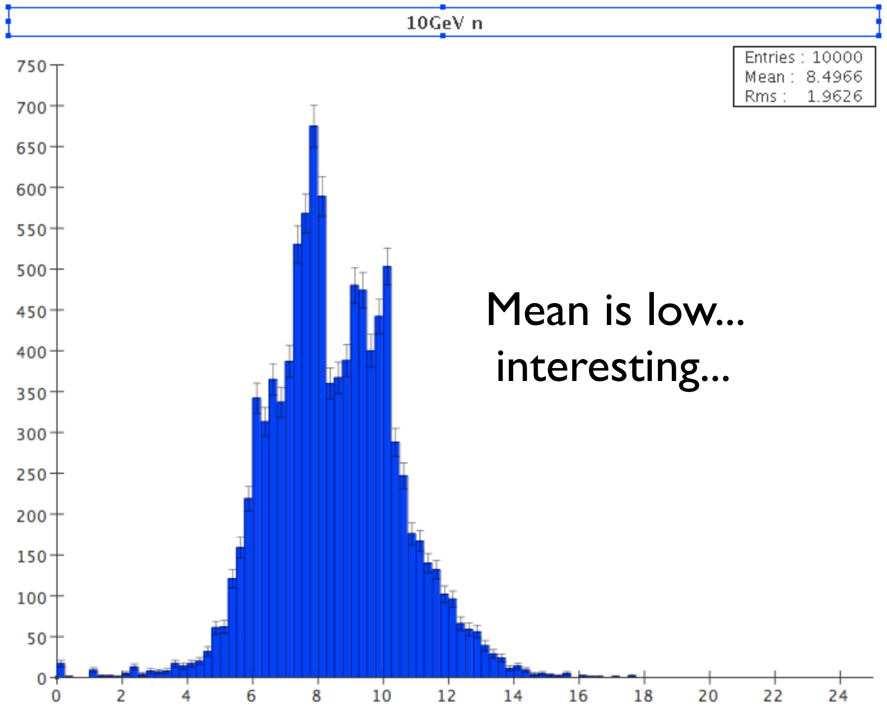
Spike at 10GeV (correct reco)

Find track + extra "neutral hadron"

Mean = 10.85 GeV

Full RMS = 1.86 GeV

10GeV n

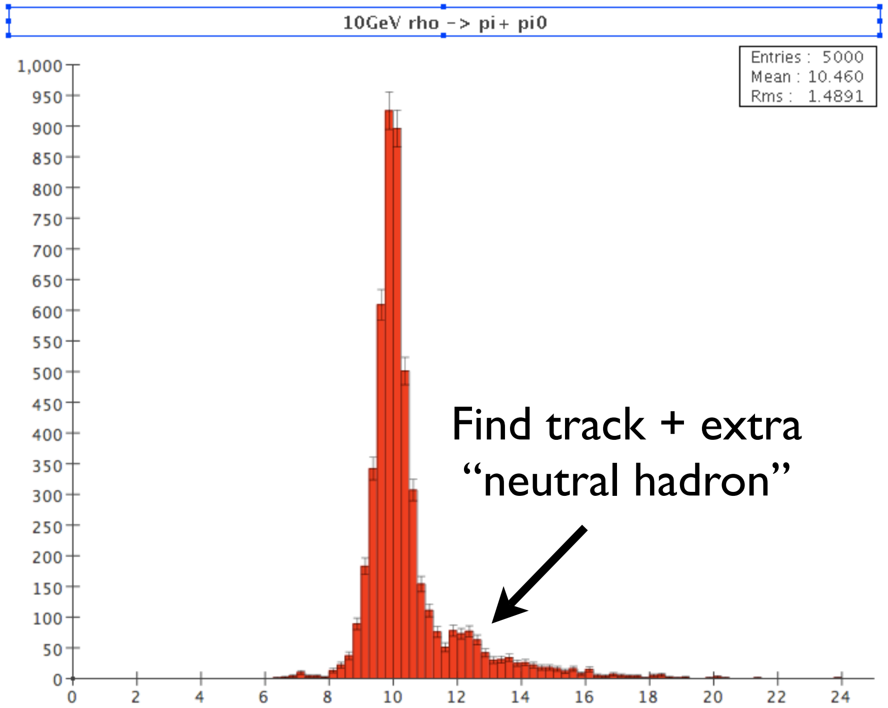


Mean is low... interesting...

