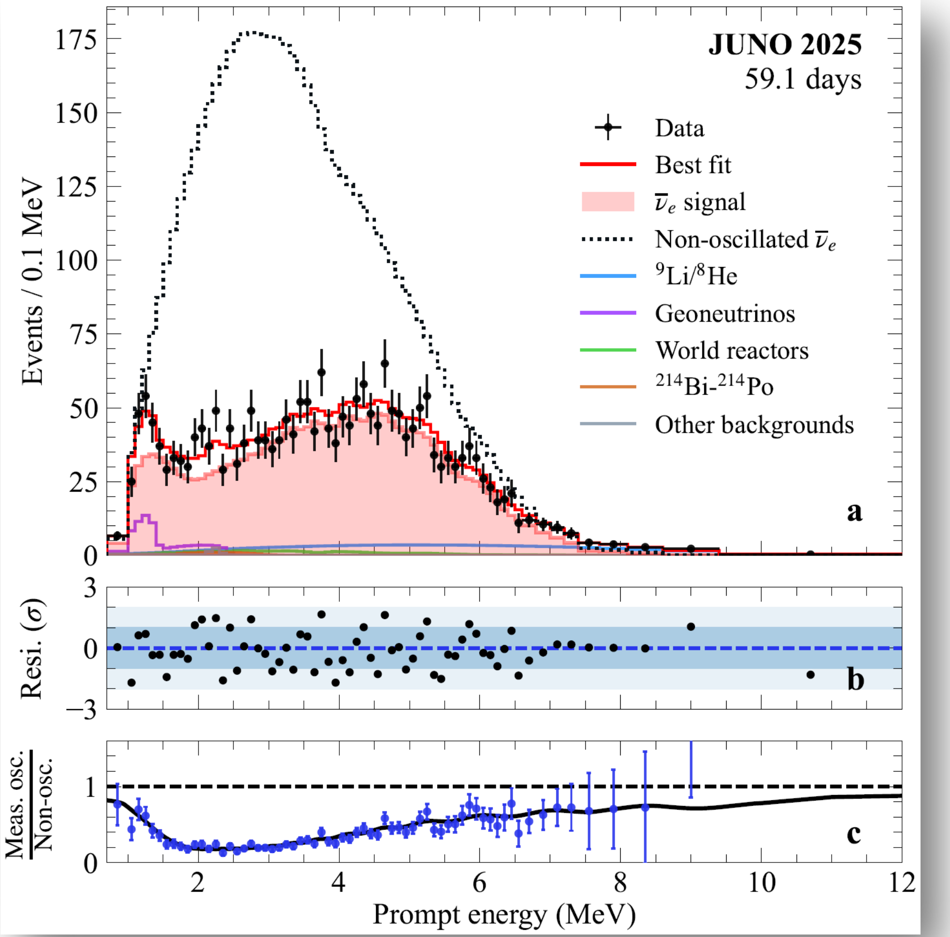
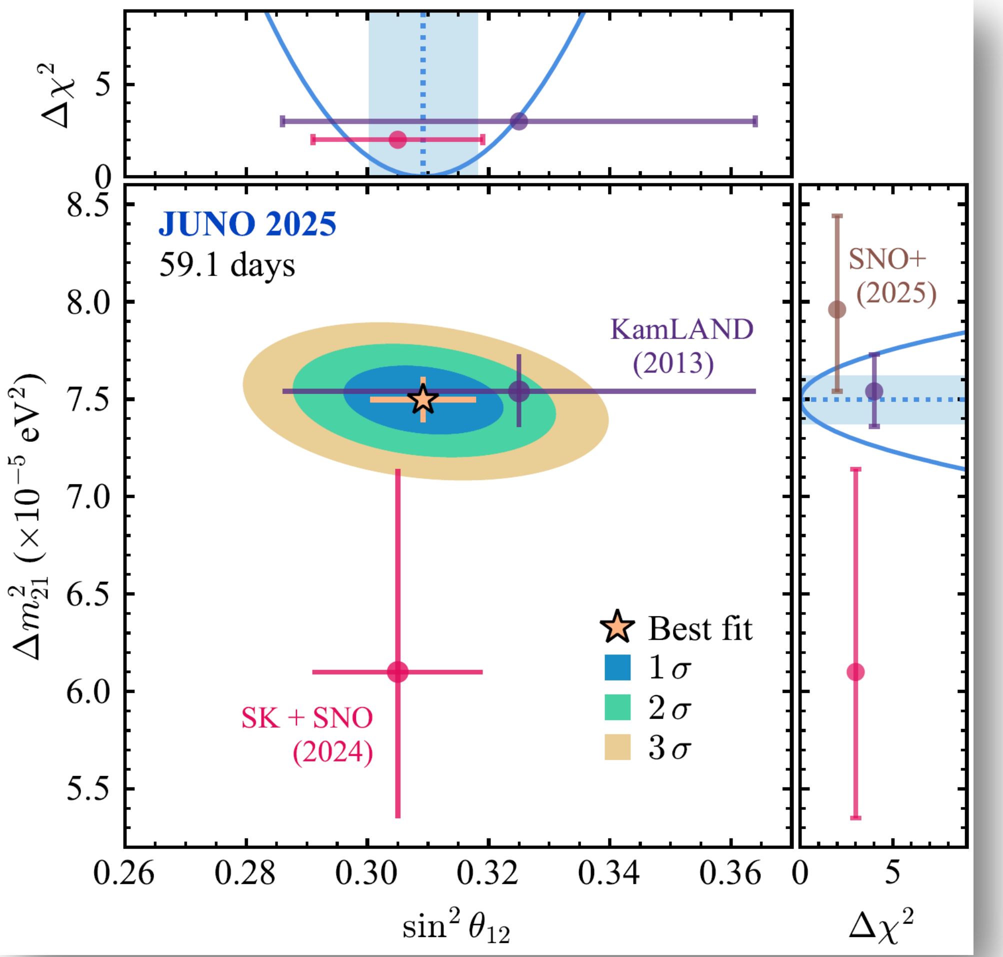
数据分析基础练习

凌家杰

2026年南山清北暑期学校

JUNO实验的首个中微子振荡结果





中微子振荡基础

味道本征态 $\quad |\nu_\alpha\rangle = \sum_{i=1}^{3} U^*_{\alpha i}\,|\nu_i\rangle \quad$ 质量本征态

$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{bmatrix} \begin{bmatrix} c_{13} & 0 & s_{13}e^{-i\delta} \\ 0 & 1 & 0 \\ -s_{13}e^{i\delta} & 0 & c_{13} \end{bmatrix} \begin{bmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

