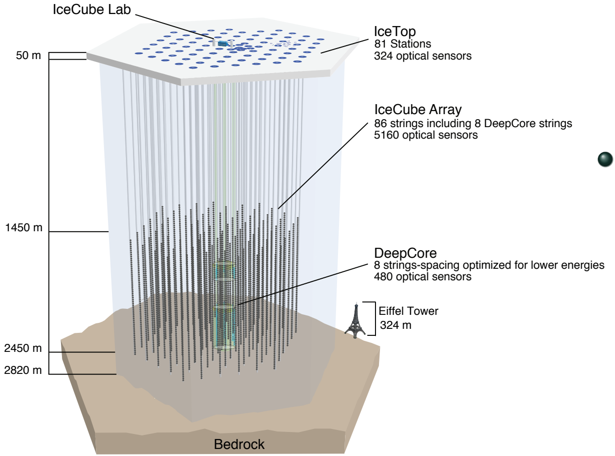
Statistical Approaches for IceCube, DeepCore, and PINGU Neutrino Oscillation Analyses

Joshua Hignight
for the IceCube-PINGU Collaboration

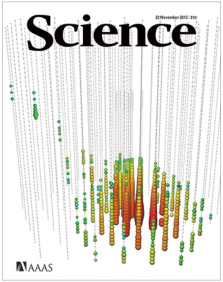
MICHIGAN STATE
UNIVERSITY

September 21^{st}, 2016

IceCube

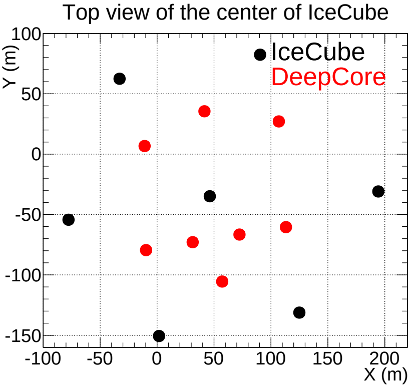


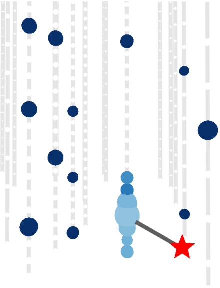
- Without DeepCore: 78 strings, 125 m string spacing, 17 m module vertical-spacing
- Optimized for (very) High Energy neutrinos



IceCube-DeepCore

- 78 strings, 125 m string spacing
- 17 m modules vertical-spacing
- 8 strings, 40-75 m string spacing
- 7 m modules vertical-spacing

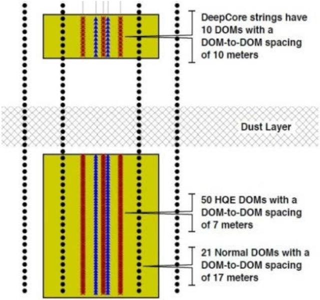




→ Typical LE ν event

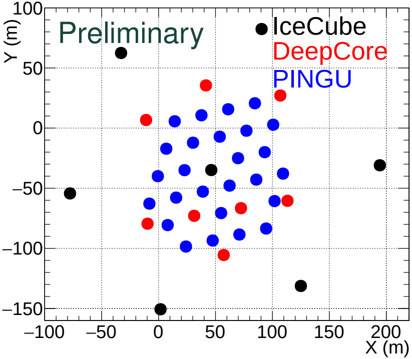
→ $E_{\nu_\mu} = 12$ GeV

(w/ $E_\mu = 8$ GeV)