Mean = 8.50 GeV

Full RMS = 1.96 GeV

10GeV ρ^+



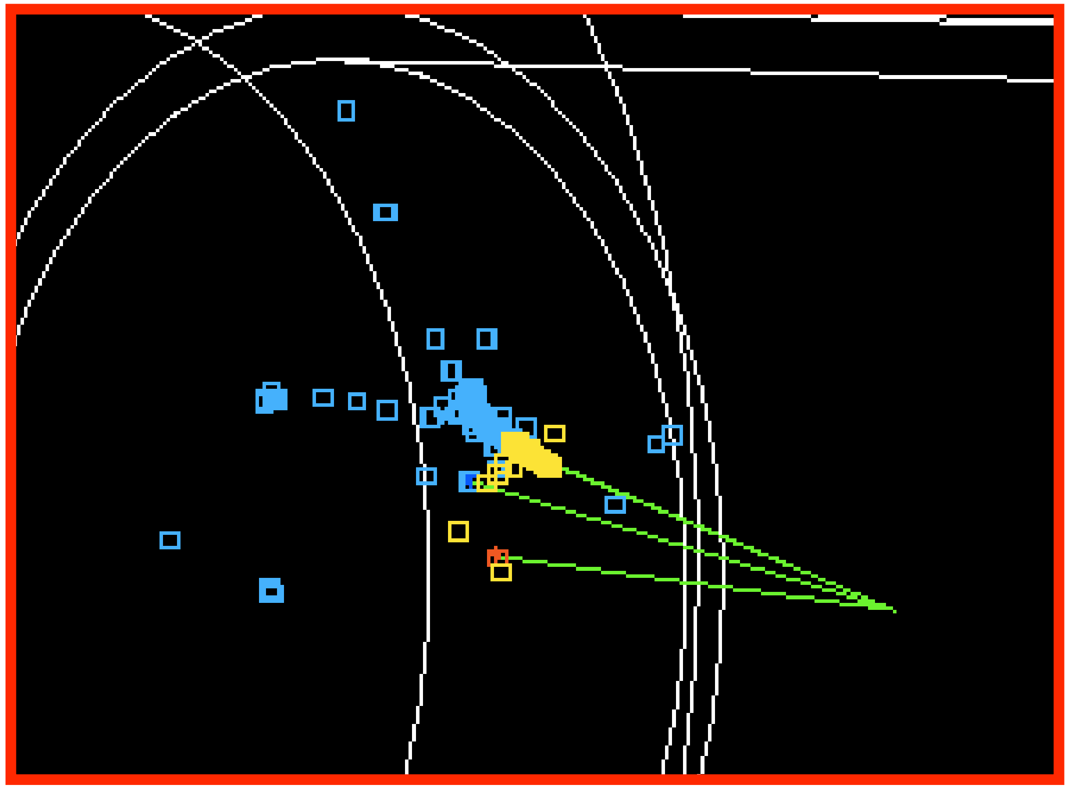
Find track + extra "neutral hadron"

