Efficient Neutrino Oscillation Parameter Inference with Gaussian Process

Ref: 10.1103/PhysRevD.101.012001

Lingge Li, Nitish Nayak, Yiwen Xiao, Jianming Bian, Pierre Baldi

UC-Irvine

New Perspective 2.0 - 2020

Neutrino Oscillation

- Neutrinos: 2 kinds of states, each comes with 3 types
  - Flavor States $(\nu_e, \nu_\mu, \nu_\tau)$ — what we observed
  - Mass Eigenstates (ν_1, ν_2, ν_3) — what in between observations
- Principle of superposition connects them via PMNS matrix, i.e.

$$\begin{bmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{bmatrix} = U_{PMNS} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix}$$

$$U_{PMNS} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13} \end{pmatrix} \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Neutrino Oscillation

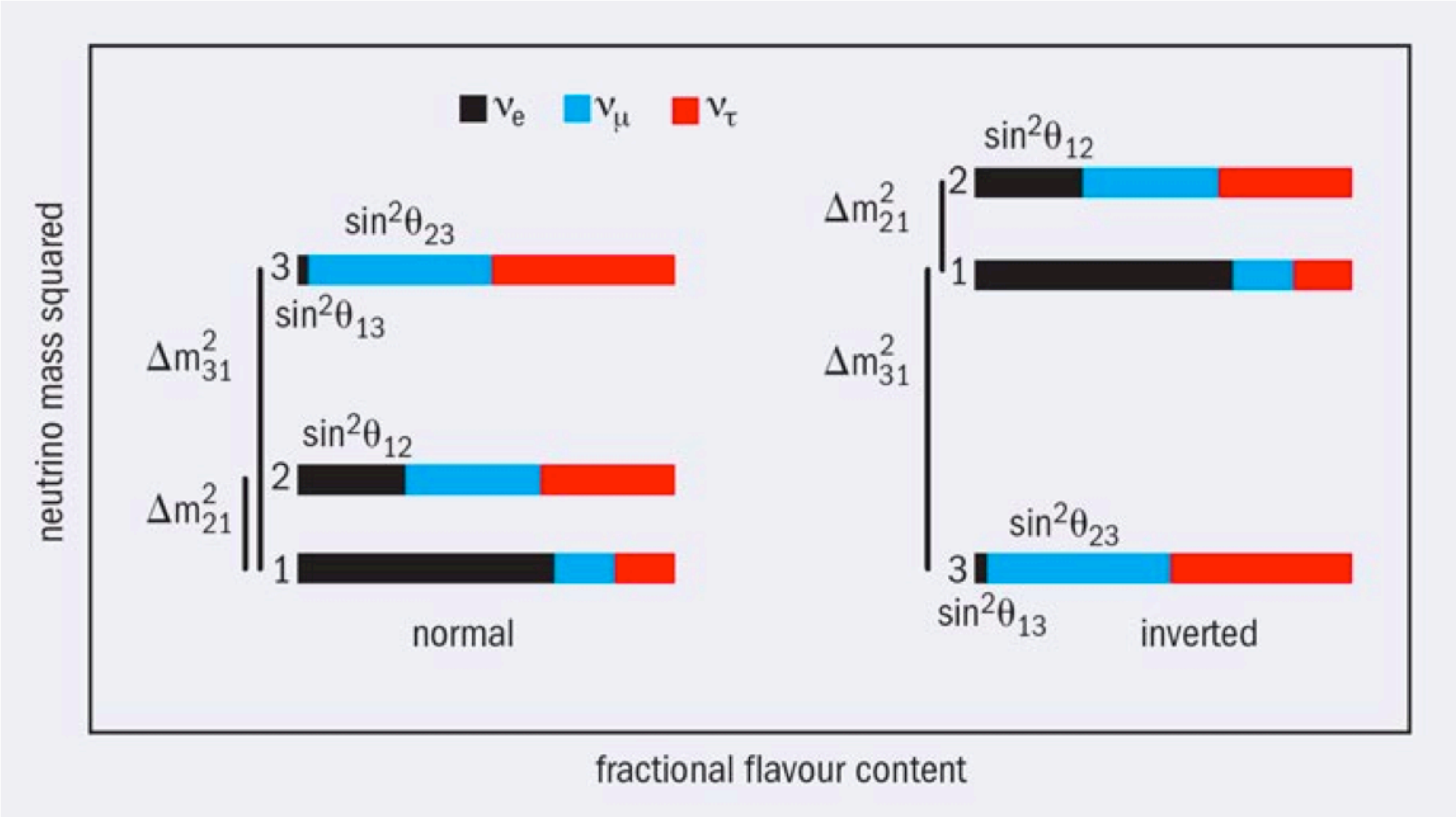
- For neutrino propagating in vacuum,

$$P(\nu_\alpha \to \nu_\beta) = \delta_{\alpha\beta} - 4\sum_{i>j}^{3} \Re(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta i}) \sin^2(\frac{\Delta m^2_{ij} L}{4E_\nu}) + 2\sum_{i>j}^{3} \Im(U^*_{\alpha i} U_{\beta i} U_{\alpha j} U^*_{\beta i}) \sin(\frac{\Delta m^2_{ij} L}{4E_\nu})$$

- Broadly, solar experiments give handle on (21) parameters, reactor experiments for θ_{13}
- Long baseline (LBL) experiments gives handle on (32) parameters
  - $P(\nu_\mu \to \nu_\mu)$ is sensitive to $\sin^2(2\theta_{23})$ and $|\Delta m^2_{32}|$
  - Non-zero θ_{13} opens up $P(\nu_\mu \to \nu_e)$ channel, sensitive to δ_{CP} , θ_{23} octant, and $sgn(\Delta m^2_{32})$

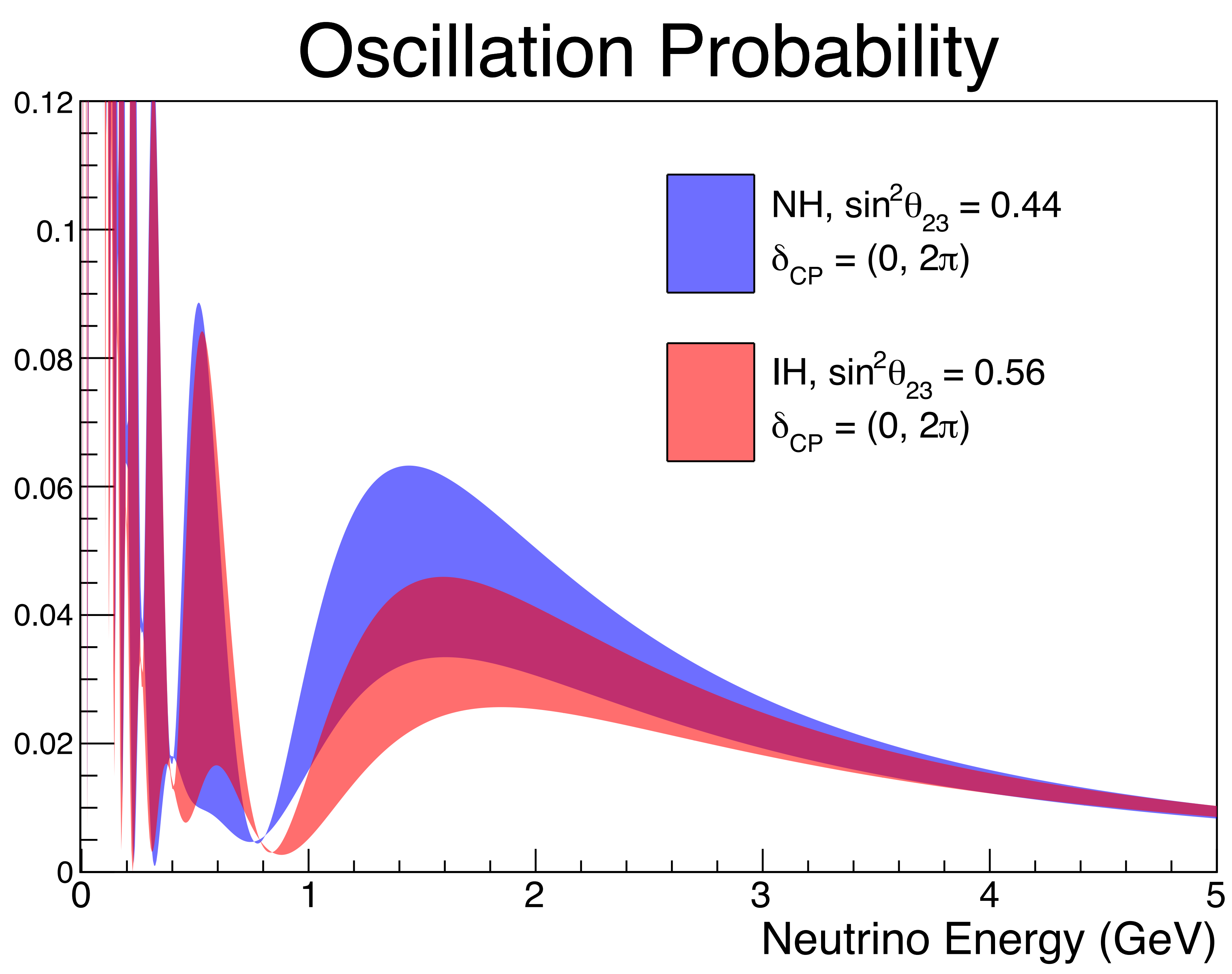
Neutrino Oscillation

- In LBL experiments, we want to know if
  - $\Delta m^2_{32} > 0$ or < 0 ? (Normal Hierarchy or Inverted Hierarchy)
    - Has implication for neutrino mass measurements
  - Octant of θ_{23} or $\theta_{23} = 45\,°$?
  - $\sin(\delta_{CP}) \neq 0$?
    - Indicate whether CP-violation exists. (matter-antimatter asymmetry)



