Imperial College London

Neutrino Physics & Statistics

–Looking forward, from 2016–

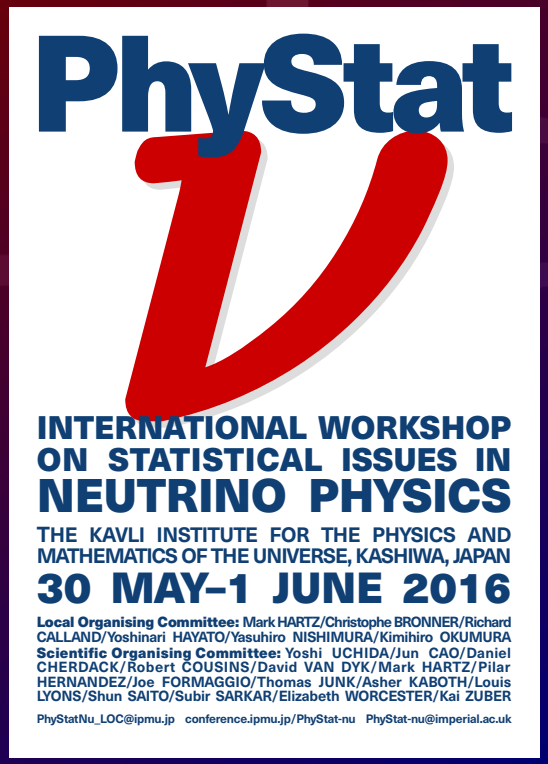
Yoshi Uchida

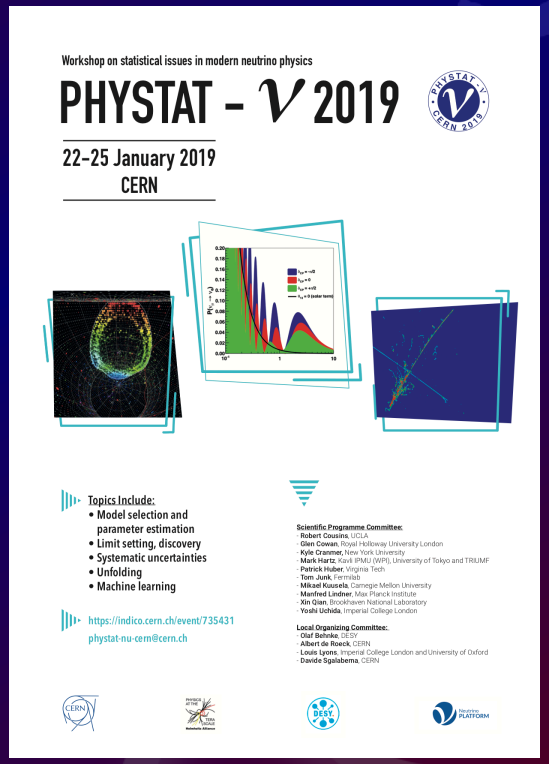
PHYSTAT-ν at CERN

22 January 2019

Disclaimer: many of the slides that are taken from past PhyStat-nu meetings represent the thoughts of the individual speakers, even when collaboration names are given in the labels for purposes of information.

PhyStat-ν

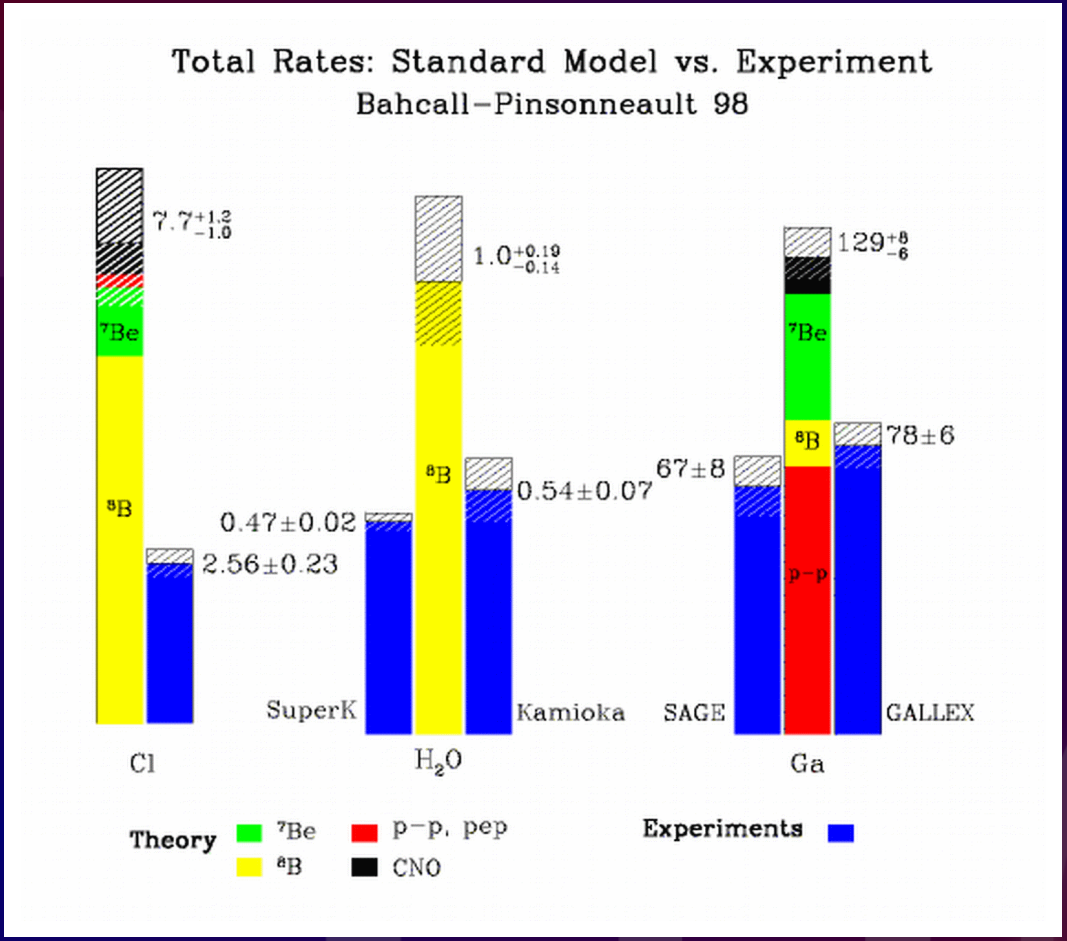




- For Statisticians
- For Neutrino Physicists
- For Non-Neutrino Particle Physicists

Early Neutrino Flux Measurements

Solar Neutrino Experiments



John Bahcall Neutrino Viewgraphs

- Yellow columns: theoretical expectation, normalised to the same height
- Blue columns: experimental results
- Shaded areas: uncertainties

Early Neutrino Flux Measurements

The Homestake Solar Neutrino Experiment

- Theoretical expectations: 9.3±1.3, 6.36, 7.64 SNUs for different calculations (right-hand axis)
- ("One FWHM" here refers to analysis cut tightness, not directly visible in the plot)

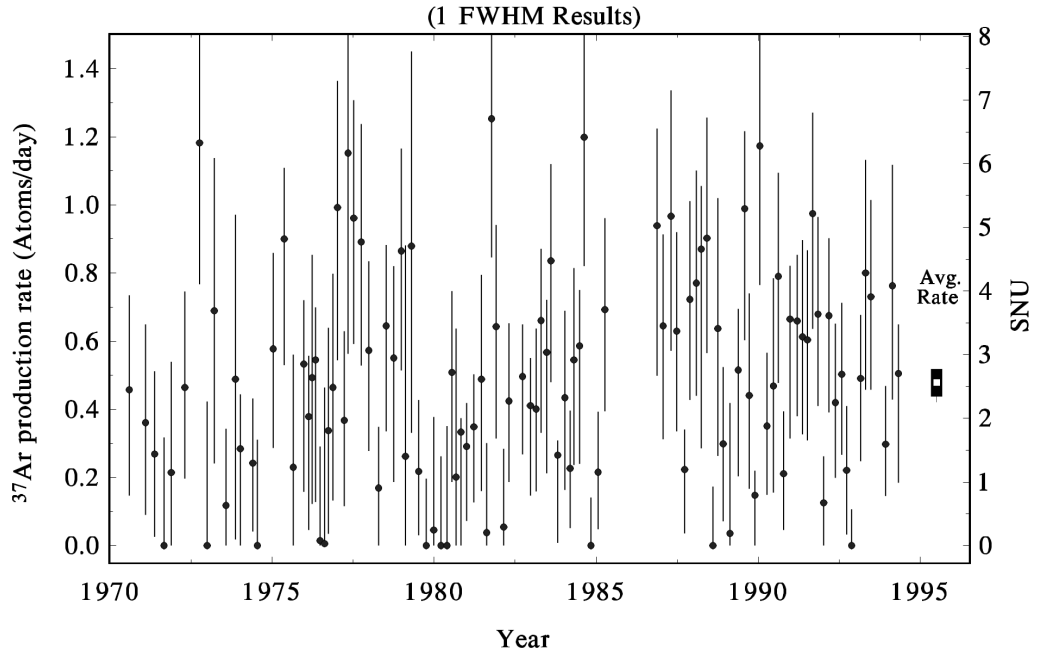


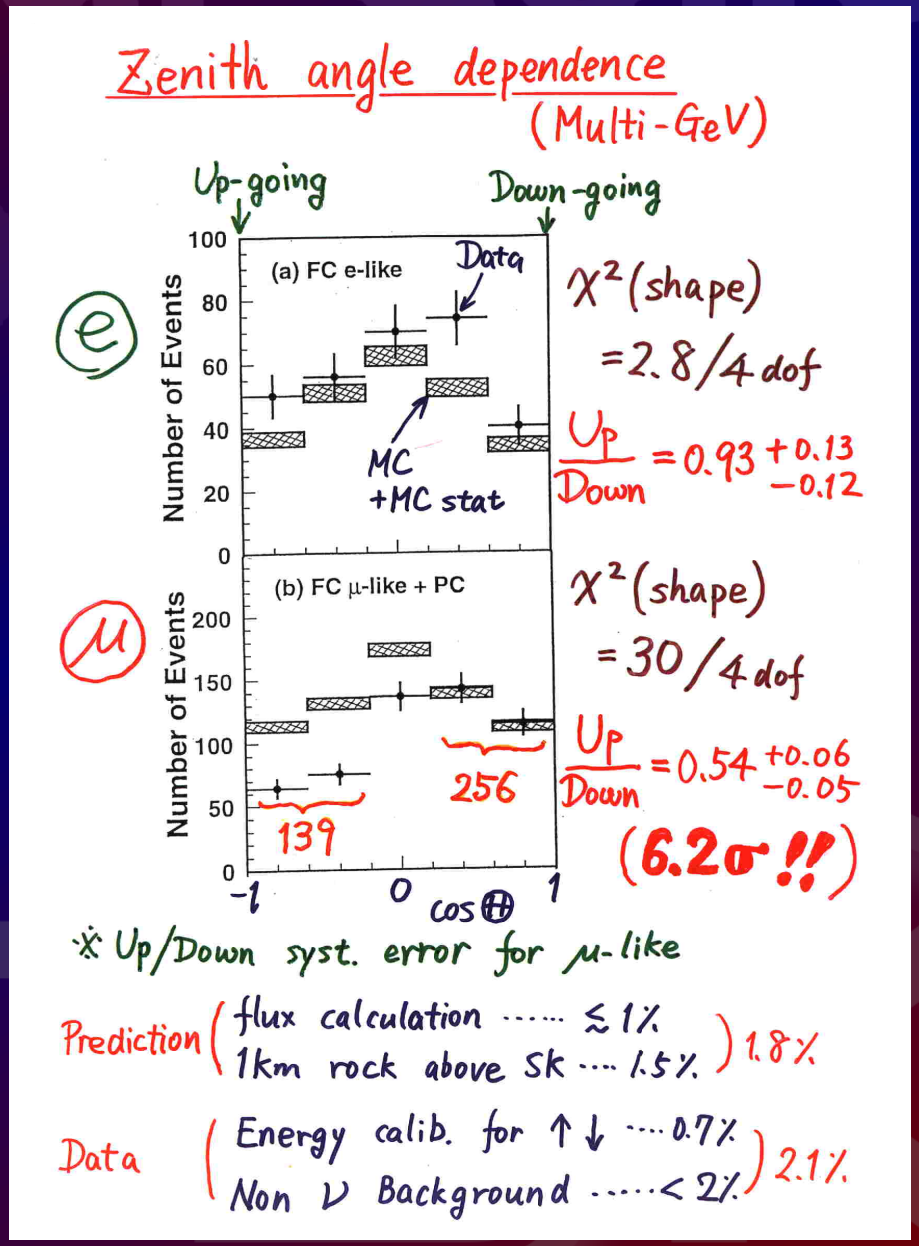
FIG. 13.—Homestake Experiment—one FWHM results. Results for 108 individual solar neutrino observations made with the Homestake chlorine detector. The production rate of ^{37}Ar shown has already had all known sources of nonsolar ^{37}Ar production subtracted from it. The errors shown for individual measurements are statistical errors only and are significantly non-Gaussian for results near zero. The error shown for the cumulative result is the combination of the statistical and systematic errors in quadrature.

