Data Unfolding with Wiener-SVD Method and Model Validation

W. Tang,[a,1] **X. Li,**[b,1] **X. Qian,**[a,2] **H. Wei**[a] **and C. Zhang,**[a]

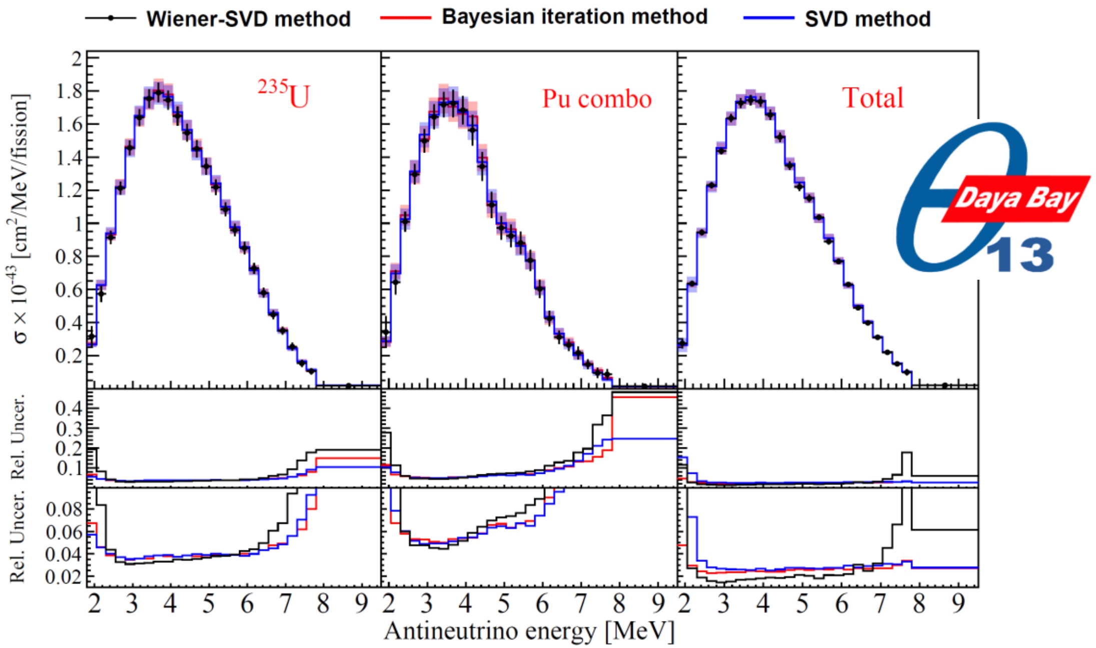
[a] *Physics Department, Brookhaven National Laboratory, Upton, NY, U.S.A.*

[b] *Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, NY, U.S.A.*

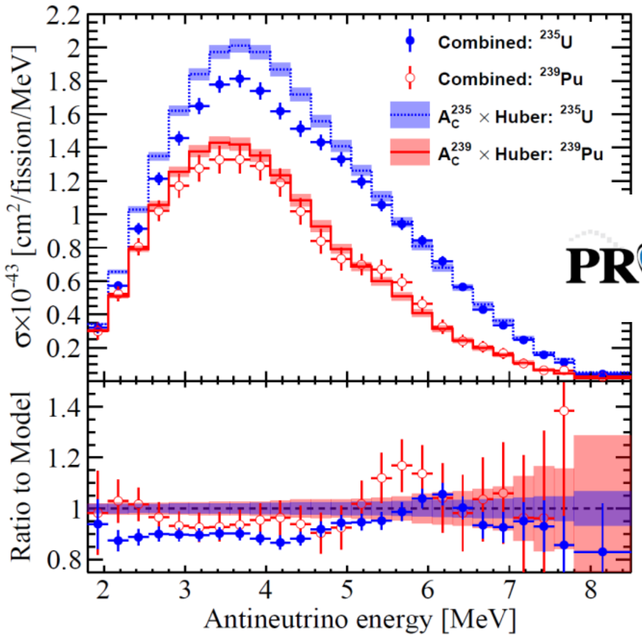
E-mail: xqian@bnl.gov

JINST 12 P10002, arXiv:1705.03568

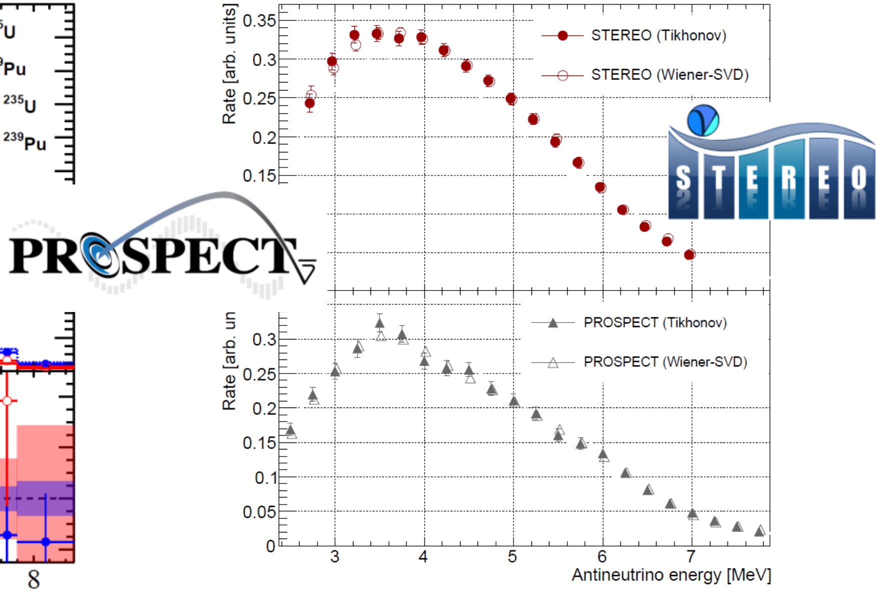
https://github.com/BNLIF/Wiener-SVD-Unfolding



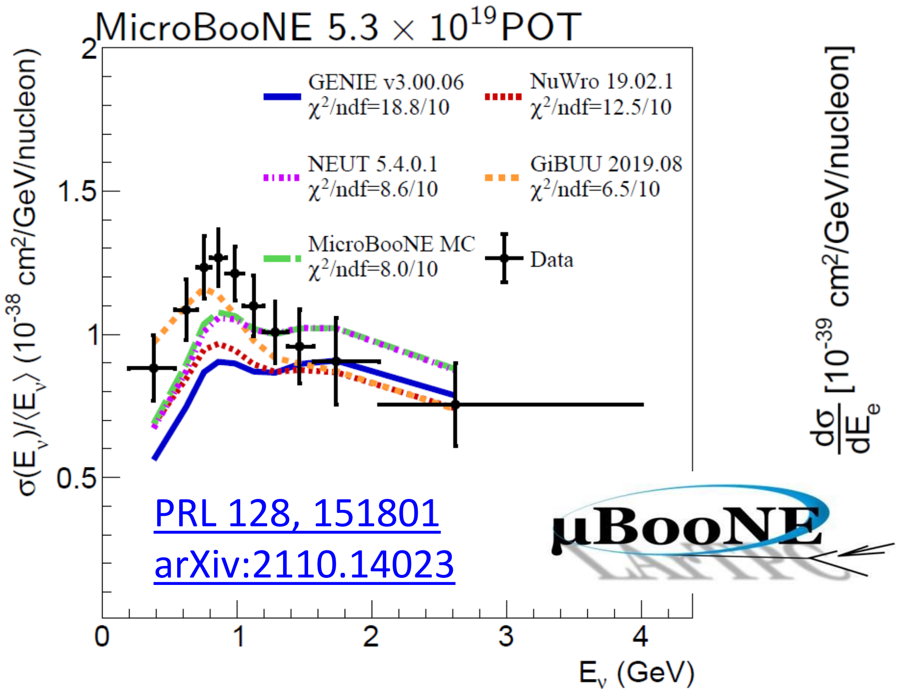
[CPC 45 073001](), [arXiv:2102.04614]()