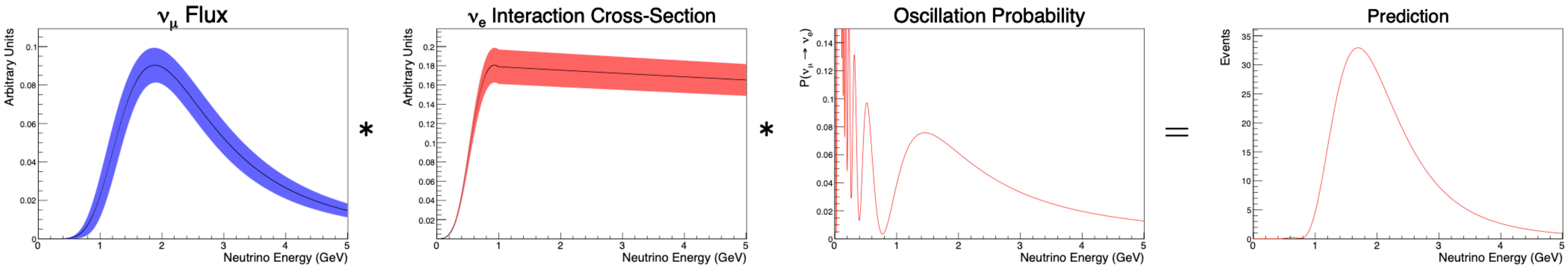
Statistic Issues

- Oscillation parameters usually measured via the Maximum Likelihood Estimation (MLE) using the PMNS model and comparing it to the observation, such as the energy spectrum.
- However, LBL experiments (T2K, NOvA) only collect a handful of statistics over years of operation.
- Oscillation Probability have complicated dependence on multiple parameters —> difficult to delineate
- Therefore, Confidence Intervals are hard to construct



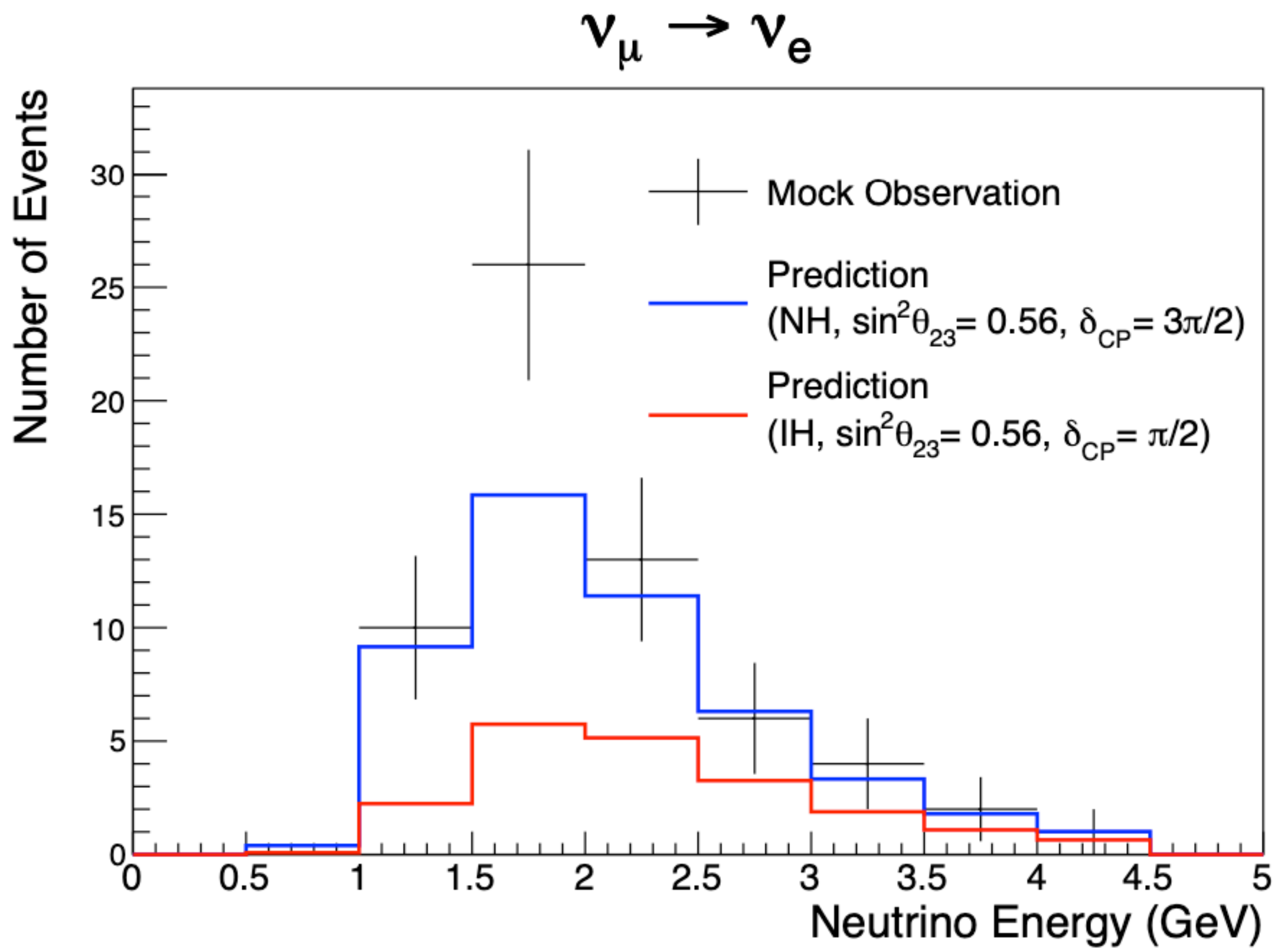
Toy Experiment

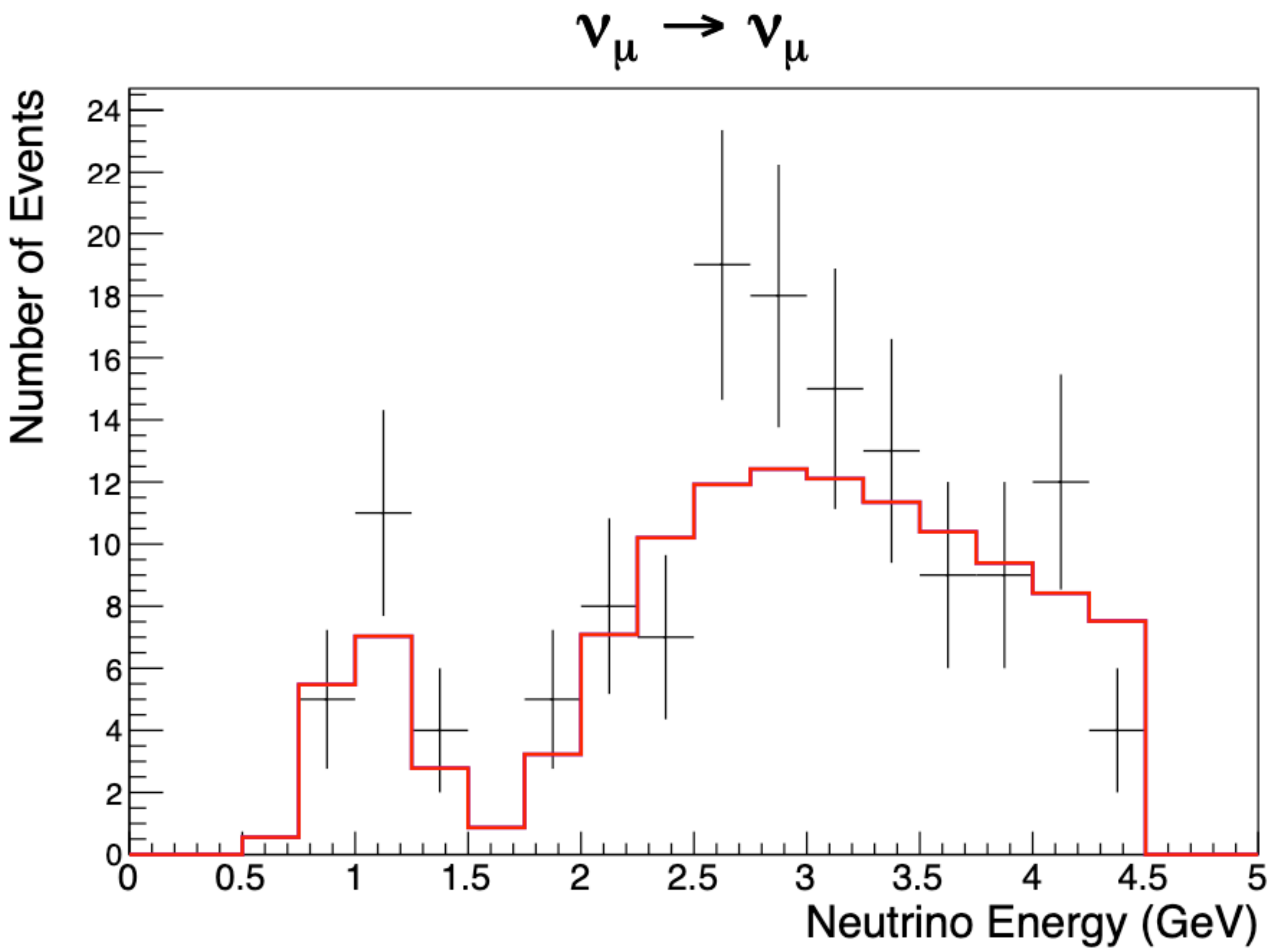
- Modeled on NOvA. Set, baseline, $L = 810$km with ν_μ flux peaking at 2 GeV
- Add, 10% normalization error on flux and cross-section model.
- Get, $\nu_\mu \rightarrow \nu_e$ prediction by multiplying toy shapes for flux, cross-section and oscillation probability.



- Simulate both appearance, $\nu_\mu \rightarrow \nu_e$, and disappearance, $\nu_\mu \rightarrow \nu_\mu$.

