Pandora Timing

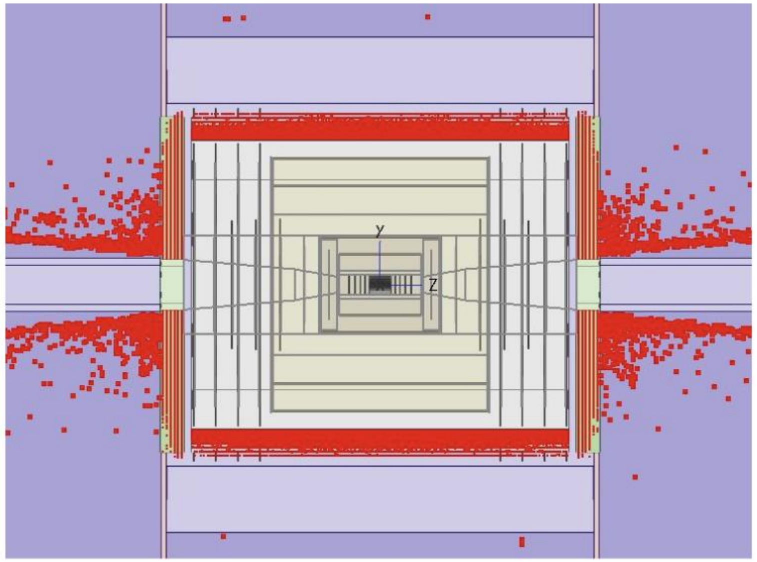
Gregory Penn
Yale University

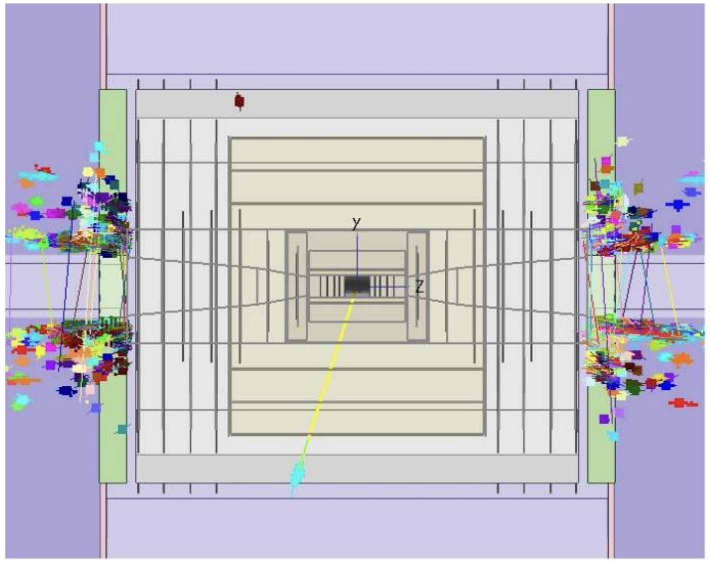
Purpose

- Reco jobs w/ BIB can take O(days) in the endcap w/ threshold fix
- Severely limits stats in the forward region
  - See Rose's slides
- A couple ideas to speed up reconstruction:
  - (Re-) tighten calo thresholds (boo! Reco performance improved after this)
  - Add some hit selection based on isolation requirements
  - Simplify reconstruction
  - Add calorimeter hit time cuts
- I've timed some of these options to see which are feasible

Hit Isolation

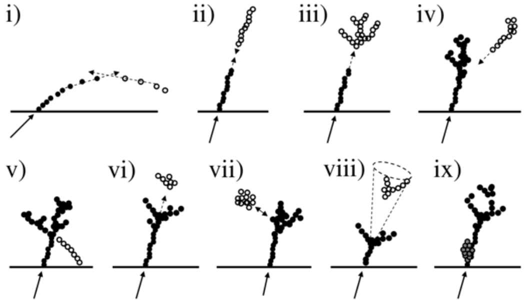
- MUSIC team optimized hit selection to reject BIB hits *before* clustering
- If done correctly, this should remove many background hits from being looped over, very few signal hits removed
- Default Pandora setting: remove hits with < 2 other hits in a 2x2 cell matrix centered on the hit (left image below)
    - MUSIC optimization: < 3 other hits in a 3x3 cell matrix centered on the hit (right image below)
- Works better in the barrel than endcap
- Also sensitive to calo cell sizes. MAIA ECal size much smaller than MUSIC. Not clear if same setting would work as well (spoiler: it did not.)





