Imperial College London

MaCh3 and Bayesian Analysis

Patrick Dunne

Imperial College London

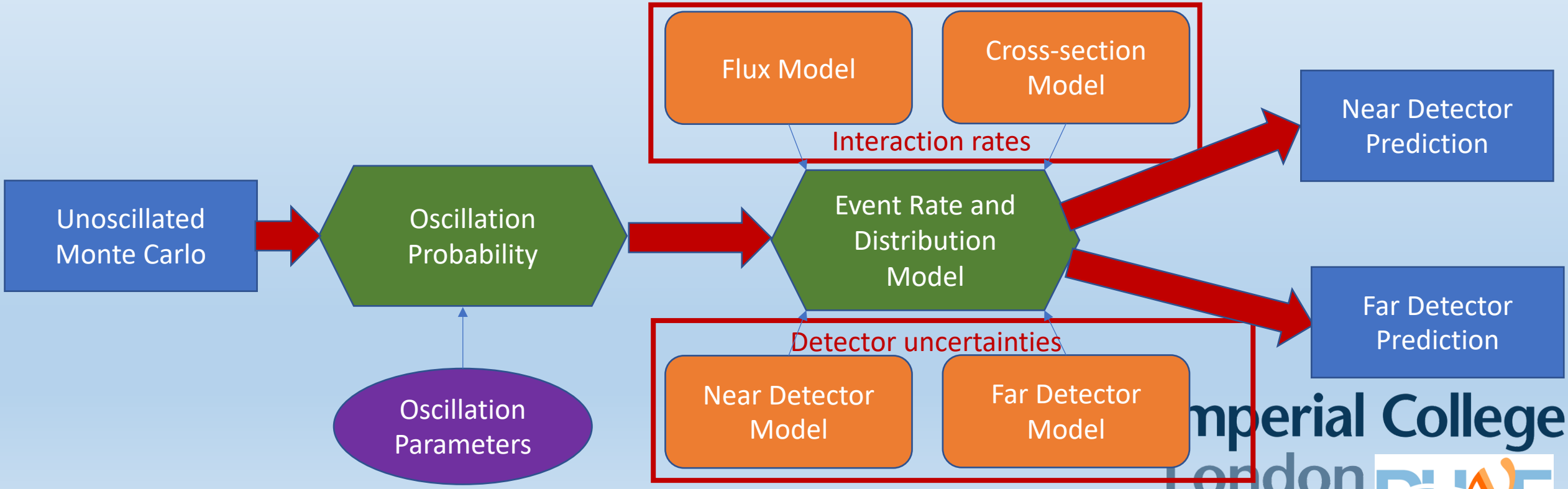
Outline

- Introduce T2K method for analysis
- How to interpret Bayesian results
- Describe MaCh3 framework
  - How do you put a new detector in?
  - How long does it take to run?

Imperial College London

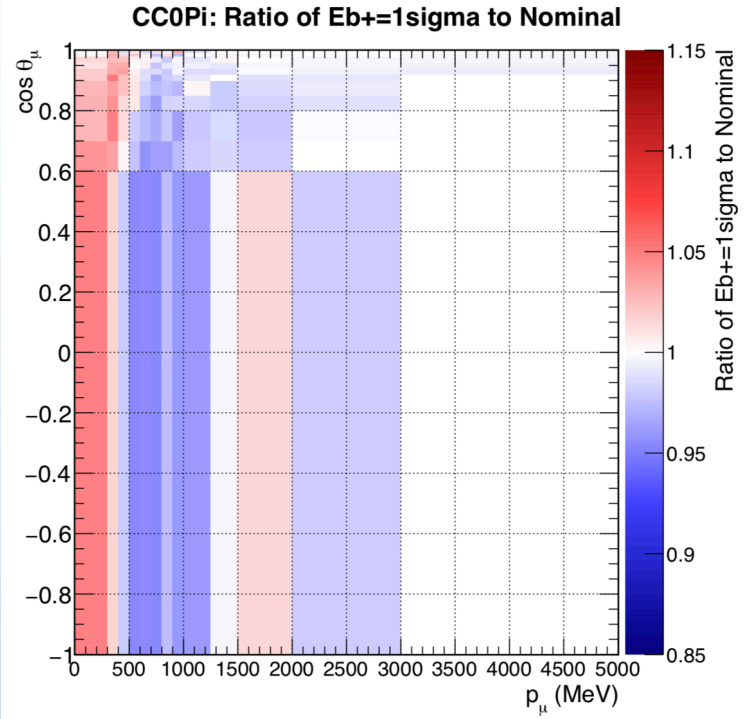
T2K oscillation parameter fit

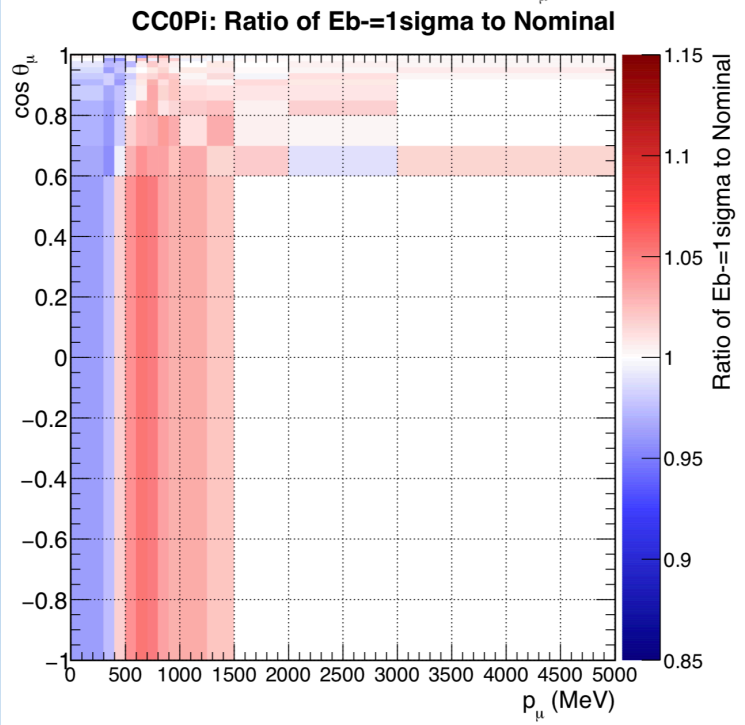
- Apply oscillation effects to Monte Carlo as a function of true E_ν
- Construct model to predict event rates and distributions at near and far detectors
- Need to ensure experiment can constrain non-oscillation elements of model
  - Cross-section model highly dependent on nuclear effects
- Do this by fitting rather than direct ND distribution to FD distribution extrapolation



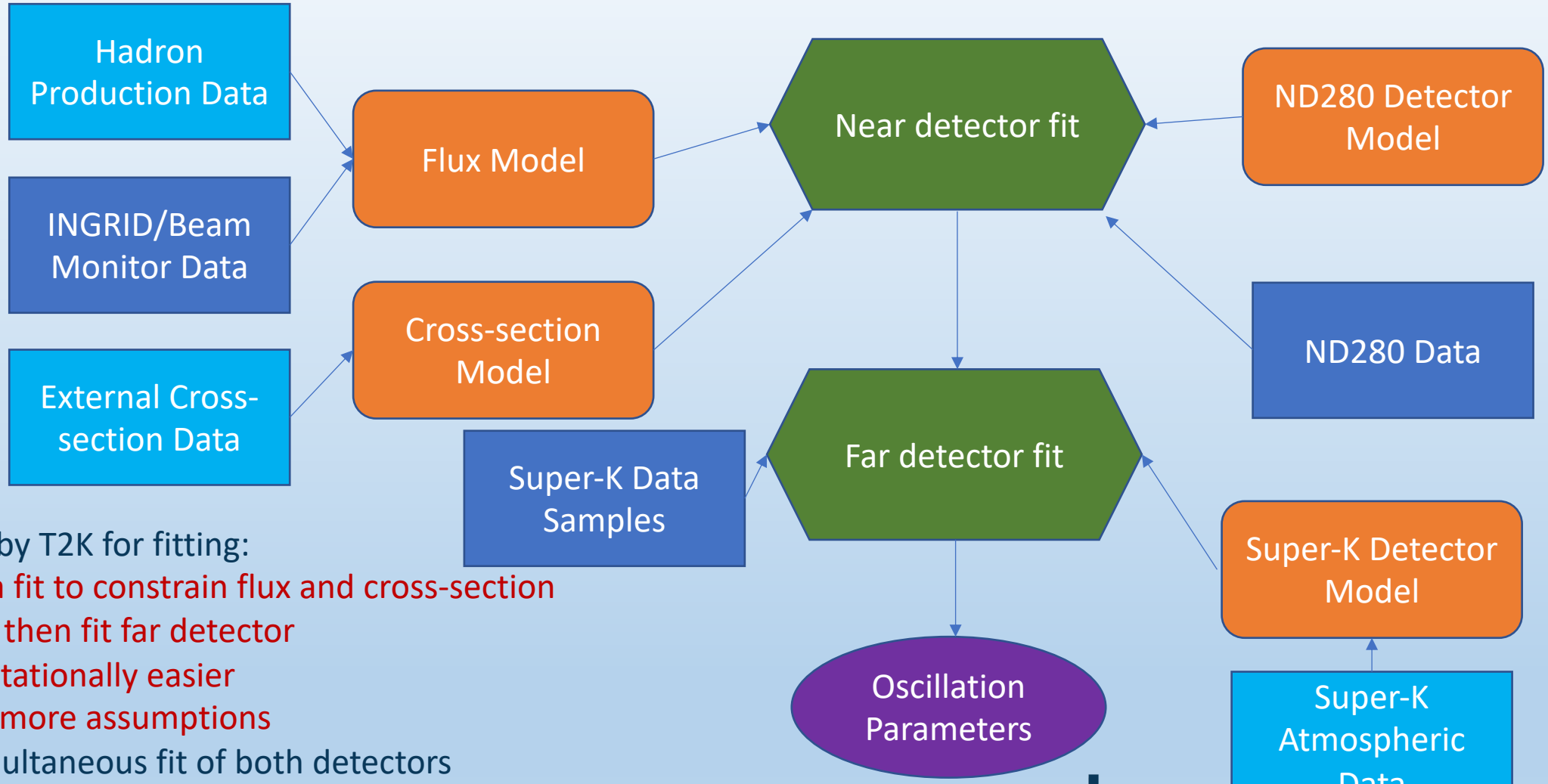
Reweighting

- MaCh3 has full access to event-by-event kinematic information
  - Enables reweights with functional forms on any event variable
  - Also enables shifts of variables e.g. Removal energy can actually change event momentum and put in a different bin
- We also have standard bin by bin (separated by mode, flavor) response functions and linear normalisations implemented





