Estimating detector systematic uncertainties for the T2K far detector

Michael Reh

On behalf of the T2K collaboration

NuFact 2024

16 September 2024

Boulder

Overview

Boulder

$$N_{\nu_\beta}^{FD} \propto \boxed{\Phi_{\nu_\beta}^{FD} \times \sigma_{\nu_\beta}^{FD}} \times \boxed{\epsilon^{FD}} \times \boxed{P_{\nu_\alpha \rightarrow \nu_\beta}}$$

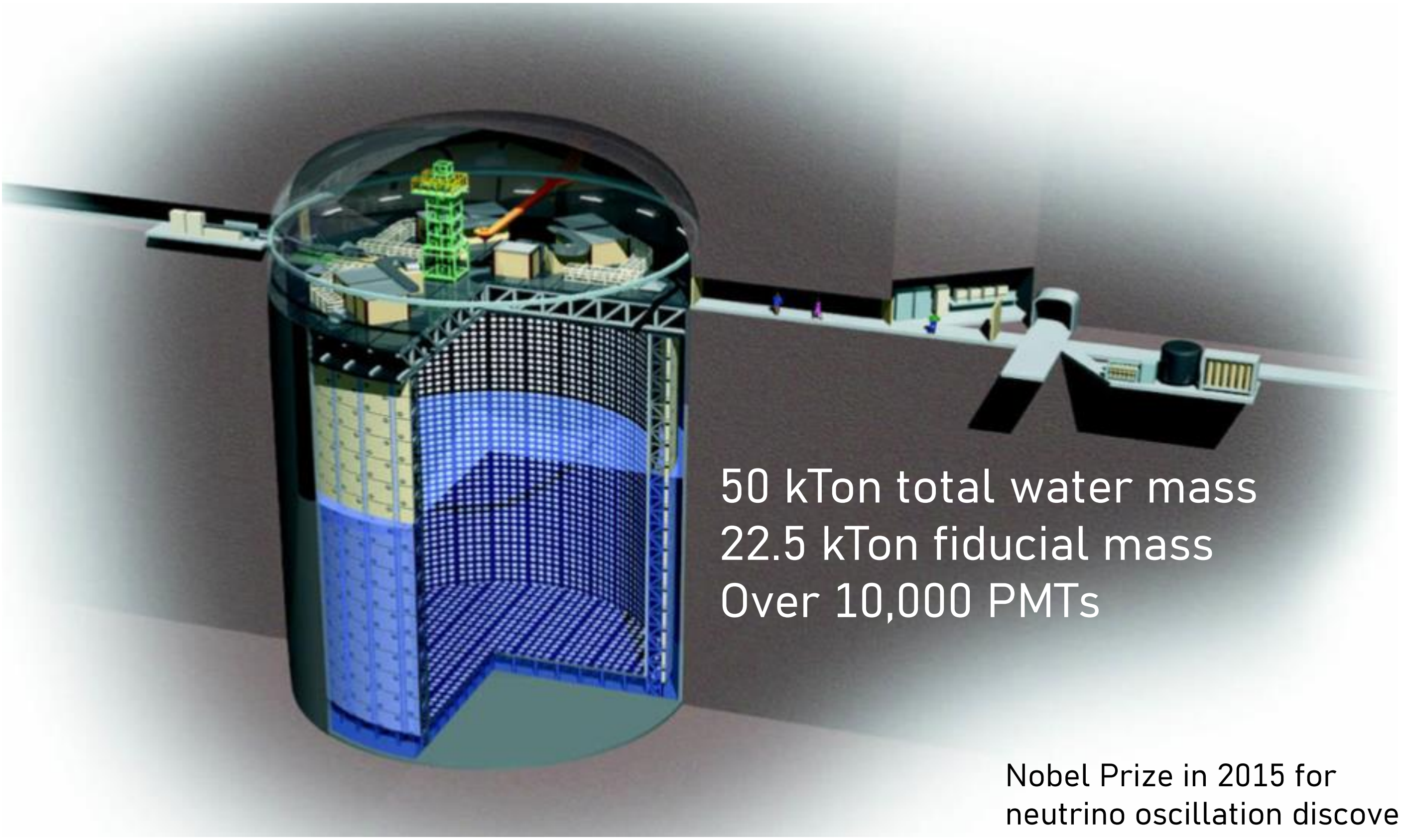
Flux and cross section

SK Detector Model

Oscillation Probability

Super-Kamiokande

- 50 kTon water Cherenkov detector
- Used in T2K as the far detector, located 295 km from the T2K beam source

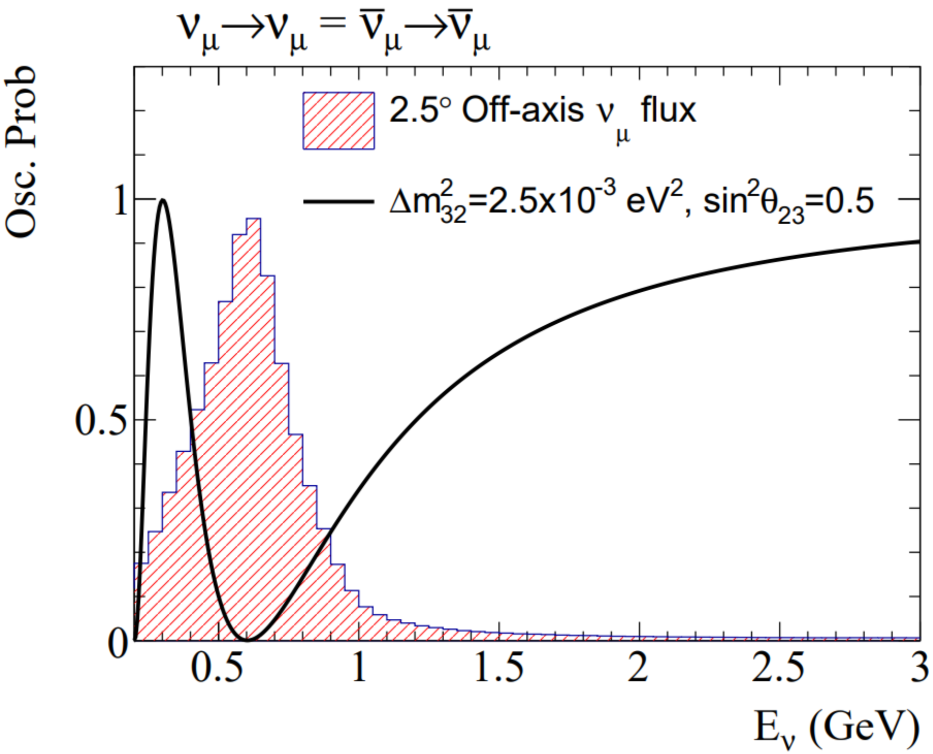


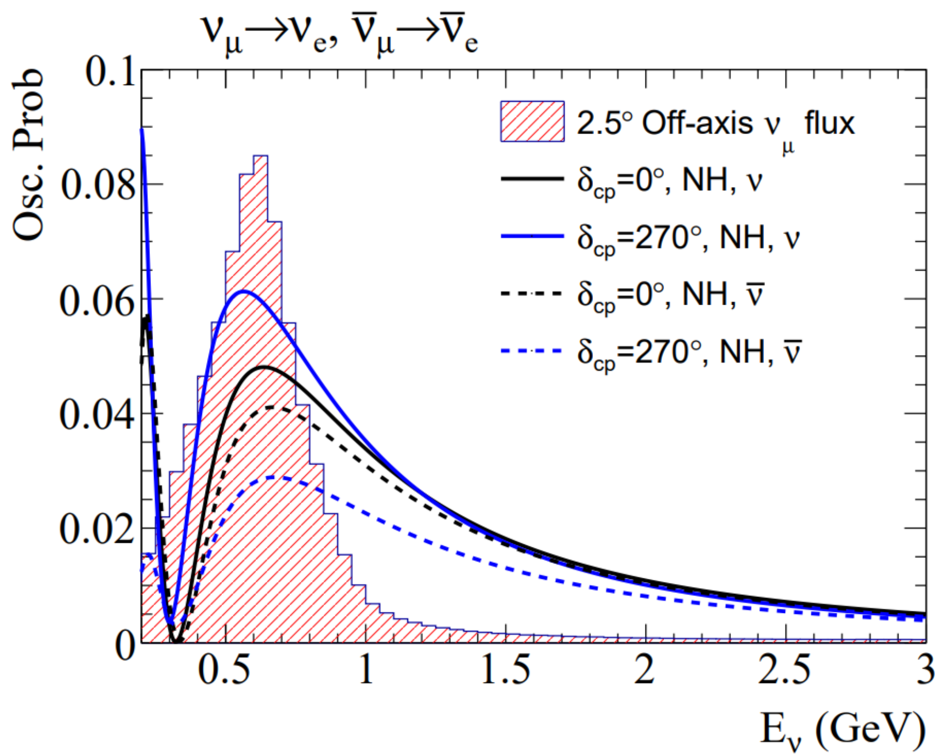
Nobel Prize in 2015 for neutrino oscillation discovery

Super-Kamiokande

Boulder

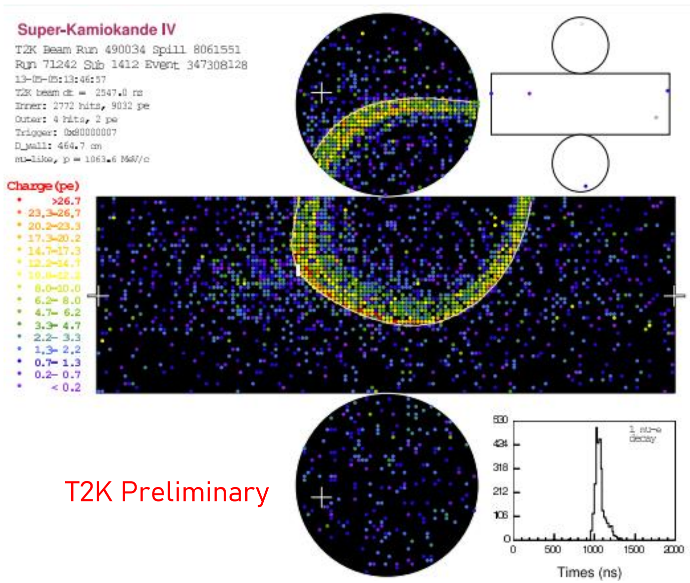
- The 295km baseline and 0.6GeV beam peak energy are designed to maximize the oscillation probability of $\nu_\mu \rightarrow \nu_e$
- T2K's oscillation analysis measures the appearance of ν_e and the disappearance of ν_μ by observing their energy spectrum at SK



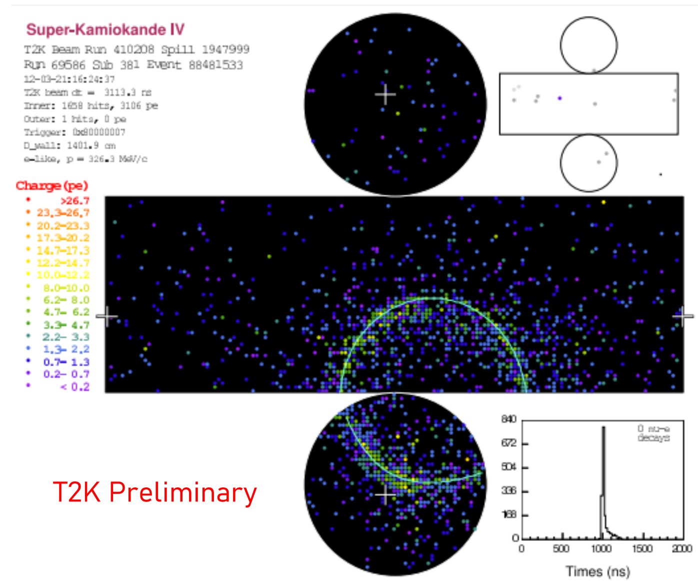


Super-Kamiokande

- SK detects Cherenkov rings
- Muon rings tend to be “sharp”; electron rings tend to be “fuzzy”
- Reconstruction algorithm predicts what particle left what ring in the detector
- How robust is the reconstruction algorithm in reality?