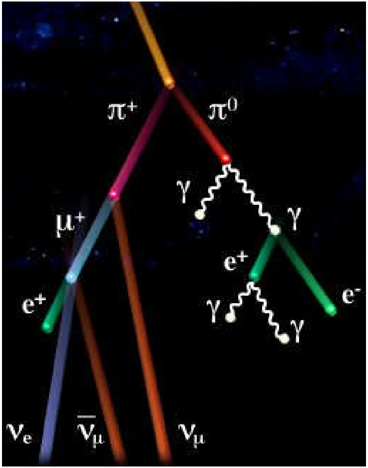


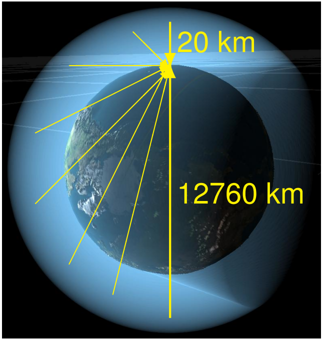
IceCube-DeepCore-PINGU

- 78 strings, 125 m string spacing
- 17 m modules vertical-spacing
- 8 strings, 75 m string spacing
- 7 m modules vertical-spacing
- 26 strings, 24 m string spacing
- 1.5 m modules vertical-spacing
  - all optical modules in clearest ice



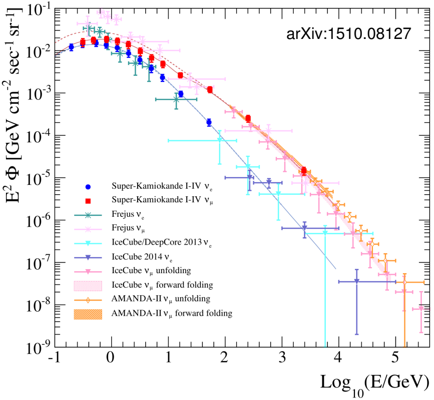
Atmospheric neutrinos

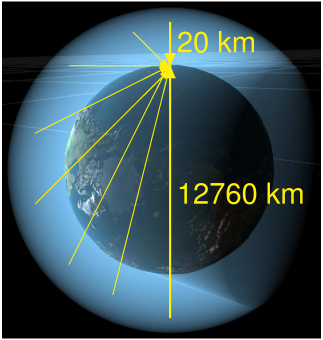




- 2:1 ratio between ν_μ:ν_e
- similar rate of ν and $\bar{\nu}$
  - however, x-sec for $\bar{\nu}$ half of ν
- various baselines (L) available

