Statistical Significance of Mass Hierarchy in Future Experiments

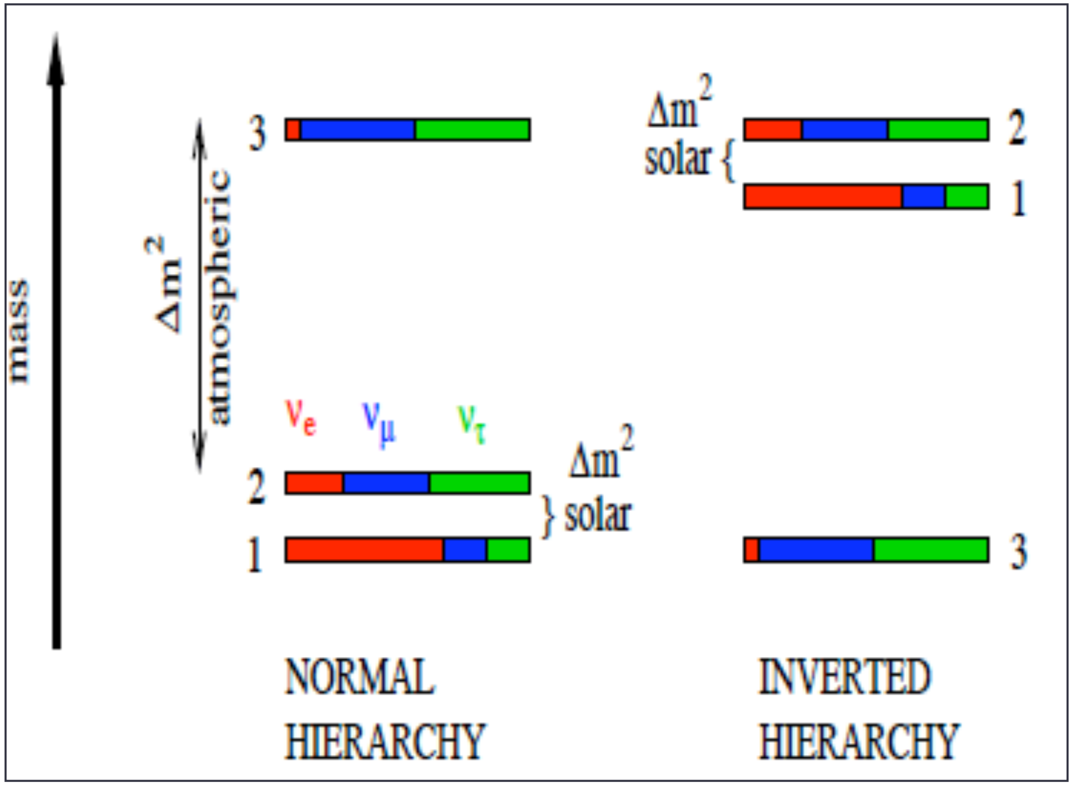
Sanjib Kumar Agarwalla

sanjib@iopb.res.in

Institute of Physics, Bhubaneswar, India

Neutrino Mass Hierarchy: Important Open Question

- The sign of Δm^2_{31} $(m_3^2 - m_1^2)$ is not known



Neutrino mass spectrum can be normal or inverted hierarchical

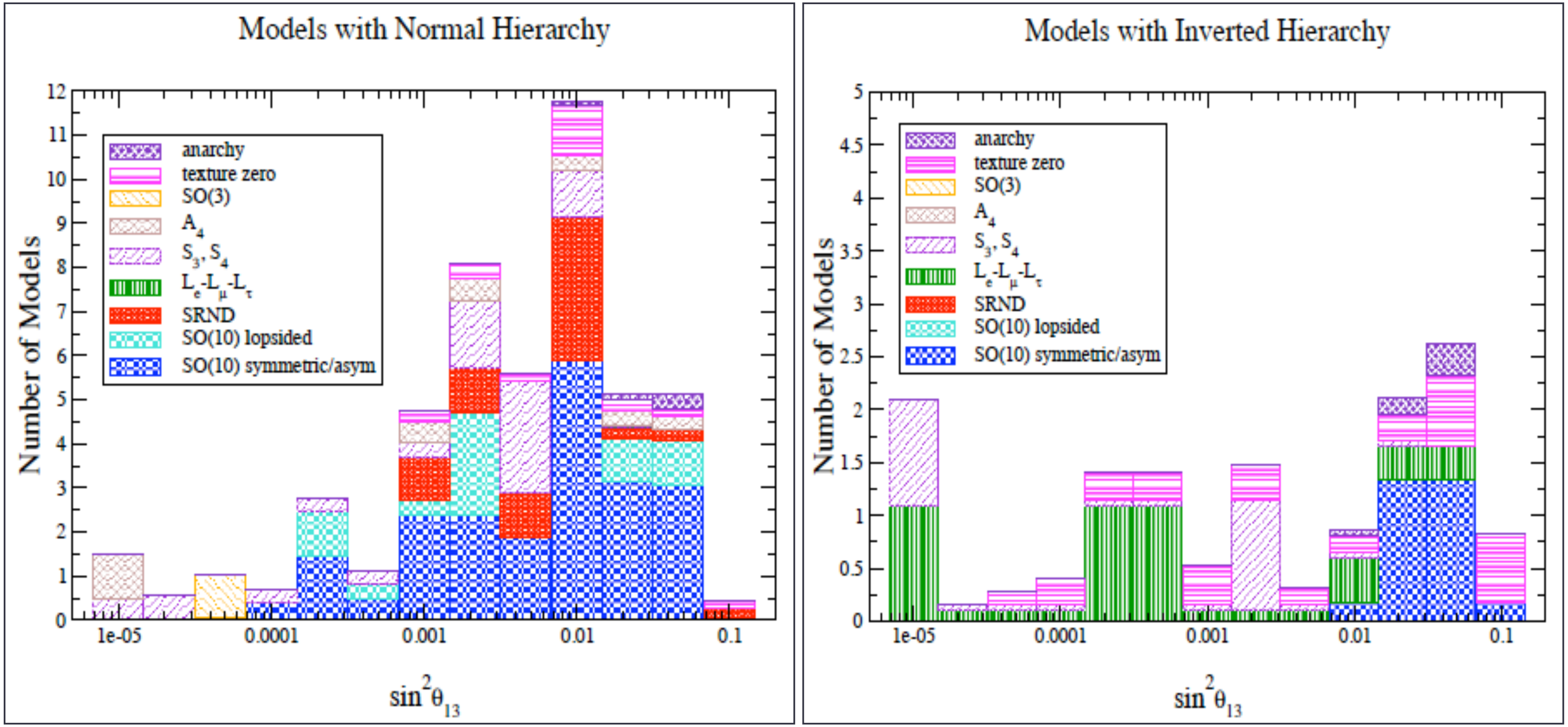
We only have a lower bound on the mass of the heaviest neutrino