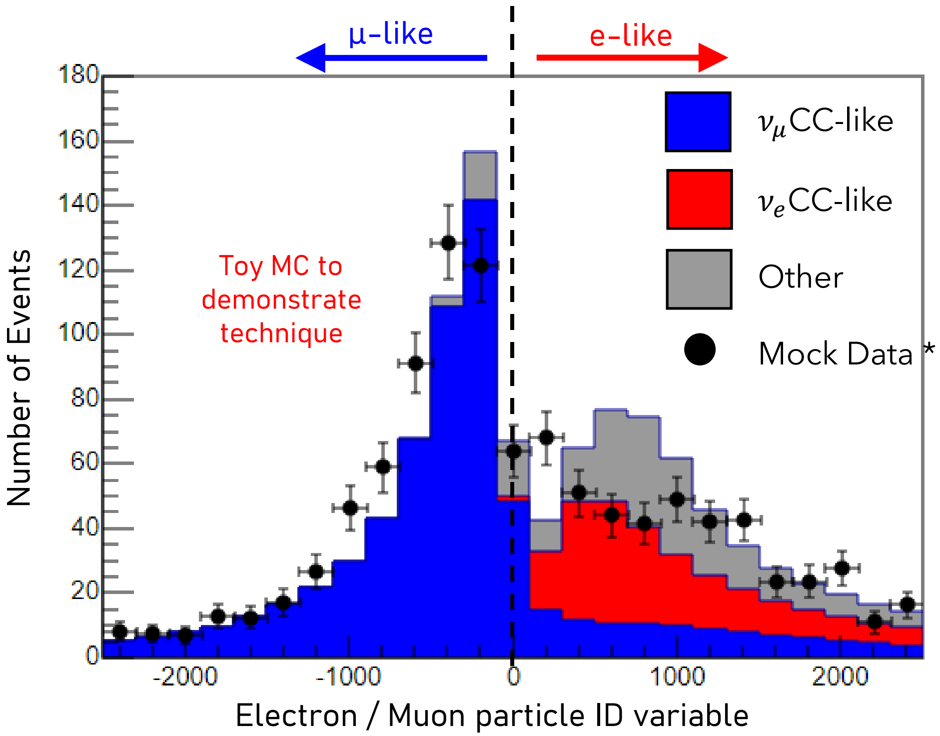
μ-like



e-like

Toy particle ID spectrum

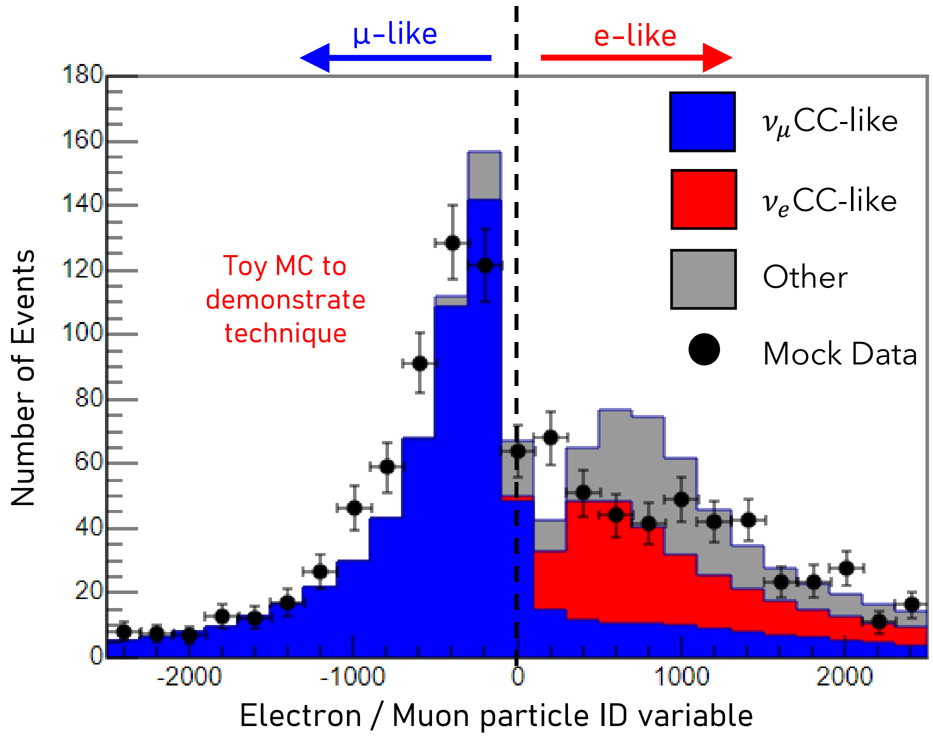
- The event reconstruction algorithm gives us log-likelihood ratio-based particle ID variables for each event
- T2K defines its analysis samples in SK (e.g., 1 ring e-like) by making cuts in these variables



*Mock Data generated by modifying nominal distributions and taking Poisson throws in each histogram bin

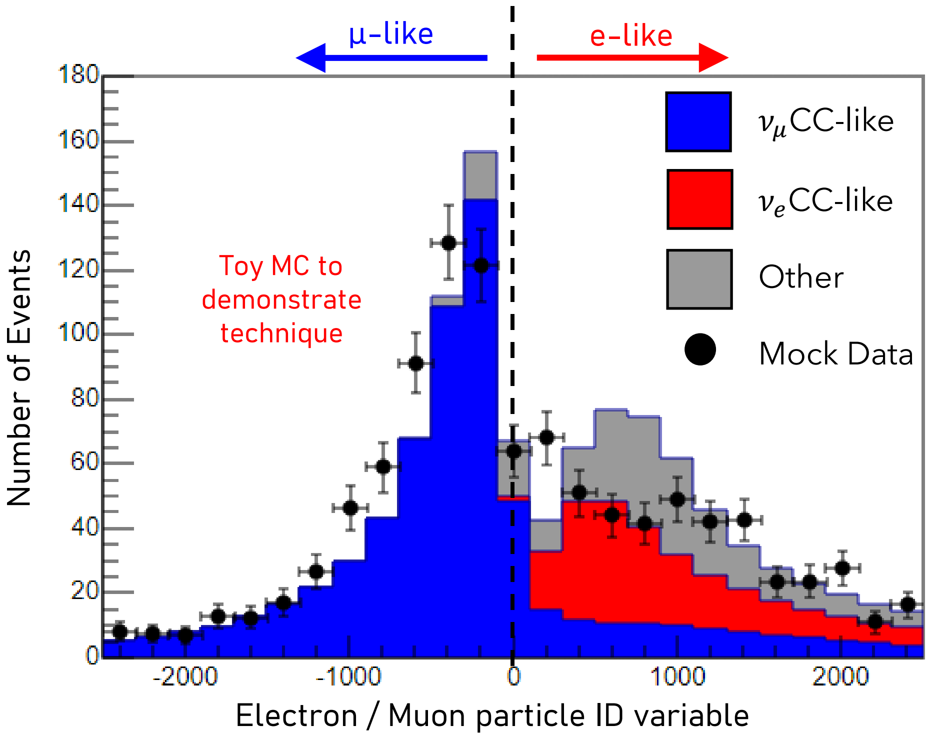
Toy particle ID spectrum

- How do we know how accurate the reconstruction algorithm is?
- What about difficult-to-model effects like light scattering intensity in water and PMT responses?
    - May affect data differently from MC!



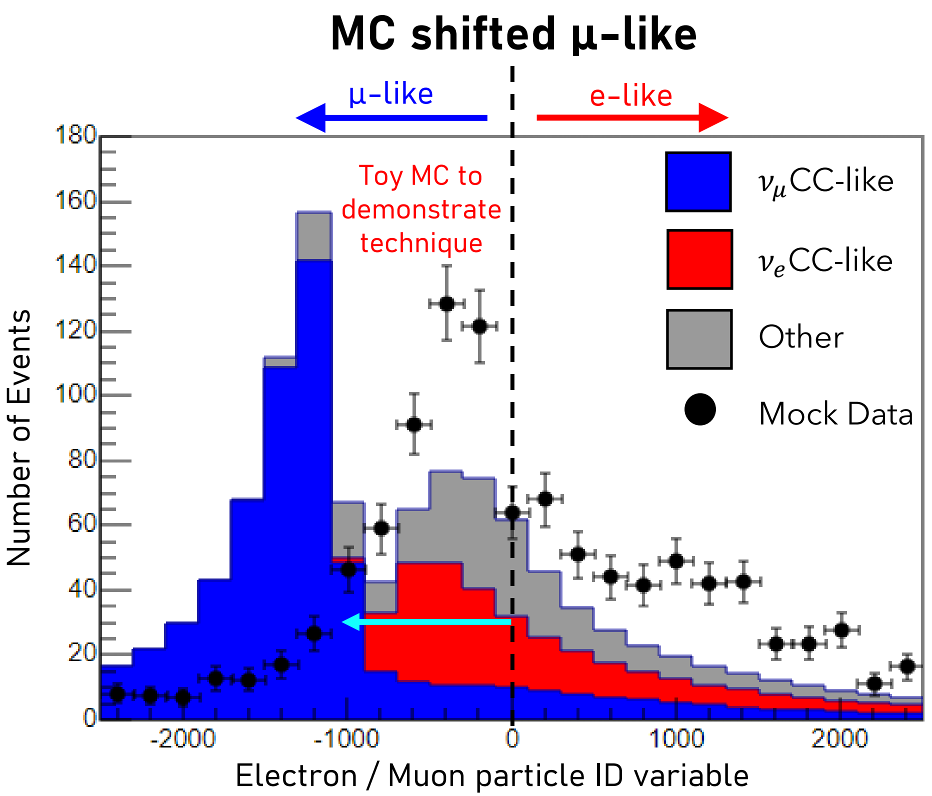
Toy particle ID spectrum

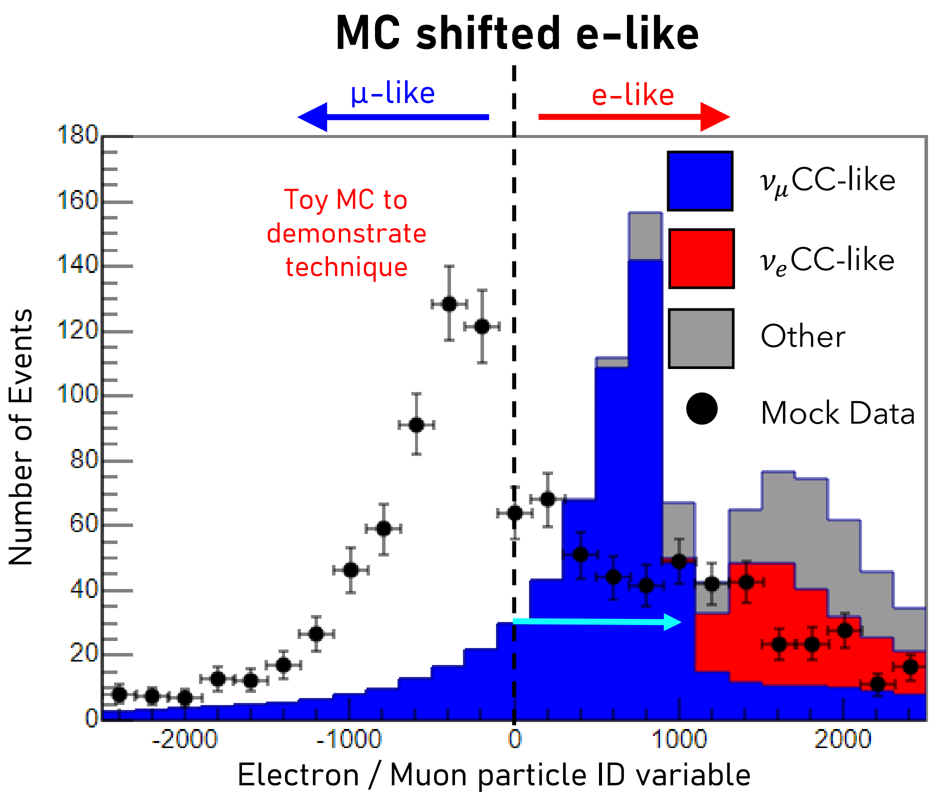
- Assume any mis-modelling will show at the particle ID level as systematic shifts in particle ID variable value