Atmospheric neutrinos

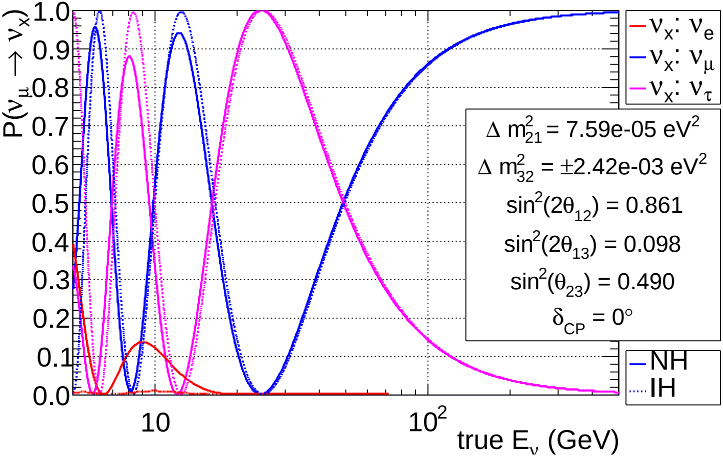




- ν energy over several orders of magnitude
- various baselines (L) available

$\Rightarrow$ wide range of L/E available for ν oscillation measurements

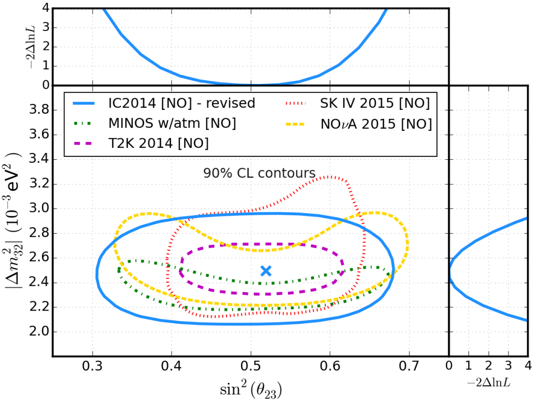
Atmospheric neutrino oscillations



- Longest baseline (L=12760 km, $\cos\theta_z = -1$) has:
  - First oscillation maxima at $\sim$ 25 GeV
  - Matter effects below $\sim$ 12 GeV
  - Potential for ν_e appearance at 8 GeV

“Atmospheric mixing” parameters by IceCube

IceCube: Phys.Rev. D 91, 072004 (2015); SK: AIP Conf. Proc. 1666, 100001 (2015)



- IceCube: fitting to data done in 2D space (E, θ_z)
  - $\chi^2/ndf = 54.9/56$

- Contours obtained using Wilk’s theorem
  - Calculate $\Delta \ln L = \ln L - \ln L_{bestfit}$ for all points in 2D parameter space
  - $\Delta \ln L$ calculated by maximizing L over nuisance parameters
  - $-2\Delta \ln L$ is asymptotically a χ^2 distribution with 2 dof.
- Side plots: profile of $\Delta \ln L$ passing through best-fit

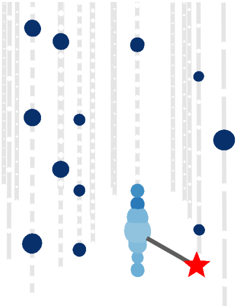
IceCube – towards future analysis

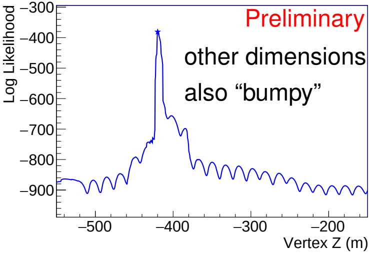
- PRD analysis focus in ν_μ CC "clean" events
  - Clear μ tracks
  - Require several non-scattered γ
  - Use only up-going events $\Rightarrow$ very small atmospheric μ contamination
  - Fits analytical formula for Cherenkov light front propagated to PMTs
- Currently working on new analysis based on new reconstruction:
  - Planning to look at full-sky
    - More atmospheric μ contamination
    - But would give us better handle on flux systematics
  - Use information from all hits in reconstruction
    - Reconstruction more sensitive to scattering
    - Unfortunately also more sensitive to noise
  - Increased presence of ν_e, ν_τ and ν NC in sample
  - And also increase significantly number of ν_μ events at final level
    - significant improvement in final result expected

IceCube – new event reconstruction

- IceCube measures Cherenkov cones in “3D”
- PMTs embedded in the parameter space creates features in their vicinity
- Natural medium also has local variations
- Low number of hits

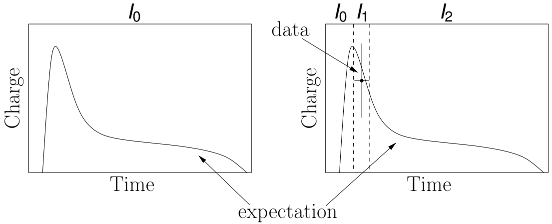
$\rightarrow$ “bumpy” likelihood space





- Need to fit 8 parameters corresponding to ν_μ DIS interaction
  - vertex (3), time, direction (2), energies of μ and hadronic cascade
- Usual minimizers do not work well
  - currently using “MultiNest”

The "event" likelihood space



$$L(D|H) = \prod_{i \in \{PMT\}} \prod_{j=0}^{n_i} Poisson\left(\int_{t \in I_j} Q^{obs} dt, \int_{t \in I_j} Q^{exp} dt\right)$$