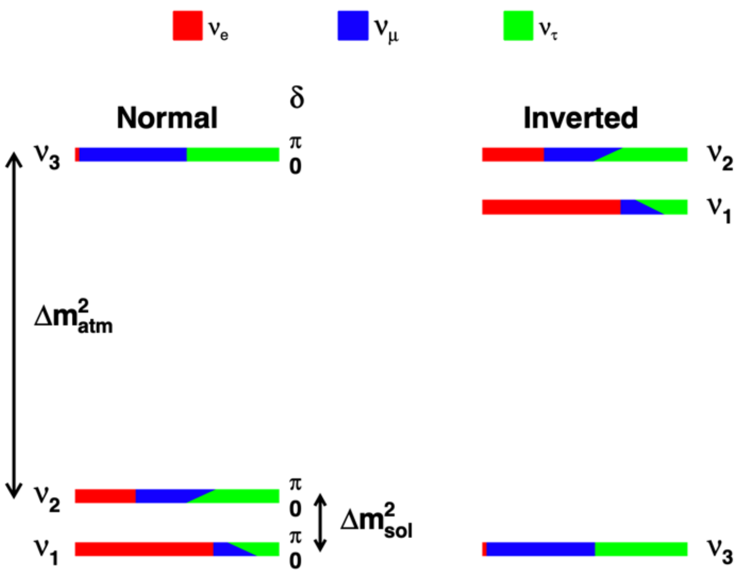
$c_{ij} = \cos\theta_{ij} \qquad s_{ij} = \sin\theta_{ij}$

三个混合角: $\theta_{23}, \theta_{13}, \theta_{12}$

$c_{12}^2 \equiv \cos^2\theta_{12} = \frac{|U_{e1}|^2}{1-|U_{e3}|^2} \qquad s_{12}^2 \equiv \sin^2\theta_{12} = \frac{|U_{e2}|^2}{1-|U_{e3}|^2}$

$s_{13}^2 \equiv \sin^2\theta_{13} = |U_{e3}|^2$

$s_{23}^2 \equiv \sin^2\theta_{23} = \frac{|U_{\mu 3}|^2}{1-|U_{e3}|^2} \qquad c_{23}^2 \equiv \cos^2\theta_{23} = \frac{|U_{\tau 3}|^2}{1-|U_{e3}|^2}$



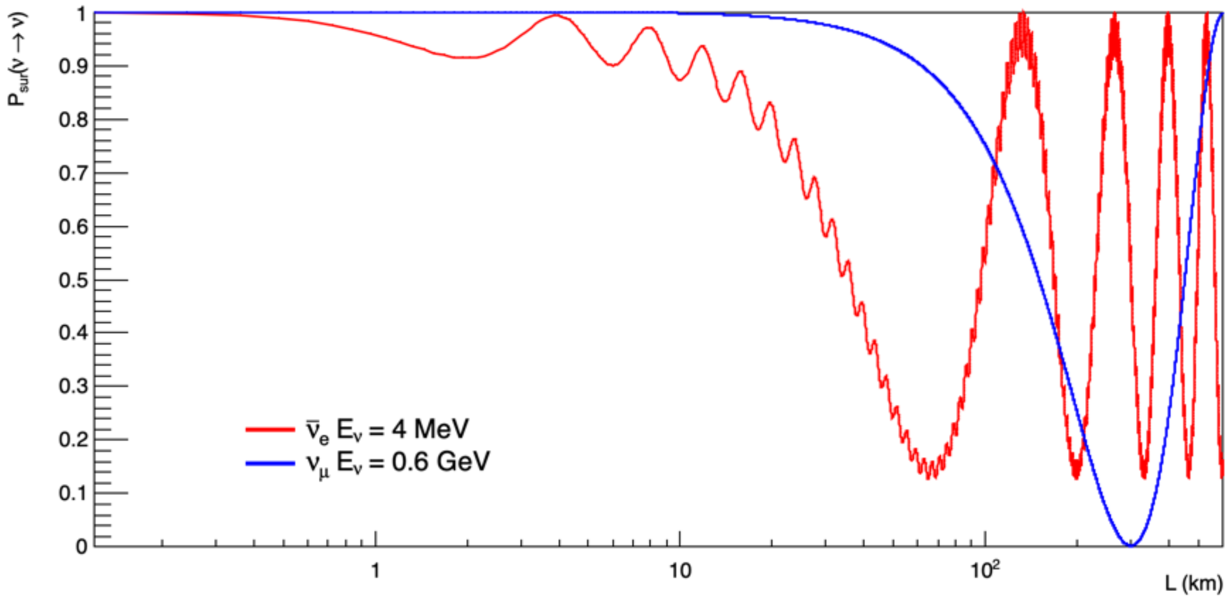
中微子振荡概率

$$U = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}$$

$$P_{\alpha\beta} = \left|\langle \nu_\beta | \nu_\alpha(t)\rangle\right|^2 = \delta_{\alpha\beta} - 4\sum_{i<j}^{3} Re\left[U_{\alpha i}U_{\beta i}^{*}U_{\alpha j}^{*}U_{\beta j}\right]\sin^2\Delta_{ij} + 2\sum_{i<j}^{3} Im\left[U_{\alpha i}U_{\beta i}^{*}U_{\alpha j}^{*}U_{\beta j}\right]\sin 2\Delta_{ij}$$

$$\Delta_{ij} = \Delta m_{ij}^2 \frac{L}{4E}$$

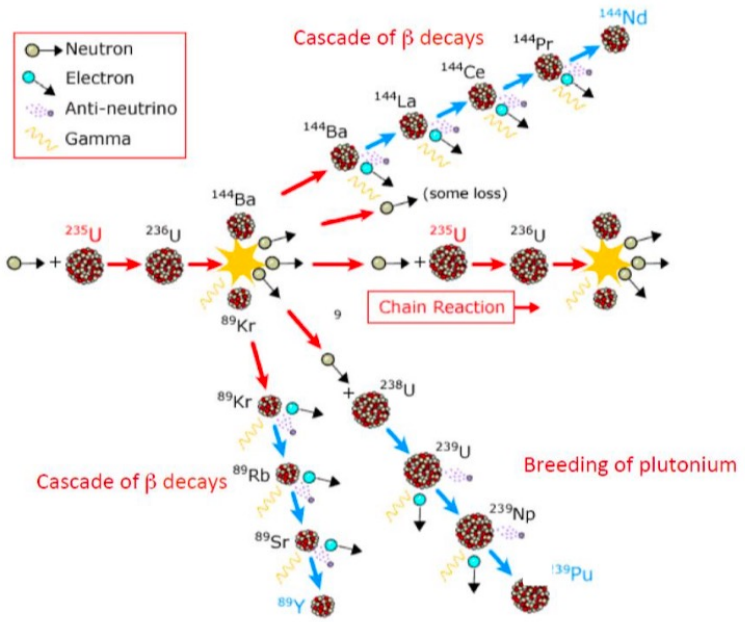
$$P_{\alpha\alpha} = 1 - 4|U_{\alpha 1}|^2|U_{\alpha 2}|^2 sin^2\left(\frac{\Delta m_{21}^2 L}{4E}\right) - 4|U_{\alpha 1}|^2|U_{\alpha 3}|^2 sin^2\left(\frac{\Delta m_{31}^2 L}{4E}\right) - 4|U_{\alpha 2}|^2|U_{\alpha 3}|^2 sin^2\left(\frac{\Delta m_{32}^2 L}{4E}\right)$$