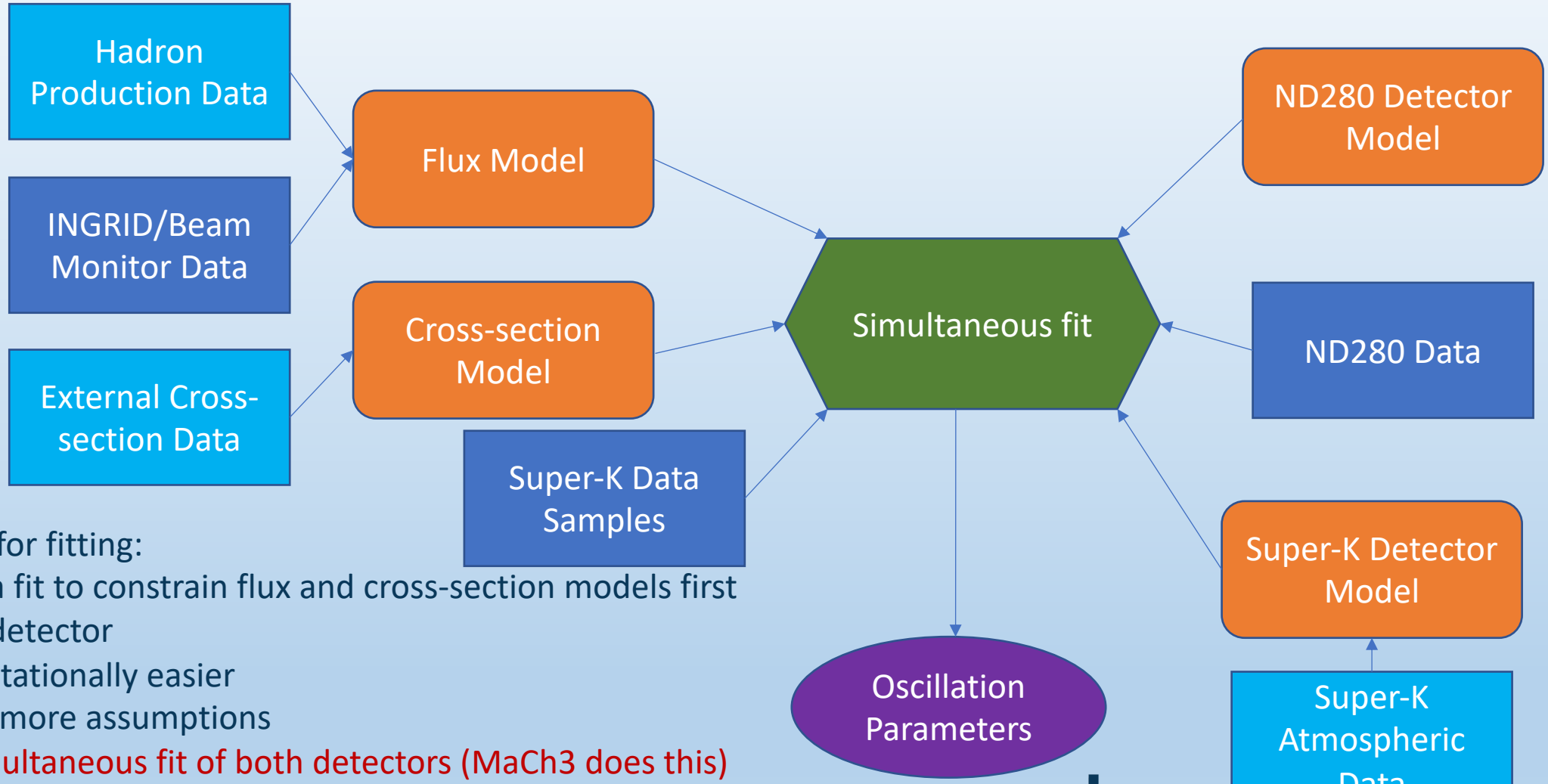
Fitting to data



Two approaches used by T2K for fitting:

1. Use ND data fit to constrain flux and cross-section models first then fit far detector
   - Computationally easier
   - Makes more assumptions
2. Perform simultaneous fit of both detectors
   - Computationally more demanding
   - Makes fewer assumptions

Fitting to data

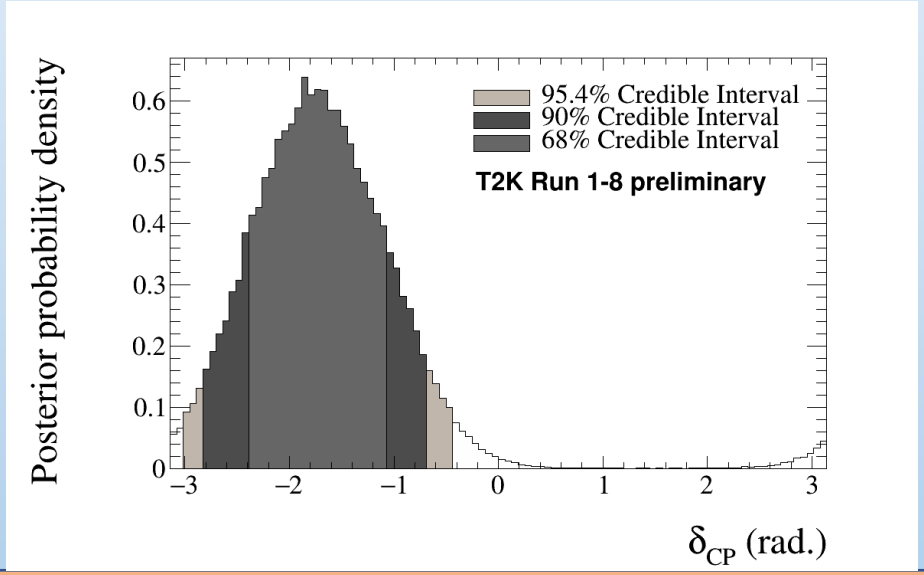


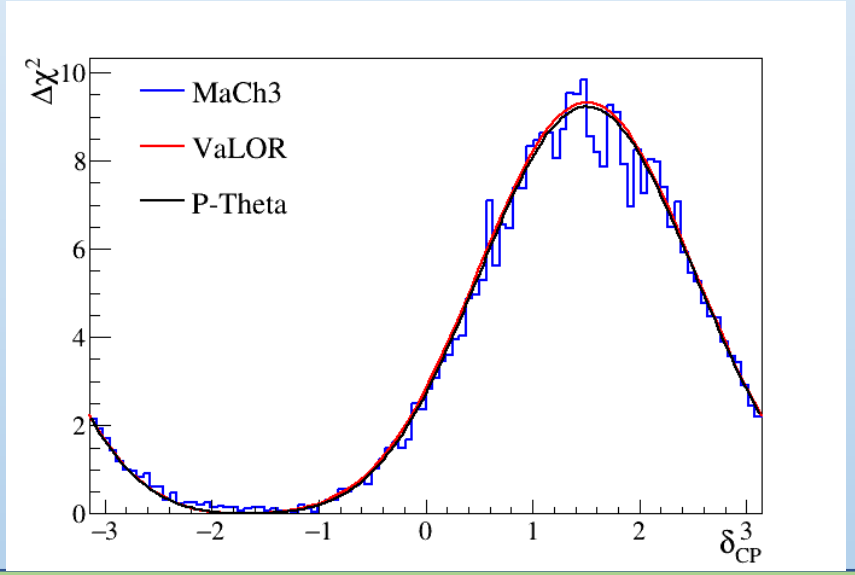
Two approaches used for fitting:

1. Use ND data fit to constrain flux and cross-section models first then fit far detector
   - Computationally easier
   - Makes more assumptions
2. Perform simultaneous fit of both detectors (MaCh3 does this)
   - Computationally more demanding
   - Makes fewer assumptions

T2K analyses

- T2K has three separate analysis frameworks: two fit near detector first and propagate, one does joint fit
- Joint fit analysis is Bayesian, one of separate fitters is frequentist and the other is a mix
- All three able to construct frequentist confidence intervals for comparisons
  - Very good agreement is seen, cross-validation highly useful for debugging





Bayesian analysis shows posterior probability density

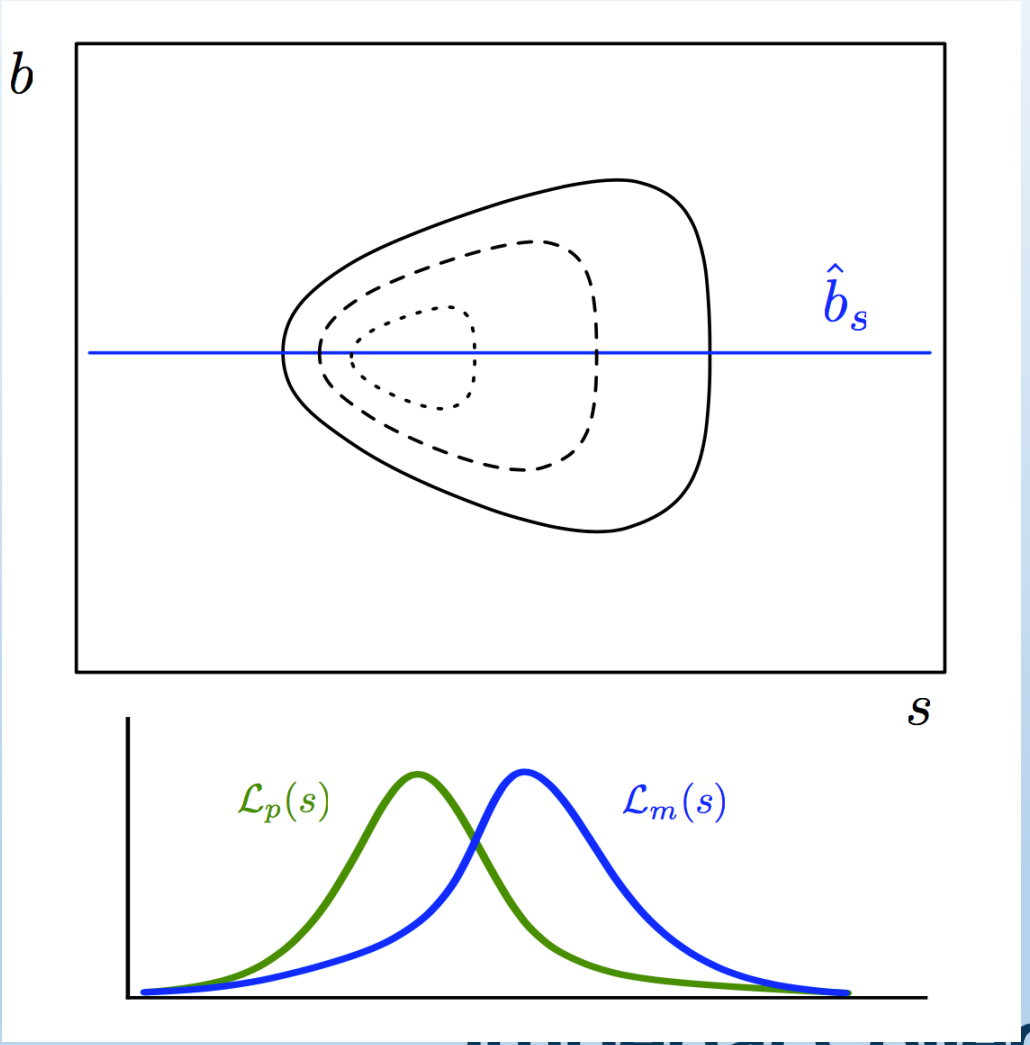
(high values mean more likely this is the "correct" parameter value)

Frequentist analyses show $\Delta\chi^2$

(low values mean better agreement with the data for this parameter value)