Homestake Solar Neutrino Experiment, ApJ 496:505526 (1998)

Early Neutrino Flux Measurements

Atmospheric Neutrinos at Super-Kamiokande

- Down-going neutrinos travel $\mathcal{O}$(10 km)
- Up-going neutrinos travel $\mathcal{O}$(10,000 km)
- Deficit seen in up-going muon neutrinos
- Flavour change into tau neutrinos (cannot produce taus as they are too heavy)

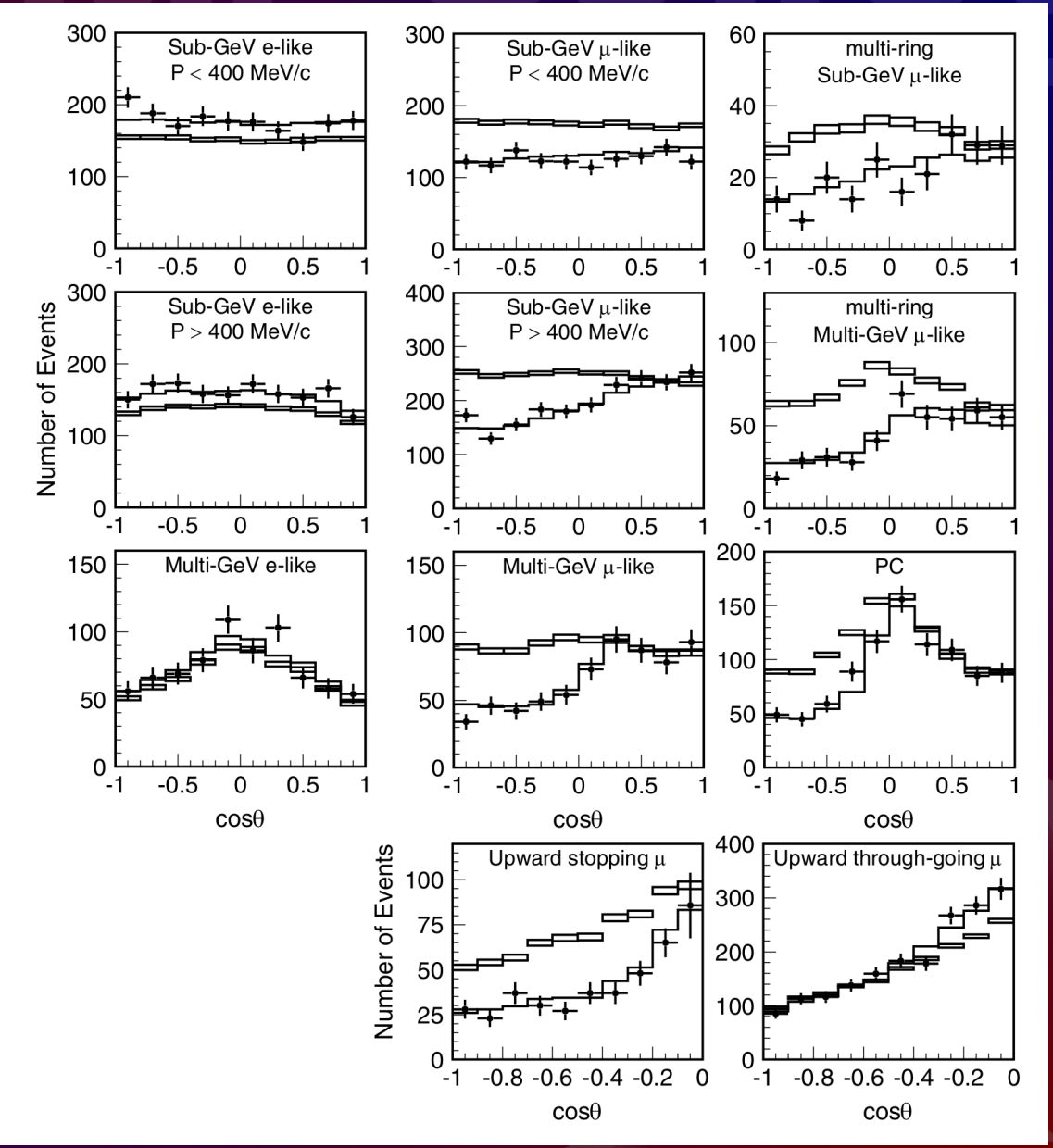


Takaaki Kajita, Super-Kamiokande Neutrino 1998

Early Neutrino Flux Measurements

Atmospheric Neutrinos at Super-Kamiokande

- Down-going neutrinos travel $\mathcal{O}$(10 km)
- Up-going neutrinos travel $\mathcal{O}$(10,000 km)
- Deficit seen in up-going muon neutrinos
- Flavour change into tau neutrinos (cannot produce taus as they are too heavy)



Super-Kamiokande (2005)

Early Neutrino Flux Measurements

KamLAND Reactor Neutrino Experiment (2002)

KamLAND should see

86.8 ± 5.6

(0.94 ± 0.85 background)

events if all antineutrinos travel to KamLAND from reactors without loss

Early Neutrino Flux Measurements

KamLAND Reactor Neutrino Experiment (2002)

KamLAND should see

86.8 ± 5.6

(0.94 ± 0.85 background)

events if all antineutrinos travel to KamLAND from reactors without loss

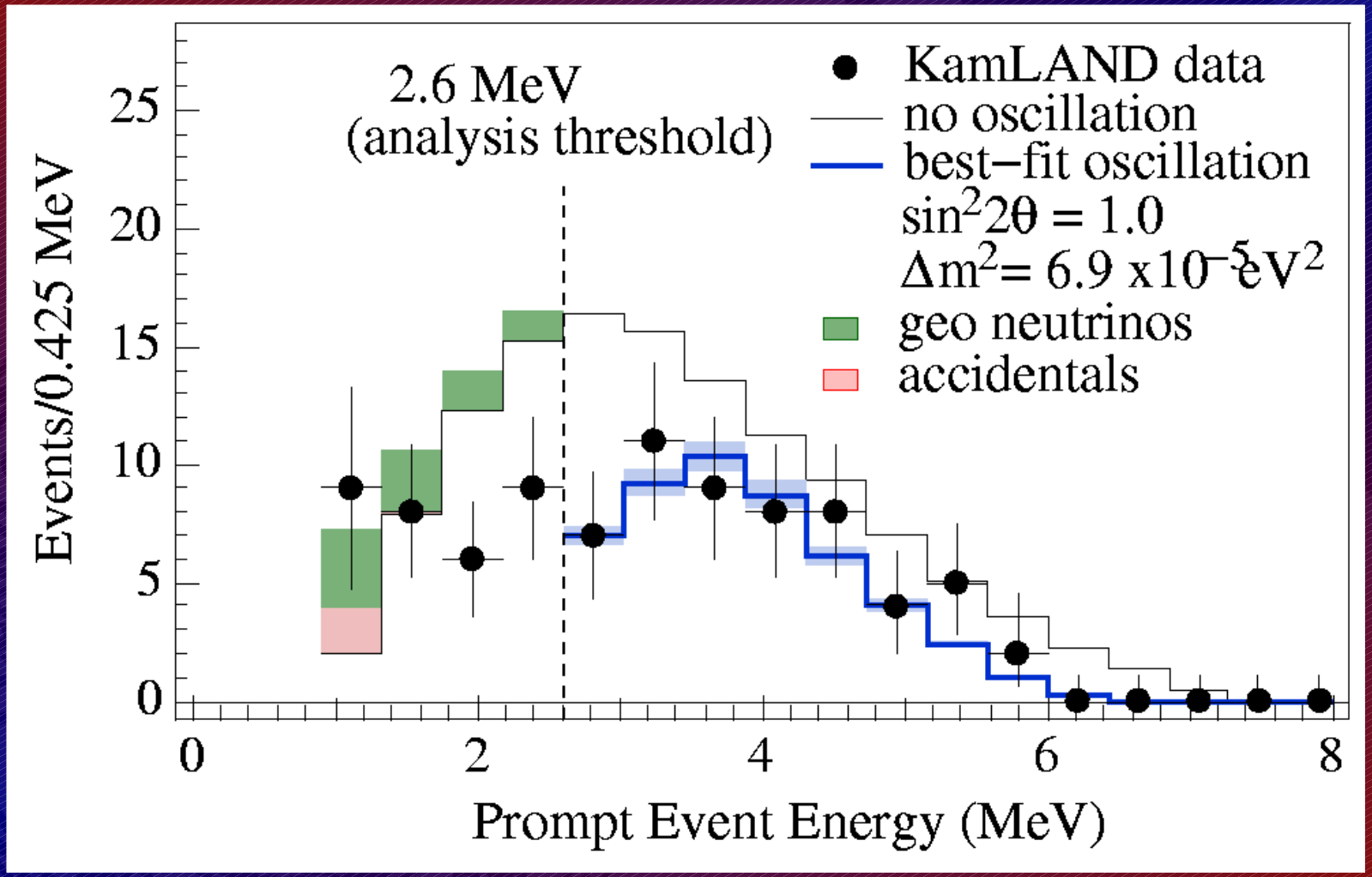
---

54

events observed

Early Neutrino Flux Measurements

KamLAND Reactor Neutrino Experiment (2002)



Presentation of Results

Neutrino Flavour Change

- These data show that neutrinos "disappear"
  - (Actually they become undetectable for kinematic reasons)
- Neutrino oscillation hypothesis
- Three types of neutrino known, but two-neutrino oscillation model worked:

$$P(\nu_A \to \nu_B) = \sin^2 2\theta \sin^2\left(1.27\Delta m^2\,[\mathrm{eV}^2]\frac{L\,[\mathrm{km}]}{E_\nu\,[\mathrm{GeV}]}\right)$$

Just two parameters sufficient to represent the oscillation effect

The amount of mixing, governed by θ

The differences of the squares of the (two) neutrino masses Δm^2

- affects the combination of Distance (L) and Energy (E) for oscillations to happen
- i.e., the experimental parameters

Two-Neutrino Oscillations

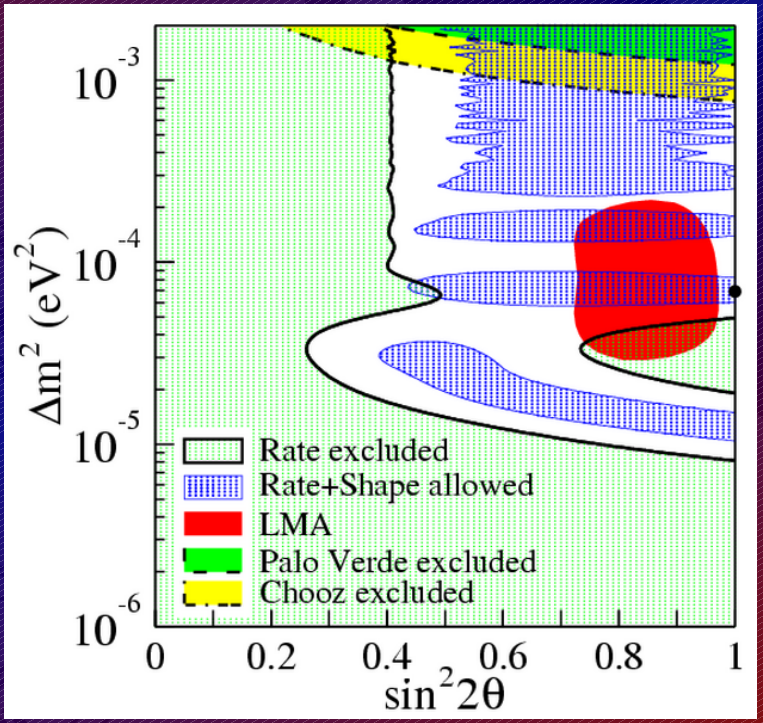
KamLAND Reactor Neutrino Experiment (2002)