Toy particle ID spectrum

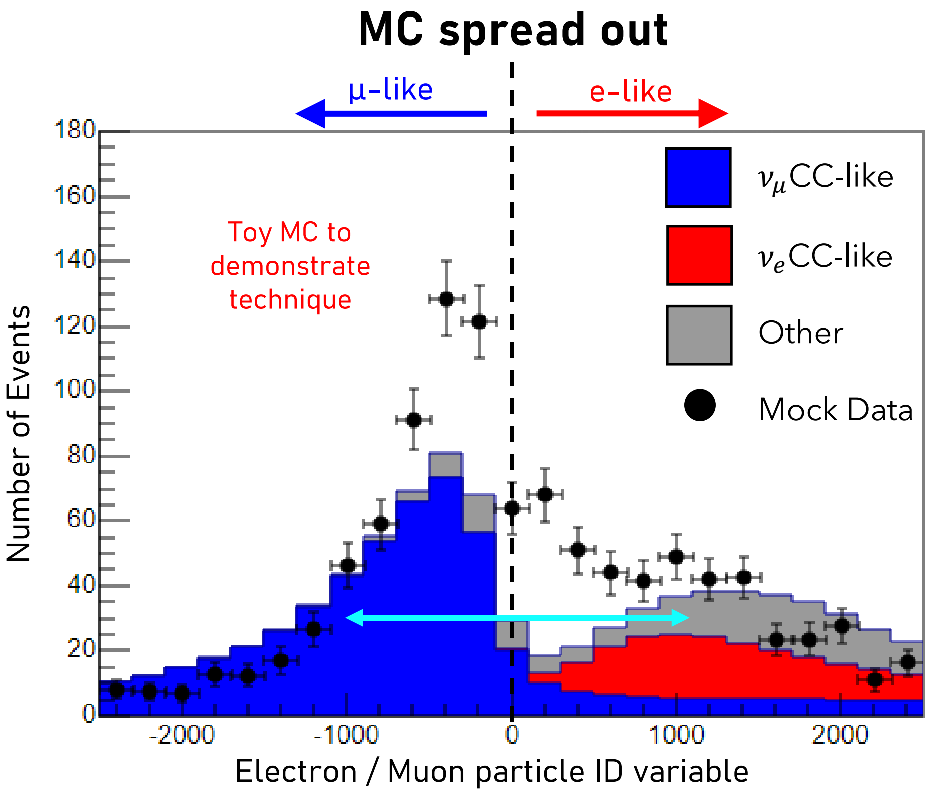
- Assume any mis-modelling will show at the particle ID level as systematic shifts in particle ID variable value,

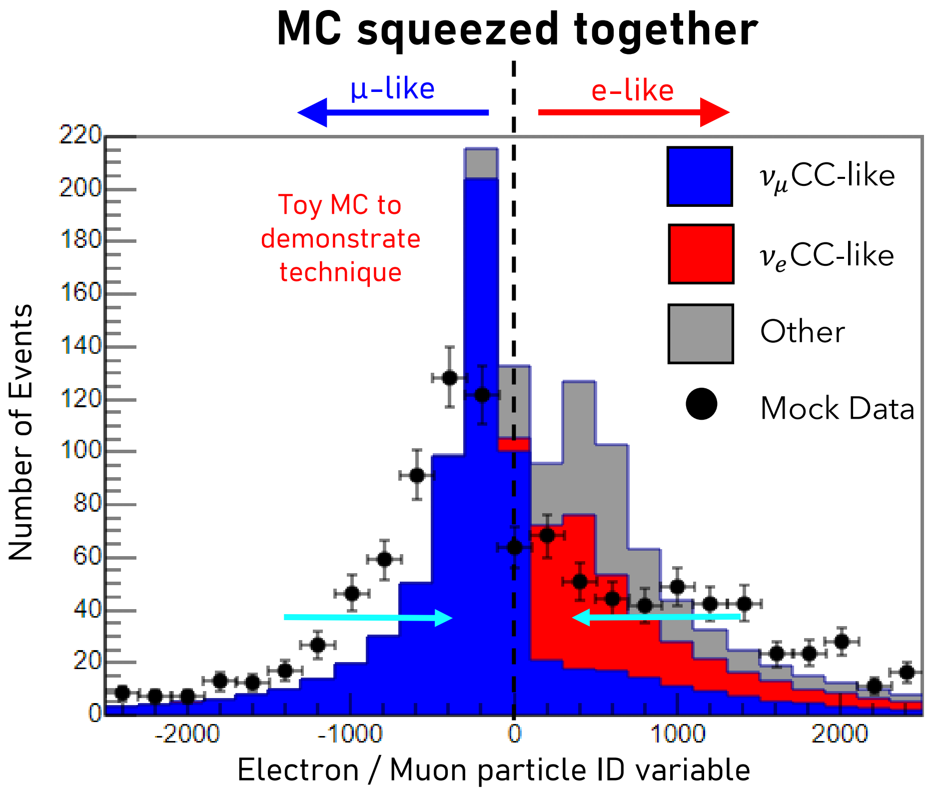




Toy particle ID spectrum

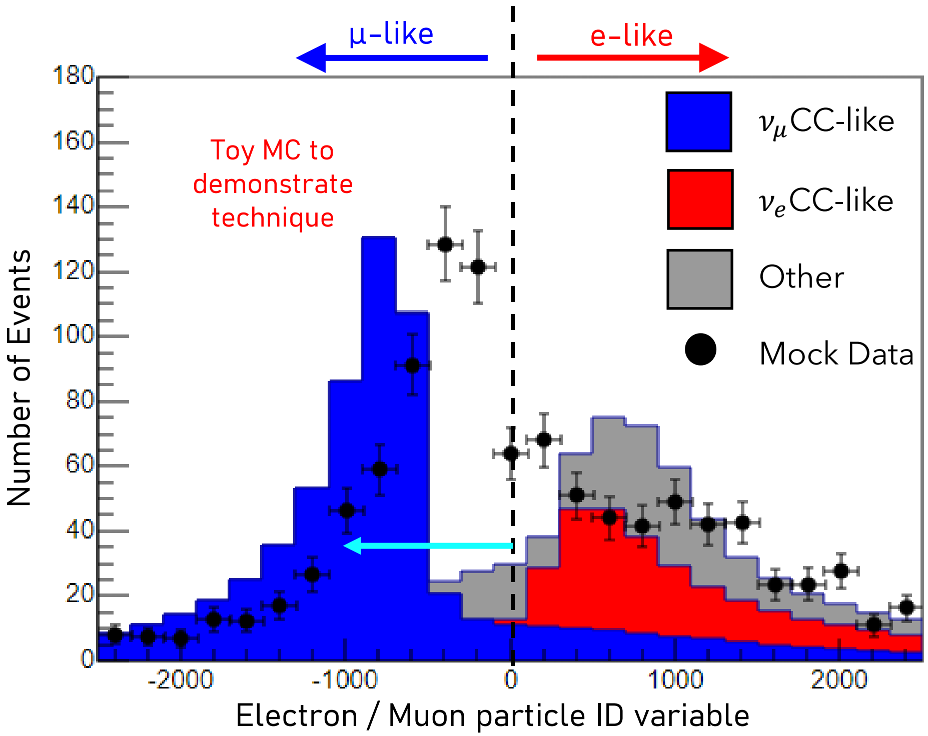
- Assume any mis-modelling will show at the particle ID level as systematic shifts in particle ID variable value,
- and/or as smearing out of the distribution





Toy particle ID spectrum

- Assume any mis-modelling will show at the particle ID level as systematic shifts in particle ID variable value,
- and/or as smearing out of the distribution
- These effects may be different for different particles in the detector



Fit parameterization

- Parameterize a fit to encode these systematic "shifts" and "smears"
  - Modify particle ID variables directly for MC

$$\underbrace{L}_{\text{Particle ID (e.g., electron/muon particle ID)}} \to \underbrace{\alpha}_{\text{Smear}} L + \underbrace{\beta}_{\text{Shift}}$$

- Move simulated event particle IDs around to try to make our simulation look like the data
- Post-fit $\{\alpha, \beta\}$ set will encode the underlying reconstructed mis-modelling between the simulation and data
- Different $\{\alpha, \beta\}$ for different event types (e.g. electron or muon) and different energy ranges

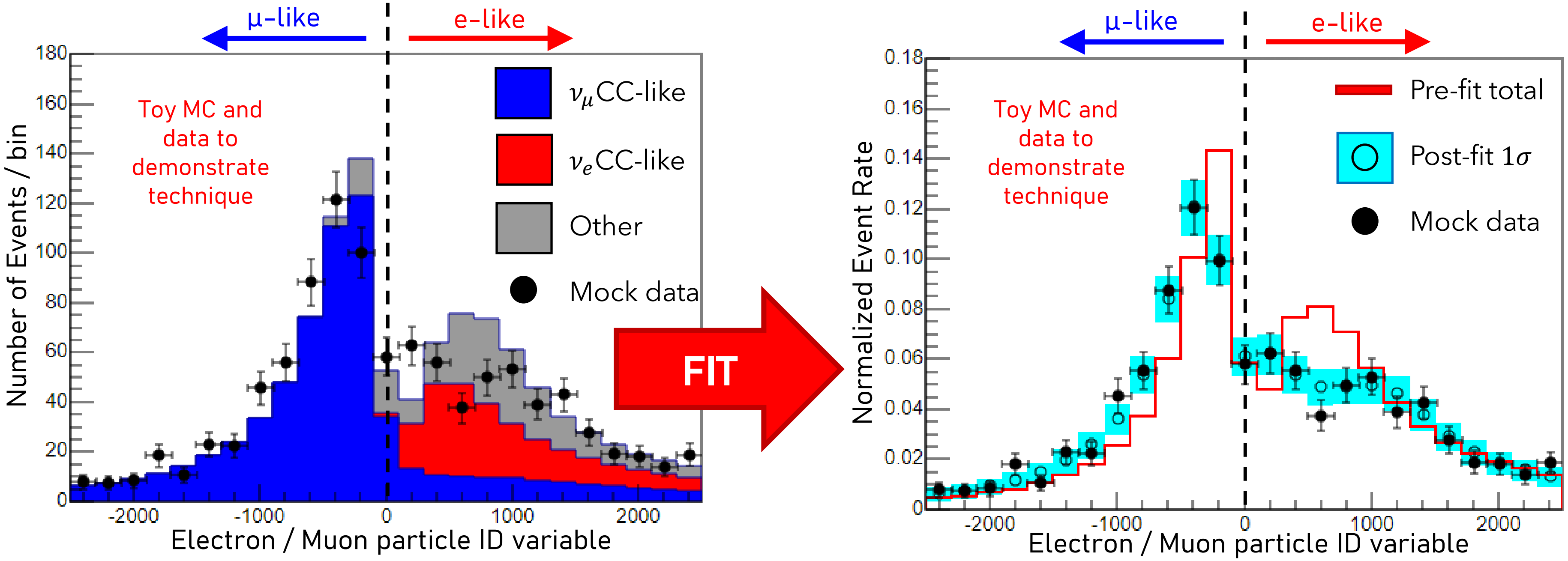
What data do we use?

