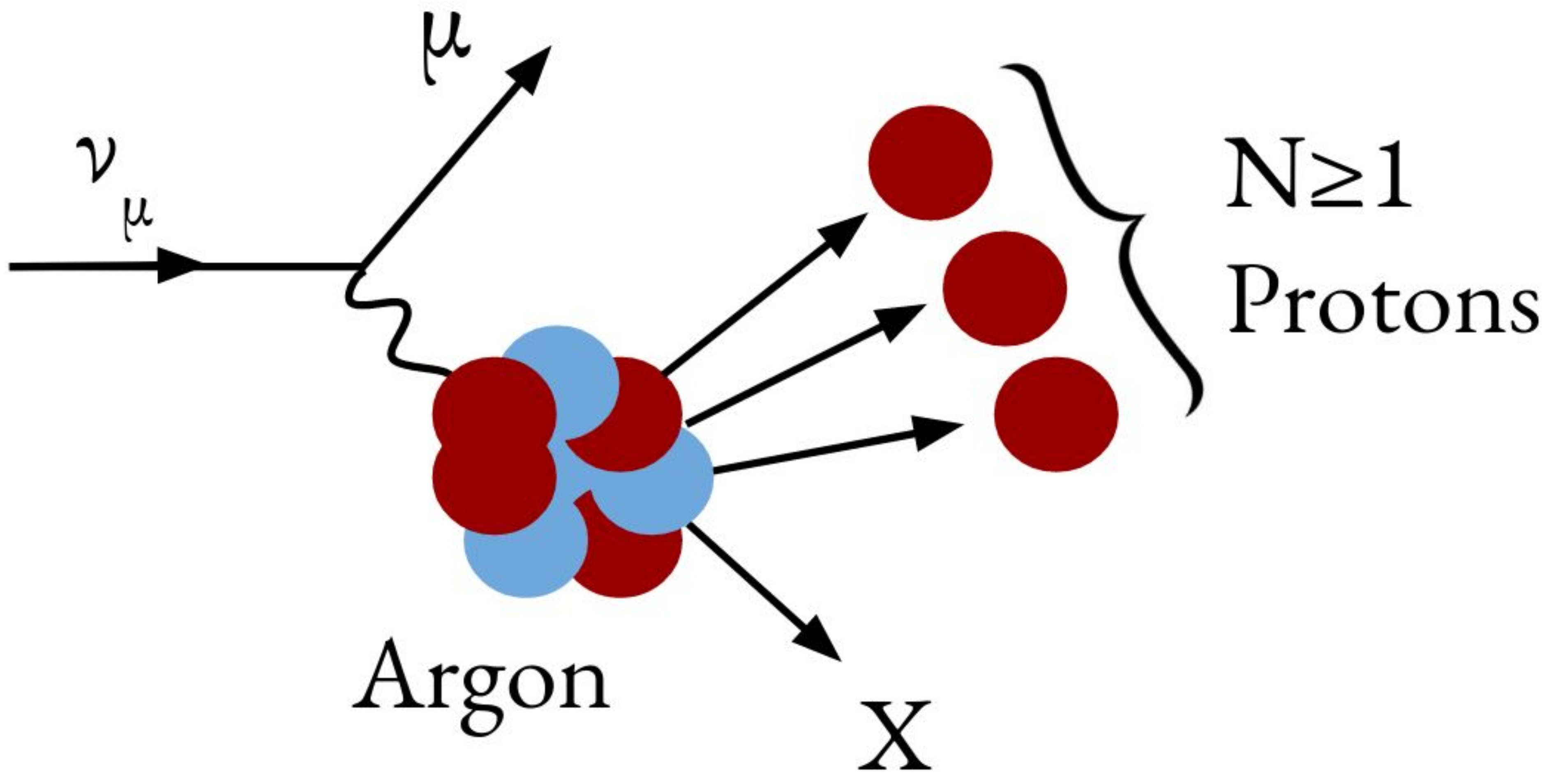
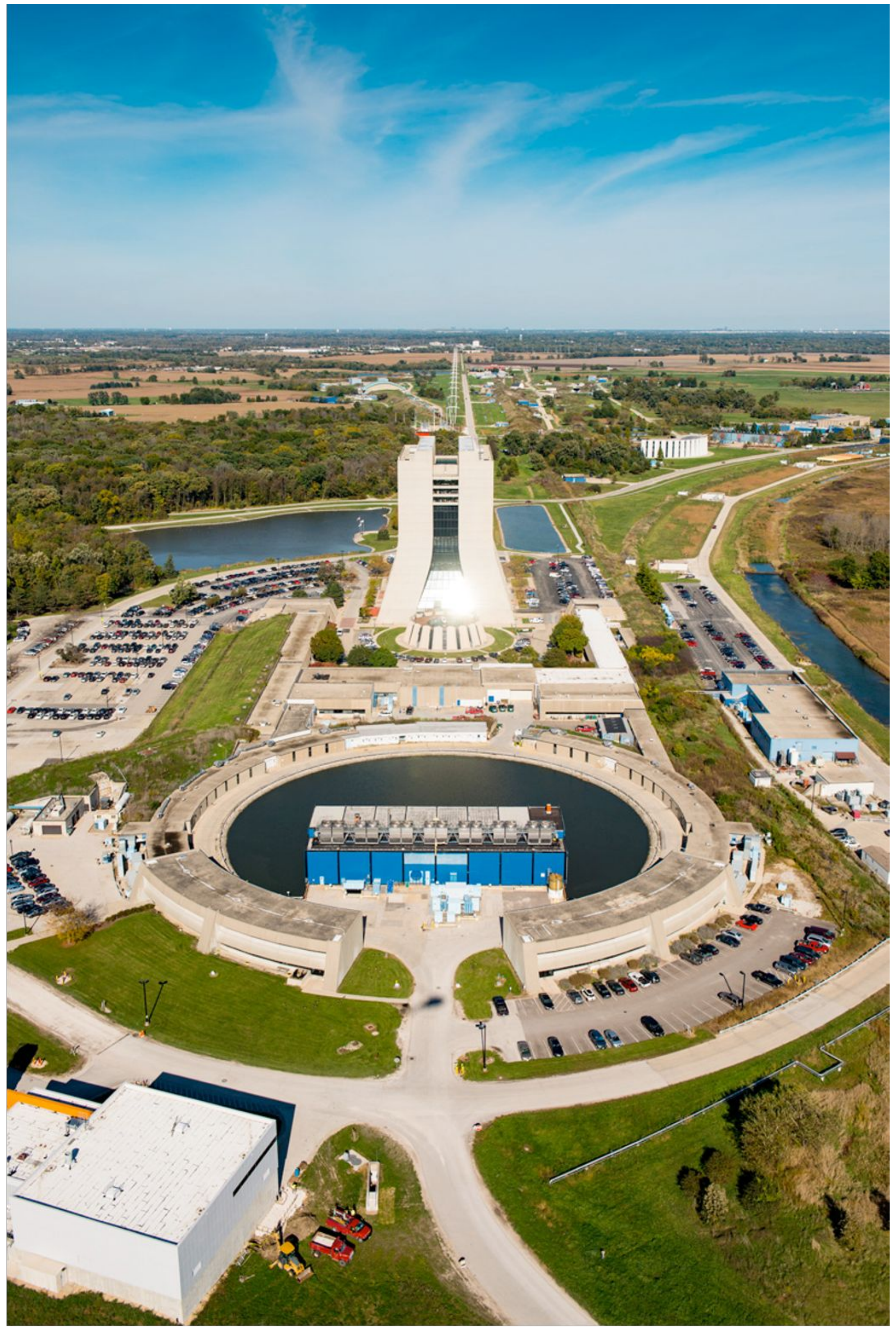
Measurement of double-differential cross sections for mesonless charged-current muon neutrino interactions on argon with final-state protons using the MicroBooNE detector





Steven Gardiner (gardiner@fnal.gov)

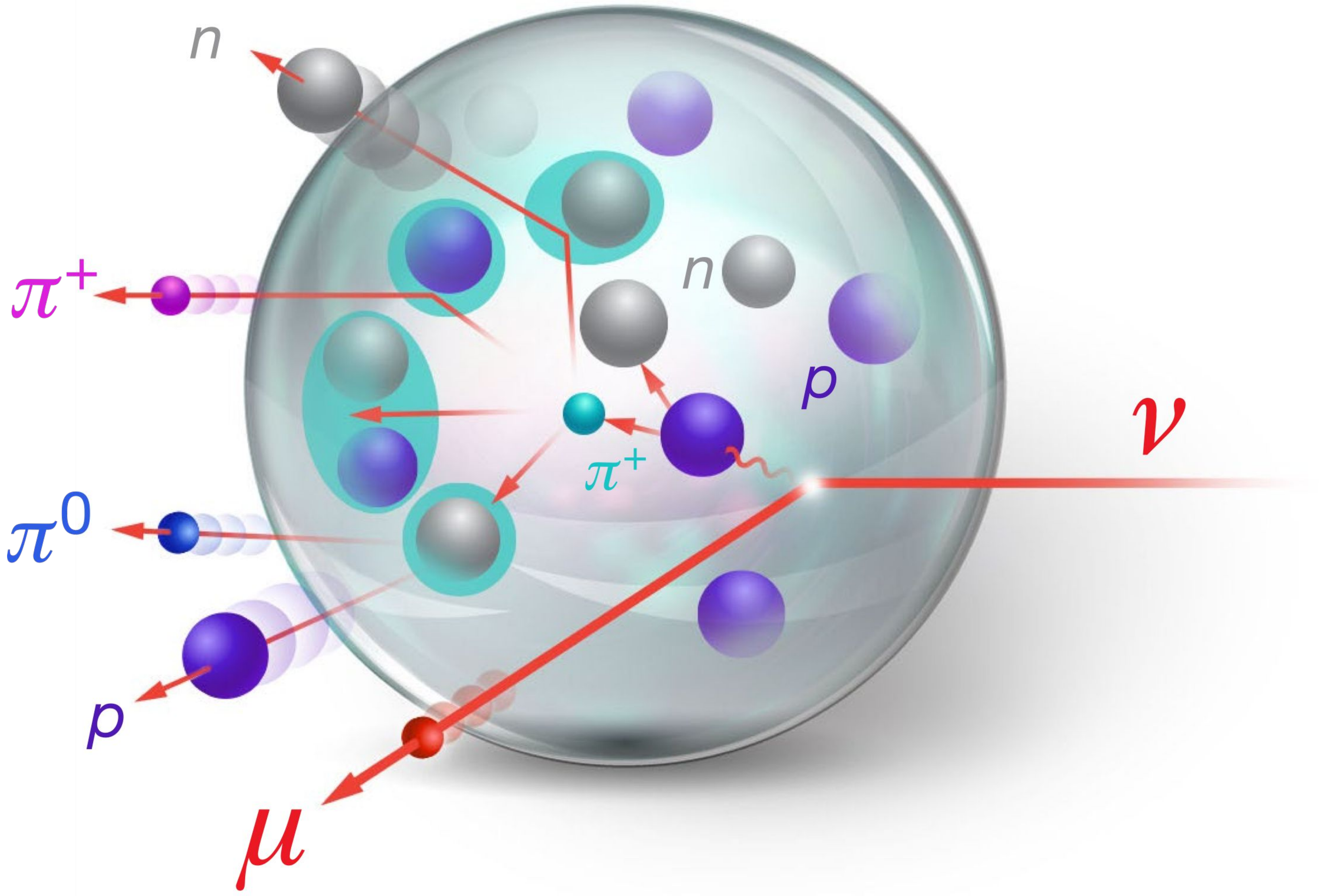
Event Generators Group Leader, Fermilab Physics Simulation Department

NuSTEC Cross-Experiment Working Group, 13 June 2024

μBooNE

Overview

- ν-A measurements designed to **improve modeling** of complex physics
  - Competing channels, nuclear effects, etc.
- This talk presents some ideas for **extending model discrimination power**
  - "Blockwise unfolding"
  - "Conditional Covariance Background Constraint" (CCBC)
- Story partially told using an **actual MicroBooNE analysis**
  - ν_μ CC0πNp cross sections
- Some iteration with organizers on content
  - Technical discussion deemed most interesting for this audience



Inspiration

- Methods proposed here arose from various discussions in MicroBooNE
  - Exploration of unfolding methods, etc.
- Catalyst for actually writing a methods paper: **CERN NuXTract workshop**
  - 2–6 October 2023
  - Clear community interest in discussing these issues
- Will not attempt to summarize those proceedings here
  - Just want to encourage similar future meetings!

https://indico.cern.ch/event/1302529/

Methods paper

- Focus on mathematical procedures used to report analysis results
  - "Cross-section Extraction"
- Short pedagogical introduction with definitions
- Survey of existing techniques in the ~GeV neutrino literature
  - Attempt to be comprehensive for accelerator experiments
  - 3 prevailing "styles"
- Blockwise unfolding and CCBC
  - Motivation
  - Recipes for current experimental use

FERMILAB-PUB-23-692-CSAID

Mathematical methods for neutrino cross-section extraction

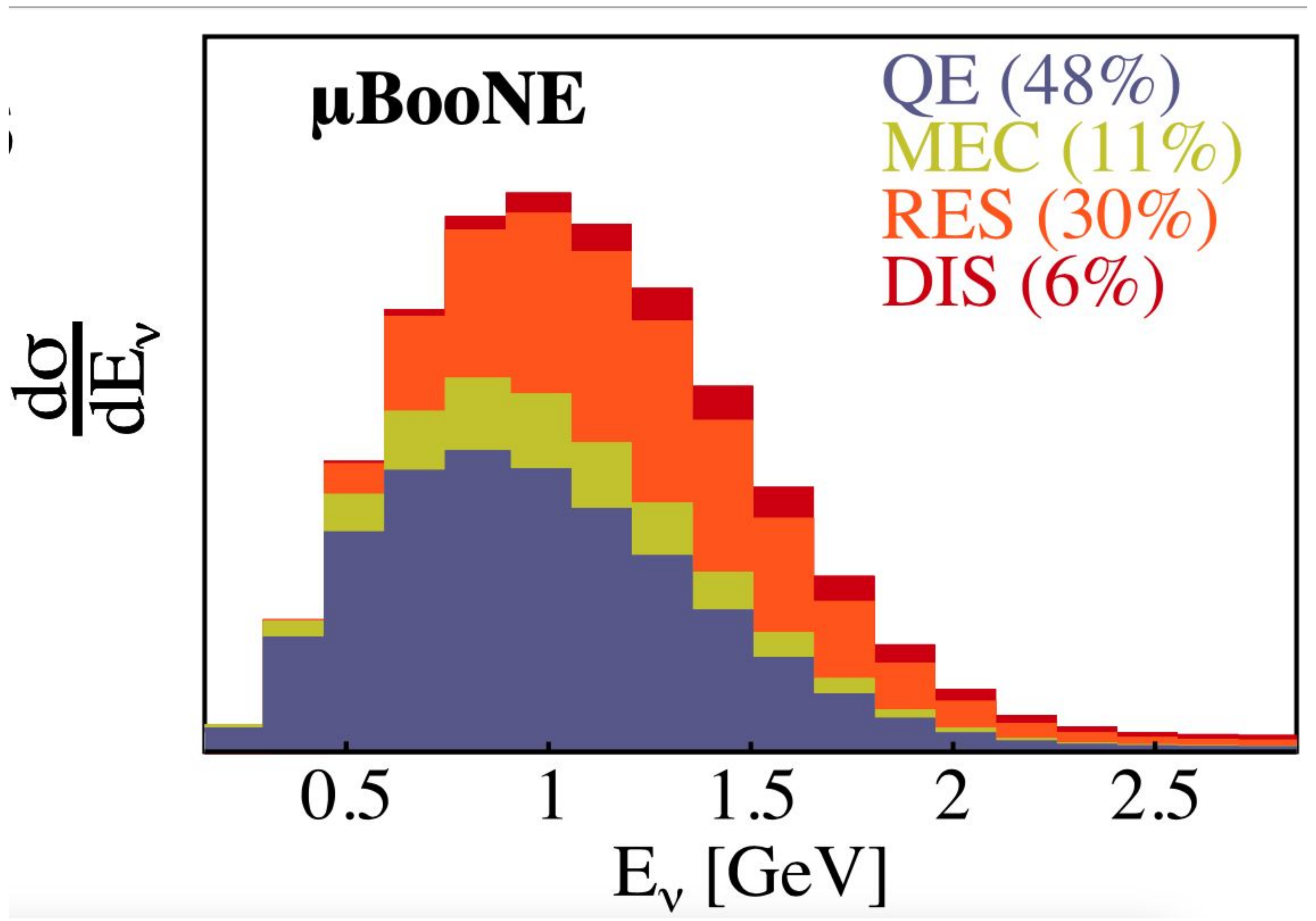
Steven Gardiner*
Fermi National Accelerator Laboratory, Batavia, Illinois 60510 USA
(Dated: January 9, 2024)

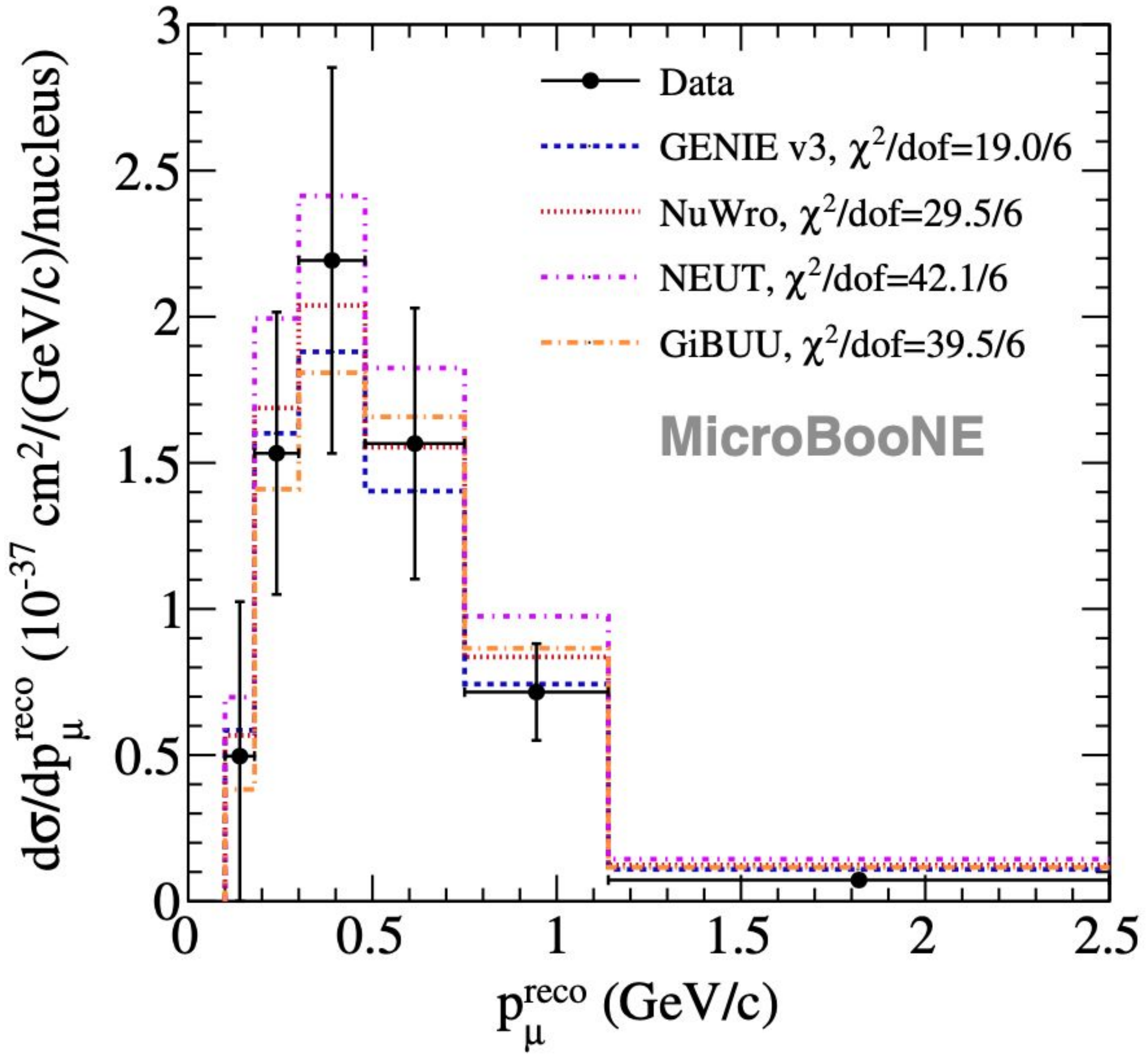
Precise modeling of neutrino-nucleus scattering is becoming increasingly important as accelerator-based oscillation experiments seek definitive answers to open questions about neutrino properties. To guide the needed model refinements, a growing number of experimental collaborations are pursuing a wide-ranging program of neutrino interaction measurements at GeV energies. A key step in most such analyses is cross-section extraction, in which measured event counts are corrected for background contamination and imperfect detector performance to yield cross-section results that are directly comparable to theoretical predictions. In this paper, I review the major approaches to cross-section extraction in the literature using representative examples from the MINERvA, MicroBooNE, and T2K experiments. I then present two mathematical techniques, blockwise unfolding and the conditional covariance background constraint, which overcome some limitations of typical cross-section extraction procedures.

