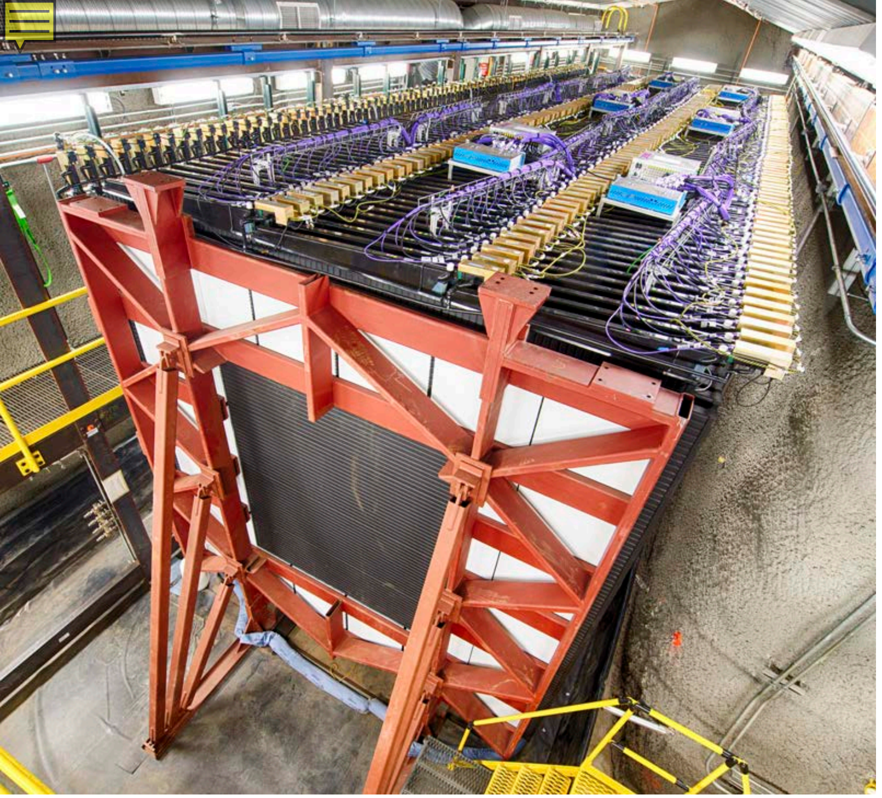
FERMILAB-SLIDES-18-111-ND

Accelerating Feldman-Cousins for NOvA using NERSC Supercomputers

Derek Doyle

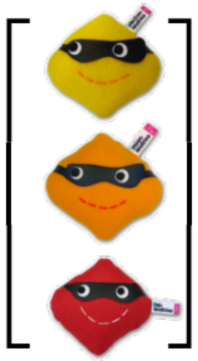
This document was prepared by [NOvA Collaboration] using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

Colorado State University

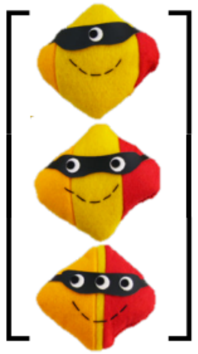


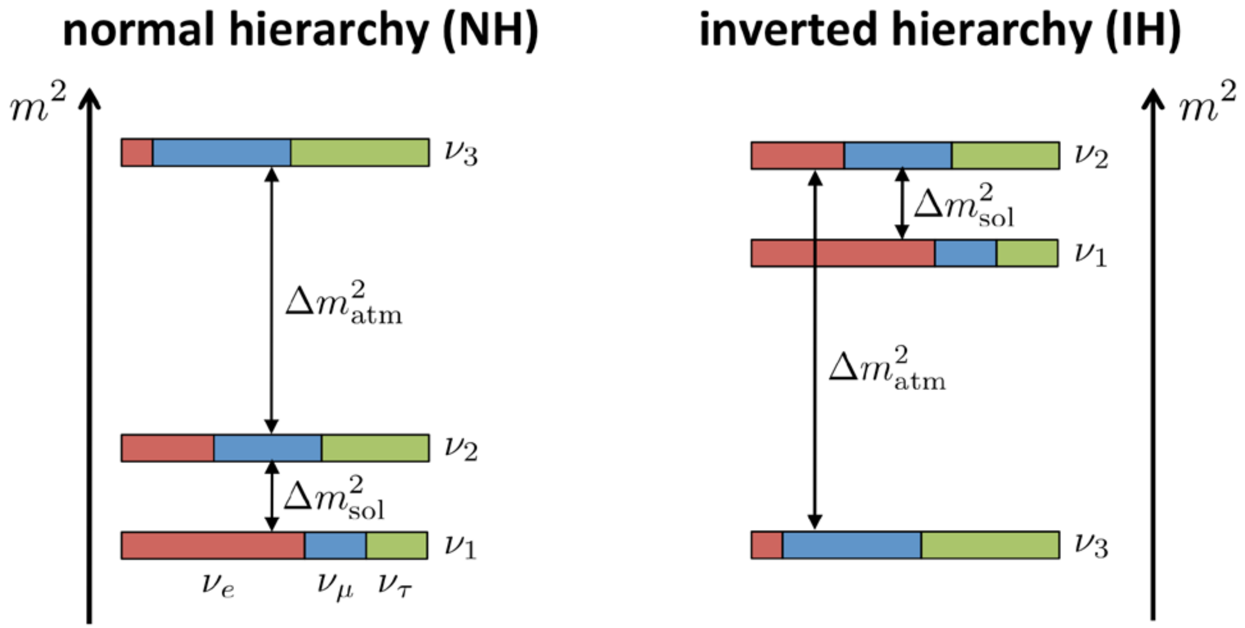
NOvA Oscillation Analysis

- Measure parameters governing neutrino oscillations through
  - $\nu_\mu \rightarrow \nu_\mu$ and $\nu_\mu \rightarrow \nu_e$
  - $\bar{\nu}_\mu \rightarrow \bar{\nu}_\mu$ and $\bar{\nu}_\mu \rightarrow \bar{\nu}_e$
- Look for CP Violation
- Determine Neutrino Mass Ordering

