

Analysis Ntuple Needs

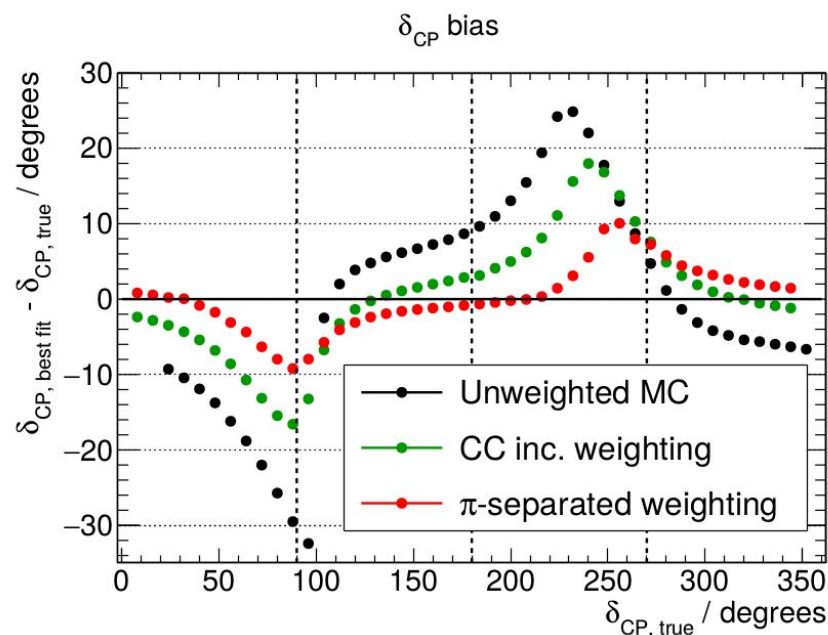
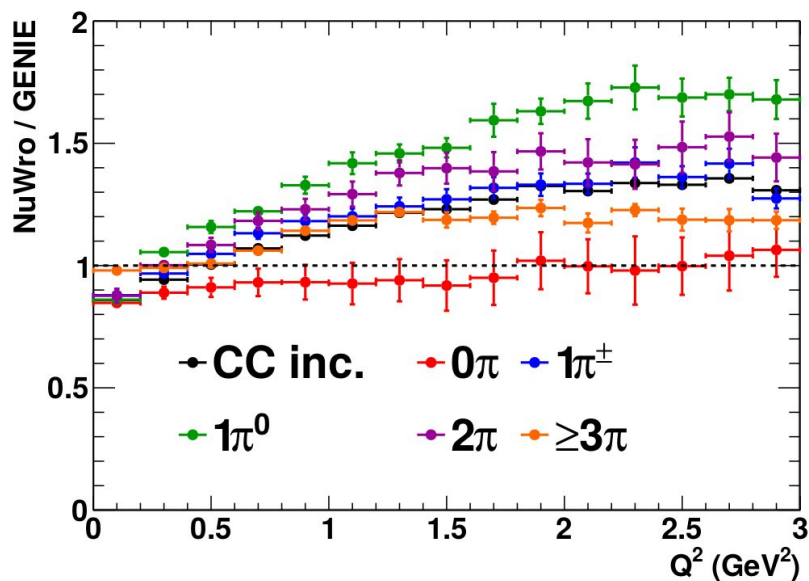


Mike Kordosky

Motivation

- We need to do more realistic studies for the TDR, to couple the design and performance of the ND to the oscillation sensitivity.
- Especially important for ND-GAr. Also, may have a longer timeline.
- A good example is the pion multiplicity study (C. Marshall, S. Jones) that appears in the ND CDR

NuWro/GENIE for various reconstructed final states (FHC)



Motivation

- The oscillation analysis will not necessarily require cross-section measurements
- But, if we can make cross-section measurements we can likely do whatever else is needed.
- So, that is my motivation: What is needed to do cross-section analyses in ND-GAr?

What is needed? (mock analysis)

- Large enough MC event samples.
- A reconstruction program.
- A mock data sample.
- A procedure to propagate systematic errors.
 - Flux, cross-sections, detector.
- A good ntuple that can be used within a light framework.

What is needed? (mock analysis)

- Large enough MC event samples. ✓
- A reconstruction program. Tracks, vertices, calorimeter clusters. ✓
- A mock data sample. ✓
- Infrastructure to propagate systematic errors. ~
- **A good ntuple** that can be used within a light framework. ?
- **Bonus:** An easily accessible event display. ✗ ?

Ntuple Needs

- The point of this discussion is to walk through what is needed.
- Some of these things “exist”.
 - In that case, do not get shirty.
 - There is a difference between existence and usefulness.

Ntuple Needs

- MC truth at the interaction level and genie final state level. Neutrino type, 4 vector, target, process type, kinematics (q^2 , etc) neutrino vertex, track PIDs and 4 vectors for final state particles
 - V0s are a special case, ideally you want their decay product too).
 - I think genie decays neutral pions itself, so the gammas appear in this record.
- The stored information should be enough to permit GENIE reweighting. This may mean storing the entire GENIE record.
 - Let's discuss this.

Ntuple Needs

- Track reconstruction info and vertex reconstruction info.
- Also, ideally the hits, but perhaps bloats the output unacceptably.
 - Tom pointed me to a 300 event file that is 500+Mb. Some pared down version of those?
- It is important to enable some early stages reconstruction R&D at analysis level.
- Also, only the final reconstruction pass is needed.
 - Tom shared a file with me that seemed to have multiple passes.
- Having the hits enables visualization, even if kind of crude.
 - Crude visualization is ∞ better than nothing

Ntuple Needs

- Track dE/dX PID info, if we have. Could be parameterized based on p and true dE/dx .
 - I believe this is what we already have? What do we have?

Ntuple Needs

- Calorimeter information that allows one to match tracks to calorimeter hits or clusters. Needs timing.
- What do we have here?
- What position information is available?
- What is, or should be, provided in the case of double ended readout?
- Consider faking the timing of exiting tracks
 - E.g., use MC truth to assign a time for any muon or other charged track exiting the GARTPC?

Ntuple Needs

- From the muon system, information from it that allows one to match tracks and decide if a track showers or punches through.
- Is there a muon system in the current geometry? Is it “read out”?

Ntuple Needs

- Ability to reweight events.
- +/- weights for different genie parameters
- Flux weights.
 - Where are we on this? Can we compute PPFX weights before ntuple generation?
 - Many random weights vs PCA weights?
- The single most useful weight is the the NuWro fake data weight.

Ntuple Needs

- What we implement should work for overlaid samples. This must be designed in from the beginning.

Ntuple Needs

- The ability to analyze the ntuple via MakeClass, at a minimum.
- Having to learn art and write code in C++ will be a huge (insurmountable) barrier for many.
- Better would be the ability to gain access to the class via an I/O library (as portable as possible) and do the work in python.
 - Or C++ , or whatever, but probably python.
 - Students are being trained in python. Aids on-boarding.
 - edepsim offers this and it's pretty nice
 - What is the easiest way to enable ntuple analysis without a large framework and its set of dependencies?
 - Could we have an I/O library that only depends on ROOT?
 - Some other solution? HDF5?