arXiv:2401.04065

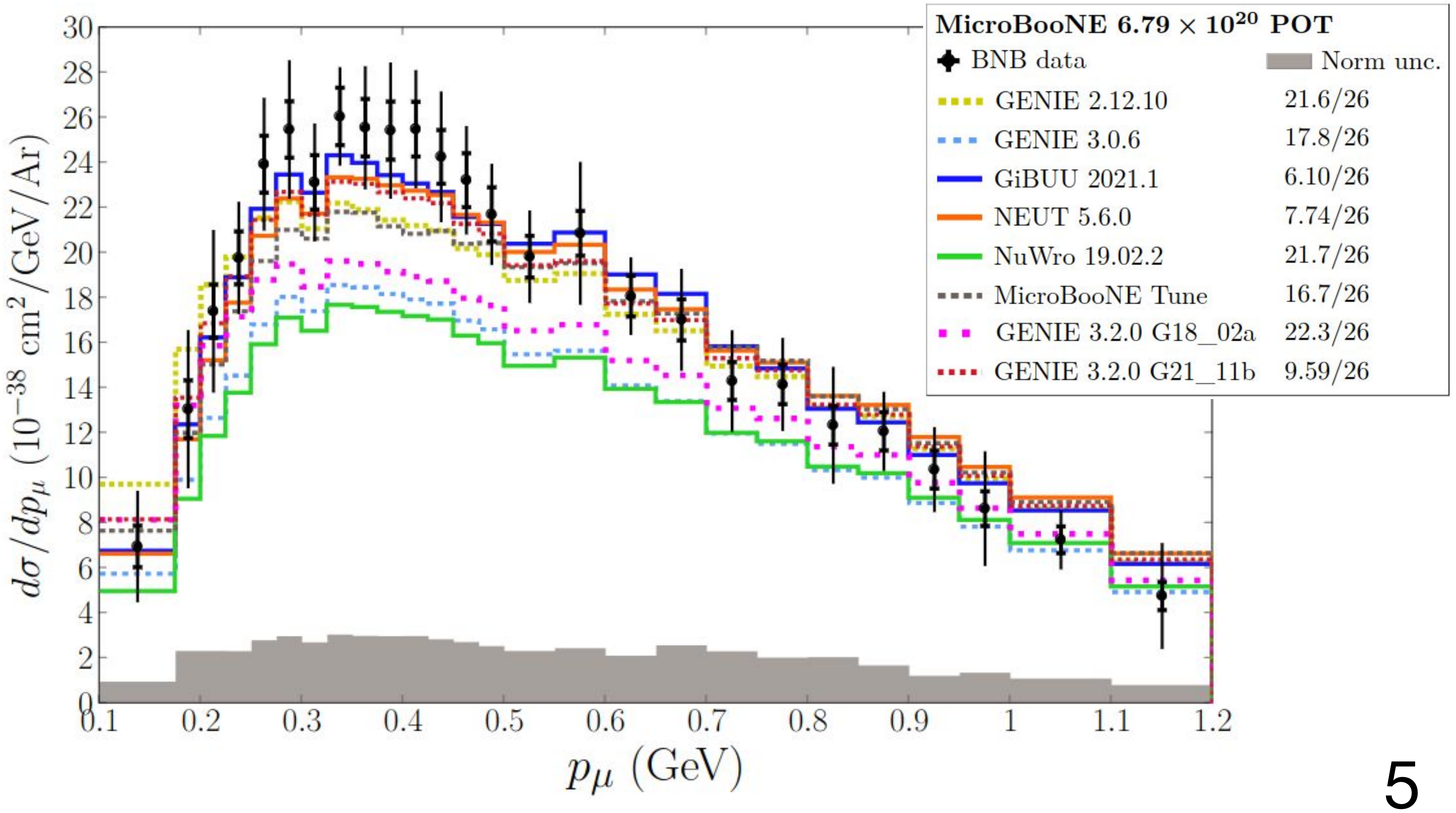
"Worked example" of blockwise unfolding in MicroBooNE analysis

- Measurement of **ν_μ CC0πNp** interactions (N ≥ 1)
  - Mostly QE+MEC with contribution from RES+FSI
  - Dominant topology at BNB energies





- **Prior analysis** from MicroBooNE: [Phys. Rev. D 102, 112013 (2020)]
  - Far greater detail now due to ~4x more data, improved systematics
- **New paper** ([arXiv:2403.19574]) provides extensive technical documentation
  - See especially lengthy supplement!



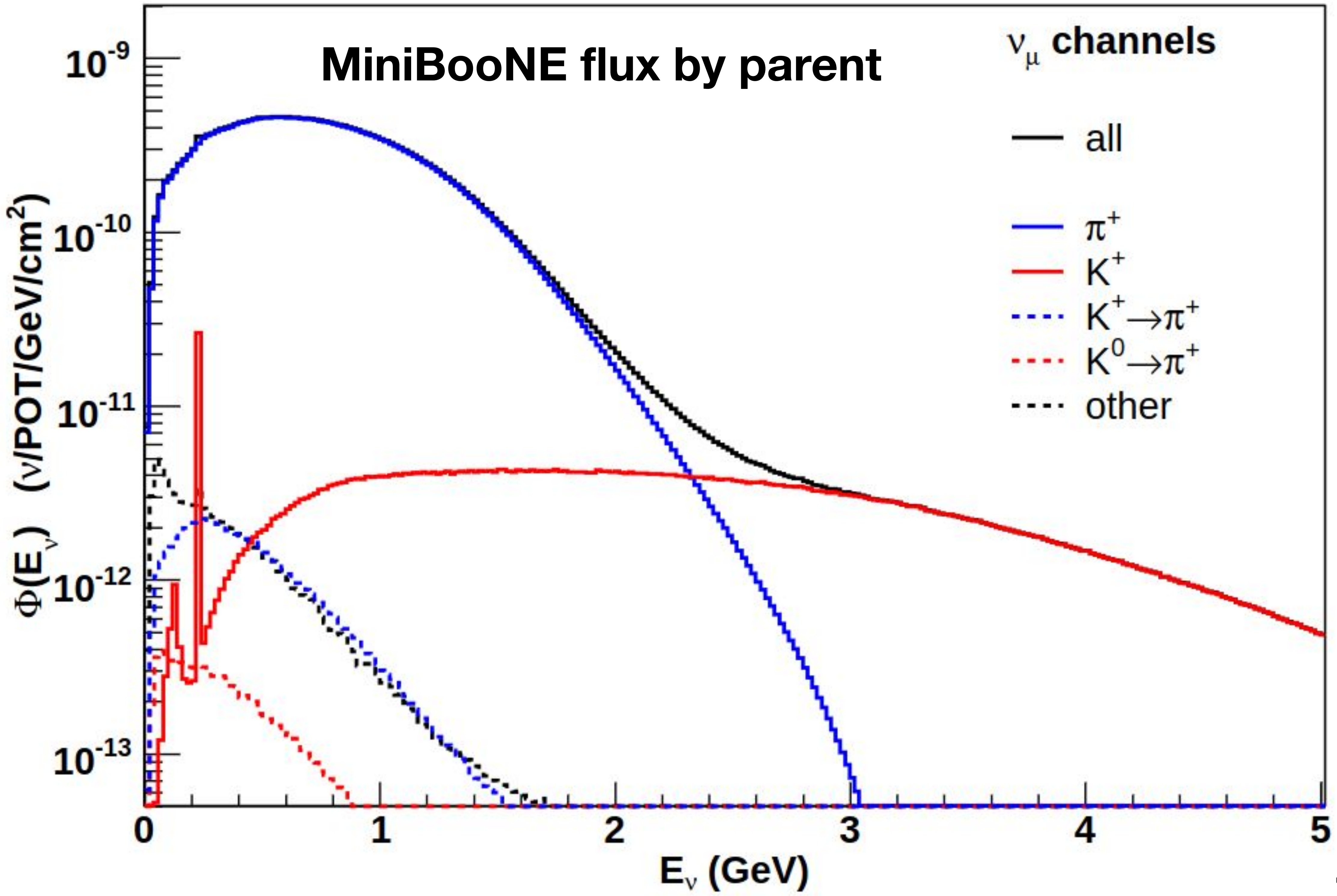
“All models are wrong, but some are useful” — George Box

The flux-averaged differential cross section

- Quantity of interest for most neutrino interaction analyses
- Folded with the beam energy spectrum
  - Flux normalization cancels out
- Avoid difficulty of direct neutrino energy reconstruction
  - Beams are generated by particle decays — broad energy range

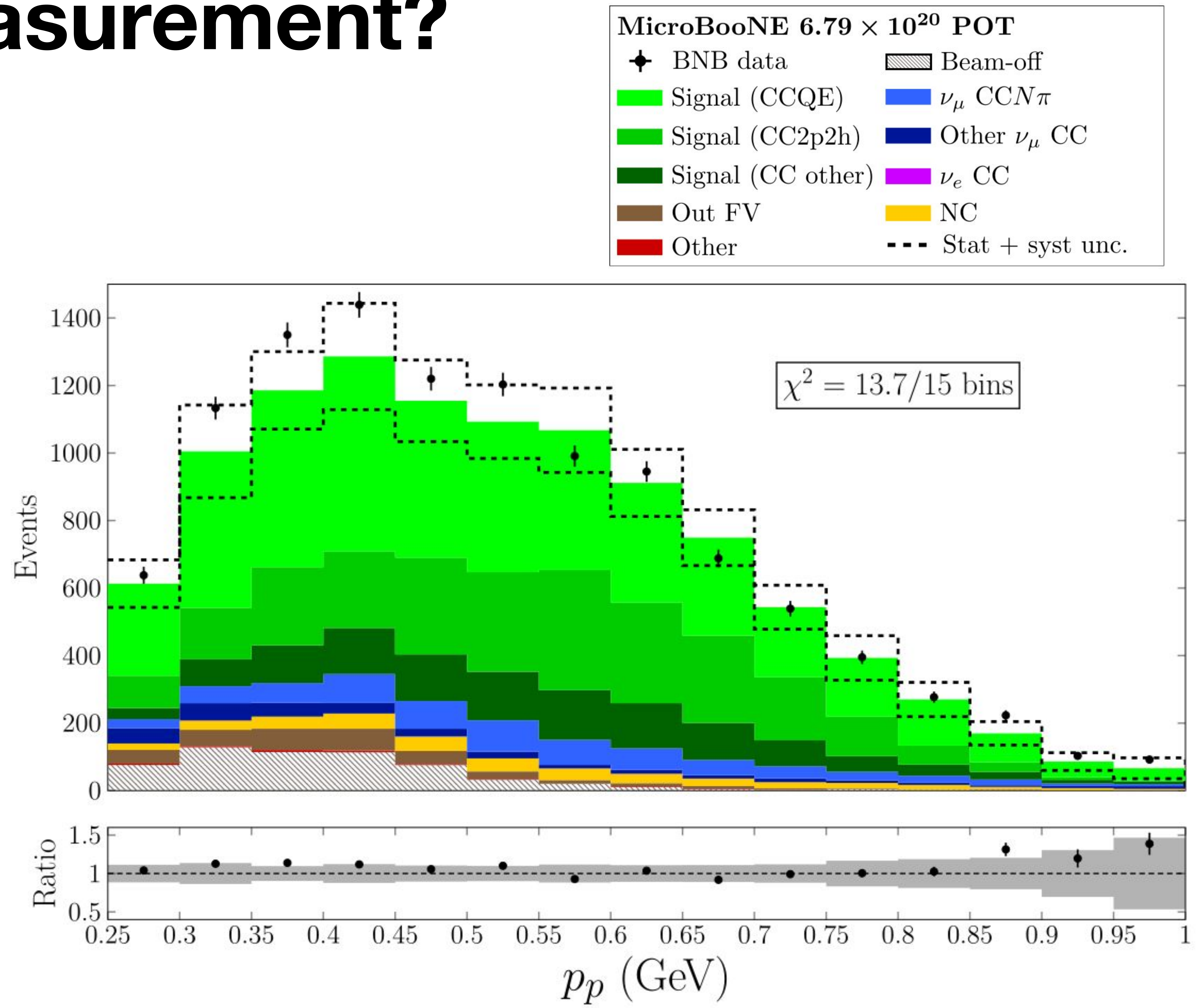
$$\left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle \equiv \frac{1}{\Phi} \int \varphi(E_\nu)\, \frac{d^n\sigma(E_\nu)}{d\mathbf{x}}\, dE_\nu$$

$$\Phi \equiv \int \varphi(E_\nu)\, dE_\nu$$



How do we perform the measurement?

- Counting experiment: bin for variable(s) of interest
- Raw event counts comparable to simulation
  - Only feasible by the experimental collaboration
- **Cross-section extraction**
  - Converts this measurement to a result anyone can use
  - Details vary across experiments
- Many subtleties, care must be taken to avoid bias

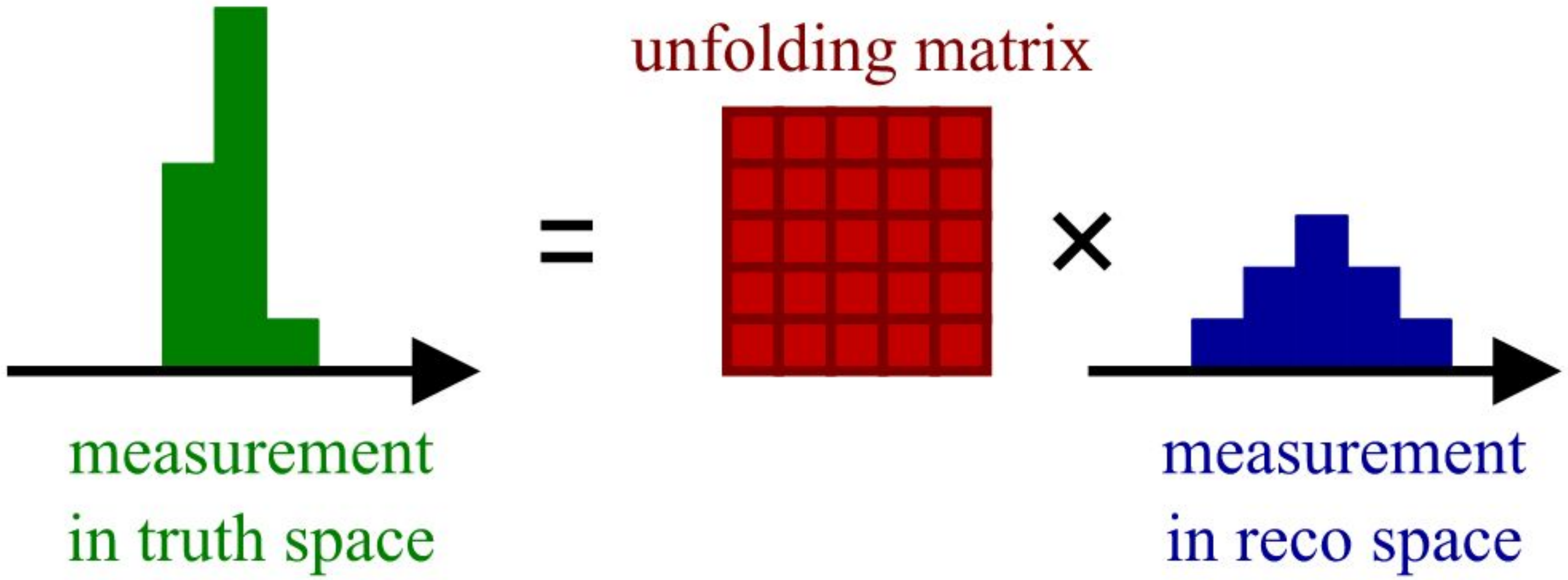


How do we perform the measurement?

- **Flux-averaged differential cross section**
  - true bins *μ*, reco bins *a*
  - Average value in true bin *μ*

$$\left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu = \frac{\sum_a U_{\mu a}\left(D_a - B_a\right)}{\Phi\, T\, \Delta \mathbf{x}_\mu}$$

- **Unfolding matrix** U accounts for inefficiency and bin migrations



- **Unfolded space ≈ true space**
  - Systematics must be considered carefully

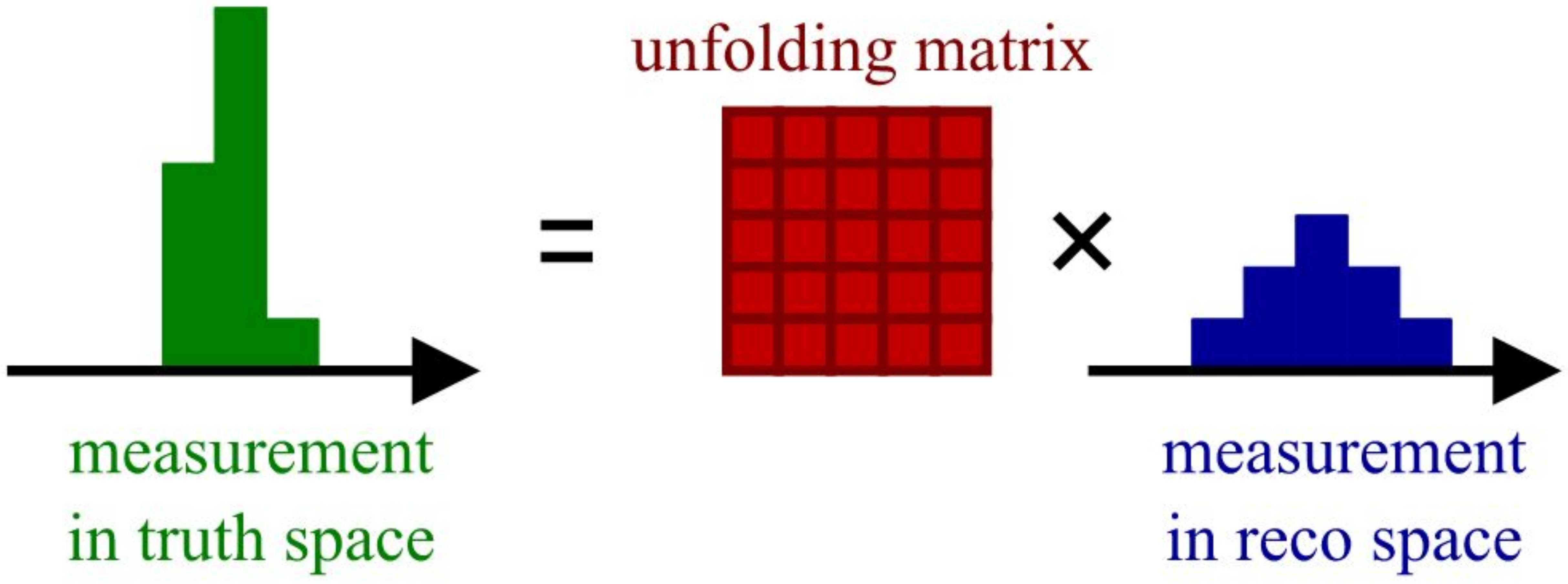
$$U_{\mu a} = \frac{P_{\mu a}}{\epsilon_\mu}$$

How do we perform the measurement?

- **Flux-averaged differential cross section**
  - true bins μ, reco bins a
  - **Average value** in true bin μ

$$\left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu \equiv \frac{1}{\Delta \mathbf{x}_\mu} \int_{\mathbf{x}_\mu}^{\mathbf{x}_{\mu+1}} \left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle d\mathbf{x}$$

- **Unfolding matrix** U accounts for inefficiency and bin migrations



- **Unfolded space ≈ true space**
  - Systematics must be considered carefully

$$U_{\mu j} = \frac{P(\mu|j)}{\epsilon_\mu}$$

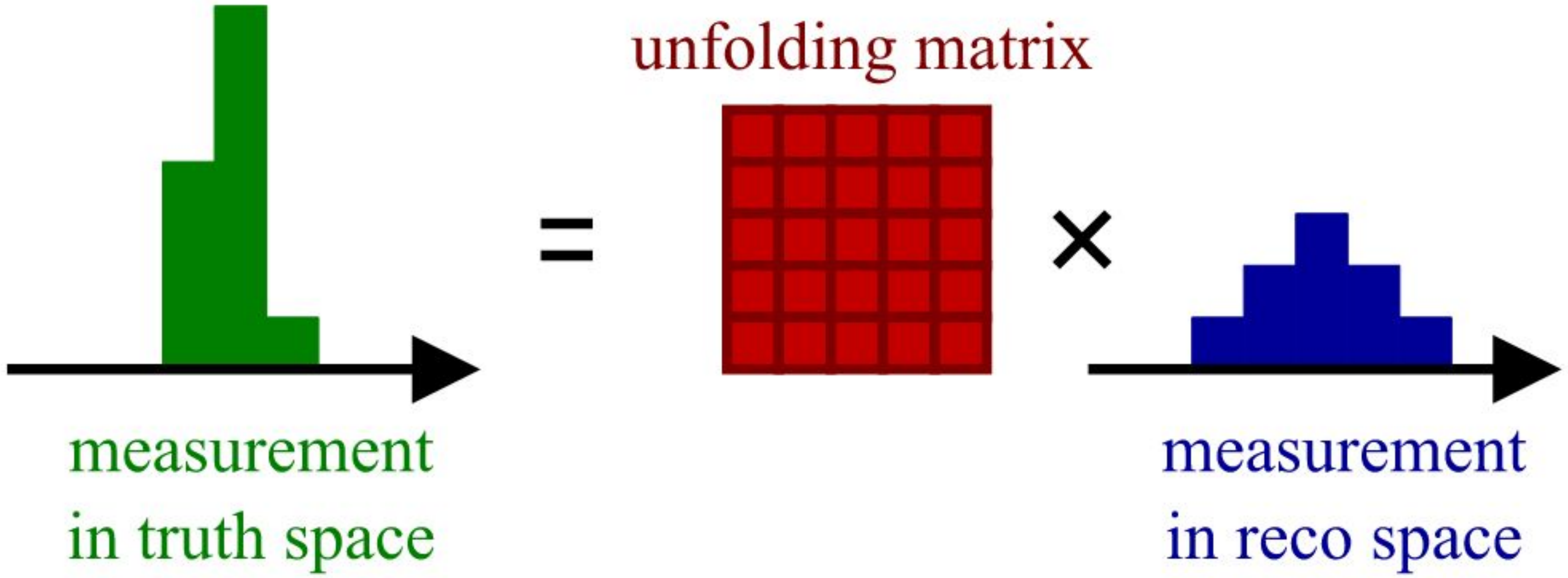
Styles of cross-section extraction

- Superficially, everyone plays the same game, but differently
  - **3 major approaches** at GeV scale, the rest are perturbations
  - Details are often not spelled out, especially for *Phys. Rev. Lett*.