- Charge expectation (Q^{exp}) distribution from spline tables
  - Spline tables account for main local/global ice properties
  - Derived from simulation
- Idea for the future: replace tables by simulated expectations
  - For every $L(D|H)$ calculation run simulation to estimate expectation
  - Can account of more detailed/evolving ice models

The MultiNest algorithm

See full description in paper by F. Feroz et al. [arXiv:0809.3437 and arXiv:1306.2144]

- MultiNest searches for maximum in multidimensional likelihood space
  - Exploration of space via ellipsoidal nested sampling
    - ★ New trials thrown in volume defined by ellipsoids obtained from distribution of previous trials $\rightarrow$ efficient sampling
    - ★ New trials accepted/rejected depending on their LH
    - ★ Posterior distributions provided could be used as error estimates
  - Natively supports multi-modal distributions
    - ★ In our case important to avoid local minima

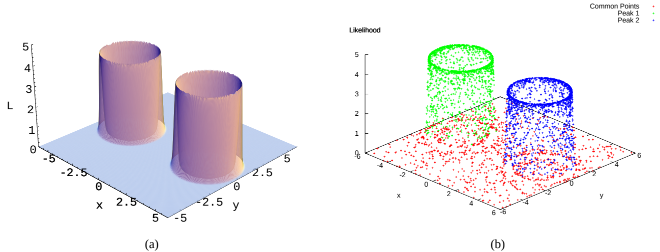
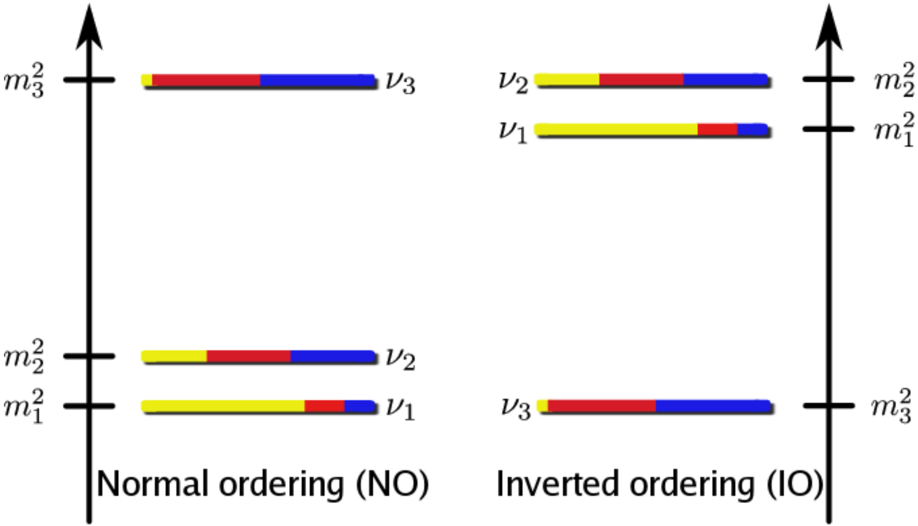


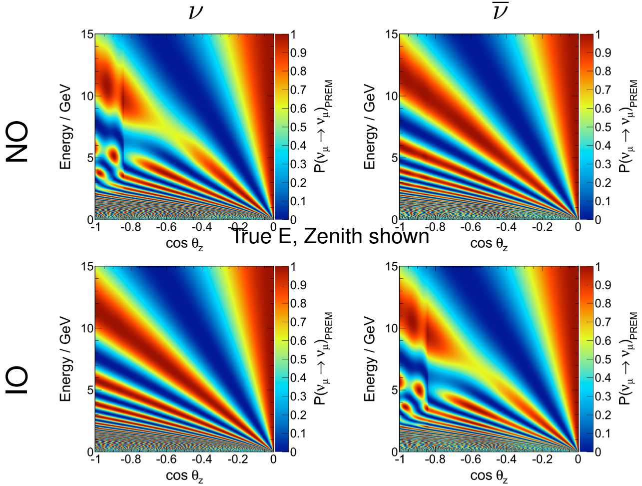
Figure 6. Toy model 2: (a) two-dimensional plot of the likelihood function defined in Eqs. (32) and (33); (b) dots denoting the points with the lowest likelihood at successive iterations of the MULTINEST algorithm. Different colours denote points assigned to different isolated modes as the algorithm progresses.

