Feldman-Cousins' ML Cousin

A Simulation-Based Inference Approach to Sterile Neutrino Global Fits

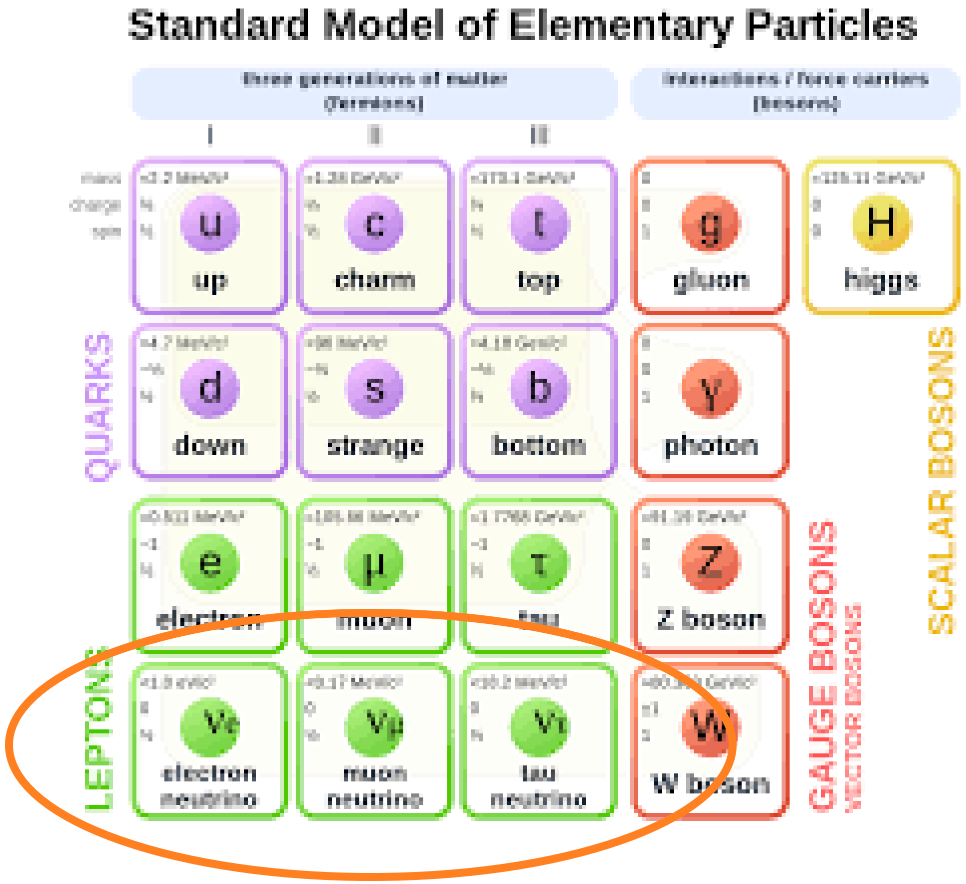
Josh Villarreal (villaj@mit.edu)

PHYSTAT 2024

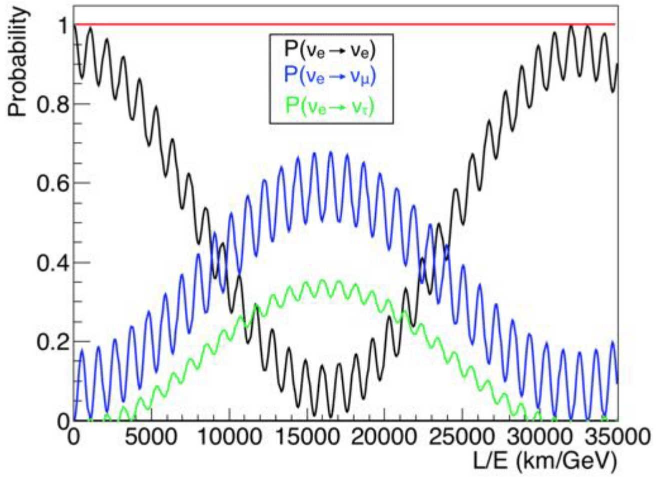
September 10, 2024

Introduction
Neutrinos and their Oscillations

- Standard model neutrinos are known to oscillate
  - Mass eigenstates determine how neutrinos propagate through space and time
  - Flavor eigenstates are determined by neutrinos' charged-current weak interactions
  - Parameterized by PMNS matrix
- Important consequence: (some) neutrinos have mass?