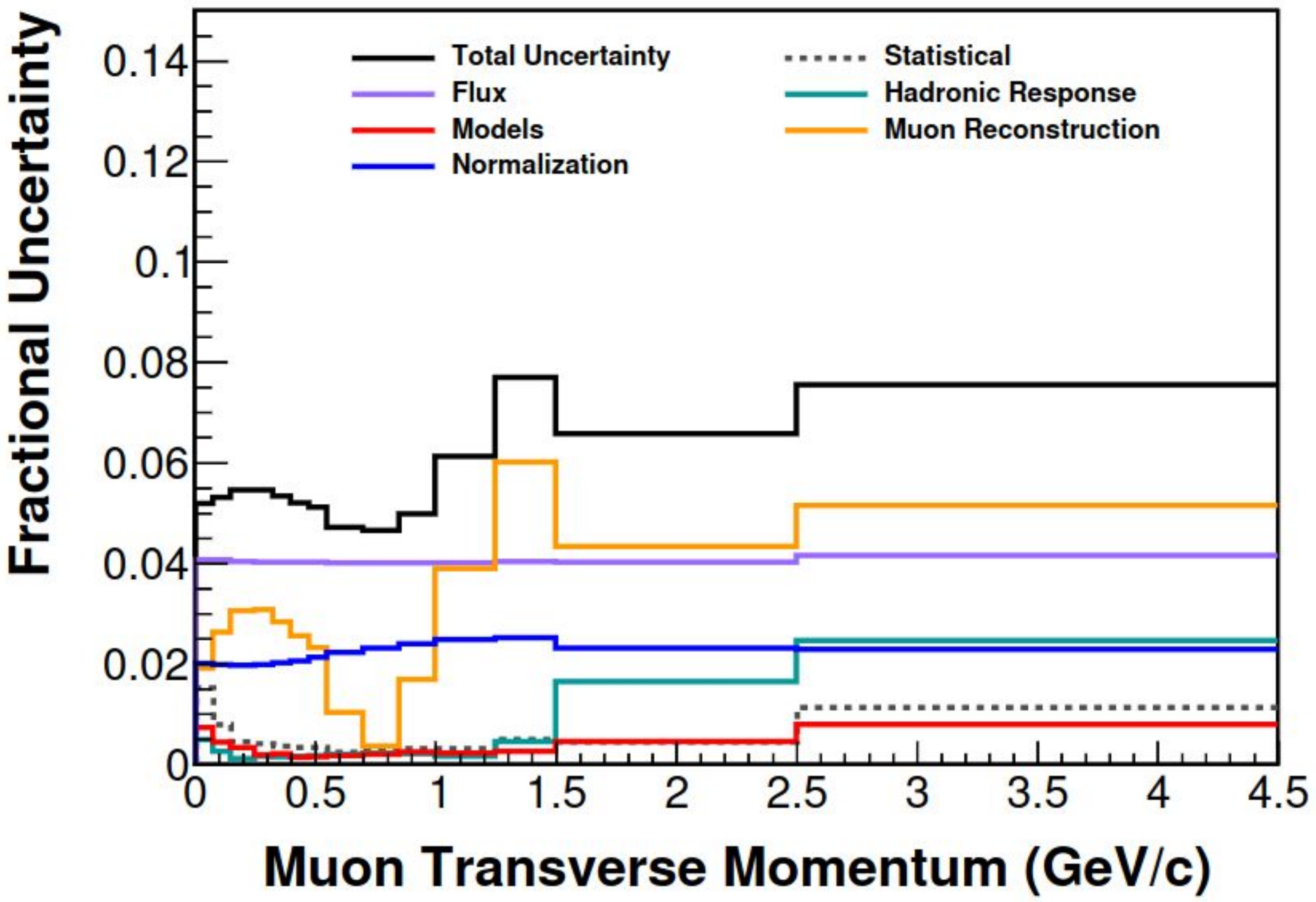
- **MINERvA**
  - D'Agostini iterative recipe for building unfolding matrix U
  - Uncertainties: repeat extraction, take spread between "universes"
- **MicroBooNE**
  - Wiener-SVD unfolding
  - Compute total covariance on event counts, propagate through unfolding
- **T2K**
  - Perform likelihood fit to event counts (huge number of parameters)
  - Uncertainties can be treated two ways
    - Repeat the fit across many universes (MINERvA-esque)
    - Vary parameters according to post-fit covariance matrix

MINERvA style

"Canonical" approach, widely used by other experiments



Phys. Rev. D 104, 092007 (2021)



D'Agostini iterative unfolding (i = 0,1,...)

$$\hat{\phi}_{\mu}^{i+1} = \sum_{a} U_{\mu a}^{i}\, d_a$$

$$U_{\mu a} = \frac{P_{\mu a}}{\epsilon_\mu} \qquad P_{\mu a}^{i} = \frac{M_{a\mu}\,\hat{\phi}_{\mu}^{i}}{\sum_\lambda M_{a\lambda}\,\hat{\phi}_{\lambda}^{i}}$$

Extraction repeated in multiple "universes"

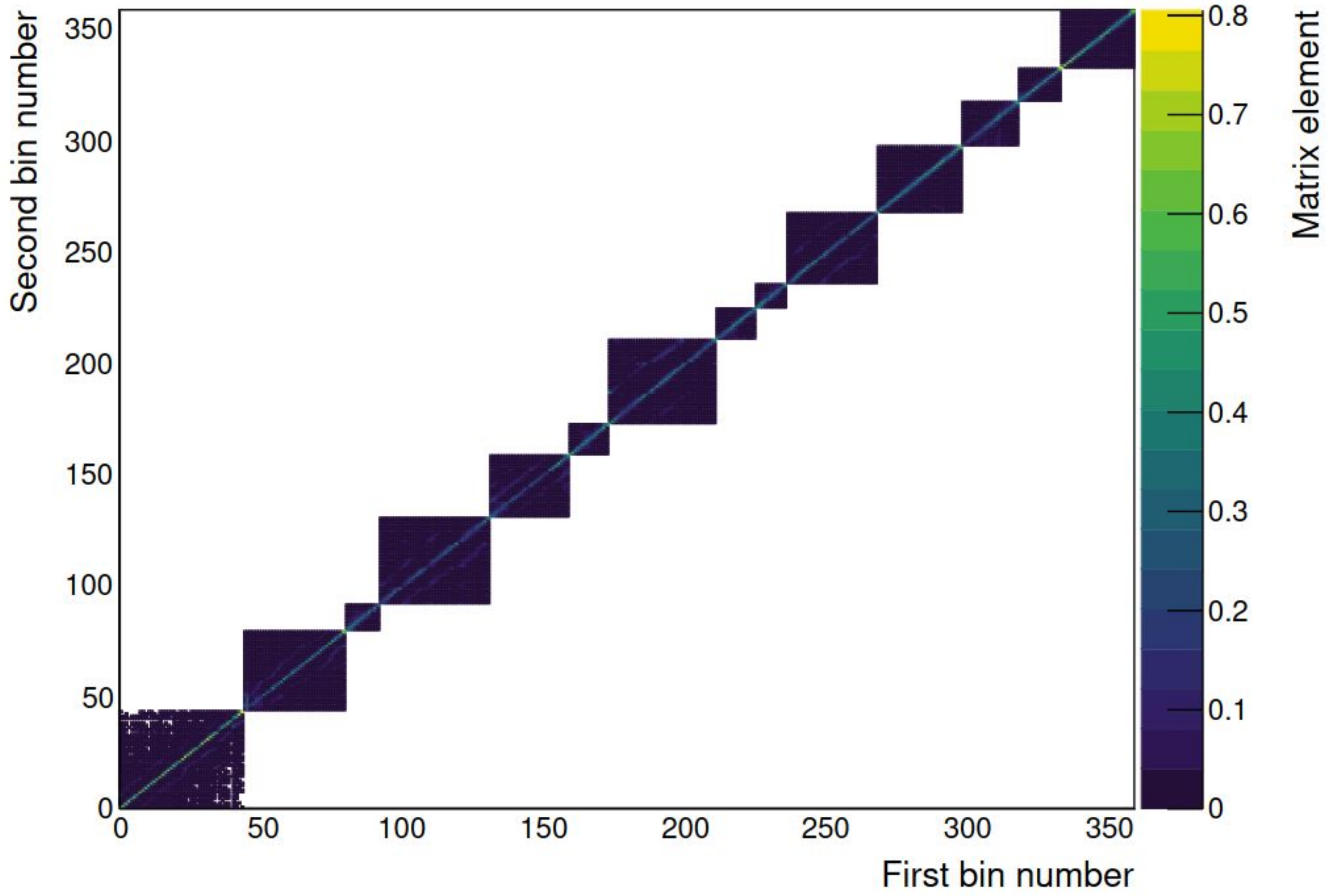
$$\mathrm{Cov}(s_\mu, s_\lambda) = \frac{1}{N_{\mathrm{univ}}} \sum_{u=1}^{N_{\mathrm{univ}}} (s_\mu^u - \bar{s}_\mu)(s_\lambda^u - \bar{s}_\lambda)$$

$$\left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu = \frac{\sum_a U_{\mu a}(D_a - B_a)}{\Phi\, T\, \Delta \mathbf{x}_\mu} \qquad s_\mu \equiv \left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu$$

MicroBooNE style

Two major differences relative to MINERvA

arXiv:2403.19574



Smear theory predictions by regularization matrix Ac

$$\Delta_{a\mu} \equiv \epsilon_\mu M_{a\mu} \qquad U^{\text{direct}} = (\Delta^T \Delta)^{-1} \Delta^T$$

$$U = A_C \cdot U^{\text{direct}} \quad \Rightarrow \quad A_C = U \cdot \Delta$$

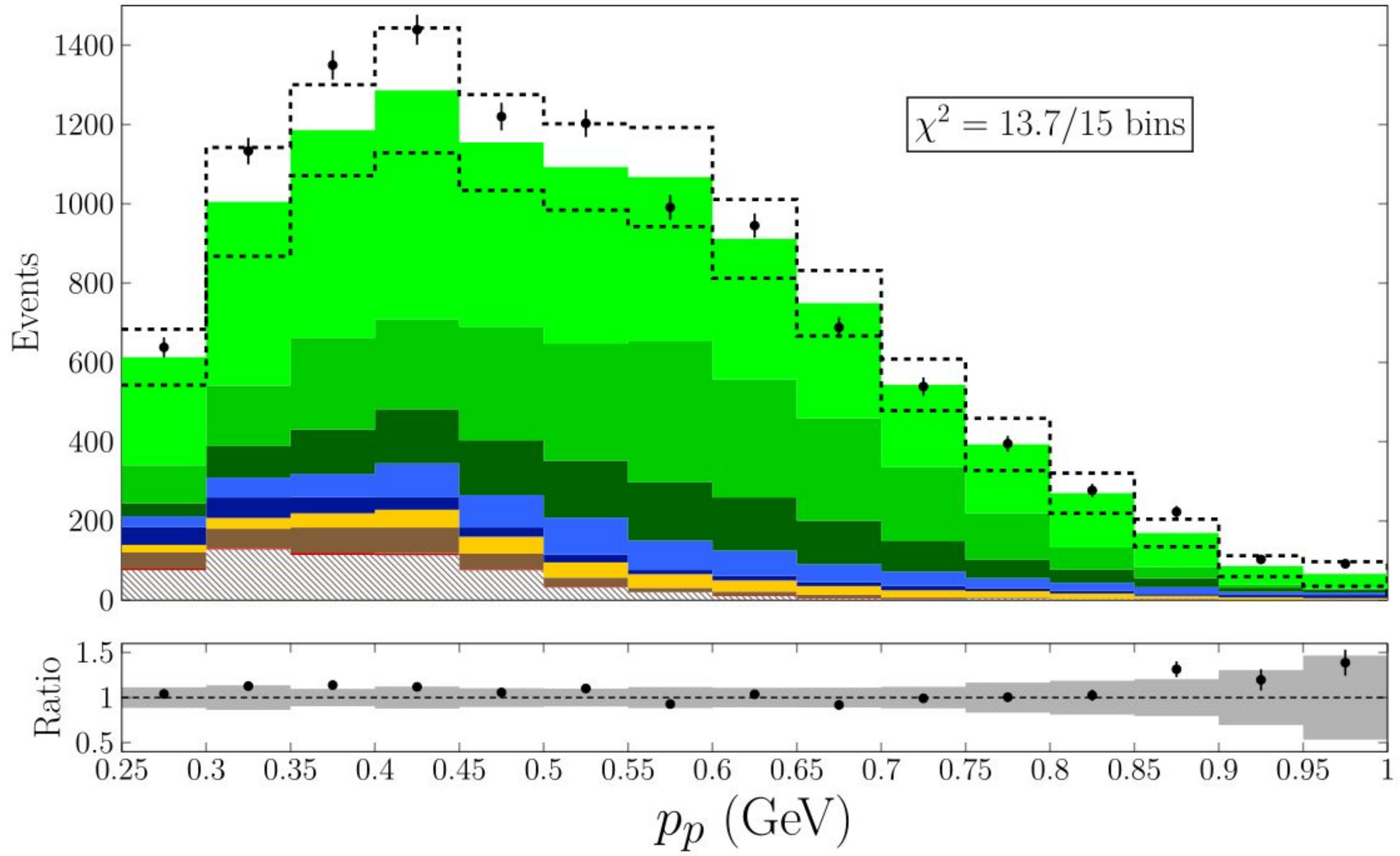
Introduced with Wiener-SVD unfolding paper, see also related article by Lukas Koch. Can use with other unfoldings.

- **Direct unfolding** = maximum likelihood estimate, but large variance
- Standard methods introduce prior information to reduce variance
  - Cost is (hopefully small) bias, "**regularization**"
- Ac allows regularization to be applied consistently to theory
  - Recovers χ^2 post-Wiener-SVD as if you didn't unfold

MicroBooNE style

Two major differences relative to MINERvA

arXiv:2403.19574



Analytic error propagation

$$\mathrm{Cov}(D_a, D_b) \approx \mathrm{Cov}(n_a, n_b) = \frac{1}{N_{\mathrm{univ}}} \sum_{u=1}^{N_{\mathrm{univ}}} \left(n_a^u - n_a^{\mathrm{CV}}\right)\left(n_b^u - n_b^{\mathrm{CV}}\right)$$

$$\mathrm{Cov}(\hat{\phi}_\mu, \hat{\phi}_\lambda) = \sum_{a,b} \mathfrak{E}_{\mu a}\, \mathrm{Cov}(d_a, d_b)\, \mathfrak{E}_{b\lambda}^T$$

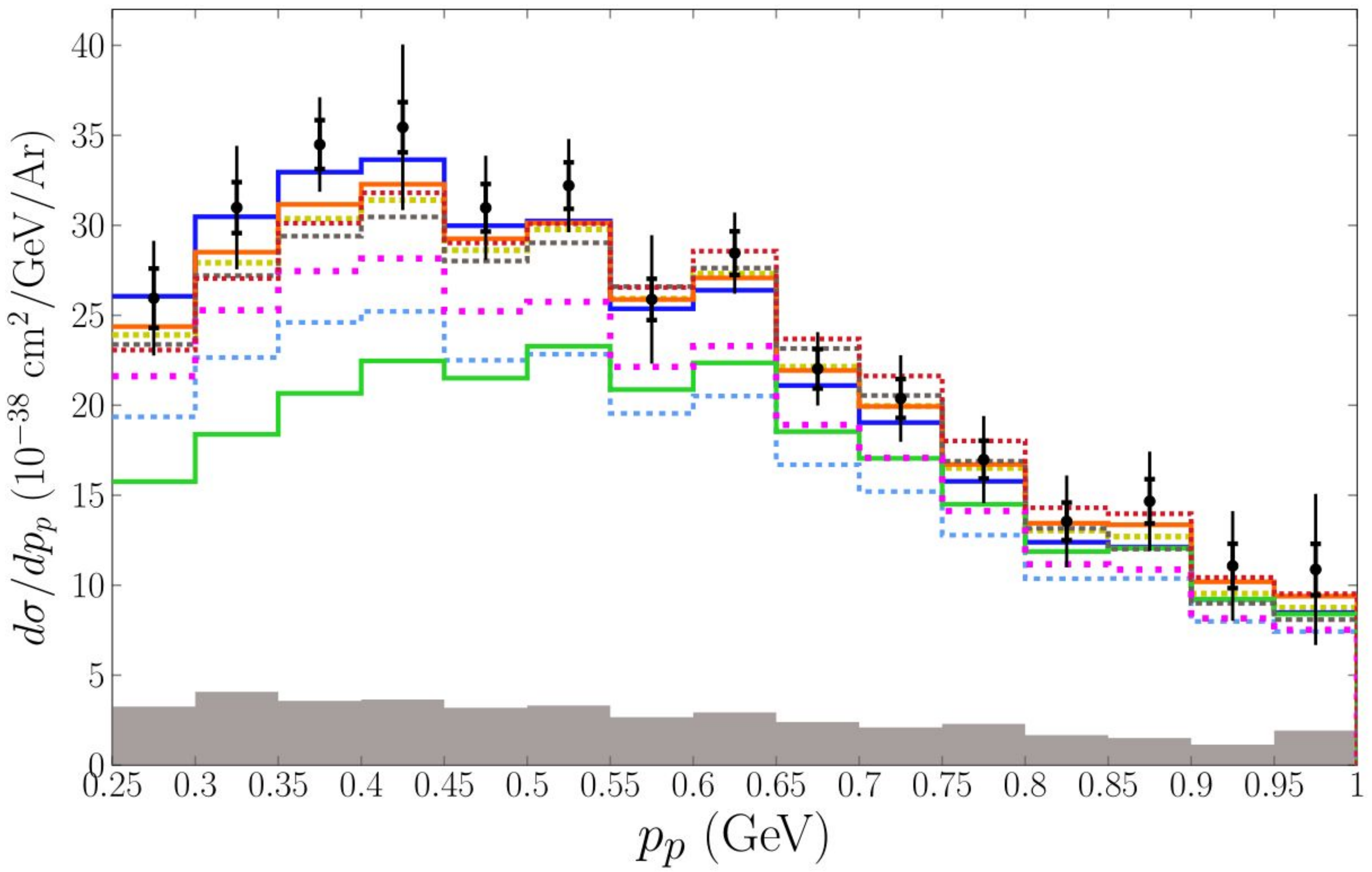
Evaluate uncertainties on reconstructed result, unfold once

- **Required** in order to apply Wiener-SVD unfolding method, compatible with others
  - Ingredients include the reconstructed-space covariance matrix (based on MC)
- **Consistent linear transformation** applied to
  - Background-subtracted data
  - Their covariances

MicroBooNE style

Two major differences relative to MINERvA

[arXiv:2403.19574](https://arxiv.org/abs/2403.19574)



Analytic error propagation

$$\mathrm{Cov}(D_a, D_b) \approx \mathrm{Cov}(n_a, n_b) = \frac{1}{N_{\mathrm{univ}}} \sum_{u=1}^{N_{\mathrm{univ}}} \left(n_a^u - n_a^{\mathrm{CV}}\right)\left(n_b^u - n_b^{\mathrm{CV}}\right)$$

$$\mathrm{Cov}(\hat{\phi}_\mu, \hat{\phi}_\lambda) = \sum_{a,b} \mathfrak{E}_{\mu a}\, \mathrm{Cov}(d_a, d_b)\, \mathfrak{E}_{b\lambda}^T$$

Evaluate uncertainties on reconstructed result, unfold once

- **Required** in order to apply Wiener-SVD unfolding method, compatible with others
  - Ingredients include the reconstructed-space covariance matrix (based on MC)
- **Consistent linear transformation** applied to
  - Background-subtracted data
  - Their covariances

An aside about the uncertainty propagation

Error propagation matrix built from partial derivatives

$$\mathrm{Cov}(\hat{\phi}_\mu, \hat{\phi}_\lambda) = \sum_{a,b} \mathfrak{E}_{\mu a}\, \mathrm{Cov}(d_a, d_b)\, \mathfrak{E}^T_{b\lambda} \qquad \mathfrak{E}_{\mu a} \equiv \frac{\partial \hat{\phi}_\mu}{\partial d_a}$$

For Wiener-SVD, this is just the unfolding matrix (**recipe does not depend on data**)

$$\mathfrak{E}_{\mu a} = U_{\mu a}$$

Conversion to cross-section units from unfolded event counts

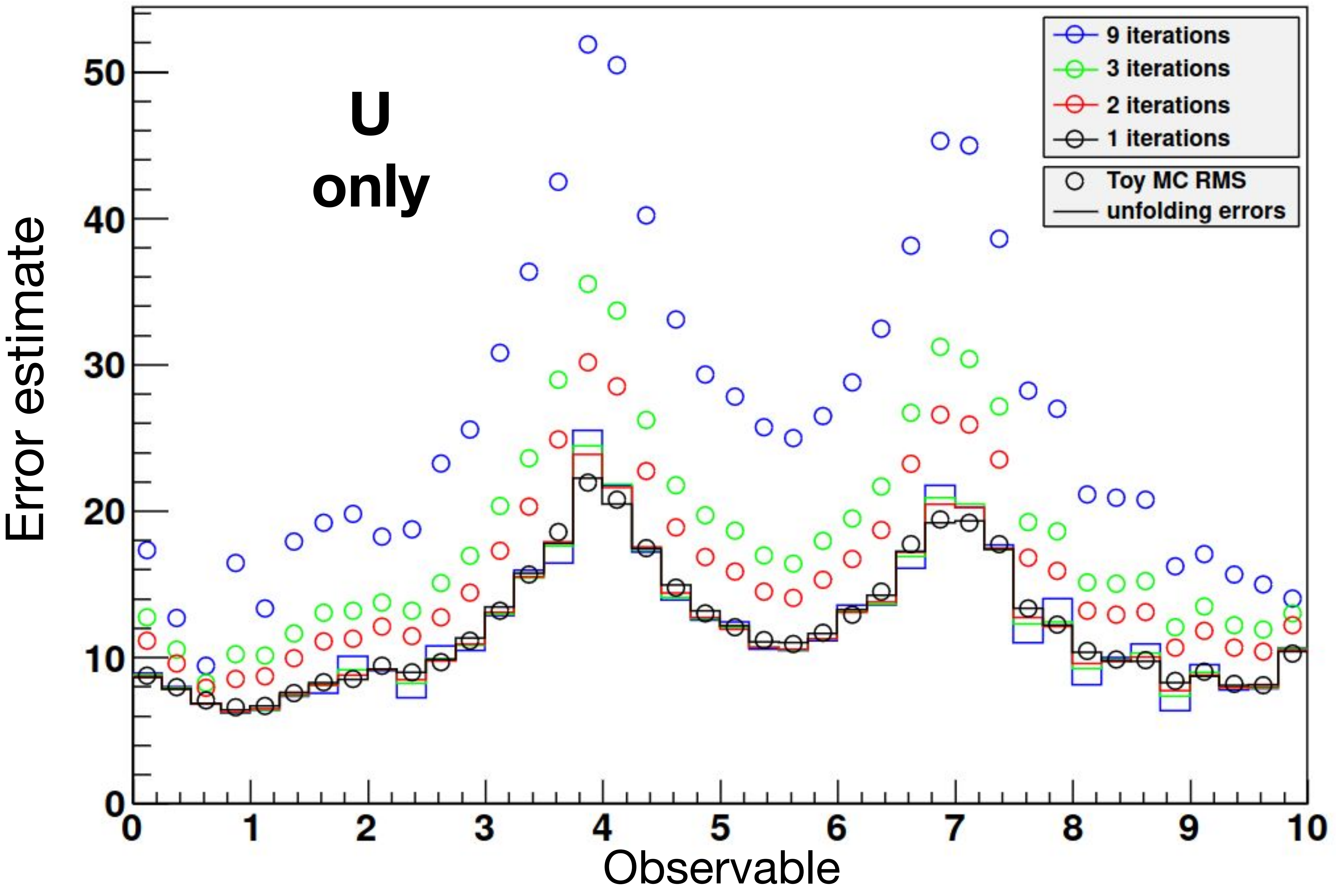
$$\mathrm{Cov}(s_\mu, s_\lambda) = \frac{\mathrm{Cov}(\hat{\phi}_\mu, \hat{\phi}_\lambda)}{\Phi^2\, T^2\, \Delta \mathbf{x}_\mu\, \Delta \mathbf{x}_\lambda} \qquad s_\mu \equiv \left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu$$