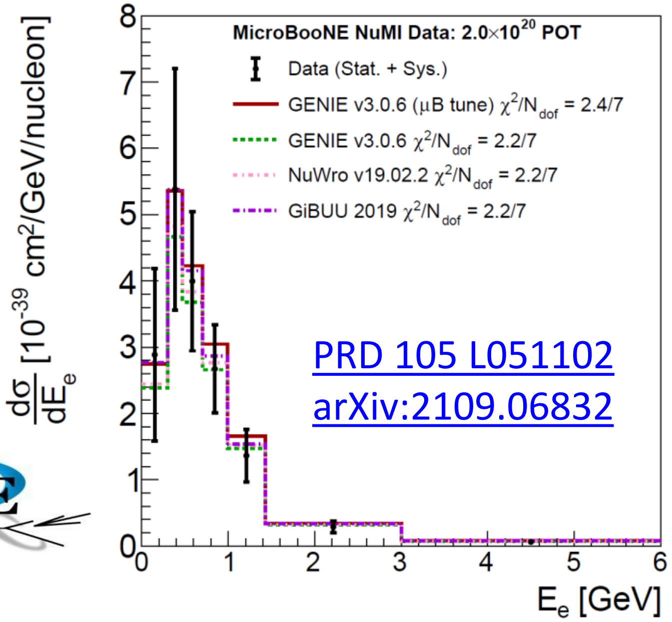
[PRL 128, 081801](), [arXiv:2106.12251]()



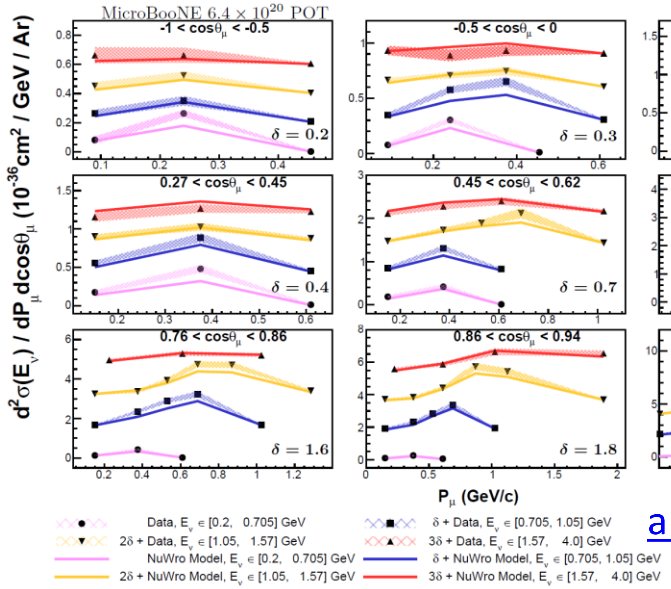
[PRL 128, 081802](), [arXiv:2107.03371]()



[PRL 128, 151801]()
[arXiv:2110.14023]()



[PRD 105 L051102]()
[arXiv:2109.06832]()



[arXiv:2307.06413]()

Introduction to Data Unfolding Problem

True distribution	:	S(x) on variable x with dimension d_S
Measured distribution	:	M(y) on variable y with $\bar{y} = R(\mathrm{x})$ and dimension d_M
Unfolding problem is		M(y) → S(x)

- Data unfolding is very general:
  - ν oscillation: x ~ ν mixing parameters, M(y) ~ distribution in recon. ν energy, $d_M \gg d_S$
  - Deconvolution: S(x) ~ ionization Q, M(y) ~ raw waveform, $d_M = d_S$ in TPC signal processing
  - LArTPC (Wire-Cell) Q/L matching: x ~ Q/L pair, M(y) ~ measured light pattern, $d_M < d_S$
  - Reactor $\bar{\nu}$ spectrum: S(x) ~ true $\bar{\nu}$ spectrum, M(y) ~ measured $\bar{\nu}$ spectrum, $d_M \geq d_S$
  - νCC Xs extraction: S(x) ~ differential Xs, M(y) ~ measured CC distributions, $d_M \geq d_S$
  - …
  - **Any physics analysis is essentially a data unfolding problem**

Special Case of $d_M \geq d_S$: Weighted Least Squares

$$T = (M - R \cdot S)^T \cdot C^{-1} \cdot (M - R \cdot S)$$

M : (vector) measurement

S : (unknown vector) signal

R : response matrix connecting signal to measurement

C : Covariance matrix describing uncertainties

→

Minimizing T leads to a solution with Linear Algebra of

$$S = (R^T \cdot C^{-1} \cdot R)^{-1} \cdot (R^T \cdot C^{-1} \cdot M)$$

and

$$C_S = (R^T \cdot C^{-1} \cdot R)^{-1}$$

A. C. Aitken Proc. R. Soc. Edinburgh 55, 42 (1935)

- Since measurements are around the expectation
  - $M = R \cdot S + N \rightarrow \hat{S} = S + (R^T \cdot C^{-1} \cdot R)^{-1} \cdot R^T \cdot C^{-1} \cdot N$
  - N : statistical and systematic uncertainties

Large fluctuations → Regularization (e.g. Wiener-SVD) is needed for intuitive results

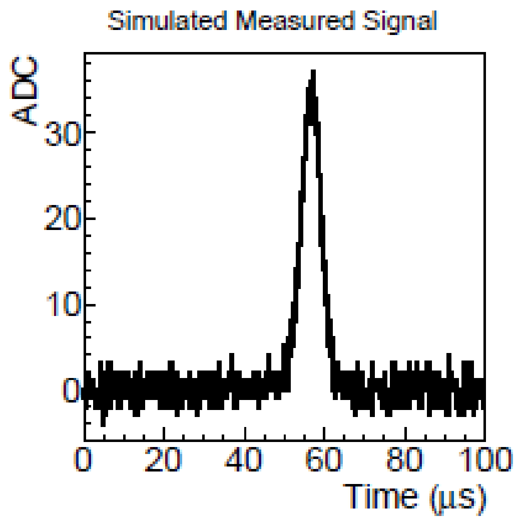
Overview: Goals of Data Unfolding

- **Reduce difference between unfolded results w.r.t. truth ($\hat{S}$ vs. S)**
- Wiener-SVD:
  - Minimize total mean squared error (MSE = bias^2 + variance) in the chosen effective frequency domain through application of the Wiener(-inspired) filter given an expected signal
  - Complete error estimation ($\delta\hat{S}$)
  - Recognize re-smearing matrix A_C ($\hat{S} \sim A_C \cdot S$)
- **Retain the maximal power to differentiate model predictions in comparing to unfolded results (P vs. $\hat{S}$)**
- Auxiliaries of Wiener-SVD:
  - Clear definition of S (e.g. nominal ν flux weighted Xs)
  - Model validation (no significant missing uncertainties by data/MC consistency)
  - Publishing A_C matrix ($\hat{S}$ $vs. A_C \cdot P$)