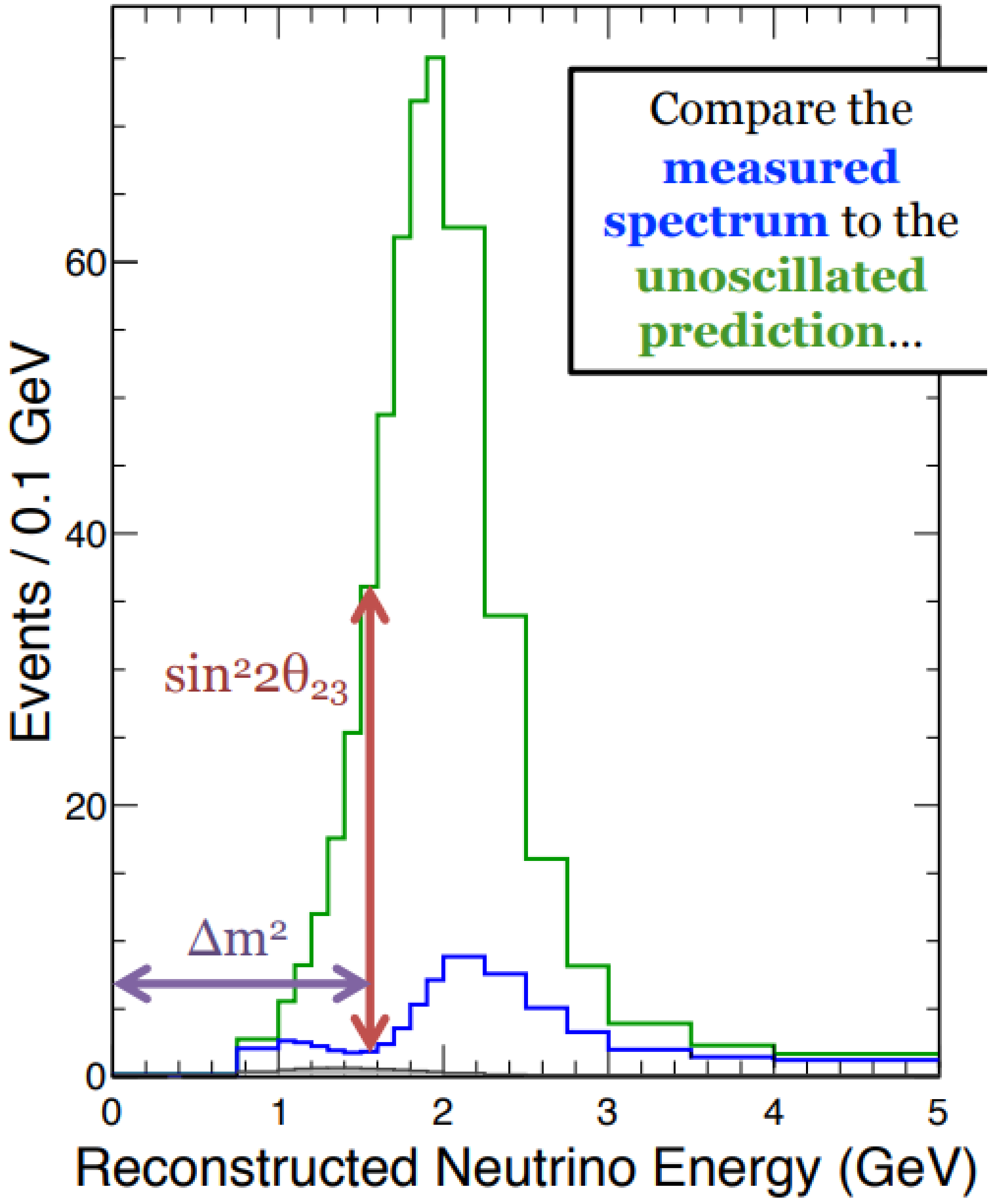


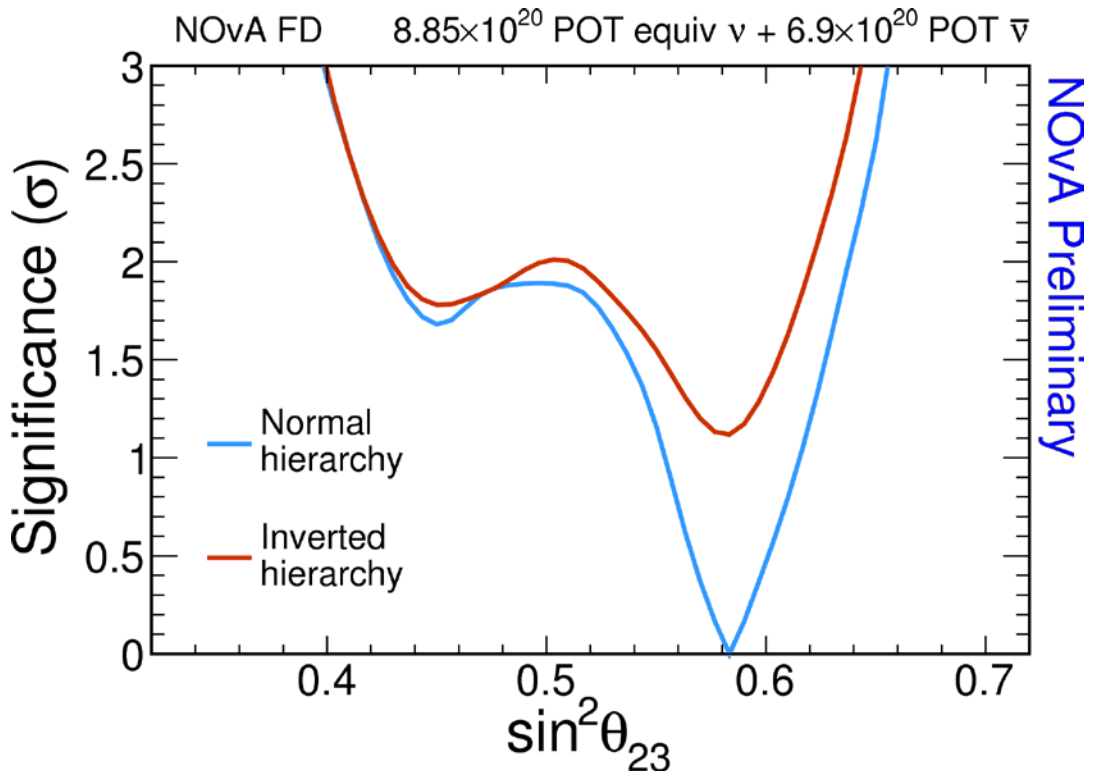
$$= R(\theta_{23}) \cdot R(\theta_{13}, \delta_{CP}) \cdot R(\theta_{12})$$





Measuring Oscillation Parameters





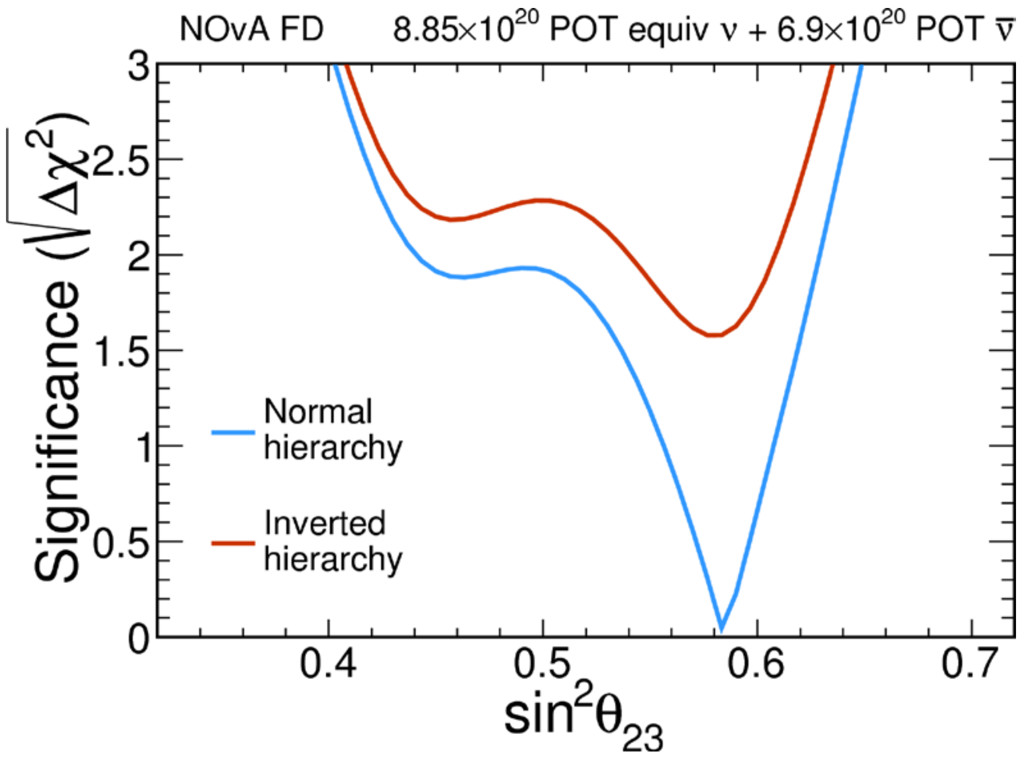
- Comparing the data to our predictions, estimate the true parameters, exclude other regions of parameter space with some degree of certainty, conventionally in terms of σ

Parameter Estimation

- Goal: Determine the likelihood that $\vec{\theta}$ are the true parameters
  - Find $\vec{\theta}$ that maximize the Likelihood function: $\mathcal{L}(\vec{\theta}_{best})$

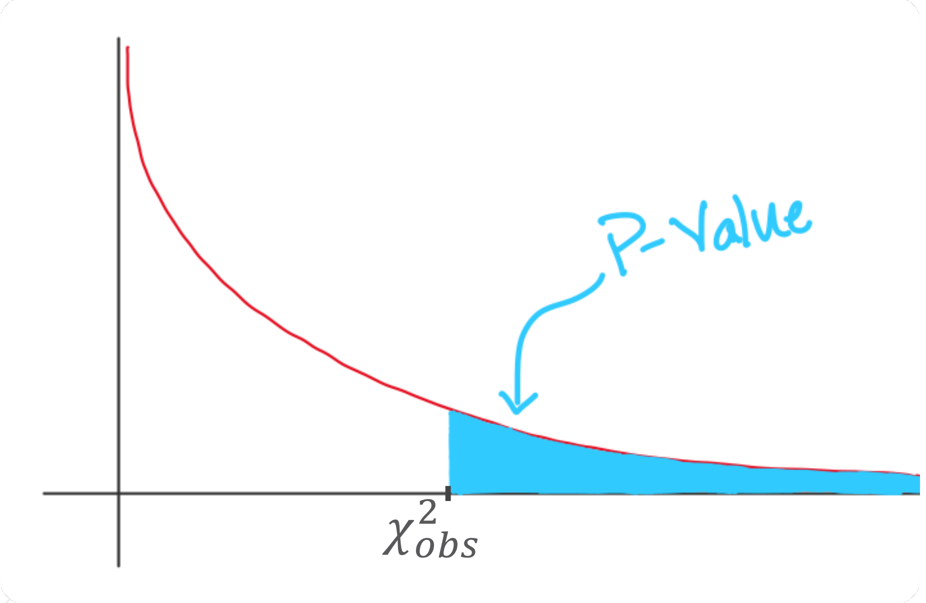
$$\chi^2(\vec{\theta}) \equiv -2\log\mathcal{L}(\vec{\theta}) = 2\sum_{i}^{bins}\left[e_i(\vec{\theta}) - o_i + o_i \log\frac{o_i}{e_i(\vec{\theta})}\right]$$