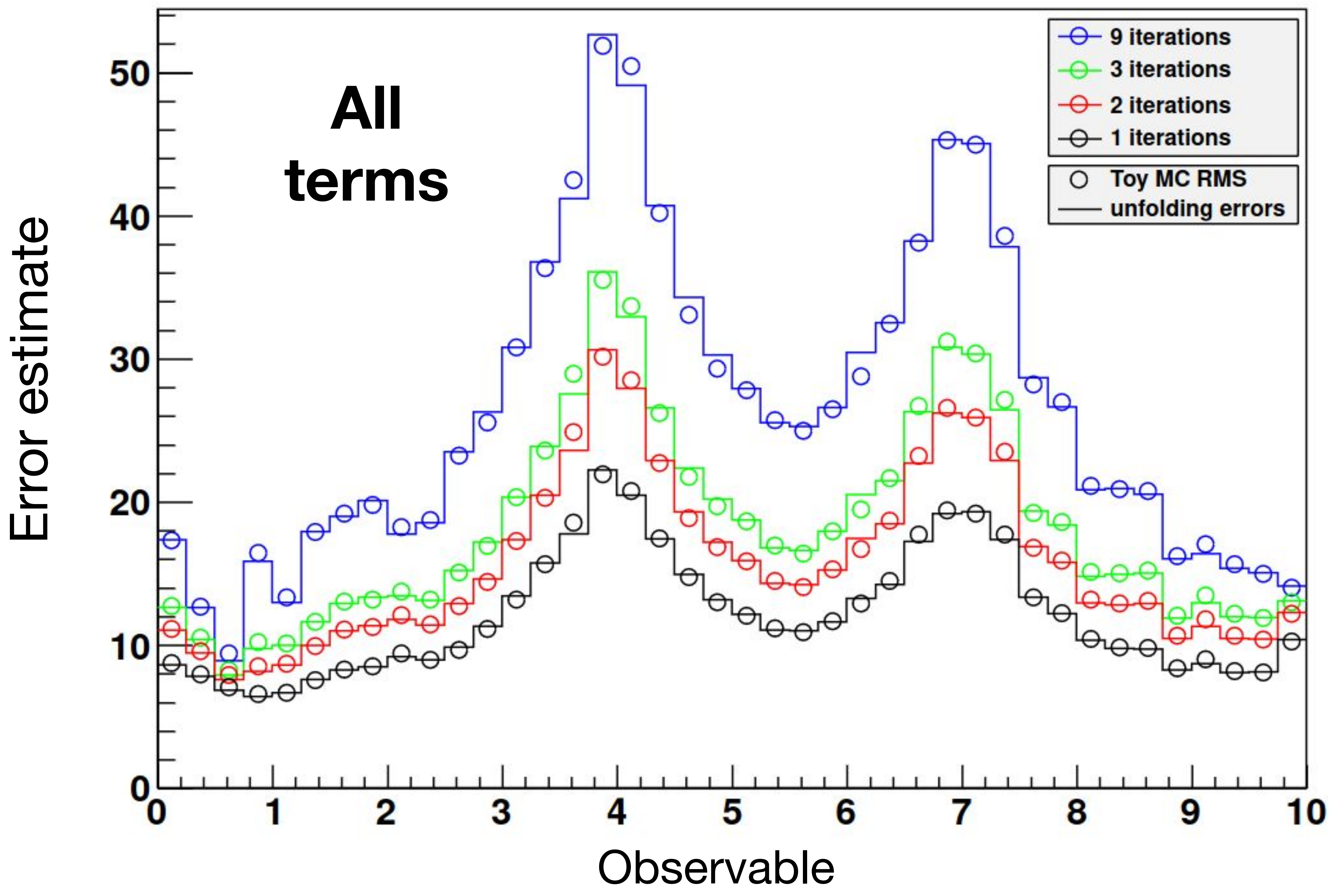
An aside about the uncertainty propagation

For D'Agostini, unfolding **depends on data** after initial iteration

$$\mathfrak{E}^{i+1}_{\mu a} = \frac{\partial \hat{\phi}^{i+1}_{\mu}}{\partial d_a} = U^i_{\mu a} + \frac{\hat{\phi}^{i+1}_{\mu}}{\hat{\phi}^{i}_{\mu}} \mathfrak{E}^i_{\mu a} - \sum_{\lambda, b} \epsilon_\lambda \frac{d_b}{\hat{\phi}^i_\lambda} U^i_{\mu b} U^i_{\lambda b} \mathfrak{E}^i_{\lambda a}$$

Neglecting extra terms (left) leads to under-coverage



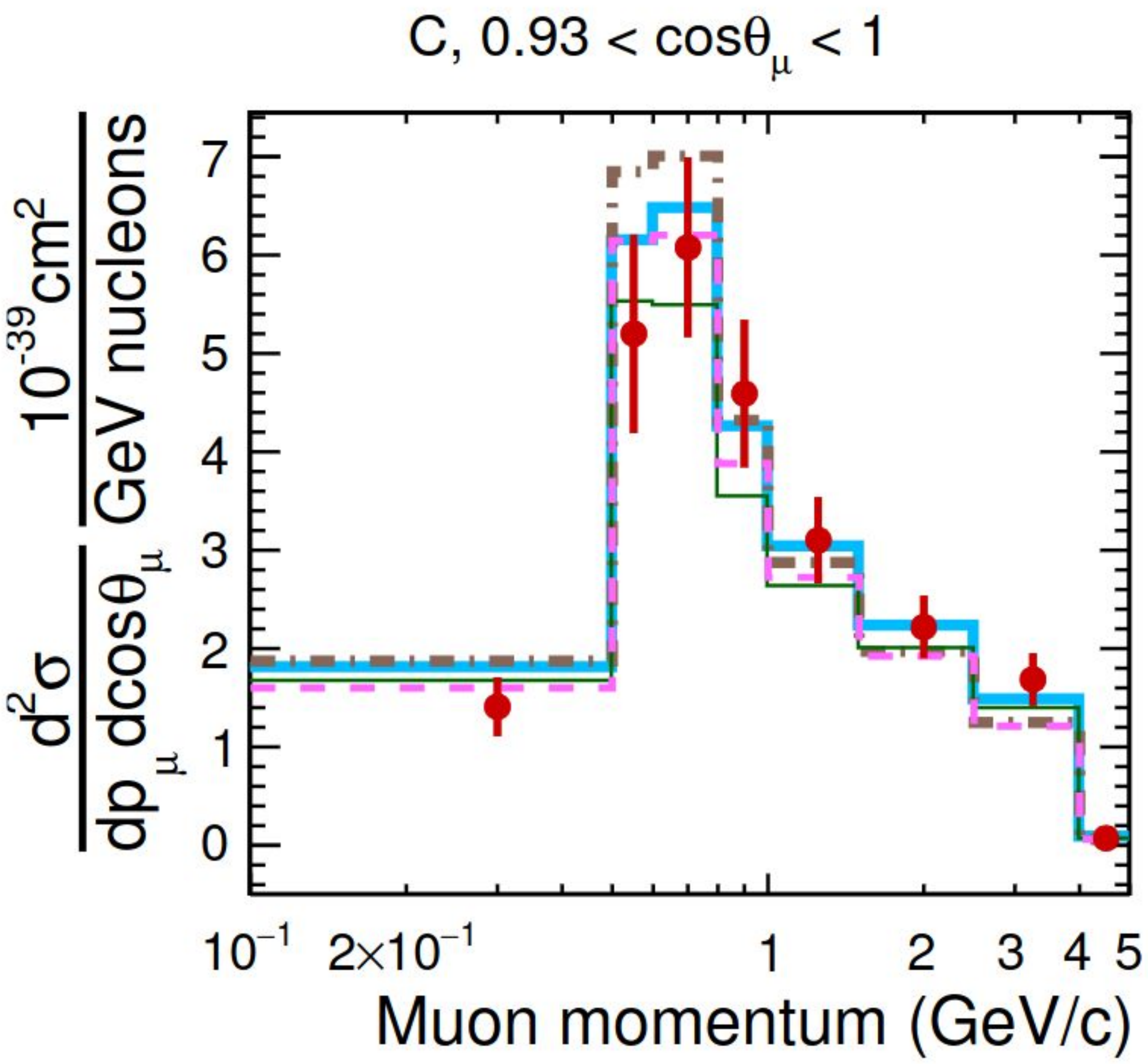


T. Adye, Proc. PHYSTAT 2011, arXiv:1105.1160

T2K style

Distinctly different approach

[Phys. Rev. D 101, 112004 (2020)]



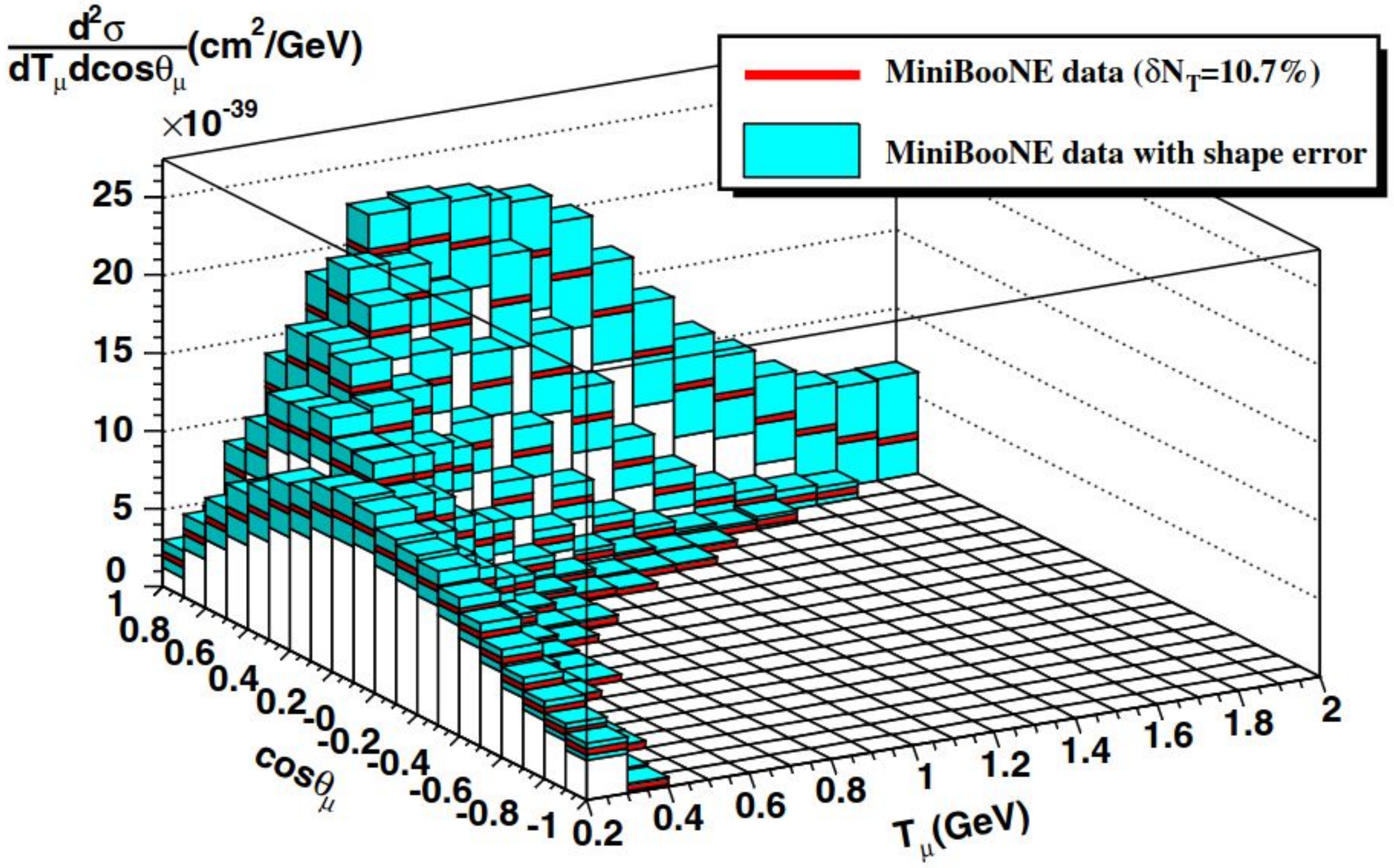
$$-2\log(\mathcal{L}) = -2\log(\mathcal{L}_{\text{stat}}) - 2\log(\mathcal{L}_{\text{syst}}) - 2\log(\mathcal{L}_{\text{reg}})$$

- Simulation compared to reconstructed data in a **binned likelihood fit**
  - Stat: Poisson likelihood, corrected for finite MC event counts
  - Syst: Prior uncertainties on model parameters
  - Reg: optional regularization term
- Includes signal scaling factors for each measurement bin, float without constraint
- Error propagation via
  - Repeated fits in each universe
  - Throws from post-fit parameter covariances

"Blockwise unfolding": motivation

Phys. Rev. D 81, 092005 (2010)

- **MiniBooNE**: pioneering neutrino experiment at Fermilab
  - Many cross-section analysis practices established
  - Key early measurements
- Several data releases report binwise uncertainties but ***not correlations***
  - Large & important
  - Both systematic (e.g., flux) and statistical (unfolding)



2D result for CH target

Problematic for quantitative comparisons (χ^2, etc.)

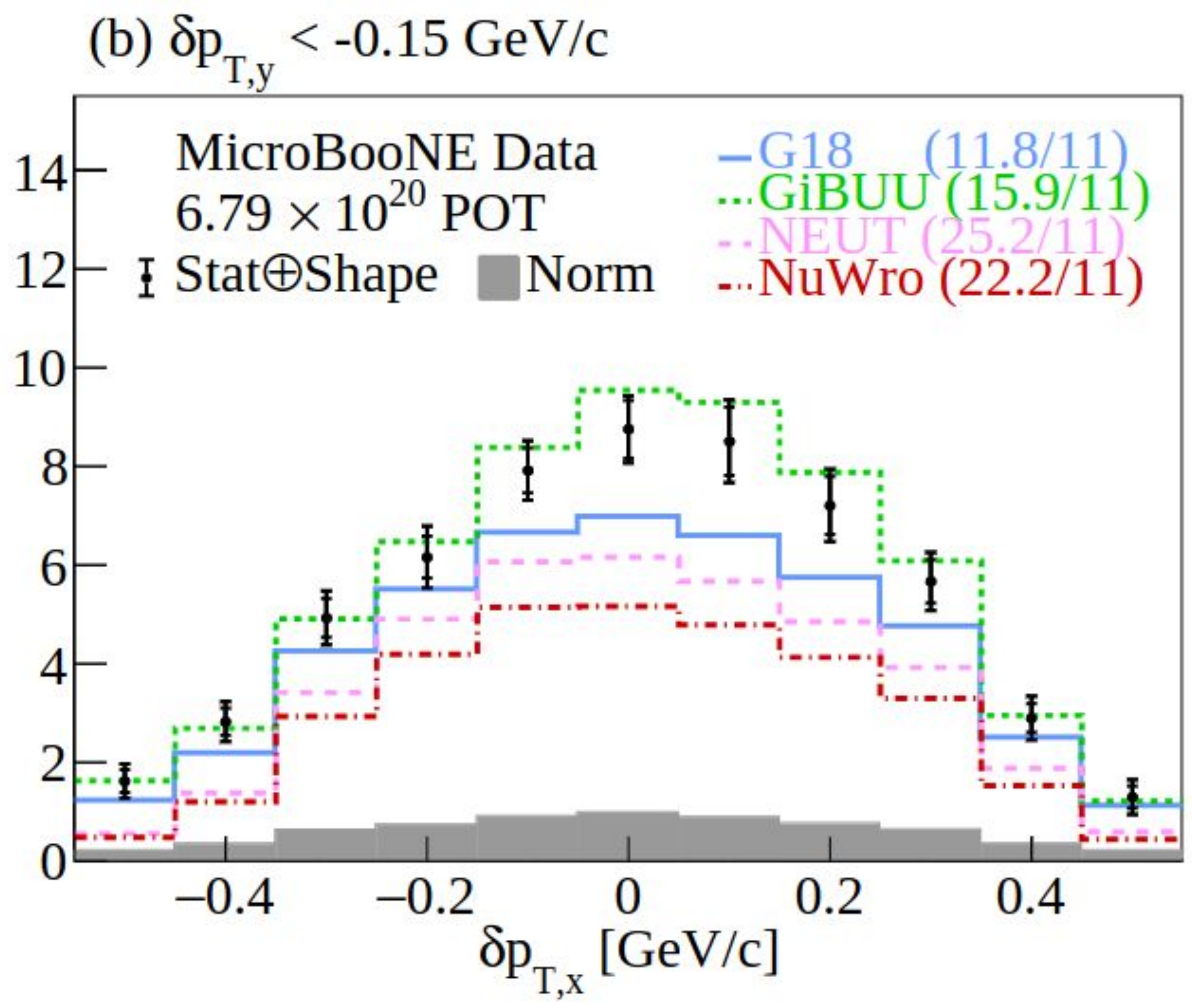
Standard practice is now to provide a full covariance matrix

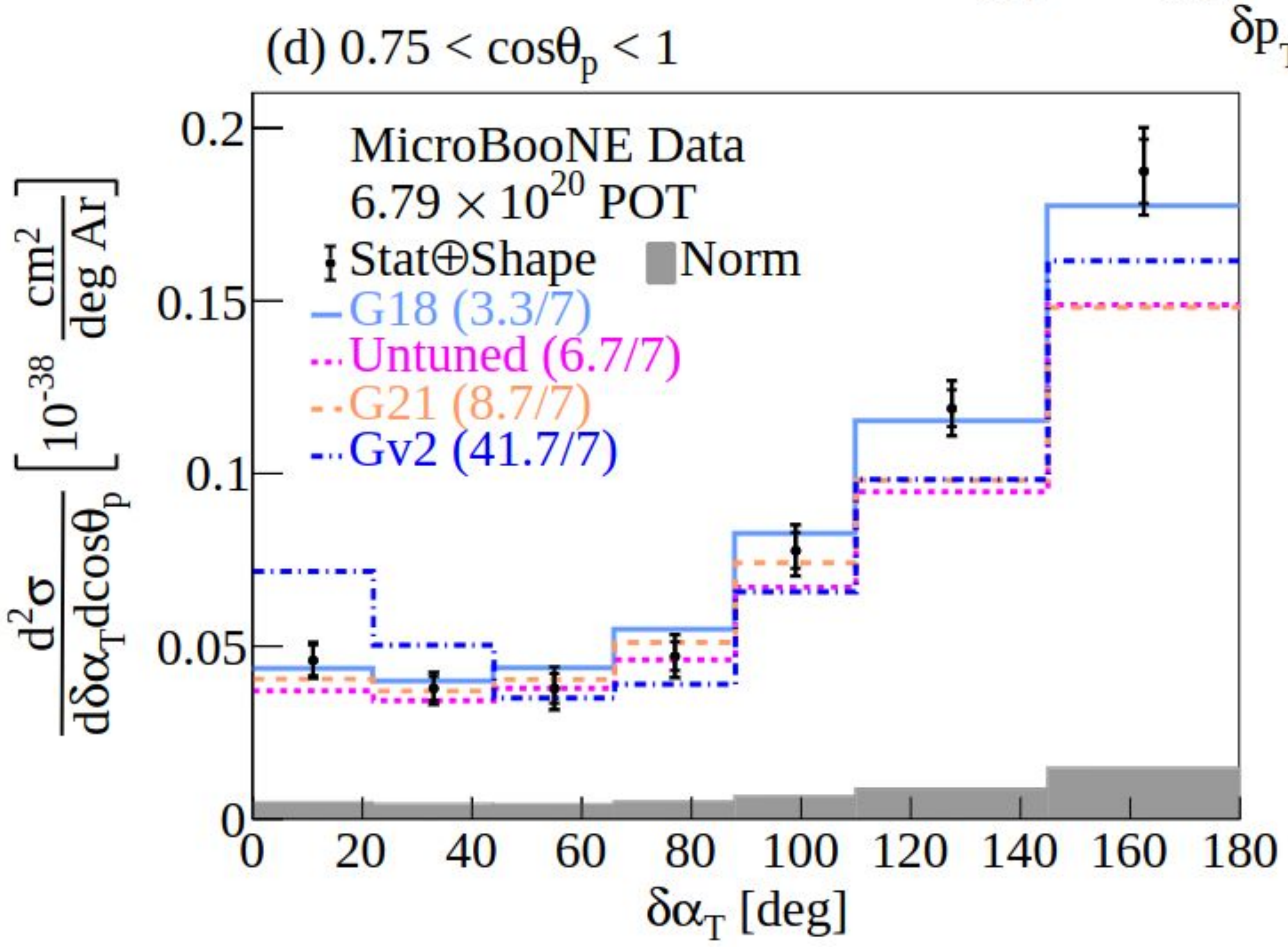
"Blockwise unfolding": motivation

Phys. Rev. D 108, 053002 (2023)