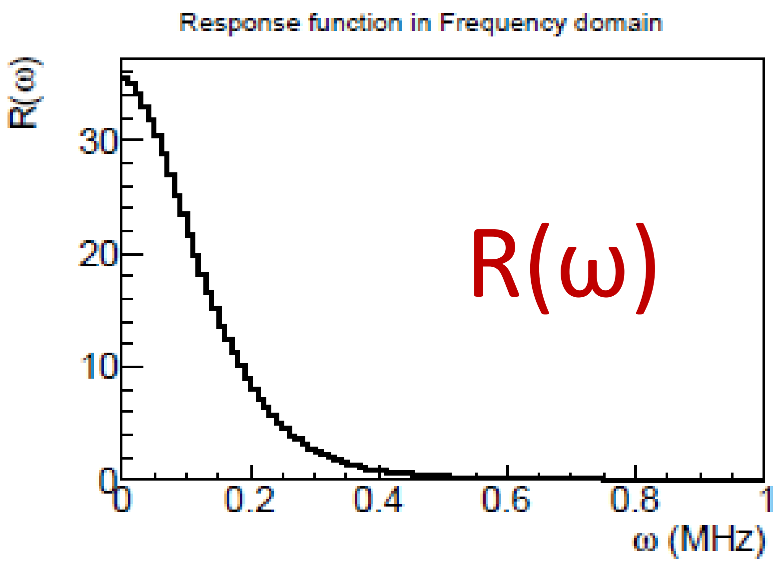
History of Wiener-SVD unfolding

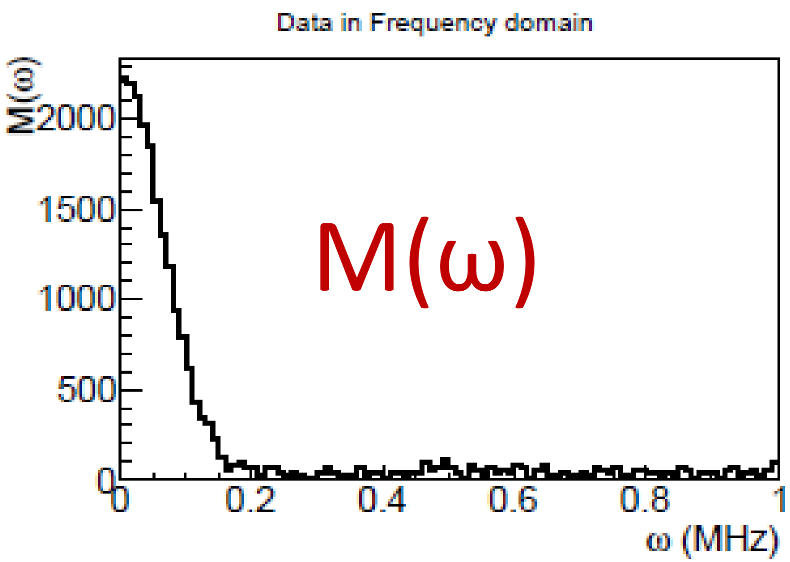
- Wiener-SVD was inspired by LArTPC signal processing
  - B. Baller [JINST 12 P07010]

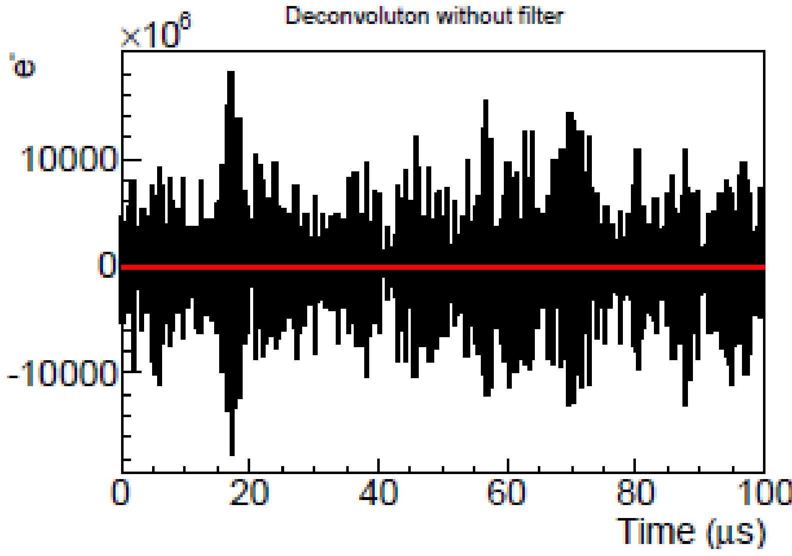
$$M(t_0) = \int_{-\infty}^{\infty} R(t-t_0)\cdot S(t)dt + N(t_0)$$



  - Apply (Fast) Fourier Transformation: $S(\omega) = \frac{M(\omega)}{R(\omega)} - \frac{N(\omega)}{R(\omega)}$







Software (Wiener) Filter

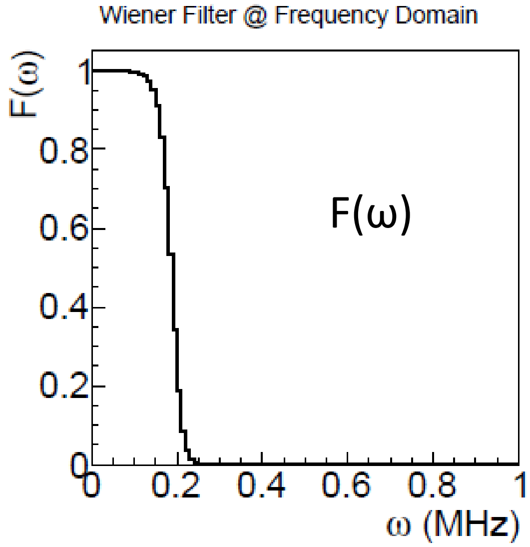
- To suppress the noise at the high frequency, a software filter is generally needed

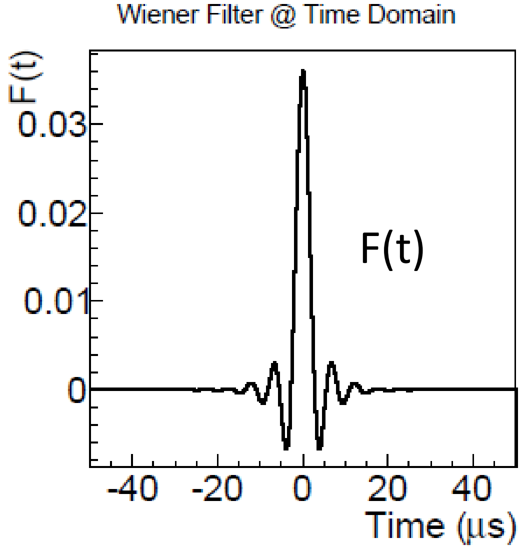
$$S(\omega) = \frac{M(\omega)}{R(\omega)} \cdot F(\omega)$$