Toy Experiment





- Toy data ($\vec{x}$) from Poisson variation with some chosen oscillation parameters.
- Including both oscillation parameters, θ, and nuisance parameters (flux and cross-sections error), δ.
- Best fit $(\hat{\theta}, \hat{\delta})$ is found by minimizing negative log-likelihood over every energy bins, i

$$-2\log L(\theta,\delta) = -2\sum_{i\in I}\log Pois(x_i; v(\theta,\delta)_i) - \sum_{i\in I} x_i + \sum_{i\in I} v(\theta,\delta)_i + \delta^2$$

Confidence Interval

- Typically, using Likelihood Ratio Test (LRT) to estimate confidence interval

$$\Delta\chi^2 = -2\log\frac{L(\theta_0)}{\arg\max_\theta L(\theta)}$$

- In asymptotic case, test statistic: $\Delta\chi^2 \sim \chi^2_k$ （Wilks Theorem)

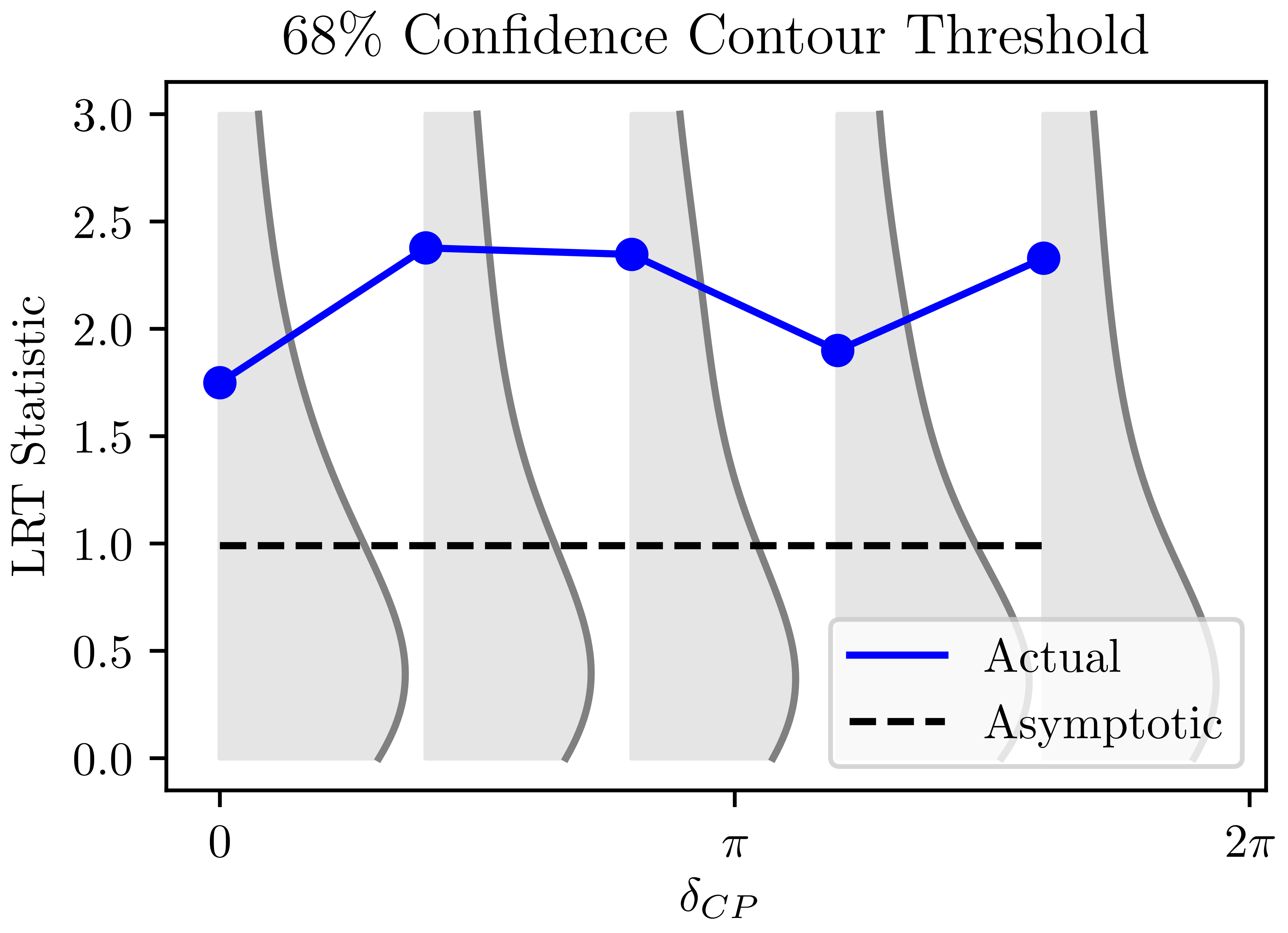
Table 38.2: Values of $\Delta\chi^2$ or $2\Delta\ln L$ corresponding to a coverage probability $1-\alpha$ in the large data sample limit, for joint estimation of m parameters.

$(1-\alpha)$ (%)	$m=1$	$m=2$	$m=3$
68.27	1.00	2.30	3.53
90.	2.71	4.61	6.25
95.	3.84	5.99	7.82
95.45	4.00	6.18	8.03
99.	6.63	9.21	11.34
99.73	9.00	11.83	14.16

From the PDG Review on Statistics

Feldman - Cousins

- Due to the small sample size in neutrino data and physical boundaries on the oscillation parameters, the asymptotic distribution is unreliable
- Explicitly simulate $\Delta\chi^2$ distribution using lots of pseudo-experiments
- Find p-value associated with $\Delta\chi^2_{data}$ for each point
- In practice, FC conducts a grid-search over the entire parameter space with many toy Monte-Carlo (MC) — Time Consuming
- Want a refined algorithm.
  - Approximating FC P-value surface non-parametrically using only a fraction of grid points



Gaussian Process

- A Gaussian Process (GP) is a special case of Bayesian learning.
- Technically, GP can be specified by a mean function, $\mu(x)$ and a covariance function (kernel), $k(x, x')$
    - Assume a collection of random models with certain probability (Priors with mean and standard deviation)

$$\begin{pmatrix} f(x) \\ f(x') \end{pmatrix} \sim \mathcal{N}(\begin{bmatrix} \mu(x) \\ \mu(x') \end{bmatrix}, \begin{bmatrix} k(x,x) & k(x,x') \\ k(x,x') & k(x',x') \end{bmatrix})$$

    - Observed data update Priors —> Posteriors (Predictions for new data)

$$f(x')|f(x) \sim \mathcal{N}(\frac{k(x,x')}{k(x,x)} f(x), k(x',x') - \frac{k(x,x')^2}{k(x,x)})$$

    - Quantifies uncertainty in model estimates (Posterior mean and standard deviation)

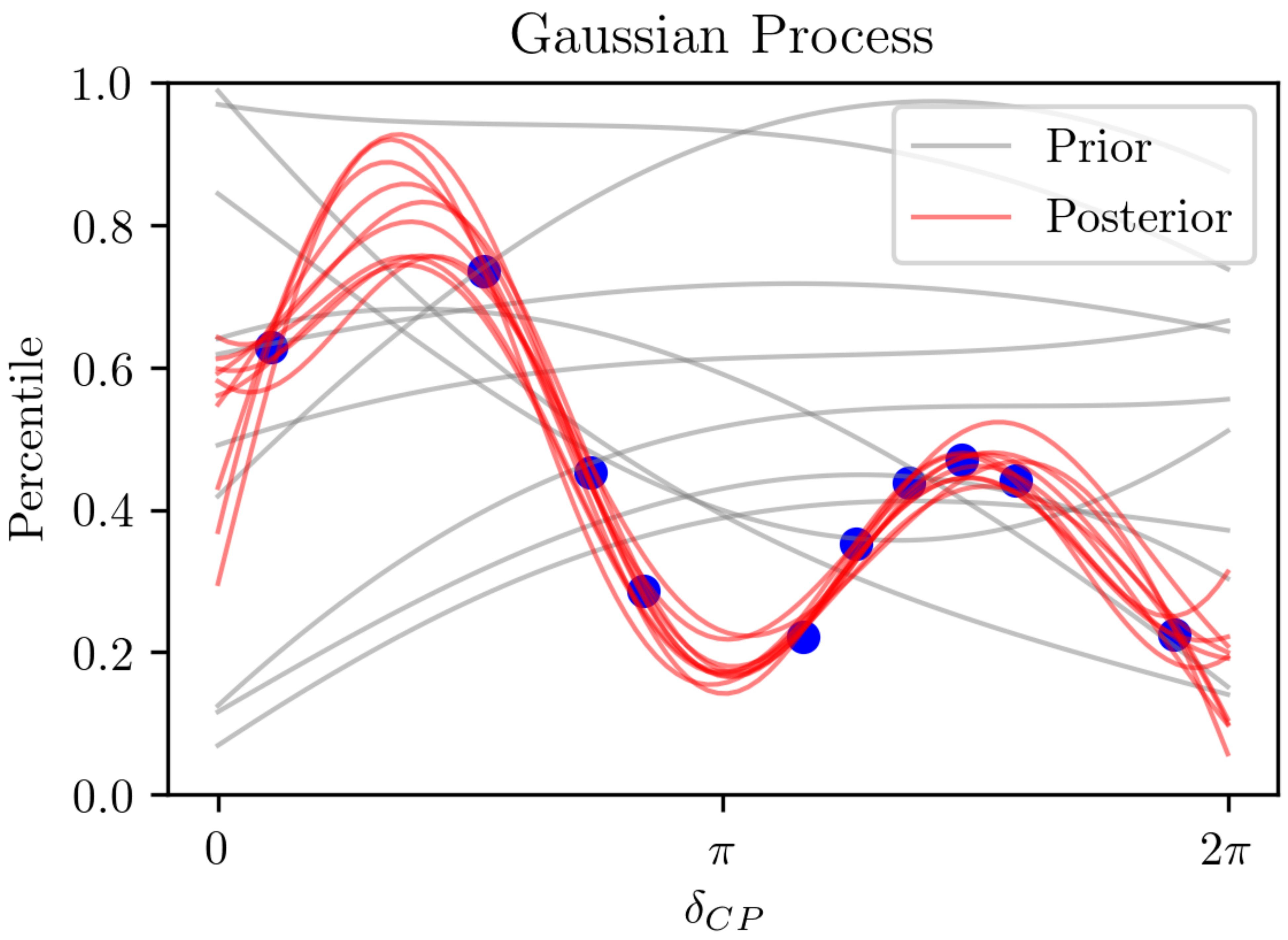
Gaussian Process

- A common choice of GP kernel k is the squared exponential radial basis function (RBF)