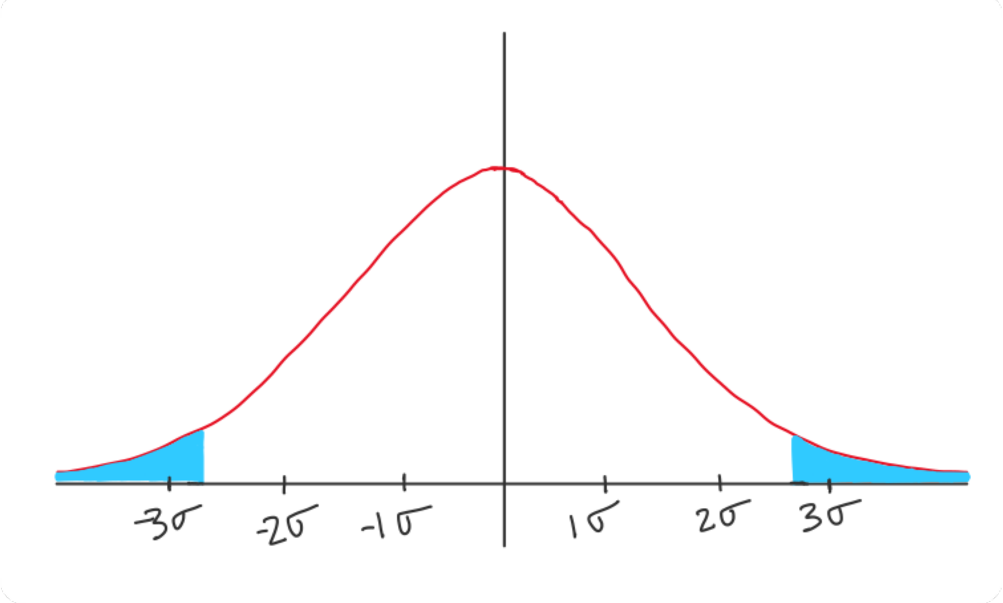
$\vec{\boldsymbol{\theta}}$: Set of Oscillation parameters
$\boldsymbol{o_i}$: Observed data
$\boldsymbol{e_i}(\vec{\boldsymbol{\theta}})$: Expected observation



- To measure the significance with which we exclude other parameters, $\vec{\theta}$, as the true parameters, we calculate $\chi^2(\vec{\theta})$ and construct the test statistic, $\Delta\chi^2 = \chi^2(\vec{\theta}) - \chi^2(\vec{\theta}_{best})$
- Conveniently, this distribution converges to a χ^2 distribution under certain assumption
- Moreover, under these assumptions, $\sigma = \sqrt{\Delta\chi^2}$

χ^2 Distribution



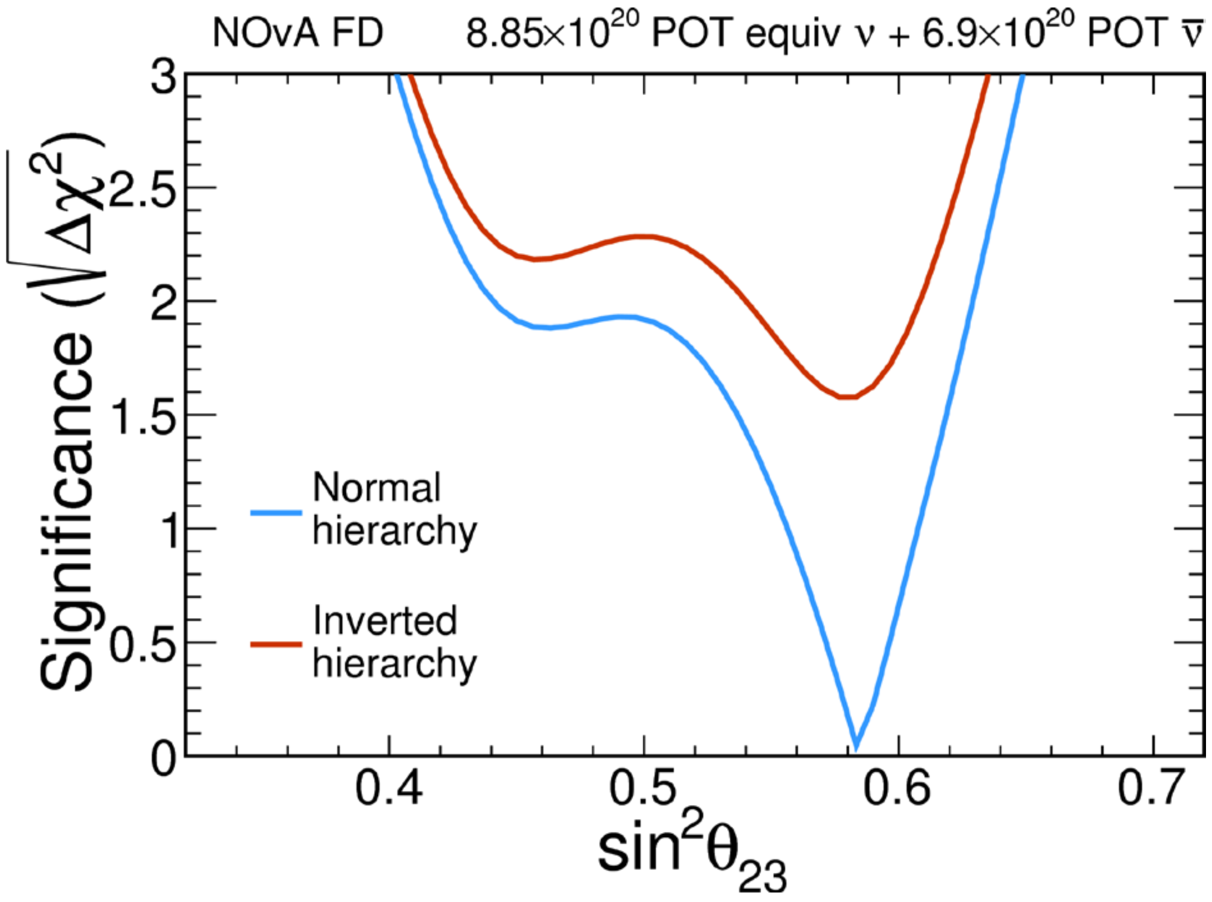


- Often used for determining goodness of fit
- Repeating experiment may result in different observations
- Want to know the probability of finding a χ^2
- P-Value: probability of finding a χ^2 at least as large as than the one you measure
- Tells you how probable it was to measure a χ^2 by accident

$$p\text{-}value = \int_{\chi^2_{obs}}^{\infty} f(\chi) d\chi$$

- Critical Value: value of test statistic corresponding to a given p-value
- Report a significance in terms of the critical value of a standard normal distribution corresponding to that p-value

Case Study: 1D Profile



- At each point in parameter space, calculate a $\Delta\chi^2(\vec{\theta})$

$$\Delta\chi^2 = \chi^2(\vec{\theta}) - \chi^2(\vec{\theta}_{best})$$

- We end up with the trajectory that minimizes $\Delta\chi^2$ along one axis
  - Profile over other parameters
- Can we use a standard χ^2?

Wilks' theorem

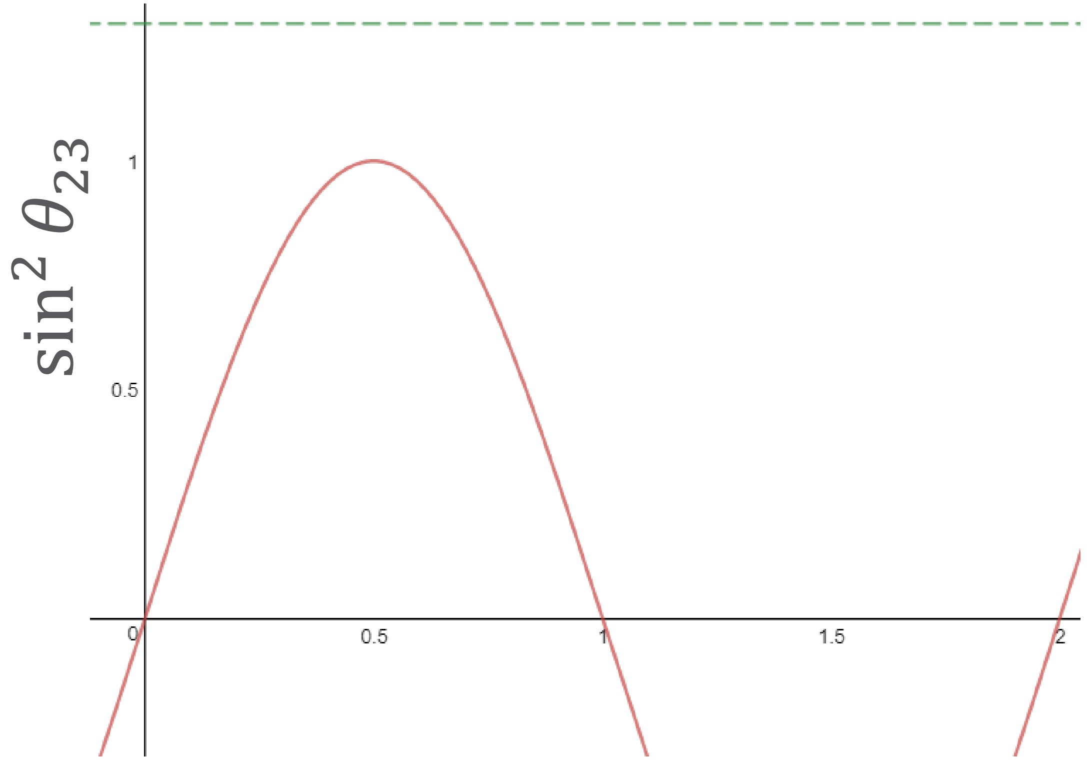
- A $\Delta\chi^2$ distribution converges to a standard χ^2 if two conditions are met:
    1. Large Statistics
    2. Parameters are sufficiently far from physical boundaries