- Can't use T2K beam data since we'd be over-fitting
- No good control samples span the full T2K kinematic range
- SK has a robust atmospheric neutrino program with data that spans the T2K energies with high statistics

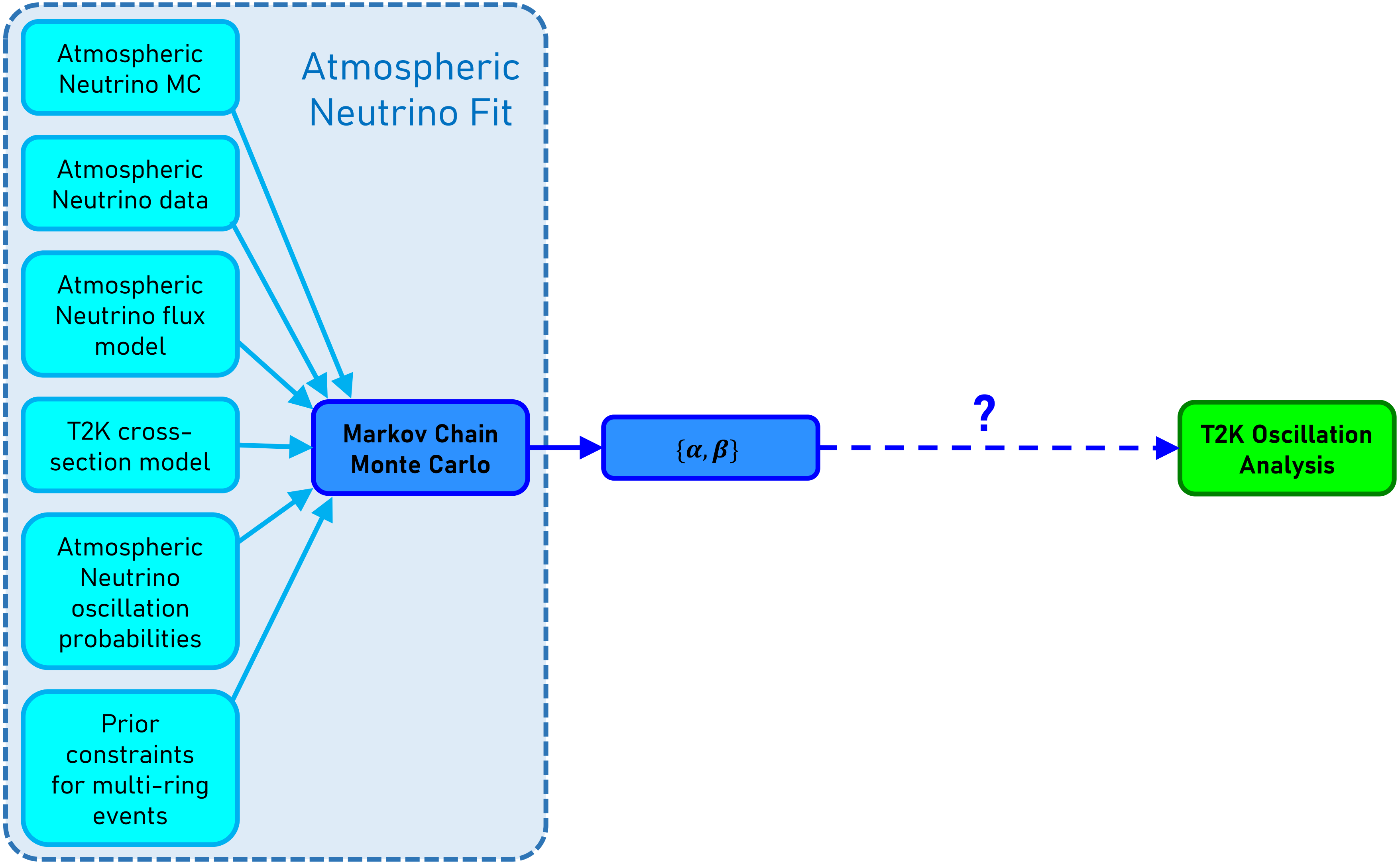
- Strategy: fit the shift and smear parameters (α, β) using SK atmospheric data and MC, then propagate the results to T2K beam data and MC

How do we fit?

- Fit the shape likelihood of the total MC histograms to the data histograms by modifying MC with $\{\alpha, \beta\}$
  - Bin histograms in terms of our particle ID variables
- Markov Chain Monte Carlo framework samples the shape likelihood
  - Results in a set of $\{\alpha, \beta\}$ that is distributed according to the likelihood distribution

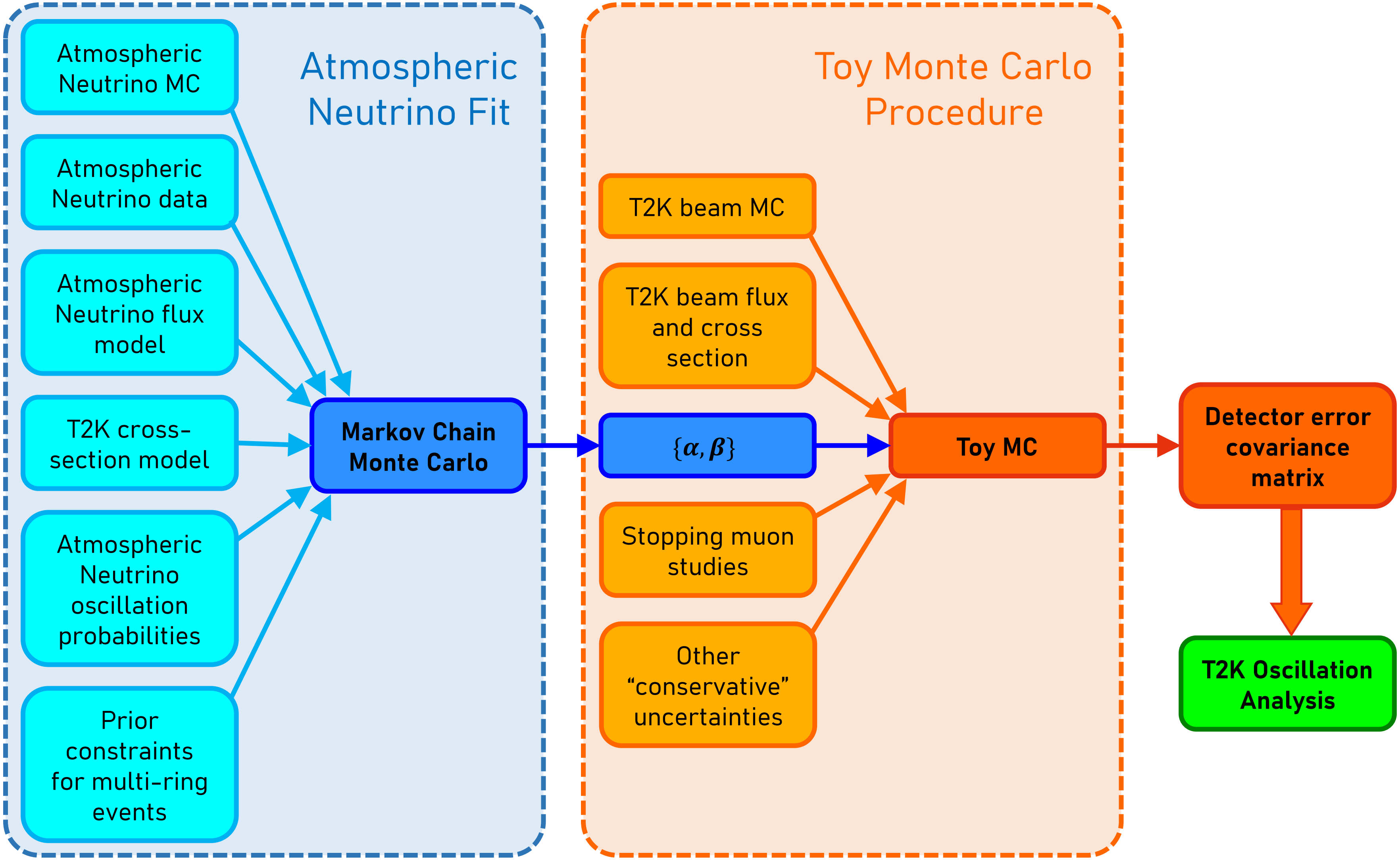


Fit overview



Fit overview

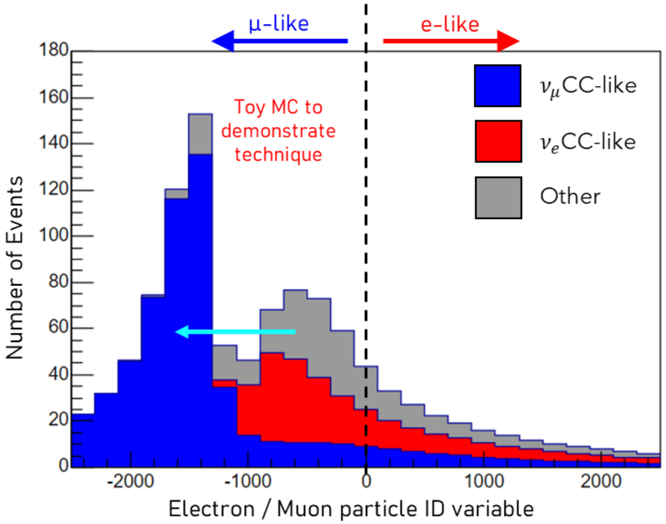
Boulder



Toy MC procedure

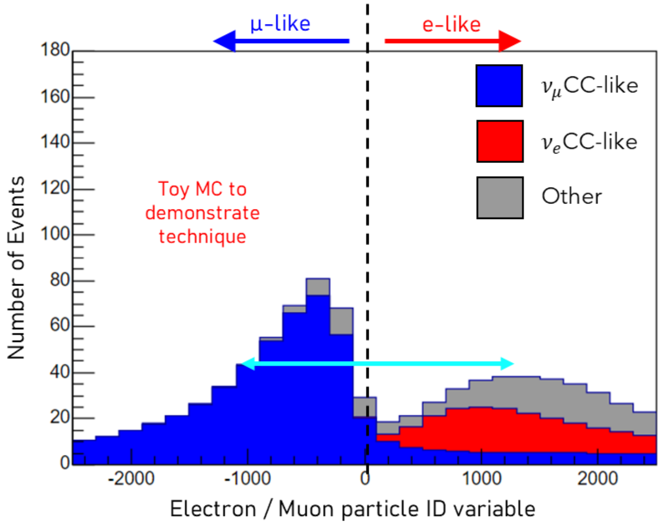
Boulder

- Take random samples from the $\{\alpha, \beta\}$ and apply them to the T2K beam MC
  - Since the T2K beam event samples are cut based on the particle ID variables, the $\{\alpha, \beta\}$ will shift some events into and out of the signal samples
  - Uncertainty in the count for each sample is encoded into a covariance matrix