反应堆中微子的产生

反应堆中微子源：U/Pu的裂变级联过程

- 常见商用反应堆功率：2.9 GW
- $1\ GW = 6.2415 \times 10^{21}\ MeV$
- ~ 200 MeV / fission， 6 $\bar{\nu}_e$ / fission
- ~ 2 x 10^{20} $\bar{\nu}_e$/GW_{th}/Sec (1/5 above IBD threshold)

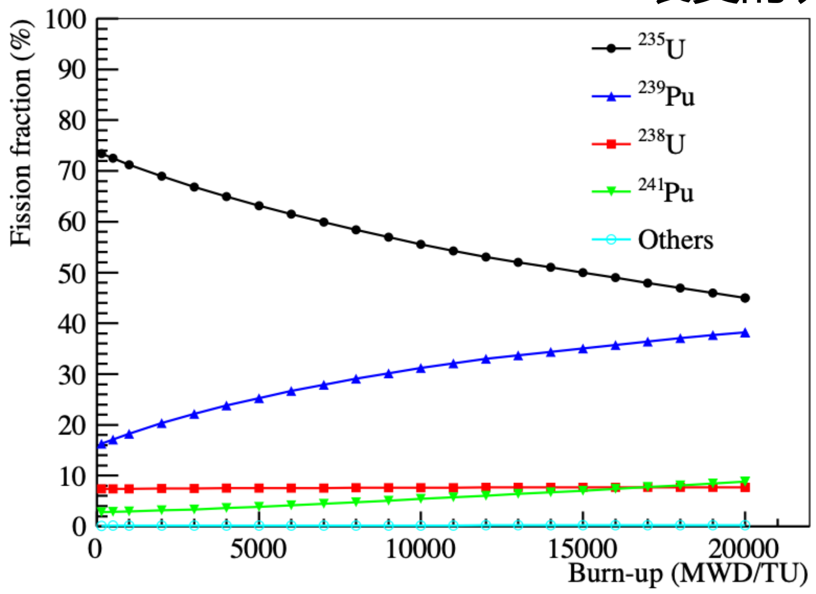


裂变产生的中微子数 $$S(E_\nu) = \sum_i^{isotopes} f_i S_i(E_\nu)$$

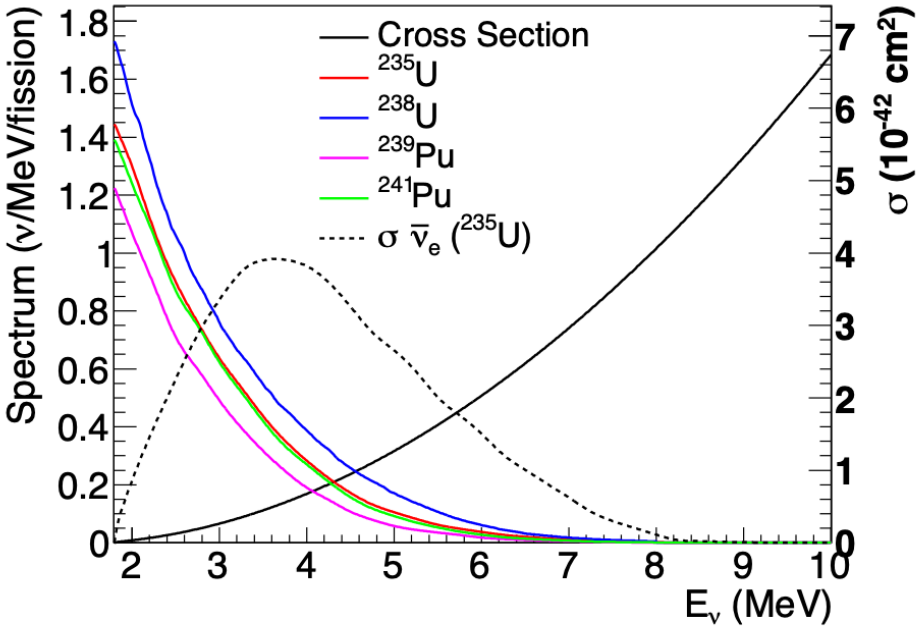
每秒产生的中微子数 $$\boxed{S(E_\nu) = \frac{W_{th}}{\sum_i (f_i/F) e_i} \sum_i^{istopes} (f_i/F) S_i(E_\nu)}$$

裂变的热功率 $$W_{th} = \sum_i f_i e_i\ , \quad F = \sum_i f_i$$

裂变的次数



中微子反应



1 IBD / (ton•GW•day) @ 1km

反贝塔衰变 (IBD) $\bar{\nu}_e + p \rightarrow n + e^+$

- 快信号: $E_{\text{prompt}} \approx E_\nu - 0.78\ MeV$
- 慢信号: nH (2.2 MeV)

$$E_\nu = E_{e^+} + m_n + T_n - m_p \approx E_{prompt} + 0.78\ MeV,$$

作业1： 中微子振荡曲线

- $sin^2\theta_{12} = 0.3$
- $sin^2\theta_{13} = 0.02$
- $\Delta m_{21}^2 = 7.5\times10^{-5} eV^2$
- $\Delta m_{32}^2 = 2.4\times10^{-3} eV^2$
- $L = 52.5\ km$
- $E: 0 - 12\ MeV$

- 画出振荡概率和能量的关系图

作业1: 预测中微子能谱

$$\frac{dN}{dE_{\rm vis}^{\rm obs}} = \frac{N_p T}{4\pi L^2}\int_{E_{\rm thr}}^{\infty} dE_\nu \frac{dN}{dE_\nu} P_{ee}(L, E_\nu) \times \sigma_{\rm IBD}(E_\nu)\, G(E_\nu - 0.8{\rm MeV} - E_{\rm vis}^{\rm obs}, \delta E_{\rm vis})$$