Two-Generation Oscillation Hypothesis

95% Confidence Level regions

"Rate" = number of events

"Shape" = energy spectrum

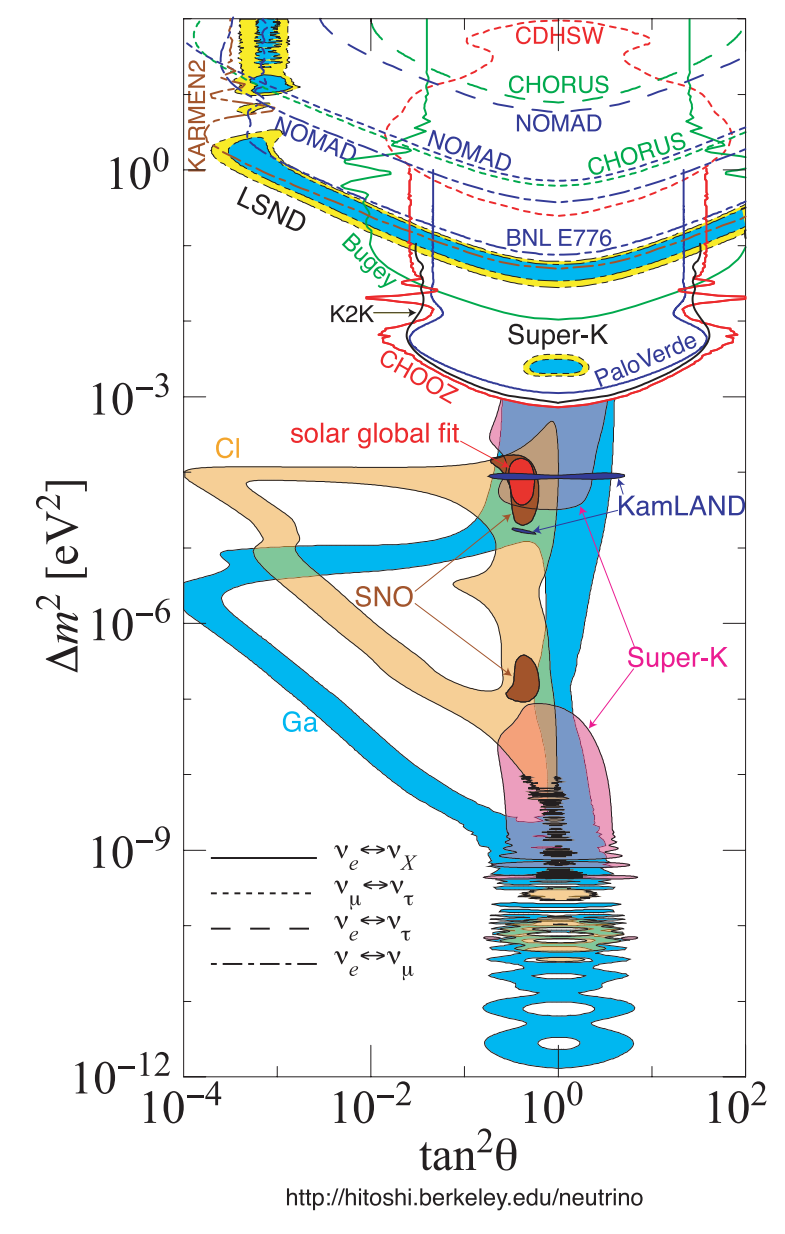


- Relatively simple $\Delta\chi^2$ analysis for two parameters at 95% C.L.
- Nuisance parameters: backgrounds, spectrum shape systematics, minimised at each raster scan point, with external constraints—**Profiling**
- No complicated nuisance parameters from the neutrino model

Two-Neutrino Oscillations

All Results (2004)

- All under the two-neutrino (two-parameter) oscillation hypothesis
- Exclusion contour lines
- Included regions filled in with colour
  - solar global fit, Super-K atmospheric, and (LSND)
  - Δm^2 values differ by orders of magnitude
    $\Rightarrow$ many of the current issues in oscillation physics



Hitoshi Murayama/PDG

Three-Neutrino Oscillations

$$U_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{23} & \sin\theta_{23} \\ 0 & -\sin\theta_{23} & \cos\theta_{23} \end{pmatrix} \times \begin{pmatrix} \cos\theta_{13} & 0 & \sin\theta_{13} e^{i\delta_{CP}} \\ 0 & 1 & 0 \\ -\sin\theta_{13} e^{-i\delta_{CP}} & 0 & \cos\theta_{13} \end{pmatrix}$$

$$\times \begin{pmatrix} \cos\theta_{12} & \sin\theta_{12} & 0 \\ -\sin\theta_{12} & \cos\theta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- Previous oscillations correspond to the "1–2" and "2–3" parameters
- The third type of oscillation would correspond to "1–3"
- Zero or very small θ_{13} combined with very different Δm^2 values for the two existing oscillations implies strong decoupling between them (Two-neutrino oscillations)
- Finite θ_{13} means two of the three Δm^2 values are very similar and oscillation effects become complicated

  $\Rightarrow$ Three-neutrino mixing

Three-Neutrino Oscillation Hypothesis

Global Neutrino Oscillation Fits (2004)

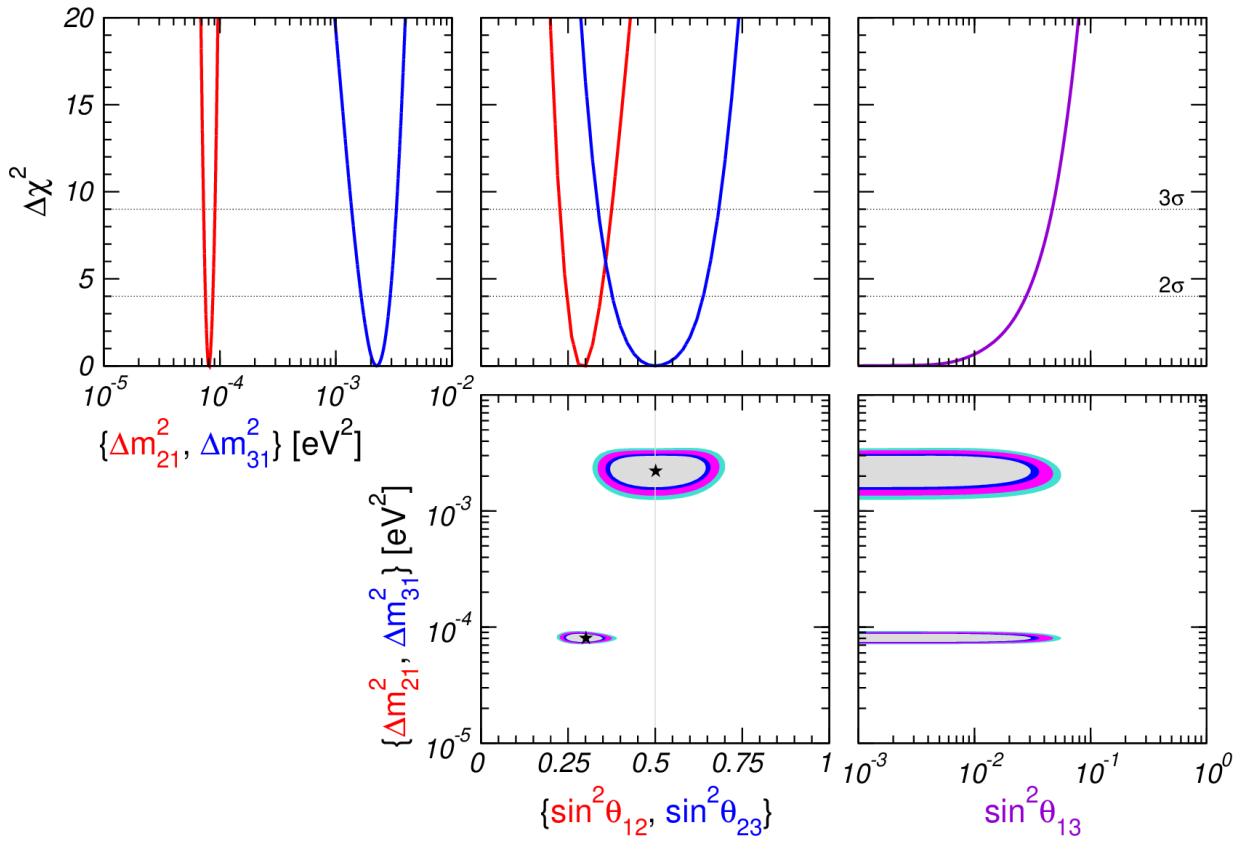
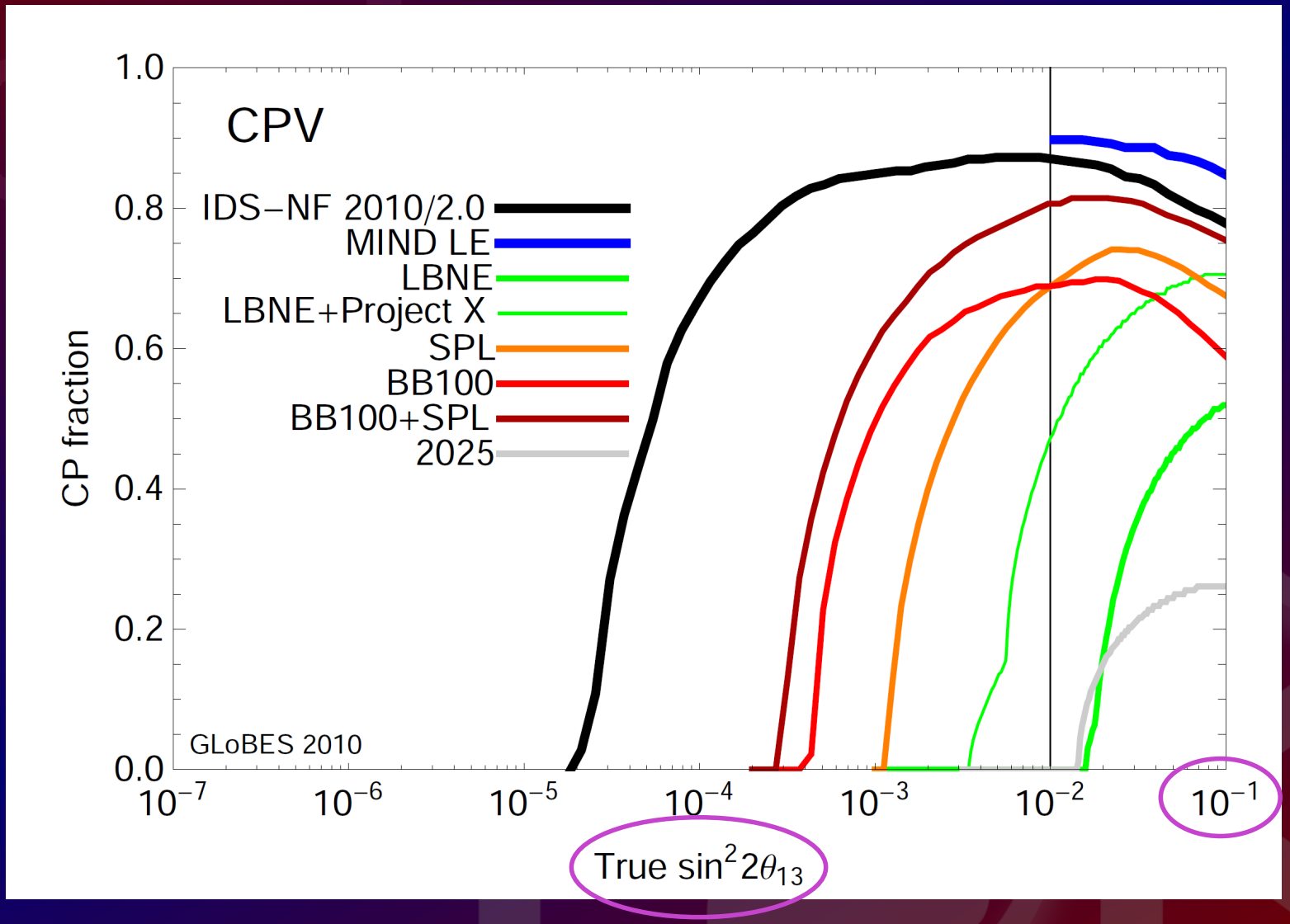


Figure 12. Projections of the allowed regions from the global oscillation data at 90%, 95%, 99%, and 3σ C.L. for 2 d.o.f. for various parameter combinations. Also shown is $\Delta\chi^2$ as a function of the oscillation parameters $\sin^2\theta_{12}, \sin^2\theta_{23}, \sin^2\theta_{13}, \Delta m^2_{21}, \Delta m^2_{31}$, minimized with respect to all undisplayed parameters.