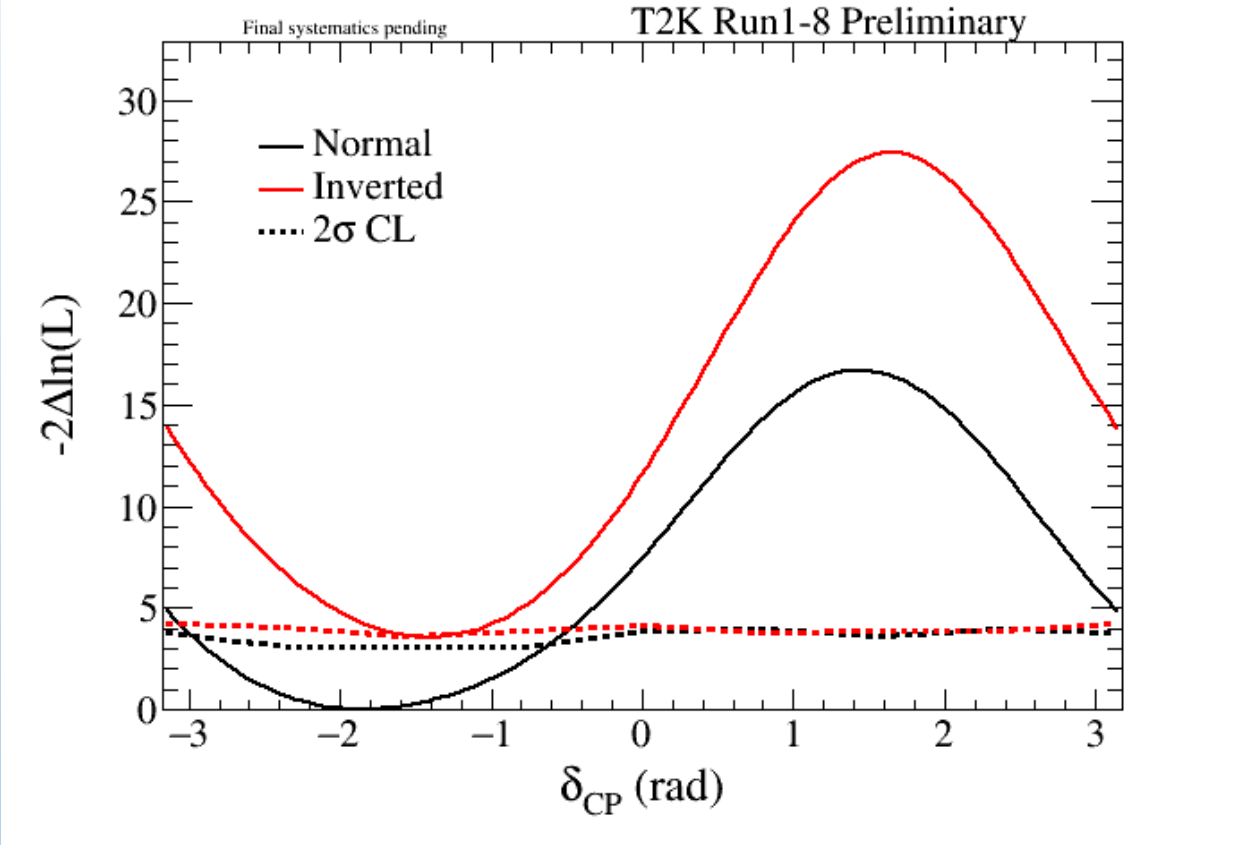
Dealing with nuisance parameters

- Likelihood has >750 parameters but want plots in ≤2 of them at once
- Two main options:
  - Profiling: Pick values of nuisance params that maximise likelihood for each set of values of parameters of interest
  - Marginalisation: Integrate over nuisance parameters (Bayesian so MaCh3 does this)
- Also finding maximum likelihood point for given osc par values is hard in 750 dimensions



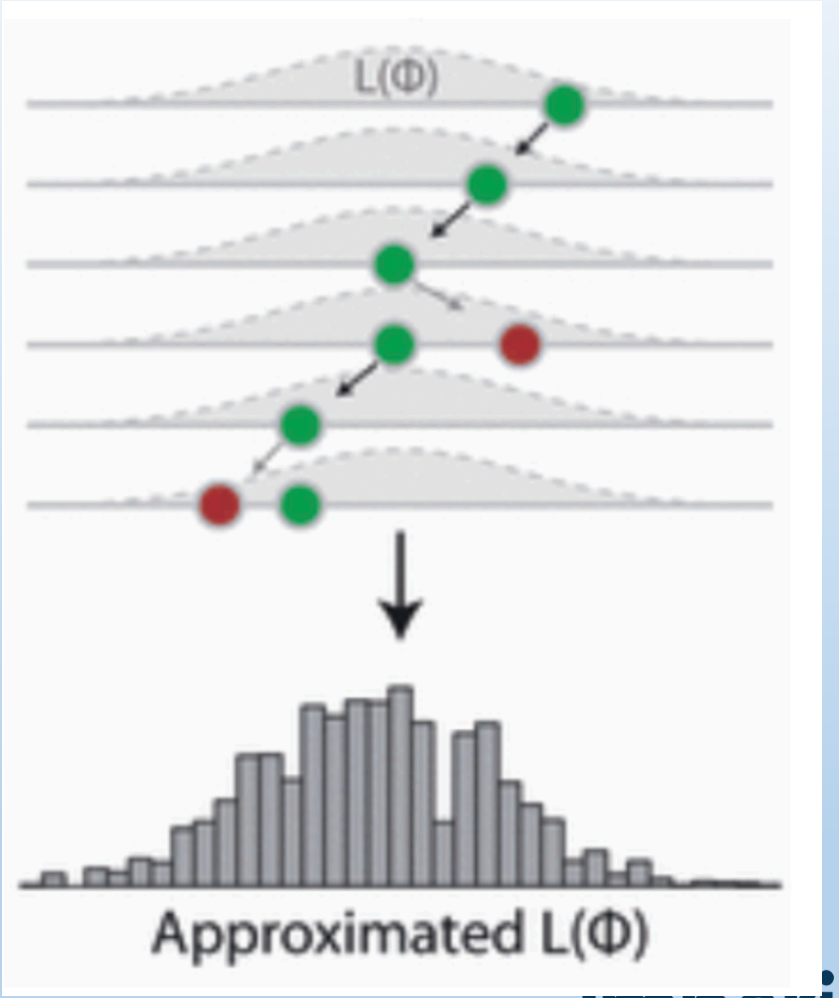
MCMC vs grid search

- Other T2K analyses use random throws of nuisance parameters from covariance matrices to marginalise
- Then do a grid search in 1D/2D calculating average $\Delta\chi^2$ across ensemble of marginalisation throws
- Use Feldman-Cousins to find critical $\Delta\chi^2$ values for δ_{CP}



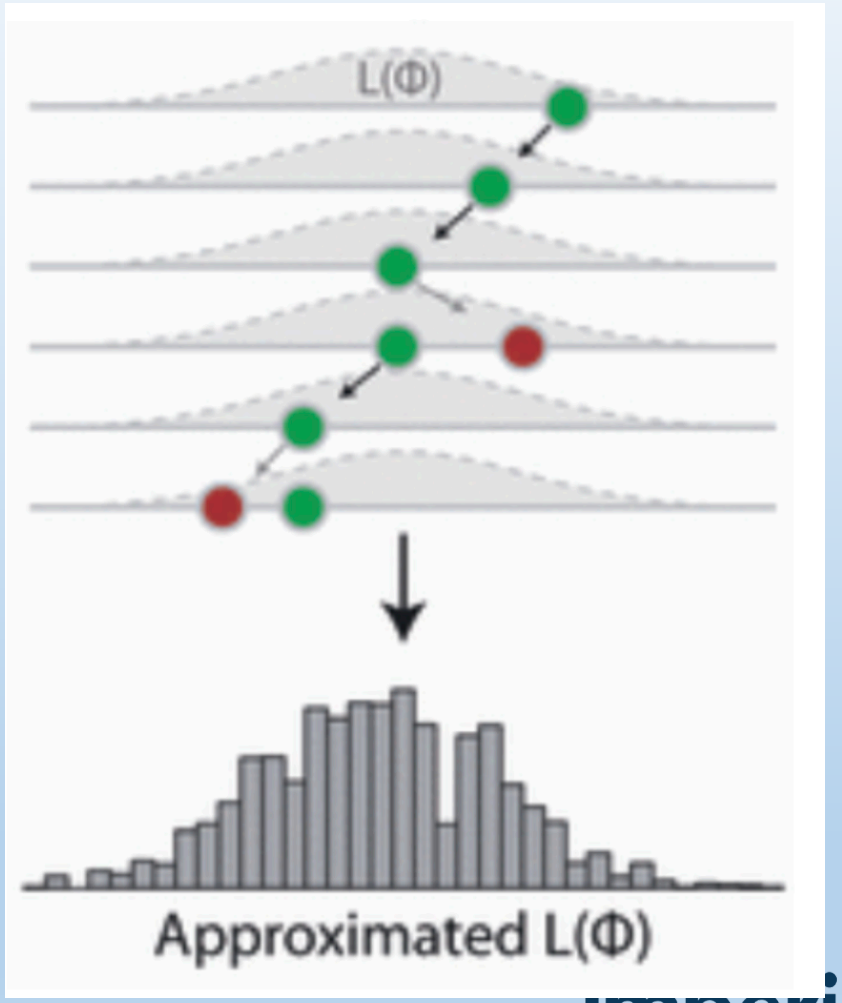
MCMC vs grid search

- MaCh3 samples likelihood space with Markov Chain MC
- Rule for stepping in parameter space ensures distribution of parameter values proportional to marginalised posterior probability
  - Targets likelihood evaluations in regions of space where likelihood is high
  - Several algorithms to choose from e.g. Metropolis-Hastings or Hamiltonian
  - MaCh3 currently uses Metropolis-Hastings and upgrading to Hamiltonian