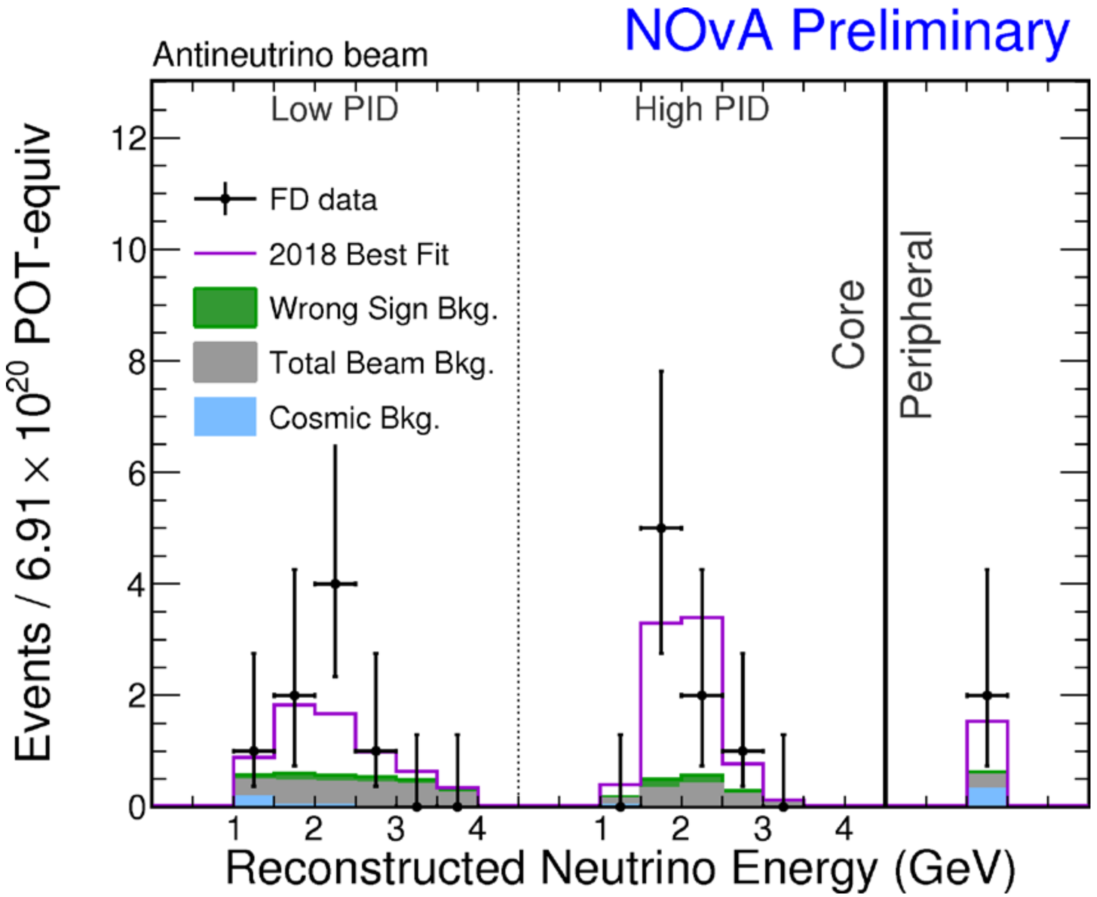
- A $\Delta\chi^2$ distribution converges to a standard χ^2 if two conditions are met:
    1. Large Statistics
    2. Parameters are sufficiently far from physical boundaries

Wilks' theorem

- A $\Delta\chi^2$ distribution converges to a standard χ^2 if two conditions are met:
  1. Large Statistics
  2. Parameters are sufficiently far from physical boundaries

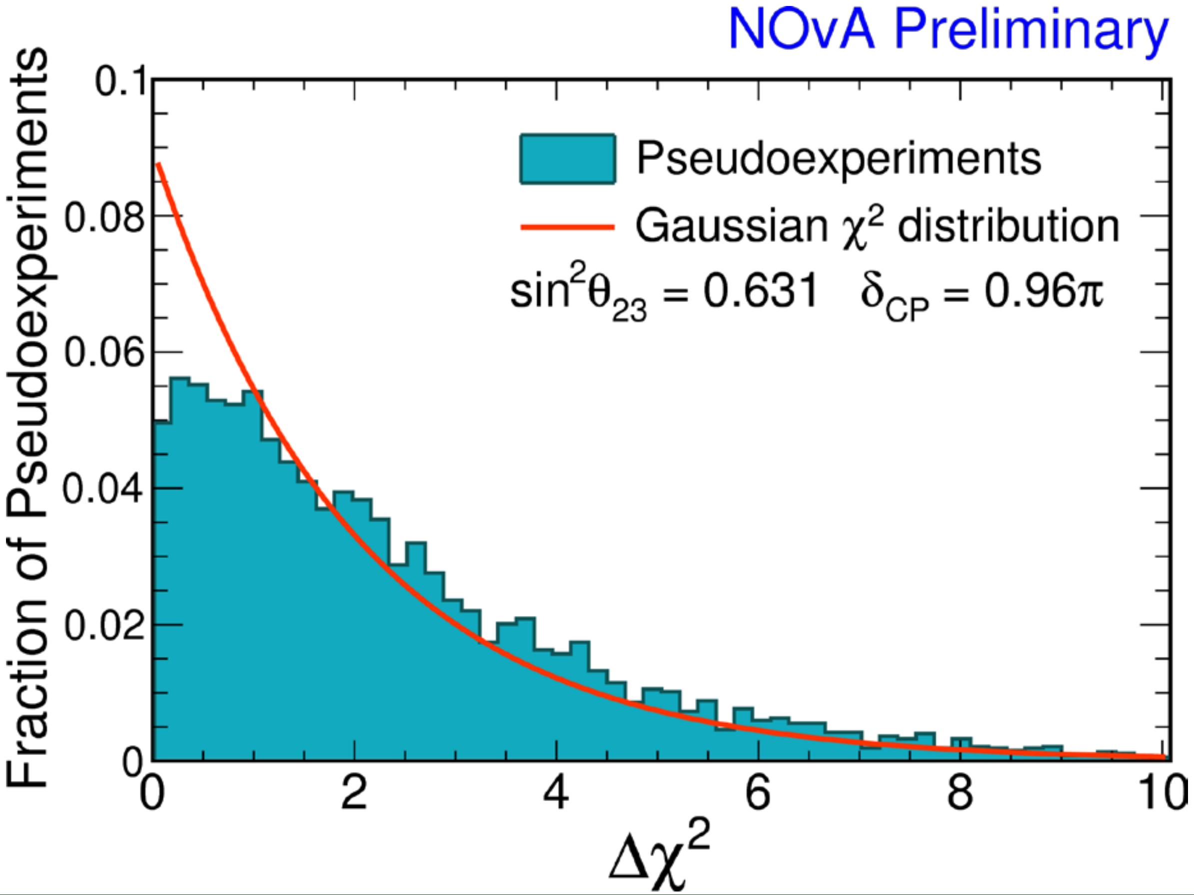
Total $\bar{\nu}_e$ Observed	18
Total Prediction	15.9
Wrong-sign	1.1
Beam Bkgd.	3.5
Cosmic Bkgd.	0.7
Total Bkgd.	5.3





Feldman-Cousins Procedure

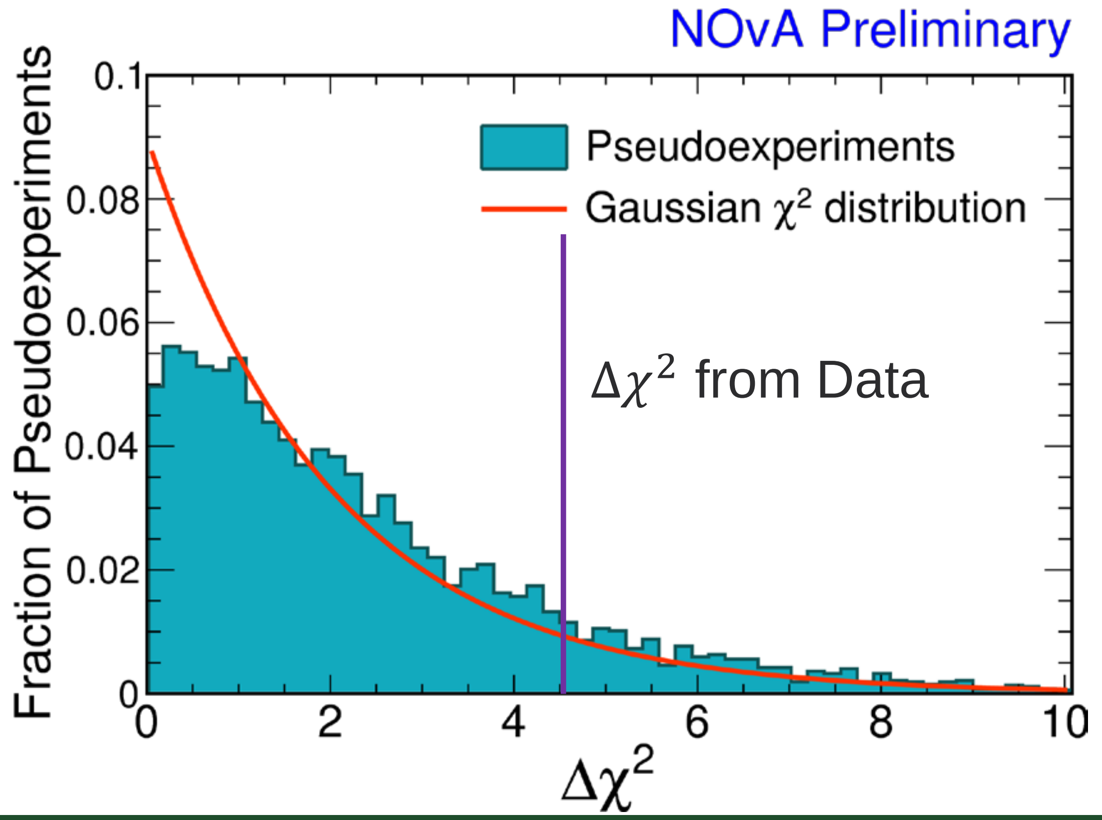
- Wilks' result is a convenience if his conditions are met
- If we could find a true $\Delta\chi^2$ distribution, we could calculate a p-value from it just like any other distribution function and quote a true significance
- So let's do that… empirically
- Recall a p-value is like proportion of your test statistic that is distributed beyond a certain value
- So if we can find a similar proportion of an empirical distribution, we can quote a significance