Maltoni et al. DOI:10.1088/1367-2630/6/1/122 (2004)

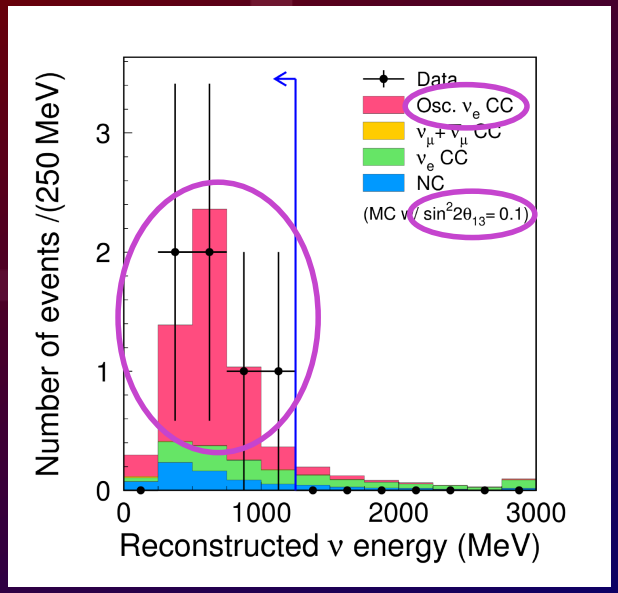
Three-Neutrino Oscillation Hypothesis

Sensitivity of Future Experiments for Different θ_{13} (2011)

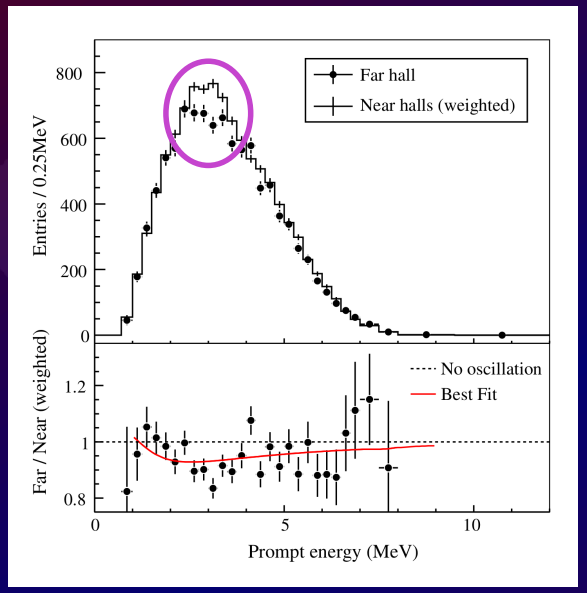


International Design Study for the Neutrino Factory, Interim Design Report (2011)

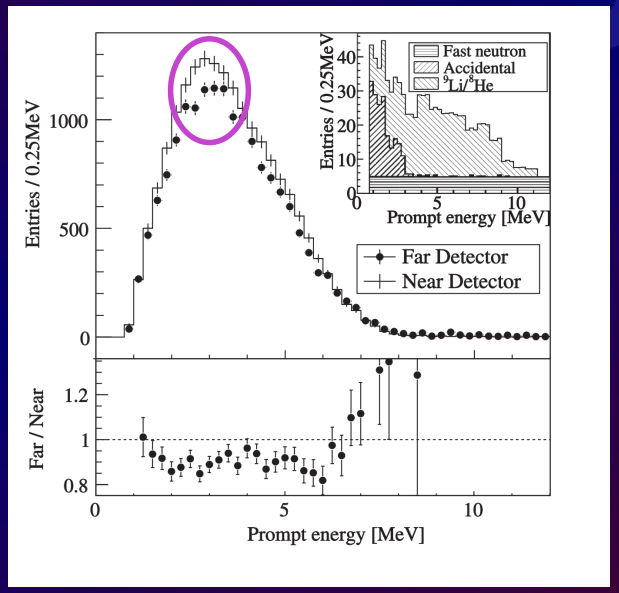
Discovery of the Third Neutrino Oscillation



T2K, Phys.Rev.Lett. 107 041801 (2011)



Daya Bay, Phys. Rev. Lett. 108, 171803 (2012)



RENO, Phys. Rev. Lett. 108, 191802 (2012)

- T2K (2011): **2.5** σ significance compared to no oscillations,
  $0.03 < \sin^2 2\theta_{13} < 0.28$ (90% C.L.)
- Daya Bay (2012): **5.2** σ,
  $\sin^2 2\theta_{13} = 0.092 \pm 0.016(\text{stat.}) \pm 0.005(\text{syst.})$
- RENO (2012): **4.9** σ,
  $\sin^2 2\theta_{13} = 0.113 \pm 0.013(\text{stat.}) \pm 0.014(\text{syst.})$

Oscillations in the Three Large Mixing-Angle Paradigm (and beyond)

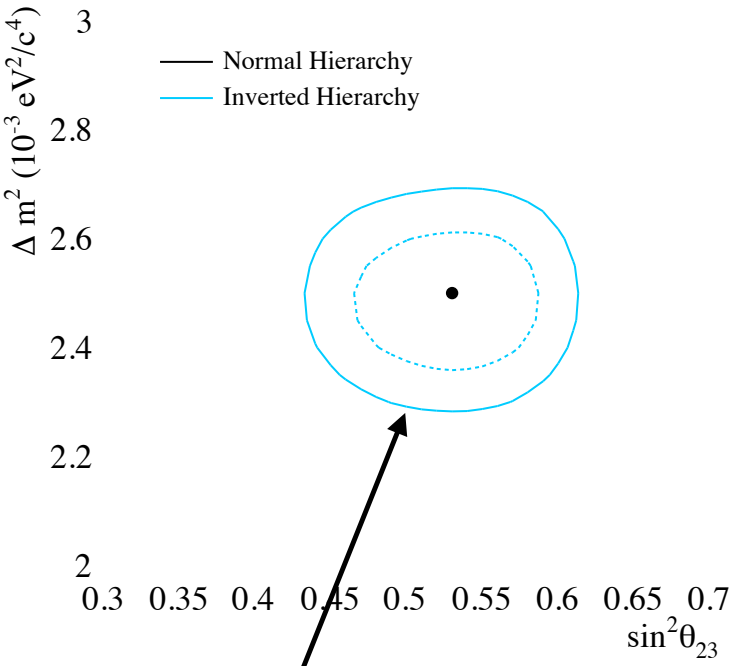
Since 2012

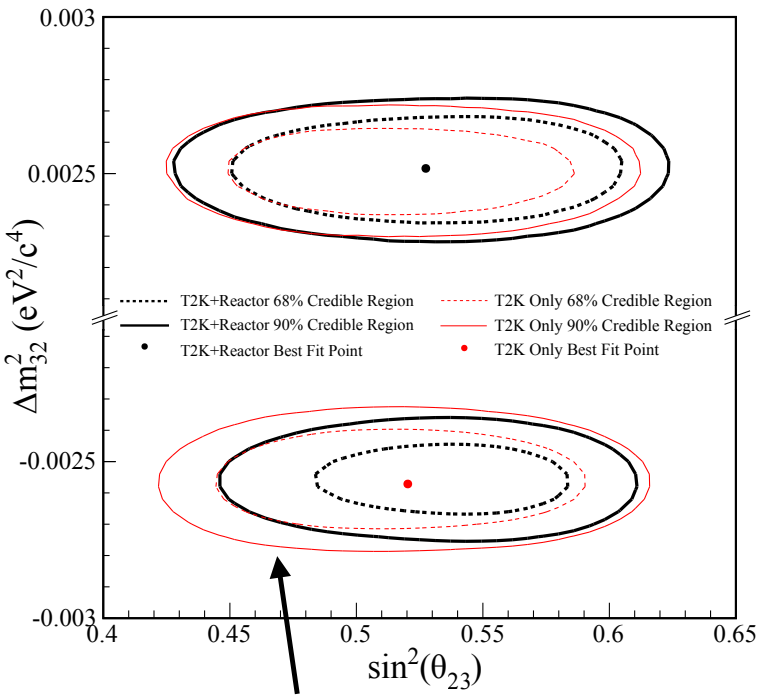
- Ten years ago we thought we might still be chasing a very small third mixing angle
- Now we know that all three mixing angles are "large":
  - $\theta_{12} \simeq 34$ degrees, $\theta_{23} \simeq 50$ degrees, $\theta_{13} \simeq 8.6$ degrees
- Clear signs of flavour change when seen: basic "rate analyses" were enough to be convincing in discovery mode
- Next steps could be more subtle
  - Determination of parameter values within the PMNS model
  - Mass Hierarchy / Ordering
  - CP-violation and δ_{CP}
  - Deviations from the Unitary 3×3 PMNS matrix
  - Oscillations into additional states
  - Non-standard interactions, Lorentz violation,....
- Consequences for statistical methods....

Three Large Mixing-Angle Paradigm

From PhyStat-ν 2016 IPMU

T2K loves all statistical methods!





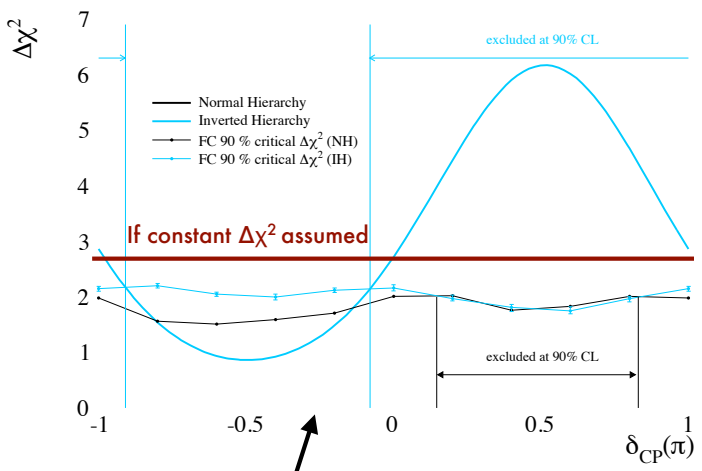
confidence level contours
constant $\Delta\chi^2$
Profiling treatment of systematics
Gradient descent

credible interval contours
highest posterior density method
Marginalization treatment of systematics
Markov Chain MC (Metropolis-Hastings)

Asher Kaboth, T2K/LBL-ν

Three Large Mixing-Angle Paradigm

From PhyStat-ν 2016 IPMU



confidence level interval
Feldman-Cousins-Highland ?
separate for NH/IH

16

A Brief Review of F-C-H

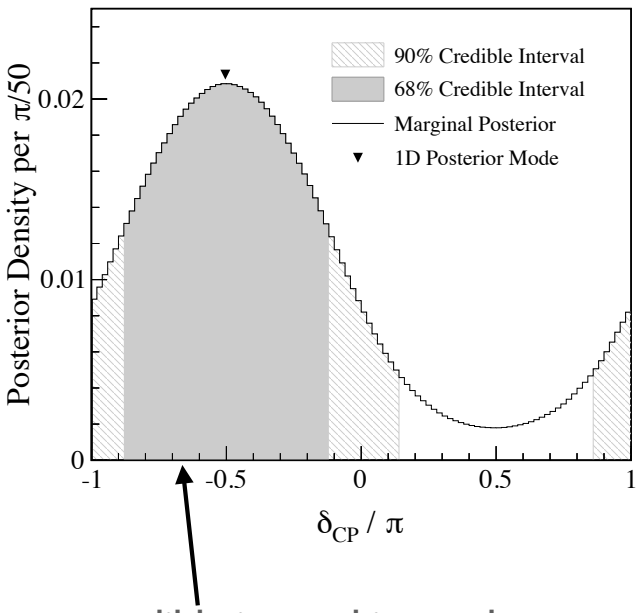
- Scan over your parameter space
- At every point, generate many toy MC experiments, including systematic parameter variations
- Fit your toys and generate the distribution of the test statistic
- Find the value of the test statistic where α% are excluded; this is the critical value of the test statistic at that point in the parameter space
- For the distribution of your data test statistic over your parameter space, draw a contour to exclude those where the data test statistic is greater than the critical value

17

Asher Kaboth, T2K/LBL-ν

Three Large Mixing-Angle Paradigm

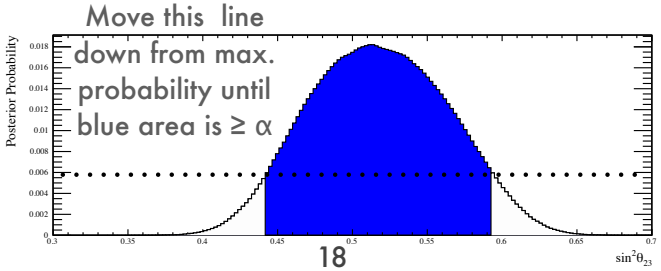
From PhyStat-ν 2016 IPMU



credible interval interval
highest posterior density method
marginalized over hierarchy

A Brief Review of HPD Credible Intervals

- Build your posterior in your parameter space of interest
- Draw your contour such that the integral of the probability inside your contour is α%, and such that the posterior density at any point inside the contour is higher than any point outside the contour
- This minimizes the size of the contour



18

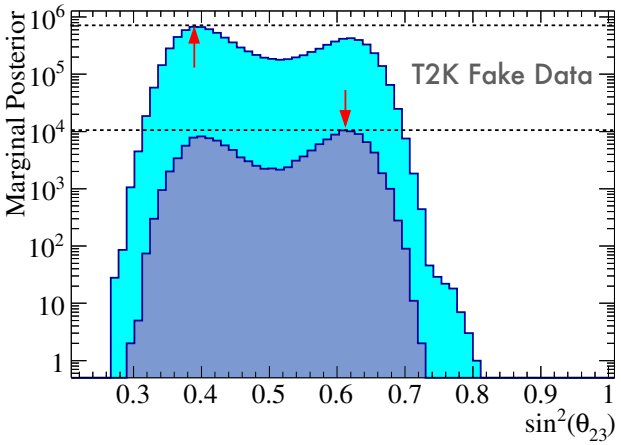
Asher Kaboth, T2K/LBL-ν

Three Large Mixing-Angle Paradigm

From PhyStat-ν 2016 IPMU

Marginalization vs Projection

- T2K found that treatment of systematics can have noticeable effects on analyses, especially when treating oscillation parameters, which are very non-gaussian
- T2K has moved to marginalization for all oscillation analyses, using both MCMC and likelihood averaging over toys



$$f_{marg}(x) = \int dy\, f(x,y)$$

$$f_{proj}(x) = \max_y f(x,y)$$

27

T2K

Asher Kaboth, T2K/LBL-ν

- Fake MC data generated at $\sin^2\theta_{23} = 0.4$
- **Marginalising** and **Profiling** give different preferred points
- Differences in meaning between Marginalising and Profiling

Presentation of Results

Elimination of nuisance parameters

- "Did you marginalise or did you profile?": Asked frequently at 2016 meetings
- **For each set of values in the parameters of interest:**
  - Profiling: Fix nuisance parameters to their best-fit points (with external constraints)
  - Marginalising: Integrate over nuisance parameters (according to their priors)
- Nuisance parameters not so problematic if they are for detector effects etc.
- But they also include physics parameters that are not well-measured yet
  - Such as θ_{23} when presenting θ_{13} and δ_{CP}; not at all Gaussian
  - Using the prior $\pi(\theta_{23})$ to integrate over
- Differences of coverage may also need to be studied

- "Never profile if you are a Bayesian"
- Many experimental systematics are also non-trivial, for example, neutrino production and interaction uncertainties
- **Q: Can we be agree on when we marginalise and when we profile?**

Frequentist and Bayesian Statistics

- From 2016 meetings: general open-mindedness towards use of Frequentist and Bayesian statistics
- Bayesian methods used quite enthusiastically for model selection, event classification and study of discrete outcomes (e.g., mass hierarchy)
- Frequentist methods may be preferred for discovery searches
- Tend to agree for parameter estimation, which is also less dependent on prior choice than model selection
- "Why throw away half your toolbox?"