DOI: 10.13140/RG.2.2.19649.56169

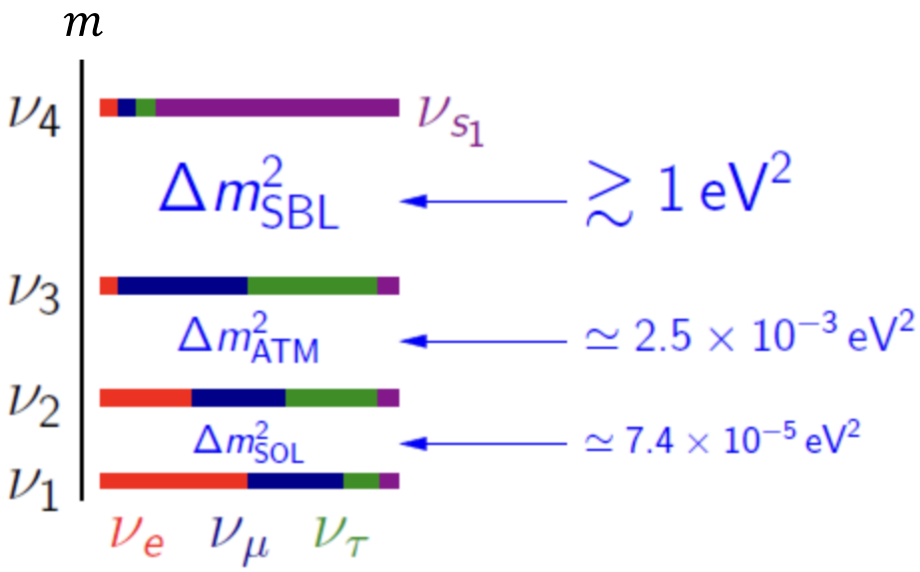


$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}$$

$$P_{\alpha\to\beta} = \delta_{\alpha\beta} - 4\sum_{j>k} \mathrm{Re}(U_{\beta j}U^*_{\alpha j}U^*_{\beta k}U_{\alpha k})\sin^2(\Delta m^2_{jk}L/4E) + 2\sum_{j>k}\mathrm{Im}(U_{\beta j}U^*_{\alpha j}U^*_{\beta k}U_{\alpha k})\sin(\Delta m^2_{jk}L/2E)$$

Introduction

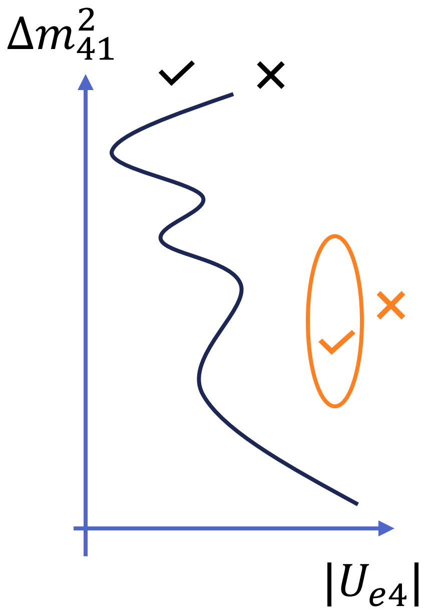
Arena: Sterile Neutrino Searches via ν_e Disappearance



- If (LH) massive neutrinos exist, then (RH) neutrinos:
    - Do not participate in the weak interaction because of $V - A$. ”Sterile”
    - May be accessible by oscillations.
- Simplest model which permits sterile neutrinos is 3+1. E.g.,

$$p(\nu_e \to \nu_e) = 1 - 4\,|U_{e4}|^2\,(1 - |U_{e4}|^2)\,\sin^2(1.27\,\Delta m^2_{41}\,L/E)$$

- As a test-bed, consider the class of experiments searching for a 3+1 sterile by electron-(anti)neutrino disappearance.

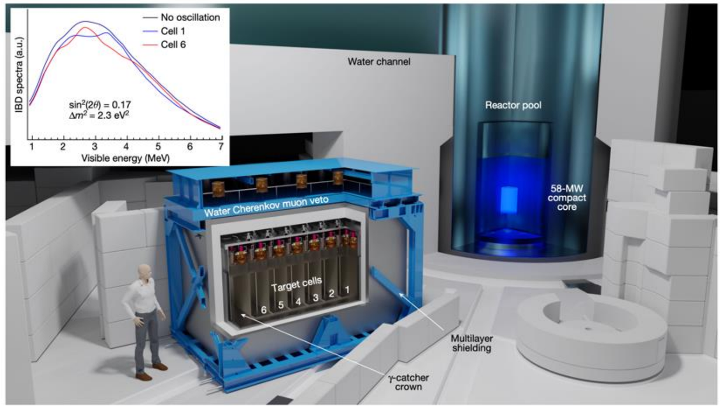


Introduction

Arena: Sterile Neutrino Searches via ν_e Disappearance

Reactor Experiments

- Scintillator detectors + nuclear reactors; measure $\overline{\nu}_e$ disappearance
- "Shape-only fit" – we correct for the flux normalization mismatch and measure oscillations directly
  - Can measure nonzero U_{e4} and $\Delta m_{41}^2 \lesssim 10\ \mathrm{eV}^2$
  - Most shape-only fits do not favor a sterile neutrino
- STEREO (below), PROSPECT, NEOS, DANSS



https://doi.org/10.1038/s41586-022-05568-2

Source Experiments

- MCi sources (e.g. ^{37}Ar and ^{51}Cr) + gallium targets
- Detectors capture neutrinos via $\nu_e\ (^{71}\mathrm{Ga},\ ^{71}\mathrm{Ge})\ e^-$; germanium atoms are periodically counted
- BEST (below), SAGE, and GALLEX each observe deficits compared to expectations, called the "gallium anomaly"