$$k(x_1, x_2) = \exp(-\frac{(x_1 - x_2)}{l^2})$$

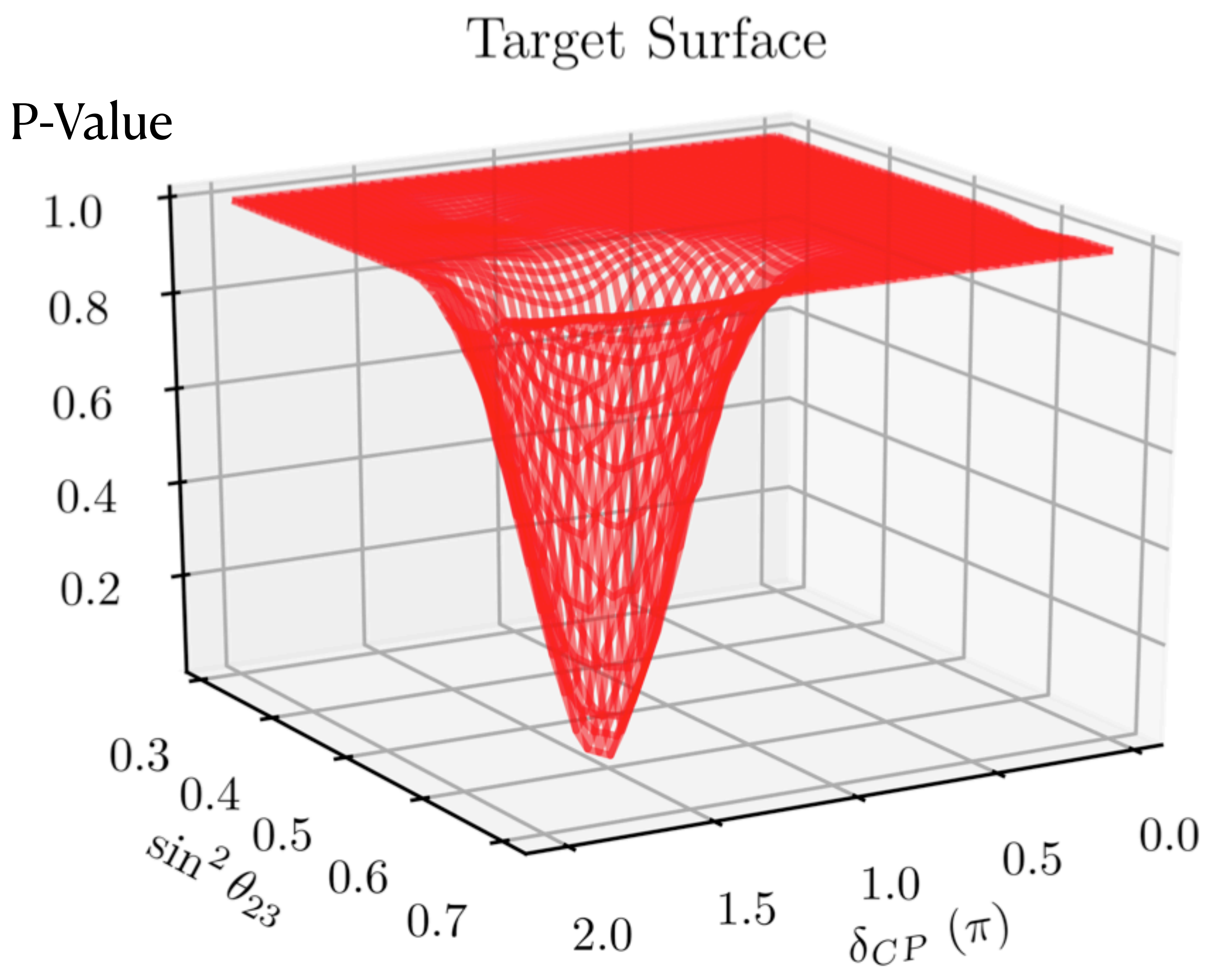
, which tells us that GP results at nearby points are highly influenced by observations at a given point while further out, they aren't.

- ***GP uses the kernel function and measured data to predict the value for an unseen point with posterior mean and posterior standard deviation***



Gaussian Process for Feldman-Cousins Method

- Fitting a GP to target p-value surface for a given contour.
- Reduce the time-cost by throwing pseudo-experiments at some point based on approximation, instead of all points in parameter space.
- Construct Confidence Interval based on the P-value surface.