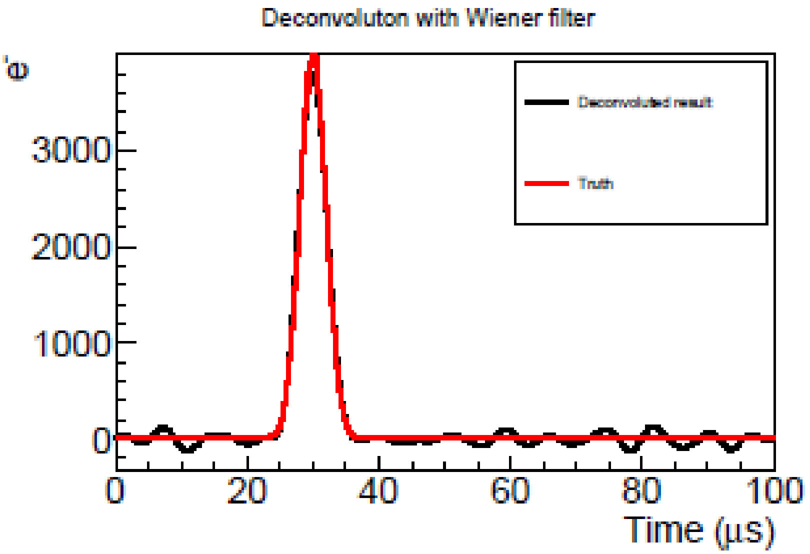
$$S_{decon}(t_0) = \int_{-\infty}^{\infty} F(t - t_0) \cdot S(t)\, dt$$

- One form of the filter is the Wiener filter using expectations of signal and noise

$$F(\omega) = \frac{\overline{S^2(\omega)}}{\overline{S^2(\omega)} + \overline{N^2(\omega)}}$$







Detector response is replaced by the filter (smearing) function!

Meaning of Wiener filter

- Wiener filter was determined by minimizing the expectation of

$$E\left[\left(F(\omega)\cdot M(\omega)-\overline{S(\omega)}\right)^2\right]$$

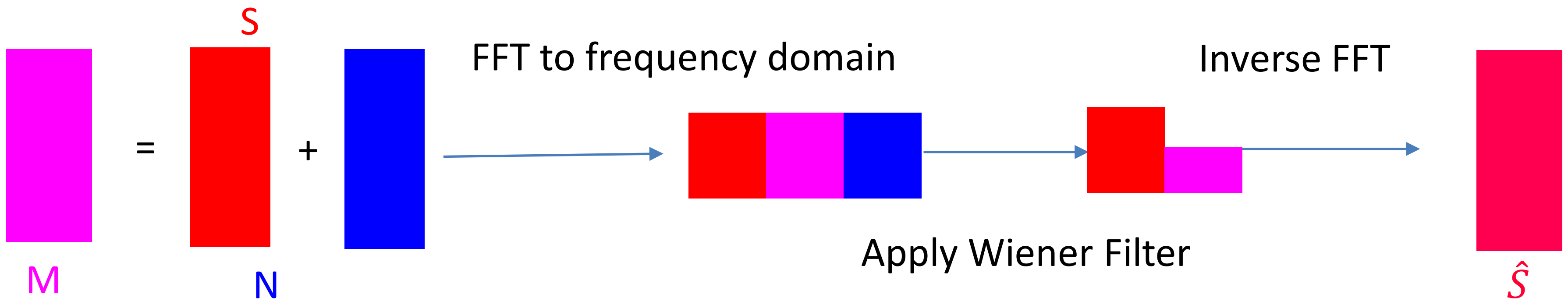
$$=E\left[\left(F(\omega)\cdot\left(\bar{S}(\omega)+N(\omega)\right)-\overline{S(\omega)}\right)^2\right]$$

M: measurement

$\bar{S}$: expectation of the signal

$$F(\omega)=\frac{\overline{S(\omega)}^2}{\overline{S(\omega)}^2+\overline{N^2(\omega)}}$$

Wiener filter is by construction to minimize the total mean squared error (MSE = bias^2 + variance) in the frequency domain



How to find a (frequency) 'domain' to maximize separating signal and noise?

SVD Unfolding

- Start with general chisquare formalism with the covariance matrix

$$\chi^2 = \sum_{i,j} \left(m_i - \sum_k r_{ik} \cdot s_k \right) \cdot Cov_{ij}^{-1} \cdot \left(m_j - \sum_k r_{jk} \cdot s_k \right)$$

m is the measured counts
r is the smearing matrix,
s is true distribution to be extracted
Cov is the covariance matrix taking into account all uncertainties

- Whitening of the chisquare

$$Cov^{-1} = Q^T \cdot Q$$

$$\chi^2 = (M - R \cdot s)^2$$

$$M = Q \cdot m$$
$$R = Q \cdot r$$

- SVD decomposition of R

$$R = U \cdot D \cdot V^T$$

$$\hat{s} = V \cdot D^{-1} \cdot U^T \cdot M \iff S(\omega) = \frac{M(\omega)}{R(\omega)}$$

Effective frequency domain determined by Cov (uncertainties) and R (response)

Frequency domain

Wiener-SVD Unfolding

$$\hat{s} = V \cdot F \cdot D^{-1} \cdot U^T \cdot M$$

⟷

$$S(\omega) = \frac{M(\omega)}{R(\omega)} \cdot F(\omega)$$

- Tikhonov regularization

$$F_{ii} = \frac{d_i^2}{d_i^2 + \tau}$$

Regularization strength τ to be varied for optimization

- Wiener regularization

$$W_{ik} = \frac{d_i^2 \cdot \left(\sum_j V_{ij}^T \cdot \bar{s}_j\right)^2}{d_i^2 \cdot \left(\sum_j V_{ij}^T \cdot \bar{s}_j\right)^2 + 1} \cdot \delta_{ik},$$

Expectation of signal $\bar{S}$ is required, no free parameter

Generalized Wiener SVD Approaches

- Instead of using amplitude of s, we can use 1st or 2nd derivative of s

$$\overline{M} = R \cdot C^{-1} \cdot C \cdot \overline{s}$$

$$\hat{s} = A_C \cdot (R^T R)^{-1} \cdot R^T \cdot M,$$

where

$$A_C = C^{-1} \cdot V_C \cdot W_C \cdot V_C^T \cdot C.$$

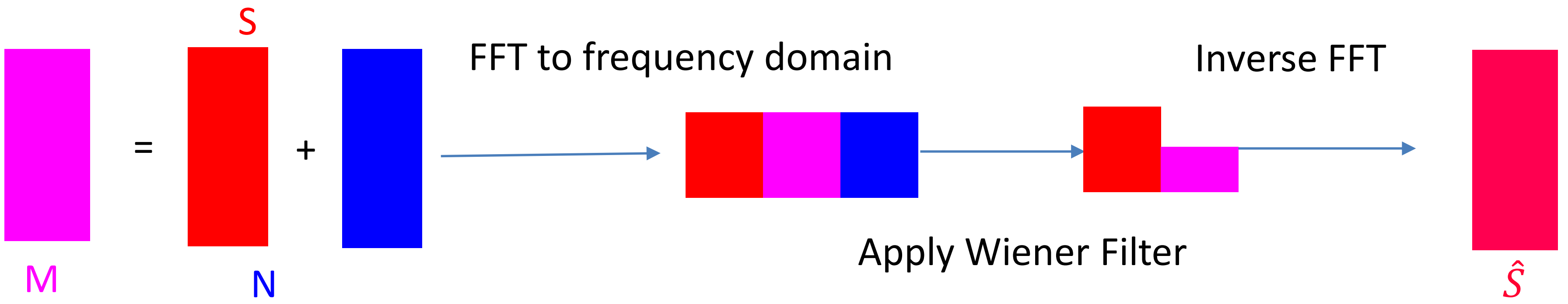
The corresponding Wiener filter would be

$$W_{ii} = \frac{d_{Ci}^2 \cdot \left(\sum_j V_{Cij}^T \cdot \left(\sum_l C_{jl} \cdot \overline{s}_l\right)\right)^2}{d_{Ci}^2 \cdot \left(\sum_j V_{Cij}^T \cdot \left(\sum_l C_{jl} \cdot \overline{s}_l\right)\right)^2 + 1},$$

$$C_0 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}, C_1 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} -1+\epsilon & 1 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2+\epsilon & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -2+\epsilon & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 & -1+\epsilon \end{bmatrix},$$

Different C gives different effective frequency domain, in which has different signal/noise separations!