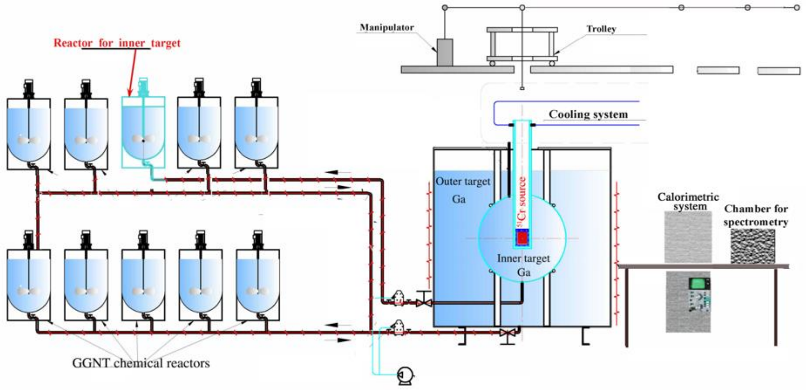


FIG. 1. The Ga target and extraction piping diagram also indicating the source handling apparatus.

https://arxiv.org/abs/2109.11482

Introduction
Wilks' Theorem and its Problems

- Conventionally:
  - Model parameters y are estimated by maximizing a likelihood function $L_x(y)$, with x observed data. That is, $\hat{y} = \mathrm{argmax}_y\, L_x(y)$.
  - One can devise a likelihood-ratio test statistic $\lambda = 2[\log L_x(\hat{y}) - \log L_x(y_0)]$ comparing the maximum likelihood to the likelihood under a null (no-oscillation) hypothesis.
  - Wilks' theorem (under some assumptions) states $\lambda \sim \chi^2_k$ with $k = \dim x$, allowing us to compute and interpret significances.
- There are two problems with sterile searches and Wilks':
  1. The null model ($|U_{e4}| = 0$) lies on the boundary of the parameter space.
  2. The oscillation frequency parameter (Δm^2_{41}) can independently scan over many local minima, with flexibility unaccounted for.
- Assuming Wilks' theorem in 3+1 fits can cause you to misinterpret the significance of your results.

Nat. Rev. Phys., 2, 245–252 (2020)

Box 2 | **Necessary conditions for Wilks' theorem**

Asymptotic
Sufficient data are collected.

Interior
Only values of the parameters of interest μ and nuisance parameters θ that are not on the boundaries of their parameter space are admitted.

Identifiable
Different values of the parameters specify distinct models.

Nested
The null hypothesis H_0 is a limiting case of the general case hypothesis H_1, for example, with some parameter constrained to a subrange of the entire parameter space.

Correct
The true model is specified either under H_0 or under H_1.

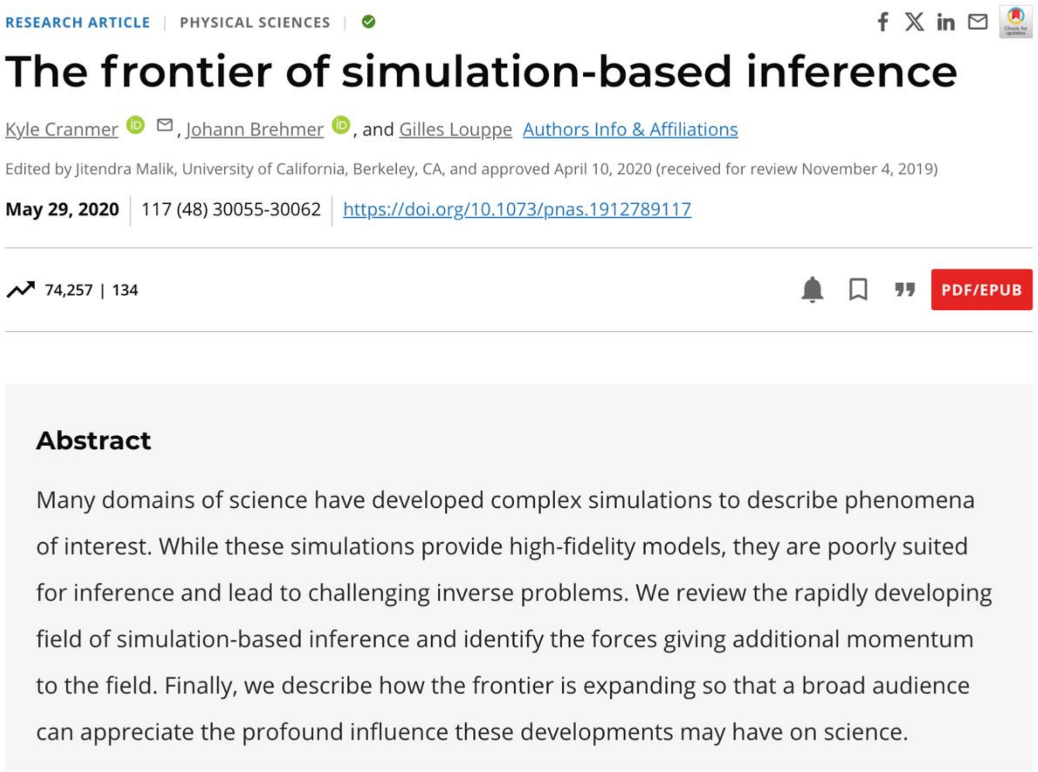
[For an excellent pedagogical overview, see 2211.06347]