(Figure extracted from arXiv:0809.3437)

Measuring the ν Mass Ordering with atmospheric ν



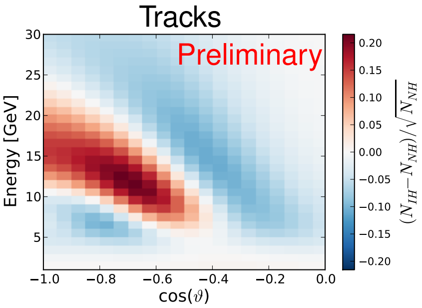
Measuring the ν Mass Ordering with atmospheric ν



- Different oscillation probabilities for ν and $\overline{\nu}$ for NO and IO
- Measure combined ν+$\overline{\nu}$
  - different cross-section $\Rightarrow$ effect doesn't vanish

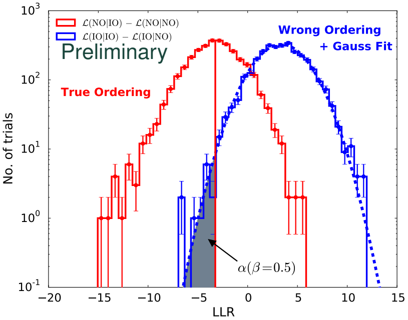
Bin-by-bin significance of mass hierarchy signature

Assuming no ν vs $\overline{\nu}$ identification



- Distinct hierarchy dependent signatures for tracks (mostly ν_μ CC) and cascades (mostly ν_e CC)
  - Intensity is statistical significance of each bin with 1 year data
  - Measurement is possible "statistically" by combining all bins – there is not one bin that would achieve that
  - Particular expected "distortion pattern" helps mitigate impact of systematics

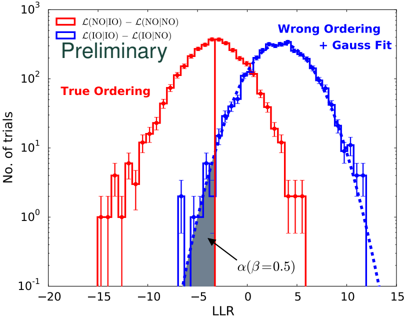
Estimating sensitivity to the NMO: Log Likelihood Ratio



1. Generate pseudo-data trial in analysis binning
   - True physics and systematics kept fixed for generation
2. Fit assuming NO and IO
3. Calculate log likelihood ratio between IO and NO

- Advantages of the method:
  - Can account for any systematic given
  - Does not pre-suppose shape of ΔLLH distribution
- Disadvantages of the method:
  - The significance “limited” by number of trials
  - Since each trial is a full fit (and given lots of trials needed) having large number of systematics can became prohibitively time consuming

Median sensitivity



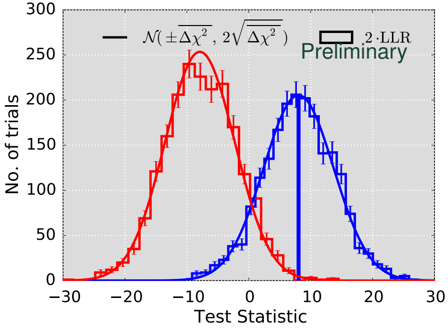
- For quantifying significance to measure ordering usually use median sensitivity
  - Widely used in literature
- "Median sensitivity" will mean that 50% of the time we can do better and 50% of the time we can do worse
- "Median sensitivity" calculated by integrating shade region under wrong ordering assumption
  - If distribution fits well Gaussian, integrate area under Gaussian curve instead of trial distribution