$$\sqrt{2.5 \cdot 10^{-3} \mathrm{eV}^2} \sim 0.05 \ \mathrm{eV}$$

We currently do not know which neutrino is the heaviest

Mass Hierarchy Discrimination : A Binary yes-or-no type question

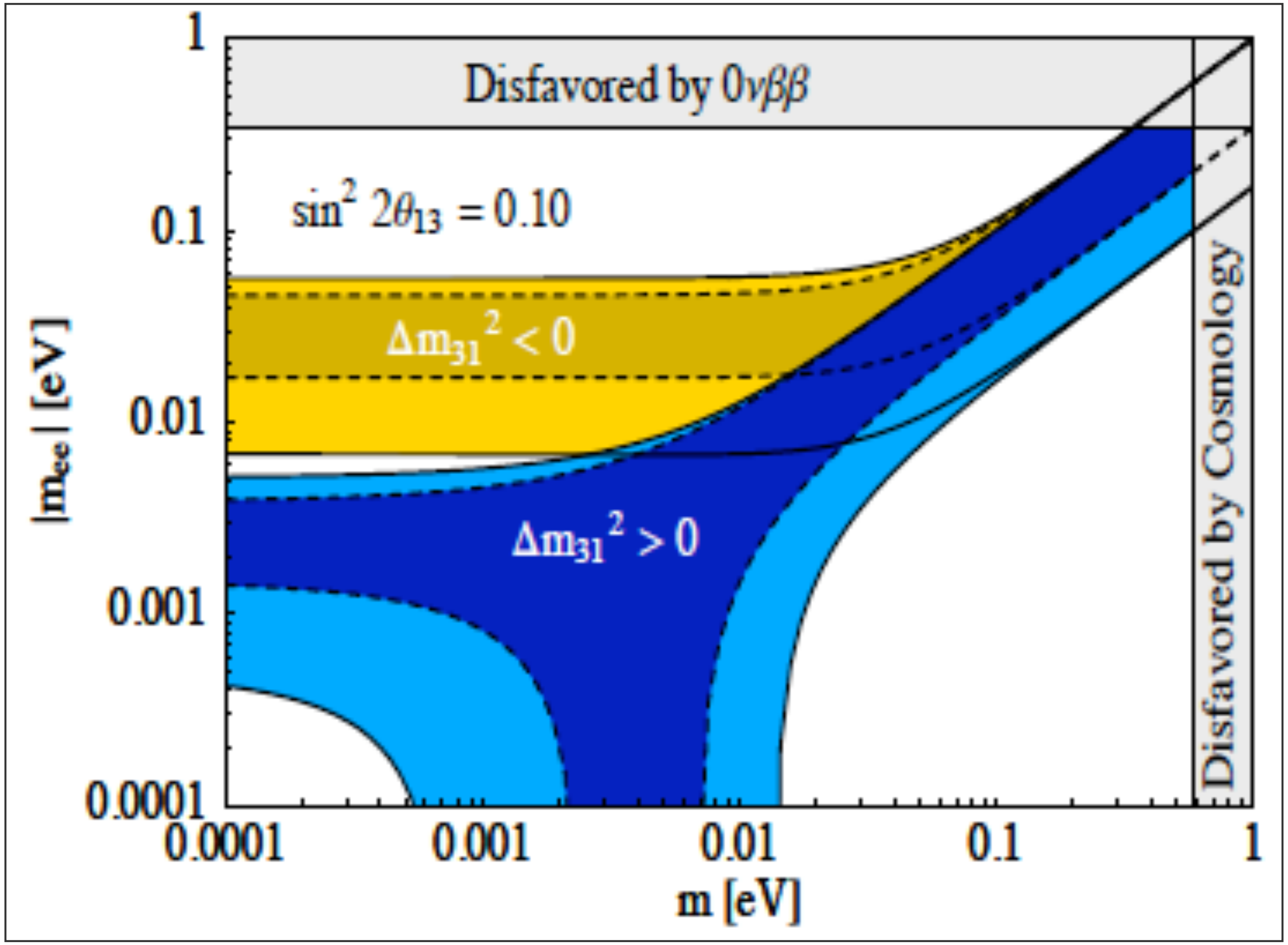
Why do we care about Neutrino Mass Ordering?



Albright and Chen, hep-ph/0608137

- Dictates the structure of neutrino mass matrix
- Essential for the underlying theory of neutrino masses and mixing
- Acts as a powerful discriminator between various neutrino mass models

Connection between 0νββ and Neutrino Mass Ordering



Lindner, Merle, Rodejohann , hep-ph/0512143

If hierarchy is inverted, and yet no 0νββ is observed in the very far future, strong hint that neutrinos are not Majorana particles

Important References

- **Statistical Evaluation of Experimental Determinations of Neutrino Mass Hierarchy**
  X. Qian, A. Tan, W. Wang, J. Ling, R. McKeown
  Phys.Rev. D86, 113011 (2012), arXiv: 1210.3651

- **Determination of mass hierarchy with medium baseline reactor neutrino experiments**
  S.-F. Ge, K. Hagiwara, N. Okamura, Y. Takaesu
  JHEP 1305, 131 (2013), arXiv: 1210.8141

- **Confidence in a Neutrino Mass Hierarchy Determination**
  E. Ciuffoli, J. Evslin, X. Zhang (2013)
  arXiv: 1305.5150

- **Neutrino mass hierarchy and electron neutrino oscillation parameters with one hundred thousand reactor events**
  F. Capozzi, E. Lisi, A. Marrone (2013)
  arXiv: 1309.1638

- **Quantifying the sensitivity of oscillation experiments to the neutrino mass ordering**
  M. Blennow, P. Coloma, P. Huber, T. Schwetz
  arXiv: 1311.1822v2

- **On the Bayesian approach to neutrino mass ordering**
  M. Blennow (2013)
  arXiv: 1311.3183

Statistical Issues Regarding Mass Hierarchy Discovery

- **In the Mass Hierarchy Determination: Only two possible results are considered (NH or IH)**

- **There are two separate but related questions:**

  a) Given real experimental data, with what significance can the MH be distinguished?

  b) When evaluating future experimental sensitivities, what is the probability that a particular experimental design will be able to determine the MH with a given significance?