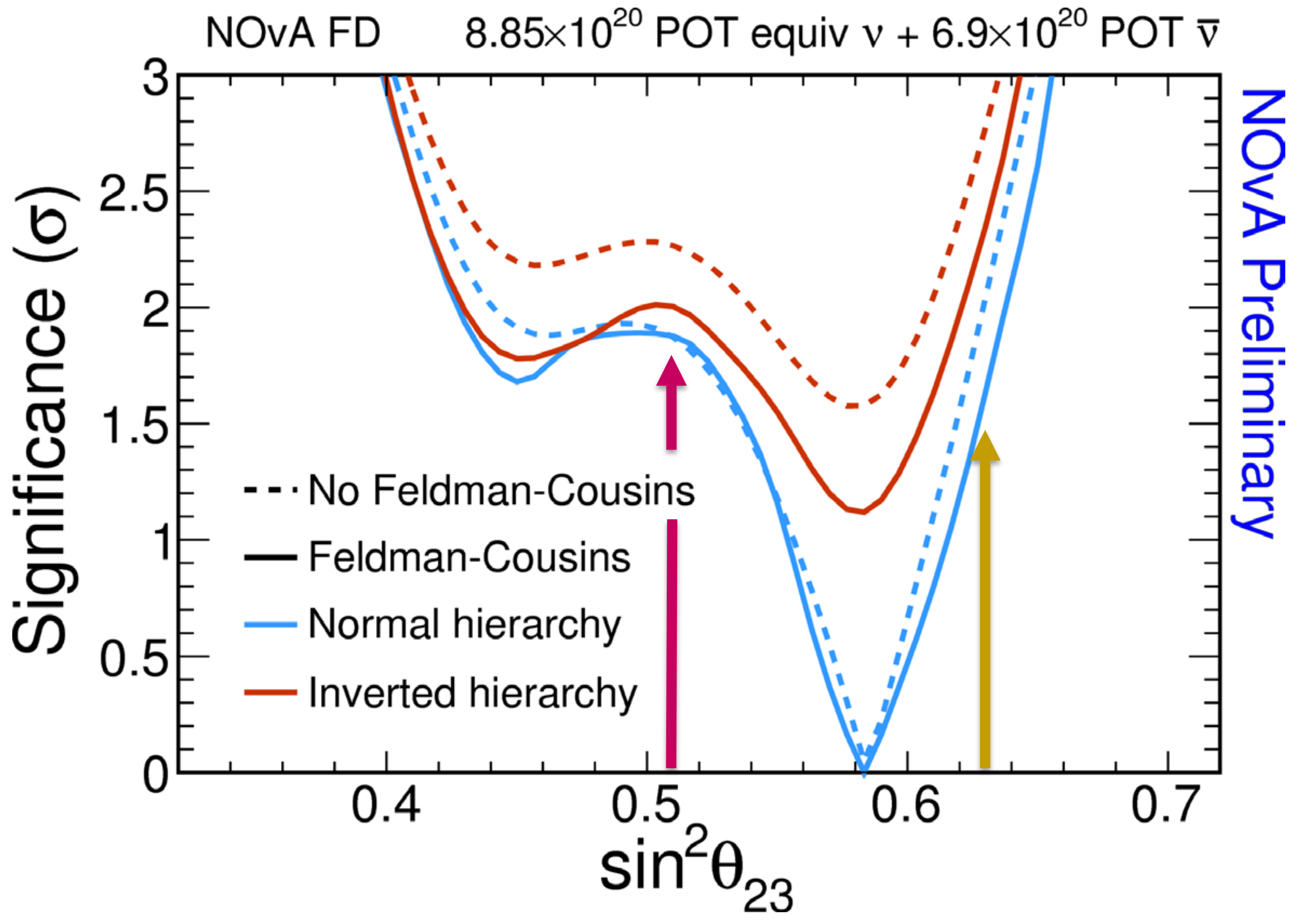
Feldman-Cousins Procedure

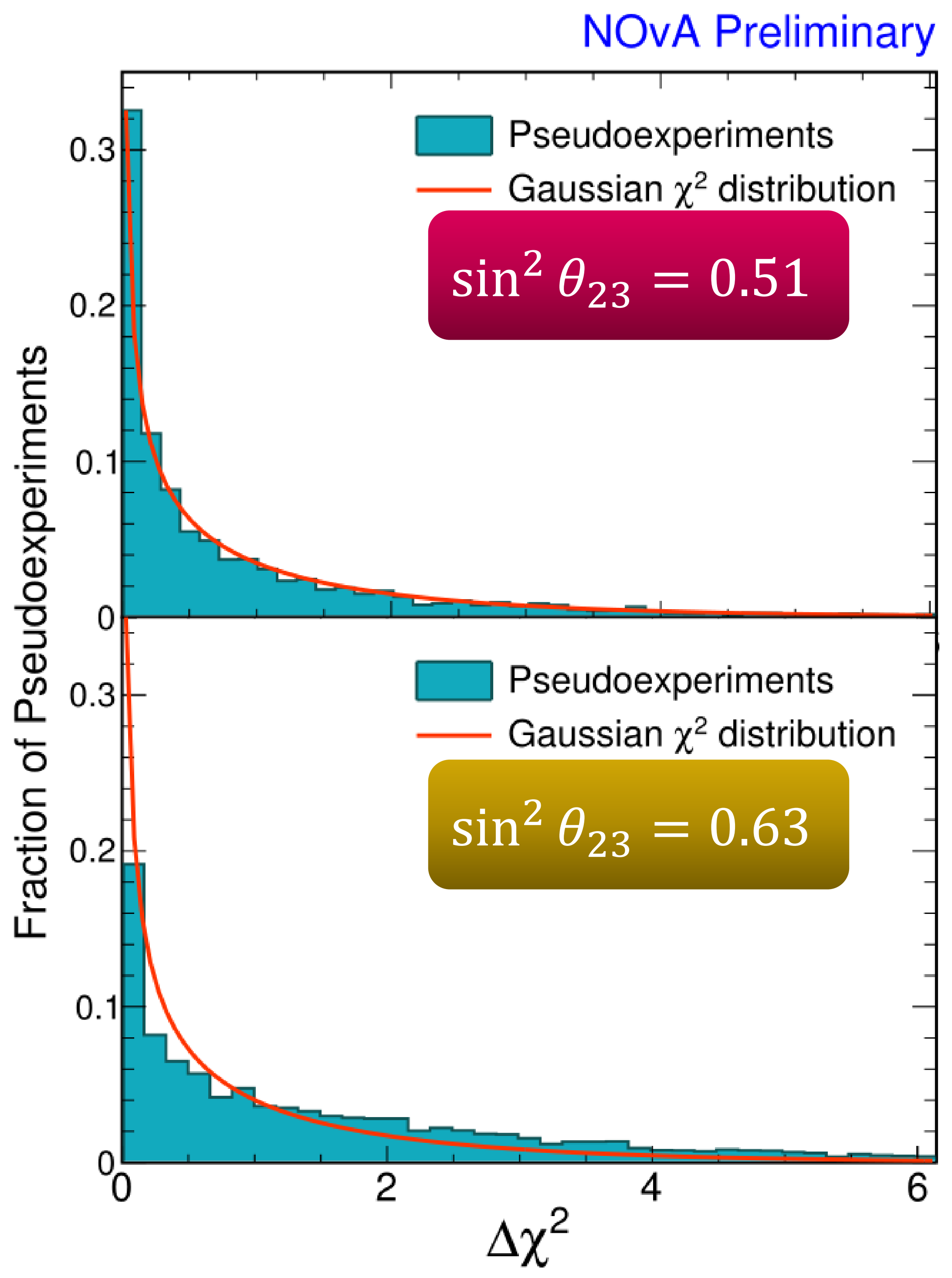
- At a given point in parameter space we want to know, “does our prediction at a set of test parameters, $\vec{\theta}_{test}$, agree with our data?”
  - We need a $\Delta\chi^2$ distribution where this is true in order to reject our hypothesis
  - i.e, In a world where the true parameters are $\vec{\theta}_{test}$, given statistical fluctuations, how would the $\Delta\chi^2$ test statistic be distributed?