Introduction
Feldman-Cousins

- Since we can't use Wilks' to do a 3+1 fit, we have to more carefully describe the distribution of our LLR test statistic λ. This is the job of the Feldman-Cousins method:
    1. Given experimental data, compute likelihood function $L_x(y) = p(x|y)$ for (all) values of the oscillation parameters $y = (U_{e4}, \Delta m^2_{41})$
    2. Given experimental data, compute value of a test statistic (e.g., $\lambda = 2[\log L_x(\hat{y}) - \log L_x(y_0)]$) for (all) values of the oscillation parameters
    3. Order points in oscillation parameter space by most to least desirable value of test statistic
    4. Add points in oscillation parameter space to the confidence region until desired the confidence is reached (computed from the likelihood $L_x(y)$)
- Major problems:
    - Computing the likelihood is often time-consuming
    - Computing the test statistic is often time-consuming
    - Doing this on a fine enough grid only complicates things further

Introduction
Simulation-Based Inference

- In the end, to describe the a best-fit point and its significance, all you need is an object like a
  - Posterior distribution $p(y|x)$, significance quantified by a *credibility region*
  - Likelihood $p(x|y)$, significance quantified by a *confidence interval*
- Advances in the field of density estimation allow us to estimate these quantities directly through machine learning, in place of (or enhancing) high-fidelity fitting procedures like MCMC to generate posterior
  - Training data acquired through simulation


RESEARCH ARTICLE | PHYSICAL SCIENCES

The frontier of simulation-based inference

Kyle Cranmer, Johann Brehmer, and Gilles Louppe Authors Info & Affiliations

Edited by Jitendra Malik, University of California, Berkeley, CA, and approved April 10, 2020 (received for review November 4, 2019)

May 29, 2020 | 117 (48) 30055-30062 | https://doi.org/10.1073/pnas.1912789117

74,257 | 134

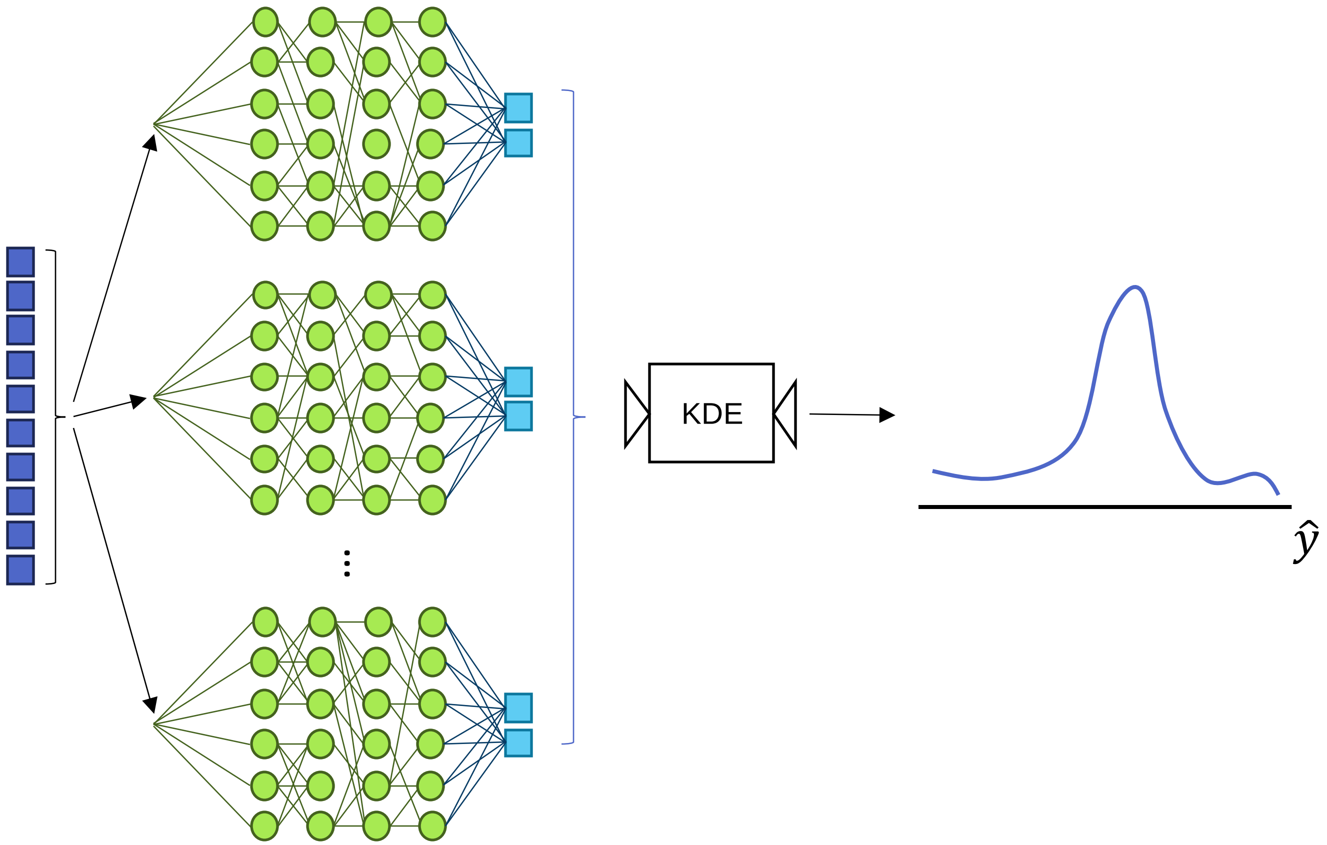
PDF/EPUB

Abstract

Many domains of science have developed complex simulations to describe phenomena of interest. While these simulations provide high-fidelity models, they are poorly suited for inference and lead to challenging inverse problems. We review the rapidly developing field of simulation-based inference and identify the forces giving additional momentum to the field. Finally, we describe how the frontier is expanding so that a broad audience can appreciate the profound influence these developments may have on science.