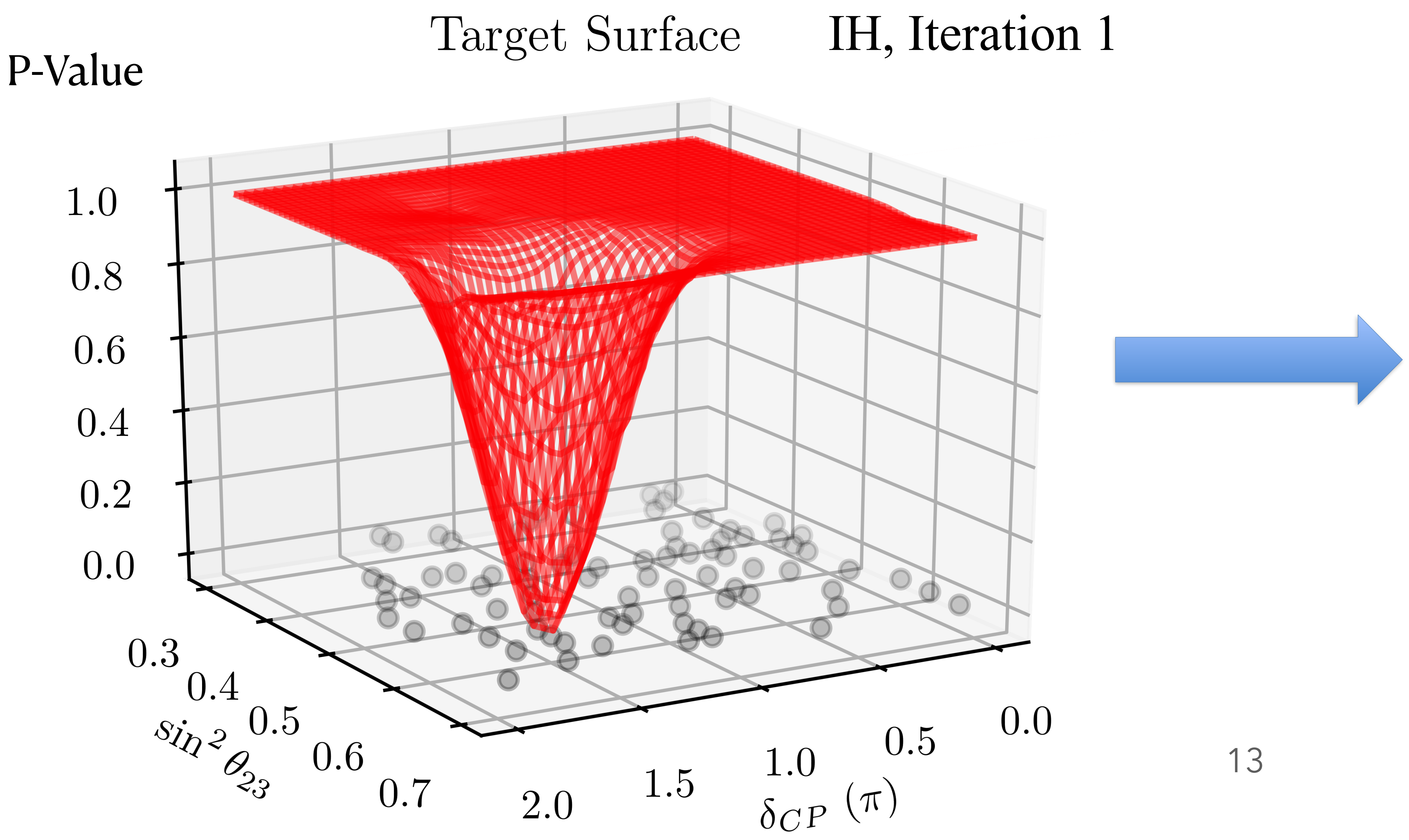


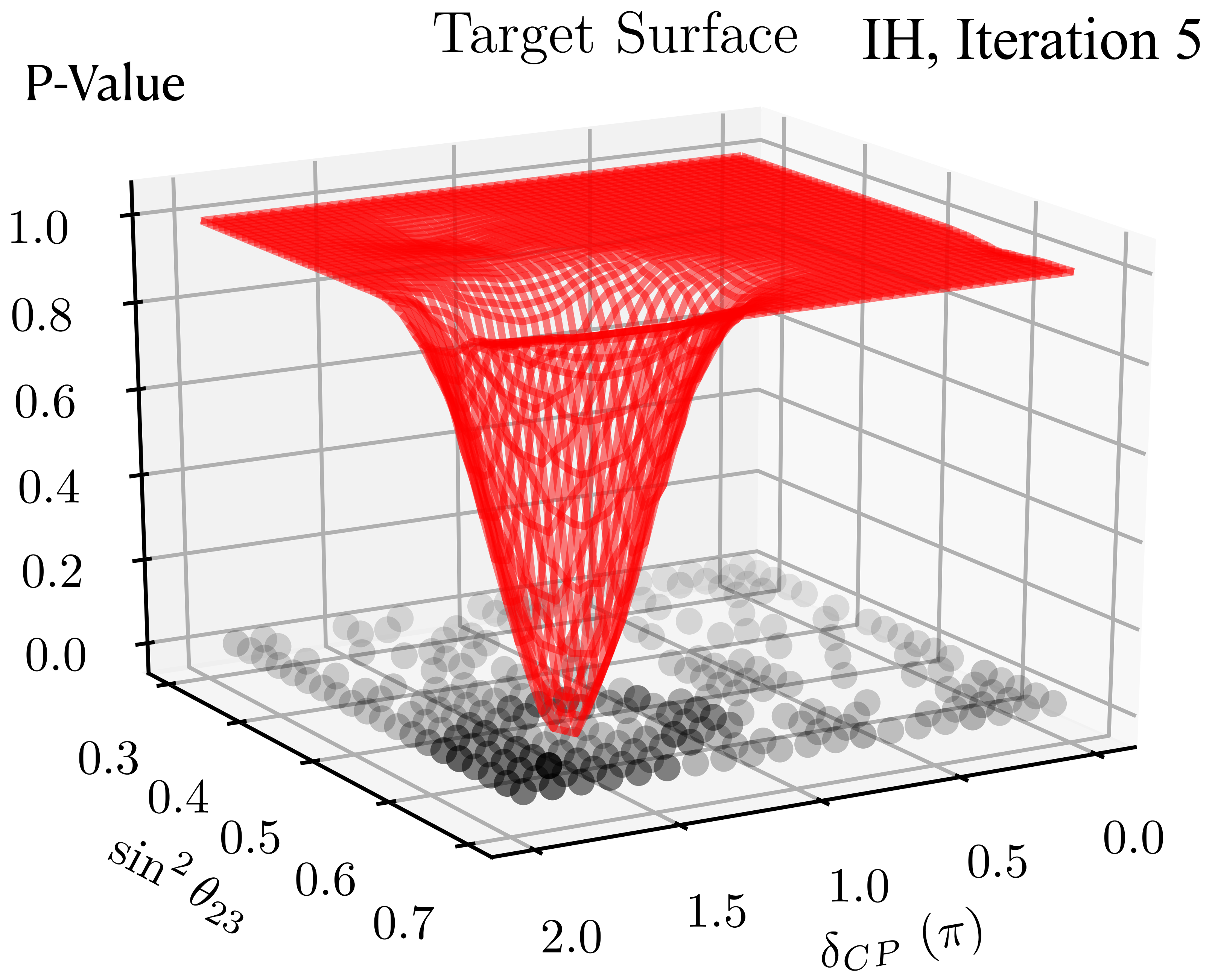
Optimized Confidence Interval Search

- Use an acquisition function that proposed new points in θ-space to explore based on GP approximated p-value surface

$$a(\theta) = \sum_{\alpha_i} \left| \frac{\hat{q}(\theta) - \alpha_i}{\sigma_{\hat{q}(\theta)}} \right|^{-1}$$

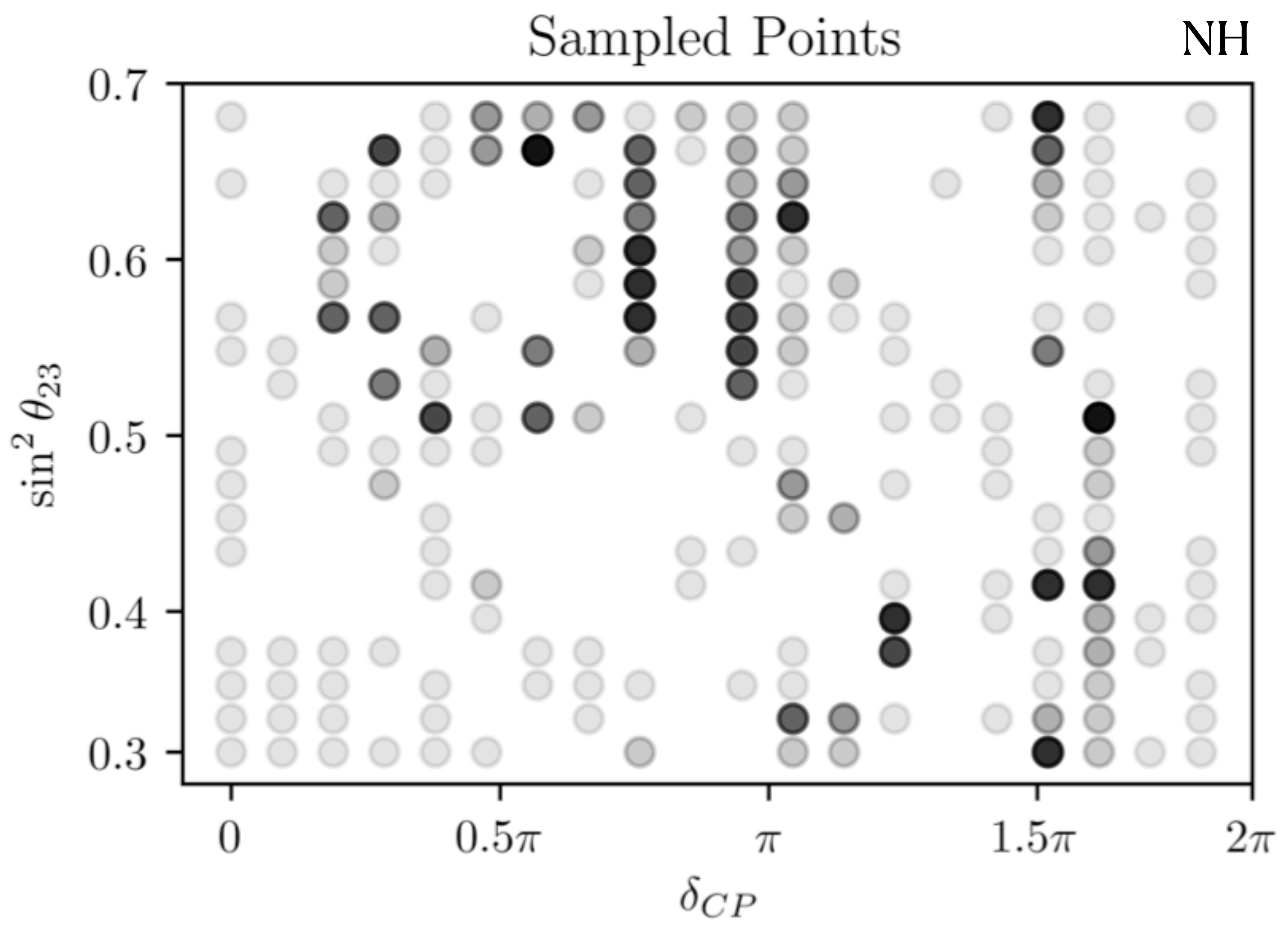
- Here, $\hat{q}(\theta)$ is GP mean, $\sigma_{\hat{q}(\theta)}$ is a GP uncertainty, α_i designates confidence levels, e.g. 68% or 90%
- $a(\theta)$ balances between exploration, i.e. reducing approximation uncertainty, and exploitation, i.e. reaching the extremum.