How to find an effective 'domain' to maximize separating signal and noise?

Wiener-SVD: Uncertainties and Regularization

- Unfolded results $\hat{S} = R_{tot} \cdot m$
  - with $R_{tot} = A_C \cdot (R^T R) \cdot R^T \cdot Q$
- Or $\hat{S} = A_C \cdot (S + (R^T R) \cdot R^T \cdot Q \cdot N)$
  - with $A_C = C^{-1} V_C W_C V_C C$
  - **Expectation $\overline{\hat{S}} = A_C \cdot \overline{S}$ (truth expectation)**
- Uncertainty of $\hat{S}$
  - $Cov_{\hat{S}} = R_{tot} \cdot Cov_m \cdot R_{tot}^T$

- Regularization language
  - Minimizing $\phi(s) = \chi^2(s) + R(s)$
- Tikhonov regularization
  - $R(s) = \tau \cdot \int \left(\frac{d^k s}{dE^k}\right)^2$
  - k=0, 1, 2 ~ amplitudes, slopes, smoothness of S
- Wiener-SVD
  - $R(s) = \frac{1}{2} \sum_i \log \frac{M_{U\,i}^2}{N^2}$

Signal in the effective frequency domain

Noise in the effective frequency domain

Key Take-away Points of Wiener-SVD

- Unfolded results are essentially $\bar{\hat{S}} = A_C \cdot \bar{S}$
  - No regularization strength τ
  - No ad-hoc unfolding uncertainty
- Wiener-SVD needs an input of expectation of signal
  - By construction, the smallest MSE
- Difference between unfolded results w.r.t. truth depends on the choice of effective frequency domain (e.g. C) and the expectation of signal
- **Retain the maximal power to differentiate model predictions in comparing to unfolded results (P vs. $\hat{S}$)**
  - Publish A_C, so that it can be applied to the predictions for comparison
- Covariance matrix of measurements need to accommodate all uncertainties → **Model validation**
  - Calibration: use (external) data to replace model
  - Validation: use (this) data to validate model

Application: Cross Section Extraction Procedure (I)

- Case study: extraction of total ν_μCC cross section as neutrino energy

Measurements | ν_μ Neutrino Flux | ν_μCC cross section | Detector response matrix | Selection efficiency | background

$$M(E_{rec}) = POT \cdot T \cdot \int F(E_\nu) \cdot \sigma(E_\nu) \cdot D(E_\nu, E_{rec}) \cdot \varepsilon(E_\nu, E_{rec}) \cdot dE_\nu + B(E_{rec})$$

E_ν: true neutrino energy E_{rec} : reconstructed neutrino energy

POT: proton on target

T: number of target nucleons

$$M_i = \sum_j S_{ij} + B_i \qquad i: \text{ bin in } E_{rec} \qquad j: \text{ bin in } E_\nu$$

Cross Section Extraction Procedure (II)

With the nominal ν_μ flux prediction $\overline{F}\left(E_{\nu\,j}\right)$, we have

$$S_{ij} = \frac{POT \cdot T \cdot \int_j F\left(E_{\nu\,j}\right)\cdot\sigma\left(E_{\nu\,j}\right)\cdot D\left(E_{\nu\,j}, E_{rec\,i}\right)\cdot\varepsilon\left(E_{\nu\,j}, E_{rec\,i}\right)\cdot dE_{\nu\,j}}{POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\,j}\right)\cdot\sigma\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}} \cdot \left(POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}\right) \cdot \left(\frac{\int_j \overline{F}\left(E_{\nu\,j}\right)\cdot\sigma\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}}{\int_j \overline{F}\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}}\right)$$

$$S_{ij} = \Delta_{ij} \cdot F_j \cdot S_j = R_{ij} \cdot S_j$$

$$\Delta_{ij} = \frac{(\text{Selected no. of events in reco. energy bin i from true energy bin j after event weights})}{(\text{Generated no. of events in truth energy bin j after event weights})}$$

Can be directly calculated with existing Monte Carlo simulations

$$F_j = POT \cdot T \cdot \int_j \overline{F}\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}$$

A constant for each reco. energy j bin related to POT, T, and nominal flux

$$S_j = \frac{\int_j \overline{F}\left(E_{\nu\,j}\right)\cdot\sigma\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}}{\int_j \overline{F}\left(E_{\nu\,j}\right)\cdot dE_{\nu\,j}}$$

Nominal flux averaged cross section in the truth energy bin j

for pros, also see discussions in L. Koch and S. Dolan PhysRevD.102.113012

This choice is crucial in simplifying the uncertainty calculation and comparisons with model calculations

Cross Section Extraction Procedure (III)

- Some of the major uncertainties
  - Flux uncertainties: reweighting from universes → impacts on F and B and then to M
  - Xs uncertainties: reweighting from universes → impacts on σ and B and then to M (different from before, since we are extract Xs here), suppressed in Xs extraction
  - Detector uncertainties: bootstrapping method based on dedicated DetVar/CV simulation → impacts on D and ε and then to M, also on B

$M = R \cdot S + B$

M : vector, number of measured events

B: vector, predicted number of background events

S : vector, nominal flux averaged Xs to be unfolded

R : (known) response matrix connecting M - B, and S

C : covariance matrix of M

→

$$S = A_C \cdot \left(R^T \cdot C^{-1} \cdot R\right)^{-1} \cdot \left(R^T \cdot C^{-1} \cdot M\right)$$

and

$$C_S = A_C \cdot \left(R^T \cdot C^{-1} \cdot R\right)^{-1} \cdot A_C^T$$

A_C : known new smearing matrix based on Wiener filter and the input Xs model

All the derivations are exact with no approximations

Model Validation

- For accelerator neutrino experiments, the validation of $D(E_\nu \rightarrow K)$ (K being kinematics variable in measurements) is important!
  - Performing measurement in the visible kinematics (e.g. K being lepton angle or lepton energy) and true (instead of nominal) neutrino flux does not avoid the problem
- To compare the unfolded Xs (at true flux) with event generator predictions,
  - Theorists provide their model of $D(E_\nu \rightarrow K)$
  - The impact of $D(E_\nu \rightarrow K)$ on the predictions depends on the neutrino spectrum and its uncertainties (extrapolation from nominal to true neutrino flux)
  - Since unfolded Xs also includes neutrino flux uncertainties, theorists cannot do a fair comparison between the unfolded Xs and the predictions without
    - Uncertainties of the neutrino flux and spectrum
    - Correlation between the neutrino flux uncertainties and unfolded Xs
- If we (experimentalists) do not do our job in validating and including the model uncertainties, **it is very difficult, if not impossible,** for theorists to take the uncertainty of $D(E_\nu \rightarrow K)$ into account!