$\alpha = 1, \quad \beta = -1000$

 # μ-like: 774, # e-like: 226



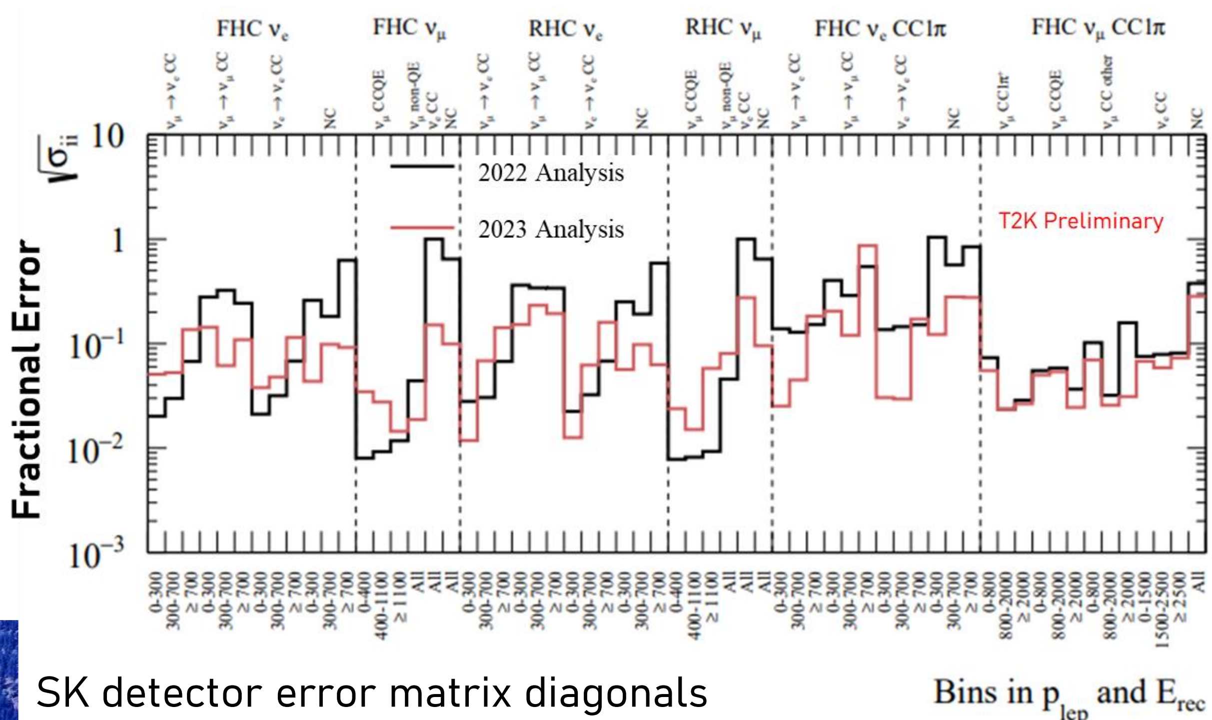
$\alpha = 2, \quad \beta = 0$

 # μ-like: 526, # e-like: 474

Count the number of ν_e and ν_μ events using many $\{\alpha, \beta\}$ from the MCMC to get error estimate

SK Detector Error Matrix

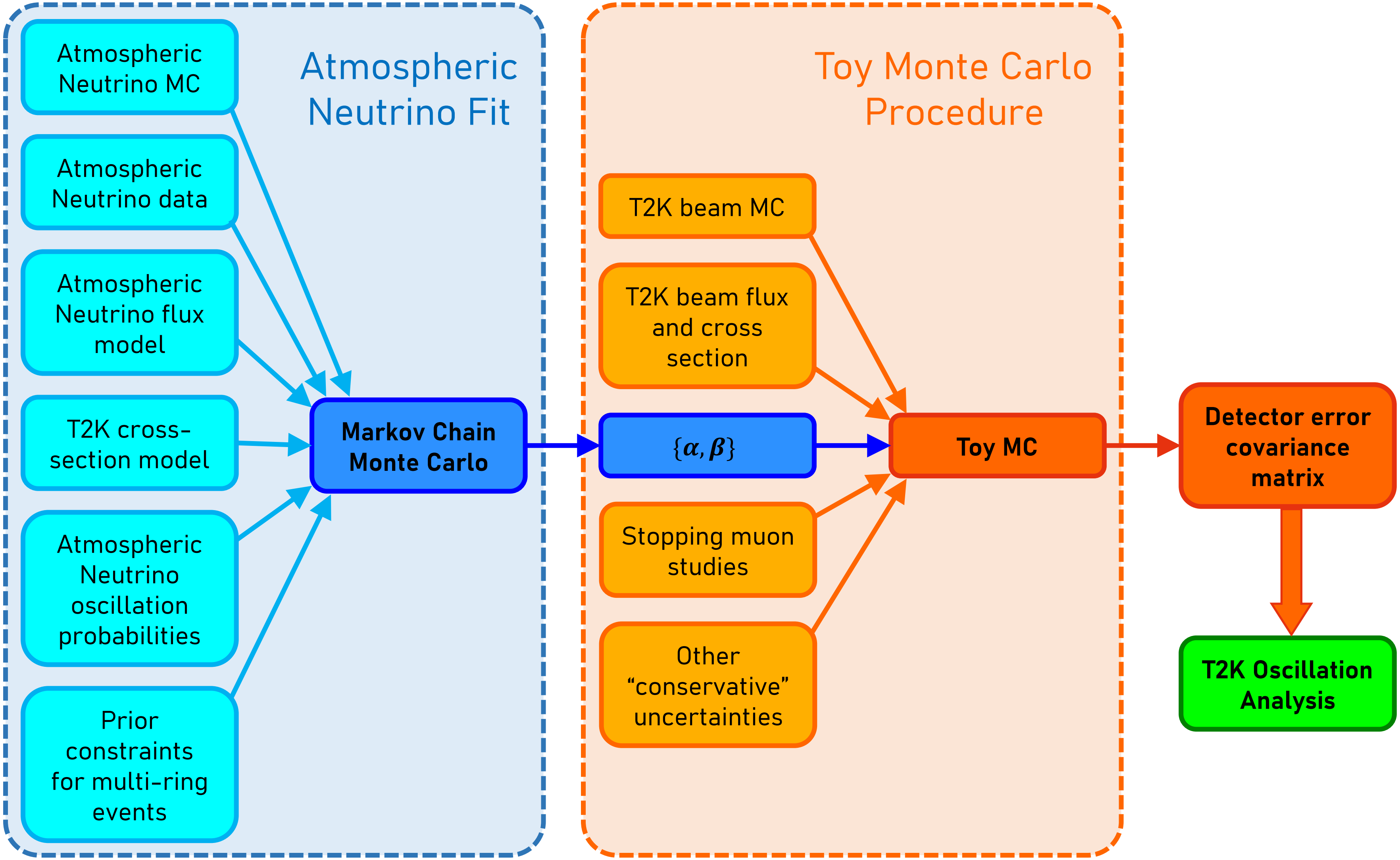
- 2023 analysis had many changes from 2022 that essentially only affected the single-ring samples in any significant way (i.e., everything but the FHC $\nu_\mu CC1\pi$ sample)
  - Removed several large "ad-hoc" errors from single-ring samples
  - Added energy-dependent binning to single-ring samples
  - Removed an intermediate covariance matrix that was used for the T2K toy MC procedure (toy MC now samples directly from MCMC posterior)



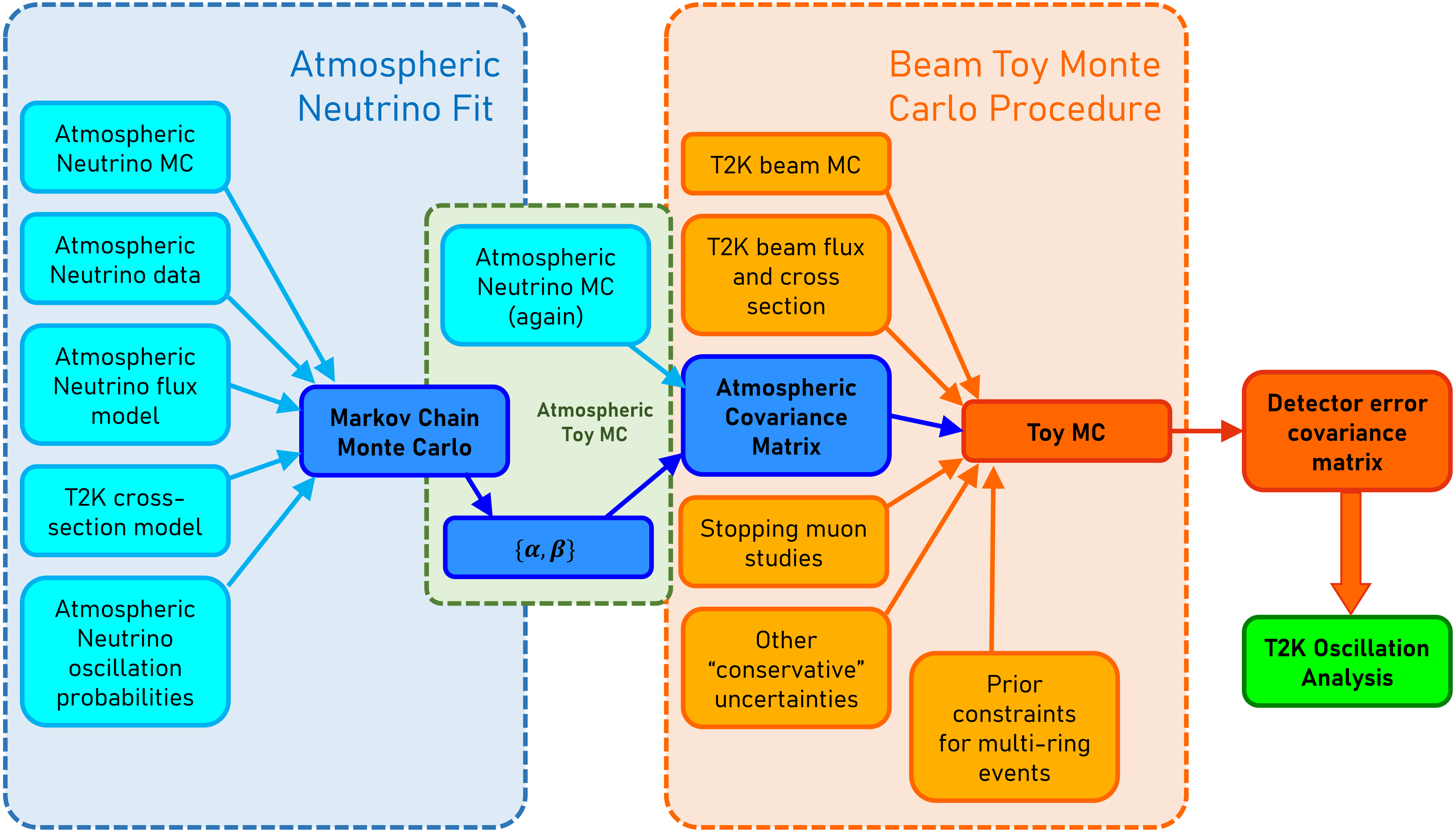
SK detector error matrix diagonals

Fit overview (2023)