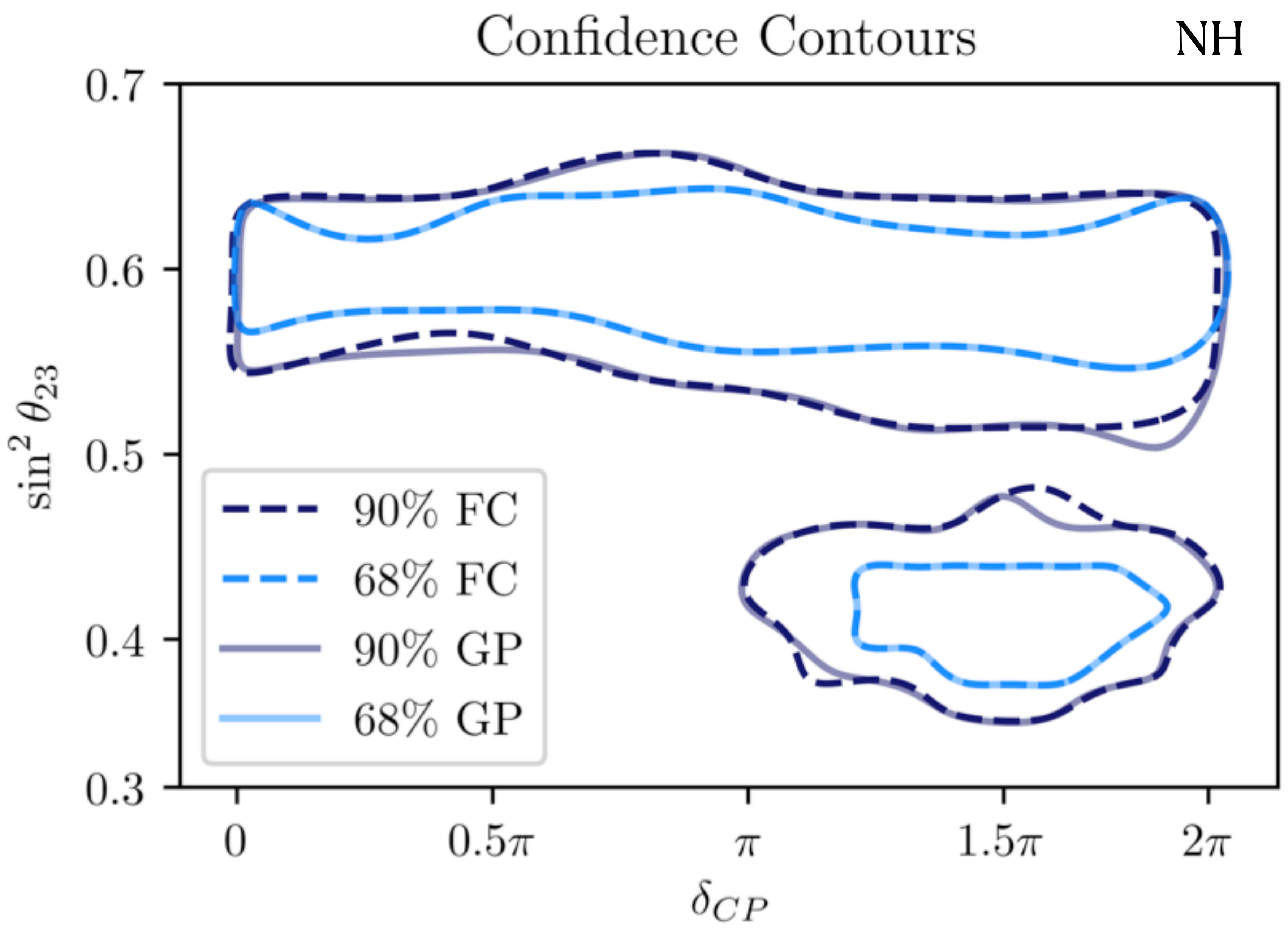




Results

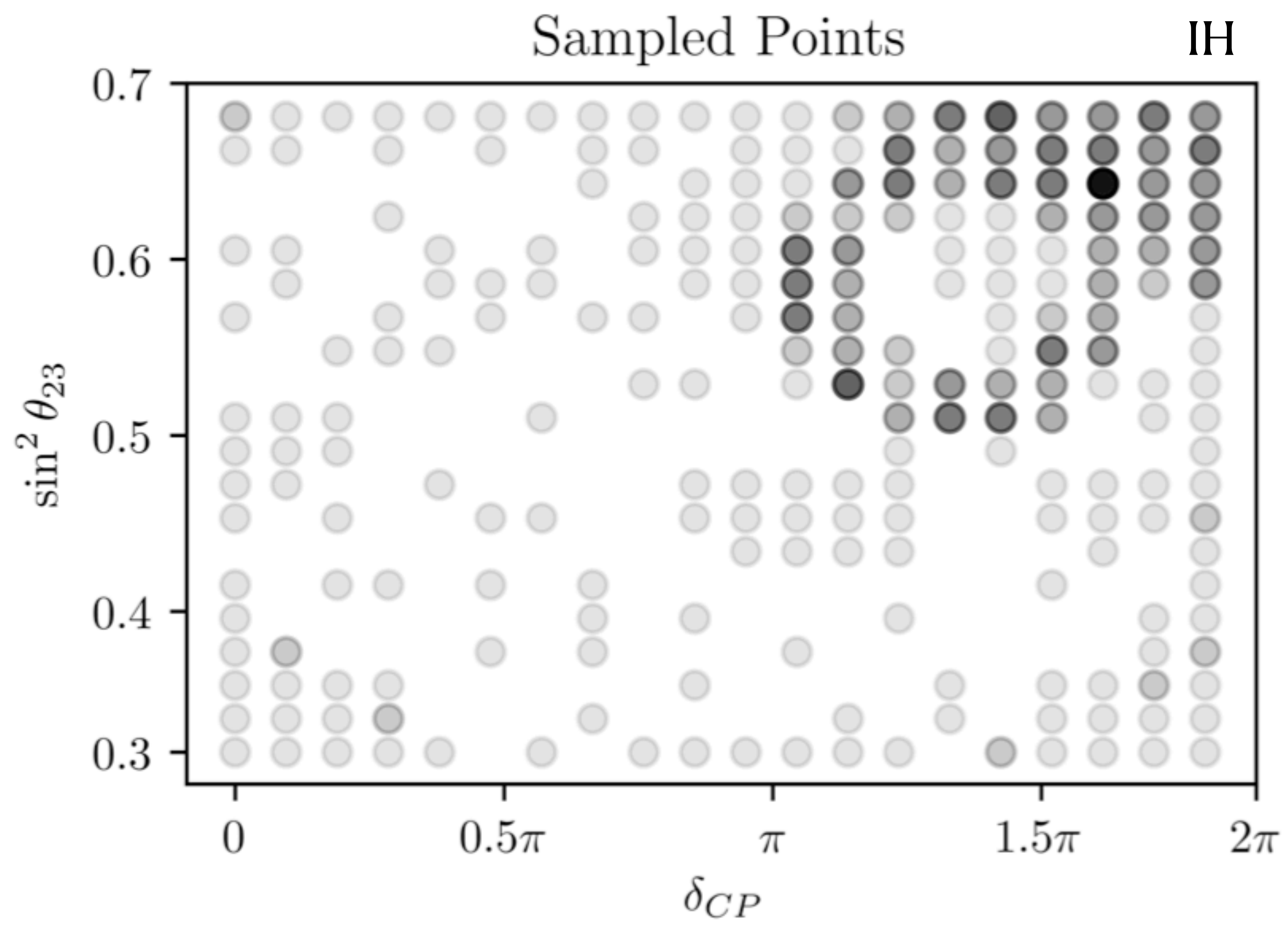
- Oscillation parameters similar to 2018 best estimate from NOvA ($\theta_{23} = 0.56$, $\Delta m^2_{32} = 2.44 * 10^{-3} eV^2$, $\delta_{CP} = 1.5\pi$)
- $\sin^2\theta_{23} - \delta_{CP}$ 68% and 90% CI for NH after 5 iterations.