- Experiments often report **multiple kinematic distributions**
  - Same analysis or complementary ones
- **Correlated uncertainties** between distributions are still not typically reported
  - All the same drawbacks as before

ν_μ CC 1p0π data from MicroBooNE



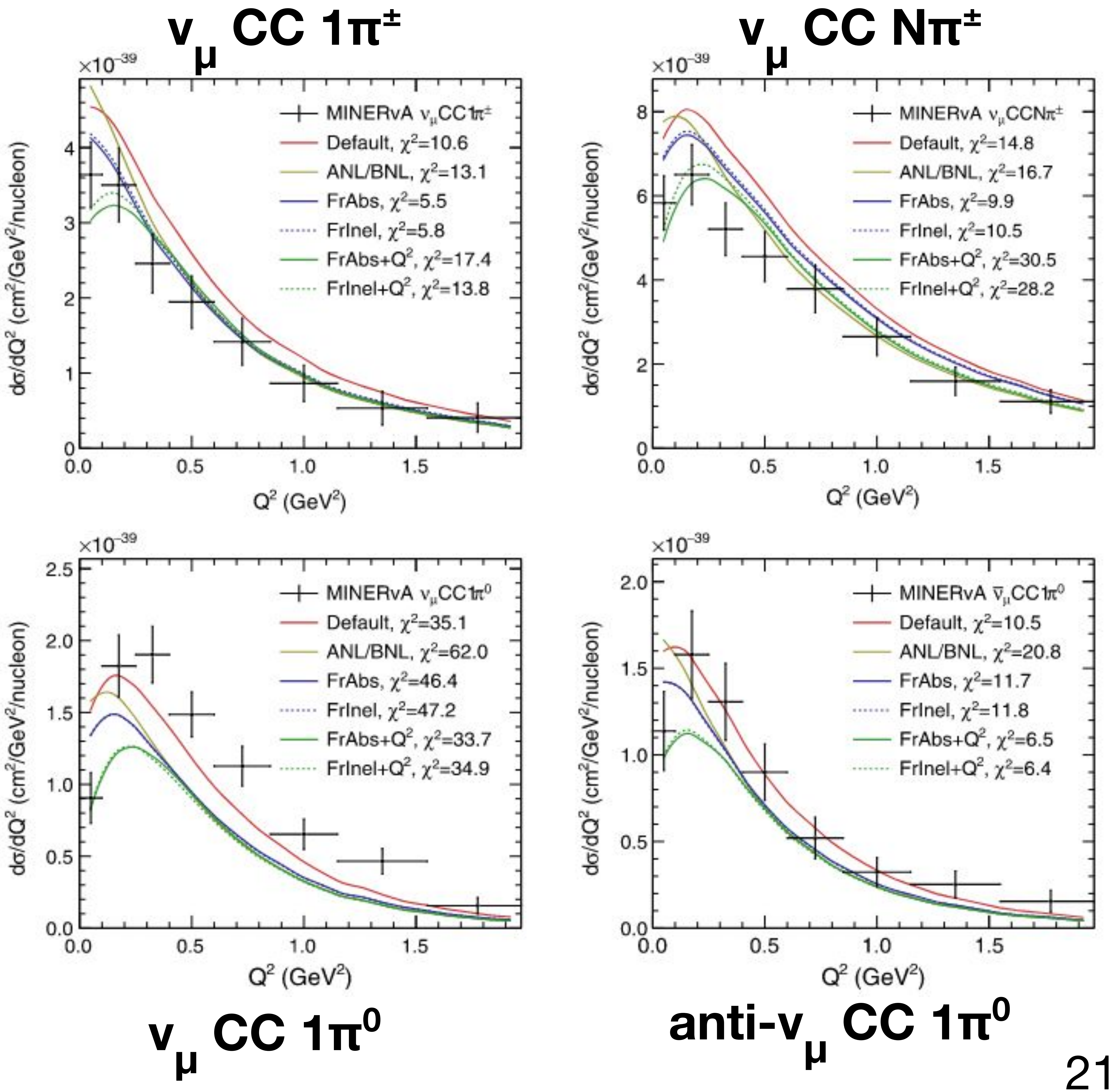


Measurements use same set of ~9000 events

"Blockwise unfolding": motivation

Phys. Rev D. 100, 072005 (2019)

- Experiments often report **multiple kinematic distributions**
  - Same analysis or complementary ones
- **Correlated uncertainties** between distributions are still not typically reported
  - All the same drawbacks as before
- **Limitations** discussed in MINERvA paper tuning GENIE to π production data



"Blockwise unfolding": motivation

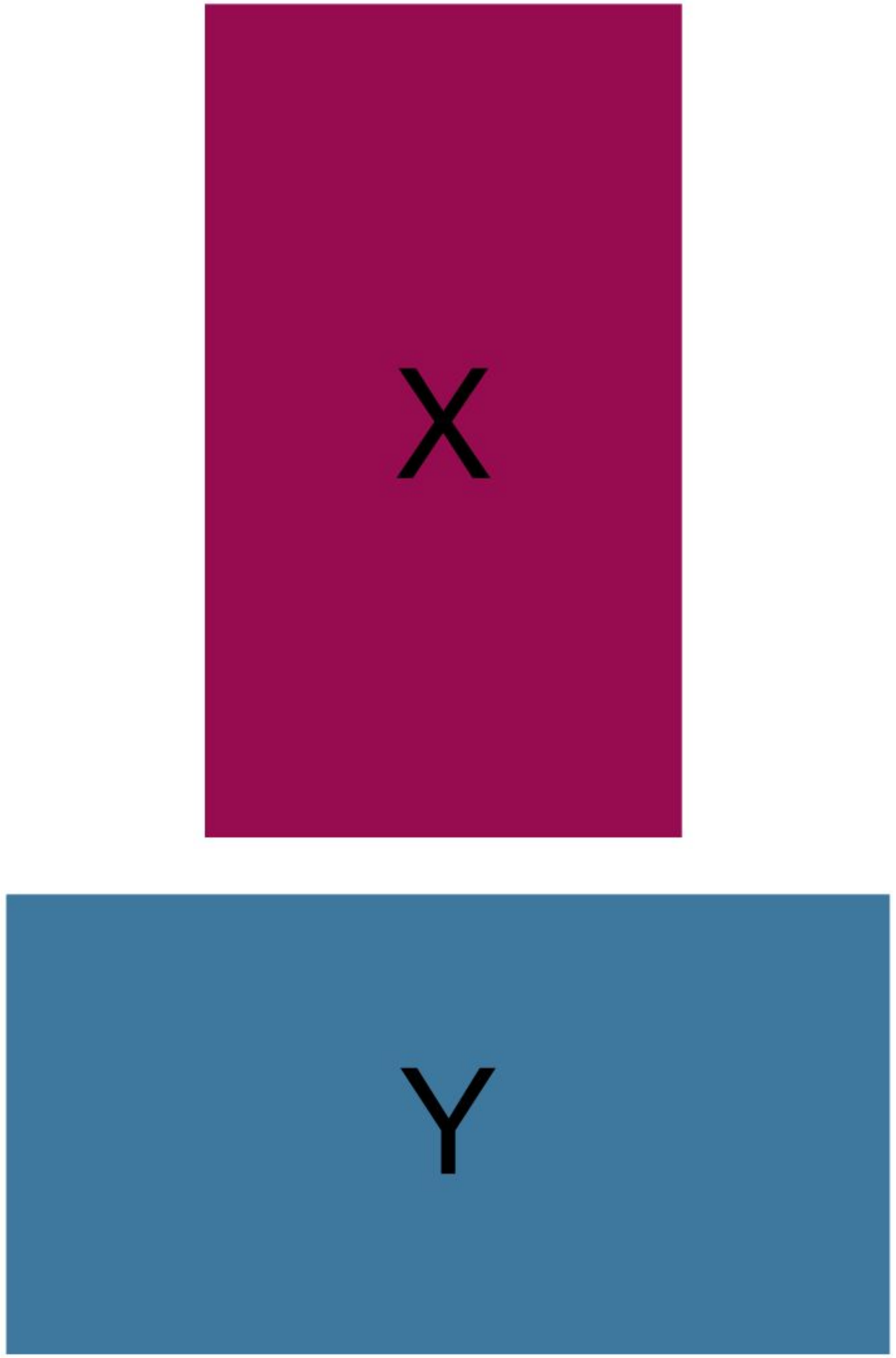
Phys. Rev D. 100, 072005 (2019)

The published cross sections are one dimensional with correlations provided between the bins within each distribution. No correlations are provided between measurements of different final states, or between different one-dimensional projections of the same measurement. These correlations are expected to be large, coming predominantly from flux and detector uncertainties. Additionally, the ν_μCC1$\pi^\pm$ event sample is a subset (~64%) of the ν_μCC$N\pi^\pm$ event sample, and including both channels introduces a statistical correlation. Not assessing correlations between the distributions, while a common practice in this field, is a limitation when tuning models to multiple datasets. It introduces a bias in the χ^2 statistic that is difficult to quantify, and requires imposing *ad hoc* uncertainties [4] as the test statistic is not expected to follow a χ^2 distribution for the given degrees of freedom.

- Not trivial to add this information after the fact
- Correlations calculable with **suitable planning ahead**
  - Maximize impact from cross-section analyses
- **Two issues**
  - Event overlaps (statistical covariances)
  - Unfolding treatment
- **Methods paper** (arXiv:2401.04065) gives recipes for solving these problems

Statistical covariances

- Events belong to multiple bins
  ⇒ correlated stat uncertainties
- **Easily calculable** if the problem is framed properly
- Arbitrary bins X and Y
  - Event count n_X in bin X follows a Poisson distribution
- Estimator for the mean: n_X
- Estimator for the variance: n_X
- Bin Y is similar. How to get the covariance?

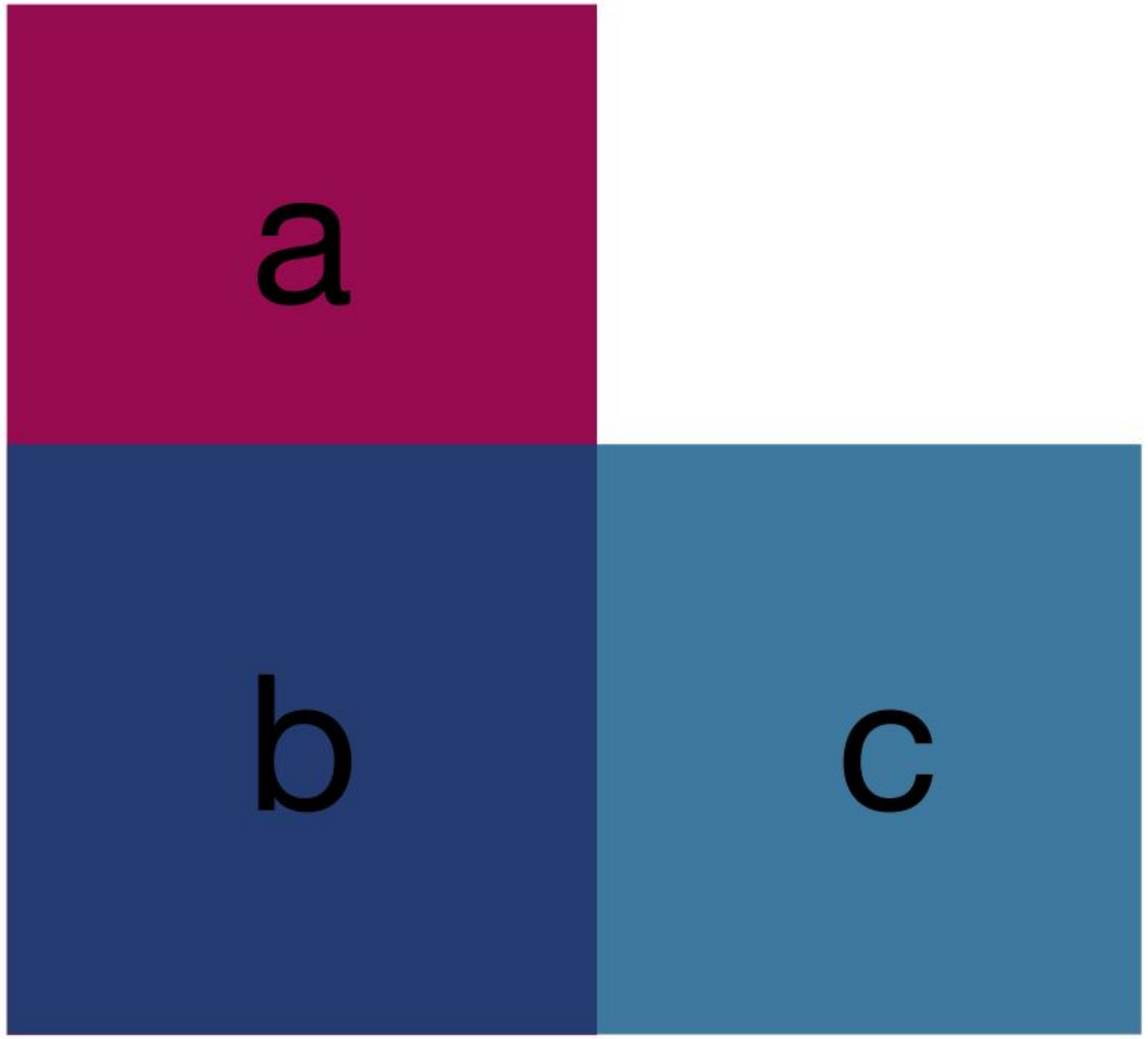


Statistical covariances

- The trick: one may always rebin 2 ⟶ 3
- Bins a, b, and c are **non-overlapping**
- Independent Poisson distributions

$$\begin{aligned}\mathrm{cov}(X, Y) &= \mathrm{cov}(a + b, b + c) \\ &= \mathrm{cov}(a, b) + \mathrm{cov}(a, c) + \mathrm{cov}(b, b) + \mathrm{cov}(b, c) \\ &= 0 + 0 + \mathrm{var}(b) + 0 \\ &\approx n_b\end{aligned}$$

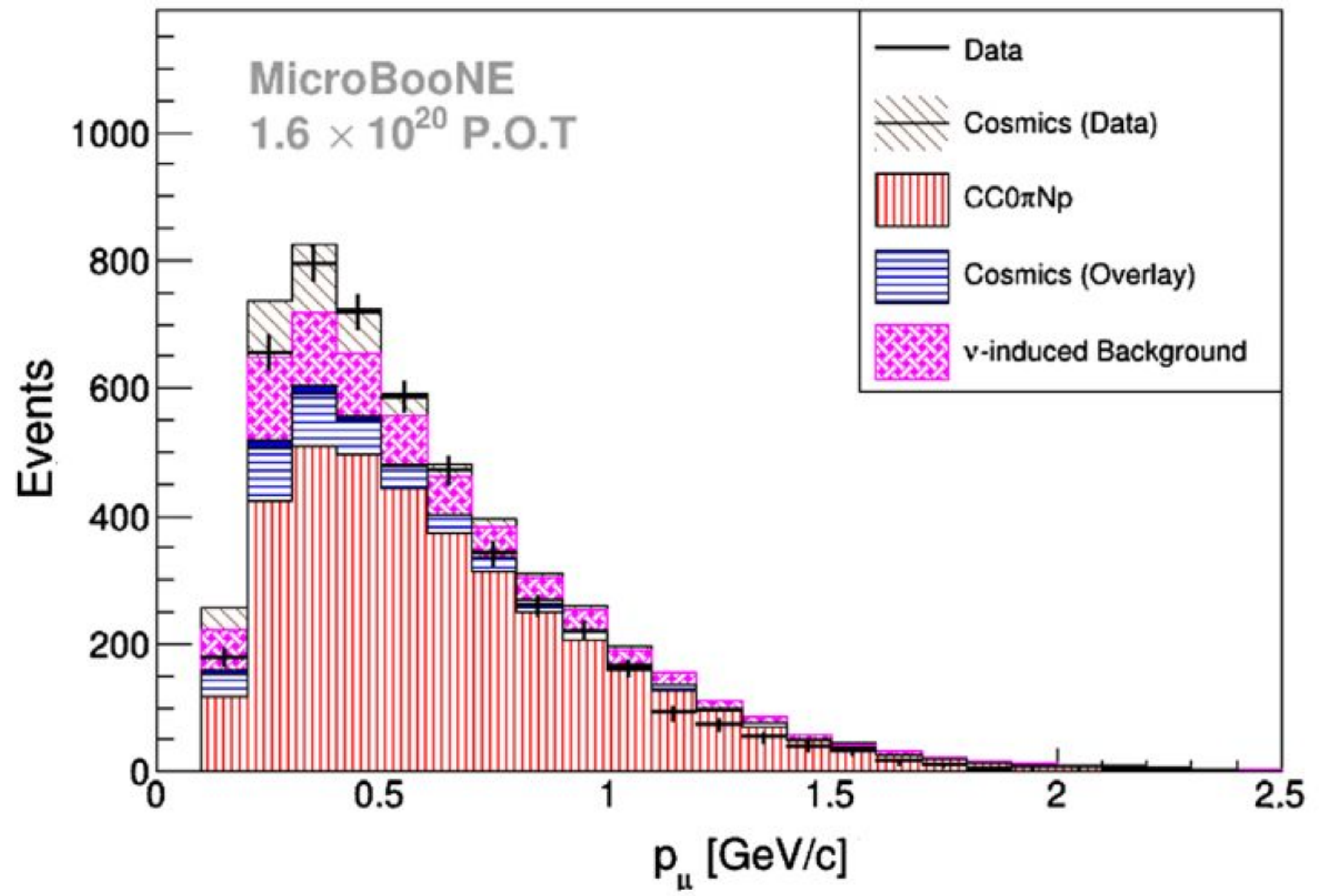
- Estimator for statistical covariance is just the **number of events that bins X and Y have in common**
- MINERvA/T2K recipe is conceptually similar, described in paper



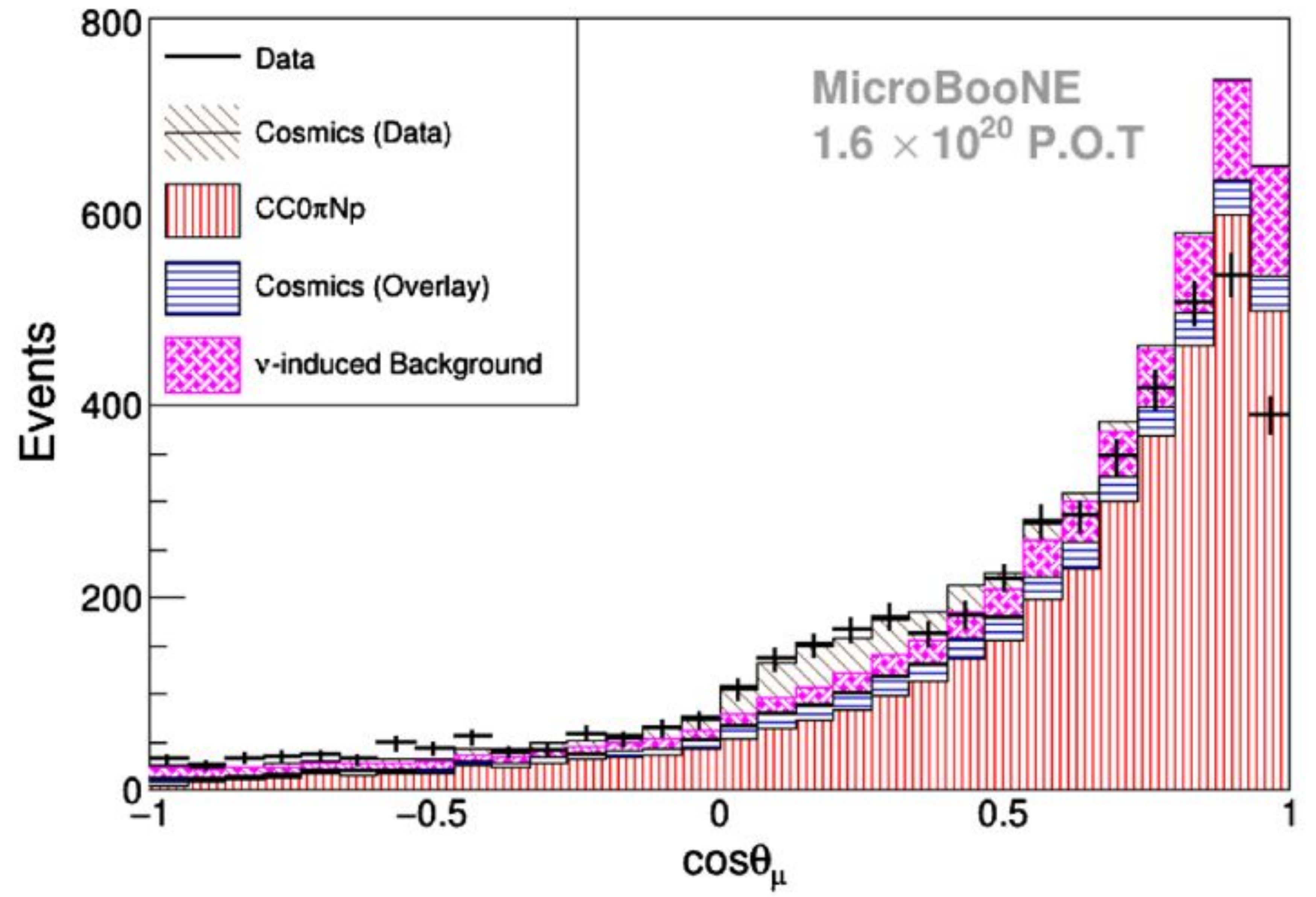
Note that this behaves as expected for X = Y as well as disjoint bins