- **Once data are in hand, a number of techniques based either within Bayesian or Frequentist make it possible to determine the level of confidence at which one MH hypothesis or the other can be ruled out.**

Neutrino community is traditionally frequentist and more accustomed to interpreting frequentist results

- **For future experiments, common practice to generate a simulated data set (for an assumed true MH) that does not include statistical fluctuations**

- **The expected sensitivity can be reported as $\overline{\Delta\chi^2}$, representative of the mean or the most likely value of $\Delta\chi^2$ that would be obtained in an ensemble of experiments for a particular true MH**

Statistical Issues in Mass Hierarchy Discovery

- In estimating the MH sensitivity for future experiments, we need to consider statistical fluctuations and variations in systematics

- If we repeat the simulations many times, a distribution of $\Delta\chi^2$ values will appear

- References (arXiv: 1210.3651 and 1311.1822) showed that the $\Delta\chi^2$ metric employed here does not follow the commonly expected χ^2 function for one degree of freedom, which has a mean of $\overline{\Delta\chi^2}$ and can be interpreted using a Gaussian distribution with a standard deviation of $\sqrt{|\overline{\Delta\chi^2}|}$

- They showed that when the observed counts in the experiment are large enough, the distribution of $\Delta\chi^2$ approximately follows a Gussian distribution with a mean and standard deviation of $\overline{\Delta\chi^2}$ and $2\sqrt{|\overline{\Delta\chi^2}|}$ respectively

Test Statistics

- A common test statistic is the χ^2 with n degrees of freedom, which describes the deviation from the expected values of the outcome of a series of measurements x_i of the normal distribution $\mathcal{N}(\mu_i, \sigma_i)$:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}$$

- The further the observations are from the expected values, i.e., the more extreme the outcome, the larger is the χ^2

- If the mean values μ_i depends on a set of p parameters θ whose values have to be estimated from the data, one usually considers the minimum of the χ^2 with respect to the parameters:

$$\chi^2_{\text{min}} = \min_{\theta} \chi^2(\theta)$$

- According to Wilk's theorem, this quantity will follow a χ^2 distribution with n-p degrees of freedom, whereas

$$\Delta\chi^2(\theta) = \chi^2(\theta) - \chi^2_{\text{min}}$$

  will have a χ^2 distribution with p degrees of freedom

- Use the data set without statistical fluctuations (Asimov data) and it works for nested hypotheses (for an example: probability to observe a non-zero θ_{13})

Test Statistics for Mass Ordering

- **One needs to select a test statistic which is well suited to distinguish between the two hypotheses H_{NO} and H_{IO}**

- **Mass Ordering is not nested, Wilk's theorem is not applicable.**
  Usual techniques relying on a Taylor expansion around a single maximum of the likelihood is not applicable in this case

- **We concentrate on the following test statistic, which is based on a log-likelihood ratio**

$$T = \min_{\theta \in \mathrm{IO}} \chi^2(\theta) - \min_{\theta \in \mathrm{NO}} \chi^2(\theta) \equiv \chi^2_{\mathrm{IO}} - \chi^2_{\mathrm{NO}}$$

Here θ is the set of neutrino oscillation parameters which are confined to a given mass ordering during the minimization

- **A crucial point in evaluating a statistical test is to know the distribution of test statistics.**
  In general this has to be estimated by explicit Monte Carlo simulations

- **Under certain conditions, the distribution of T can be derived analytically and corresponds to a normal distribution (the Gaussian case):**

$$T = \mathcal{N}(\pm T_0, 2\sqrt{T_0})$$ **[+ (-) sign holds for true NO (IO)]**

T_0 = value for Asimov data set without any statistical fluctuations

Frequentist Methods

- **Hypothesis testing: Test hypothesis (H) and alternative hypothesis (H′)**
  Choose a test statistic T to check whether data can reject the test or null hypothesis H

- **Once the distribution of T is known under the assumption of H being true, we decide to reject H at confidence level (CL) 1 – α** if $T > T_c^\alpha$, where the critical value T_c^α is defined by

$$\int_{T_c^\alpha}^{\infty} p(T|H)dT = \alpha$$

with p(T|H) being the probability distribution function of T given that H is true

- **α is the probability of making an "error of the first kind" i.e., rejecting H although it is true**

- **The conversion between nσ and the value of α (using a double sided Gaussian test):**

$$\alpha(n) = \frac{2}{\sqrt{2\pi}} \int_n^{\infty} dx\, e^{-x^2/2} = \mathrm{erfc}\left(\frac{n}{\sqrt{2}}\right) \quad \Leftrightarrow \quad n = \sqrt{2}\,\mathrm{erfc}^{-1}(\alpha)$$

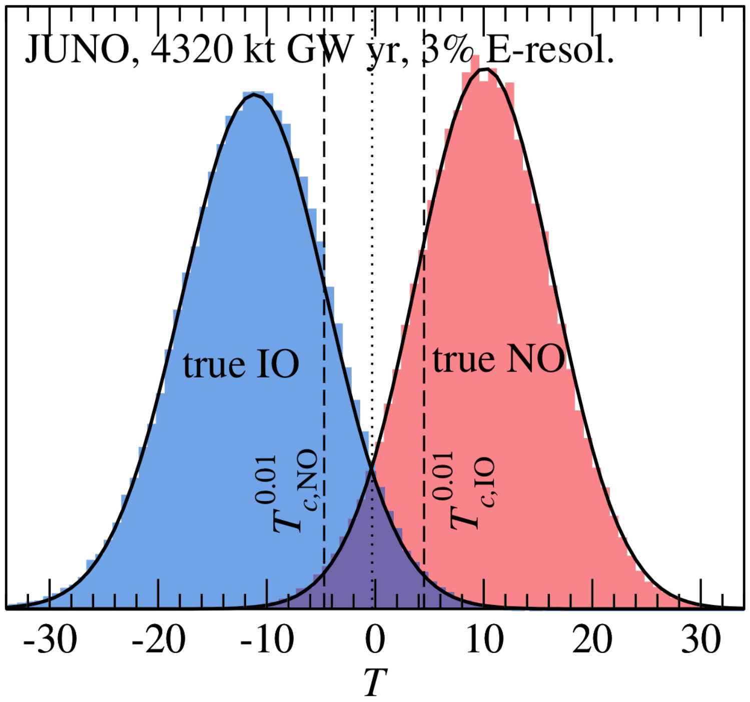
It implies that we identify 1σ, 2σ, 3σ with a CL (1 - α) of 68.27%, 95.45%, 99.73%, respectively