Building an ML-Based (Global) Fitter
Naïve Approach

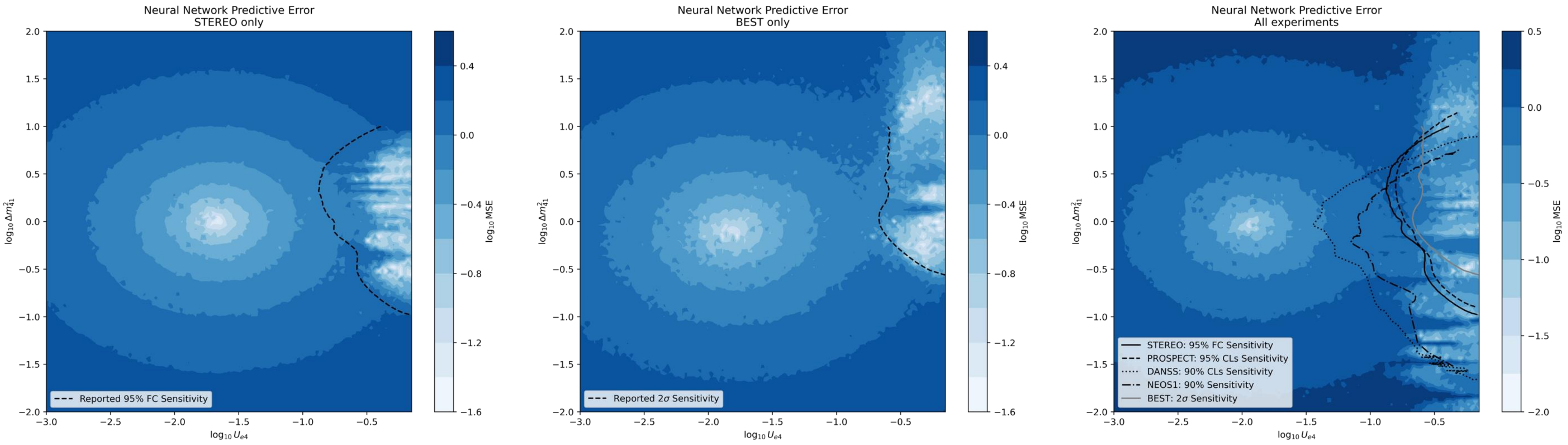
- Gal and Ghahramani (2015) show that the dropout procedure, traditionally used for regularization of neural networks, can be repurposed to approximate predictive uncertainties. The idea: *Train a neural network with dropout before each hidden layer, and leave dropout on at inference time.*
  - Each time a prediction is made from the neural network, a different subnetwork is randomly and independently chosen.
  - Effectively asking for predictions from an ensemble of smaller neural networks.
  - Obtain prediction PDF by fitting a KDE to many prediction samples.



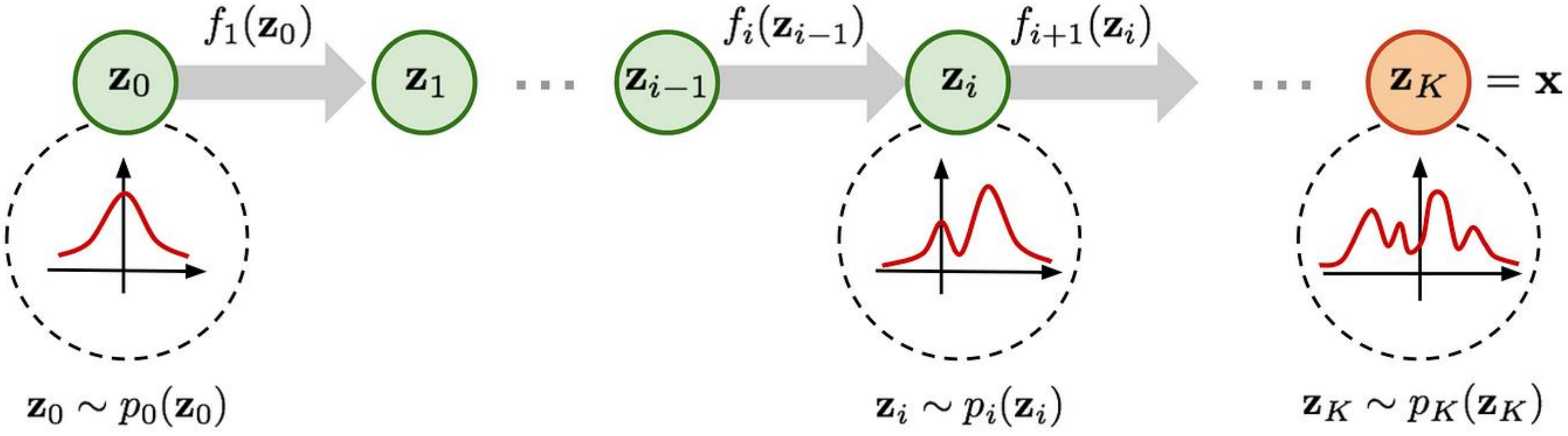
Building an ML-Based (Global) Fitter

Mode Collapse

- Misleading structures emerge in the predictive accuracies of these effective network ensembles.
  - If the NNs are shown MC data thrown from a physical model to which the experiment(s) is (are) insensitive, the NN just guesses the $R^2 = 0$ choice (an average).
- Offers a natural "embedding"