Neutrino Flux per fission:

$$\begin{aligned}\phi(E_\nu) = \; & f_{^{235}U} \exp\left(0.870 - 0.160E_\nu - 0.091E_\nu^2\right) \\ & + f_{^{239}P_u} \exp\left(0.896 - 0.239E_\nu - 0.0981E_\nu^2\right) \\ & + f_{^{238}U} \exp\left(0.976 - 0.162E_\nu - 0.0790E_\nu^2\right) \\ & + f_{^{241}P_u} \exp\left(0.793 - 0.080E_\nu - 0.1085E_\nu^2\right),\end{aligned}$$

Thermal Power: 20GW$_{\rm th}$

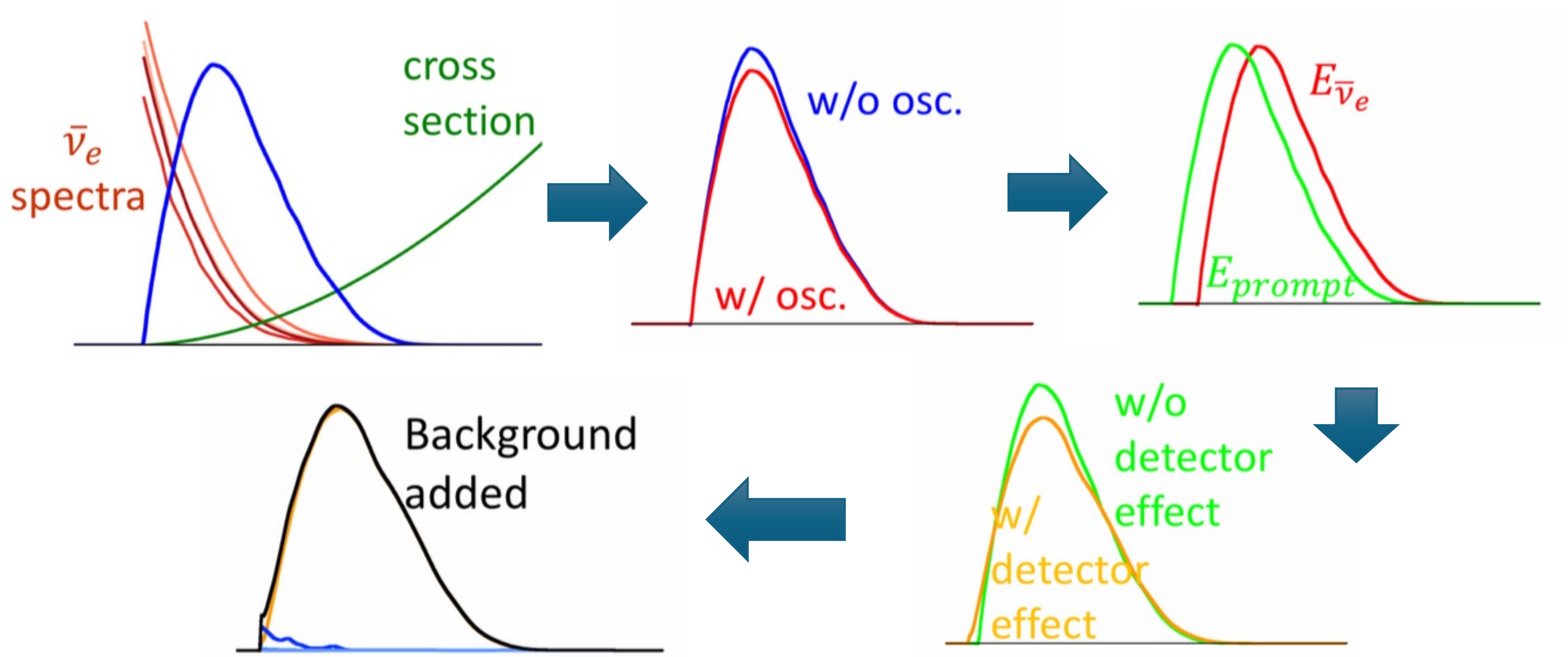
Fission Fraction: f_{U235} = 0.58, f_{Pu239} = 0.29, f_{U238} = 0.08, f_{Pu241} = 0.05

Fission Energy: ε_{U235} = 202.36 MeV, ε_{Pu239} = 211.12 MeV,

ε_{U238} = 205.99 MeV, ε_{Pu241} = 214.26 MeV

- IBD 反应截面: $$\sigma_{\mathrm{IBD}} = 0.0952 \left(\frac{E_e \, p_e}{1\mathrm{MeV}^2} \right) \times 10^{-42} \, \mathrm{cm}^2,$$
- 2万吨液闪的质子数:　$N_p = 1 \times 10^{33}$
- T = 5年
- 探测效率: 80%
- $E_{vis} \sim E_e + m_e \sim E_\nu - 0.78$ MeV
- 能量分辨率: $$\frac{\delta E}{E} = \frac{3\%}{\sqrt{E(MeV)}}$$

能谱预测



$\bar{\nu}_e$
spectra
cross
section
w/o osc.
w/ osc.
$E_{\bar{\nu}_e}$
E_{prompt}
w/o
detector
effect
w/
detector
effect
Background
added

振荡参数拟合步骤

- For given oscillation parameters

$$\theta = (\theta_{13}, \theta_{12}, \Delta m^2_{31}, \Delta m^2_{21})$$

- Suppose a given experiment divides the range of observation into N bins. The outcome is reported in number of observed events in each bin n_i.
- Expect Poisson distribution $\mu_i(\theta)$ for the number of events in each bin

$$\mu_i(\theta) = N_i(\theta) + B_i$$

建立 χ^2

To build χ^2:

$$\chi^2(\theta) = \sum_{i=1}^{N} \frac{[\mu_i(\theta) - n_i]^2}{n_i} \qquad \text{or} \qquad \chi^2(\theta) = \sum_{i=1}^{N} \frac{[\mu_i(\theta) - n_i]^2}{\mu_i(\theta)}$$

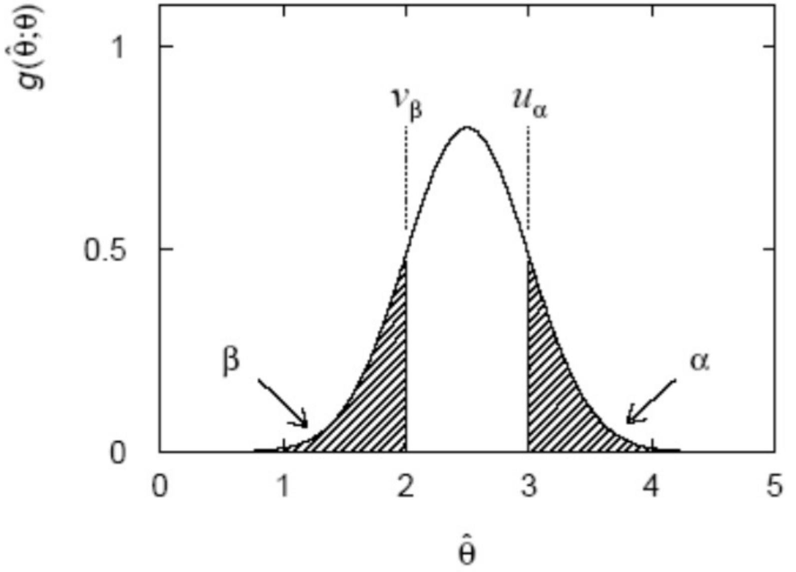
If the number of events is small in some bins, binned log maximal likelihood χ^2 function, (Poisson χ^2)

$$\chi^2(\theta) = 2\sum_{i=1}^{N}\left[\mu_i(\theta) - n_i + n_i \log \frac{n_i}{\mu_i(\theta)}\right]$$