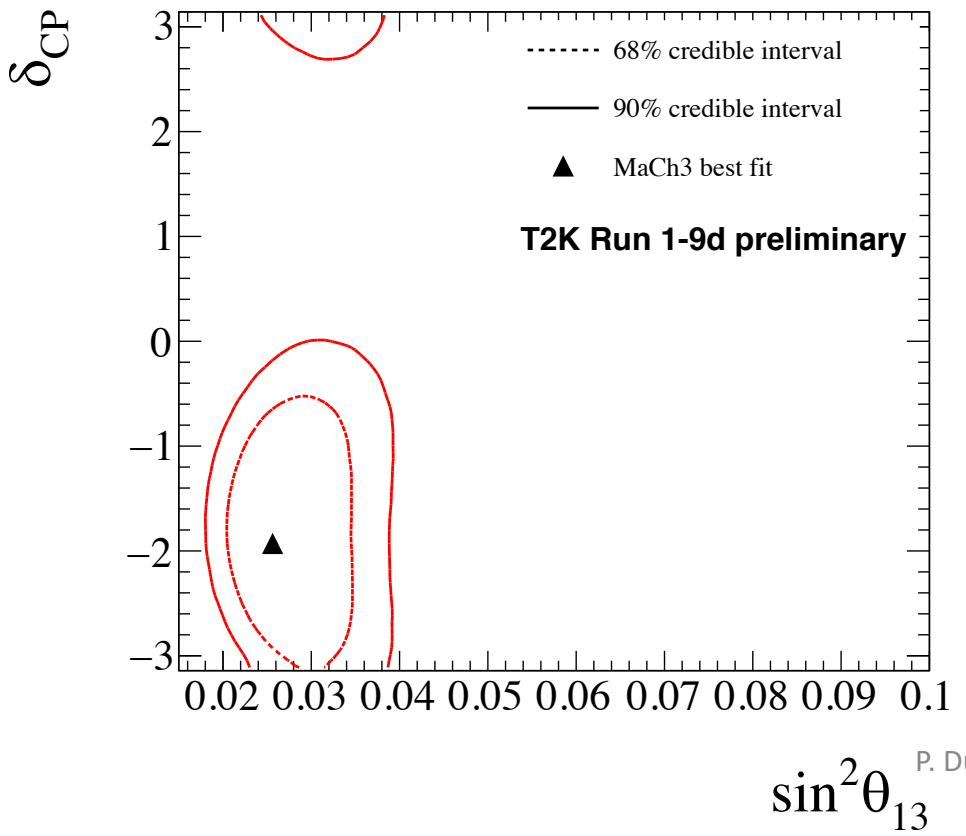
MCMC vs grid search

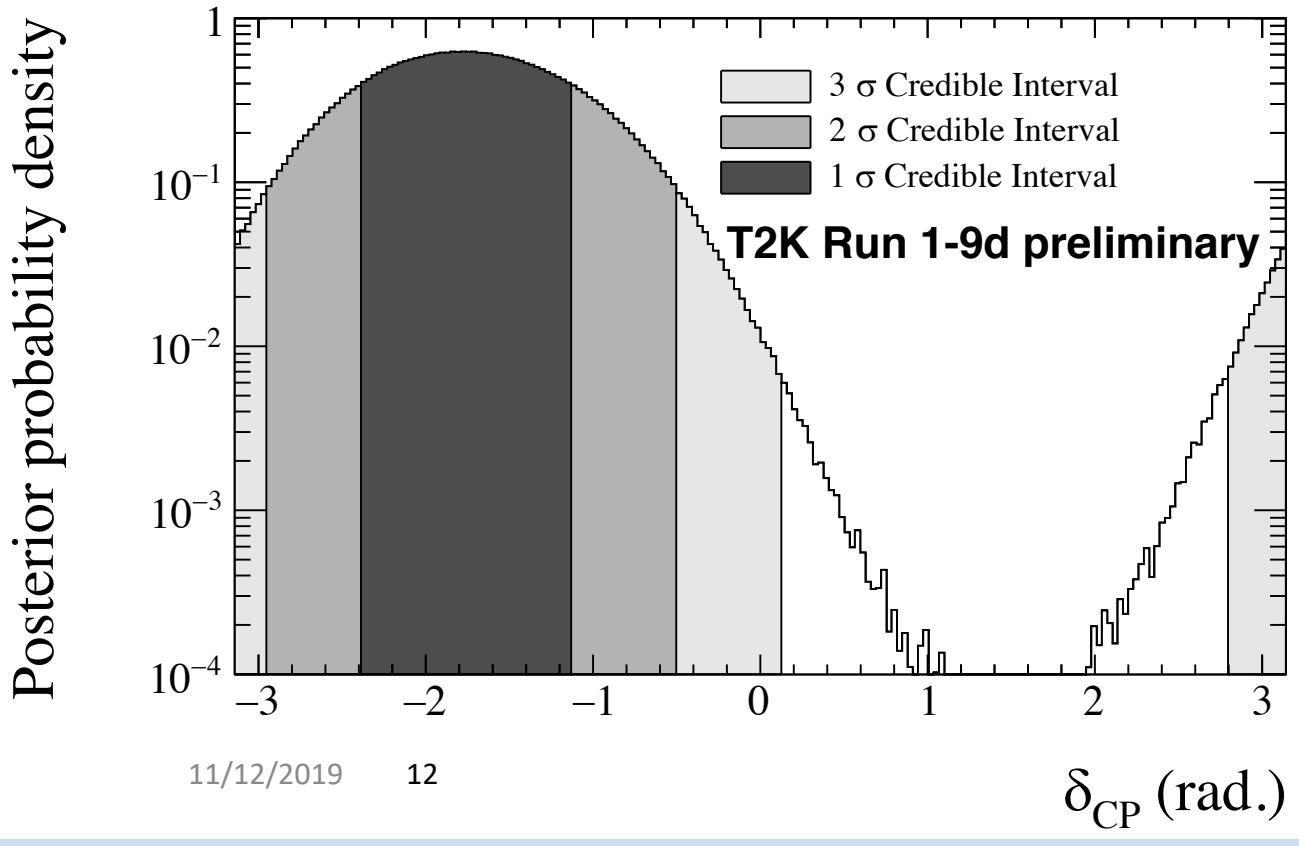
- Output of MCMC is a large number of ‘steps’
  - Each step is a vector of the values of all parameter for the step
- Creating 1D/2D histograms of any combinations of parameters gives the posterior probability for those parameters



Appearance parameter constraints

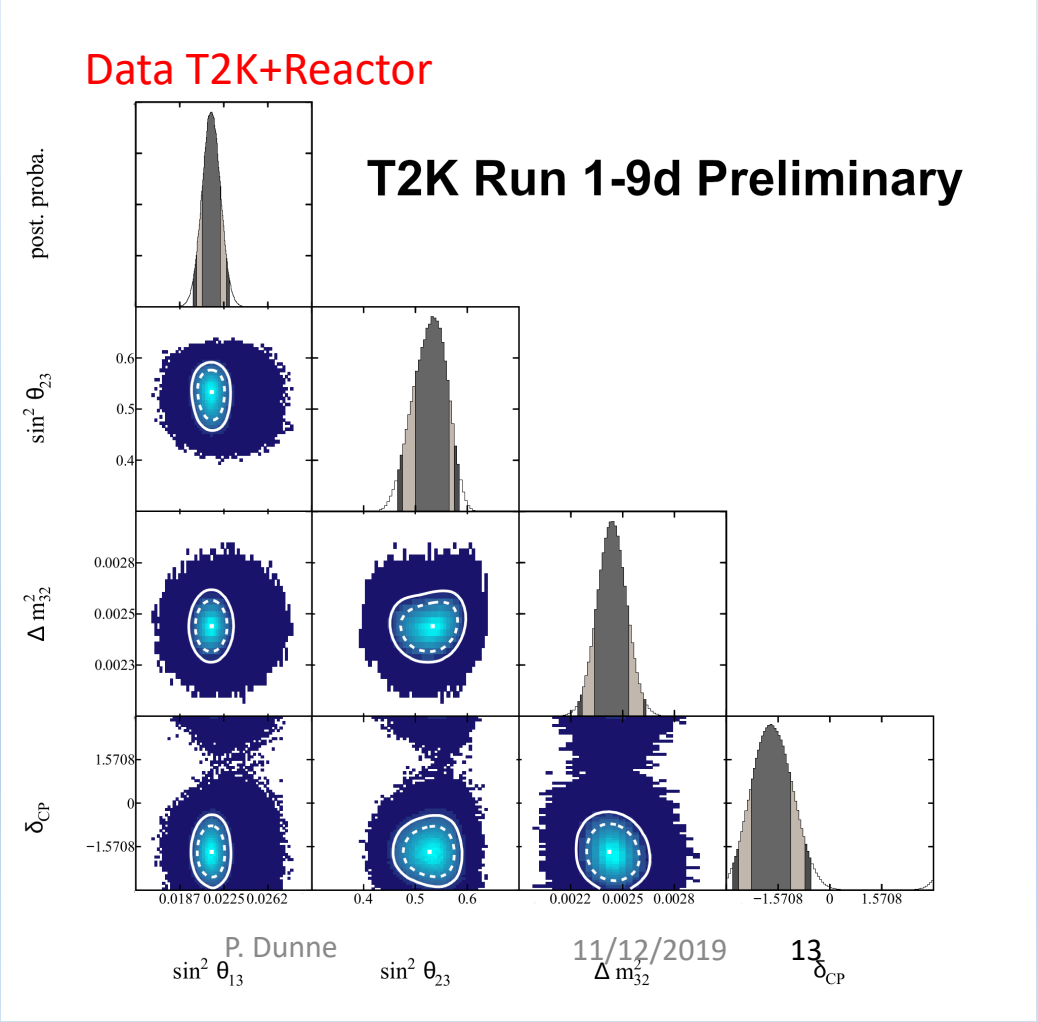
- Make contours by taking bins with most steps (i.e. highest prob) until you have 68/95/99.73…% of the probability inside your contour
- Don’t get a multidimensional best fit, but do get highest prob bin per par





Other variable combinations

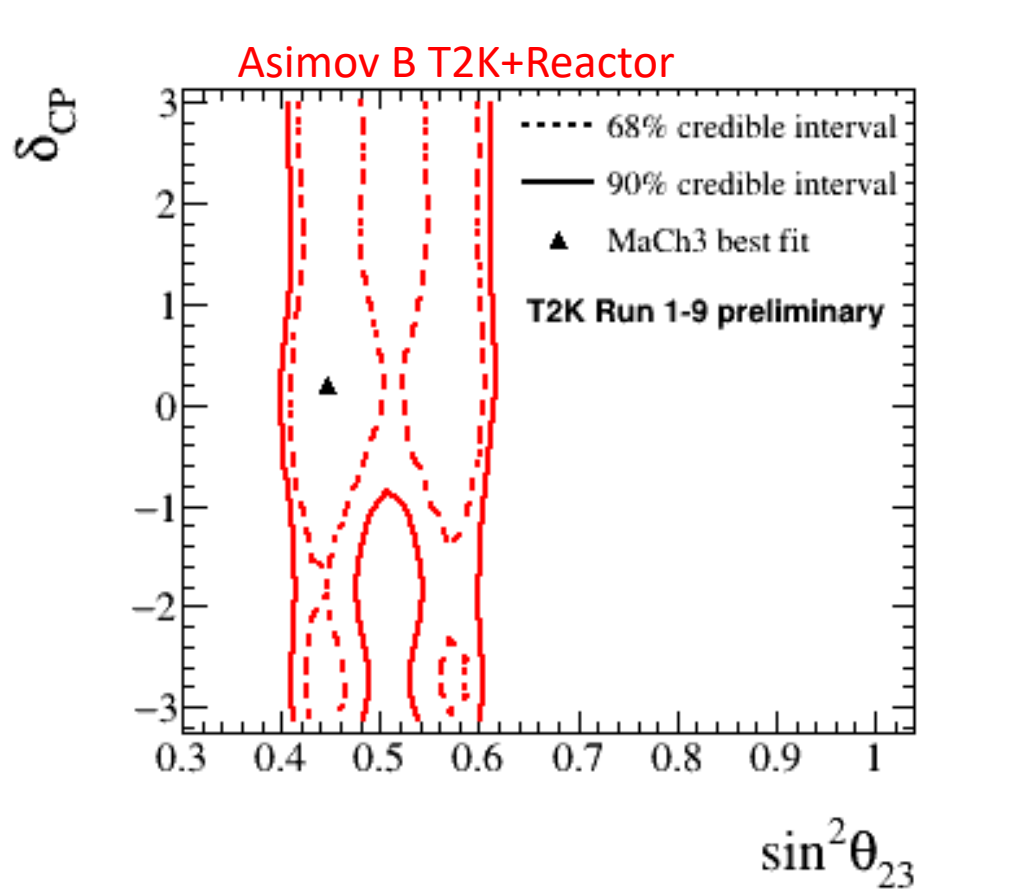
- Markov Chain samples all variables simultaneously
- Can compare other combinations with no extra computing