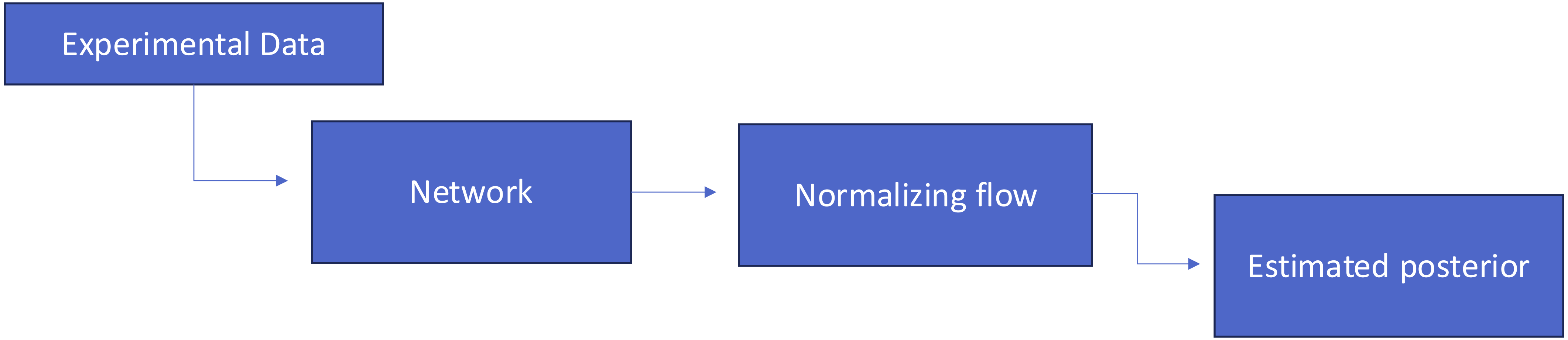
Building an ML-Based (Global) Fitter
Conditional Density Estimation with Normalizing Flows



https://ankurdhuriya.medium.com/what-are-normalizing-flows-ce7ccd222ee7

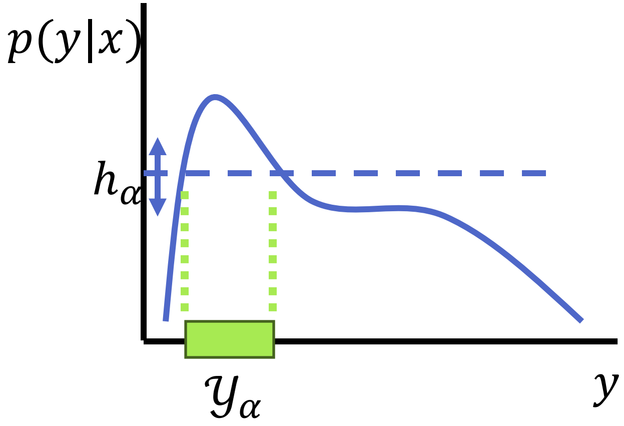
- Normalizing flows learn a sequence of invertible transforms from a base distribution to some arbitrary distribution
- Frequently used for generative modeling, but have obvious applications to density estimation.
- Use NFs to ”correct” neural network predictions to compensate for mode collapse

Building an ML-Based (Global) Fitter
The FCMLC Procedure



- With a posterior, can find 2D credibility region $\mathcal{Y}_\alpha = \{y: p(y|x) > h_\alpha\}$ satisfying

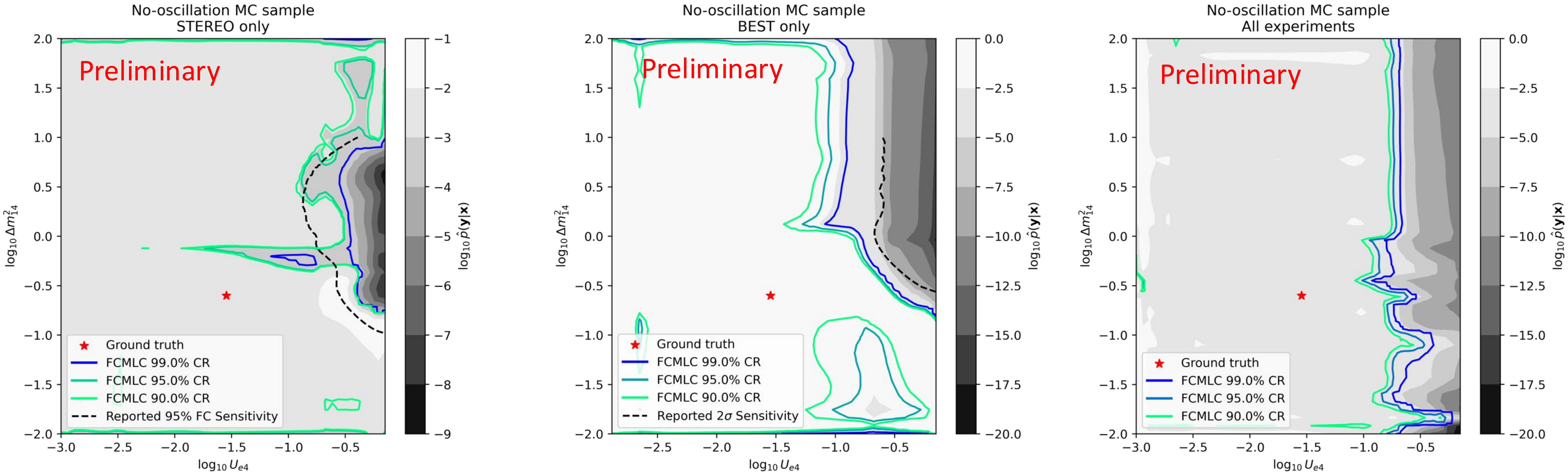
$$\int_{y \in \mathcal{Y}_\alpha} p(y|x)\, dy = \alpha$$