- INSPIRE HEP numbers for papers with "Neutrino" in the title:
  - 1548 mention "Frequentist" and variants
  - 2816 mention "Bayesian" and variants
  - 560 mention both
  - 35044 mention neither!

Use of Bayesian Statistics

Mass Hierarchy

Strong Bayesian evidence for the normal neutrino hierarchy

Fergus Simpson,[a] **Raul Jimenez,**[a,b] **Carlos Pena-Garay**[c,d] **and Licia Verde**[a,b]

Abstract. The configuration of the three neutrino masses can take two forms, known as the normal and inverted hierarchies. We compute the Bayesian evidence associated with these two hierarchies. Previous studies found a mild preference for the normal hierarchy, and this was driven by the asymmetric manner in which cosmological data has confined the available parameter space. Here we identify the presence of a second asymmetry, which is imposed by data from neutrino oscillations. By combining constraints on the squared-mass splittings [1] with the limit on the sum of neutrino masses of $\Sigma m_\nu < 0.13\,\mathrm{eV}$ [2], and using a minimally informative prior on the masses, we infer odds of 42:1 in favour of the normal hierarchy, which is classified as "strong" in the Jeffreys' scale. We explore how these odds may evolve in light of higher precision cosmological data, and discuss the implications of this finding with regards to the nature of neutrinos. Finally the individual masses are inferred to be $m_1 = 3.80^{+26.2}_{-3.73}\,\mathrm{meV}; m_2 = 8.8^{+18}_{-1.2}\,\mathrm{meV}; m_3 = 50.4^{+5.8}_{-1.2}\,\mathrm{meV}$ (95% credible intervals).

- "minimally-informative prior on the masses"
- "infer **odds of 42:1 in favour of the normal hierarchy**"
- "which is classified as 'strong' on the Jeffreys' scale"

Objective Bayesian analysis of neutrino masses and hierarchy

Alan F. Heavens,[a] **Elena Sellentin**[b]

Abstract. Given the precision of current neutrino data, priors still impact noticeably the constraints on neutrino masses and their hierarchy. To avoid our understanding of neutrinos being driven by prior assumptions, we construct a prior that is mathematically minimally informative. Using the constructed uninformative prior, we find that the normal hierarchy is favoured but with inconclusive posterior odds of 5.1:1. Better data is hence needed before the neutrino masses and their hierarchy can be well constrained. We find that the next decade of cosmological data should provide conclusive evidence if the normal hierarchy with negligible minimum mass is correct, and if the uncertainty in the sum of neutrino masses drops below 0.025 eV. On the other hand, if neutrinos obey the inverted hierarchy, achieving strong evidence will be difficult with the same uncertainties. Our uninformative prior was constructed from principles of the Objective Bayesian approach. The prior is called a reference prior and is minimally informative in the specific sense that the information gain after collection of data is maximised. The prior is computed for the combination of neutrino oscillation data and cosmological data and still applies if the data improve.

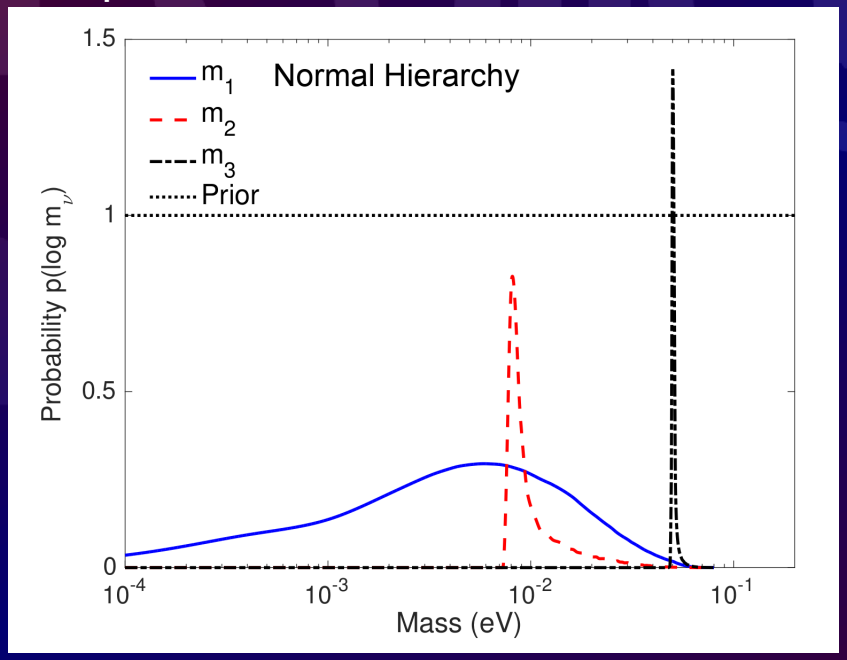
- "we construct a prior that is minimally-informative"
- "we find that the **normal hierarchy is favoured but with inconclusive posterior odds of 5.1:1**"
- "our uninformative prior was constructed from principles of the Objective Bayesian approach"

Use of Bayesian Statistics

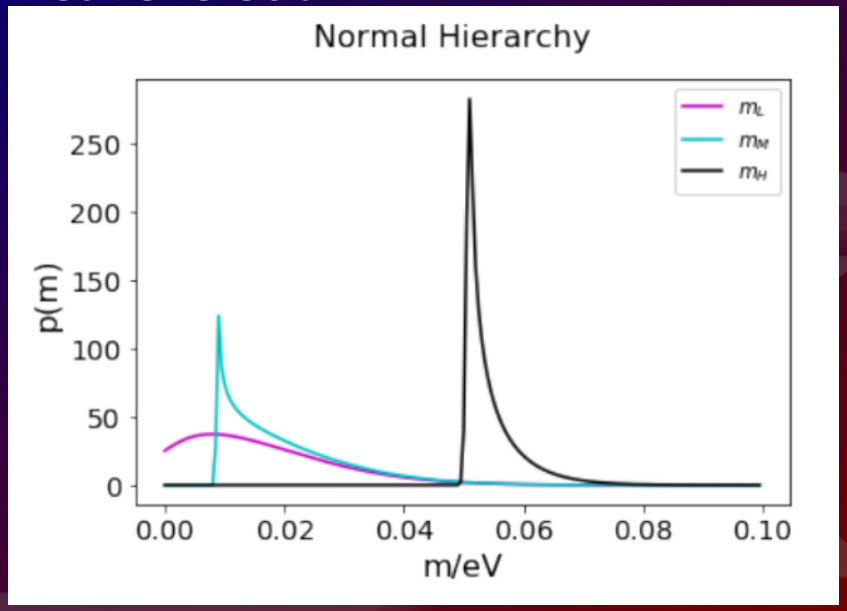
Mass Hierarchy

- Conflicting conclusions for model selection
- Quite similar results for parameter fitting
- Simpson uses logarithms of masses

Simpson et al.:



Heavens et al.:



The Neutrino Mass Hierarchy/Ordering

- Two disjiont hypotheses
- $\Delta S = \chi^2_{\rm IH} - \chi^2_{\rm NH}$
- Wilk's Theorem does not apply
- Bayes factor:

$$\frac{p(\text{NH}|\text{data})}{p(\text{IH}|\text{data})} = \frac{\pi(\text{NH})}{\pi(\text{IH})}\frac{p(\text{data}|\text{NH})}{p(\text{data}|\text{IH})}$$

Unbiased prior between two hierarchies usually used

- CL_s sometimes used to report results

Distribution of $\Delta\chi^2$

The binary classification function

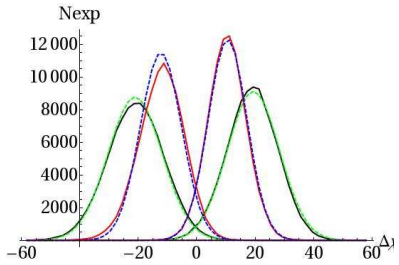
$$\Delta\chi^2 = \chi^2_{\rm IH} - \chi^2_{\rm NH}$$

does not follow a one-degree-of-freedom χ^2 distribution
(for example: it is not always > 0)
Not the first case in physics: see, for ex. Cousins *et al.*, JHEP 2005
Discrimination between spin-0 and spin-2 resonances at LHC

In the absence of degeneracies, to a good approximation it follows a Gaussian distribution, with

$$\mu = \pm\overline{\Delta\chi^2} \qquad \sigma = 2\sqrt{\overline{\Delta\chi^2}}$$

Right: distribution of $\Delta\chi^2$ using different assumptions on the precision on Δm^2_{32}. Dashed curves: results of a JUNO MC simulation



EC, Evslin and Zhang JHEP 2014

Disjoint Hypotheses: Approaches

The neutrino literature has so far used three statistical approaches to the MH determination:

1) **Cox 1961; 1962**:
Test each hypotheses (frequentist approach), then compare the likelihood. Possible to accept or reject each hypothesis separately
In the neutrino literature: (Qian et al. PRD 2012; Blennow et al. JHEP 2014)

2) **Hotelling 1940; Vuoung 1989**:
Consider the hypothesis that both H_1 and H_2 are equally effective. Test this new hypothesis (frequentist approach).
In the neutrino literature: (Capozzi, Lisi and Marrone PRD 2014)

3) **Jeffreys 1935; 1961; Kass and Raftery 1995:**
Bayesian Model Selection with the Bayes factor
In the neutrino literature: (Qian et al. PRD 2012; EC, Evslin and Zhang JHEP 2014; Blennow JHEP 2014)

DISCLAIMER

There are no ''right'' an ''wrong'' definitions of sensitivity of an experiments; we will discuss several possible definitions, stressing the relations with quantities of physical interest

Neutrino Cross Sections

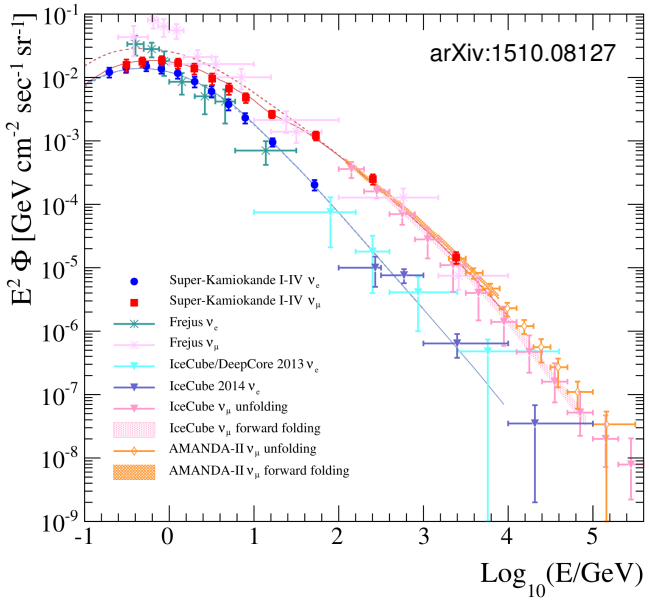
Measurement, Reporting and Use

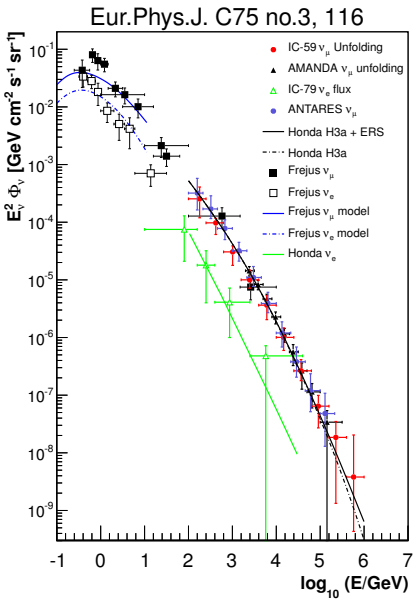
- Much discussion on "Cross Sections" from the GeV to the PeV range; lack of event-by-event neutrino energy information and difficult of reconstruction makes these a leading source of systematics in measurements
  - **Extraction of information from neutrino events** including **Unfolding**
  - **Presentation of cross section data**, for current and future experiments
  - **Use of cross section data**, from current and past experiments