Model Validation Tools: Goodness-of-Fit Tests

Global/Local GoF Tests

- χ^2**/ndf** calculated from the full systematics (flux, Xs, detector, MC statistics) and statistics

$$\chi^2 = (M-P)^T \times Cov_{full}^{-1}(M,P) \times (M-P)$$

- Perform decomposition on the $\mathrm{Cov}_{\mathrm{full}}$ so that one can examine deviation on each (independent) eigen vectors
  - $Cov_{full} = Q^T \cdot D \cdot Q$ D: diagonal, Q: unitary
  - $\chi^2 = [Q \cdot (M-P)]^T \cdot D^{-1} \cdot [Q \cdot (M-P)]$

$$\chi^2 = \sum_i \chi_i^2 = \sum_i \frac{(m_i - p_i)^2}{d_i^2}$$

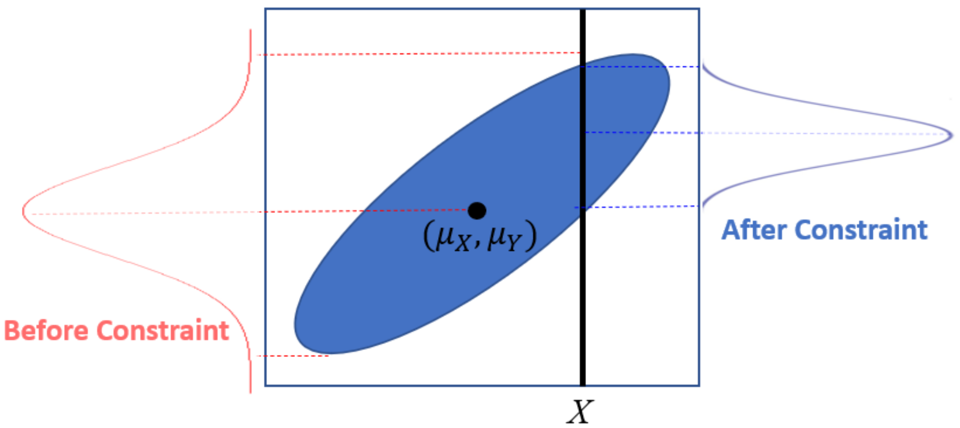
Conditional Constraining Procedure

Conditional expectation & covariance

$$\mu_{X,Y} = \begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \qquad \Sigma_{X,Y} = \begin{pmatrix}\Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY}\end{pmatrix}$$

$$\mu_{Y|X} = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X)$$

$$\boldsymbol{\Sigma_{Y|X} = \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}}$$



* Estimate correlated statistical uncertainty with bootstrapping (sampling w/ replacement)

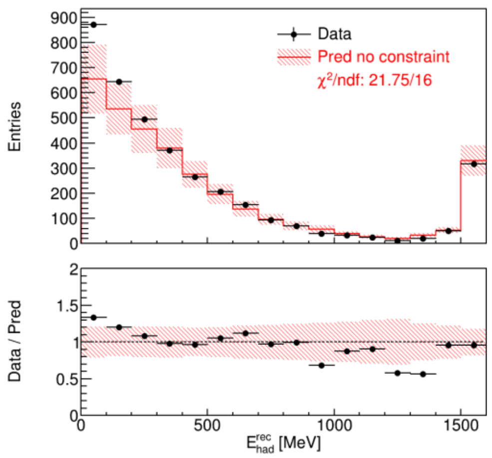
Model Validation: $M(\mathbf{E}^{\mathbf{rec}}_{\mathbf{had}})$ vs. $\mu(\mathbf{E}^{\mathbf{rec}}_{\mathbf{had}} \mid \mathbf{E}_{\nu}, \mathbf{E}^{\mathbf{rec}}_{\mu})$

$\mathbf{E}_{\nu}$ → Neutrino flux modeling

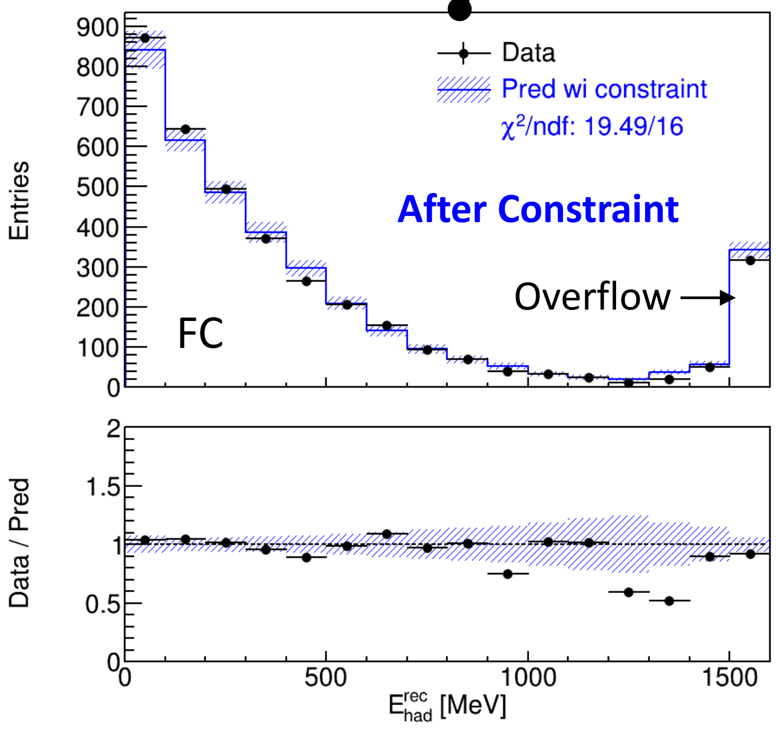
$\mathbf{E}^{\mathbf{rec}}_{\mu}$ → Measurement of muon kinematics

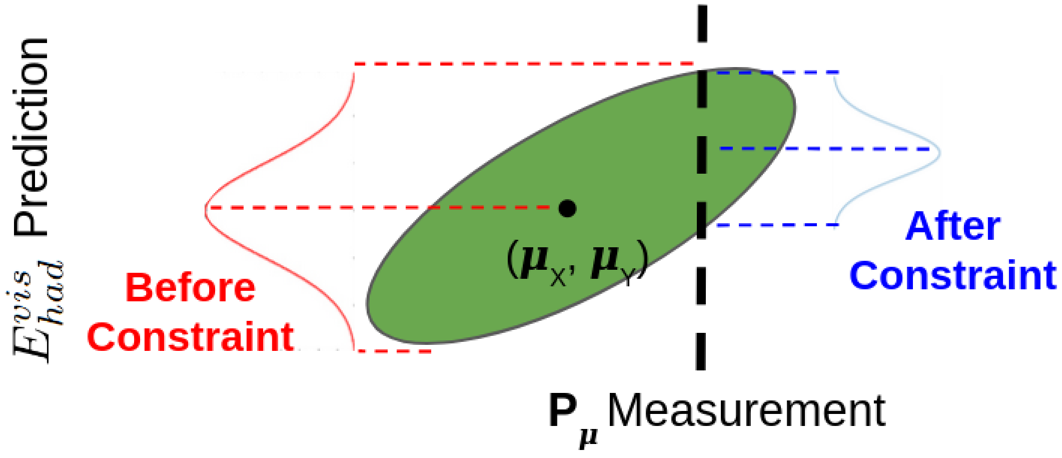
- New method to <u>validate modeling of neutrino energy</u> reconstruction given separated lepton and hadronic energy measurements in LArTPC

$$\mathbf{E}_{\nu} = \mathbf{E}_{\mu} + \mathbf{E}_{\mathbf{had,vis}} + \mathbf{E}_{\mathbf{had,missing}}$$



Excess at low hadronic energy indicates mis-modeling of missing energy?