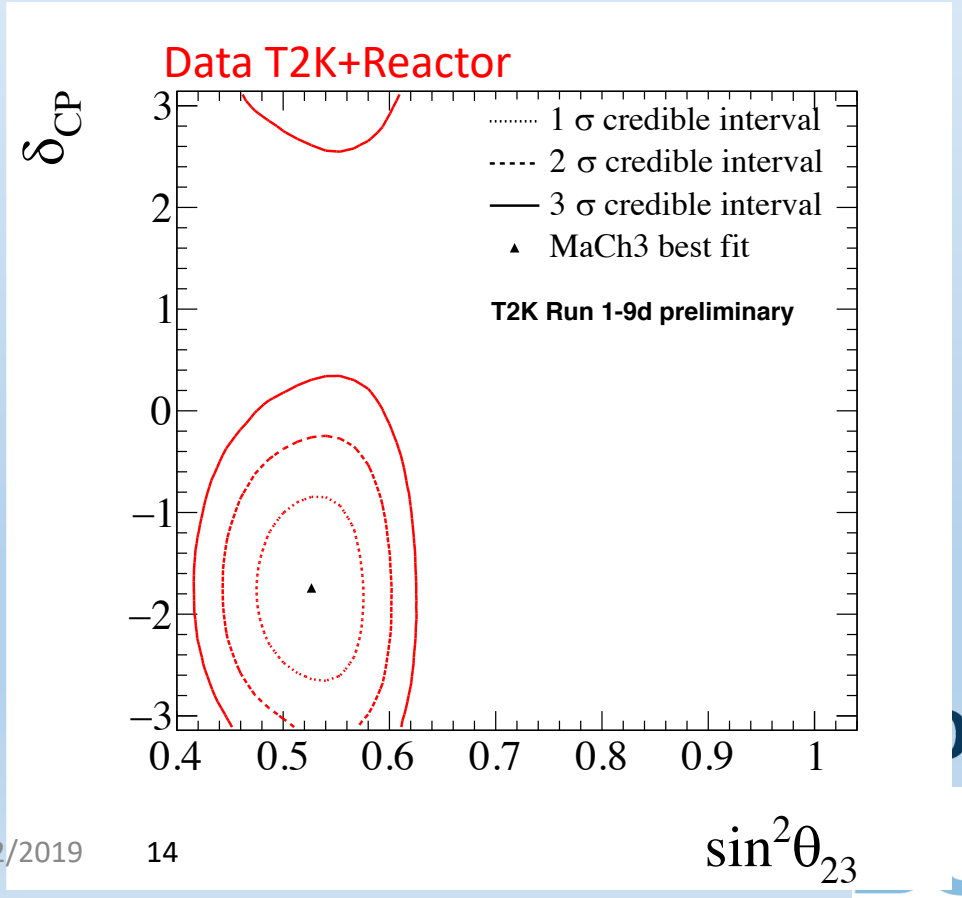


Imperial College London

Other variable combinations

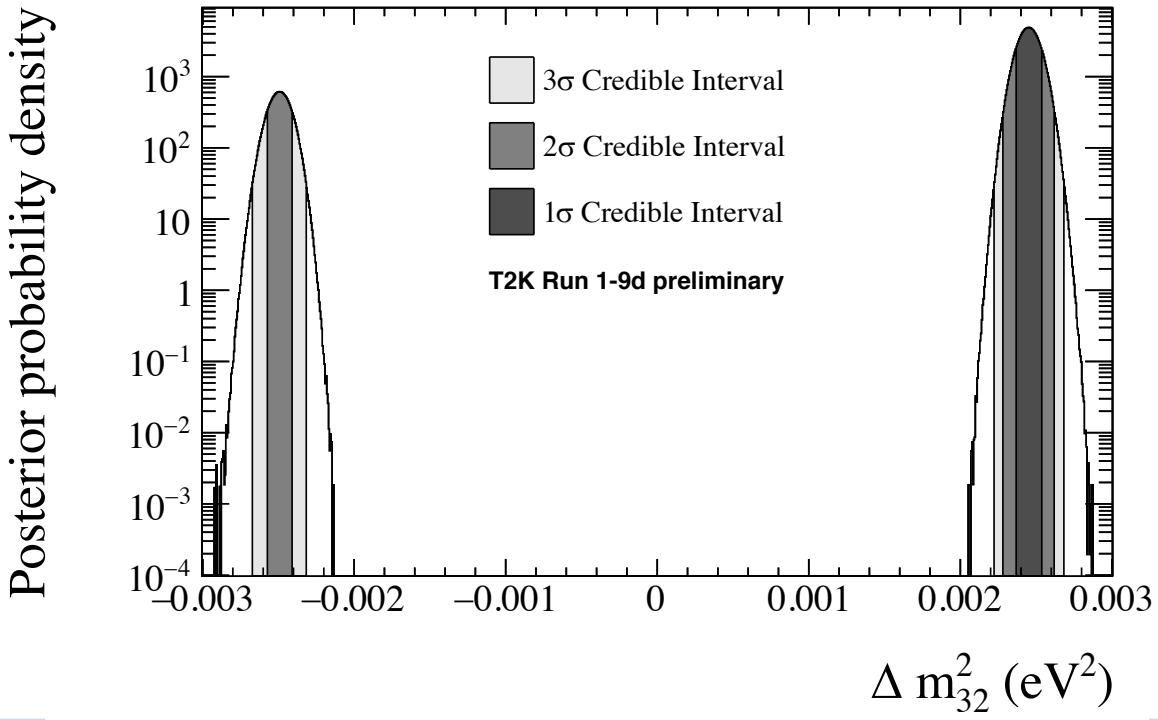
- Markov Chain samples all variables simultaneously
- Can compare other combinations with no extra computing





Mass Hierarchy results

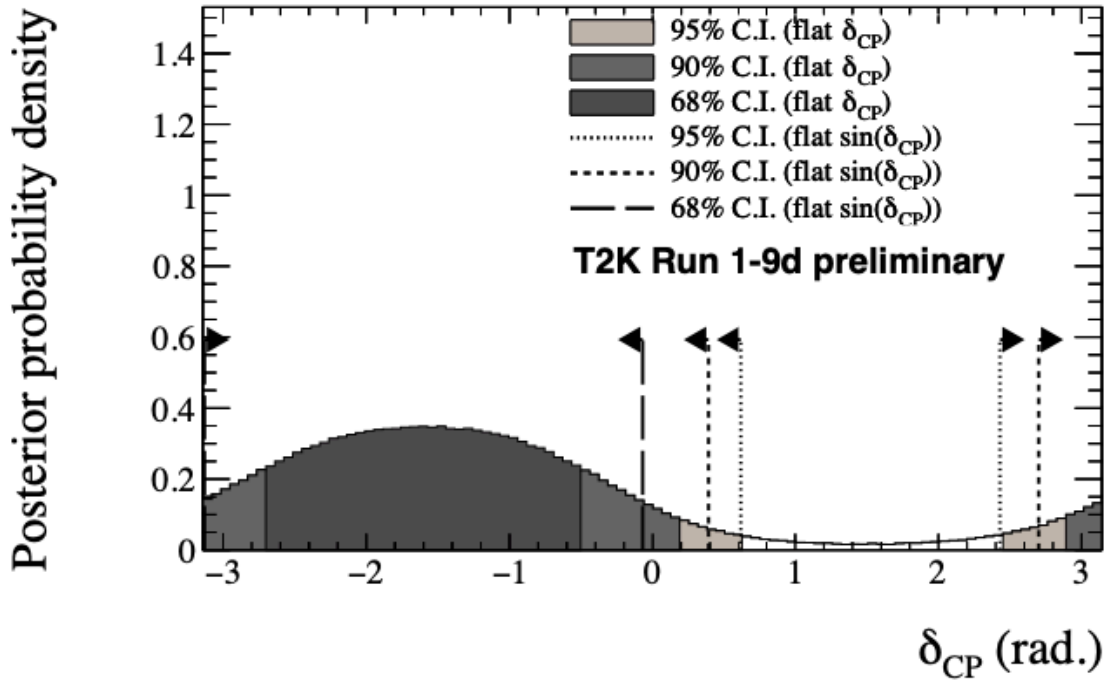
- On T2K for each step we assign a 50% probability that the proposal will be in the other hierarchy
  - With enough data this 50% is overcome by data preference
  - Preference for each hierarchy is given by fraction of steps that lie in each
- Both hierarchies run in a single fit
- Also don't have to choose octant

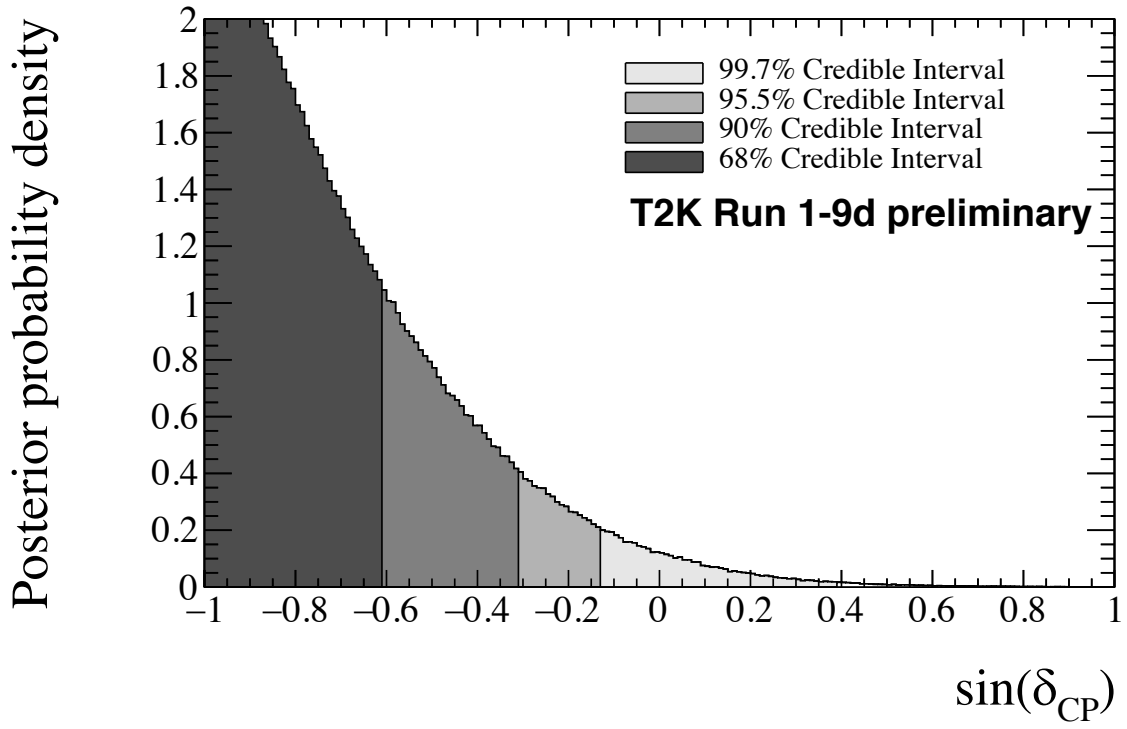


	$\sin^2\theta_{23} < 0.5$	$\sin^2\theta_{23} > 0.5$	Sum
NH ($\Delta m^2_{32} > 0$)	0.184	0.705	0.889
IH ($\Delta m^2_{32} < 0$)	0.021	0.090	0.111
Sum	0.205	0.795	1