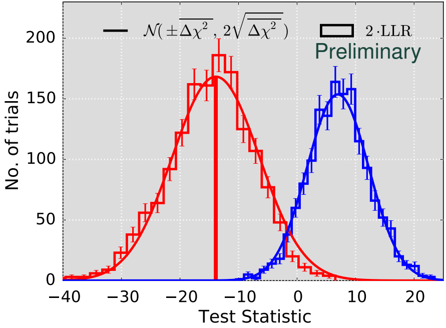
Estimating sensitivity to the NMO: $\Delta\chi^2$ method

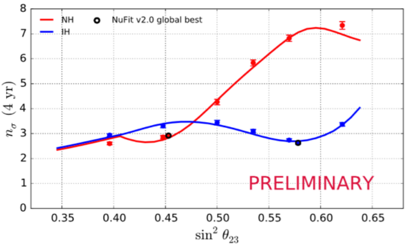
1. Get expected number of events in analysis binning
   - True physics and systematics kept fixed as in LLR method
   - But, no Poisson fluctuations applied
2. Calculate minimal $\overline{\Delta\chi^2}$ for the WO
   - $\overline{\Delta\chi^2} = \min_{p \in WO} \sum_i \left(\frac{\mu_i^{TO}(p_0) - \mu_i^{WO}(p)}{\sigma_i} \right)^2$
   - $\Delta\chi^2$ is Gaussian distributed with mean $\pm\overline{\Delta\chi^2}$ and sigma $2\sqrt{\overline{\Delta\chi^2}}$
3. Evaluate distribution of $\Delta\chi^2$ for NO and IO
   $\Rightarrow$ correspond to the LLR trial distribution

- Advantages of the method:
  - Linear systematics are extremely fast to be computed
  - Even with non-linear systematics still much faster than LLR
- Disadvantage of the method:
  - Intrinsic assumption of gaussianity of final distribution
  - Not possible to include non-centered priors

Comparing Test Statistic of LLR and $\Delta\chi^2$







- Good agreement between TS

$\Rightarrow$ sensitivities in agreement

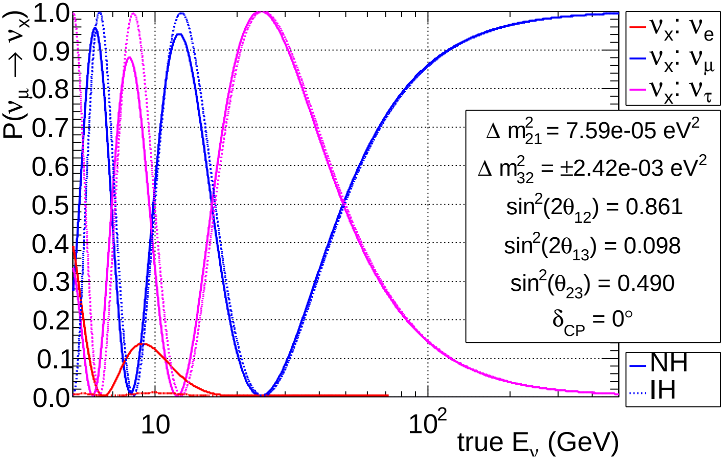
  - lines from $\Delta\chi^2$
  - points from LLR

Summary

- Various different techniques used for reconstruction
  - sometimes different tools used as minimizers:
    - MultiNest used to avoid local-minima by exploring L space
- Measurements using very different statistical techniques
  - Statistically evaluate presence of components in sample
  - From $-2\Delta \ln L$ obtain contours via Wilks theorem
  - LogLikelihood Ratio, $\Delta\chi^2$ to distinguish between hypothesis
- All these techniques used with main (physics) goal of measuring ν oscillations