Also see N. Nayak's talk on Wednesday

Measured muon kinematics are used to constrain the overall model (flux, cross section, etc.) for hadronic energy

Summary

- **Wiener-SVD**, by construction, gives the **smallest MSE**
  - **No regularization strength τ**
  - **No ad-hoc unfolding uncertainty**

Unfolded results $\hat{S} = R_{tot} \cdot m$
with $R_{tot} = A_C \cdot (R^T R) \cdot R^T \cdot Q$
Uncertainty of $\hat{S}$ $Cov_{\hat{S}} = R_{tot} \cdot Cov_m \cdot R_{tot}^T$

- The quality of Wiener-SVD unfolding depends on the choice of effective frequency domain and the input signal expectation
- Retain the maximal power to differentiate model predictions in comparing to unfolded results
  - **Clear definition of S (e.g. nominal ν flux weighted Xs)**
  - **Model validation (no significant missing uncertainties by data/MC comparison)**
  - **Publishing A_C matrix ($\widehat{S}$ $vs.$ $A_C \cdot P$)**

Acknowledgement

- Thank London Cooper-Troendle, Ben Bogard, Lee Hagaman, Matt Toups and others for their help in preparing this talk

Summary of different Apporaches

- Linear Algebra is a very powerful tool in experimental data analysis
  - **Least square approach**:

Applications	Selected Examples
Signal Processing	MicroBooNE TPC and Daya Bay PMT waveform analysis
Data unfolding	EXO-200, Wiener-SVD
Event Reconstruction	Wire-Cell 3D imaging, MicroBooNE ν selection

  - **Iterative approach**:

Applications	Selected Examples
Numerical solutions	Coordinate Descent, BCGSTAB
Data unfolding	ML-EM (Bayesian unfold) in NEXT, ν Xs
Reducing bias	Iterative Least Weighted Squares
Approx. NL fit	Trajectory & dQ/dx fit in LArTPC

$N_{constraint} \geq N_{unknown}$:

i) Matrix Inversion (modest N)

ii) Numerical Solution (large N)

iii) Fast Fourier Trans. (Toeplitz)

$N_{constraint} < N_{unknown}$:

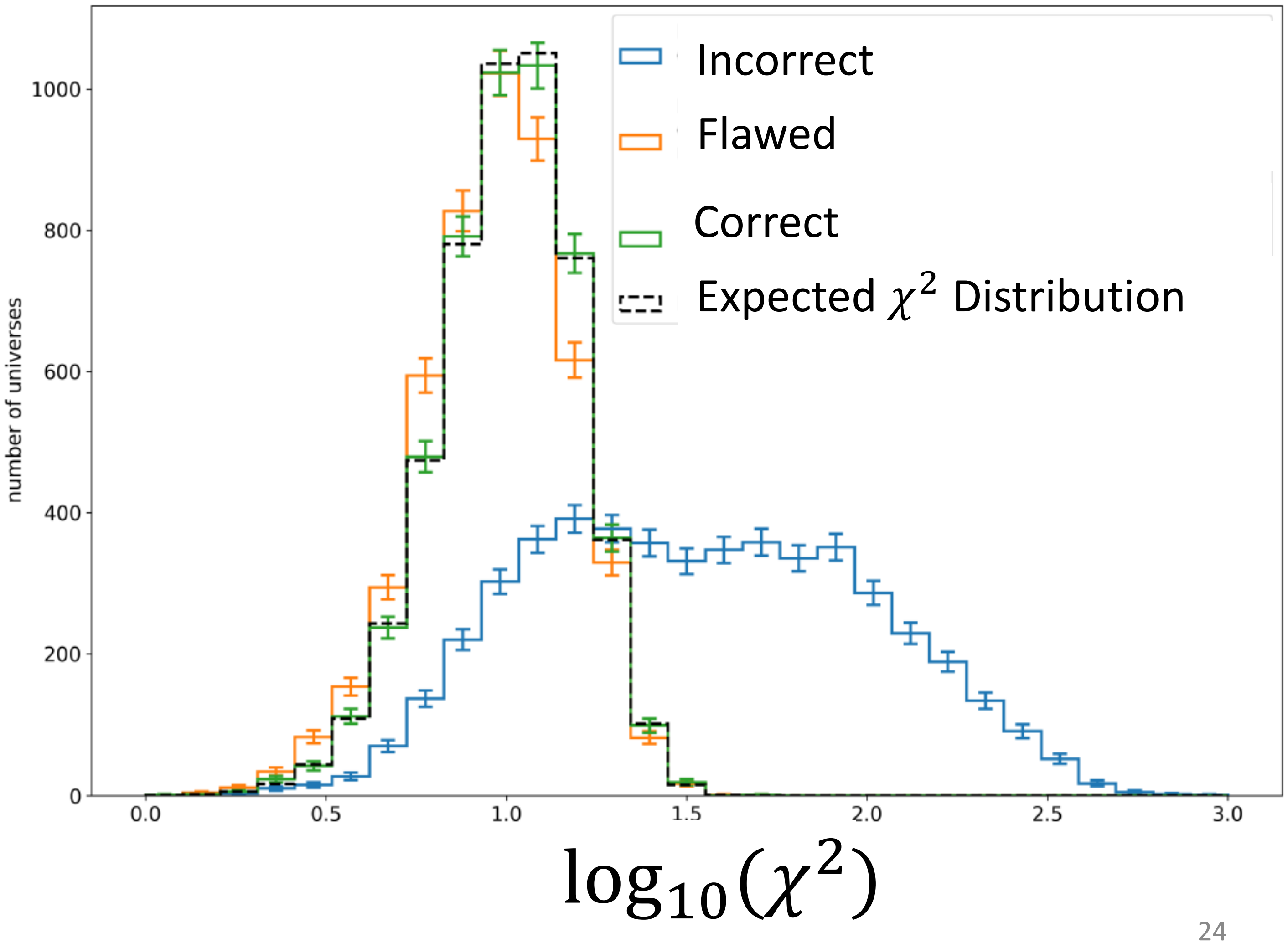
i) Compressed Sensing L1 reg. (Sparsity and Positivity info.)

ii) ML-EM iterative approach (Prior and Positivity info.)

Issue of Extracting Xs at Real Flux with FDS

- **Incorrect:** Not considering the neutrino flux uncertainty in making predictions
- **Flawed:** Ignore correlation of flux uncertainty between prediction and Xs
  - Theorists need to learn how to use the reported neutrino flux uncertainties to extrapolate to real flux
- **Correct:** Consider correlations
  - No experiment report such correlation so far