Event Unfolding

From PhyStat-ν 2016 IPMU

And going back to measuring the ν flux





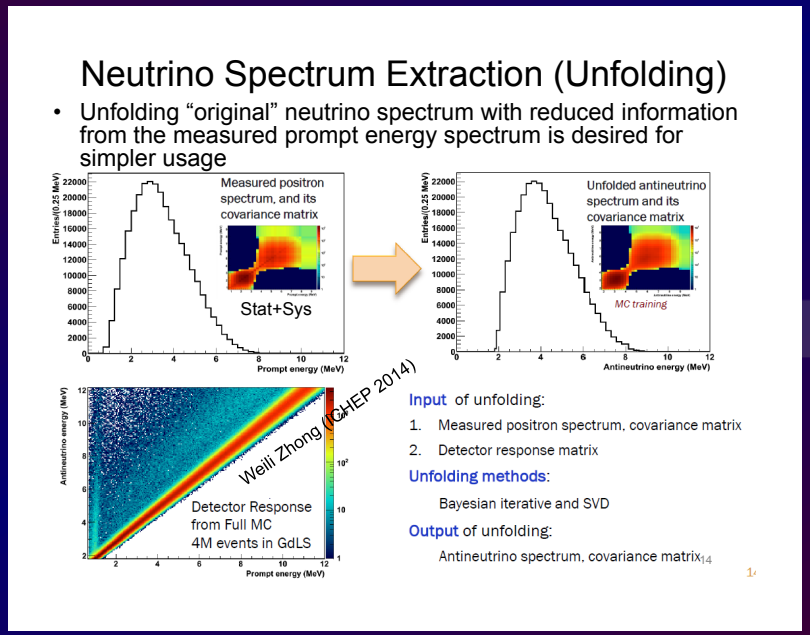
- While ν flux is input for oscillation analysis, we **also** measure it
- Most common techniques:
  - Forward folding
    - ★ fit model-parameters $\Rightarrow$ bands accounting for uncertainties
  - Unfolding
    - ★ model-independent method

JP de Andr, IceCube/Atmospheric ν

Event Unfolding

From PhyStat-ν IPMU & FNAL

- Much discussion on Unfolding and Forward Folding at IPMU
- Led to dedicated talk at Fermilab and a full session here
- **Q: Can/should we move away from unfolding?**



Xin Qian, Daya Bay/Reactor ν

Unfolding of the energy spectrum

- The general idea of unfolding is to obtain the true ν energy spectrum ($f(E)$) from the measured data ($g(y)$)
- Those quantities are related by a "response matrix" ($A(y, E)$):

$$g(y) = A(y, E) f(E)$$

- We would like to invert $A(y, E)$ to obtain f from g
  - However this is an ill defined problem
  - Even when we can invert A the solution might be unstable
- Different techniques about how to go about doing the unfolding
  - For invertible A, add regularisation terms to avoid instability
    - Regularization term corresponds to a priori information about smoothness of true solution
    - Could introduce bias in solution, weight of "regularization" needs to be determined so bias smaller than statistical errors
  - "Bayesian Unfolding"

"Bayesian Unfolding"

notation from Nucl.Instrum.Meth. A362 (1995) 487-498

- Goal: Identify the "causes" (C_i) that produced the observable effects (E_j)
- Starting point: Bayes theorem

$$P(C_i|E_j) = \frac{P(E_j|C_i)P(C_i)}{P(E_j)} = \frac{P(E_j|C_i)P(C_i)}{\sum_l P(E_j|C_l)P(C_l)}$$

- Number of events from C_i ($\hat{n}(C_i)$) given the number of events in E ($n(E_j)$)

$$\hat{n}(C_i) = \frac{1}{\epsilon_i} \sum_j n(E_j) P(C_i|E_j) = \sum_j M_{ij} n(E_j), \quad \epsilon_i \neq 0$$

($\epsilon_i = \sum_j P(C_i|E_j)$ is efficiency of detecting C_i and M_{ij} is unfolding matrix)

- Unfolding process:
  1. Choose initial $P_0(C)$, can be uniform
  2. Calculate $\hat{n}(C)$ and $\hat{P}(C) = P(C_i|n(E))$
  3. Compare $\hat{n}(C)$ and $n_0(C) = P_0(C)N_{obs}$
  4. Replace P_0, n_0 with $\hat{P}, \hat{n}$ and start over at 2 until converged
     - $\hat{P}$ can be smoothened to avoid "overfitting"
     - Uncertainties calculated after final M_{ij} and $n(E_j)$
- Note [from R. Cousins]: this is not truly a Bayesian technique
  - Also called: "Expectation Maximization" or "Lucy-Richardson"

JP de Andr, IceCube/Atmospheric ν

JP de Andr, IceCube/Atmospheric ν

Event Unfolding

From PhyStat-ν 2016 IPMU

Unfolding instability example

From V. Blobel, “Unfolding Methods in High-energy Physics Experiments”

Example: Unfolding of a distribution of a discrete variable. The case $n = m = 20$ is assumed, with the following response matrix:

$$A = \begin{pmatrix} 0.75 & 0.25 & 0 & & & \\ 0.25 & 0.50 & 0.25 & 0 & & \\ 0 & 0.25 & 0.50 & 0.25 & 0 & \\ & 0 & 0.25 & 0.50 & 0.25 & \\ & & 0 & 0.25 & 0.50 & \\ & & & & & \ddots \end{pmatrix}$$

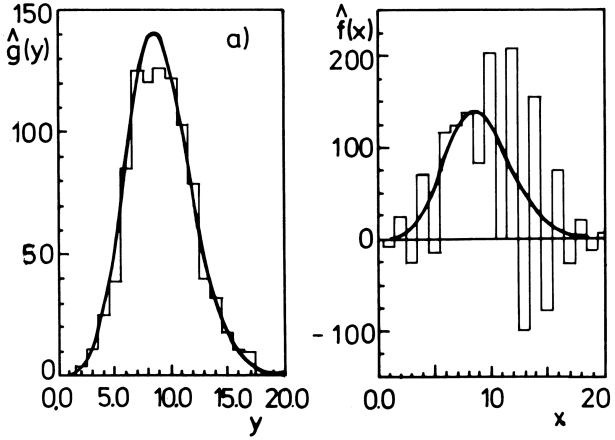


Figure 1. Distribution of the measured quantity y (a) and oscillating result of unfolding (b) using equation (1.05), shown as histograms. The original dependence is shown as a curve in both cases.

JP de Andr, IceCube/Atmospheric ν

Presentation of Cross-Section Data

- Distinguish between σ model and detector model
  - Any MC-derived quantity is, of course, model-dependent
- Restricting unsmearing, BG corrections, and efficiencies to detector MC quantities, not cross section processes is probably the best we can do
- This is why we should always publish final state particle cross sections first, and then proceed to process measurements, etc.
- Can we move to generator-less measurements??

Absolute flux-averaged differential cross section formula

U_{ij} :unsmearing matrix

N_j^{obs} : data

$$\hat{N}_i = \frac{\sum_j U_{ij}(N_j^{obs} - B_j)}{\epsilon_i}$$

ε_i :efficiency

b_j : background

$$\frac{d\sigma}{dx_i} = \frac{1}{T\Phi_\nu}\frac{\hat{N}_i}{\Delta x_i}$$

T :integrated target number

Φ :integrated ν-flux

Morgan Wascko, Cross Section Measurements, PhyStat-ν 2016 IPMU

- Many **model-dependencies** when extracting "true" information

  $\Rightarrow$ need for model-independent measurements/results/presentation of data
- Provide later physicists with clean data to use with new models
- Confront models with many independent data sets

Presentation of Cross-Section Data

From data to cross sections

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x_i} = \frac{\sum_j \widetilde{U}_{ij}\,(N_j - B_j)}{\Phi_\nu T \Delta x_i \epsilon_i},$$

where N_j = the number of selected events,
B_j = predicted background events in reconstructed bin j,
Φ_ν = total integrated flux,
T = number of target nuclei per unit area,
Δx_i = width of the true bin, ϵ_i = selection efficiency,
and $\widetilde{U}_{ij}$ = the unfolding matrix

- Many **model-dependencies** when extracting "true" information

  $\Rightarrow$ need for model-independent measurements/results/presentation of data
- Provide later physicists with clean data to use with new models
- Confront models with many independent data sets

Presentation of Cross-Section Data

From PhyStat-ν 2016 IPMU

- Discussion on how to move away from model-dependencies when presenting cross-section experiment results
- General agreement, and focus of discussion was on unfolding and avoiding data reduction
- Idea 1 is the same as providing likelihood functions ($\rightarrow$ concerns for higher dimensions), and avoiding unfolding
- Idea 2 can be related to Generative Modelling (see later)
- **Q: What are good ways to present our data with future physicists in mind?**

Idea 1:

- Experimentalists use *uncorrected* data to perform parameters fits.
  - We use detector MC to naturally handle efficiencies and smearing,
    - by comparing smeared MC samples after cuts to data after cuts.
  - We use MC samples/reweighting techniques to adjust MC until it matches data.
- Why not publish uncorrected data along with appropriate smearing and efficiency functions?
  - Turn detector MC into a lookup table.

Idea 2:

- Apply efficiency/smearing corrections to individual events.
- Publish an ntuple of events: reconstructed final state particles.
  - Each event comes with a cross-section weight derived with POT numbers, flux & detector MC.
  - Allows one to make a plot giving cross sections instead of number of events.

➡ Gives unprecedented knowledge to future analysers since it would allow analysis of <u>new variables</u>

Use of Cross-Section Data

Event Rate: $R(\vec{x}') = \int \mathrm{d}E_\nu \, \Phi(E_\nu) \cdot \sigma(E_\nu, \vec{x}) \cdot \epsilon(\vec{x}, \vec{x}') \cdot P(\nu_A \to \nu_B; E_\nu)$,

with the **cross section** $\sigma(E_\nu, \vec{x})$
(which relates E_ν with the true outgoing particle kinematics, $\vec{x}$)
and the detector response $\epsilon(\vec{x}, \vec{x}')$ and the flux $\Phi(E_\nu)$

Questions we want to answer

1. How should we calculate goodness of fit and select a model with limited information about the data?
2. How should we define the parameter errors for that model?
3. Should we exclude any datasets from our fits and how should we decide which?

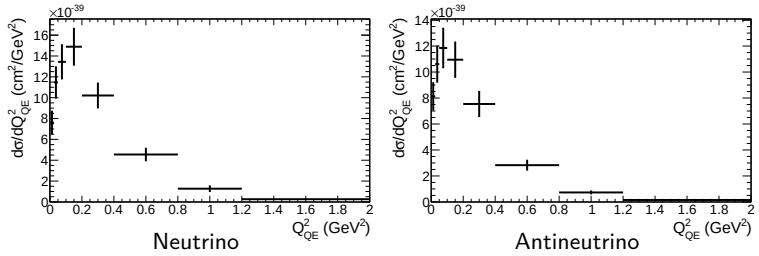
I will reiterate these and plead for help at the end.
But please interrupt at any point!

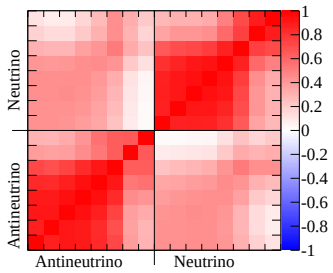
Callum Wilkinson, Combining Cross-Section Data, PhyStat-ν 2016 IPMU

Use of Cross-Section Data

- Experiments provide results in different ways, which have evolved over the years
- Data can go back decades, with much information lost
- Background subtraction often involves use of different models/generators

MINERνA CCQE data

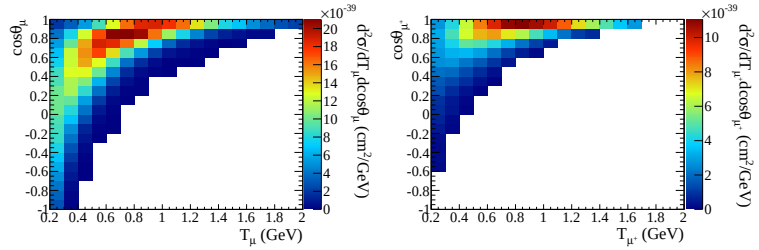




Correlation matrix

- **Unfolded distributions** differential in Q^2_{QE} (derived from the muon kinematics), on ~CH target
- **Background subtracted**.
- MINERνA provide the full covariance.

