PFA for SiD: Where we are and where we aren't

Mat Charles
The University of Iowa

Foreword

This talk will be short, plain, and won't have many pictures or plots.

$$t_{\text{development}} + t_{\text{benchmarking}} + t_{\text{talk preparation}} = \text{const} \ll 1$$

Disclaimer: Concerns based mainly on performance of my PFA implementation.

What is the goal?

To produce lists of reconstructed final-state particles without cheating which are good enough to use in physics benchmarking & analysis.

This immediately throws up questions:

- Physics benchmarking: Which channels?
- Good enough: How good is that?
- Without cheating: What do we do in the meantime?
- Final-state particles: A whole other can of worms...
- Which detector? Digital or analog readout? etc etc etc

Where are we?

To short-cut these questions, we pick:

- Baseline detector sid01, treat HCAL as digital
- Cheat on tracks in various ways
- Figure of merit: rms_{90} of dijet invariant mass residuals in simple events ($e^+e^- \rightarrow Z(qq)\ Z(\nu\nu)$, $|\cos\theta_q|<0.8$)
- Good enough: rms_{90} ~ 3 GeV (maybe 4 GeV)

Critical to show that we can do at least this much

- This is a specific, relatively easy case -- algorithm isn't close to done until we can solve it.
- This is the minimum threshold. We have to reach it to prove that the SiD is viable. Can we show this in time for LOI?

Where are we?

To short-cut these problems, we pick:

- Baseline detector sid01, treat HCAL as digital
- Cheat on tracks in various ways
- Figure of merit: rms_{90} of dijet invariant mass residuals in simple events ($e^+e^- \rightarrow Z(qq)\ Z(\nu\nu)$, $|\cos\theta_q|<0.8$)
- Good enough: $\mathrm{rms}_{90} \sim 3$ to 4 GeV

We don't have a PFA which gets 4 GeV reproducibly.

- I know mine does worse: $\mathrm{rms}_{90} \sim 5.9$ GeV.
- Other code exists, but isn't always easy to benchmark. (Ron Cassell is the gold standard for reproducibility!)

So now what?

What can we do? (Short-term)

Getting to threshold any way we can:

- Current PFA developers should keep plugging away and hope for a Eureka moment. Odds are not good (it's a marathon, not a sprint) but you never know...
- PandoraPFA has been shown to perform well for LDC00. There's a good chance it will be adequate (if not optimal) for sid01. If we can get it working properly on sid01 [not trivial] and show adequate performance, we prove that the hardware is viable & are no longer dead in the water.

Both efforts are ongoing in parallel.

What can we do? (Long-term)

Need a PFA for optimization, physics studies, etc.
Three basic options:

- Add tightly focused effort: Coherent, concentrated effort (ideally full-time PFA developers).
- Add diffuse effort: Template approach is supposed to allow new people to try out algorithms & work on pieces. Not much has filtered back yet.
- Switch to PandoraPFA. This is not a panacea: Pandora is optimized for a very different detector & may give non-optimal/misleading results. (Also technical headache.)
  - This was part of Caroline's point in the emails. But perhaps if we were involved as developers rather than users, it would be better...

Beware of premature optimization

- We want to have the best detector we can, and quantitative input from PFA will be essential.
- But until we show that our PFA gives adequate, consistent, and well-understood performance, comparisons are meaningless & potentially misleading.
  - PFA is a complex interacting system; until you're close to the minimum, you can't change one piece in isolation.
- For example, I compared detector variants with different inner ECAL radii and found:

  acme0605 with r=125cm: rms_{90} = 5.29 GeV
  acme0605 with r=150cm: rms_{90} = 5.54 GeV
  acme0605 with r=175cm: rms_{90} = 5.69 GeV

  That doesn't [necessarily] mean a bigger detector is worse, it means I don't understand my track-matching yet...