Boulder



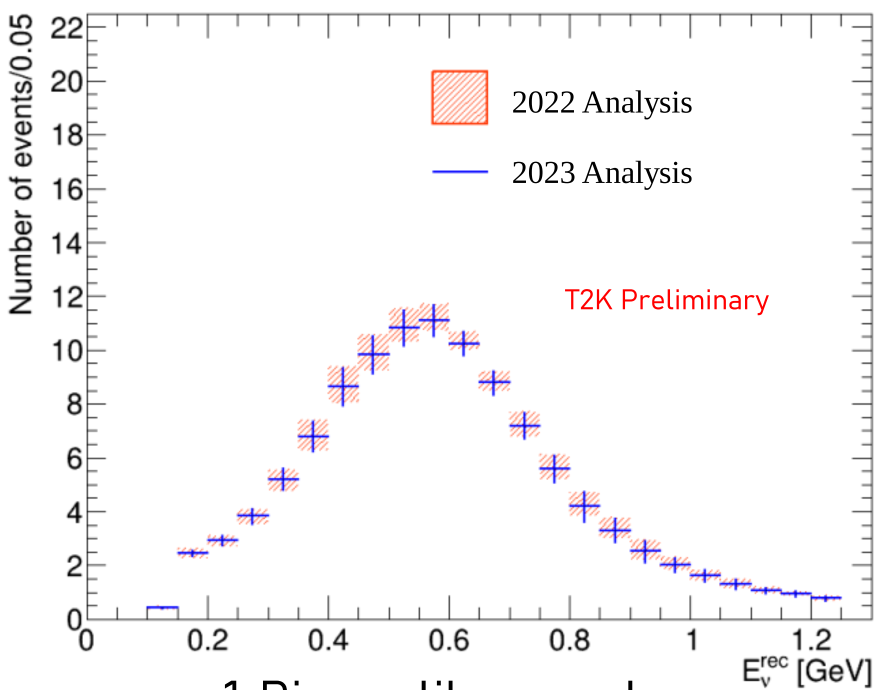
Fit overview (2022)



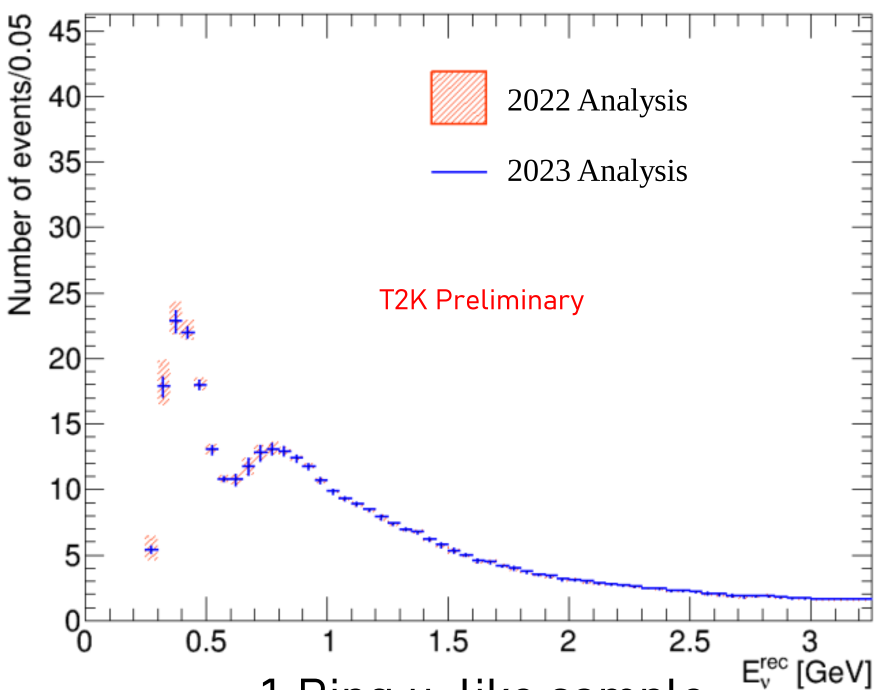
How is the error matrix used?

Boulder

- The covariance matrix provides an uncertainty on the number of events in each T2K-FD sample
- When T2K performs its oscillation analysis, the number of events in different T2K-FD samples are allowed to increase and decrease as determined by this uncertainty matrix



1 Ring e-like sample

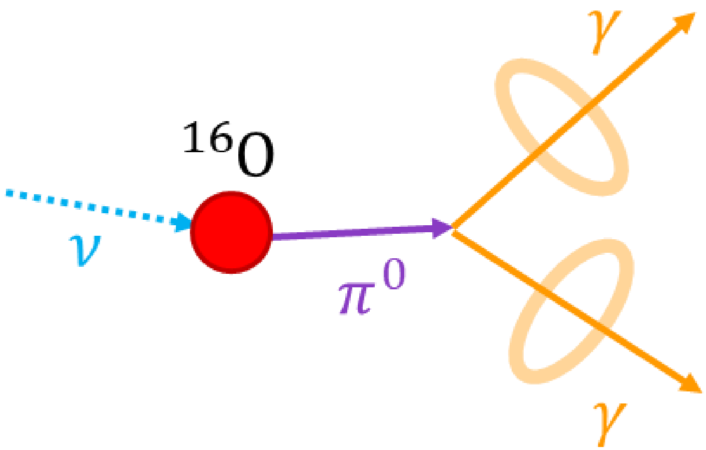


1 Ring μ-like sample

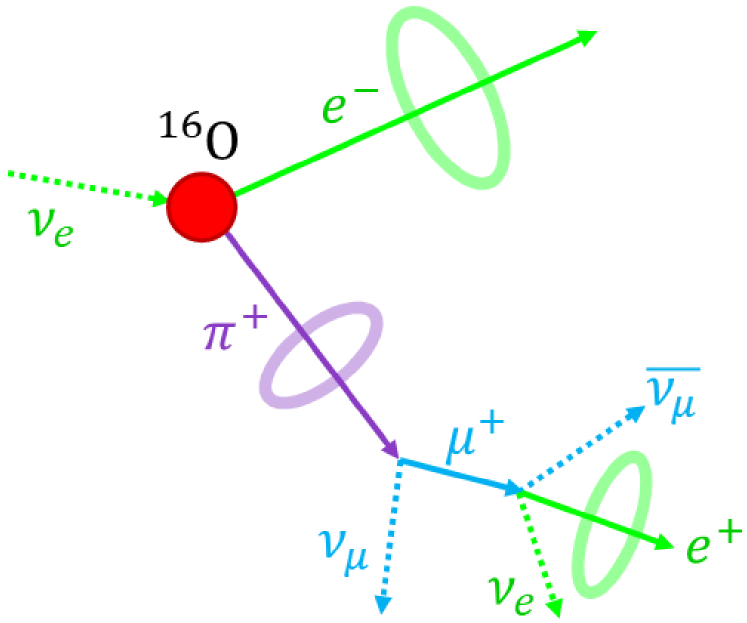
Posterior-predictive plots of T2K MC corresponding to 21.428×10^{20} POT (neutrino mode)

New additions post-2023

- Adding new particle ID variables for selecting future multi-ring T2K analysis samples
  - A new $NC\pi^0$ sample will help constrain π^0 backgrounds in ν_e samples
  - A new $\nu_e CC1\pi^+$ sample will add more ν_e signal statistics



$NC\pi^0$ cartoon



$\nu_e CC1\pi^+$ cartoon

New additions post-2023

Boulder

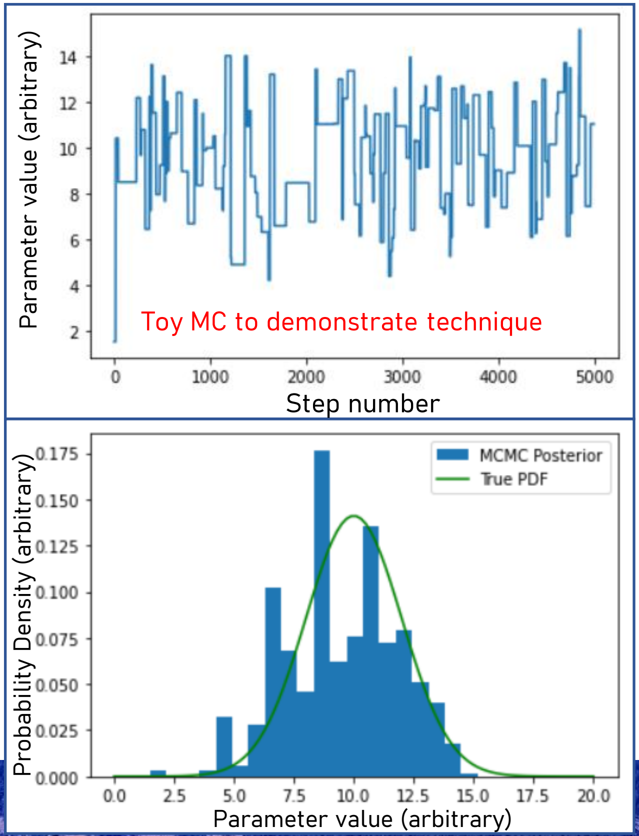
- New MC production with new interaction model
  - Previously Relativistic Fermi Gas-based model (Improved constraints on neutrino mixing from the T2K experiment with 3.3x10^21 protons on target)
  - Now use Spectral Function-based model (Measurements of neutrino oscillation parameters from the T2K experiment using 3.6x10^21 protons on target)

New additions post-2023

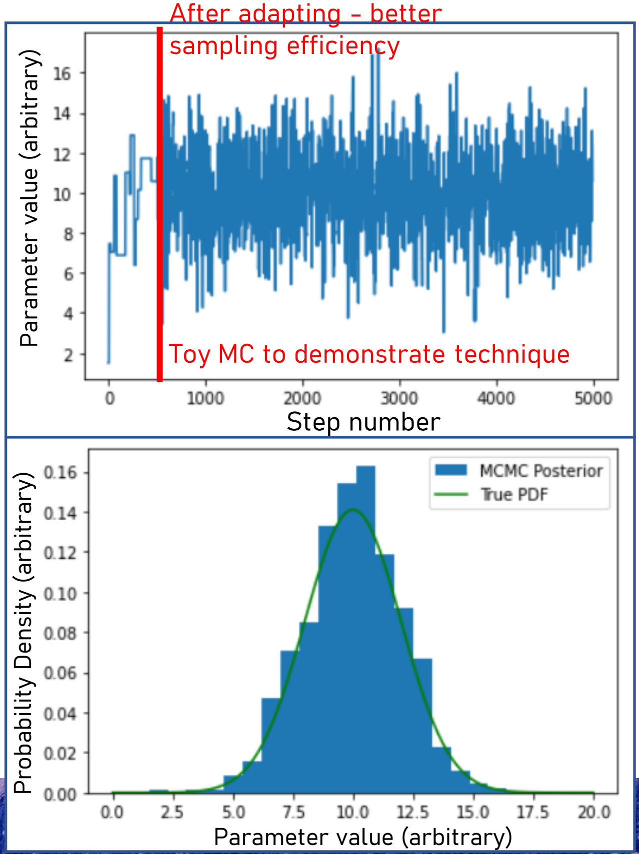
Boulder

- Overhaul of code to increase speed
  - Introduced multithreading and reduced code bloat
  - Previously tuned MCMC by hand, now use [Adaptive Metropolis]() to automatically tune MCMC
  - ~10x faster from OA2023 to now!