Unfolding with correlated uncertainties

- Group bins belonging to the same kinematic distribution in a "**block**"
- An event should belong to a maximum of one reco bin and one true bin in each block → avoids double-counting
- Observables can be abstracted away by working in "bin number space"
  - Trivially generalizes to 2D, 3D, etc.
- Example:
  - Bins 0-19 represent pμ → block #0
  - Bins 20-49 represent cosθμ → block #1



Phys. Rev. D 102, 112013 (2020)



A "blockwise" unfolding matrix

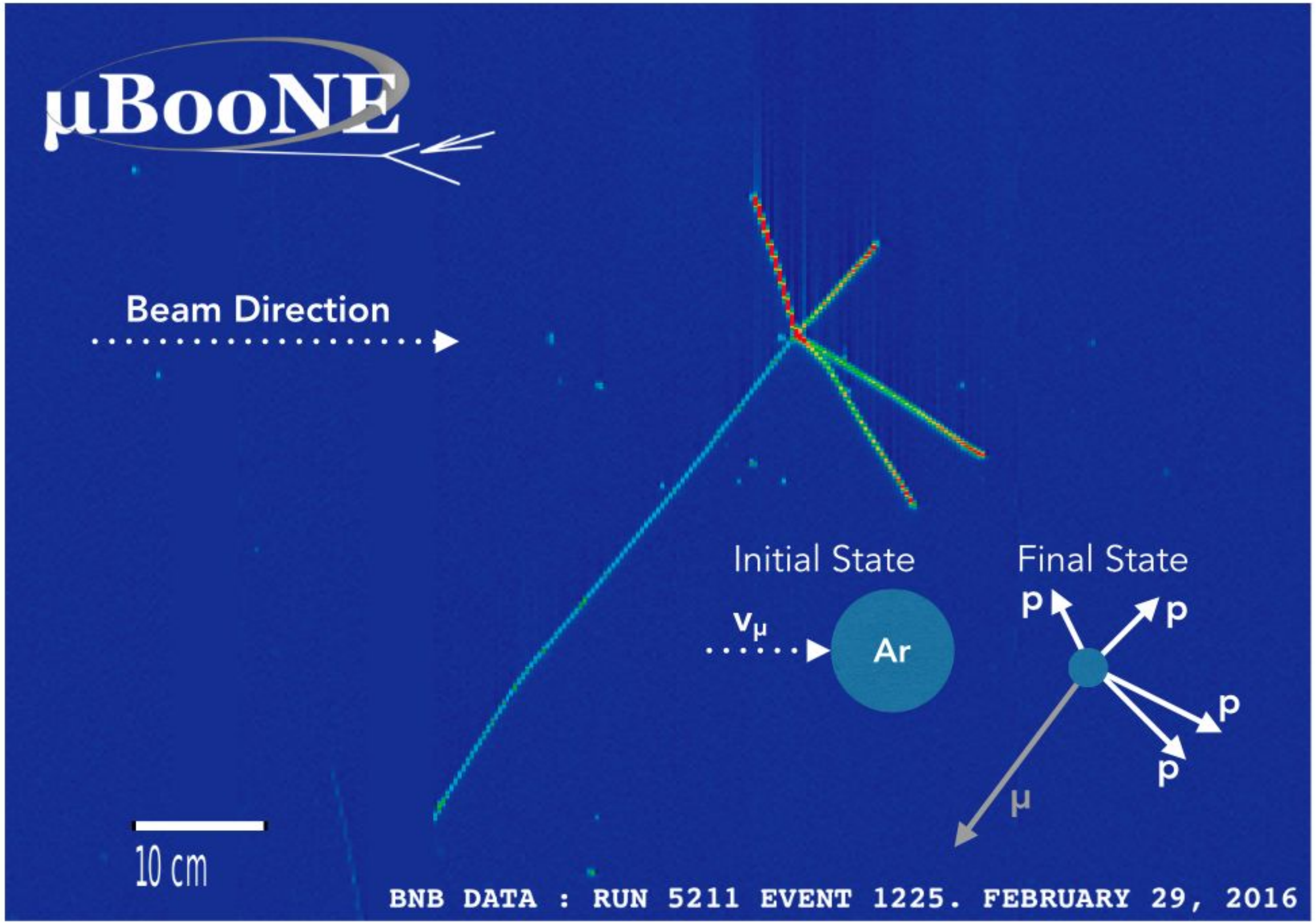
- Build an unfolding matrix U_b for the b-th block according to one's preferred approach
- Overall unfolding matrix U is block-diagonal
- Results for individual blocks are the **same** as for stand-alone measurements of each
- This organization allows reporting of correlated uncertainties between all bins in all blocks
  - Details depend on extraction style, but fully documented in paper

$$U = \bigoplus_{b=0} U_b = U_0 \oplus U_1 \oplus \cdots = \begin{pmatrix} U_0 & 0 & 0 & \dots \\ 0 & U_1 & 0 & \dots \\ 0 & 0 & \ddots & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Signal for MicroBooNE ν_μ CC0πNp

- ν_μ CC on Ar, at least one final-state proton
- Zero (anti)mesons
- $p_\mu \in [0.1, 1.2]$ GeV/c
- $p_p \in [$**0.25**$, 1.0]$ GeV/c
- Restricted phase space motivation similar to 1p analysis
- The p_p limit only applies to the **leading proton**

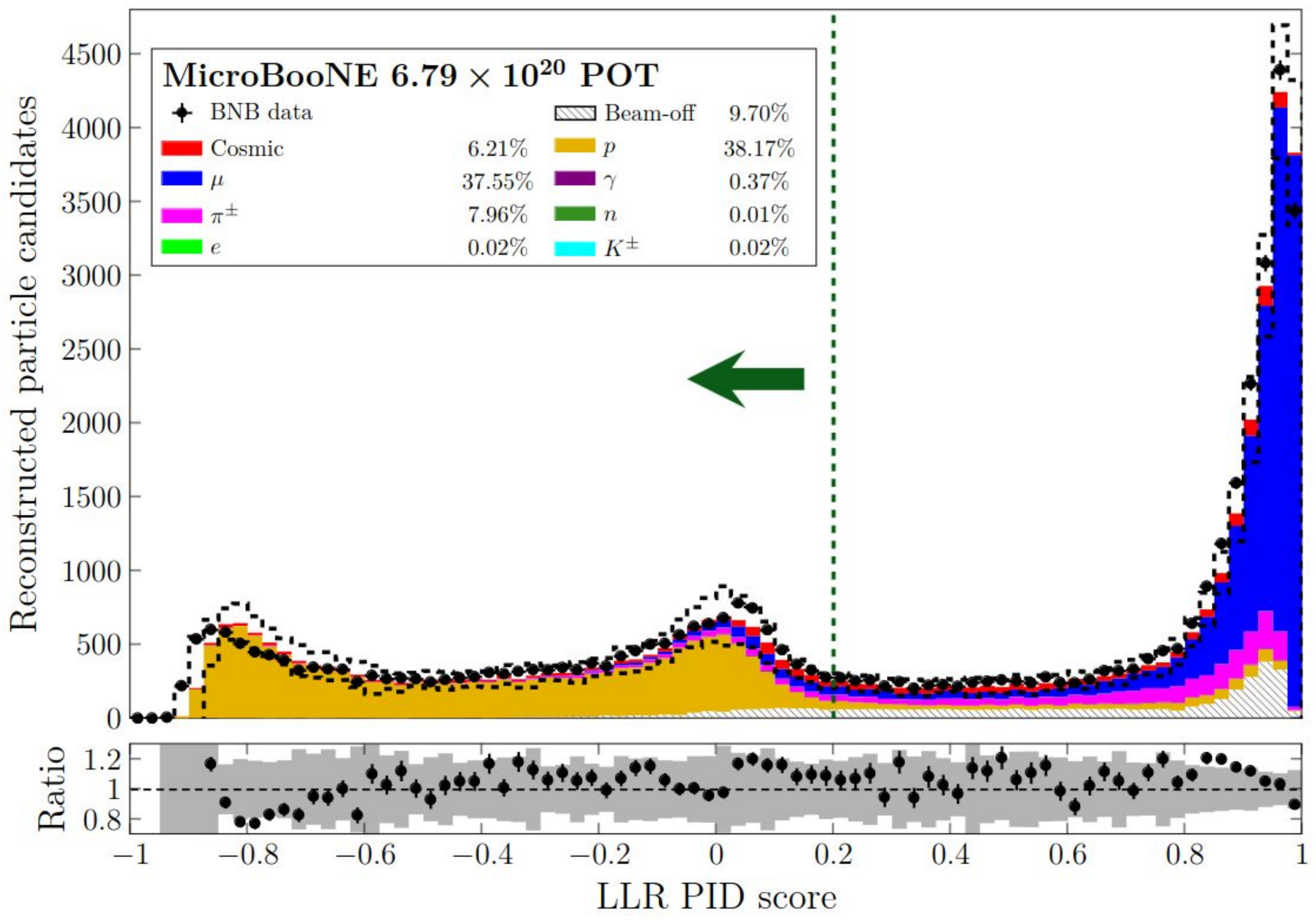
This 4p (!) candidate event is selected



Event selection

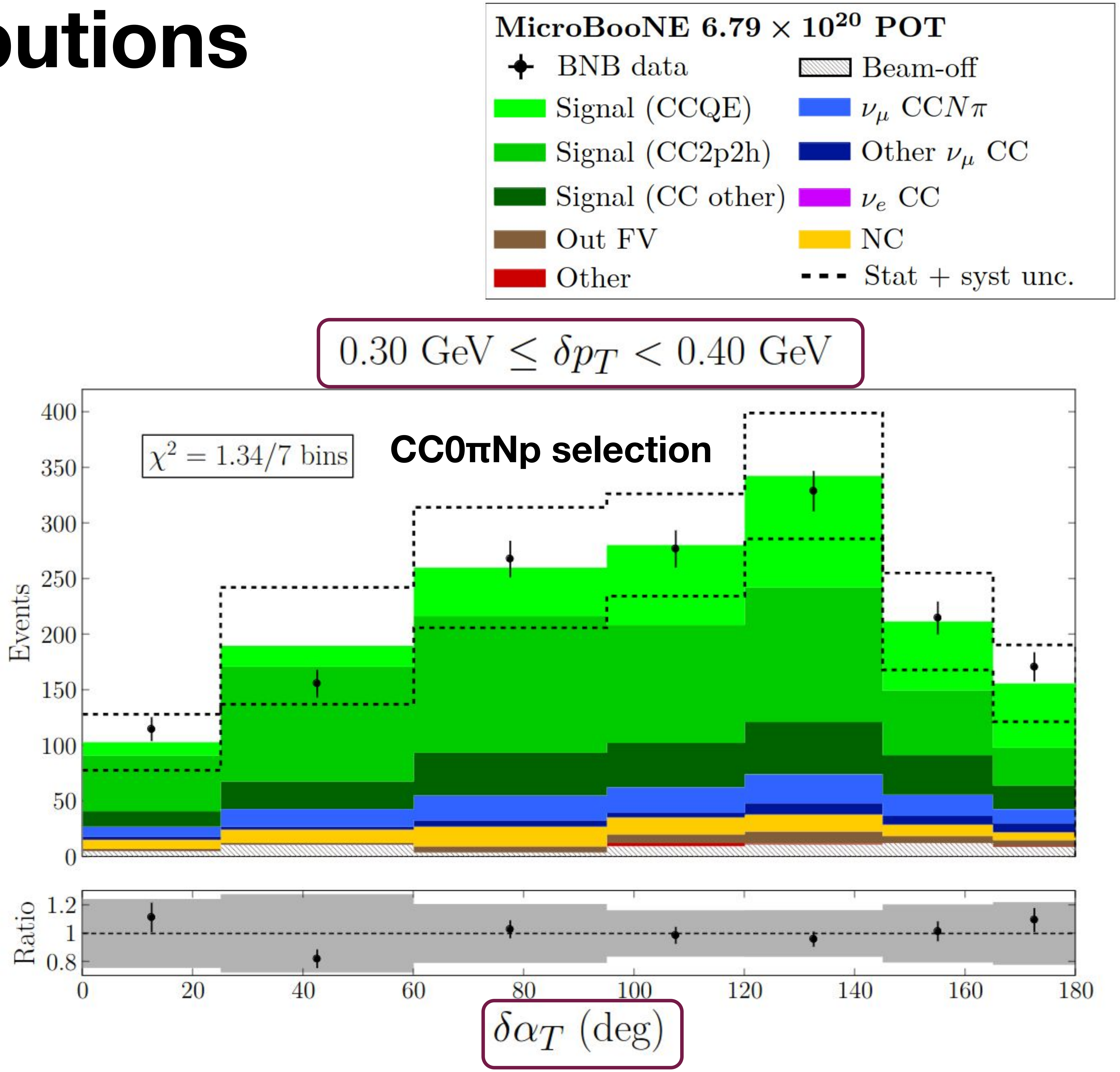
- Implemented using automated Pandora reconstruction: [Eur. Phys. J. C 78, 82 (2018)]
- Series of 12 cuts:
  - Find a ν-induced μ
  - that is well-reconstructed
  - and accompanied only by p
- Overall performance
  - 12.3% efficiency
  - 78.5% purity

Proton identification



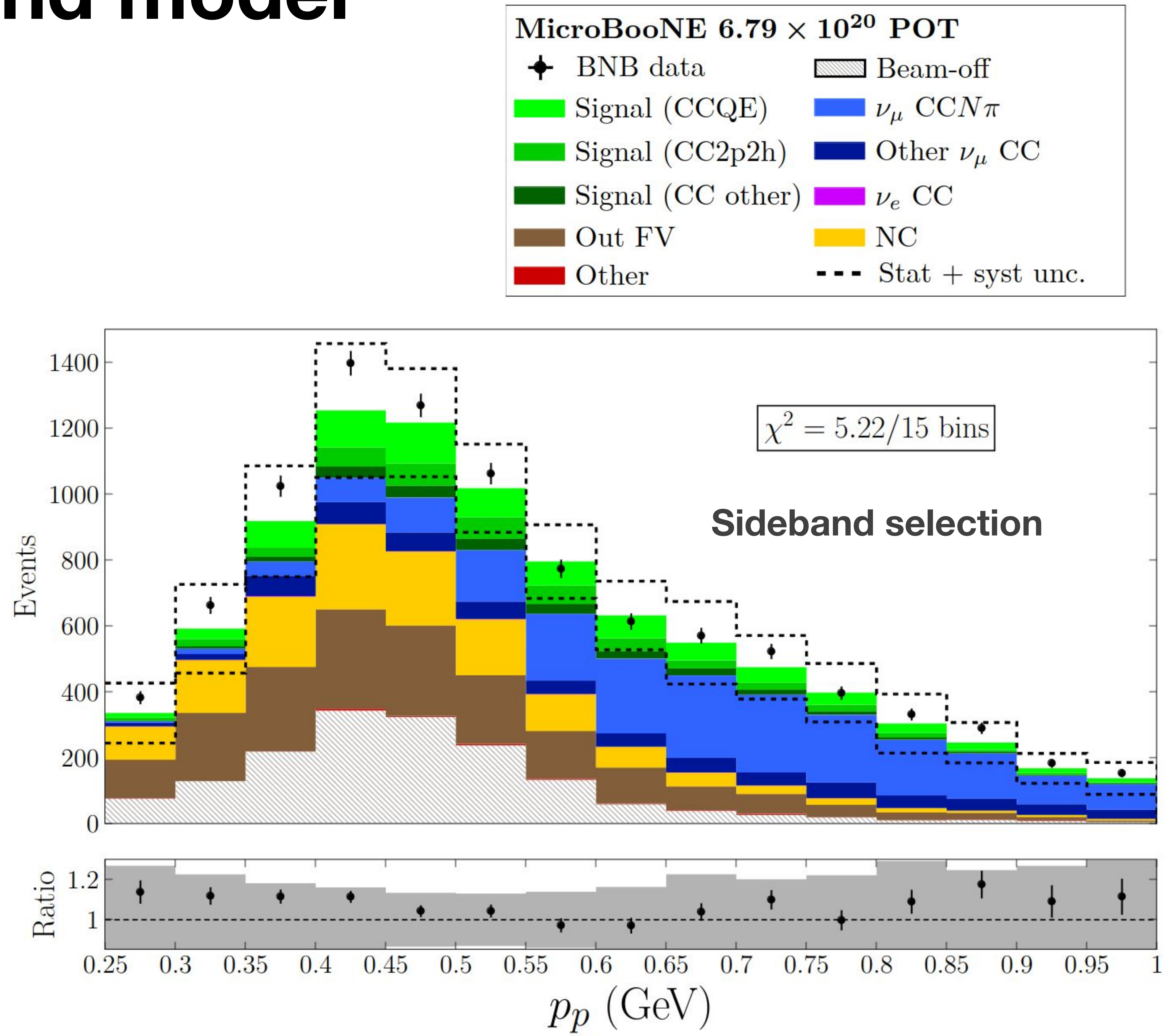
Reconstructed event distributions

- Agreement reasonable within uncertainties (χ^2 = 355 / 359 bins)
- 3 dominant backgrounds:
  - **Out of Fiducial Volume** (Out FV)
  - **Neutral-current** (NC)
  - **Pion production** (ν_μ CCNπ)
- Alternate selections made to enhance each, check background prediction



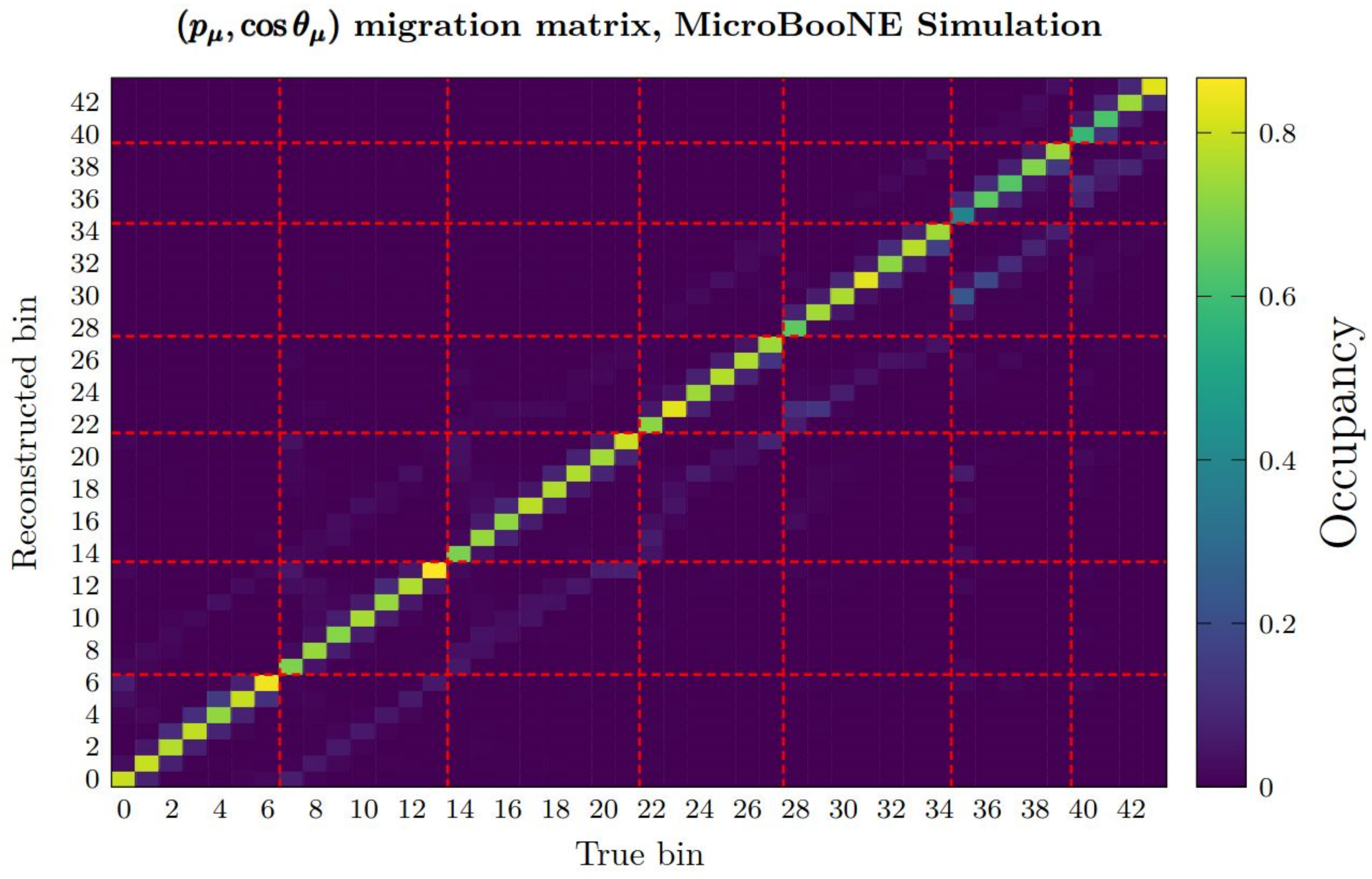
Sideband test of background model

- Logical OR of 3 alternate selections plotted for 359 bins
- **Out FV** and **NC** important at low p_p, **π production** at high p_p
- Satisfactory agreement everywhere in phase space (χ^2 = 178 / 359 bins)
- GENIE-based model used unaltered
- Full sideband results in supplement and data release
  - Includes all systematic universes