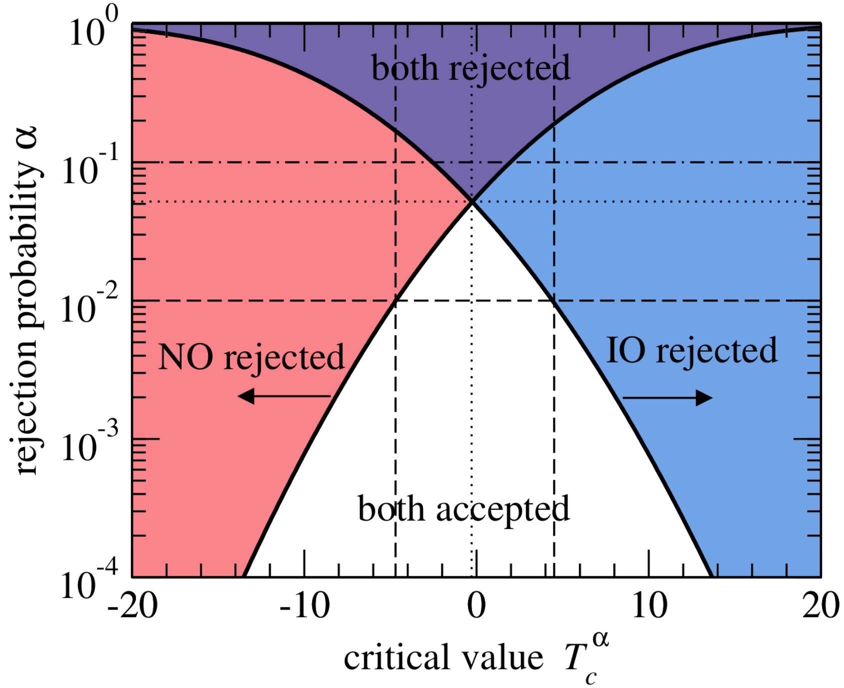
One-sided Gaussian limit: $n_{\text{1-sided}} = \sqrt{2}\,\mathrm{erfc}^{-1}(2\alpha)$ **, CL of 84.14%, 97.73%, 99.87% for 1σ, 2σ, 3σ**

- **β is the probability of making an "error of the second kind" i.e., accepting H although it is not true**

$$\beta = P(T < T_c^\alpha|H') = \int_{-\infty}^{T_c^\alpha} p(T|H')dT$$

Testing both the Mass Hierarchies





Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

Left: Distribution of the test statistics T. Histograms show results of the 10^5 MC simulations. Black curves corresponds to Gaussian approximation

Right: The value of α as a function of the critical value T_c^{α} required for rejecting IO (Blue) and NO (Red). In the purple region both mass ordering are rejected at the CL (1 - α). In the white region both orderings are consistent with data at the CL (1 – α)

Curves for testing the different orderings cross around α = 5.2%, indicated by the dotted lines. It represents the unique CL for which the experiment will rule out exactly one of the orderings

Median Sensitivity

- **The most common performance indicator used for the normal mass ordering determination is:**

$$T_0^{\rm NO}(\theta_0) = \min_{\theta\in{\rm IO}} \sum_i \frac{[\mu_i^{\rm NO}(\theta_0) - \mu_i^{\rm IO}(\theta)]^2}{\sigma_i^2}$$

- **The data x_i are replaced by the predicted observables $\mu_i(\theta_0)$ at true parameter values θ_0**
  No statistical fluctuations included, it is representative for an "average" experiment

- **T_0 is then evaluated assuming a χ^2 distribution with 1 dof in order to quote a CL with which a given mass ordering can be identified: this is known as "Standard Sensitivity"**

- **To define an average experiment, one possibility is to calculate the CL (1 – α) at which a false hypothesis can be rejected with a probability of 50%, i.e., β = 0.5**

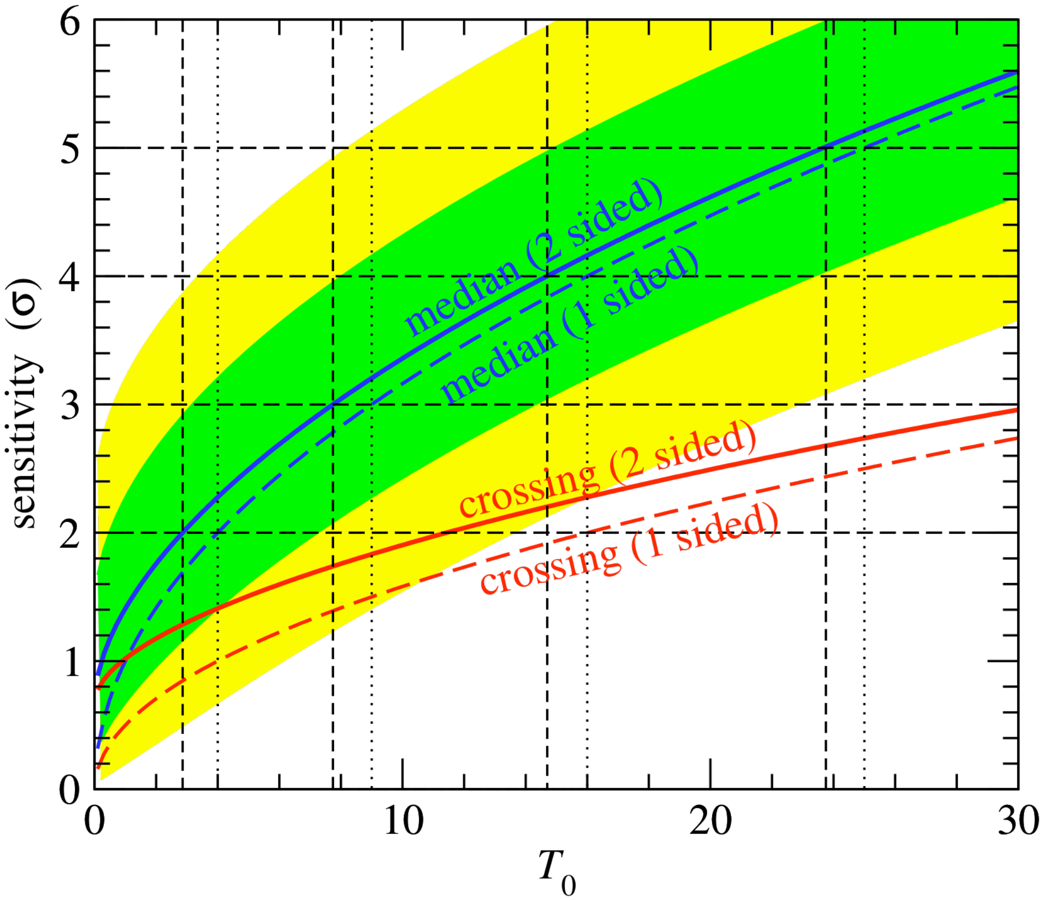
The probability α (β = 0.5) is called the "median" sensitivity:

$$n = \sqrt{2}\,{\rm erfc}^{-1}\left[\frac{1}{2}\,{\rm erfc}\left(\sqrt{\frac{T_0}{2}}\right)\right] \qquad \text{(median sensitivity)}$$