MiniBooNE CCQE data



- **Unfolded distributions** differential in muon angle, $\cos\theta_\mu$, and kinetic energy, T_μ on a ~CH_2 target.
- **Background subtracted**.
- MiniBooNE only released the diagonals of the shape-only covariance matrix and the normalization uncertainty for each dataset:
  - ν: $\pm 10.7\%$
  - $\bar{\nu}$: $\pm 13.0\%$
- **No information is available about correlations.**

Use of Cross-Section Data

Model Selection Methods and Discussion

- Single χ^2 fit across all data, varying model parameters and normalisations
  - using correlations where available
  - inconclusive and inconsistent results
- Parameter Goodness of Fit* (PGof) which compares the best fit across all datasets with when each dataset is at its best fit point
  - assessing the χ^2 given the difference in degrees of freedom (so assumes Gaussian statistics)
  - Arbitrary scaling of errors was needed

- Suggestion to use Bayesian Hierarchical Modelling-based meta-analysis of data
  - e.g., Hippel et al. The Power of Principled Bayesian Methods in the Study of Stellar Evolution
  - application of this to neutrino cross section data not clear
  - mentioned in context of Gravitational Waves at Fermilab in 2016
- General advice otherwise was to fit everything and assess
- **Q: How should one properly make use of past and current data?**

(* Testing the statistical compatibility of independent data sets M. Maltoni and T. Schwetz Phys. Rev. D 68, 033020)

Generative Modelling

Frequentist Treatment

- A single likelihood which connects all the parameters to the data

---

- "Decoupled science from statistics"
- Can apply any statistical procedure, given the likelihood

- **Q: Can we introduce these ideas more explicitly in our thinking?**

Frequentist methods use the likelihood to identify parameters compatible with the data *on average*.

$$\hat{\theta}(\mathcal{D})$$

$$\mathcal{L}(\theta) = \int \mathcal{L}(\hat{\theta}(\mathcal{D}), \theta)\pi(\mathcal{D} \mid \theta)\, \mathrm{d}\mathcal{D}$$

In general, nuisance parameters have to enter into the estimator design.

$$\hat{\theta}(\mathcal{D})$$

$$\mathcal{L}(\theta, \widetilde{\theta}) = \int \mathcal{L}(\hat{\theta}(\mathcal{D}), \theta)\pi(\mathcal{D} \mid \theta, \widetilde{\theta})\mathrm{d}\mathcal{D}$$

$$\mathcal{L}(\theta)_{\max} = \max_{\widetilde{\theta}} \mathcal{L}(\theta, \widetilde{\theta})$$

In *some* applications, relevant parameters can be isolated with techniques such as profiling.

$$\hat{\widetilde{\theta}}(\theta) = \operatorname*{argmax}_{\widetilde{\theta}} \pi(\mathcal{D} \mid \theta, \widetilde{\theta})$$

$$\hat{\pi}(\mathcal{D} \mid \theta) = \pi(\mathcal{D} \mid \theta, \hat{\widetilde{\theta}}(\theta))$$

Michael Betancourt, Generative Modellin PhyStat-ν 2016 IPMU

Generative Modelling

Bayesian Treatment

- A single likelihood which connects all the parameters to the data

---

- "Decoupled science from statistics"
- Can apply any statistical procedure, given the likelihood

- **Q: Can we introduce these ideas more explicitly in our thinking?**

Bayesian inference builds upon frequentist inference by treating the data *and* the parameters as uncertain.

$$\pi(\mathcal{D} \mid \theta)$$

Adding a prior distribution "inverts" the likelihood, identifying the parameters consistent with the data.

$$\pi(\theta \mid \mathcal{D}) \propto \pi(\mathcal{D} \mid \theta)\, \pi(\theta)$$

Relevant parameters are isolated by marginalizing nuisance parameters out of the posterior.

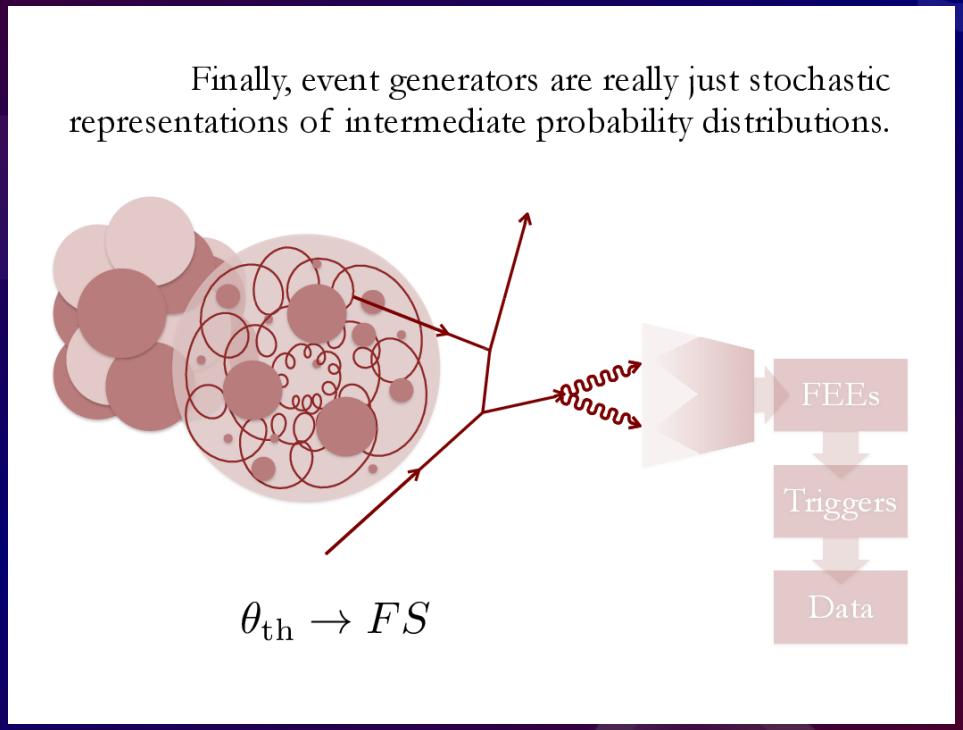
$$\pi(\theta \mid \mathcal{D}) = \int \pi(\theta, \widetilde{\theta} \mid \mathcal{D}) \mathrm{d}\widetilde{\theta}$$

Michael Betancourt, Generative Modelling PhyStat-ν 2016 IPMU

Generative Modelling

Event Generators and Likelihood Factors

- Generators take some theoretical parameters and generate an ensemble of random final states samples
- Samples are a representations of a probability distribution
- The likelihood distribution inherently exists

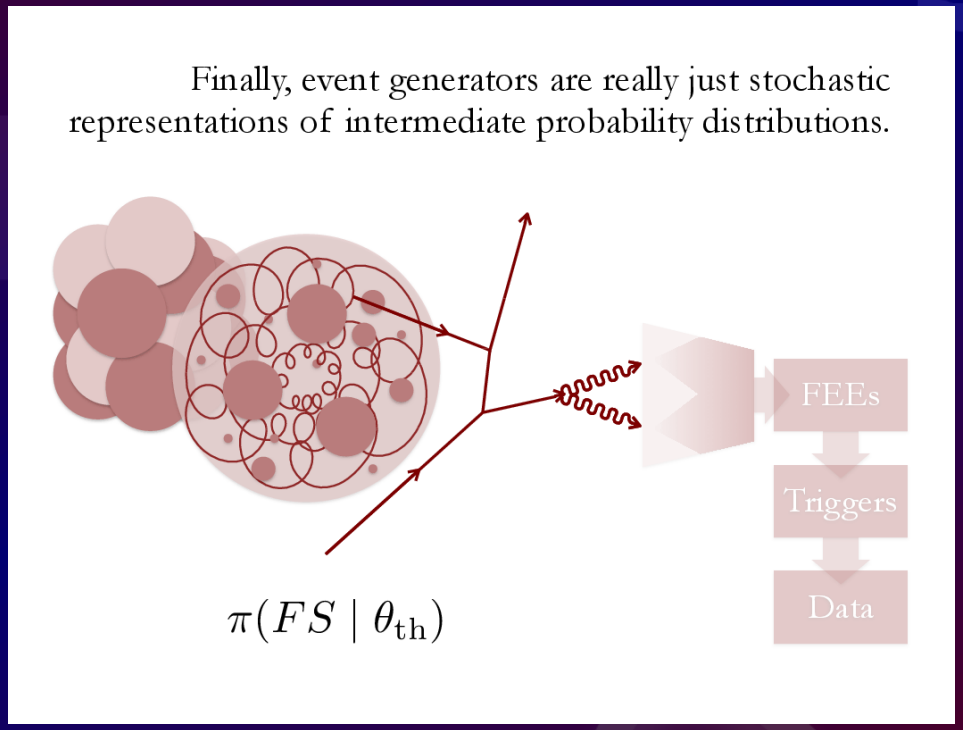


Michael Betancourt, Generative Modelling PhyStat-ν 2016 IPMU

Generative Modelling

Event Generators and Likelihood Factors

- Generators take some theoretical parameters and generate an ensemble of random final states samples
- Samples are a representations of a probability distribution
- The likelihood distribution inherently exists



Michael Betancourt, Generative Modelling PhyStat-ν 2016 IPMU

- This distribution is an approximation to the probabilities produced by the quantum mechanical processes between the neutrino, nucleus and final state
- Approximations have uncertainties
- Ideally, these uncertainties would be represented in the likelihoods, and hence the generators

Direct Mass Measurements

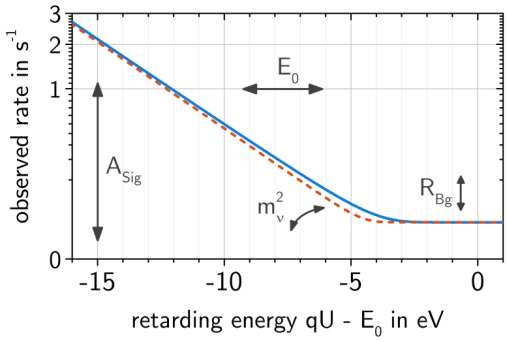
Measurement Parameters

Standard 4 parameter model fit

- Number of observed events N_i at retarding step i with retarding voltage U_i (recent setups):

$$N_i \propto t_i \cdot \left[A_{\mathrm{sig}} \cdot \int_{qU_i}^{\infty} R(U_i) \cdot \frac{d\Gamma}{dE}\left(m_{\nu_e}^2, E_0\right) dE + R_{\mathrm{bg}} \right]$$

- Parameter of interest:
  - Squared neutrino mass $m_{\nu_e}^2$
- Nuisance parameters:
  - Tritium endpoint energy E_0
  - Signal amplitude A_{sig}
  - Mean background rate R_{bg}



- $A_{\mathrm{sig}}, E_0, R_{\mathrm{bg}}$ are correlated with $m_{\nu_e}^2$ and apriori not known well enough

12

Marco Kleesiek, Direct Neutrino Mass Measurements, PhyStat-ν 2006 IPMU

- Shape of spectrum gives m_ν^2 information
- But at least three other poorly-known nuisance parameters

Direct Mass Measurements

Systematics in Frequentist $\Delta\chi^2$ Analyses

Data analysis – systematic errors

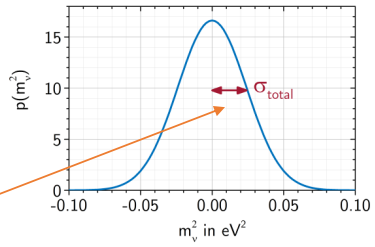
- Common practice for systematic errors:
  - Evaluate shift of the estimate $\Delta\hat{m}_\nu^2$ for each systematic effect / variable + covariances using Monte Carlos
  - Add systematic errors in quadrature, incl. correlation terms:

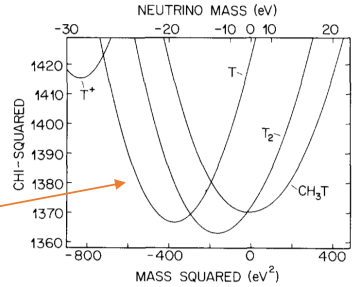
$$\sigma_{\text{syst}}(m_\nu^2) = \sqrt{\Delta_1^2 + \Delta_2^2 + \rho(\Delta_1\Delta_2) + \cdots}$$

  - Combine statistical and systematic error in quadrature:

$$\sigma(m_\nu^2) = \sqrt{\sigma_{\text{stat}}^2 + \sigma_{\text{syst}}^2}$$

- Publish distinct results if a systematic is 'not parameterizable'
  - E.g. different final state calculations





Fritschi et al. (1986) Zürich

16

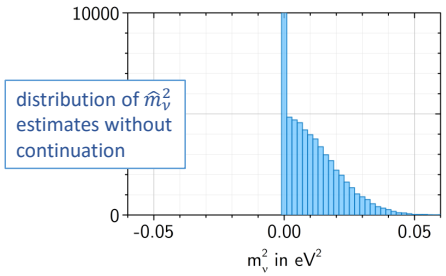
Marco Kleesiek, Direct Neutrino Mass Measurements, PhyStat-ν 2006 IPMU

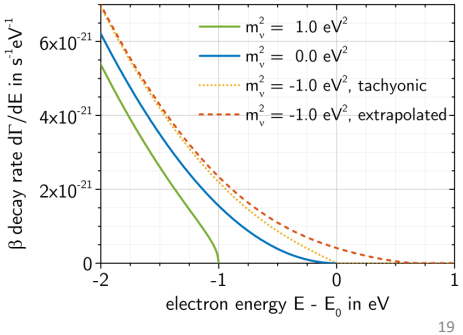
- Evaluate shifts in m_ν^2 when each systematic is varied, and add in quadrature
- Different (non-parametrised) results for different daughter nuclei have been published in the past

Direct Mass Measurements

Issue of a Negative Mass-Squared

- Required to reflect statistical fluctuations of observed data
  - χ^2 statistic / goodness-of-fit
  - Frequentist confidence interval construction
- Demands arbitrary non-physical continuation of the model
  - Symmetrical continuous likelihood / χ^2 around minimum
- Not required in bayesian methods
  - Choose prior with $m_\nu^2 \geq 0$.