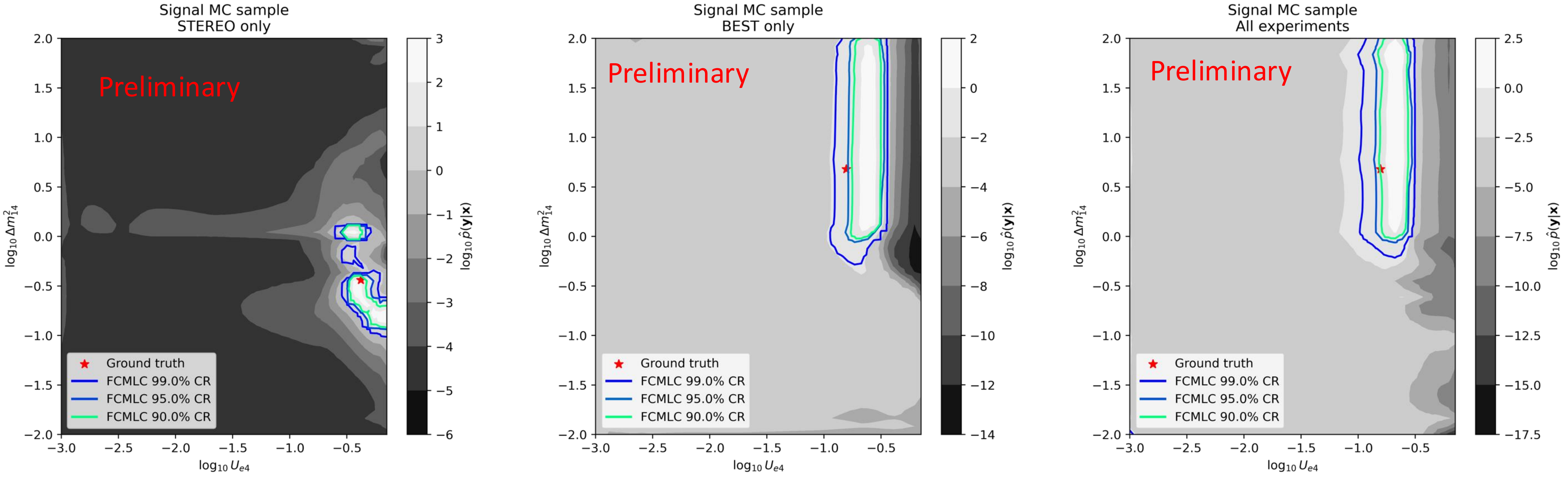
Results

Fits on Simulated Experimental Data

- Sensitivities (dashed black lines) come from conventional fits



Results
Fits on Simulated Experimental Data



Discussion
FCMLC Pros and Cons

- Pro: The SBI approach is much faster than traditional fitting methods.
  - With a NN and a NF trained, posterior density estimation takes ~5 minutes and can run in a Jupyter notebook.
  - Together with generation of MC data, training, and FCMLC evaluation, ~2.5K CPU hours
  - By comparison, Feldman-Cousins takes ~510K CPU for the experiments considered in this work.
- Pro: Runtime of FCMLC is ~constant with respect to the inclusion of more experiments.
- Pro: SBI does not assume Gaussian uncertainties, and it is straightforward to set up simulations to capture this.
- Con: Hyperparameter optimization can further extend the overhead of FCMLC.
- Con: FCMLC is sensitive to the way in which you generate your MC data.
- Con: FCMLC's posterior density and Feldman-Cousins answer slightly different statistical questions.

Summary

Conclusions and Future Directions

- SBI fits nicely in the realm of particle physics global fits, alleviating some of the computational barriers currently in place.
- These methods are plug-and-play; pick your favorite density estimator and try it out. As the field of deep-learning based methods for density estimation develops, SBI will mature.
- Using a technique like FCMLC may be able to seed higher-fidelity fitting frameworks.
- Simulation-based likelihood / likelihood ratio estimation may be able to super-charge Feldman-Cousins.