Priors

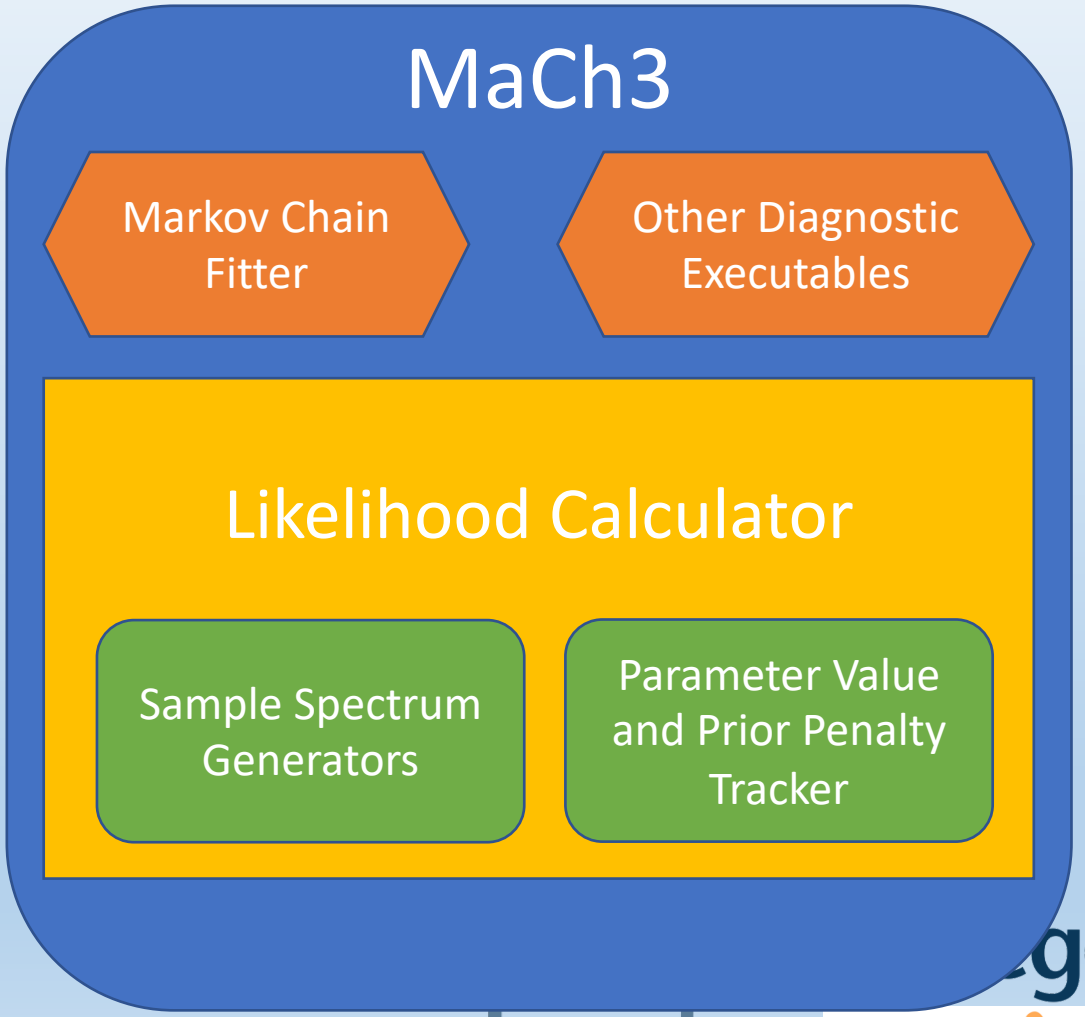
- Bayesian analysis requires a choice of prior (quite a few frequentist ones do too)
- As long as prior doesn't strongly favour region with no steps you can reweight existing chain to change prior





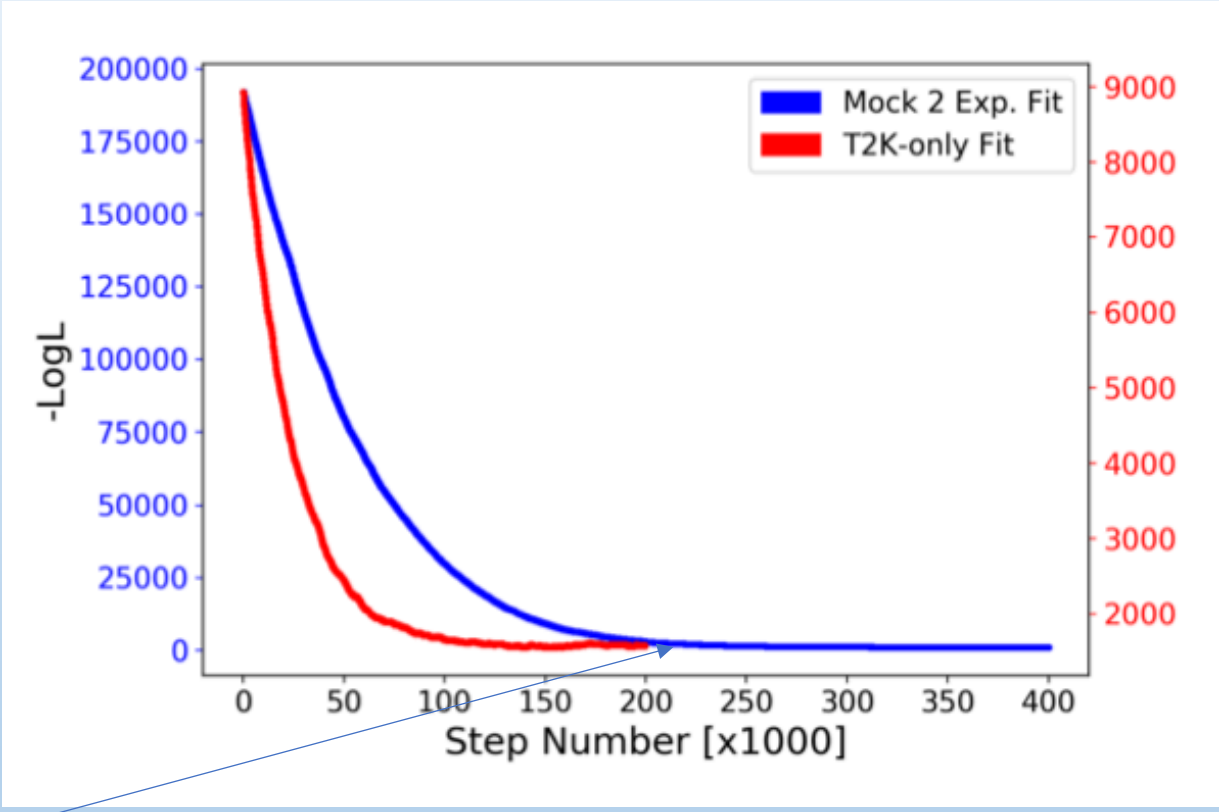
Code Structure

- Code is modular, can add any number of samples/parameters to the fit (defined in executable)
- Sample spectrum generator:
  - Code that gives expected distribution of a sample as a function of parameters
- Parameter tracker:
  - Calculates systematic penalty terms and tells spectrum generators what parameter values are
- We'd need to make DUNE sample spectrum generators ~few thousand lines of code per detector



Computing time needed

- Determined by three things:
  - Time to perform a likelihood evaluation
  - Autocorrelation between parameters
  - Desired number of steps in excluded region
- Also chains take some time to start properly sampling from the likelihood if started at a random point called “burn in”



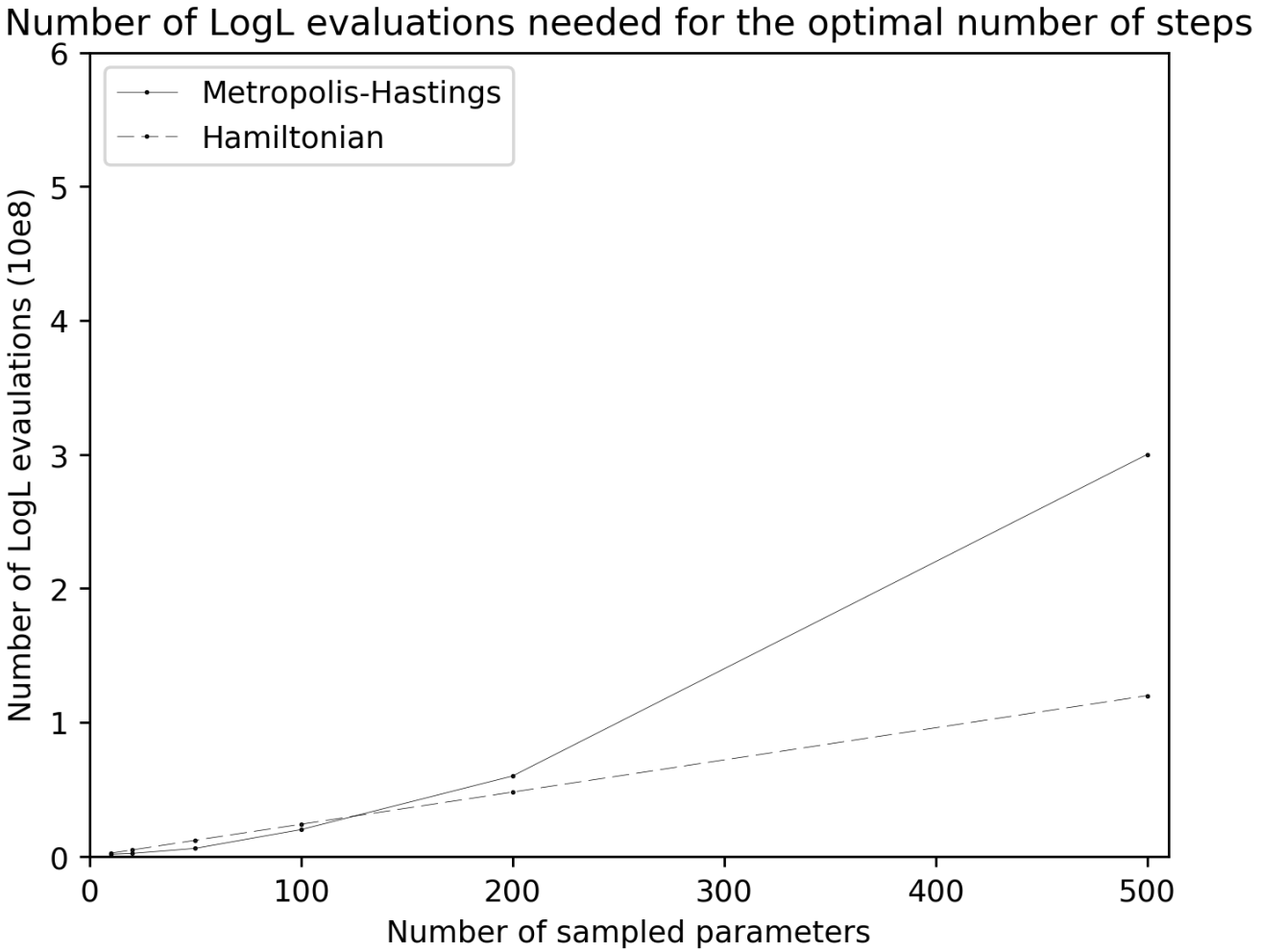
Computing time

- Time per step for T2K is ~0.05s for ND and 0.5 for SK
  - We have ~750 parameters, 19 samples
  - Depends heavily on how LLH evaluation is implemented
- Autocorrelation is the number of steps before you have an uncorrelated sample from the likelihood
  - ~10,000 steps for T2K
  - Depends on number of parameters and tuning of step proposal function
- Number of uncorrelated steps in excluded region
  - If you want to do a result at X% significance need statistical errors on number of steps outside interval to be small enough
- **Total time for all MaCh3 fits for Nature 3 sigma results 30,000 CPU hours**