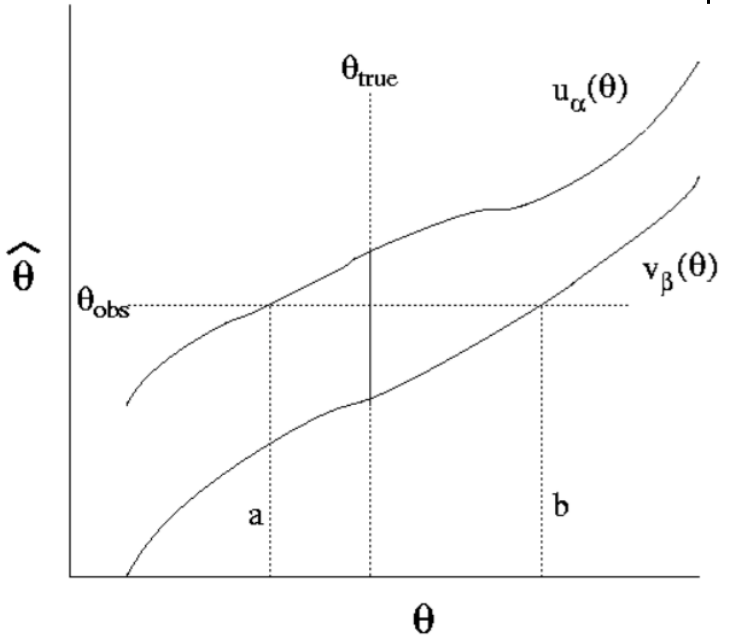
Or with covariance matrix V:

$$\chi^2 = 2\sum_{i,j=1}^{N}[\mu_i(\theta) - n_i]V_{ij}^{-1}[\mu_j(\theta) - n_j]$$

频率学中的置信区间



Confidence belt, $u_\alpha(\theta)$ and $v_\beta(\theta)$



Neyman Confidence interval [a, b]

parameter: θ

estimator: $\hat{\theta}$

Confidence level $P\left(v_\beta(\theta) \le \hat{\theta} \le u_\alpha(\theta)\right) = 1 - \alpha - \beta$

HEP/NP error convention: 68.3%

利用 χ²函数

In the "χ² approximation"

$$\chi^2(\theta) \quad = \quad \chi^2_{\min}\left(\hat{\theta}\right) \quad + \quad \Delta\chi^2(\theta)$$

$$N \qquad\qquad N-P \qquad\qquad P$$

	Parameter estimation, goodness of fit	Confidence interval

- The parameter values $\hat{\theta}_\alpha$ which minimize the χ^2 (usually called "best fit values") are estimators of the "true values"
- X^2_{min} follows a χ^2-distribution with N-P d.o.f. and can be used to evaluate the goodness of fit.
- The $\Delta\chi^2$ relative to the minimum follows a χ^2-distribution with P d.o.f can be used to determine confidence intervals (or regions) for the parameters θ

利用似然函数获得近似置信区间

Suppose we test parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$ using the likelihood ratio

$$t_\theta = -2\ln\left(\frac{L(\theta)}{L(\hat{\theta})}\right)$$

The p-value of $\boldsymbol{\theta}$ $\quad p_\theta = \int_{t_{\theta,obs}}^{\infty} f(t_\theta \mid \theta)\, dt_\theta$

From Wilks' theorem, $f(t_\theta|\boldsymbol{\theta}) \sim \chi_n^2$

The boundary of confidence region in $\boldsymbol{\theta}$ space is

$$\ln L(\theta) = \ln L(\hat{\theta}) - \frac{1}{2} F_{\chi_n^2}^{-1}(1-\alpha)$$

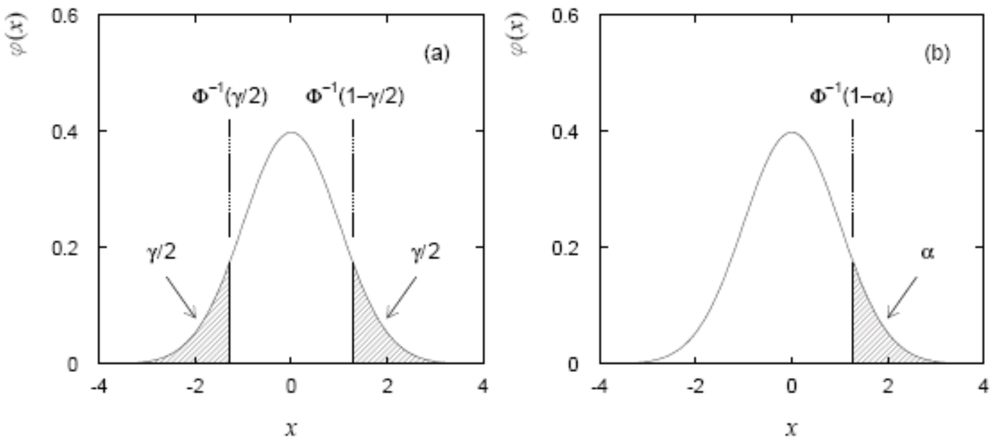
利用 $\Delta\chi^2$ 来建立置信区间

A P-dimensional region in the space θ at given CL is obtained by requiring $\Delta\chi^2(\theta) < X(CL)$ (contours in $\Delta\chi^2$)

d.o.f. \ CL	68%(1σ)	90%	95%(2σ)	99%	99.73%(3σ)
1	1	2.71	4	6.64	9
2	2.28	4.61	5.99	9.21	11.8
3	3.51	6.25	7.82	11.4	14.2

$$F^{-1}_{\chi^2_n}(1-\alpha)$$

Central (2-sided) Upper/lower limit (1-sided)



TMinuit 例子

- Example Code: https://root.cern.ch/doc/master/Ifit_8C.html
- Reference: https://www-glast.slac.stanford.edu/software/root/GRUG/docs/Feature/GRUGminuit.pdf