The sensitivity α for which the critical values are the same for both orderings:

$$\alpha = \frac{1}{2}\,{\rm erfc}\left(\frac{T_0^{\rm NO} + T_0^{\rm IO}}{\sqrt{8T_0^{\rm NO}} + \sqrt{8T_0^{\rm IO}}}\right) \approx \frac{1}{2}\,{\rm erfc}\left(\frac{1}{2}\sqrt{\frac{T_0}{2}}\right) \qquad (T_c^{\rm NO} = T_c^{\rm IO})$$

Median Sensitivity for Simple Hypotheses



Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

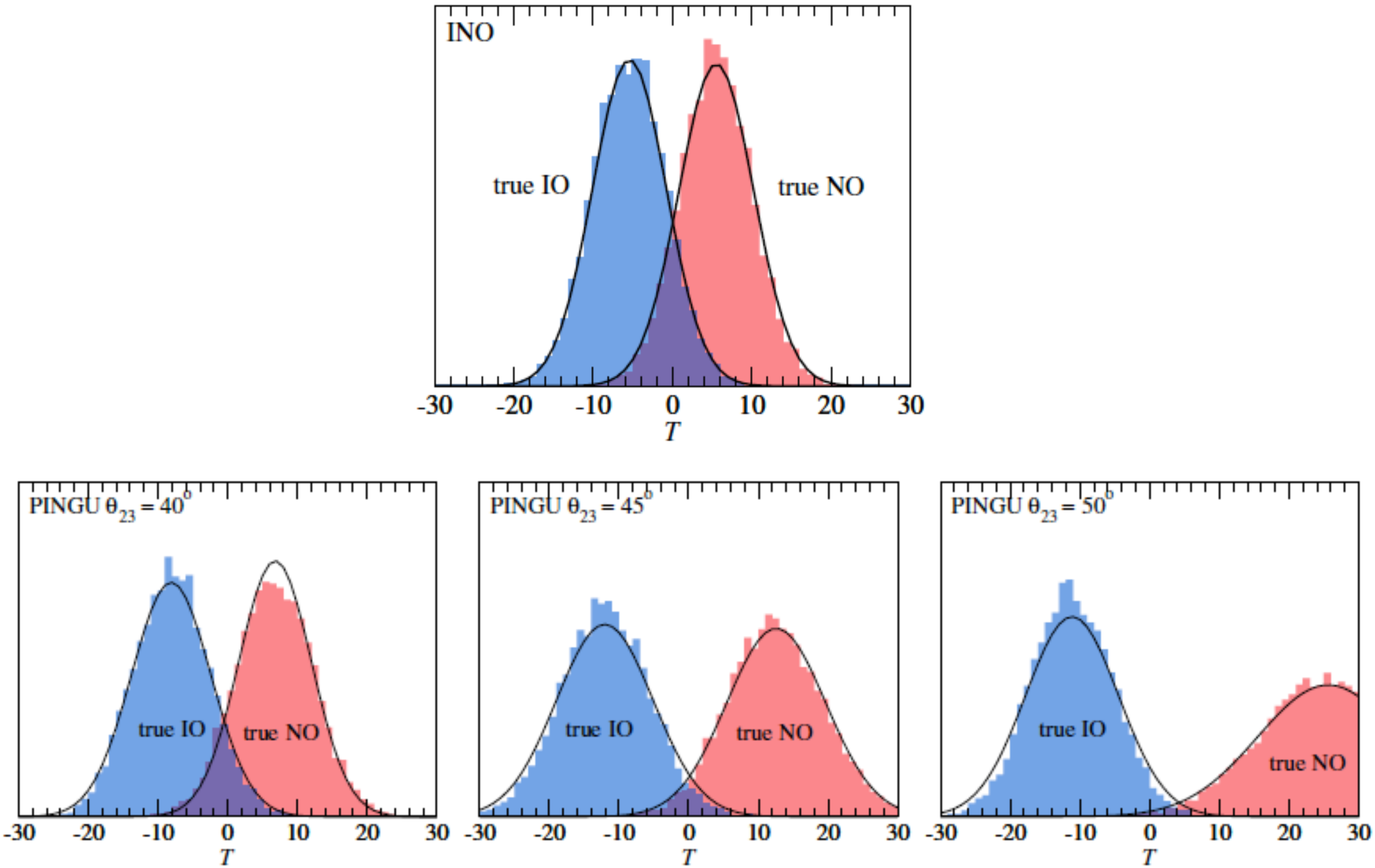
- Hypotheses are not parameter dependent
- Neyman Pearson lemma is applicable which means that the test based on the likelihood ratio is most powerful
- Applicable to reactor experiments

Median sensitivity ($\beta = 0.5$) as a function of T_0. The curves labeled 'crossing' show the sensitivity corresponding to the condition $T_c^{NO} = T_c^{IO}$

Standard sensitivity: $n = \sqrt{T_0}$ and crossing sensitivity: $n = \sqrt{T_0}/2$