Pandora Settings

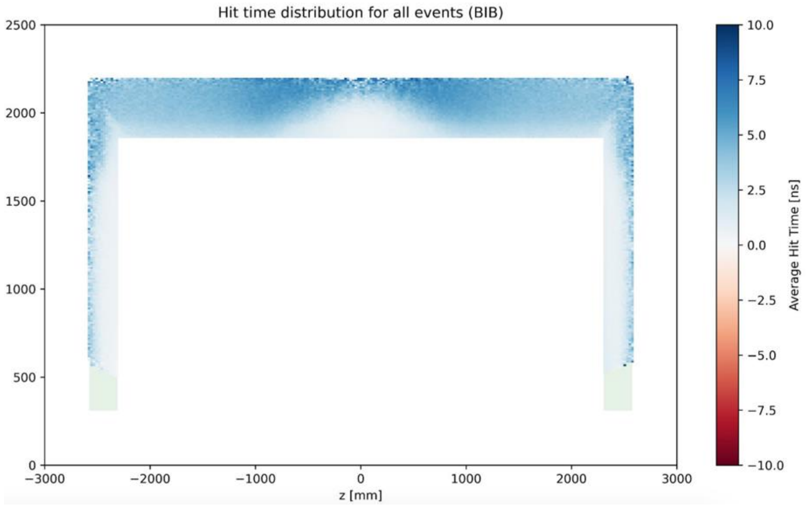
Disclaimer: I still have only skimmed these files! I may be wrong about their purpose.

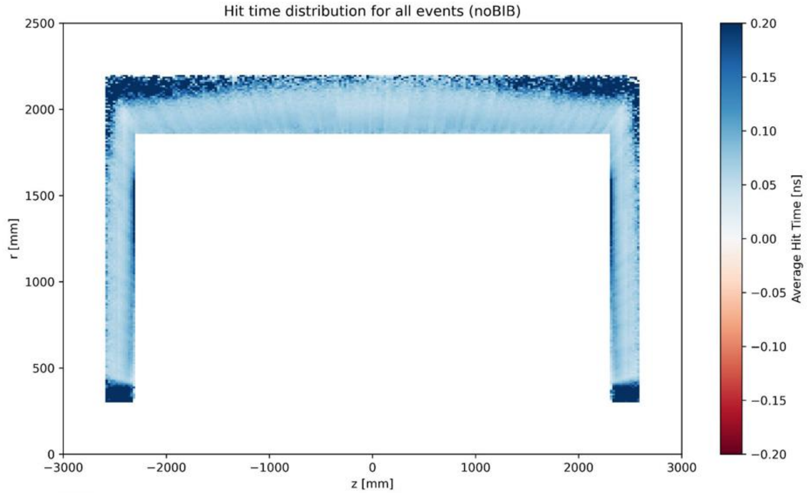


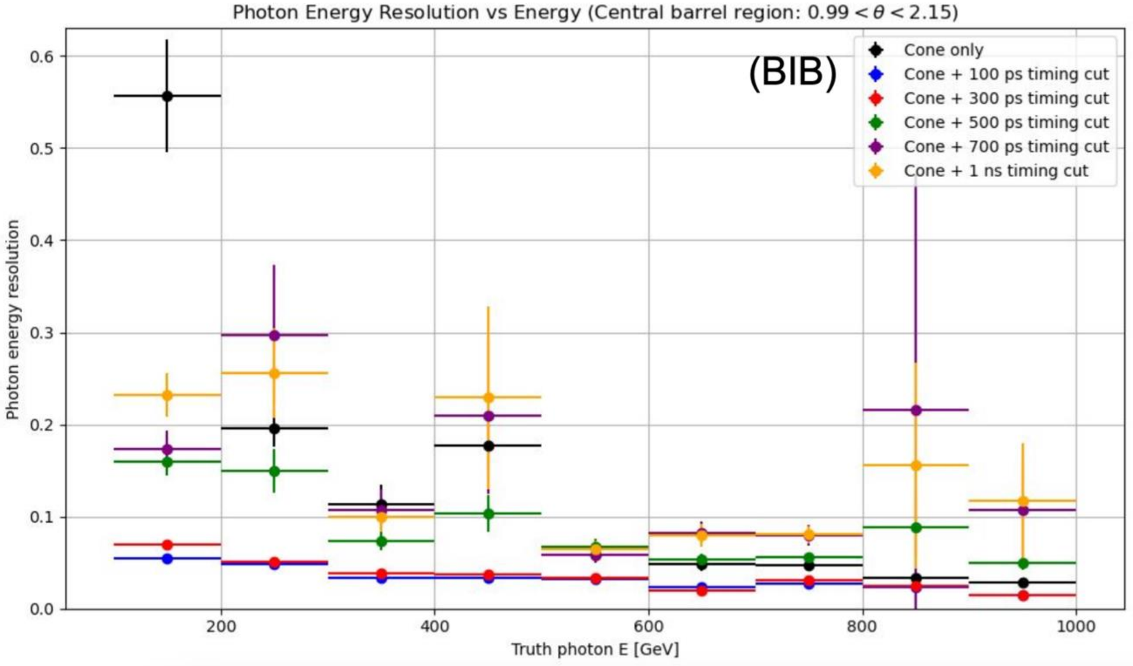
- Larry’s config runs (in addition to the bare minimum):
  - Fast photon clustering
    - “Quick” photon ID algorithm. These “fast ID’d photons” are not considered in the rest of particle flow. If the photon ID is good, this could actually save time.
  - Topological merging algorithms:
    - BrokenTracks (figure ii): Intends to connect clusters separated by “dead” region. Fits line to cluster, searches for other clusters along that line
    - ShowerMipMerging (figure iii, iv) (algorithm 1, algorithm 2): Identifies MIP-like clusters, attempts to merge them. Not sure what this exactly does just from skimming the code. The two algorithms look similar.
    - ConeBasedMerging (figure viii): Opens cone around MIP fit, merges particles with sufficient number of hits in cone
    - ProximityBasedMerging (maybe figure vii): Merges nearby clusters. Can also account for clusters along a track projected through the calorimeter.
  - Fragment Removal algorithms (Main, Neutral, Photon, PhotonSplitting, PhotonFragmentMerging).
    - I left these in a few months back, as I noticed the number of clusters reconstructed in an event w/o BIB was rarely = 1 often without them. But no extensive test on calorimeter energy resolution w/ and w/o this were performed.