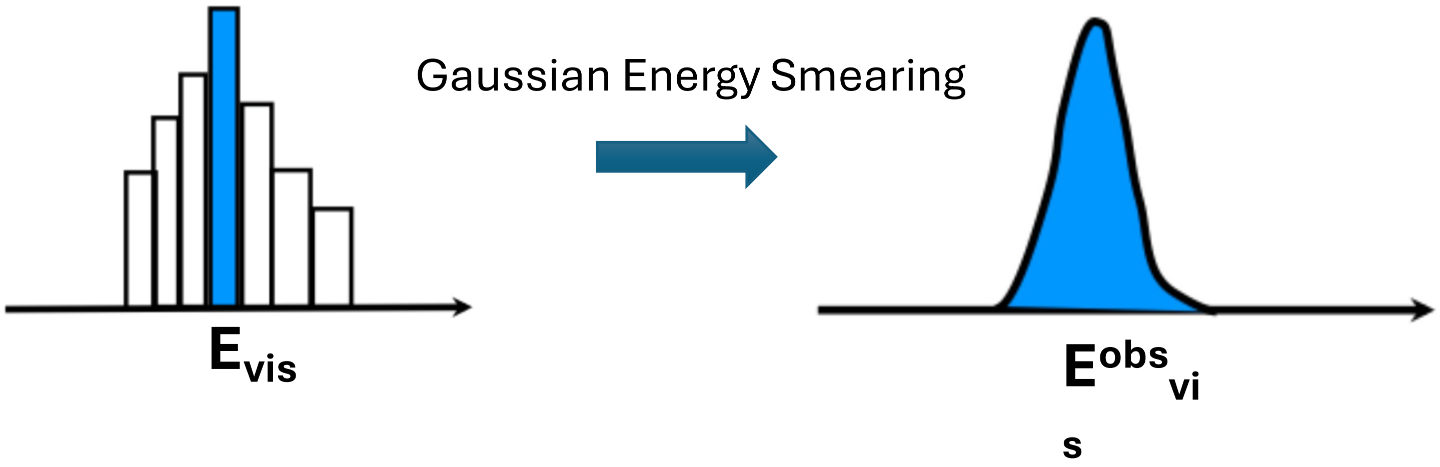
```
void fcn(int &npar, double *gin, double &f, double *par, int iflag)   {
     const int nbins = 5;
     int i;
     //calculate chisquare
    double chisq = 0;
    double delta;
    for (i=0;i<nbins; i++) {
         delta = (z[i]-func(x[i],y[i],par))/errorz[i];
         chisq += delta*delta;
    }
    f = chisq;
}
```

一般单能函数的转换

$$\frac{dN}{dE_\nu}(E_\nu) \to \frac{dN}{dE_{rec}}(E_{rec})$$

$$\frac{dN}{dE_{rec}}(E_{rec}) = \frac{1}{\frac{dE_{rec}(E_\nu)}{dE_\nu}} \frac{dN}{dE_\nu}(E_\nu)$$

能量分辨率的影响



能量响应卷积

$$\frac{dN}{dE_{rec}} = \int \frac{dN}{dE_\nu} \otimes R(E_{rec}|E_\nu)dE_\nu$$

$$n_i(E_{rec}) = \sum_{j=1}^{E_\nu\ bins} R_{ij}(E_{rec}|E_\nu)m_j(E_\nu) = \sum_{j=1}^{E_\nu\ bins} \frac{R_{ij}(E_{rec}, E_\nu)}{\sum_{i=1}^{E_{rec}\ bins} R_{ij}(E_{rec}, E_\nu)} m_j(E_\nu)$$

E_rec　　　　能量响应矩阵　　　　E_nu

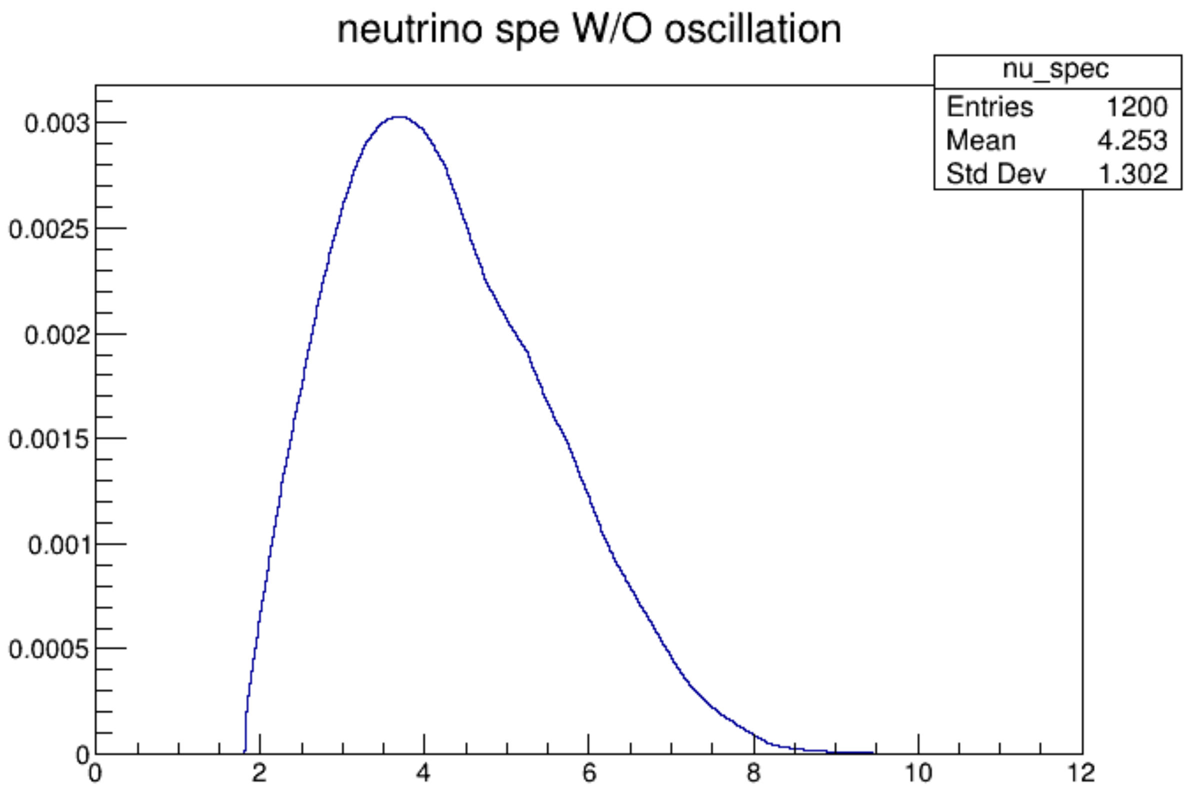
$$\begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix}$$