Backup

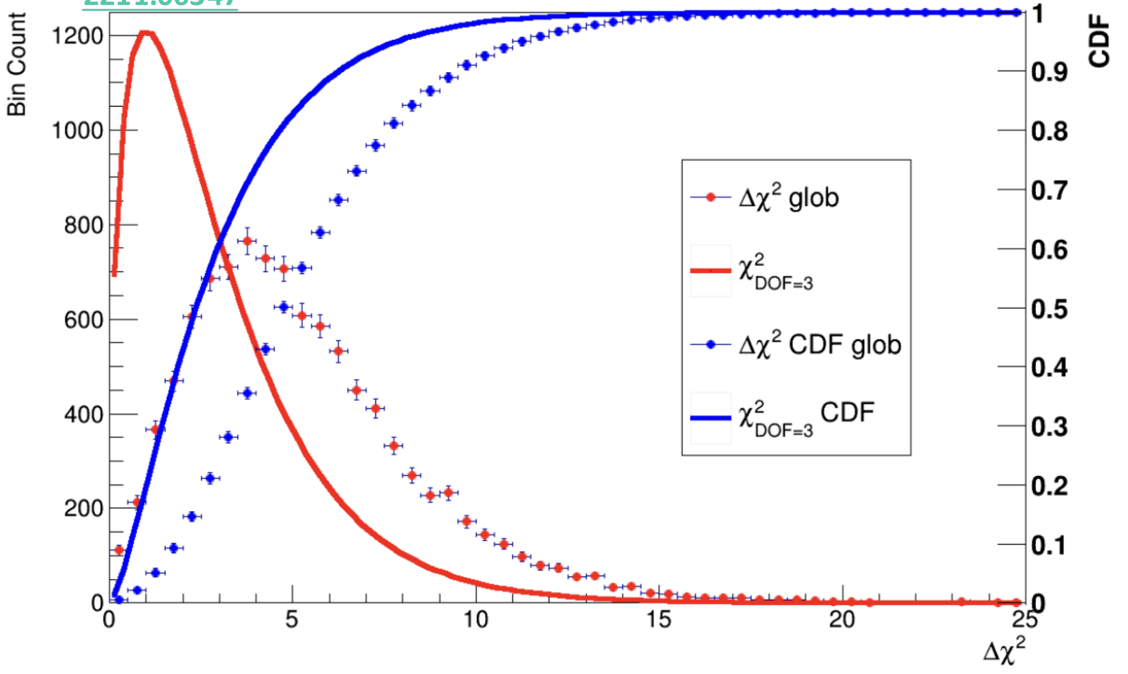
The Wilks' Trap

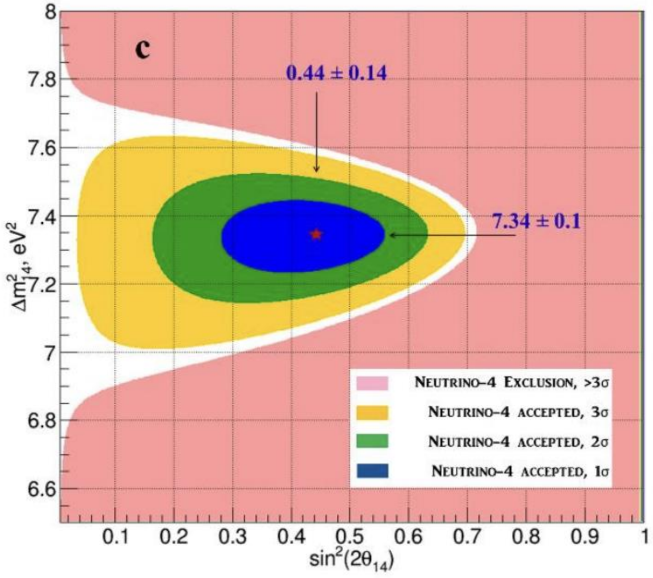
The first observation of effect of oscillation in Neutrino-4 experiment on search for sterile neutrino (continuation)

A.P. Serebrov[1], R.M. Samoilov[1], V.G. Ivochkin[1], A.K. Fomin[1], A.O. Polyushkin[1], V.G. Zinoviev[1], P.V. Neustroev[1], V.L. Golovtsov[1], A.V. Chernyj[1], O.M. Zherebtsov[1], M.E. Chaikovskii[1], M.E. Zaytsev[1, 3], A.A.Gerasimov[1], A.L. Petelin[2], A.L. Izhutov[2], A.A.Tuzov[2], S.A. Sazontov[2], M.O. Gromov[2], V.V. Afanasiev[2],

1. *NRC "KI" Petersburg Nuclear Physics Institute, Gatchina, Russia*
2. *JSC "SSC Research Institute of Atomic Reactors", Dimitrovgrad, Russia*
3. *Dimitrovgrad Engineering and Technological Institute MEPhI, Dimitrovgrad, Russia*

2211.06347





Neutrino-4 anomaly: Oscillations or fluctuations?

C. Giunti [a], Y.F. Li [b,c], C.A. Ternes [a], Y.Y. Zhang [b,c,*]

[a] *Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[b] *Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China*
[c] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

"Using a more reliable Monte Carlo simulation of a large set of Neutrino-4-like data, we found that the statistical significance of the Neutrino-4 short-baseline neutrino oscillation signal decreases to about 2.2σ."