Calo Hit Time Cuts

- Following studies from Alyna
- 300 ps calo hit time cuts removes a large fraction of BIB hits, comparatively few signal
- Improves photon energy resolution
- Should reduce runtime of PandoraPFA
- Implemented in external CaloHitSelector processor
    - Cut on earliest subhit time, corrected for time of flight
    - Applied in ECal only







Test info

- I ran 7 jobs:
  - No reconstruction at all (for timing the rest of the job). Includes:
    - Overlay of BIB
    - Digitization
    - Tracking
  - Basic Pandora: Only the necessary algorithms w/o parameter optimization (like what MUSIC did on slide 13 here)
  - Larry’s config (config)
  - “Larry + tight calo hit isolation”(config)
    - Uses definition from MUSIC (slide 3)
  - “Reduced Topological Merging” (config)
    - Removed BrokenTracks, ShowerMipMerging2, and ConeBasedMerging (very vibe based, no strong reason)
  - Larry’s config without any fragment removal algorithms (config)
  - Larry’s config with ECal hit time cuts (script)
- All test were run on the same *one* event of a charged pion going into the endcap region
  - With BIB files copied from /ospool/uc-shared/project/futurecolliders/data/fmeloni/DataMuC_MAIA_v0/v5
  - Example script

Time for each config

- Pandora time = total time – time for “no reco” job
- Only relative times are helpful – job times depend on machine
- These are rough estimates, but general trends make sense
  - Rounding to nearest 15 min

Reco Job Type	Total Time	Pandora Time
No reconstruction	3 hrs, 45 min	---
Basic Pandora	4 hrs, 45 min	1 hr
Larry’s Config	9 hrs	5 hrs, 15 min
Larry + tight calo hit isolation	9 hrs, 45 min	6 hrs
Reduced Topological Merging	8 hrs, 30 min	4 hrs, 45 min
No fragment removal	6 hrs, 15 min	2 hrs, 30 min
Calo hit timing cut	6 hrs, 15 min	2 hrs, 30 min

Conclusions

- Basic Pandora can run in significantly less time than the configs used for v5
- Tightening calo hit isolation increased run time
  - Loop over additional calo layers not worth the number of hits removed in later steps
- Fragment removal seems to be very slow! Removing all fragment removal halved the runtime of Pandora
- Topological cluster merging algorithms seem to be relatively cheap to run
- Remaining algorithms not tested (Fast photon clustering + ShowerMipMerging + ProximityBasedMerging) run in ~ 1hr
- Promising direction: Implementing ECal hit time thresholds before reconstruction

Reco Job Type	Total Time	Pandora Time
No reconstruction	3 hrs, 45 min	---
Basic Pandora	4 hrs, 45 min	1 hr
Larry’s Config	9 hrs	5 hrs, 15 min
Larry + tight calo hit isolation	9 hrs, 45 min	6 hrs
Reduced Topological Merging	8 hrs, 30 min	4 hrs, 45 min
No fragment removal	6 hrs, 15 min	2 hrs, 30 min
Calo hit timing cut	6 hrs, 15 min	2 hrs, 30 min