1. Generate data at θ_{test} that is Poisson-fluctuated (pseudoexperiment) in order to get value of $\Delta\chi^2$
   a. Do this *many* times to build a distribution
2. Find the proportion of pseudoexperiments above observed $\Delta\chi^2$
3. Quote a significance based off that p-value



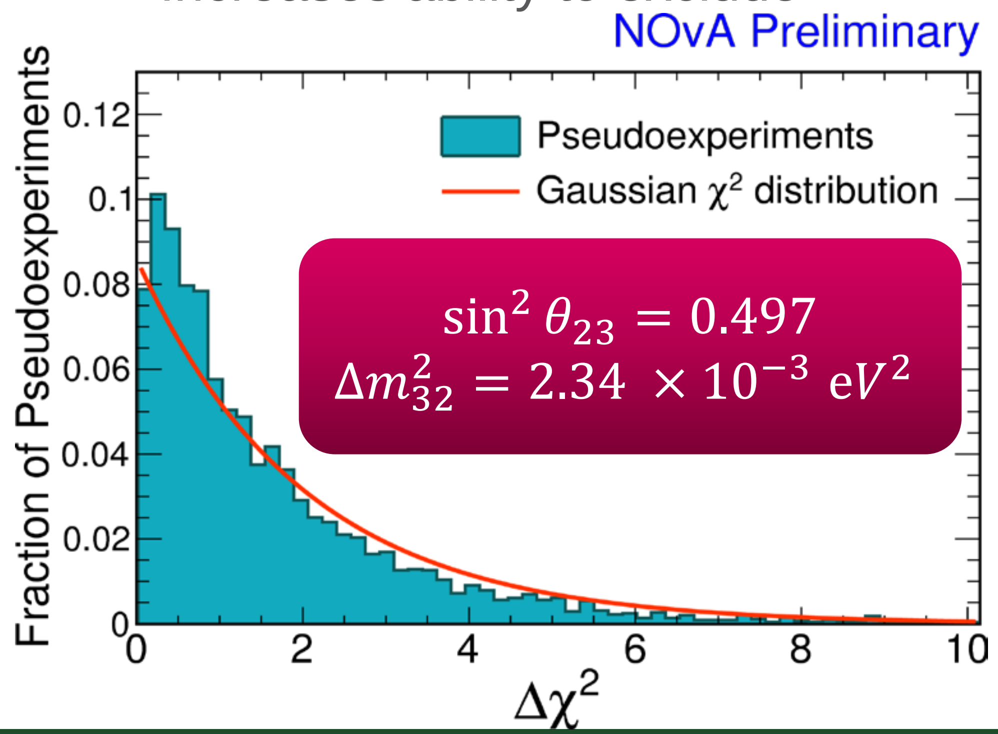
Correcting a profile



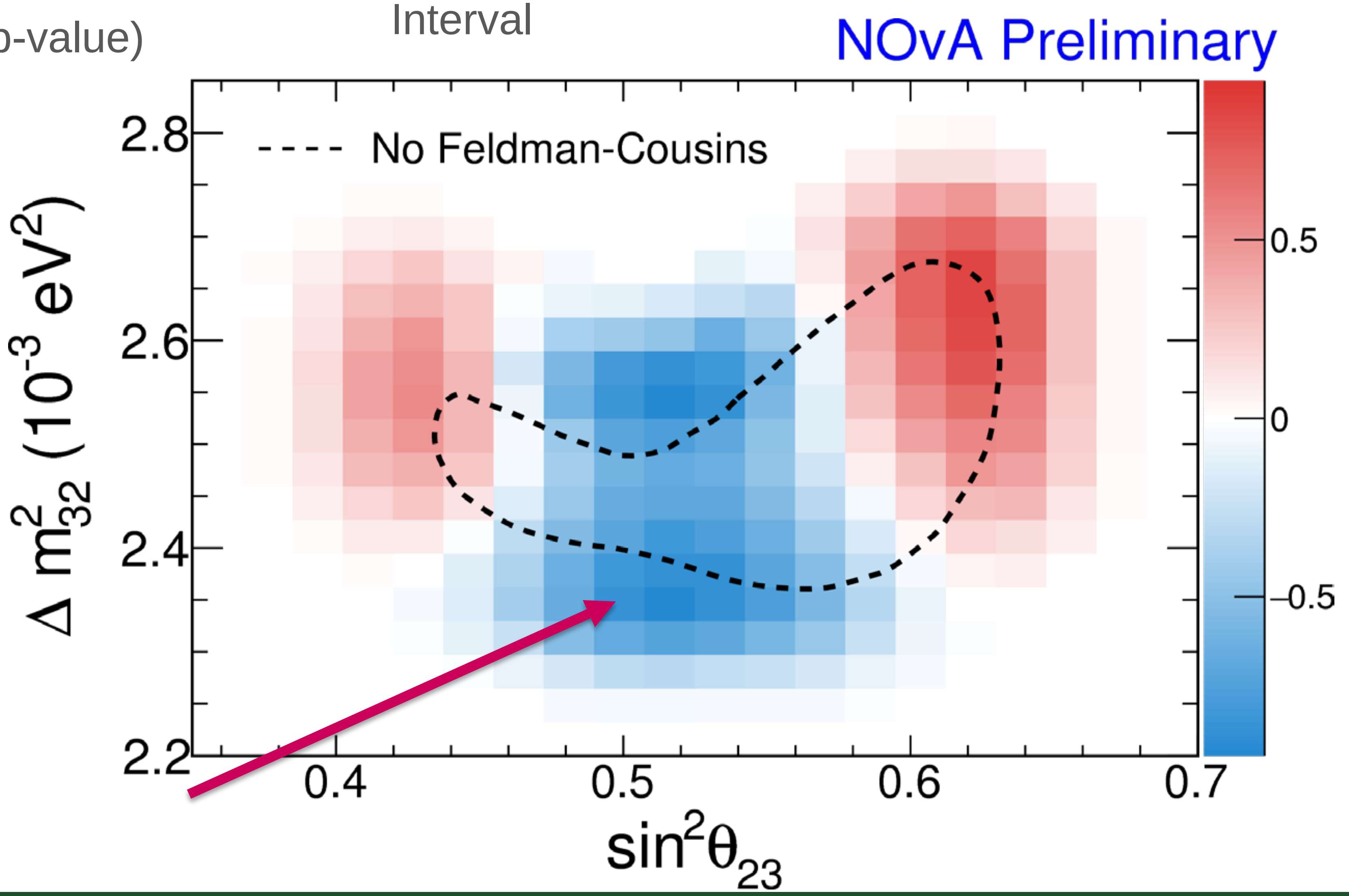


Correcting a contour

- The same thing can be done in 2D
  - Like 1D profiles: Now we find the surface in parameter space that minimizes $\Delta\chi^2$
  - Draw lines of constant significance (p-value)
- Left-skewed distribution:
  - Decreases $\Delta\chi^2_{crit}$
  - Increases ability to exclude

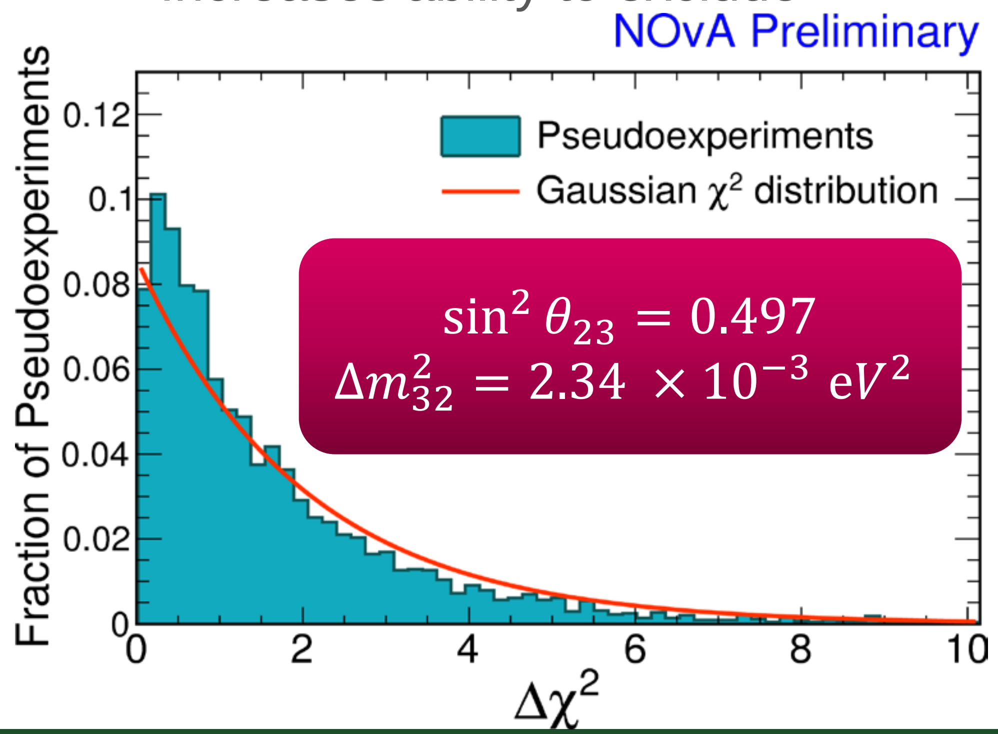


Difference between empirical and gaussian $\Delta\chi^2_{crit}$ for a 90% Confidence Interval