Poorly tuned MCMC



Adaptive MCMC



Last word

- The uncertainties are calculated from SK atmospheric neutrino data and MC, then applied to T2K beam MC
- Uncertainties are parameterized using a unique set of “shift” and “smear” parameters
- Error analysis focuses on modifying T2K particle ID variables

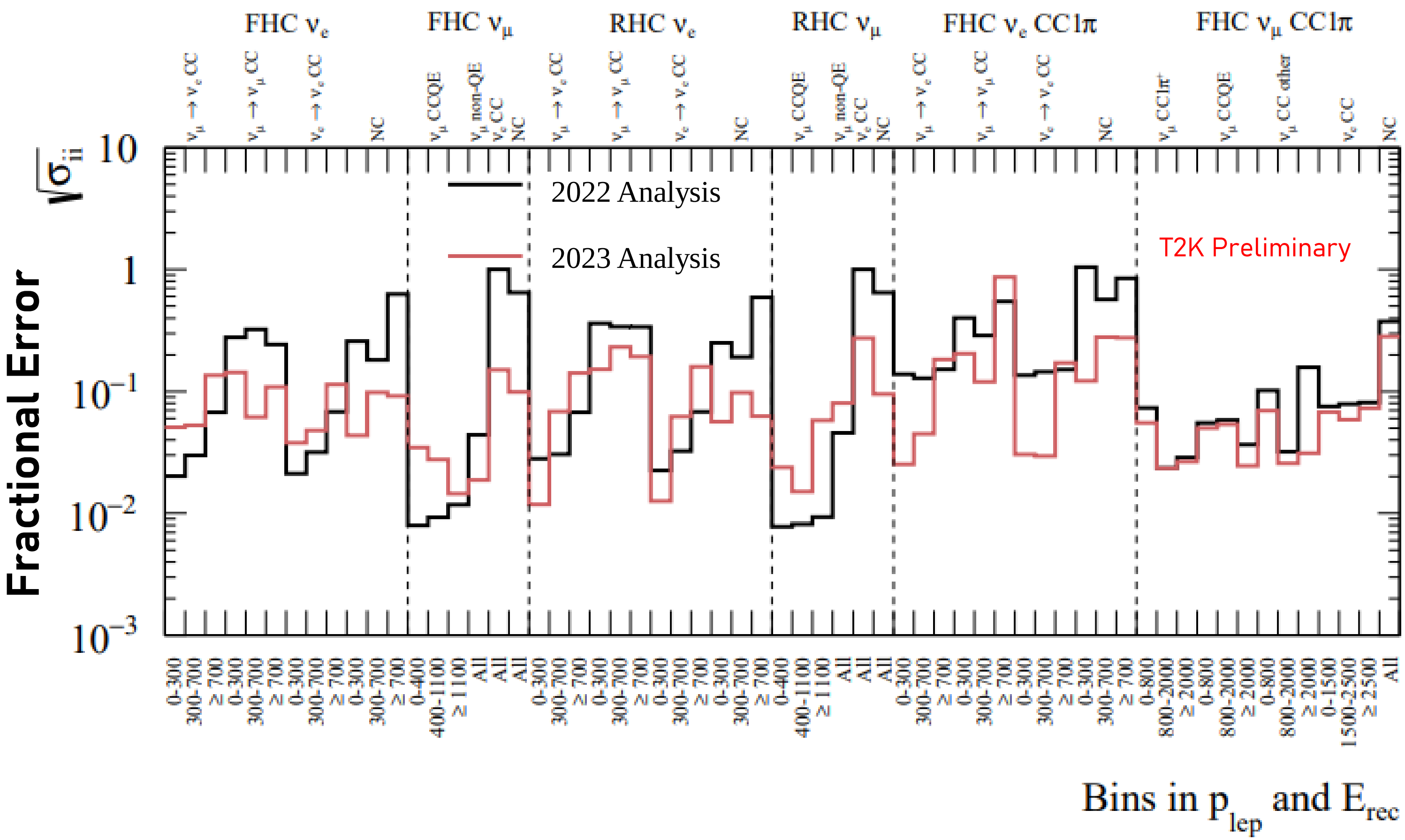
- New SK uncertainties will be one of the main updates for T2K’s upcoming oscillation analysis, in the absence of new data
- Stay tuned for future T2K results!

End

Backup

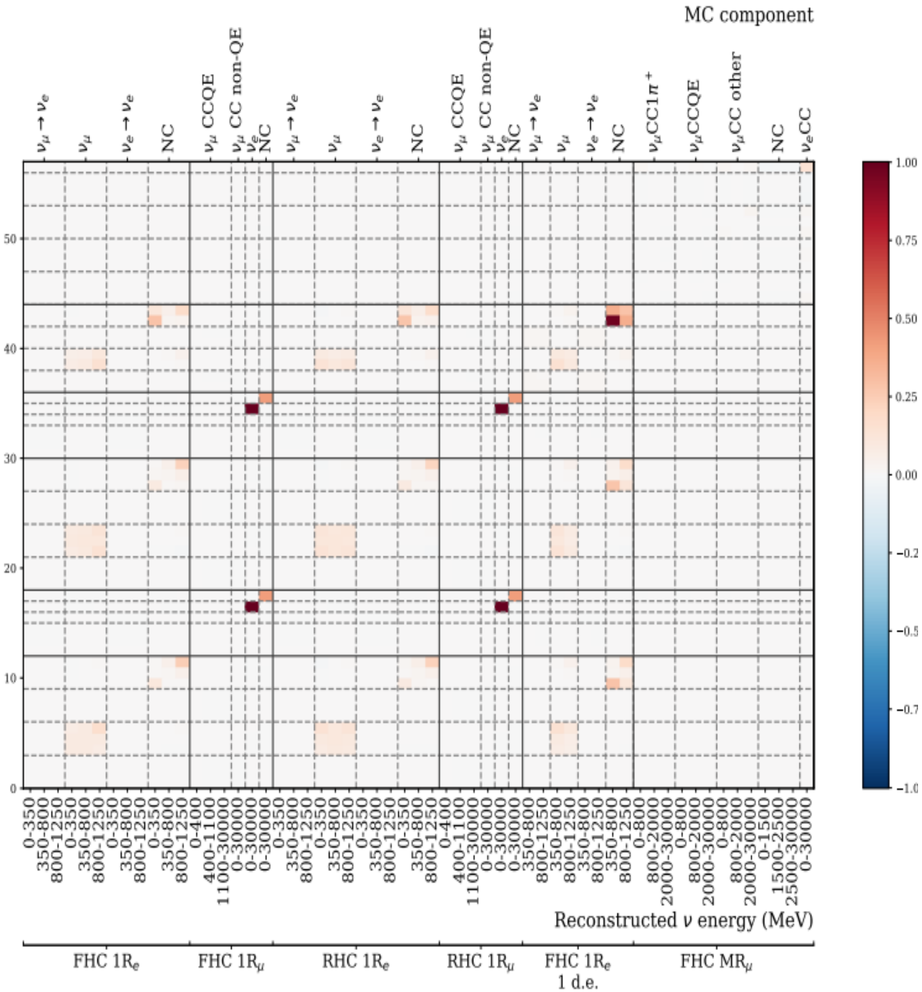
SK Detector Error Matrix

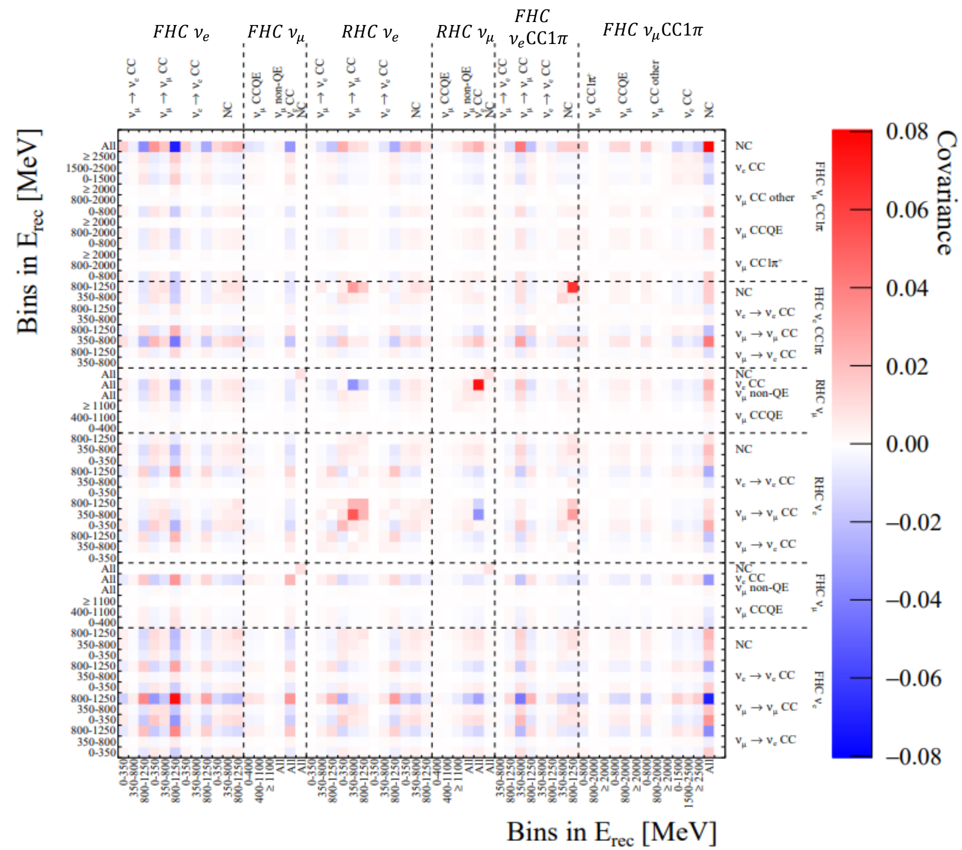
- Example SK detector error matrix diagonal values, comparing the 2022 analysis and 2023 analysis