Imperial College London

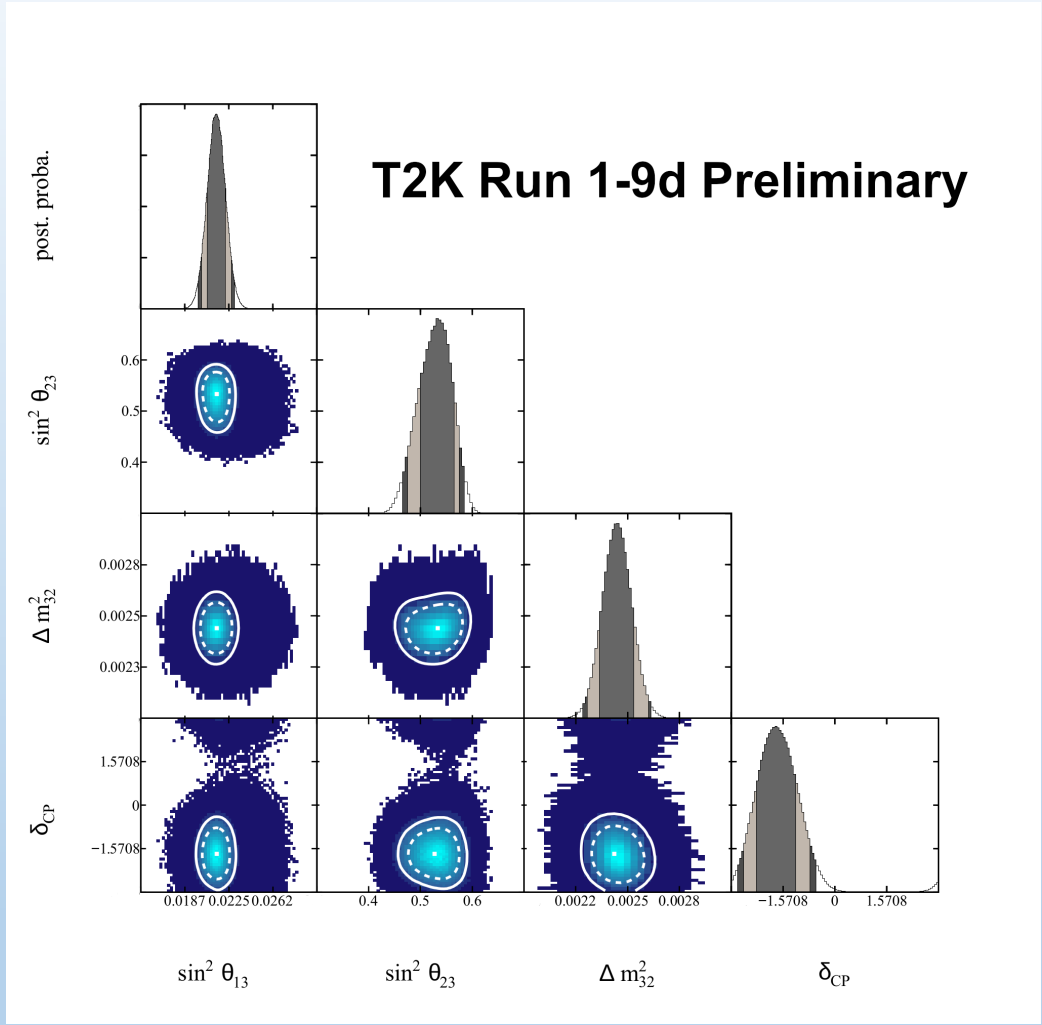
Scaling with parameters

- Time to fit does increase with number of parameters
- Hamiltonian is approximately linear so not too bad and Metropolis-Hastings not much worse
- Both increase less quickly than profile likelihood



Summary

- Markov Chains provide significant efficiency improvements by targeting where throws are carried out (usually used for Bayesian results)
- MaCh3 is a flexible analysis framework for Bayesian oscillation analyses
  - All current MaCh3 institutes are on DUNE
  - Three most experienced developers are all DUNE collaborators and several existing group members have expressed interest in DUNE-MaCh3
  - T2K has used it for inclusion of a high dimensional, many sample ND likelihood