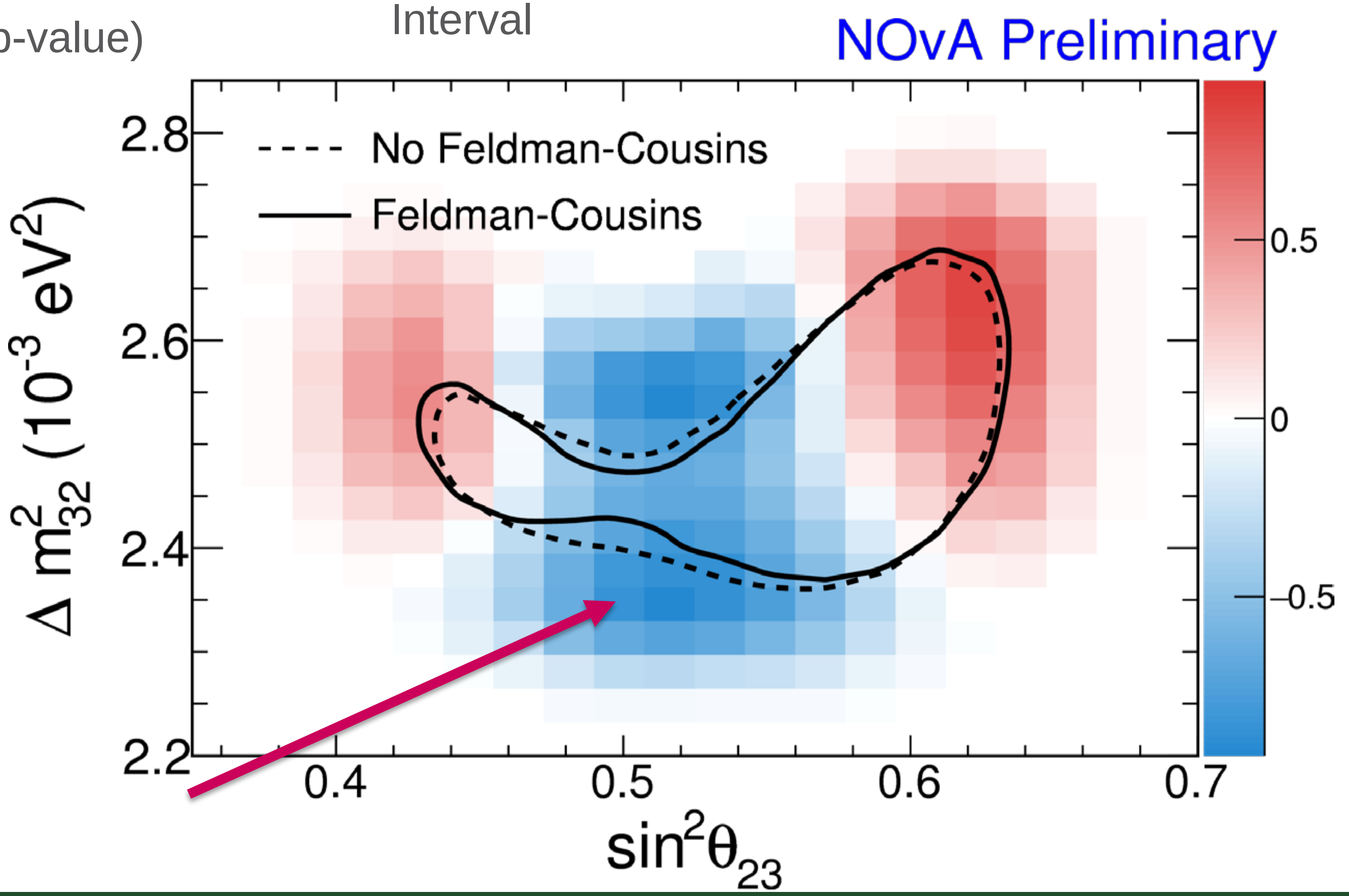


Correcting a contour

- The same thing can be done in 2D
  - Like 1D profiles: Now we find the surface in parameter space that minimizes $\Delta\chi^2$
  - Draw lines of constant significance (p-value)
- Left-skewed distribution:
  - Decreases $\Delta\chi^2_{crit}$
  - Increases ability to exclude



Difference between empirical and gaussian $\Delta\chi^2_{crit}$ for a 90% Confidence Interval



Computational considerations

- We need to generate $\Delta\chi^2$ distributions at each point in parameter space that we're sampling
  - Profiles: 1,210 bins
  - Contours: 471 bins
- In each bin, we need at least 3,000 pseudoexperiments in $\Delta\chi^2$ to generate an accurate empirical distribution
- In previous years, this has been done on the FermiGrid with results in ~4 weeks
- With the addition of antineutrino data and a longer list of systematics, the FermiGrid was no longer an option

Required No. Bins	Required No. Pseudoexperiments
1,681	6,724,000

NERSC – Lawrence Berkeley National Lab

(National Energy Research Scientific Computing)

CORI – Phase I

- Intel "Haswell" at 2.3 GHz
- Total Cores: 76,416
- Capable of 2.81 PFlops
- 32 cores share 128 GB memory

- *Each core can support 2 FC jobs*

CORI – Phase II

- Intel "Knights Landing" (KNL) at 1.4 GHz
- Total Cores: 658,784
- Capable of 29.5 PFlops
- 68 cores share 96 GB memory

- *Each core can support 1 FC job*

The NERSC Advantage

First Run

- Peaked over 1 million cores
  - Largest Condor pool ever!
- Noticed the fits were misbehaving due to added complexity
- *NERSC allowed us to correct for this unforeseen complication and quickly turn around results*

Second Run

- Peaked ~700,000 cores
  - Second Largest Condor pool ever!
- Over 9 million total points
- 36 hours

The NERSC Advantage

2017 ν Analysis*

- Fitting 5 histograms
- 4 profiles
- 4 contours
- ~ 4 weeks to completion (FermiGrid)
- 2 minutes per $\Delta\chi^2$ (NERSC)
- Reproduced on NERSC in a few hours

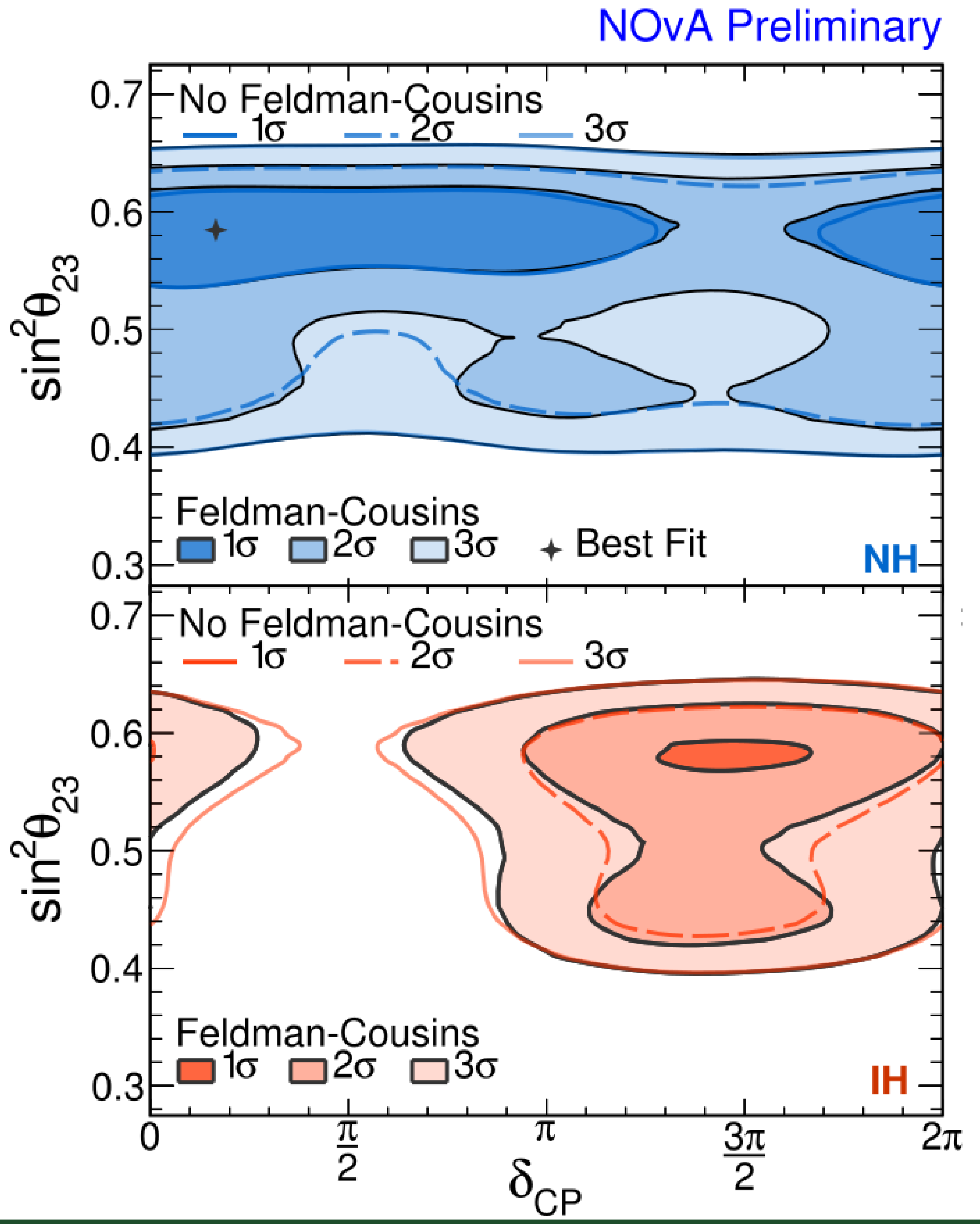
2018 $\overline{\nu} + \nu$ Analysis

- Fitting 10 histograms
- 10 profiles
- 4 contours
- 36 hours to completion (NERSC)
- 10 minutes per $\Delta\chi^2$
- 36 hours to completion (NERSC)
- ~ 5 months on FermiGrid