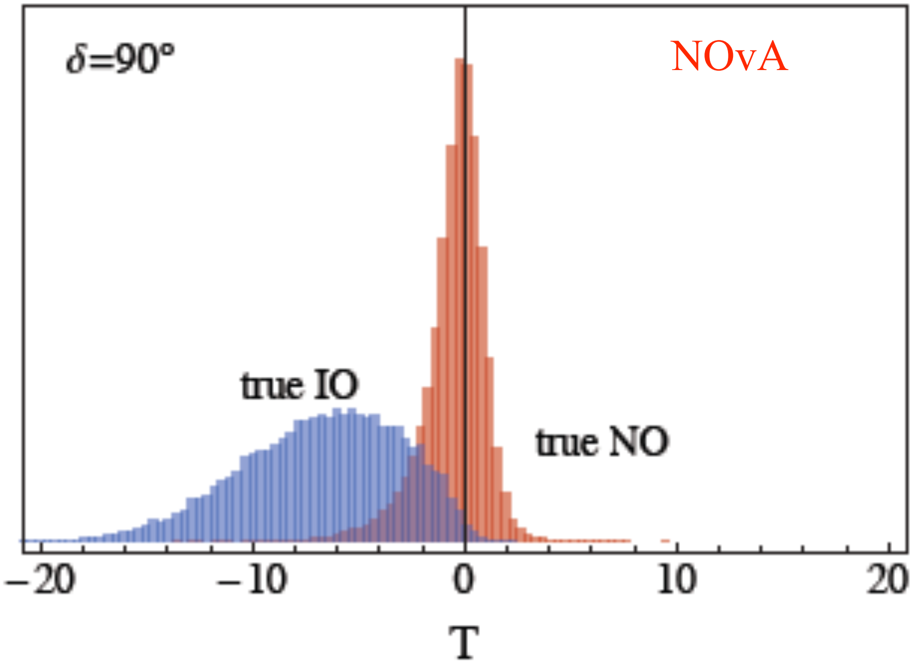
Green bands: $\beta = 1/2 \pm 0.6827/2$ Yellow bands: $\beta = 1/2 \pm 0.9545/2$