Mean = 10.46 GeV

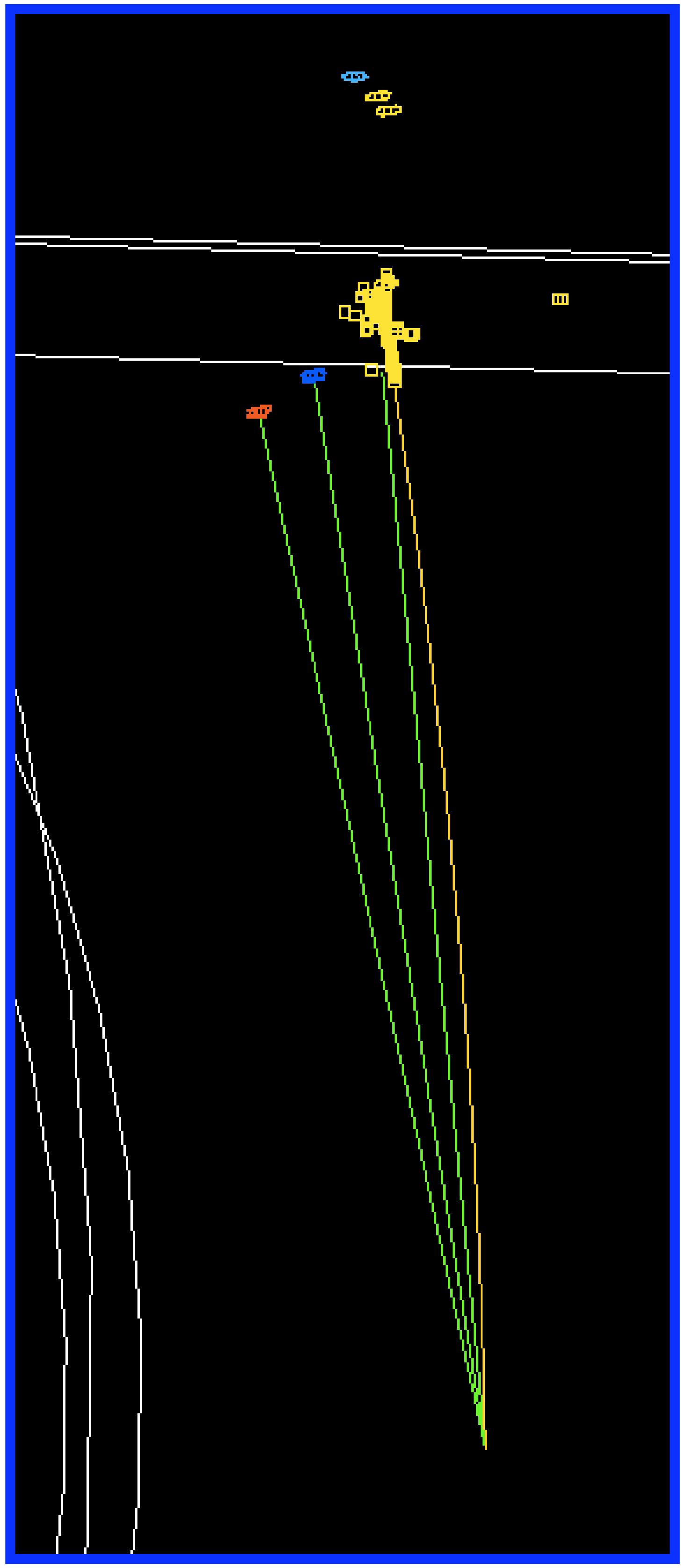
Full RMS = 1.49 GeV

Most π^+ are well reconstructed... but there is enough of a tail that full RMS is nearly as bad as for neutral hadrons!

Two example single π^+ events



Cluster was split into pieces and neither matched E/p of track ⇒ pure calorimetry used.



Main cluster found... along with a fragment identified as neutral hadron. Energy overcounted by ~ 2 GeV.

Diagnosis & treatment

- Clustering fails for a fraction of hadronic clusters
- Currently dodging these with E/p cut... but this is killing the resolution. Need to fix properly.
- Not making full use of the information!
  - Cluster topology & E/p & number of tracks per cluster
  - Using these individually... but there's more power in an iterative approach, correcting problem for individual showers
- Treatment: iterative reclustering

Sketched algorithm idea

- Take previous clustering as initial state
  - Skeletons, built from MIPs and clumps
  - Halo of energy around the above
  - Photon candidates
  - Other blobs
  - Isolated hits
- Look again at linkage
  - Cluster pieces connected directly to tracks (seeds)
  - ... and indirect connections (fragment→clump→MIP→track)
  - Assign a "score" to each link; keep if score above threshold
- If E/p is wrong, recluster with looser/tighter threshold
  - ... and iterate since neighbours will be affected too
- Add more obvious nearby clusters if consistent with E/p