19

Marco Kleesiek, Direct Neutrino Mass Measurements, PhyStat-ν 2006 IPMU

- Discussion regarding m_ν^2 being allowed to be negative or not
- Need to clearly distinguish the experimental fit statistic which can go negative, with the model parameter m_ν^2 which is not defined below 0
- Louis' "flux of solar neutrinos" and "temperature of the Sun"
- Bayesian analyses used more for cases where sterile neutrinos are included etc., but overall framework is consistent if it it were used for standard neutrino mass measurement

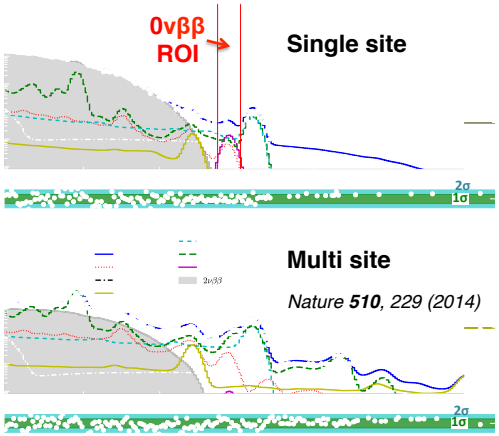
Search for Neutrinoless Double Beta Decay

Bump-Hunting

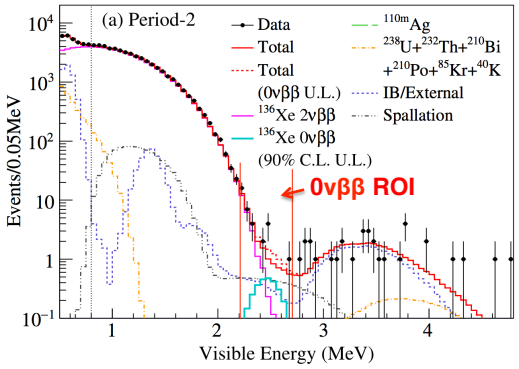
0νββ detection techniques (cont'd)

- Alternative technique uses large, self-shielded detectors (EXO, KamLand-Zen, SNO+)
- Poorer energy resolution, but multidimensional fit to event location, topology, and energy used to constrain 0νββ

EXO-200 spectrum (2014):



KamLAND-Zen spectrum (2016):



PRL 117, 082503 (2016)

David Moore, Statistical Issues for Neutrinoless Double Beta Decay, PhyStat-ν FNAL

- Discovery of neutrinoless double-beta decay would be a paradigm-shifting discovery in our field
- Looking for a "bump" at the end of the total energy spectrum
- Blind analyses are commonly used
- Valuable data highlights importance of analysis methods

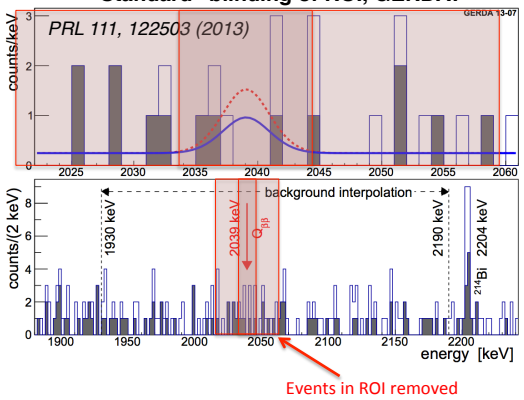
Search for Neutrinoless Double Beta Decay

Bump-Hunting

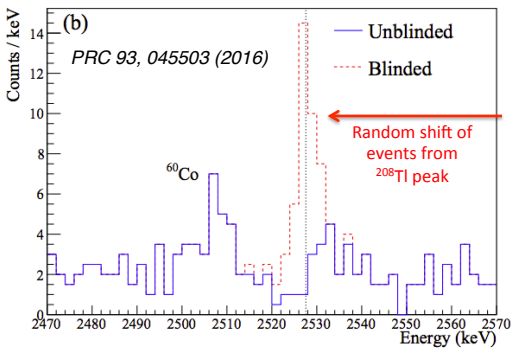
Blinding

- All experiments are looking for a small number of events (typically <10 per dataset), so most implement blinding to avoid any statistical bias
- Standard method removes all events in the 0νββ ROI
- Alternative method "salts" an artificial peak at the Q-value, which allows data quality issues to be identified

"Standard" blinding of ROI, GERDA:



Example of "salting" technique, CUORE:



David Moore, Statistical Issues for Neutrinoless Double Beta Decay, PhyStat-ν FNAL

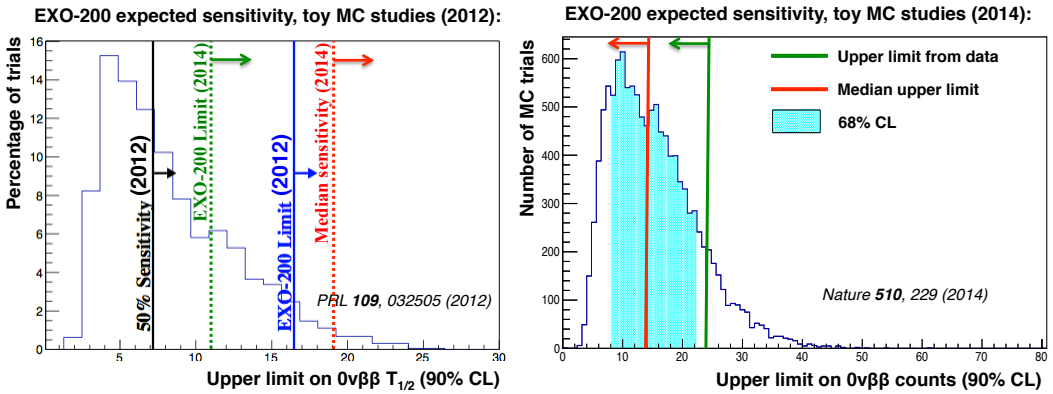
- Discovery of neutrinoless double-beta decay would be a paradigm-shifting discovery in our field
- Looking for a "bump" at the end of the total energy spectrum
- Blind analyses are commonly used
- Valuable data highlights importance of analysis methods

Search for Neutrinoless Double Beta Decay

Bump-Hunting

Statistical Issues

- Many statistical issues for 0νββ (typically <10 events/dataset) common to other rare event searches (see e.g. NIM A **774**, 103 [2015])
- Frequentist confidence intervals are most commonly presented, although some experiments report Bayesian credible intervals with flat prior
- Large statistical fluctuations common, e.g. EXO-200:



David Moore, Statistical Issues for Neutrinoless Double Beta Decay, PhyStat-ν FNAL

- Discovery of neutrinoless double-beta decay would be a paradigm-shifting discovery in our field
- Looking for a "bump" at the end of the total energy spectrum
- Blind analyses are commonly used
- Valuable data highlights importance of analysis methods

Search for Neutrinoless Double Beta Decay

Bump-Hunting

- Discovery of neutrinoless double-beta decay would be a paradigm-shifting discovery in our field
- Looking for a "bump" at the end of the total energy spectrum
- Blind analyses are commonly used
- Valuable data highlights importance of analysis methods

The Meaning of n–Sigma Significances

Search	Degree of surprise	Impact	LEE	Systematics	Number of σ
Higgs search	Medium	Very high	Mass	Medium	5
Single top	No	Low	No	No	3
SUSY	Yes	Very high	Very large	Yes	7
B_s oscillations	Medium/low	Medium	Δm	No	4
Neutrino oscillations	Medium	High	$sin^2(2\theta), \Delta m^2$	No	4
$B_s \to \mu\mu$	No	Low/Medium	No	Medium	3
Pentaquark	Yes	High/very high	M, decay mode	Medium	7
$(g-2)_\mu$ anomaly	Yes	High	No	Yes	4
H spin $\neq 0$	Yes	High	No	Medium	5
4th generation q, l, ν	Yes	High	M, mode	No	6
$v_\nu > c$	Enormous	Enormous	No	Yes	>8
Dark matter (direct)	Medium	High	Medium	Yes	5
Dark energy	Yes	Very high	Strength	Yes	5
Grav waves	No	High	Enormous	Yes	7

Table 1: Summary of some searches for new phenomena, with suggested numerical values for the number of σ that might be appropriate for claiming a discovery.

Louis Lyons, Discovering the Significance of 5 sigma

- "Five-sigma" or one-in-1.7-million is not sacred; often infeasible to actually demonstrate without using approximations or asymptotics
- Many potential discoveries are very different from an "unexpected bump" in some spectrum
- Understanding the look-elsewhere-effect is critical
- "Two 3.5σ results are better than one 5σ result"; "Calibrated 3.5σ result better than uncalibrated 5σ"
- Suggestion to discuss beforehand, in the community, what sigma-values (or equivalent) are appropriate

Experimental Design

Bayesian Decision Theory

Bayesian experimental design

- Uncertainties predicting data and capabilities of future experiment
  - Which model?
  - Which parameter values?
  - Which data realization?
- *Expected utility* of experiment e

$$\mathrm{EU}(e) = \int \Pr(D_f|D_p, e)U(D_f, e)\mathrm{d}D_f$$

- Utility U = usefulness of observed data D_f in experiment e
- Future data D_f predicted using present data D_p, e.g.,

$$\Pr(D_f|D_p, e) = \int \Pr(D_f|\boldsymbol{\Theta}, e)\Pr(\boldsymbol{\Theta}|D_p)\mathrm{d}\boldsymbol{\Theta}$$

  – Minimizing prior sensitivity
- Frequentist requires performing tons of analyses for fixed parameter values – no way to combine them

Johannes Bergstrom, Three-Neutrino Mixing Fits

- Here $\boldsymbol{e}$ is an experiment, $\boldsymbol{D_p}$ is the present data, which is used to predict the future data, $\boldsymbol{D_f}$, using the posterior the parameters $\boldsymbol{\Theta}$
- Integrating over the parameters is key to obtaining the prediction for the future data for an experiment

Experimental Design

Bayesian Decision Theory

Bayesian experimental design

- Uncertainties predicting data and capabilities of future experiment
  - Which model?
  - Which parameter values?
  - Which data realization?
- *Expected utility* of experiment e

$$\mathrm{EU}(e) = \int \Pr(D_f|D_p, e) U(D_f, e) \mathrm{d}D_f$$

- Utility U = usefulness of observed data D_f in experiment e
- Future data D_f predicted using present data D_p, e.g.,

$$\Pr(D_f|D_p, e) = \int \Pr(D_f|\Theta, e) \Pr(\Theta|D_p) \mathrm{d}\Theta$$

  – Minimizing prior sensitivity
- Frequentist requires performing tons of analyses for fixed parameter values – no way to combine them

Johannes Bergstrom, Three-Neutrino Mixing Fits

- Here e is an experiment, D_p is the present data, which is used to predict the future data, D_f, using the posterior the parameters Θ
- Integrating over the parameters is key to obtaining the prediction for the future data for an experiment

Experimental Design

Bayesian Decision Theory

Choice of utility

- Estimation
  - Size of error/Fisher matrix
  - Information gain on parameters
- Testing/signal detection
  - $\Delta\chi^2$
  - Prob observing $\Delta\chi^2 > C$
  - $\log B$
  - Prob observing $\log B > C$
  - Information gain on model space (MO!)

$$\sum_{O=\mathrm{NO,IO}} \Pr(O|\mathbf{D}) \log \frac{\Pr(O|\mathbf{D})}{\Pr(O)}$$

Trotta, Kunz, Liddle: arXiv:1012.3195
Loredo: astro-ph/0409386
von Toussaint, Rev. Mod. Phys. 83 (2011)

Johannes Bergstrom, Three-Neutrino Mixing Fits

- The "Utility" is what determines the final Expected Utility of an particular experiment
- Formal design arguments also made in context of JUNO and sensitivity to the Mass Hierarchy

- Blind Analyses
  - Mentioned explicitly in the context of IceCube at both IPMU and Fermilab, where there was a dedicated talk, and Double Beta-Decay
  - Most earlier discovery experiments did not use blind analyses
  - Currently not being used universally—but important to consider for discovery physics, given how valuable our data are
- Machine Learning
  - At IPMU, only mentioned in COMET ($\mu^- + N \rightarrow e^- + N$) context; discussion about difficulty to incorporate into a Bayesian analysis unless generative modelling is used
  - At Fermilab some MINERvA examples shown as part of a general talk
- Bayesian Event Reconstruction
  - Described in context of Super-Kamiokande at IPMU
  - Non-parametric—number of parameters is variable
- Computational Methods and Tools
  - MCMC methods: Hamiltonian, Affine Invariant Parallel Tempering, Reversible Jump etc.
  - BUGS/OpenBUGS: a system for specifying Bayesian models and generating MCMC posterior samples
  - Stan: statistical modelling and high-performance statistical computation

PHYSTAT-ν

As an Ongoing Forum

- In collider experiments, PHYSTAT is associated with experiments' Statistics Committees
  - answering questions and providing guidance to analysers
- In neutrino physics, work is spread across a larger number of smaller collaborations
- PHYSTAT-ν could provide a shared forum for questions for analysers and discussions, and advice from experts
- Possibilities include:
  - Regular seminars, tutorials, perhaps jointly across all PHYSTATs
  - Constant electronic presence
  - "Good practice" document maintained by PHYSTAT-ν experts, tailored to current experiments' needs
  - Occasional workshops such as this one
  - ....
- The organisers are open to suggestions!