能量响应矩阵

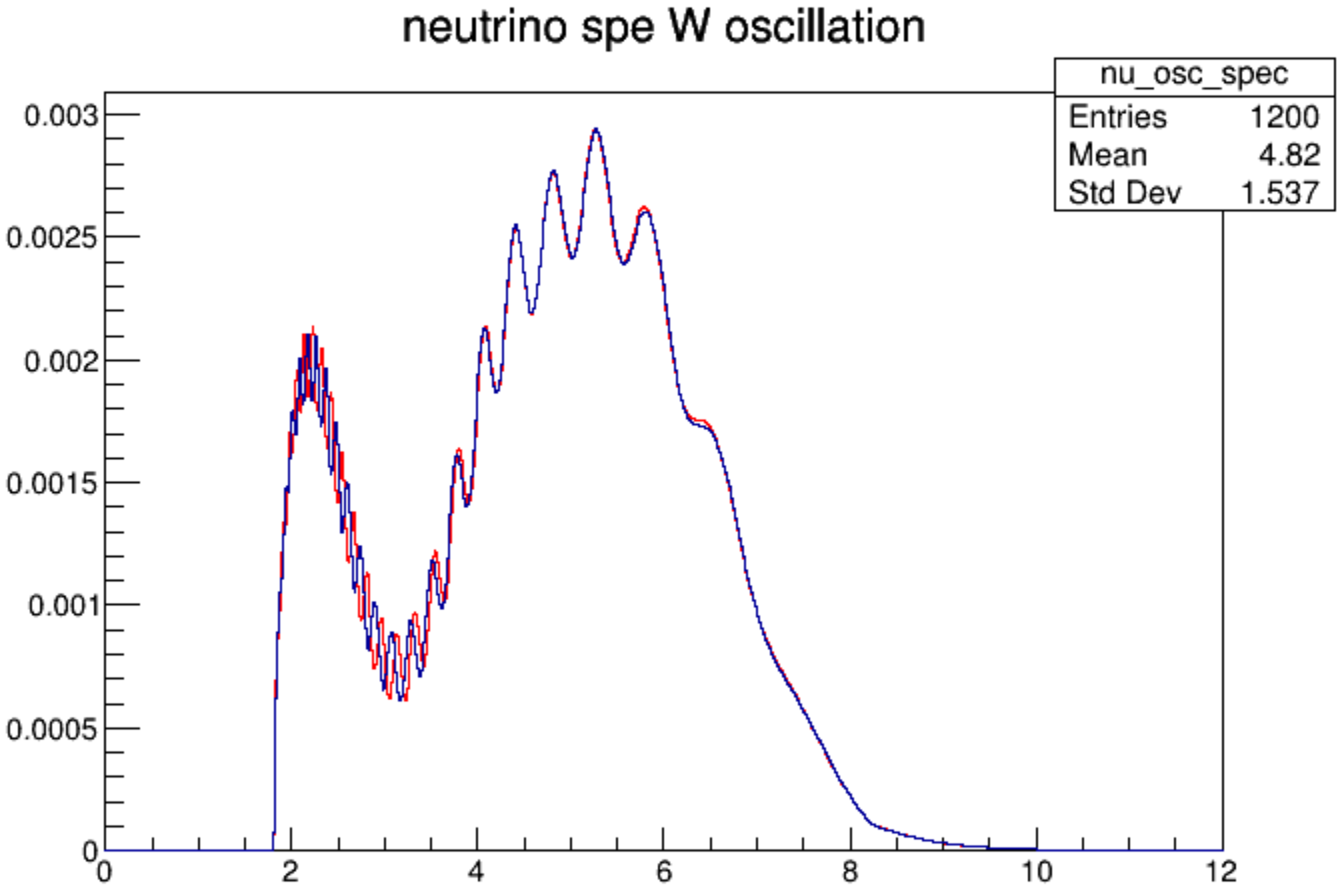
$$f(E_{rec}|E_\nu) = \frac{f(E_{rec}, E_\nu)}{f_{E_\nu}(E_\nu)} = \frac{f(E_{rec}, E_\nu)}{\int f(E_{rec}, E_\nu) dE_{rec}}$$

$$R(E_{rec}|E_\nu) = \frac{R(E_{rec}, E_\nu)}{\sum_{E_{rec}} R(E_{rec}, E_\nu)}$$