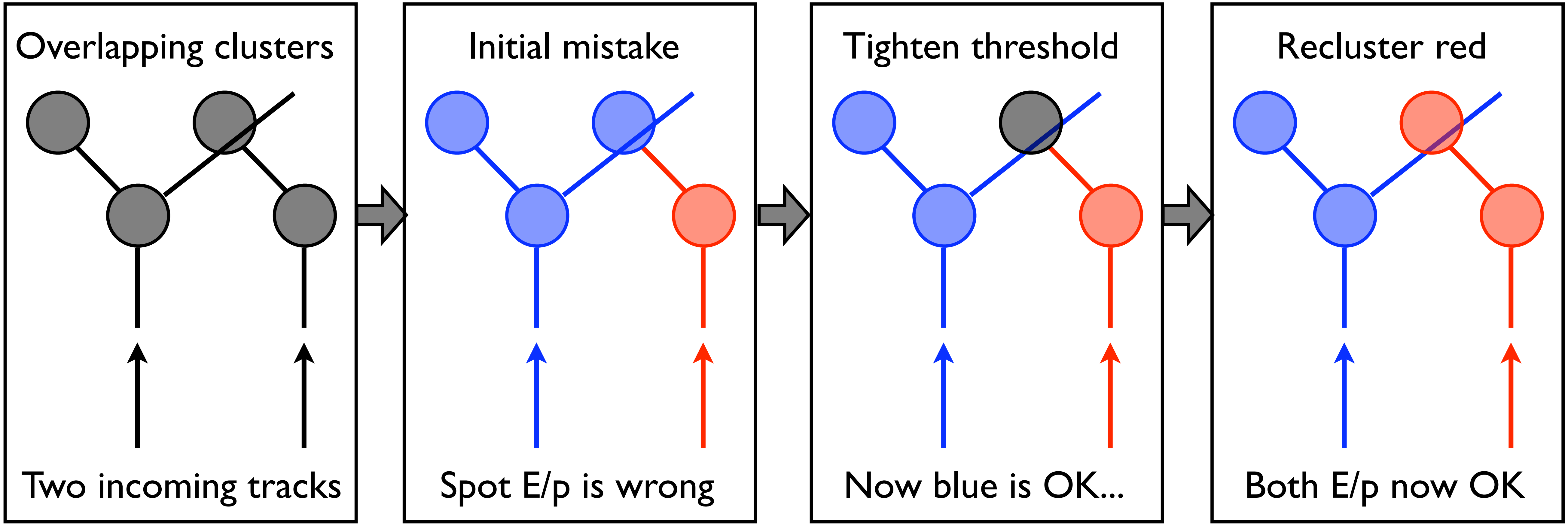
Sketched algorithm idea

- Take previous clustering as initial state
  - Skeletons, built from MIPs and clumps
  - Halo of energy around the above
  - Photon candidates
  - Other blobs
  - Isolated hits

The hard bit!

- Look again at linkage
  - Cluster pieces connected directly to tracks (seeds)
  - ... and indirect connections (fragment→clump→MIP→track)
  - Assign a “score” to each link; keep if score above threshold
- If E/p is wrong, recluster with looser/tighter threshold
  - ... and iterate since neighbours will be affected too
- Add more obvious nearby clusters if consistent with E/p