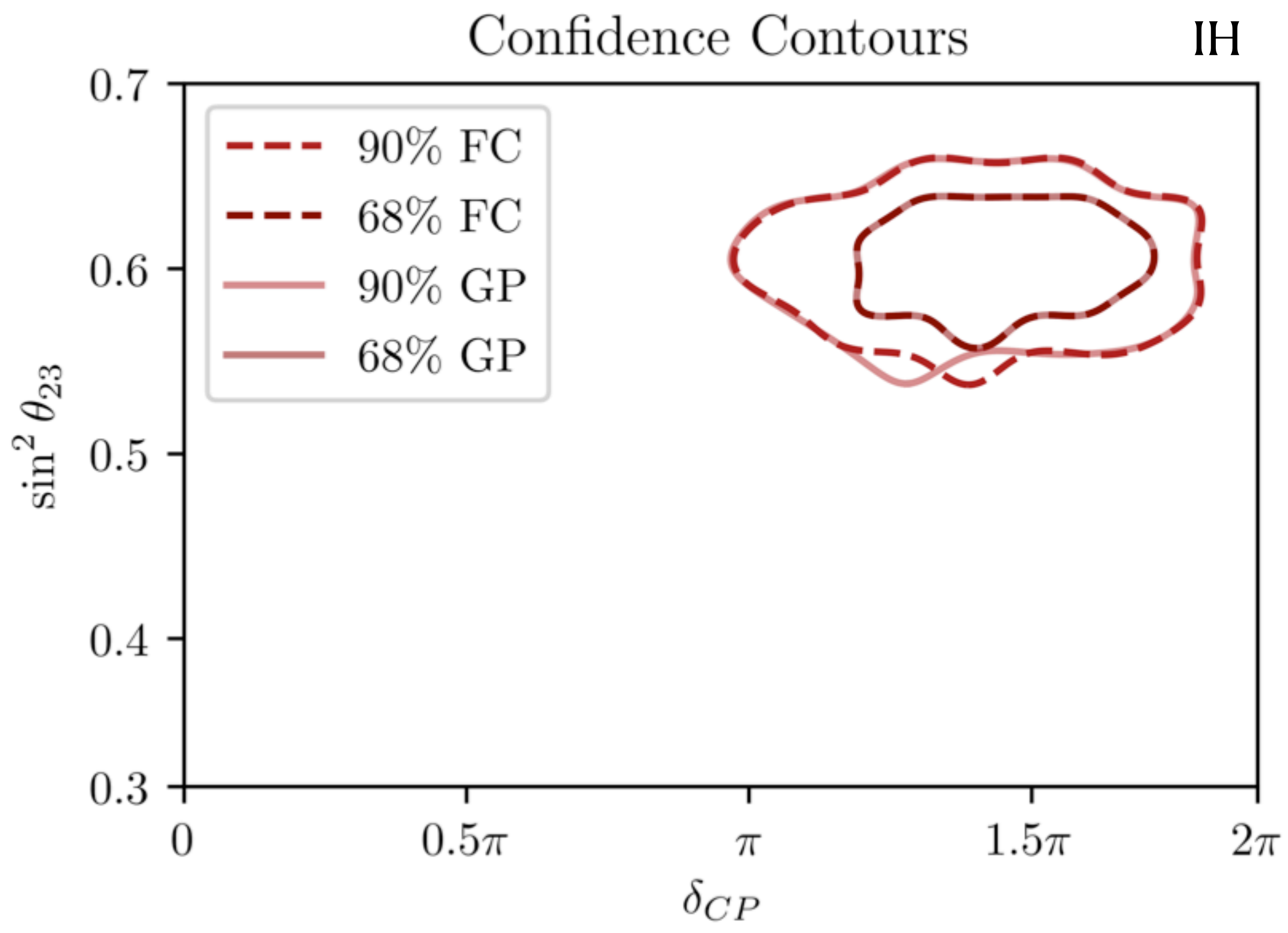




Results

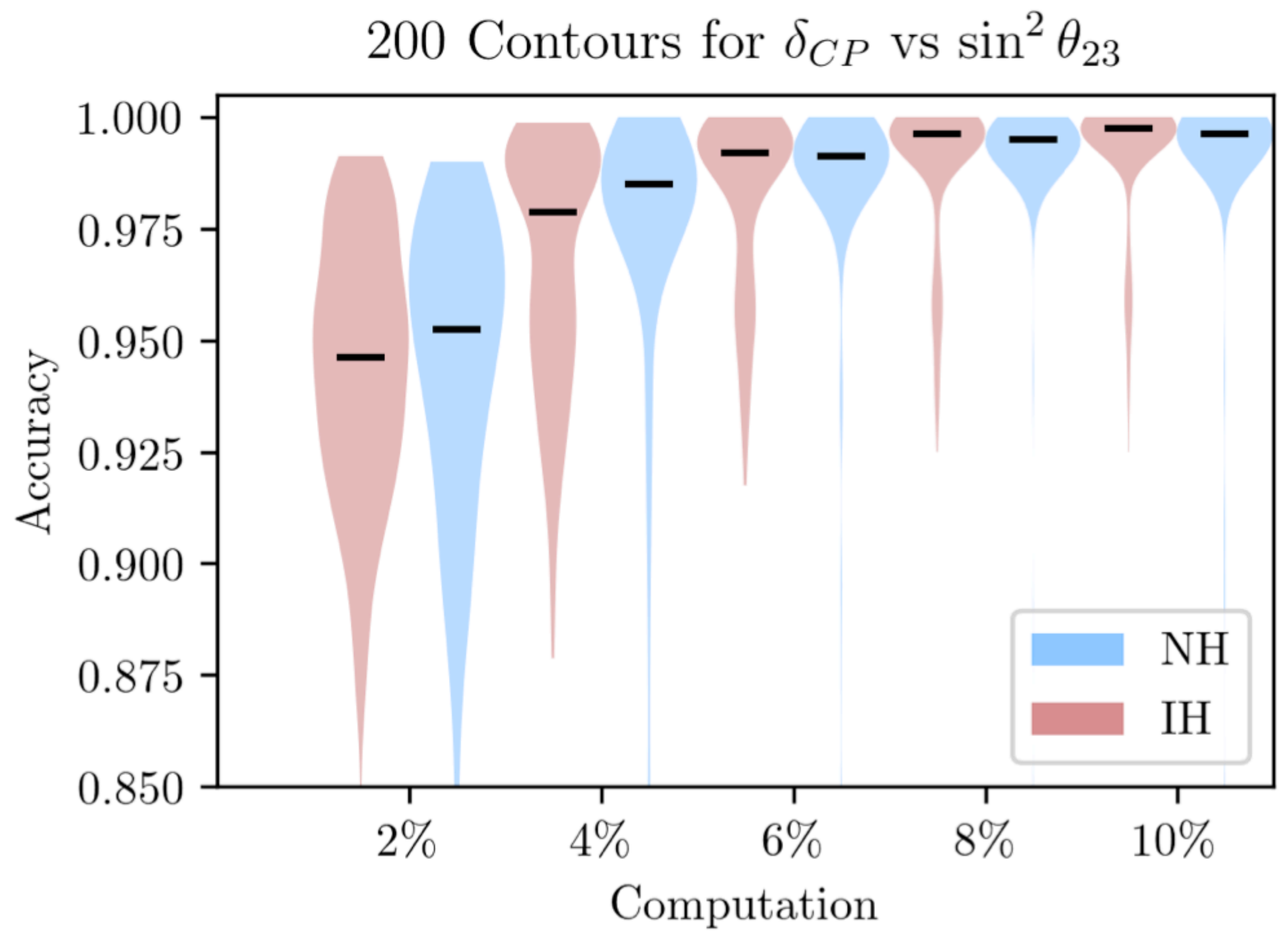
- Oscillation parameters similar to 2018 best estimate from NOvA ($\theta_{23} = 0.56$, $\Delta m^2_{32} = 2.44 * 10^{-3} eV^2$, $\delta_{CP} = 1.5\pi$)
- $\sin^2\theta_{23} - \delta_{CP}$ 68% and 90% CI for IH after 5 iterations.





Results

- Use classification accuracy of all grid points, taking FC result as truth, to evaluate performance
- Progress shows the search algorithm converges to the FC value ~ $10\times$ faster for 2D case
- Median Accuracy for for 2D is $> 99.5\%$ for both NH and IH



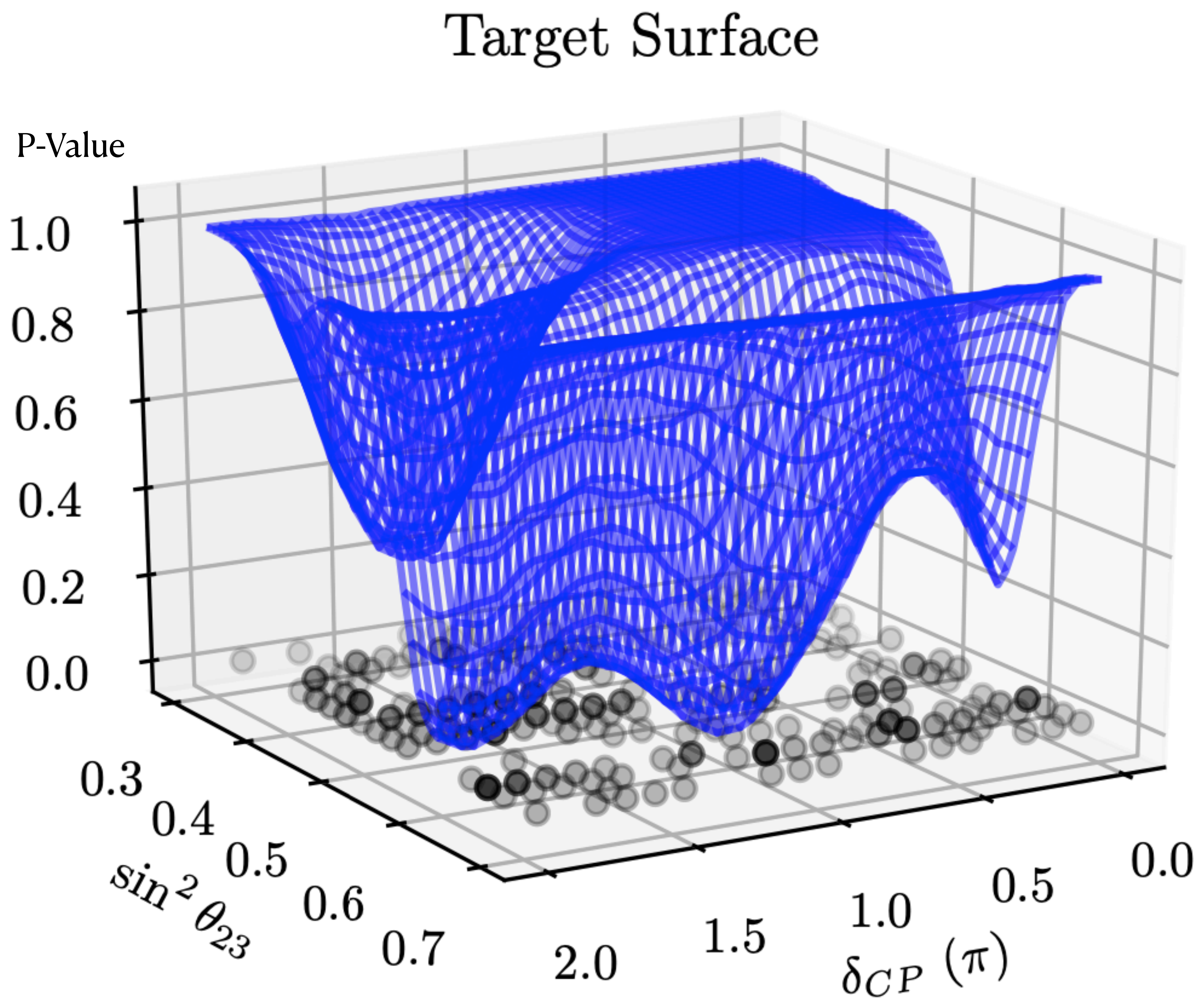
Summary

- Neutrino Oscillation experiments provide interesting test case for estimating frequentist confidence intervals.
- LBL experiments typically proceed via Feldman-Cousins
- However, simulating $\Delta\chi^2$ distribution across multi-dimensional parameter space requires huge computational source
- We've studied Gaussian Process on a toy LBL set-up
- Helps us estimate frequentist contour edges to quite a high accuracy without having to sample the entire parameter space.
- See publication for more details : Phys.Rev.D 101 (2020) 1, 012001
- All code with illustrative notebooks here : https://github.com/nitish-nayak/ToyNuOscCI, maintained by Lingge (linggeli7@gmail.com), Nitish (nayakb@uci.edu), and Yiwen (yiwenx7@uci.edu)

Thanks!