Backup slides

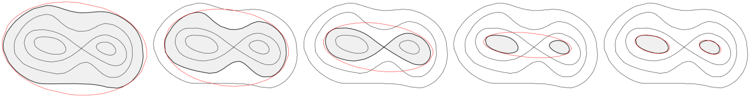
Atmospheric neutrino oscillations



- Longest baseline (L=12760 km, cos $\theta_z = -1$) has:
  - First oscillation maxima at $\sim$ 25 GeV
  - Matter effects below $\sim$ 12 GeV
  - Potential for ν_e appearance at 8 GeV

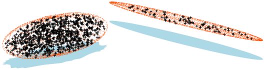
More plots from MultiNest

From arXiv:0809.3437

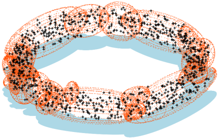


(a) (b) (c) (d) (e)

Figure 2. Cartoon of ellipsoidal nested sampling from a simple bimodal distribution. In (a) we see that the ellipsoid represents a good bound to the active region. In (b)-(d), as we nest inward we can see that the acceptance rate will rapidly decrease as the bound steadily worsens. Figure (e) illustrates the increase in efficiency obtained by sampling from each clustered region separately.

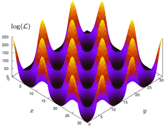


(a)

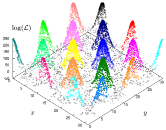


(b)

Figure 3. Illustrations of the ellipsoidal decompositions returned by Algorithm 1: the points given as input are overlaid on the resulting ellipsoids. 1000 points were sampled uniformly from: (a) two non-intersecting ellipsoids; and (b) a torus.



(a)

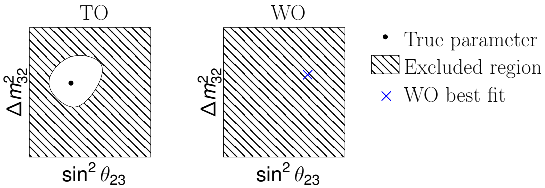


(b)

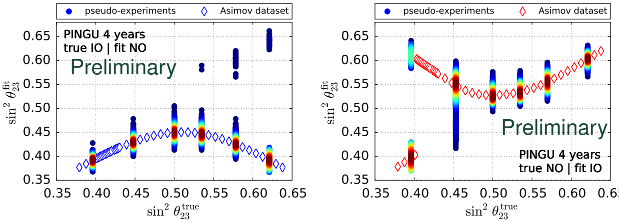
Figure 5. Toy model 1: (a) two-dimensional plot of the likelihood function defined in Eq. 31; (b) dots denoting the points with the lowest likelihood at successive iterations of the MultiNest algorithm. Different colours denote points assigned to different isolated modes as the algorithm progresses.

Excluding an ordering

- To say we measure the true ordering (TO) at a given CL we want to be able to exclude the wrong ordering (WO) for any value of the oscillation parameters



- Testing every point of the WO parameter space too costly
  - WO best-fit gives parameters of "maximum confusion" (used to get WO trial distribution)



Unfolding

- Interesting resources:
  - Presentation on unfolding in HEP: `http://mkuusela.web.cern.ch/mkuusela/ETH_workshop_July_2014/slides.pdf`
  - V. Blobel, “Unfolding Methods in High-energy Physics Experiments” at `https://cds.cern.ch/record/157405?ln=en`
  - D’Agostini, Nucl.Instrum.Meth. A362 (1995) 487-498

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:

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

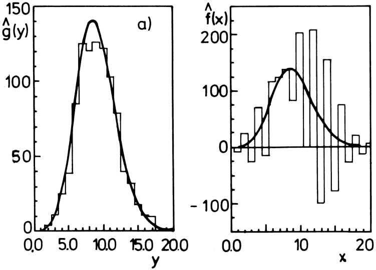


Figure 1. Distribution of the measured quantity y (a) and oscillating result of unfolding (b) using equation (1.05), shown as histograms. The original dependence is shown as a curve in both cases.