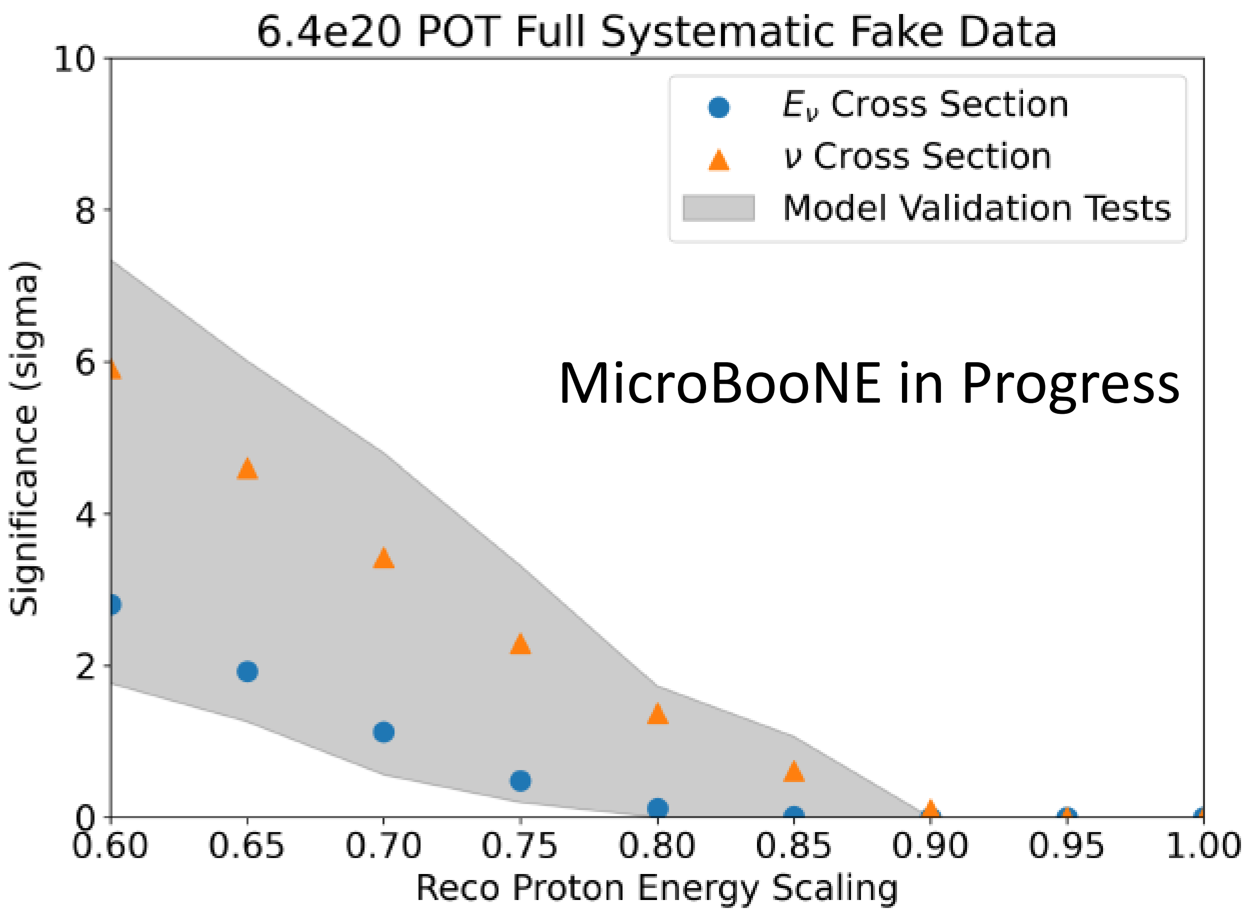
Comparisons between Xs prediction and extraction



Sensitivity of Model Validation and Xs Extraction

		Full systematic uncertainties			
	POT	XS Extraction σ			Model Val. σ
		E_ν	E_μ	ν	$E_{\rm had}\vert(E_\mu,\theta_\mu)$
GENIE v2	7.24E20	0.2	0.02	2.9	6.8
GENIE v3	5.33E19	$\sim$0	$\sim$0	$\sim$0	-
NuWro	6.11E20	0.01	0.02	0.1	0.01

- FDS with either different event generators or artificially creation all support the expectation that the model validation is more sensitive to mismodeling than the Xs extraction



Conclusion holds for (signal-only) Xs-only systematics and works for other systematic uncertainties (flux, detector …)

Outcome of “pre-data” interaction model fake data tests

Matt Toups

Outcome of “post-data” model validation tests

	Interaction model systematics are inadequate	Interaction model systematics are adequate
Full systematics are inadequate	Both methods would have specific concerns about bias in the cross section extraction	Indication of a potential for bias in the cross section extraction from an unknown source, not necessarily due to the interaction model (possible Type-I error, mitigated if extracting cross sections a function of “directly observable” quantities)
Full systematics are adequate	Indication of a potential for bias due to the interaction model that the data itself indicates is only realized as a subdominant effect in the cross section extraction (possible Type-II error)	Both methods would not have specific concerns about bias in the cross section extraction

50 Years of Quantum Chromodynamics

Franz Gross[a,1,2], Eberhard Klempt[b,3],

Stanley J. Brodsky[c,4], Andrzej J. Buras[c,5], Volker D. Burkert[c,1], Gudrun Heinrich[c,6], Karl Jakobs[c,7], Curtis A. Meyer[c,8], Kostas Orginos[c,1,2], Michael Strickland[c,9], Johanna Stachel[c,10], Giulia Zanderighi[c,11,12],

Nora Brambilla[5,12,13], Peter Braun-Munzinger[10,14], Daniel Britzger[11], Simon Capstick[15], Tom Cohen[16], Volker Crede[15], Martha Constantinou[17], Christine Davies[18], Luigi Del Debbio[19], Achim Denig[20], Carleton DeTar[21], Alexandre Deur[1], Yuri Dokshitzer[22,23], Hans Günter Dosch[10] Jozef Dudek[1,2], Monica Dunford[24], Evgeny Epelbaum[25], Miguel A. Escobedo[26], Harald Fritzsch[d,27], Kenji Fukushima[28], Paolo Gambino[11,29], Dag Gillberg[30,31], Steven Gottlieb[32], Per Grafstrom[33], Massimiliano Grazzini[34], Boris Grube[1], Alexey Guskov[35], Toru Iijima[36], Xiangdong Ji[16], Frithjof Karsch[37], Stefan Kluth[11], John B. Kogut[38,39], Frank Krauss[40], Shunzo Kumano[41,42], Derek Leinweber[43], Heinrich Leutwyler[44], Hai-Bo Li[45], Yang Li[46], Bogdan Malaescu[47], Chiara Mariotti[48], Pieter Maris[49], Simone Marzani[50], Wally Melnitchouk[1], Johan Messchendorp[51], Harvey Meyer[20], Ryan Edward Mitchell[52], Chandan Mondal[53], Frank Nerling[51,54,55], Sebastian Neubert[3], Marco Pappagallo[56], Saori Pastore[57], José R. Peláez[58], Andrew Puckett[59], Jianwei Qiu[1,2], Klaus Rabbertz[60], Alberto Ramos[61], Patrizia Rossi[1,62], Anar Rustamov[51,63], Andreas Schäfer[64], Stefan Scherer[65], Matthias Schindler[66], Steven Schramm[67], Mikhail Shifman[68], Edward Shuryak[69], Torbjörn Sjöstrand[70], George Sterman[71], Iain W. Stewart[72], Joachim Stroth[51,54,55], Eric Swanson[73], Guy F. de Téramond[74], Ulrike Thoma[3], Antonio Vairo[75], Danny van Dyk[40], James Vary[49], Javier Virto[76,77], Marcel Vos[78], Christian Weiss[1], Markus Wobisch[79] Sau Lan Wu[80], Christopher Young[81], Feng Yuan[82], Xingbo Zhao[53], Xiaorong Zhou[46]

https://arxiv.org/abs/2212.11107

~730 pages of review

About the Role of Event Generator, Model Uncertainties and QCD in the non-perturbative region

The usage of event generators as vehicles in achieving calibration naturally raised another question regarding the model dependence, given the effective-model nature of the event generators. On one hand, it is unlikely the parameters available in the event generators are sufficient in describing interaction modes in the complete phase space. On the other hand, the conservative uncertainties assigned on these parameters may lead to overestimation of systematic uncertainties. Before QCD is solved in the non-perturbative regime, one may never know the exact state of event generators in describing the nature between these two points. In order to mitigate this issue, we propose the model validation procedure that allows us to test whether the model together with its uncertainties can describe the data in a self-consistent manner [2, 20, 21]. During the model validation procedure,