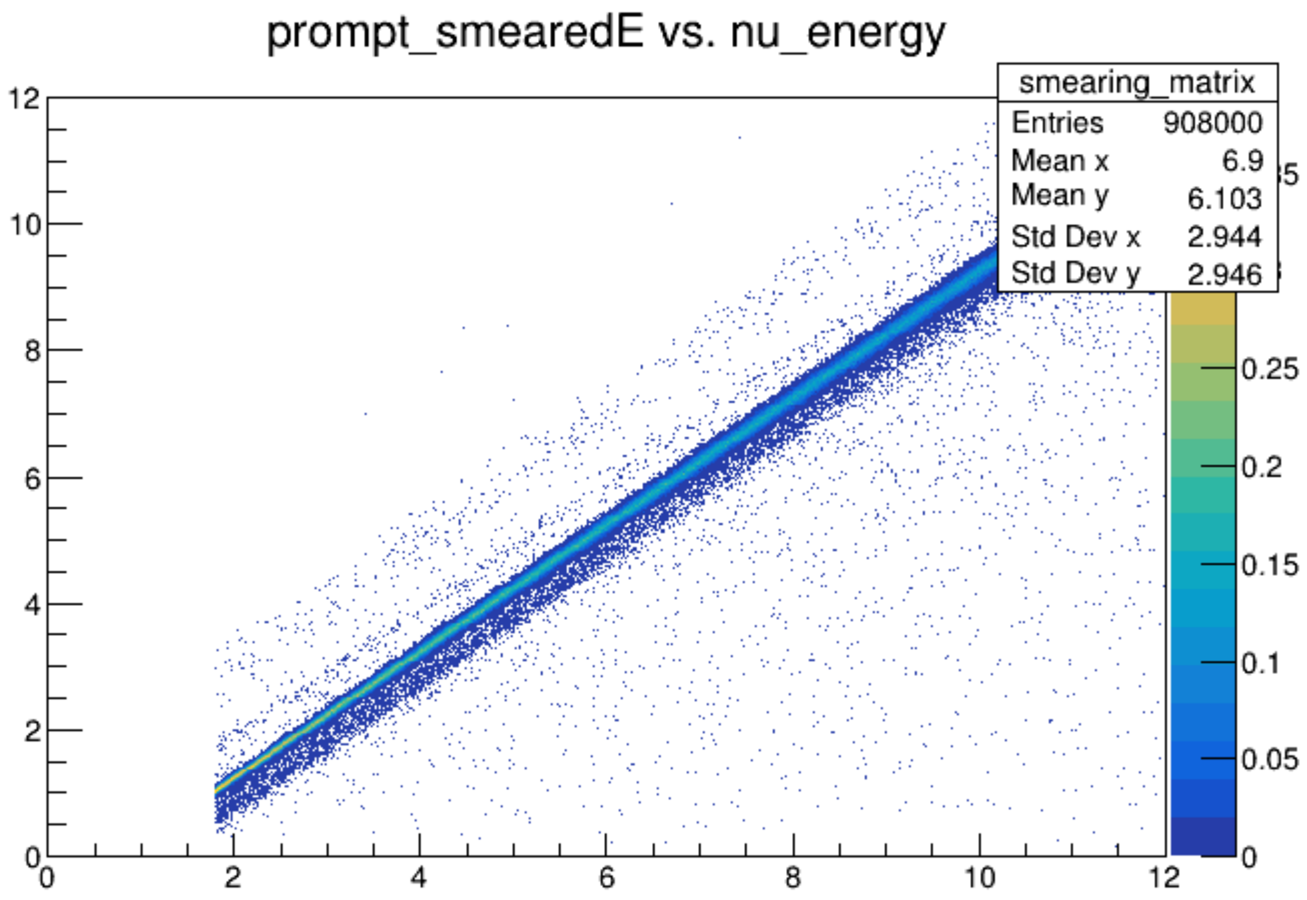
中微子非振荡能谱 $\frac{dN}{dE_\nu}(E_\nu)$



中微子振荡能谱 $\frac{dN}{dE_\nu}(E_\nu)P(L,E_\nu)$

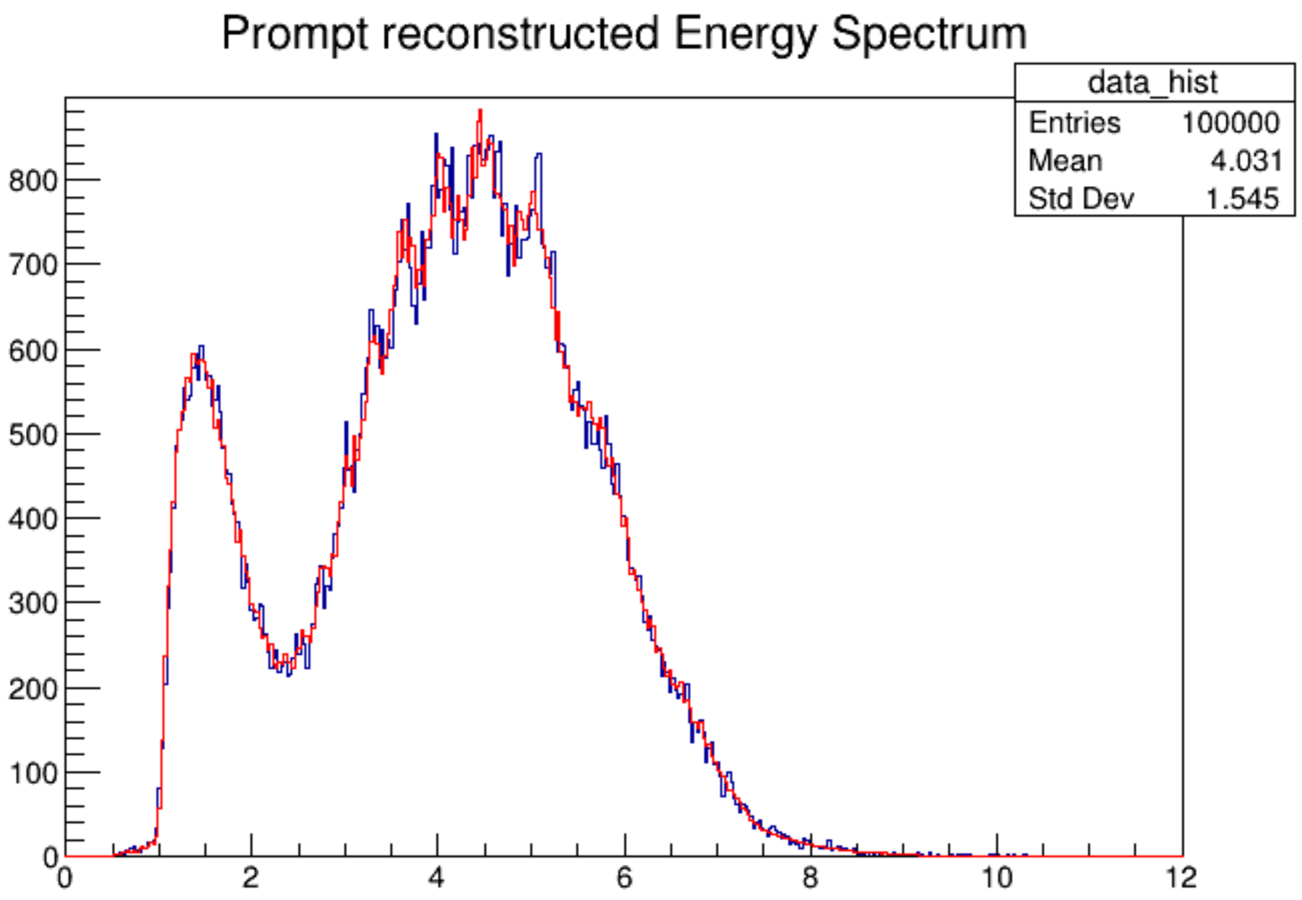


能量响应矩阵



prompt_smearedE vs. nu_energy
smearing_matrix
Entries 908000
Mean x 6.9
Mean y 6.103
Std Dev x 2.944
Std Dev y 2.946
0
2
4
6
8
10
12
0.25
0.2
0.15
0.1
0.05

快信号重建能谱$\frac{dN}{dE_{rec}}(E_{rec})$



作业2：拟合振荡参数

- File1：Nu_spec_OSC_noME.root

totalprob2：无振荡的中微子能谱输入 （需要归一化）

- File2: IBD_flat500k.root

Geant4 模拟的中微子平谱和对应的快信号事例文件 (可以用它产生探测器能量响应矩阵：nu_energy vs. prompt_smearedE)

- File3: IBD_OSC100k_noME.root (建议自己用Geant4来产生)

用中微子振荡能谱的产生子和Geant4模拟后的数据，可以当成真实数据来做振荡参数的拟合(prompt_smearedE 能谱)

作业：

① 利用Tminuit，分别在正序假设下，测量振荡参数$sin^2\theta_{12}$和Δm^2_{21}（2个）的中心拟合值和对应误差

② 分别在正反序假设下，比较最佳拟合所对应的chi2值