* arXiv:1806.00096 [hep-ex]

Results and Conclusions

- Feldman-Cousins Corrections have a large effect on our results
- Low statistics and physical boundaries require an empirical approach to stating a significance
- Part of a larger project: HEP Data Analytics on HPC
  - Pythia8 and detector simulation tuning for LHC experiments
  - Improve performance and precision of neutrino cross-section analyses
- We hope to develop tools made available to all members of the HEP community in achieving similar advances



Thank you

Colorado State University

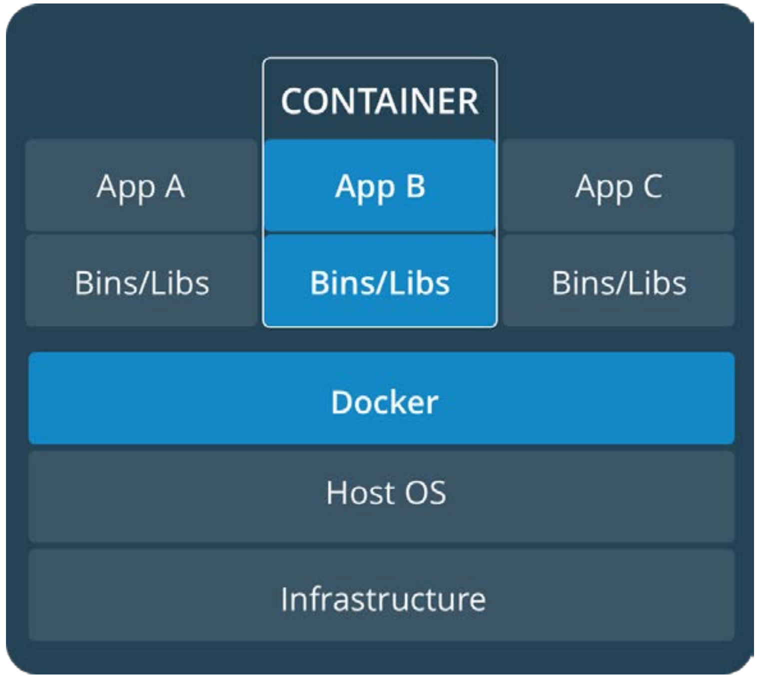
U.S. DEPARTMENT OF ENERGY | Office of Science

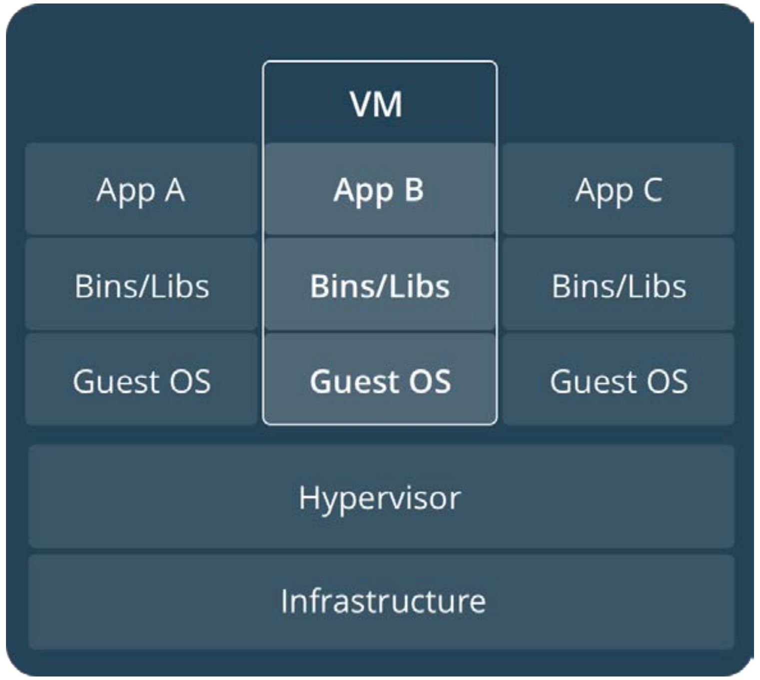
NERSC National Energy Research Scientific Computing Center

NOvA

How we run on CORI

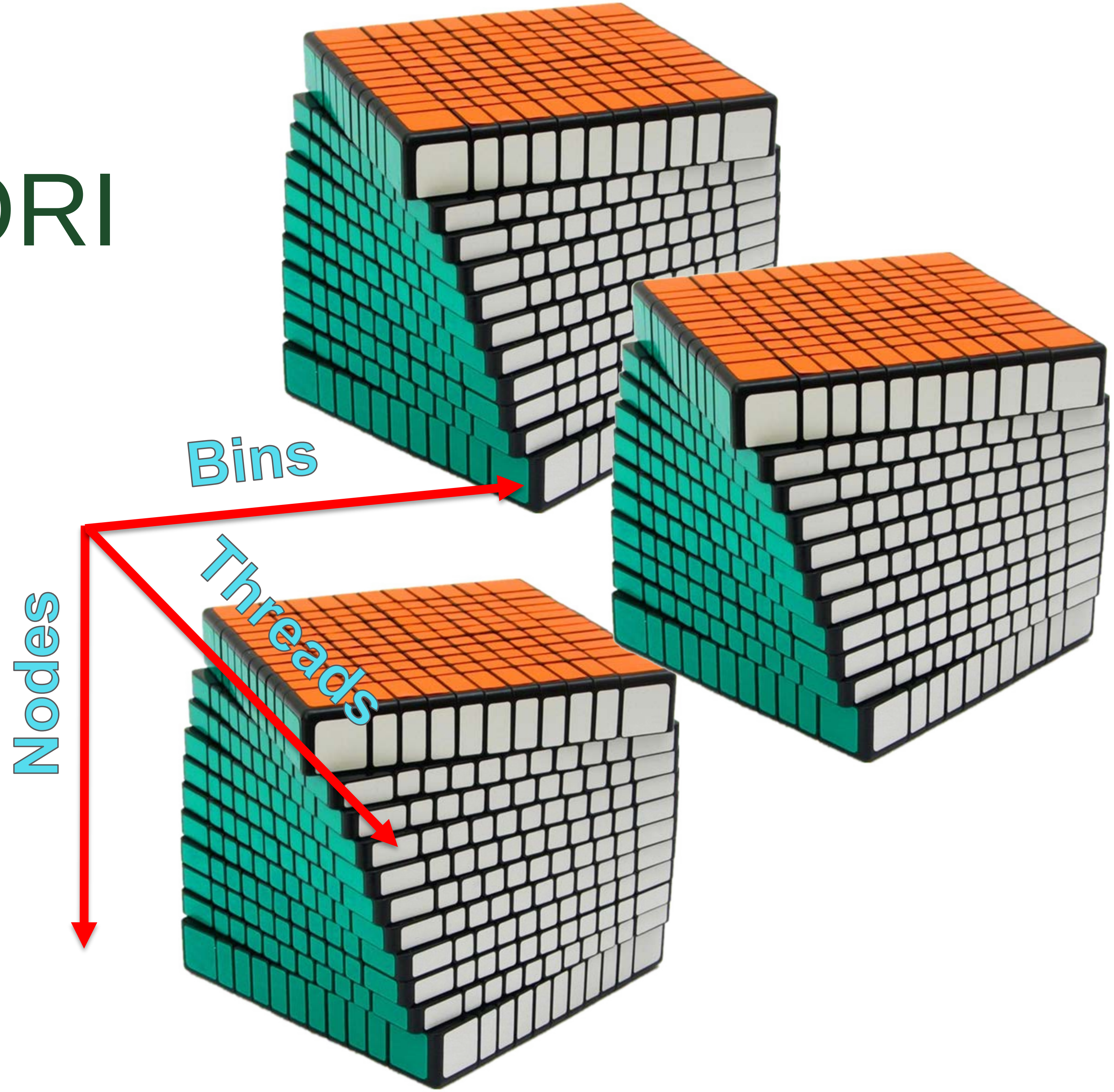
- Docker Images: Like a Virtual Machine, but has a direct line to the machine's kernel
- What is in our image:
  - Slf6.7
  - NOvASoft
  - Spectra/inputs to the FC root macro
- Jobs submitted to CORI mount the image
  - Each job can run a separate image instance, called a *container,* on available threads





How we run on CORI

- Submit to CORI like a giant Rubix Cube
  - Each compute node has a number of threads that each run a separate instance of the macro
  - Each macro runs over all bins few FC points in each bin
  - Scale by requesting a large number of nodes
- Wait in regular queue: ~ 3 days for big jobs
- Make a reservation: instant access to requested resources



Inverted Hierarchy Rejection

- H_o: $\Delta m^2_{32} < 0$
- Need a distribution where this is true
- Profile out dependencies on the value of Δm^2_{32}
- Let $\Delta\chi^2 = 0$ if $\vec{\theta}$ is in Inverted Hierarchy

Observed $\Delta\chi^2$	Significance
2.47	1.77