Some thought experiments



Overlapping clusters

Two incoming tracks

Initial mistake

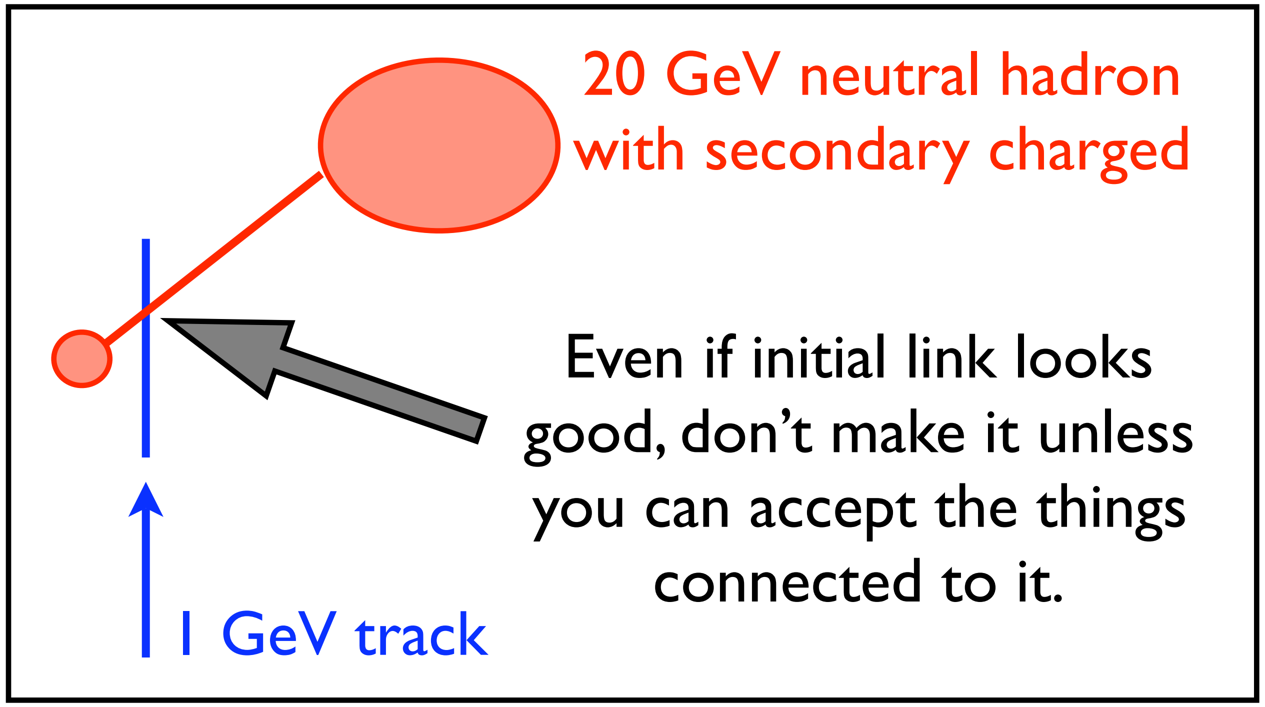
Spot E/p is wrong

Tighten threshold

Now blue is OK...

Recluster red

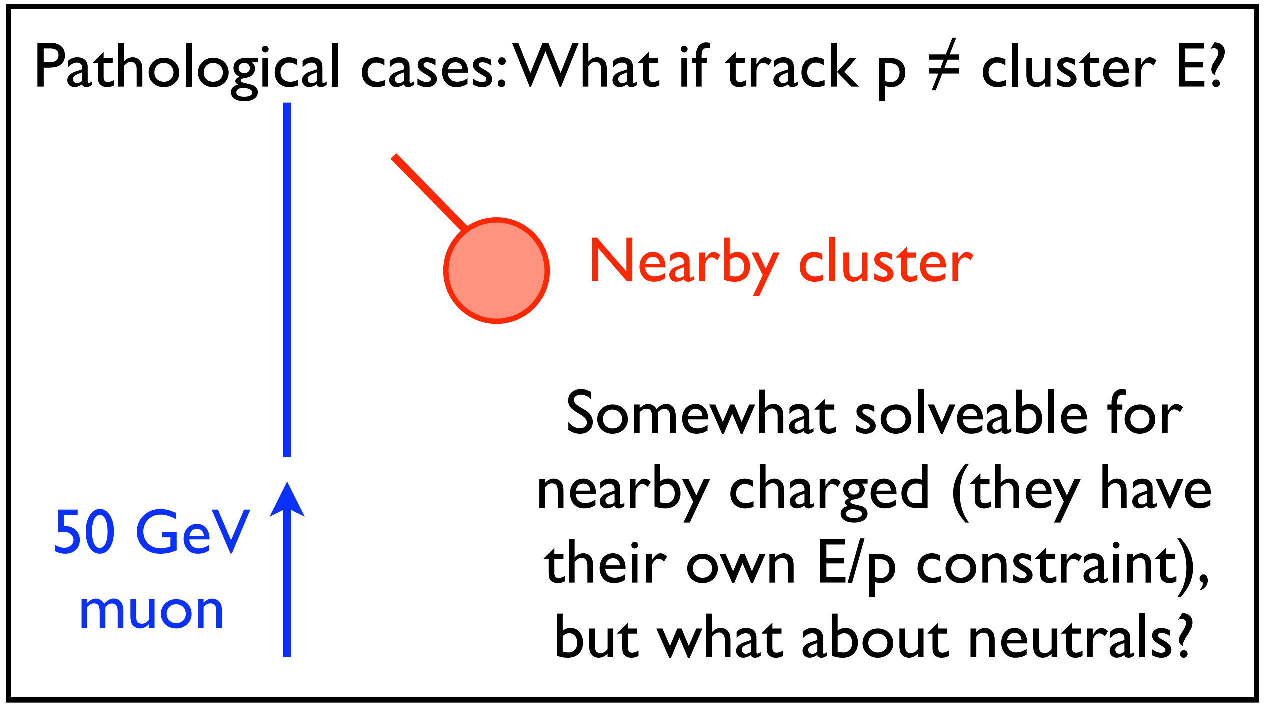
Both E/p now OK



20 GeV neutral hadron with secondary charged

Even if initial link looks good, don't make it unless you can accept the things connected to it.

1 GeV track



Pathological cases: What if track p ≠ cluster E?

Nearby cluster

50 GeV muon

Somewhat solveable for nearby charged (they have their own E/p constraint), but what about neutrals?

Last words

- PFA still under development
- I found a bunch of problems and fixed/tested them...
  - Track extrapolation
  - Photon finding
  - Track matching

Had help from Ron here

- ... but the overall performance is still not good.
- Clustering seems to be the (?) problem
- Trying more sophisticated clustering strategy
- Early-draft code exists, but not ready or tested yet