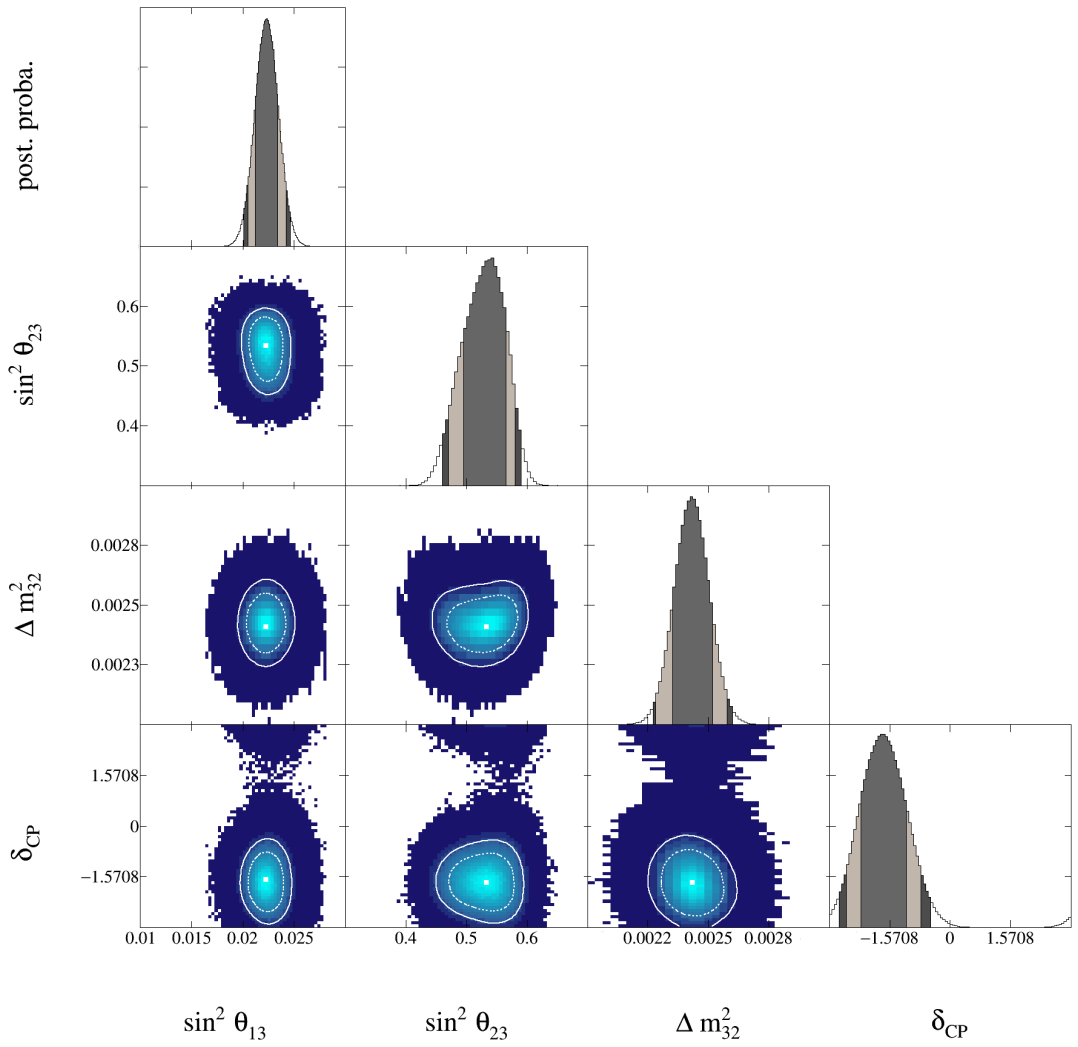


Imperial College London

Backup

Imperial College London

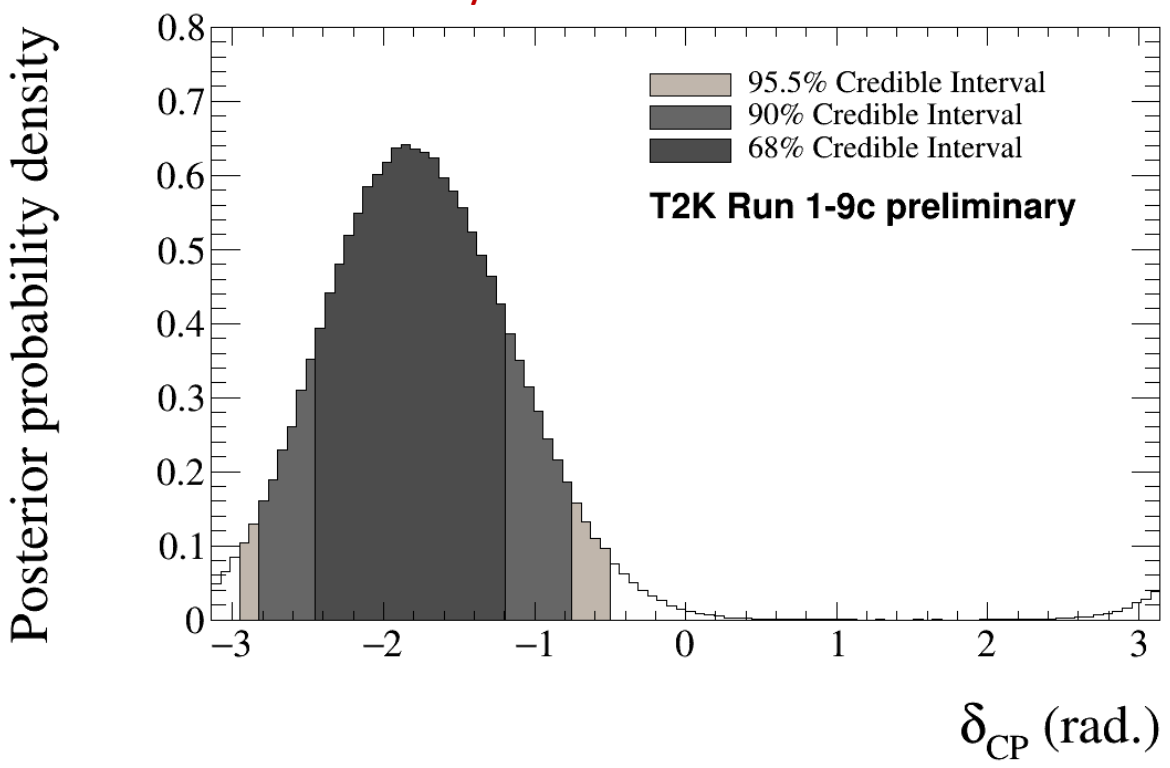
T2K +reactor constraint IH



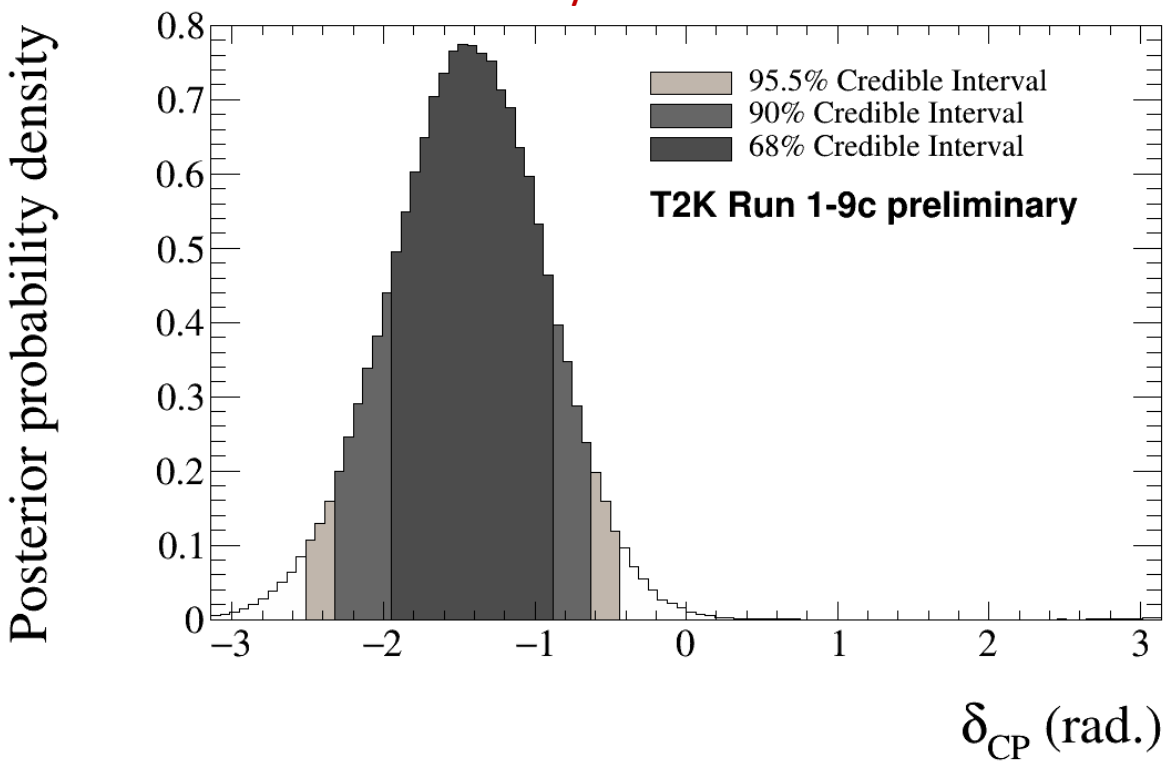
Imperial College London

Dcp split by hierarchy- T2K+reactor

Normal hierarchy

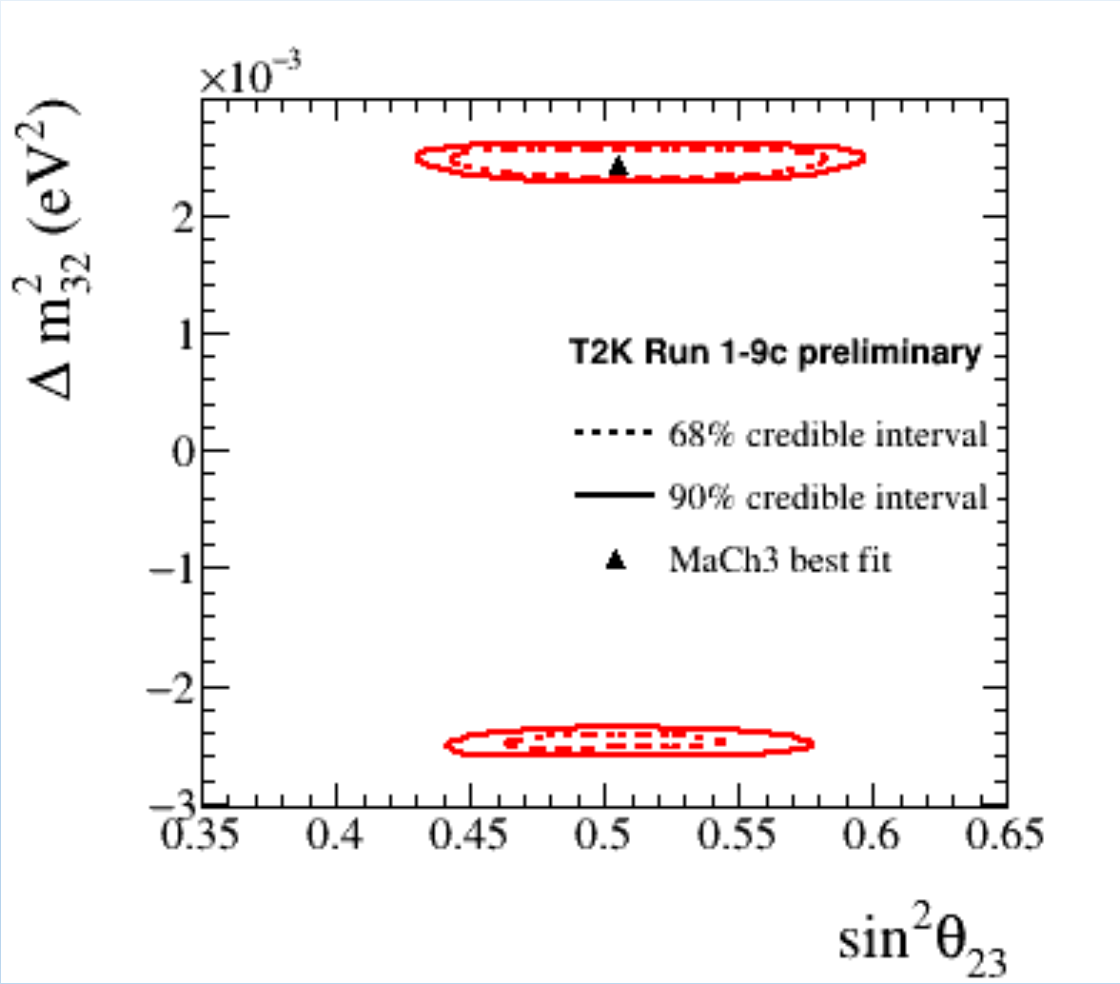


Inverted hierarchy

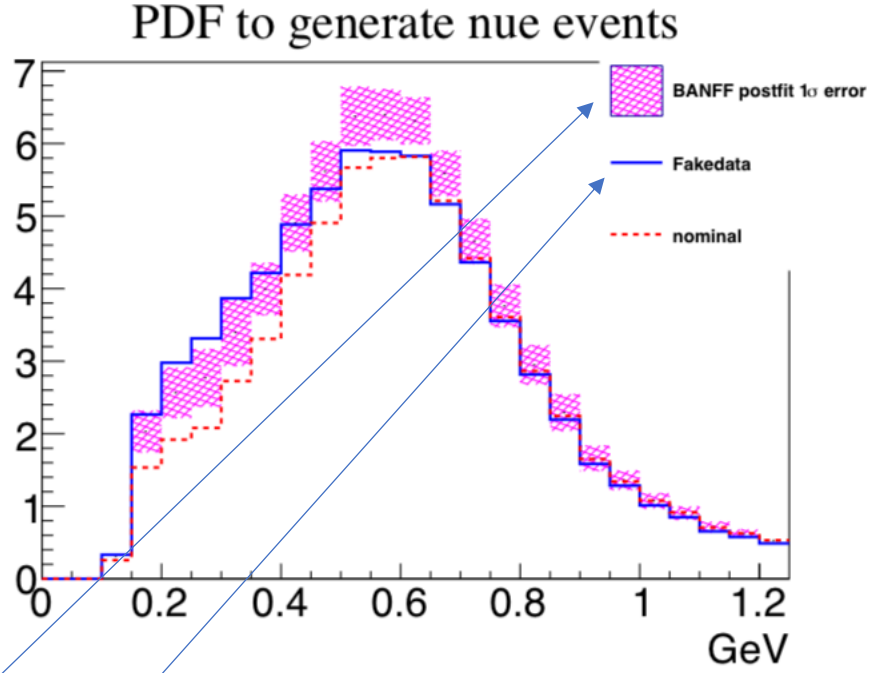
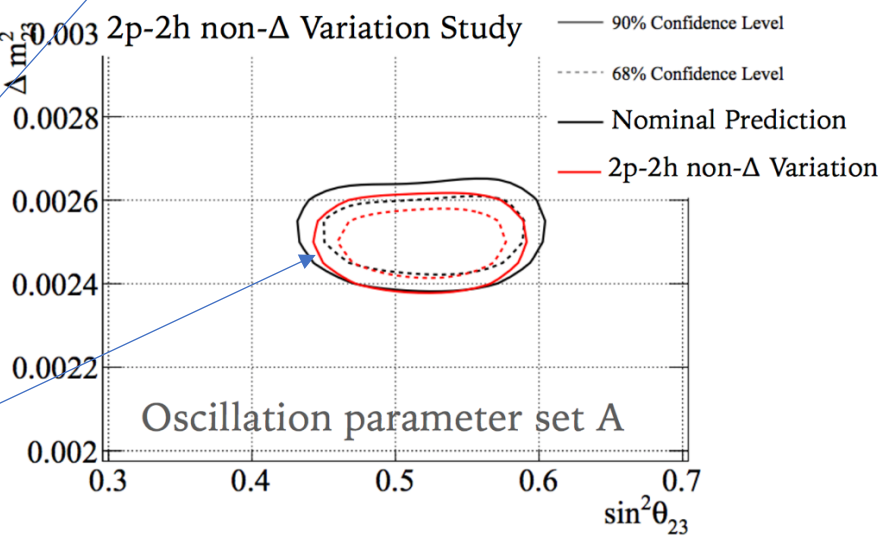


Imperial College London

T2K data only disappearance parameters



Imperial College London

Simulated data method

- Check robustness of results to neutrino interaction model by using our model to fit ``simulated data”
- Simulated data are generated in two ways
  1. `Data-driven’: Inflate one interaction mode to account for differences between current model prediction and existing data
  2. Model choices: generate data using other models implemented in generator but not used in oscillation analysis and refit
- Fit simulated ND data, propagate constraint to SK
- Fit SK simulated data using ND constrained xsec model
- Compare fit to simulated data to nominal model Asimov
- If getting the interaction model wrong leads to significantly different constraints: further investigation