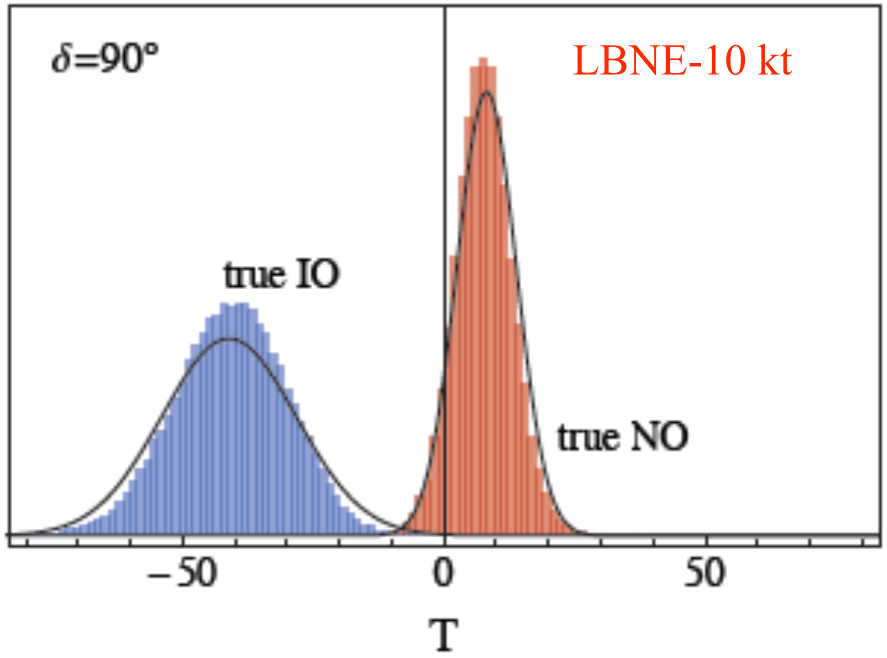
Atmospheric Experiments



Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

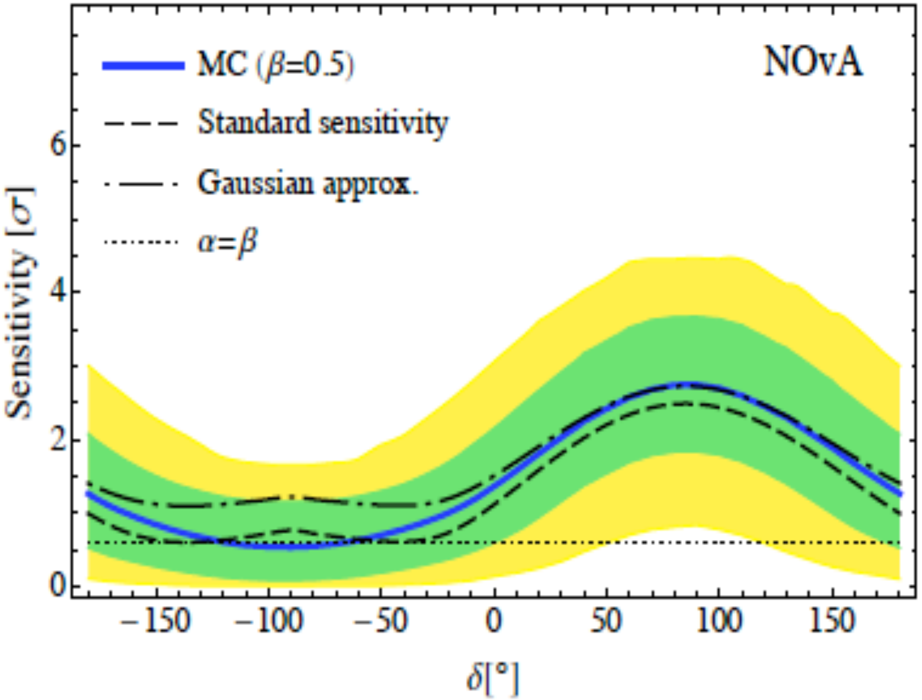
Accelerator Experiments

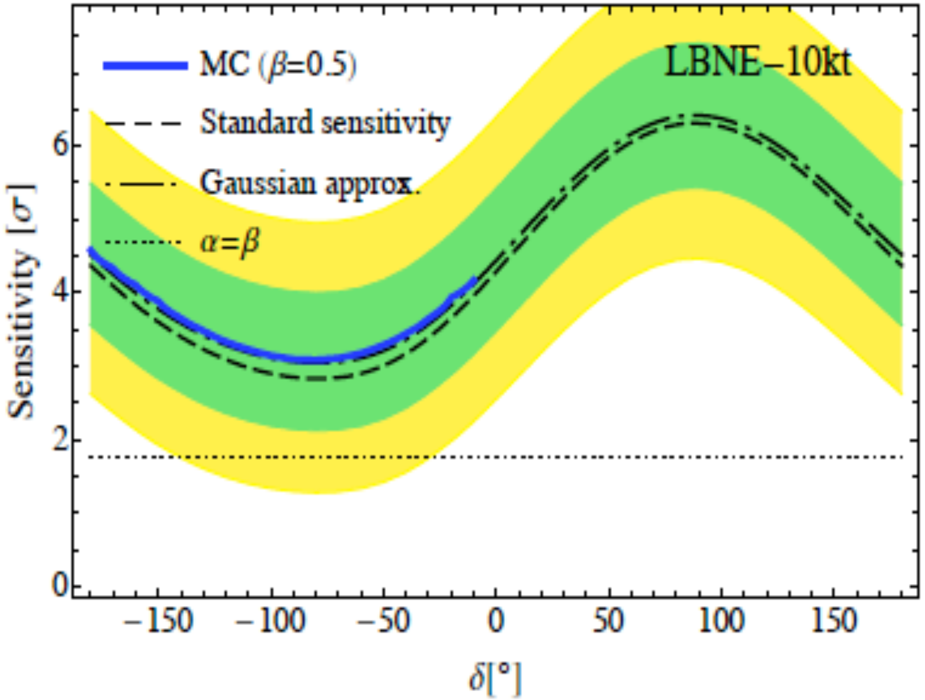




Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

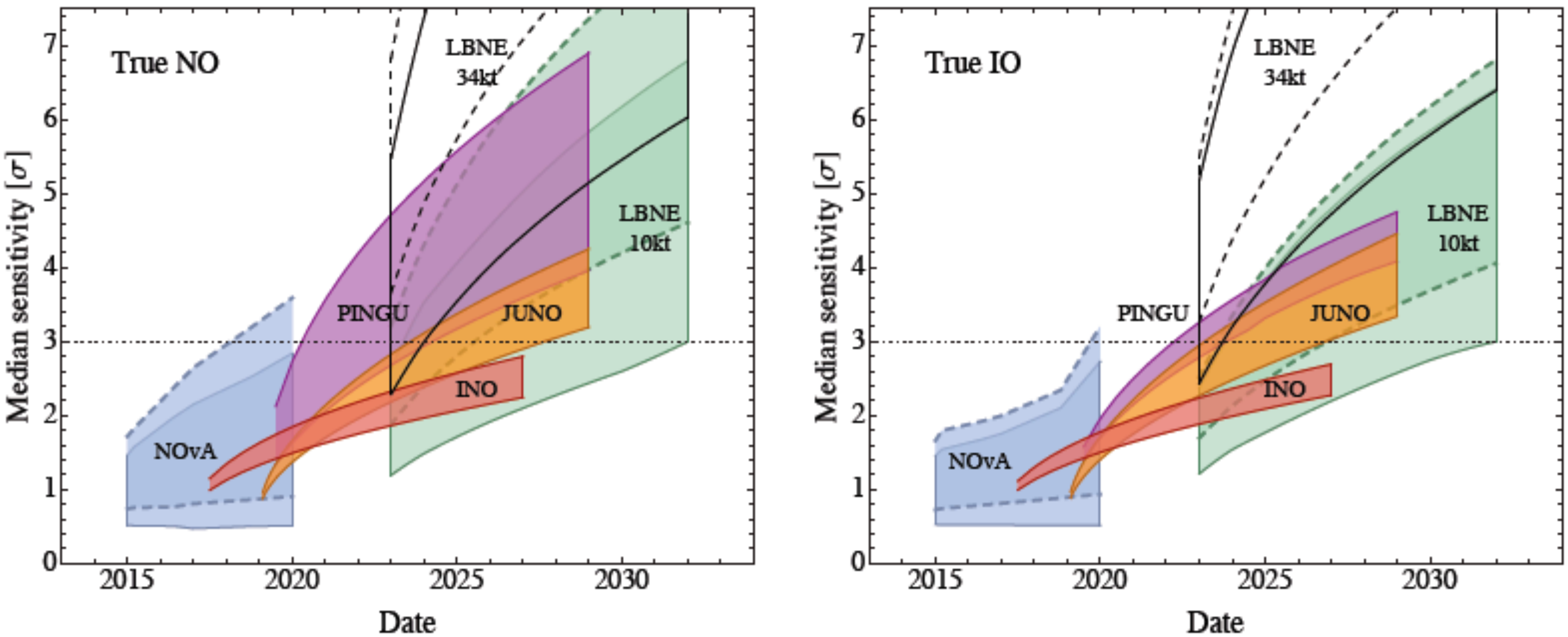
- **Important to Notice: Not always Gaussian**
- **Typical for low counting experiments**
- **Need to perform MC studies for accuracy**
- **Rejection power depends on the true parameters**





Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

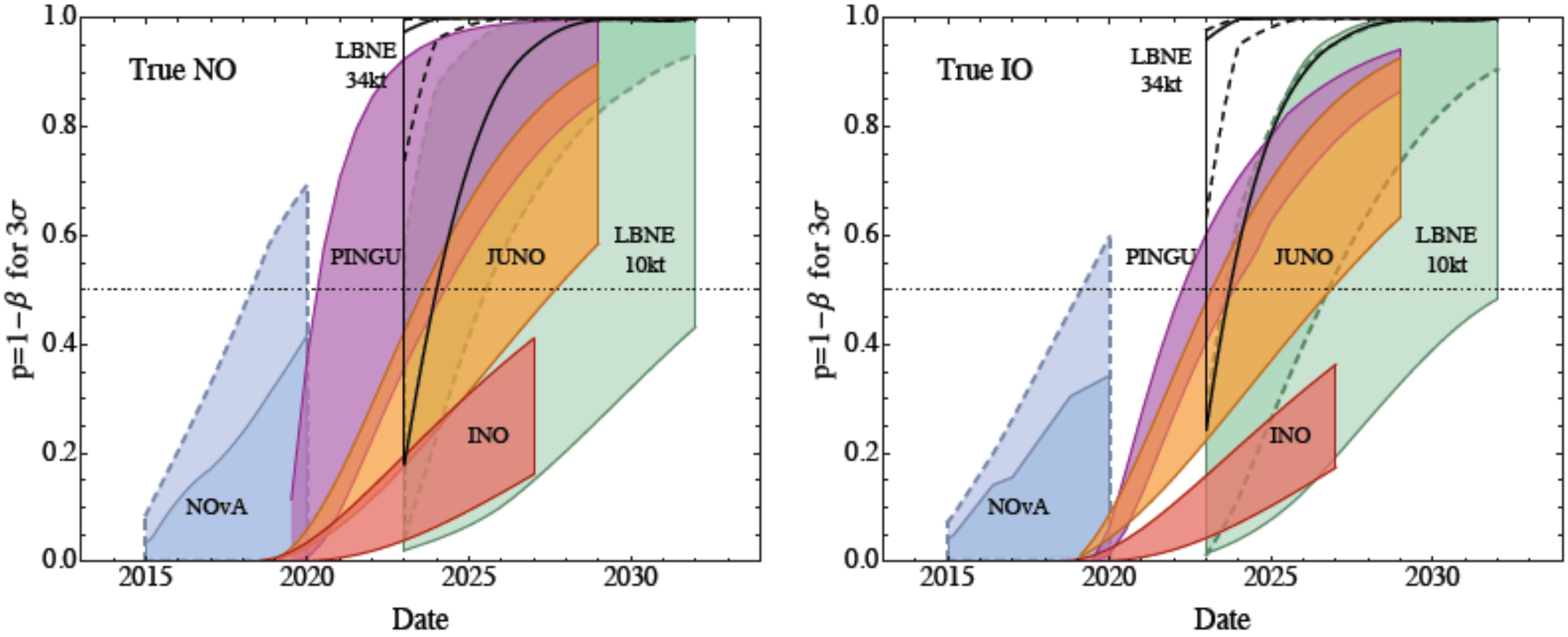
Comparison of Experiments – Median Sensitivity



Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

Bands have different meanings:
For NOvA and LBNE: Different true values of CP phases
For INO and PINGU: 2-3 mixing angle between 40 degree and 50 degree
For JUNO: Energy resolution between 3% and 3.5%

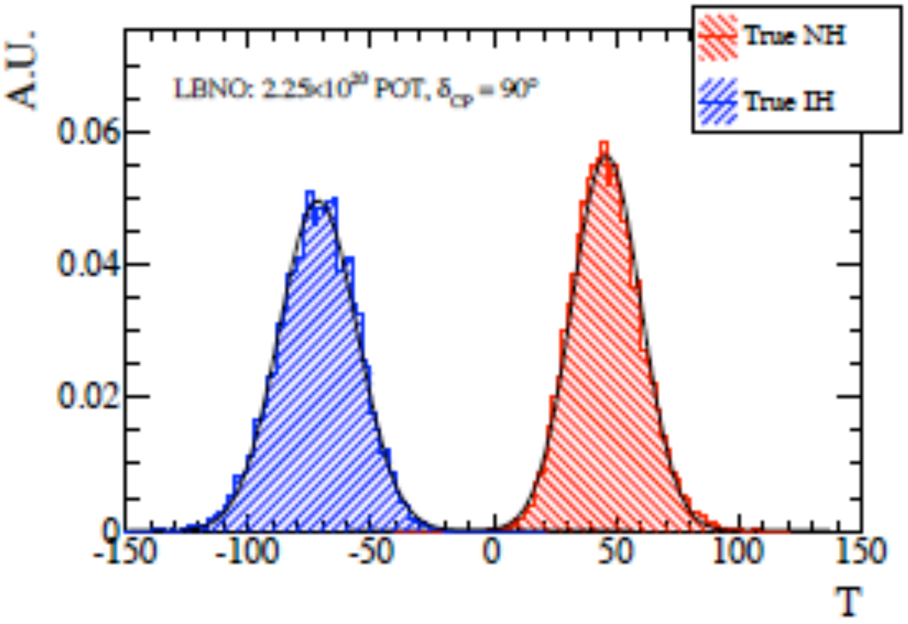
Probability of Rejecting Wrong ordering at 3σ

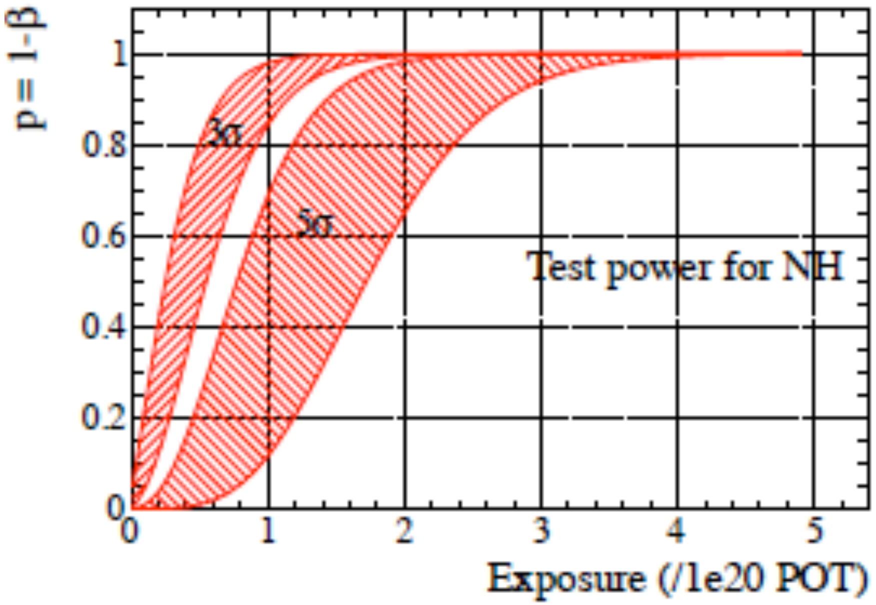


Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

Bands have different meanings:
For NOvA and LBNE: Different true values of CP phases
For INO and PINGU: 2-3 mixing angle between 40 degree and 50 degree
For JUNO: Energy resolution between 3% and 3.5%

Results for LBNO





LBNO Collaboration, arXiv: 1312.6520

Sensitivity Measures for MH

T_0	std. sens.	median sens.	crossing sens.	β for 3σ	68.27% range	95.45% range
9	99.73% (3.0σ)	99.87% (3.2σ)	93.32% (1.8σ)	0.41	$2.3\sigma - 4.2\sigma$	$1.4\sigma - 5.1\sigma$
16	99.9937% (4.0σ)	99.9968% (4.2σ)	97.72% (2.3σ)	0.11	$3.2\sigma - 5.1\sigma$	$2.3\sigma - 6.1\sigma$
25	99.999943% (5.0σ)	99.999971% (5.1σ)	99.38% (2.7σ)	0.013	$4.2\sigma - 6.1\sigma$	$3.2\sigma - 7.1\sigma$

Blennow, Coloma, Huber, Schwetz, arXiv:1311.1822v2

Sensitivity measures for neutrino mass hierarchy in the Gaussian approximation assuming $T_0^{\rm NO} = T_0^{\rm IO}$

Final Words

The sensitivity obtained by using the standard method of taking the square-root of the $\Delta\chi^2$ without statistical fluctuations is very close to the median sensitivity obtained within the Gaussian approximation for the test statistics T

Thank You