A few more words of caution

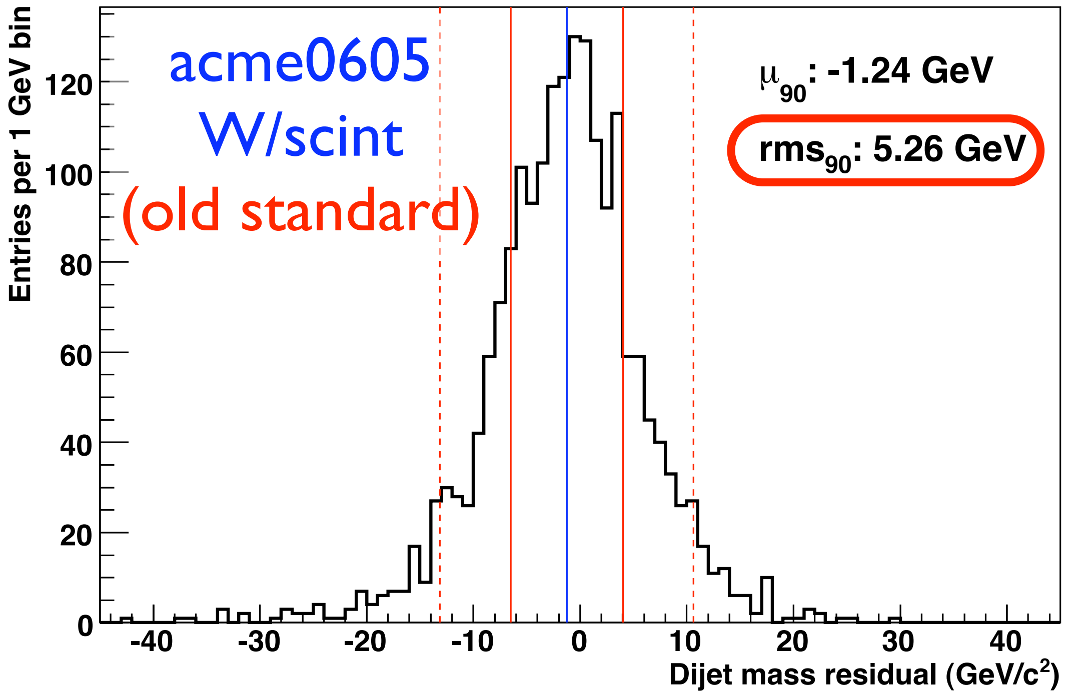
- Algorithm development is heavily influenced by the choice of figure of merit
    - e.g. focus on Z → qq (q=u,d,s) means we don't think much about leptons, heavy quark jets, etc

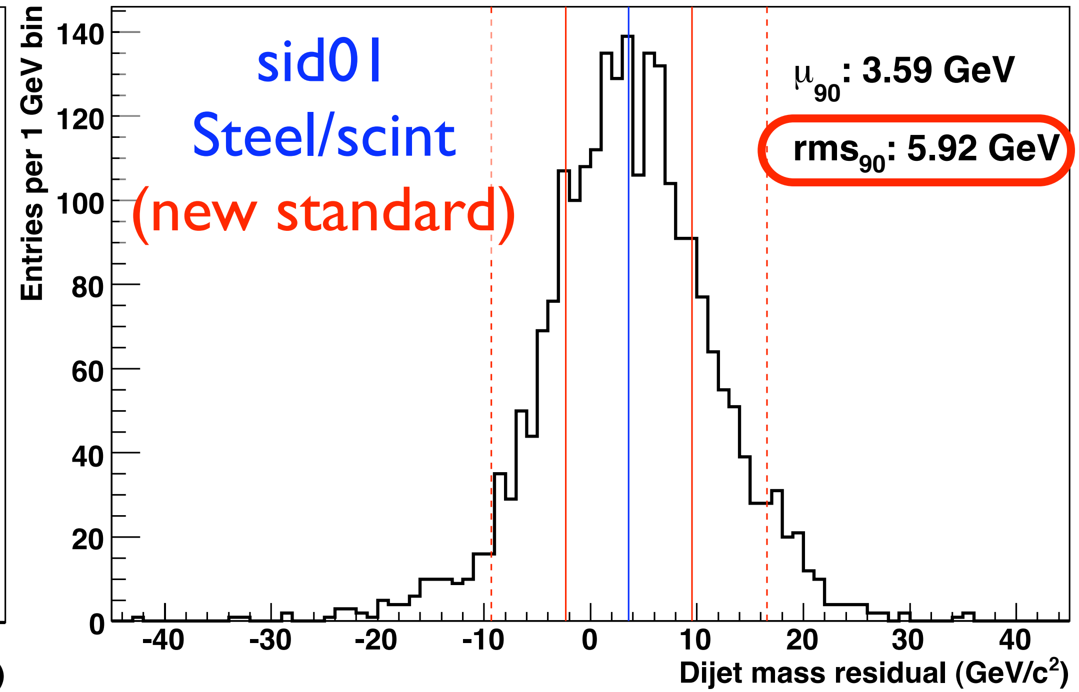
  For now that's OK -- first priority is to solve at least one case -- but we'll need to return to this.