Unfolding with regularization

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

Input pdf and data

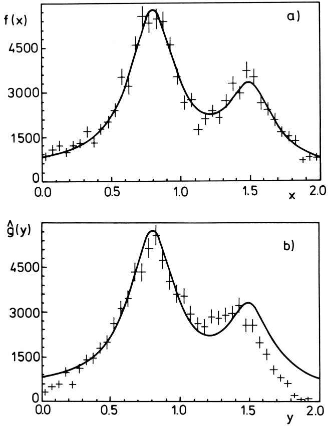


Figure 8. Histogram of the generated data (a) before and (b) after simulation of acceptance, transformation and resolution. The original function is shown as curve.

Unfolding result

- Using B-splines for regularization

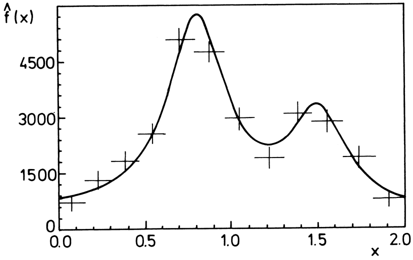


Figure 11. Result of unfolding with regularization, shown as data points together with the original function. The horizontal bar gives the range, over which the data points represents the average.

"Bayesian Unfolding" example

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

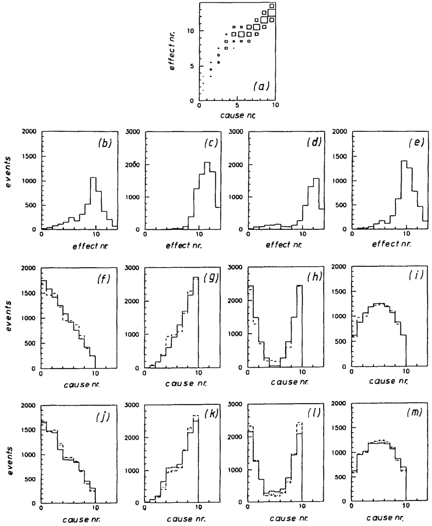


Fig. 1. (a) Visualization of the smearing matrix $\mathbf{S}_1$ (see text): in the abscissa and the ordinate are the true and the measured quantities, respectively; the box area is proportional to the migration probability. (b–e) Smeared distributions of f_{1-4} (see text) obtained from 10000 generated events. (f–i) True distributions of f_{1-4} (solid lines) compared with the results of a 3-step unfolding (dashed lines). (j–m) Unfolded distributions after the first, second and third step (solid, dotted and dashed lines) of f_{1-4}.

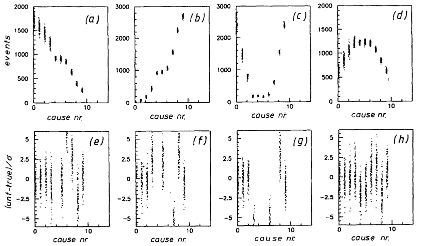


Fig. 4. Results obtained by a 3-step unfolding of 100 independent data sets, each based on 10000 generated true events smeared with $\mathbf{S}_1$: (a–b) distribution on the unfolded numbers for true distributions f_{1-4}; (e–h) distribution of the difference between the unfolded and the true numbers, divided by the extimated standard deviation, for true distributions f_{1-4}.

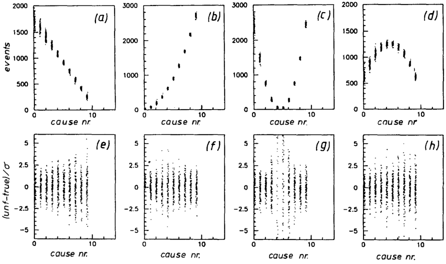


Fig. 8. Same as Fig. 4, but with a 20-step unfolding and smoothing the probability distribution between the steps.