Backup

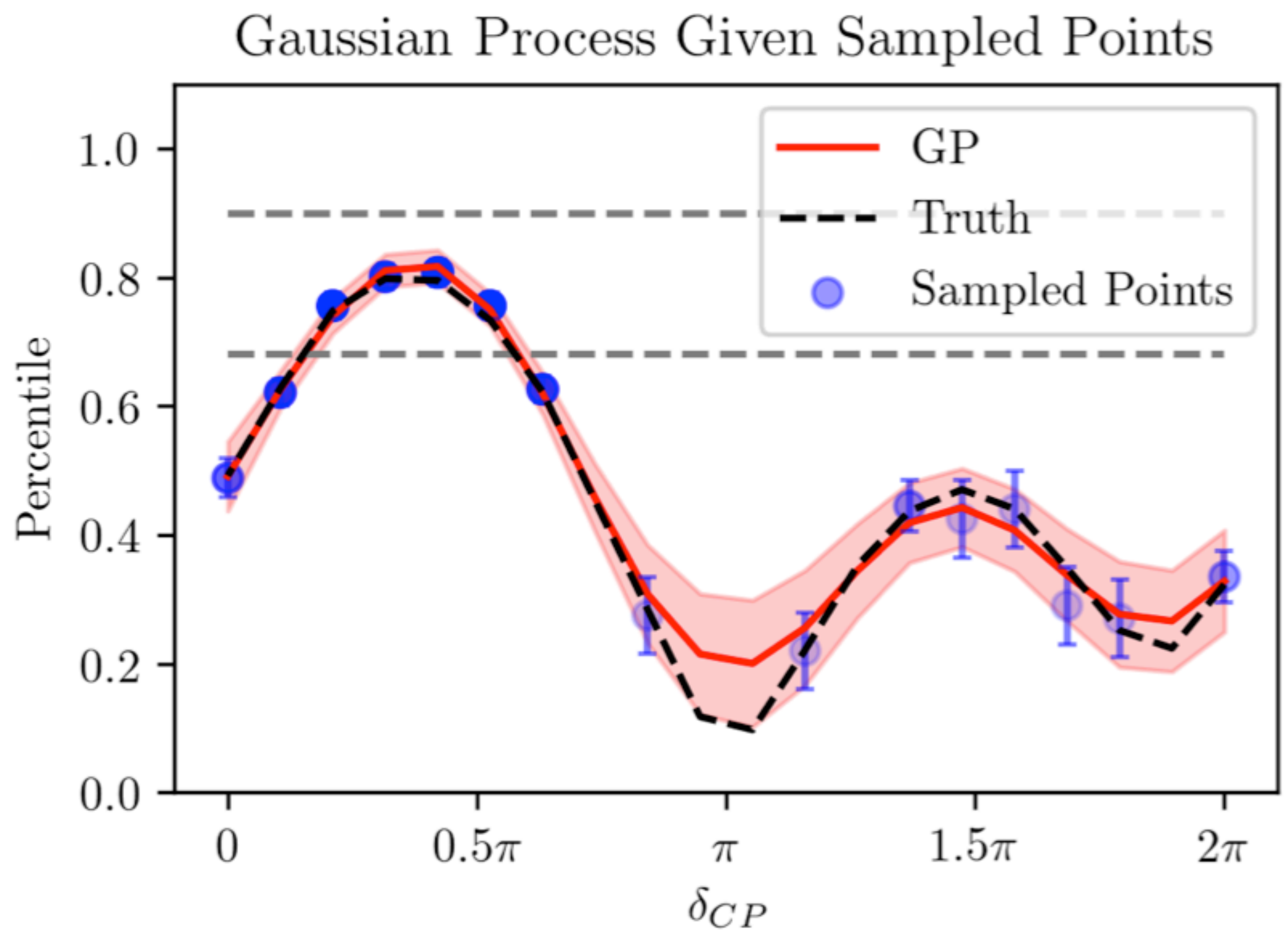
Results

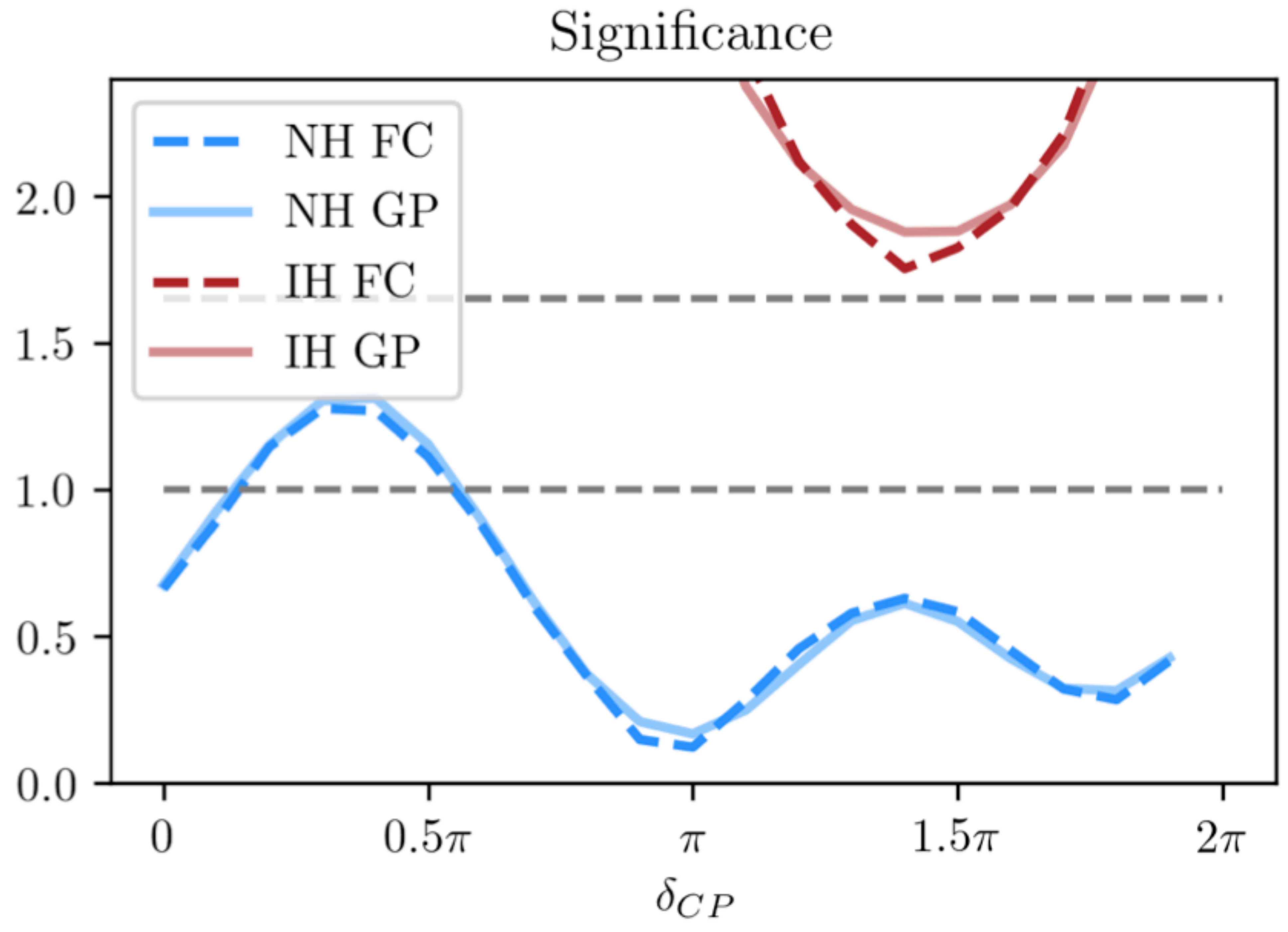


Target Surface
P-Value
1.0
0.8
0.6
0.4
0.2
0.0
0.3
0.4
0.5
0.6
0.7
$\sin^2\theta_{23}$
2.0
1.5
1.0
0.5
0.0
δ_{CP} (π)

Results

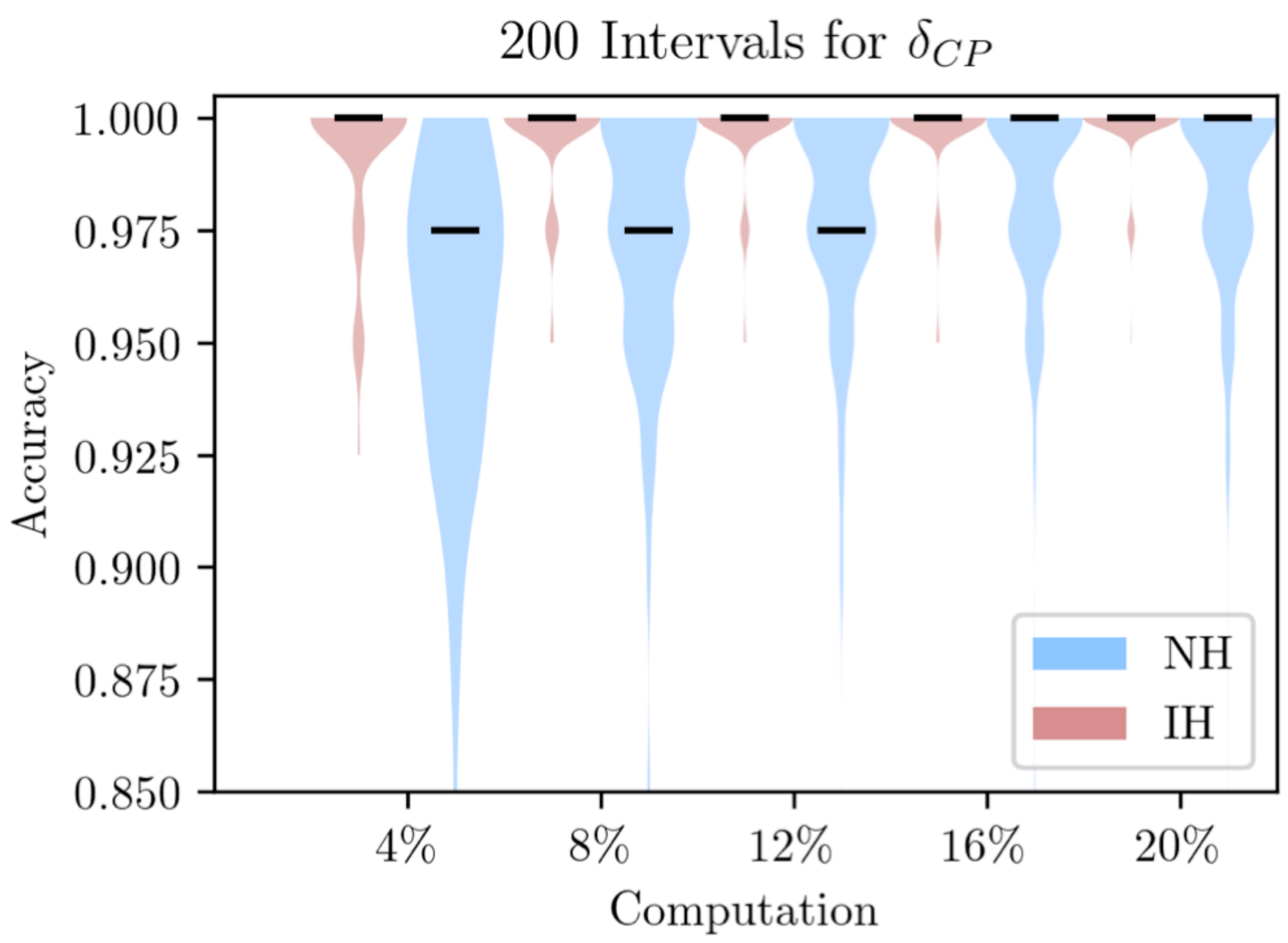
- Oscillation parameters similar to 2018 best estimate from NOvA ($\theta_{23} = 0.56$, $\Delta m^2_{32} = 2.44 * 10^{-3} eV^2$, $\delta_{CP} = 1.5\pi$)
- Significance of rejecting δ_{CP} only after 5 iterations.





Results

- Use classification accuracy of all grid points, taking FC result as truth, to evaluate performance
- Progress shows the search algorithm converges to the FC value ~ $5\times$ faster for 1D case.
- Median Accuracy for 1D is 100% for both NH and IH



Pseudo-Code

Algorithm 1 $\mathcal{GP}$ iterative confidence contour finding

```
for each iteration t = 1, 2, ... do
    Propose new points in parameter space arg max_θ a(θ)
    for each point θ' do
        Simulate likelihood ratio distribution
        for k = 1, 2, ... do
            Perform a pseudo experiment
            Maximize the likelihood with respect to (θ, δ)
            Maximize the likelihood with constraint θ = θ'
        end for
        Obtain critical value c(θ')
    end for
    Update GP approximation ĉ(θ)
    Update confidence contours
end for
```