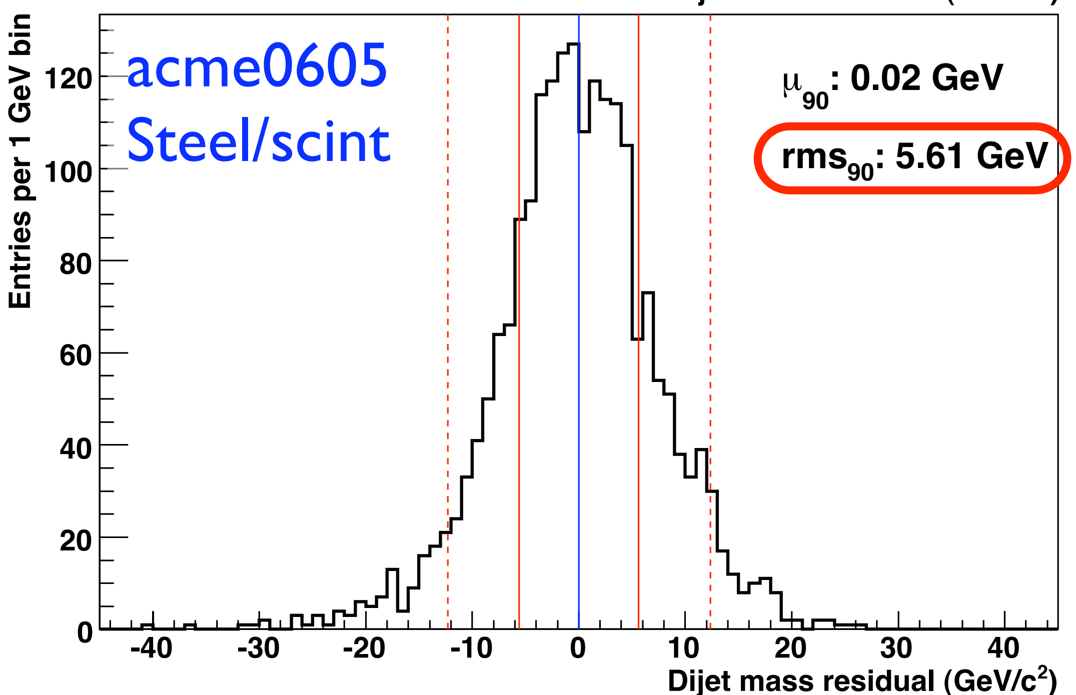
- Definition of tracks/final-state particles can make a significant difference
    - Some very important work going on here (Ron, Steve, Rob)
- Some effects that are minor now will become important as performance improves
    - e.g. Helical approximation for tracks

A couple of plots

... for everybody's favourite event type: $e^+e^- \rightarrow Z(qq)\,Z(\nu\nu)$ @ 500 GeV







Something changed here & I lost ground.

It's not obvious what -- material details, extra layer in ECAL, artifact of algorithm, ...

Summary

- We do not yet have a public PFA which gives adequate performance reproducibly in our chosen benchmark.
  - Still hoping for a pleasant surprise at FNAL!
- We are trying to get PandoraPFA going. Want to see relative and absolute performance on sid01.
- On the present trajectory, there is a very real risk that we won't have adequate PFA performance in time to write the LOI (i.e. spring, maybe summer).
  - What can be done to mitigate this?
  - What is the fallback strategy?