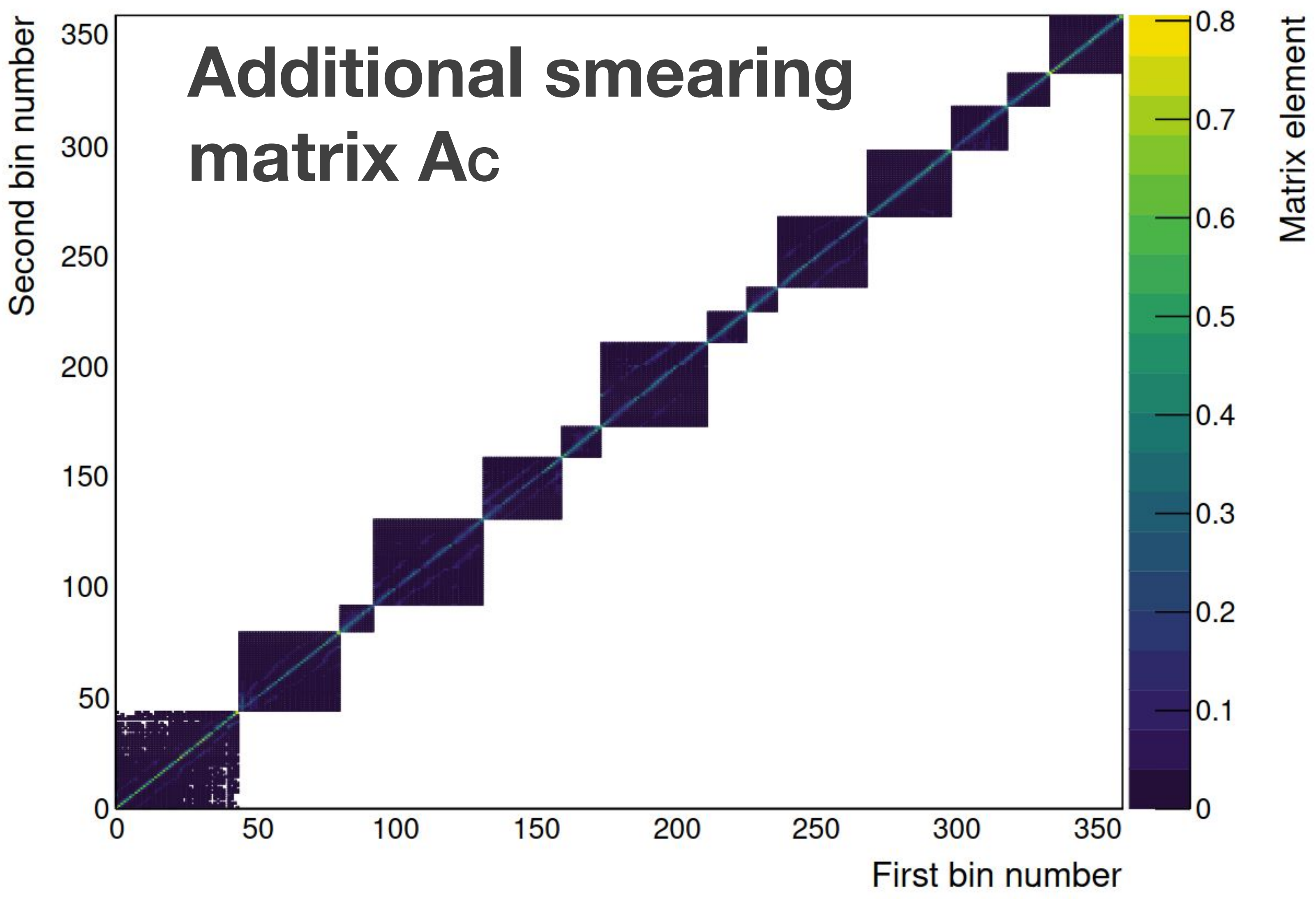


Unfolding

- D'Agostini method used for each of 14 blocks of bins
  - 2-5 iterations depending on specific distribution
  - Validated with mock data
- Additional smearing matrix
  - Supplied in data release for new model comparisons
  - Computed via the formalism described earlier in the talk



Red dashed lines denote pμ bin boundaries

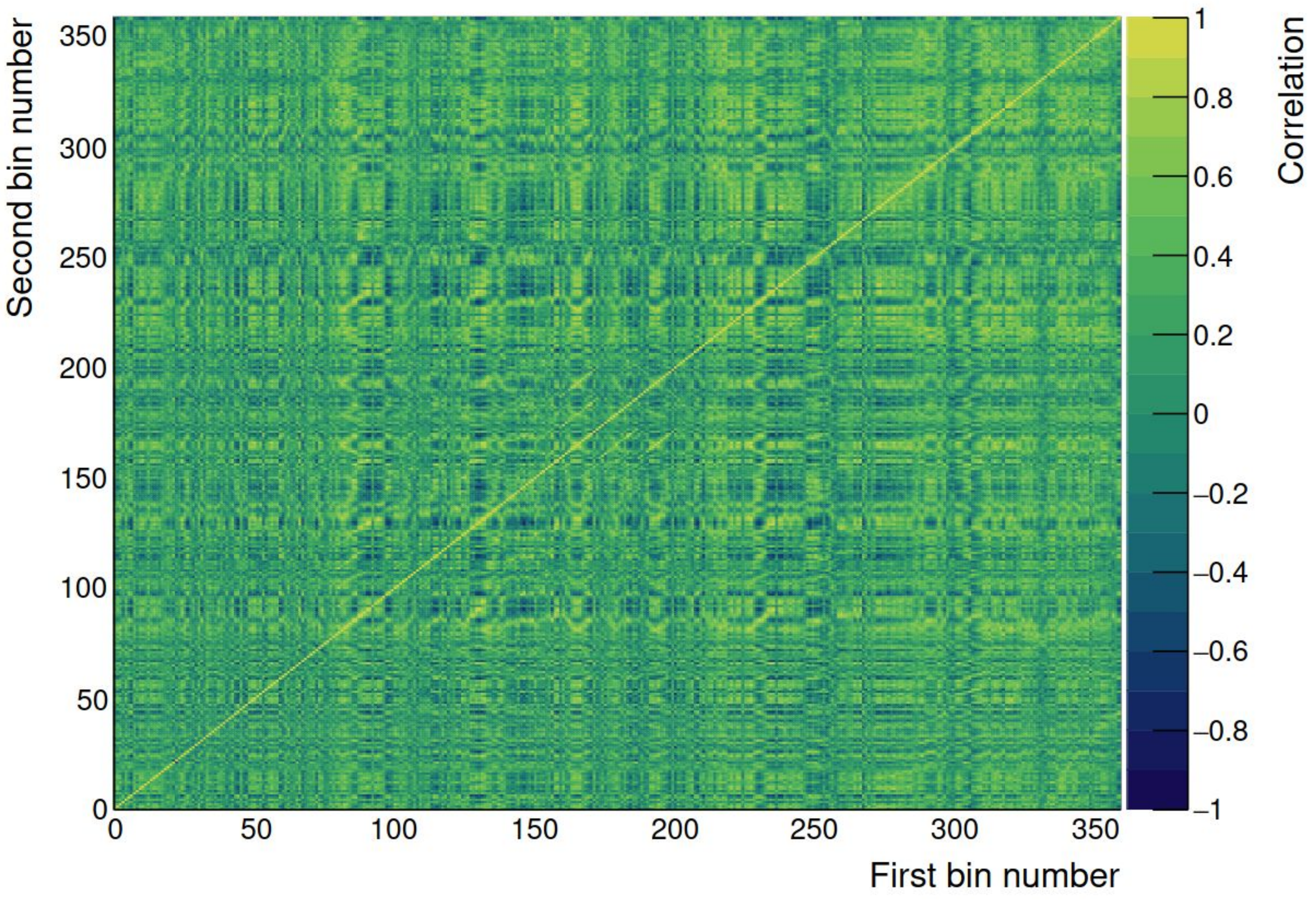


White regions correspond to matrix elements that are identically zero

Inter-distribution correlations

- Enables χ^2 comparisons to entire data set
- **Annoying detail:** differential cross sections vary in their units
    - Can lead to confusion when reporting covariances
- **Recommendation** from [arXiv:2401.04065](https://arxiv.org/abs/2401.04065) implemented
    - Re-express as *total* cross sections per bin

Total correlation matrix for measured CC0πNp cross sections



$$\langle \sigma \rangle_\mu = \left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu \cdot \Delta\mathbf{x}_\mu \qquad \langle \sigma \rangle_\mu = \frac{\hat{\phi}_\mu}{\Phi_\mu \, T_\mu} = \frac{\sum_a U_{\mu a}\,(D_a - B_a)}{\Phi_\mu \, T_\mu}$$

$$\text{Cov}\Big(\langle\sigma\rangle_\mu, \langle\sigma\rangle_\lambda\Big) = \frac{\sum_{a,b} \mathfrak{E}_{\mu a}\,\text{Cov}(D_a, D_b)\,\mathfrak{E}_{\lambda b}}{\Phi_\mu\,\Phi_\lambda\,T_\mu\,T_\lambda}$$

Inter-distribution correlations

- Enables χ^2 comparisons to entire data set
- **Annoying detail:** differential cross sections vary in their units
  - Can lead to confusion when reporting covariances
- **Recommendation** from [arXiv:2401.04065] implemented
  - Re-express as *total* cross sections per bin

Summary table of final results from data release

TABLE I: Measured flux-averaged CC0πNp total cross sections

bin number	total cross section (10^{-38} cm^2/Ar)	stat. unc. (10^{-38} cm^2/Ar)	total unc. (10^{-38} cm^2/Ar)
0	0.491	0.041	0.110
1	0.305	0.034	0.077
2	0.249	0.033	0.079
3	0.225	0.030	0.059
4	0.147	0.024	0.055
5	0.191	0.024	0.061
6	0.071	0.020	0.073
7	0.281	0.025	0.048
8	0.260	0.024	0.046
9	0.233	0.024	0.060
10	0.258	0.026	0.062
11	0.186	0.021	0.036
12	0.183	0.018	0.042
13	0.091	0.014	0.029
14	0.264	0.028	0.079

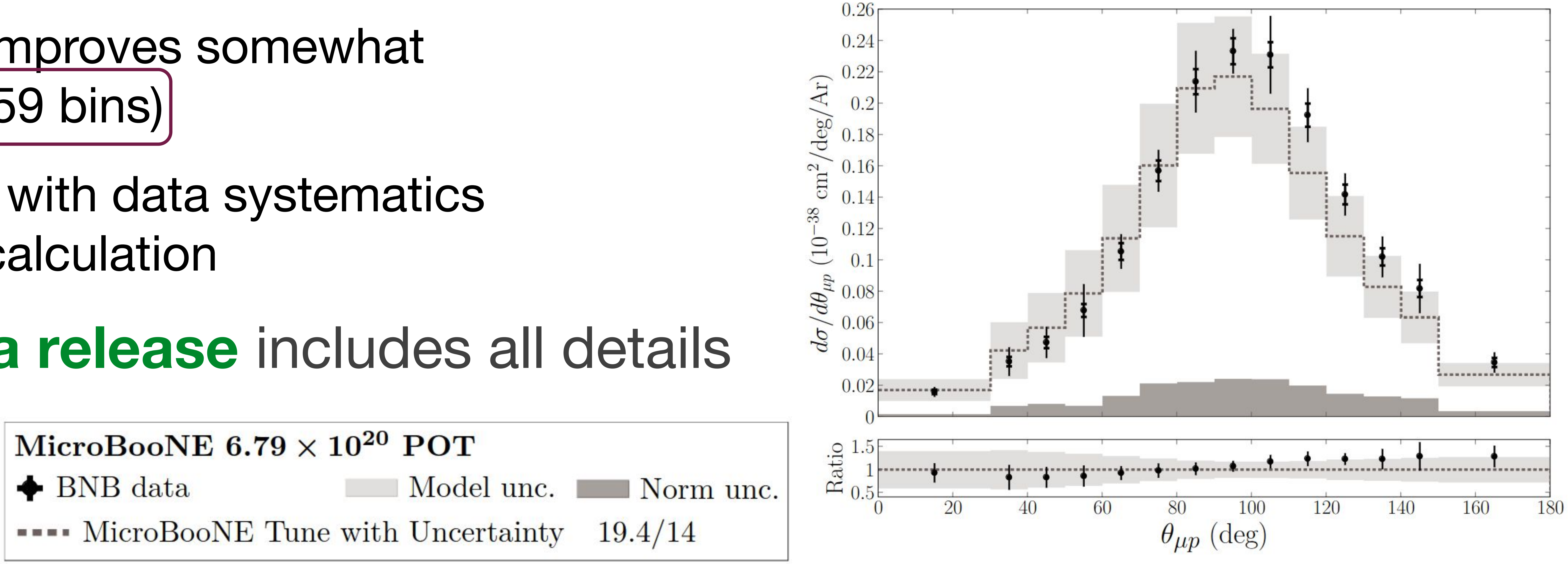
$$\mathrm{Cov}\Big(\langle\sigma\rangle_\mu, \langle\sigma\rangle_\lambda\Big) = \frac{\sum_{a,b} \mathfrak{E}_{\mu a}\, \mathrm{Cov}(D_a, D_b)\, \mathfrak{E}_{\lambda b}}{\Phi_\mu\, \Phi_\lambda\, T_\mu\, T_\lambda}$$

$$\langle\sigma\rangle_\mu = \left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu \cdot \Delta\mathbf{x}_\mu \qquad \langle\sigma\rangle_\mu = \frac{\hat{\phi}_\mu}{\Phi_\mu\, T_\mu} = \frac{\sum_a U_{\mu a}\,(D_a - B_a)}{\Phi_\mu\, T_\mu}$$

Comparisons enabled by this treatment

- Universal **room for improvement** in comparisons to full data set
- MicroBooNE Tune model uncertainty shown for comparisons in supplement
  - Agreement improves somewhat (χ^2 = 979 / 359 bins)
  - Correlations with data systematics included in calculation
- **Extended data release** includes all details

Model	χ^2 / 359 bins
GENIE 3.0.6	1859
NEUT 5.6.0	2582
MicroBooNE Tune	2673
GENIE 3.2.0 G21_11b	2947
GiBUU 2021.1	4836
NuWro 19.02.1	5315
GENIE 3.2.0 G18_02a	5724
GENIE 2.12.10	7799



"Showing our work" in the supplement

- Basic data release
  - Cross-section results, MicroBooNE flux
  - Overall and partial covariance matrices
  - A_C and example scripts for model comparisons
- Extended data release
  - All information needed to revisit unfolding, uncertainty propagation
  - Stat covariances and systematic universes
  - Script to re-generate covariances between signal bins, sidebands, and the MicroBooNE Tune prediction

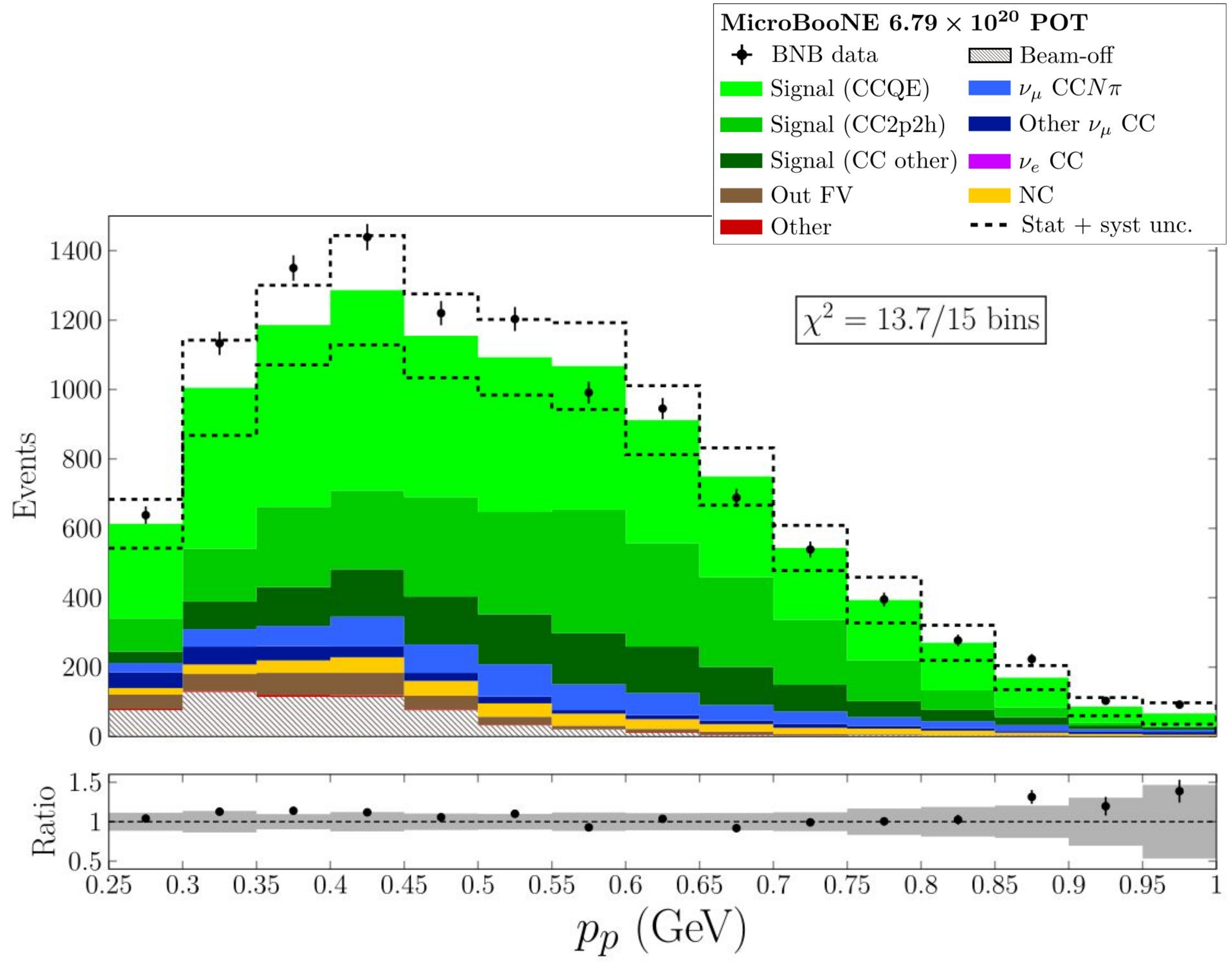
Outlook for the blockwise unfolding technique

- Theorists and generator developers can fit to all measured distributions simultaneously
  - Increases discrimination power of the data: can the model describe the correlations as well as each individual block?
- No need for ad hoc estimates of flux-related covariances, etc.
  - All uncertainties come from the experiment itself
- Potential for **inter-analysis covariances** with two ingredients:
  - Bookkeeping for event overlaps (statistical uncertainties)
  - Consistent systematic variations
- **Latest MicroBooNE analyses** report model goodness-of-fit χ^2 over hundreds of bins in this way
  - See also arXiv:2402.19281, arXiv:2402.19216, arXiv:2404.10948

Background control samples

- Minimizing model dependence is critical
  - We want to learn about Nature, not our simulation!
- Risk of biasing the measurement in both the unfolding (U) and **background subtraction** (B)
  - Sometimes we have to rely on the prediction
  - **Is it good enough to do this?** If not, how do we fix it?

$$\left\langle \frac{d^n\sigma}{d\mathbf{x}} \right\rangle_\mu = \frac{\sum_a U_{\mu a}\left(D_a - B_a\right)}{\Phi\, T\, \Delta \mathbf{x}_\mu}$$

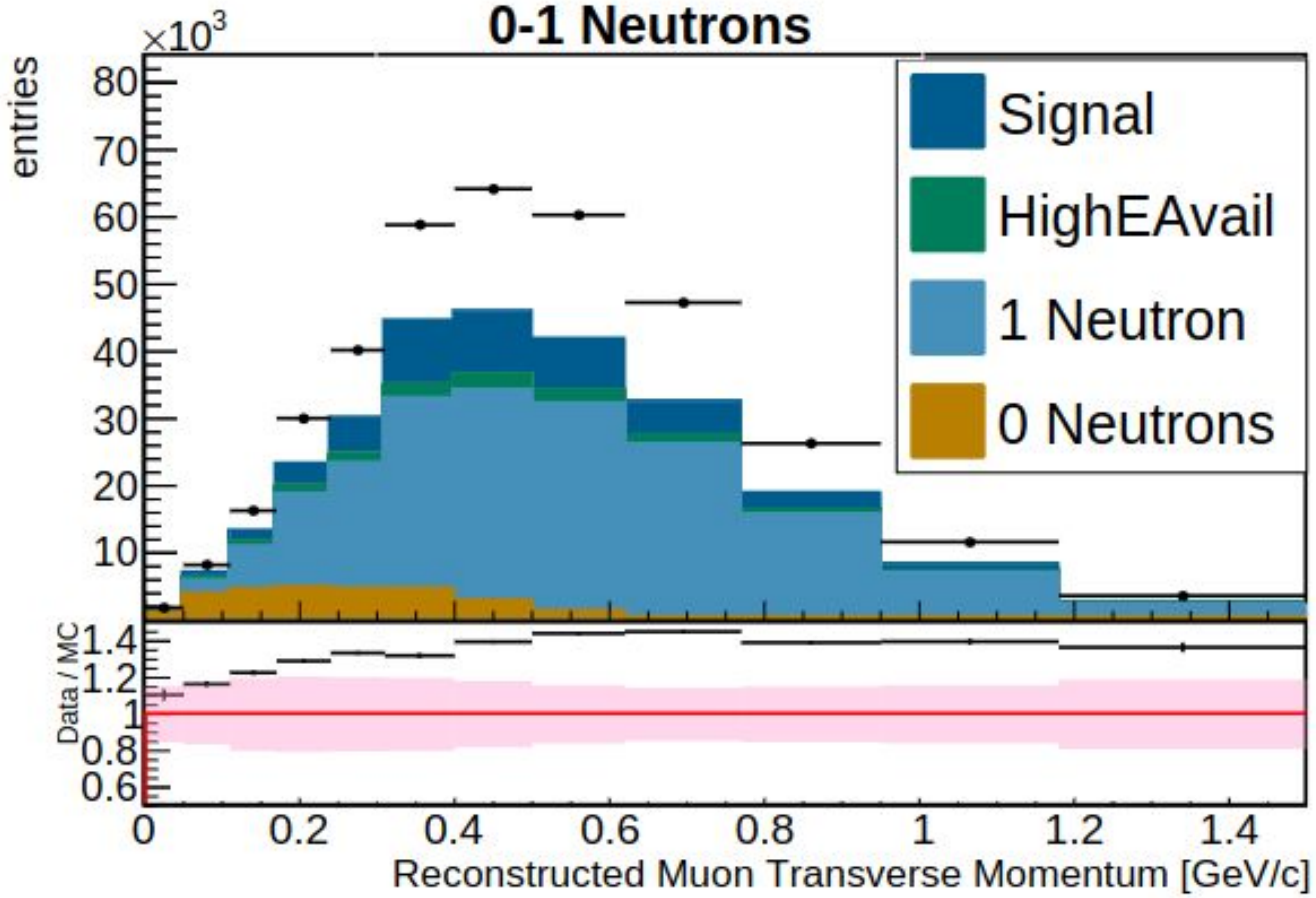


Background control samples

[Phys. Rev. D 108, 112010 (2023)]