SK Detector Error Matrix

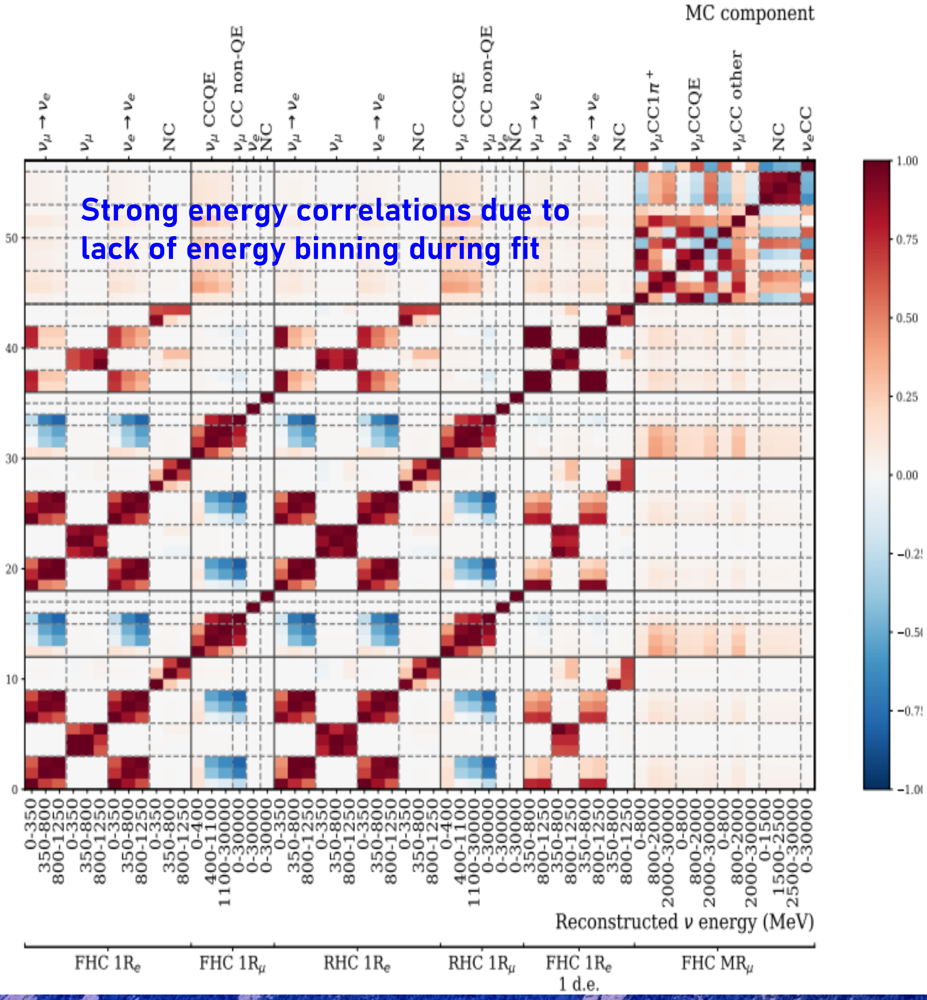
- 2022 analysis covariance matrix (left) vs 2023 (right)

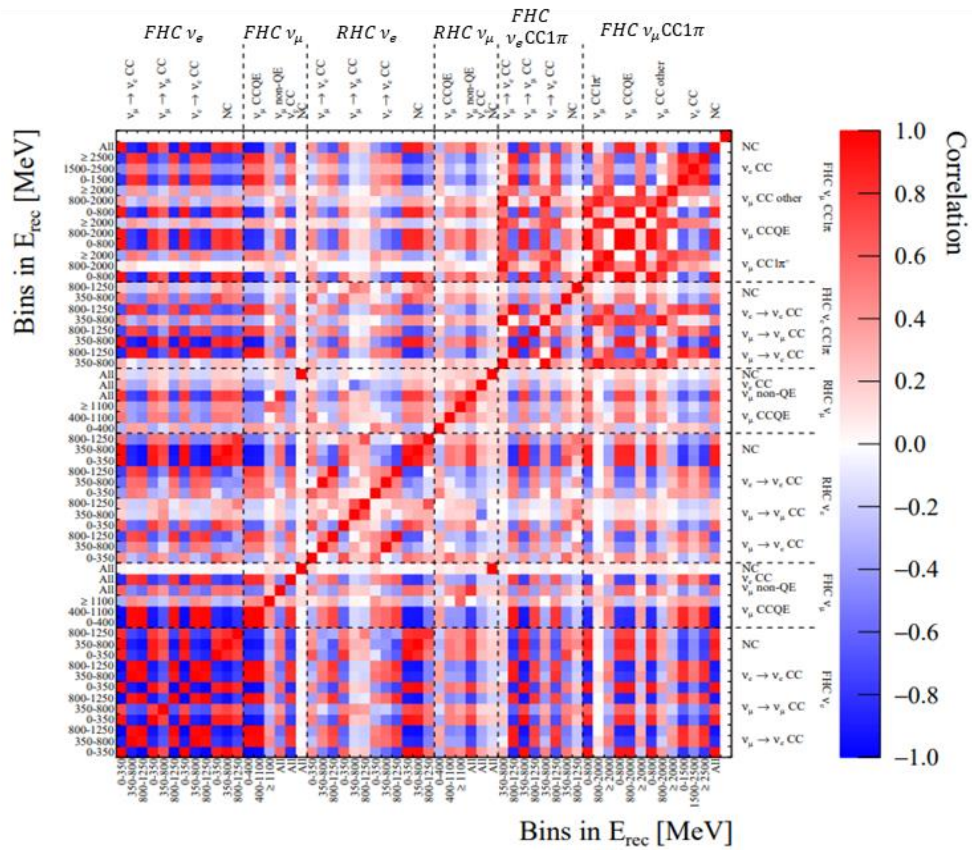




SK Detector Error Matrix

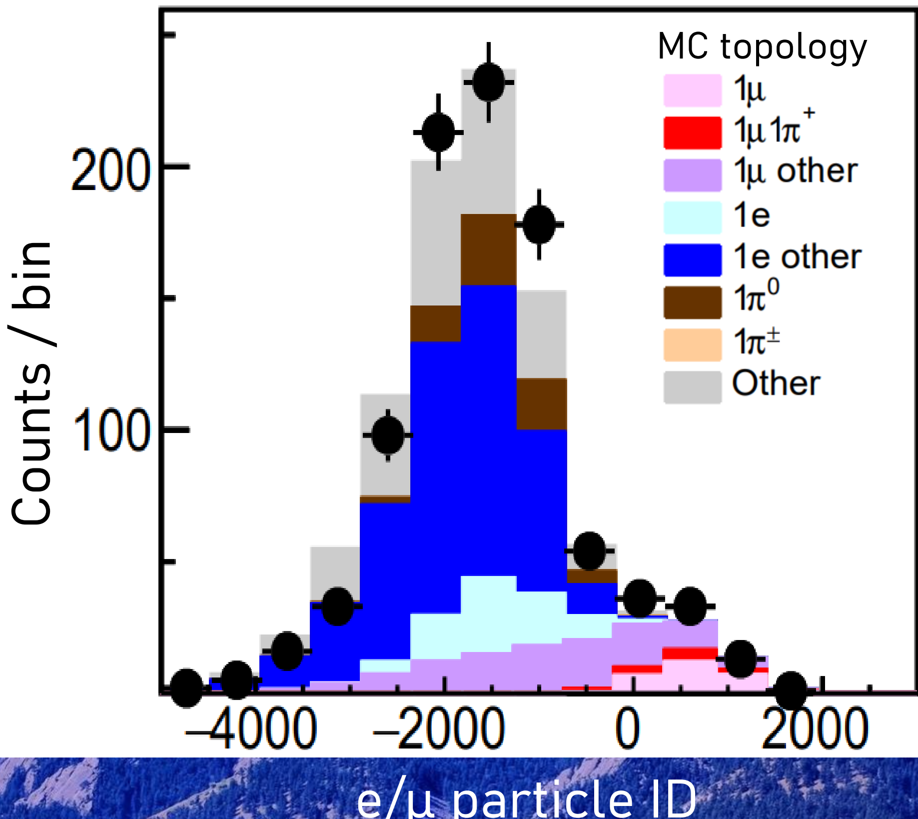
- 2022 analysis correlation matrix (left) vs 2023 (right)

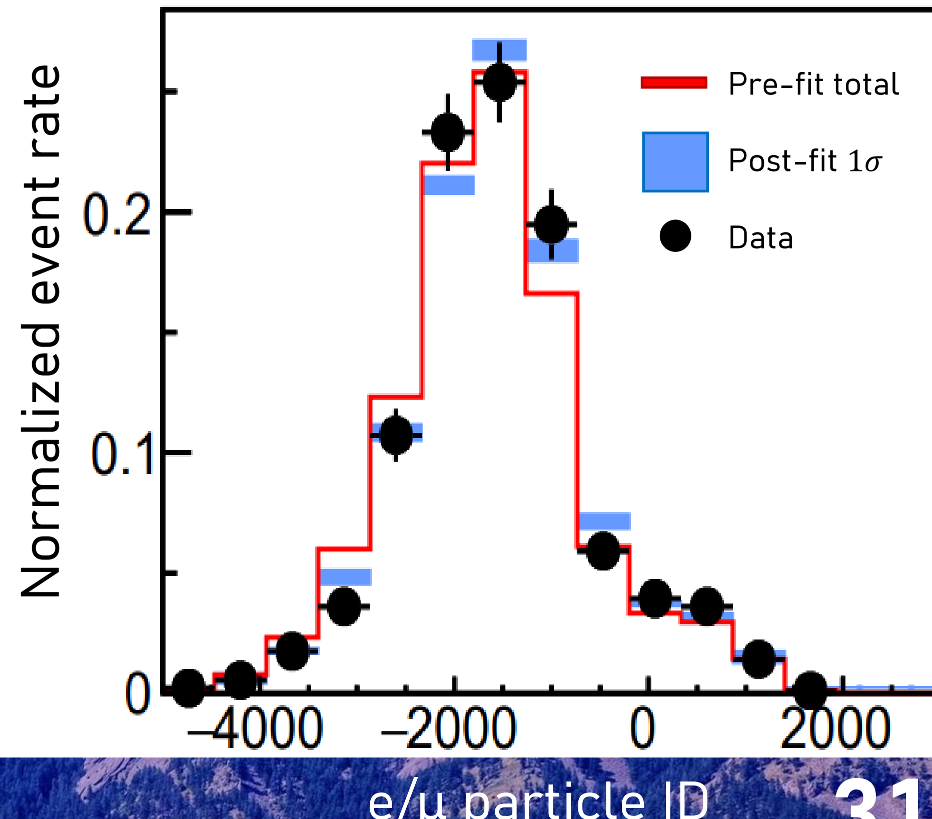




Real data / MC comparison

- Example data / MC comparison plot for events with…
  - More than 1 Cherenkov ring
  - 0 delayed Michel electrons
  - Visible energy between 700 and 1330 MeV

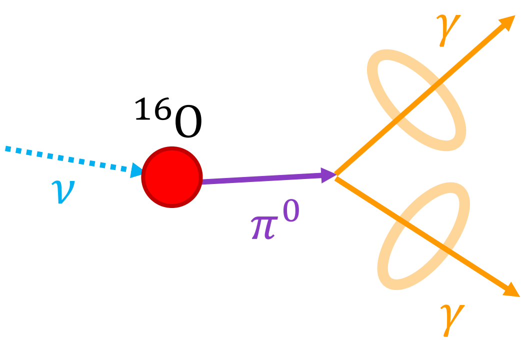




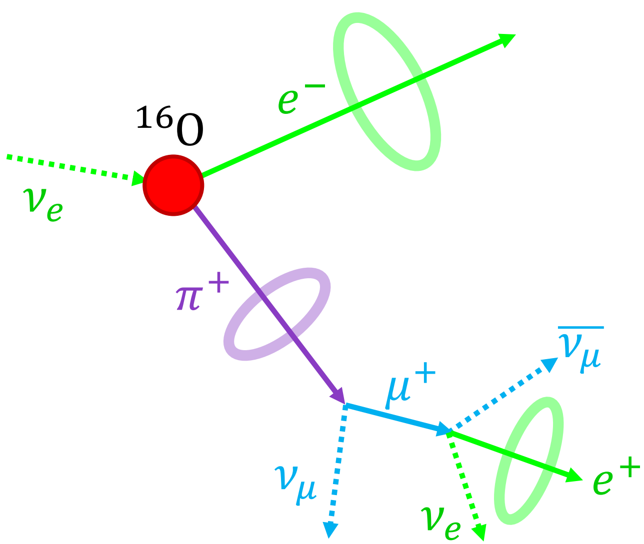
New additions

Boulder

- Adding new particle ID variables for future multi-ring T2K analysis samples ($\nu_e CC1\pi^+$, $NC\pi^0$)



$NC\pi^0$ cartoon



$\nu_e CC1\pi^+$ cartoon