- **Control samples**: **check/correct background model** based on parallel measurement
  - Background-enhanced selection
- Also often referred to as "**sidebands**"
  - I use the terms interchangeably in the paper
- I propose a semi-new way of using these for cross-section analyses

anti-ν_μ CC 2+ neutrons (MINERvA)



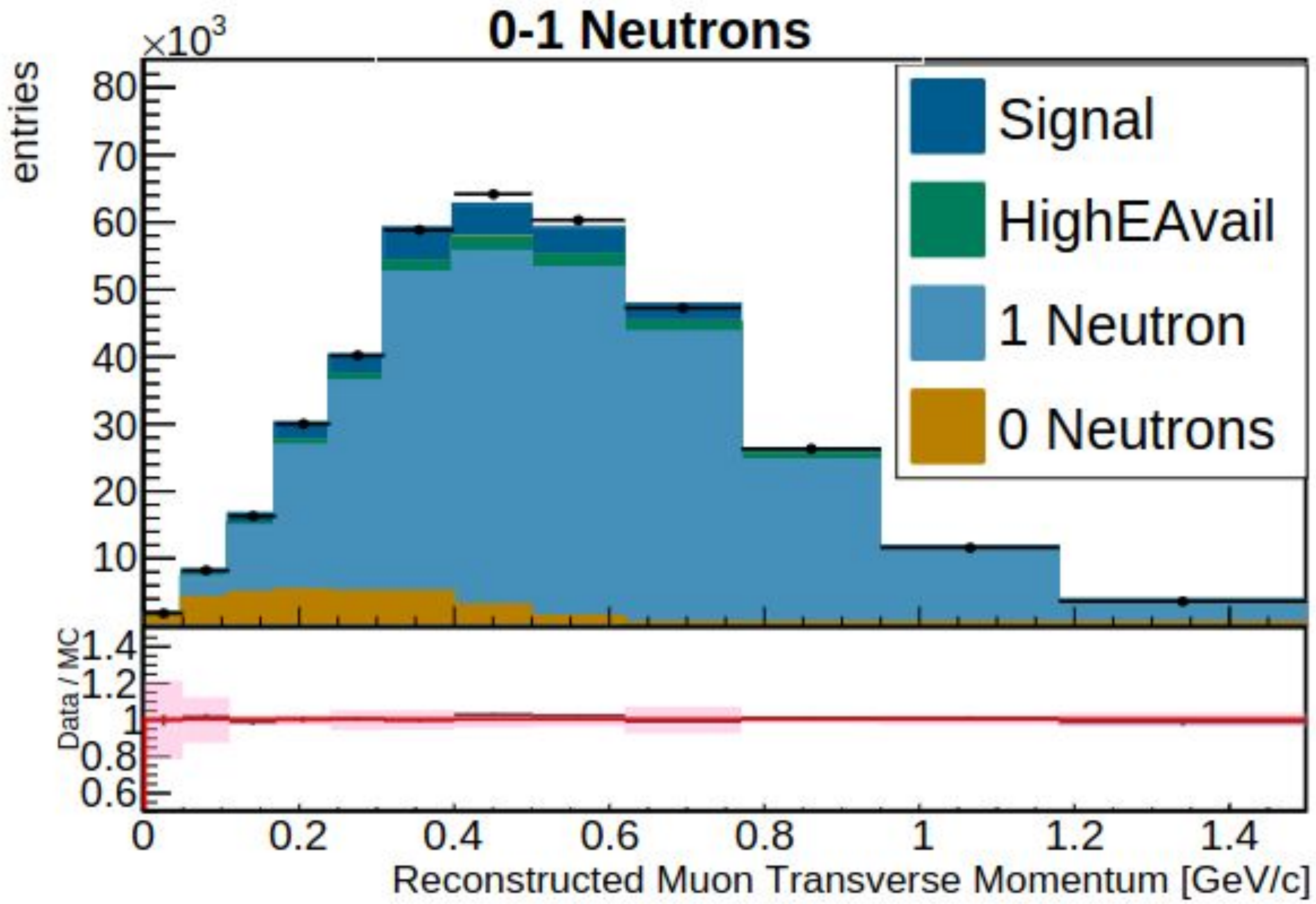
Few-neutron sideband (**pre-fit**)

Background control samples

[Phys. Rev. D 108, 112010 (2023)]

- **Control samples**: **check/correct background model** based on parallel measurement
  - Background-enhanced selection
- Also often referred to as "**sidebands**"
  - I use the terms interchangeably in the paper
- I propose a semi-new way of using these for cross-section analyses

anti-ν_μ CC 2+ neutrons (MINERvA)



Few-neutron sideband (post-fit)

Use by experiments

- **T2K** gets background model constraints "for free"
  - Just include bins from the sideband(s) in the fit!
- **MINERvA**: normalization scale factor approach
  - **Pre-fit**: $\alpha_þ = 1$ for all background classes þ
  - **Post-fit** values obtained from sidebands
    - Details vary widely
  - Shape from simulation unaltered*
  - Assumes 100% correlation between $\alpha_þ$ in sidebands and **signal region**

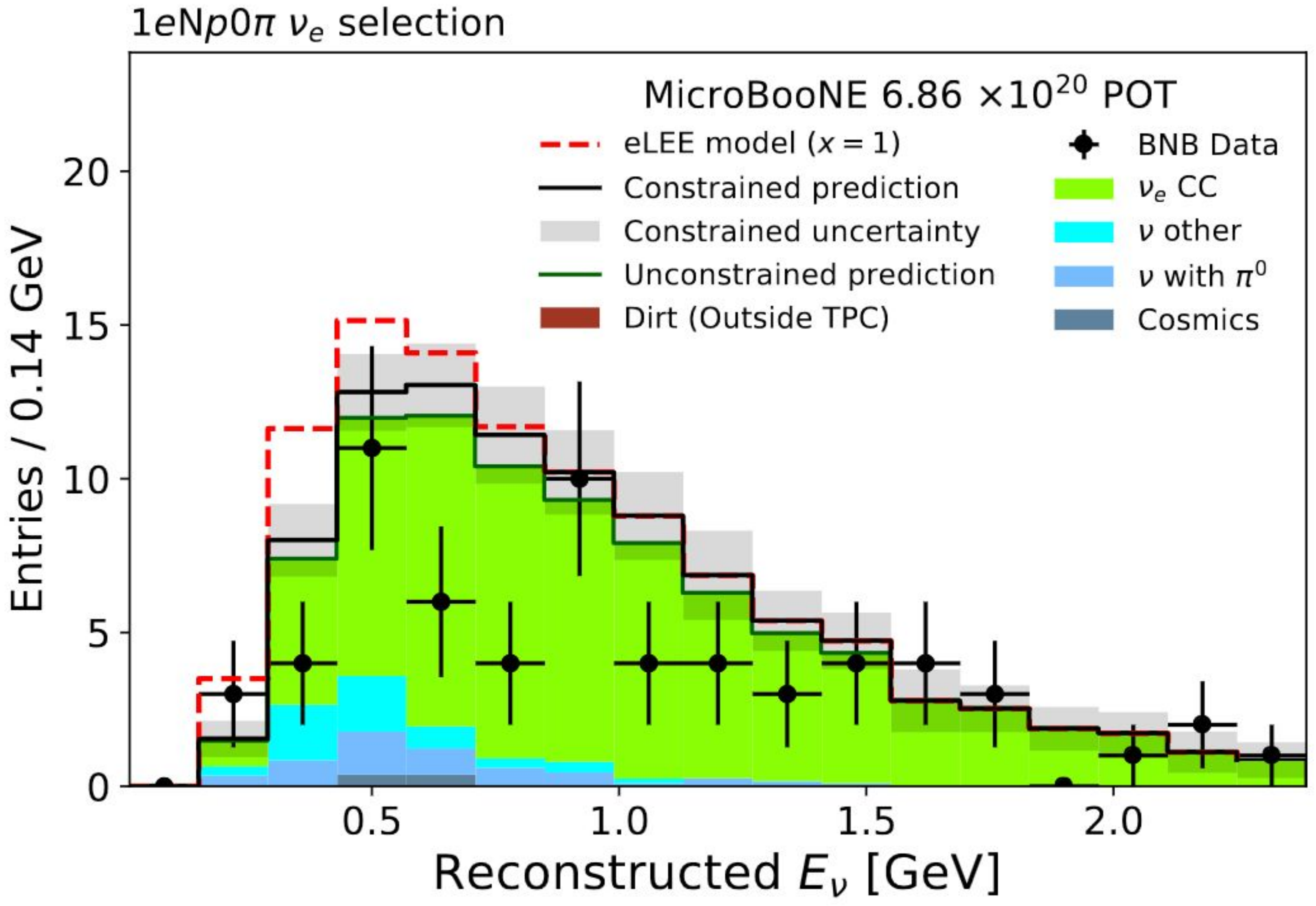
$$B_a = \sum_{þ} \alpha_þ \, B_{aþ}$$

- **MicroBooNE**: no sidebands used as a constraint for any multi-bin cross-section result so far
  - I **generalize and improve** a method used for single-bin η analysis

Data-driven constraint in MicroBooNE LEE analyses

- MicroBooNE built to investigate anomalous excess of ν_e-like events seen by MiniBooNE at low energies ("**LEE**")
- First results October 2021
  - Data prefer **no excess**
- Judged relative to prediction of "MicroBooNE GENIE tune" with **data-driven, analysis-specific adjustments**
- All based on a **conditional covariance** treatment

Phys. Rev. D 105, 112004 (2022)



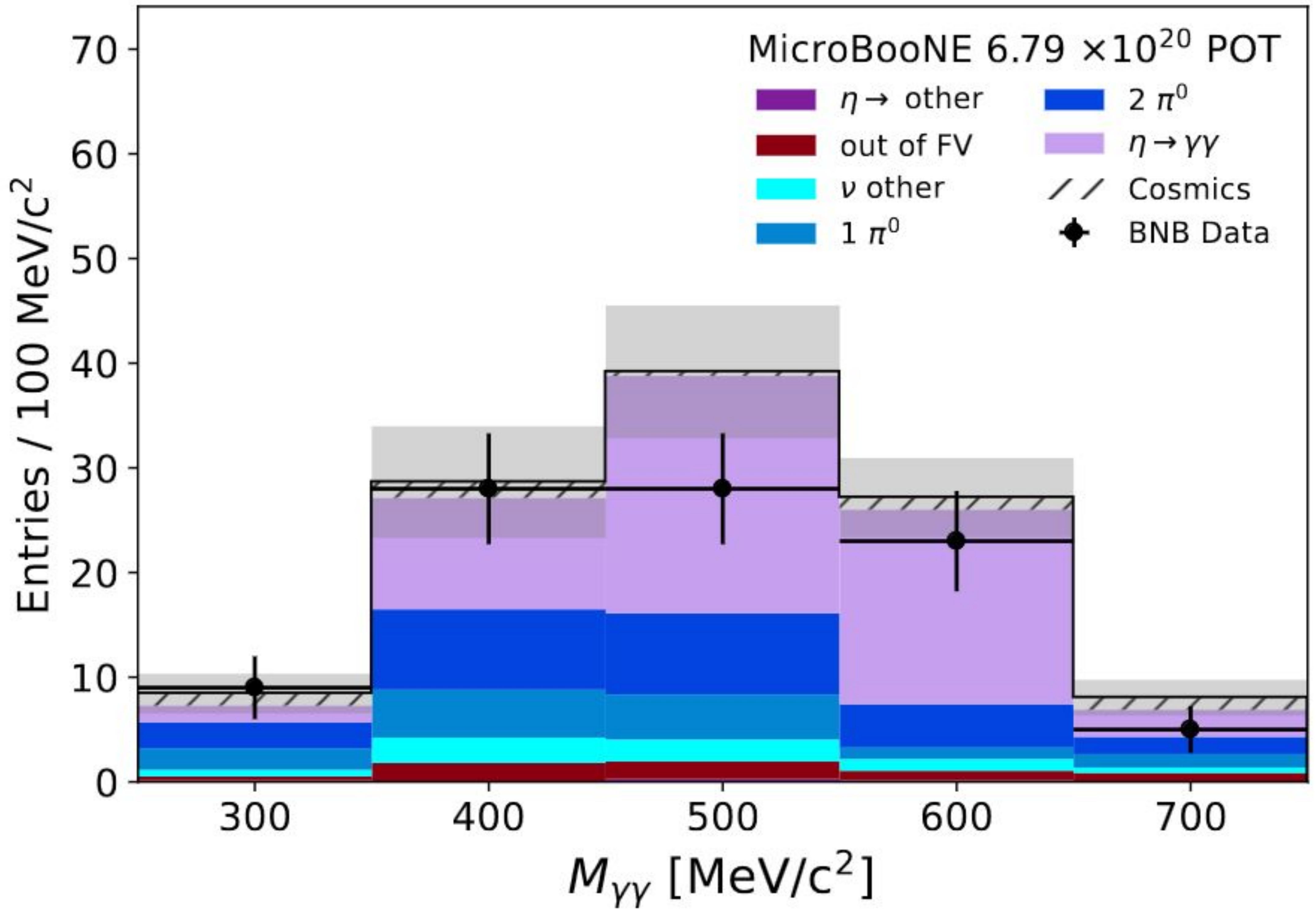
$$m^{e\,\text{constrained}} = m^e + C^{e\mu}(C^{\mu\mu})^{-1}(n^\mu - m^\mu)$$

$$C^{ee\,\text{constrained}} = C^{ee} - C^{e\mu}(C^{\mu\mu})^{-1}C^{\mu e}$$

Use for a background model constraint

Phys. Rev. Lett. 132, 151801 (2024)

- MicroBooNE η production study
  - Signal is two photons with the η invariant mass
- Dominant backgrounds are single- and multi-π^0 production
  - Each constrained separately with a single sideband bin
- I **generalize this procedure** for multiple bins and simultaneous fits to multiple backgrounds
  - Treatment suitable for MicroBooNE-style extraction



$$N_{MC}^{S,constrained} = N_{MC}^{S} + \frac{\sigma^{corr}}{(\sigma^{B})^2} \times \left(N_{data}^{B} - N_{MC}^{B}\right)$$

$$\left(\sigma^{S,\text{ constrained}}\right)^2 = \left(\sigma^{S}\right)^2 - \frac{\left(\sigma^{corr}\right)^2}{\left(\sigma^{B}\right)^2}$$

Can also be adapted to MINERvA's style (no 100% correlation assumption)

Conditional Covariance Background Constraint (CCBC)

Form a vector Y of predicted total (n) and background-only (B) event counts, compute covariances similarly to the the usual way

$$\mathbf{Y} \equiv \begin{pmatrix} \mathbf{n}_S \\ \mathbf{B}_S \\ \mathbf{n}_C \end{pmatrix} \qquad V_{\mathbf{YY}} = \begin{pmatrix} V_{\mathbf{n}_S\mathbf{n}_S} & V_{\mathbf{n}_S\mathbf{B}_S} & V_{\mathbf{n}_S\mathbf{n}_C} \\ V_{\mathbf{B}_S\mathbf{n}_S} & V_{\mathbf{B}_S\mathbf{B}_S} & V_{\mathbf{B}_S\mathbf{n}_C} \\ V_{\mathbf{n}_C\mathbf{n}_S} & V_{\mathbf{n}_C\mathbf{B}_S} & V_{\mathbf{n}_C\mathbf{n}_C} \end{pmatrix}$$

Use the observed control sample (C) event counts to constrain those in the signal region

$$\mathbf{B}_S^{\mathrm{constr}} = \mathbf{B}_S^{\mathrm{CV}} + V_{\mathbf{B}_S\mathbf{n}_C} \cdot V_{\mathbf{n}_C\mathbf{n}_C}^{-1} \cdot \left(\mathbf{D}_C - \mathbf{n}_C^{\mathrm{CV}}\right)$$

$$\mathbf{n}_S^{\mathrm{constr}} = \mathbf{n}_S^{\mathrm{CV}} + V_{\mathbf{n}_S\mathbf{n}_C} \cdot V_{\mathbf{n}_C\mathbf{n}_C}^{-1} \cdot \left(\mathbf{D}_C - \mathbf{n}_C^{\mathrm{CV}}\right)$$

$$V_{\mathbf{n}_S\mathbf{n}_S}^{\mathrm{constr}} = V_{\mathbf{n}_S\mathbf{n}_S} - V_{\mathbf{n}_S\mathbf{n}_C} \cdot V_{\mathbf{n}_C\mathbf{n}_C}^{-1} \cdot V_{\mathbf{n}_S\mathbf{n}_C}^{T}$$

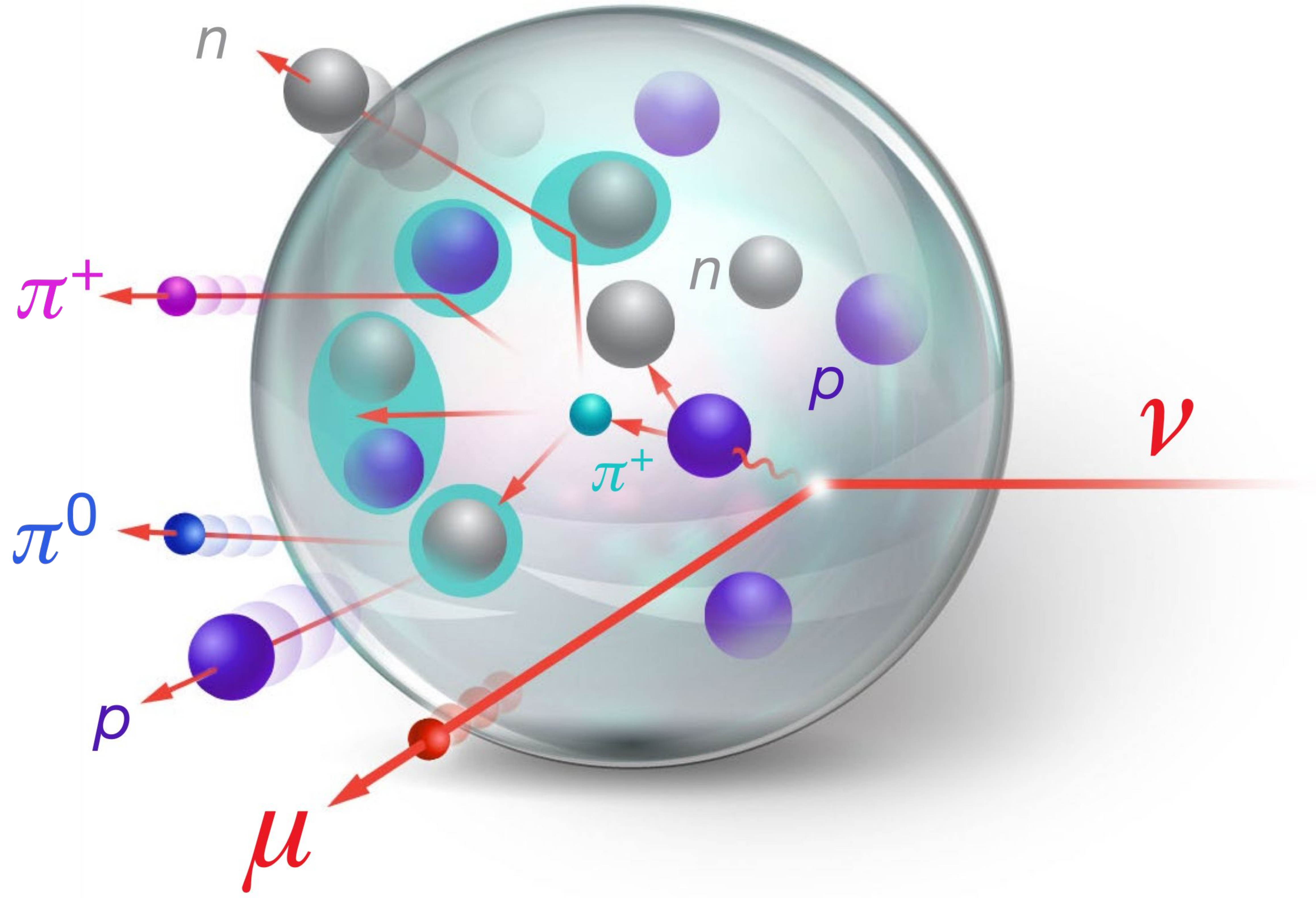
Can check data/MC agreement post-constraint for sanity. Use constrained B and V as input to MicroBooNE-style extraction

Outlook for the CCBC

- Provides a data-driven background constraint for the MicroBooNE style
  - Can potentially be adapted for use in MINERvA context
  - Still requires building reco-space covariances
- Being tried out in MicroBooNE, not yet used in any public result
- Allows the **full simulation to inform assumed relationship** between sideband/signal regions
  - Shouldn't trust blindly, can re-assess goodness of fit after constraint
- Akin to what T2K gets "for free" by **including sidebands in likelihood fit**
  - Compatible with matrix-inversion strategies for unfolding
- Offered as an idea to the community, also encouragement for further exploration in MicroBooNE and elsewhere

Conclusion

- Recent paper (arXiv:2401.04065) proposes some adjustments to how we extract neutrino cross section data
- "**Blockwise unfolding**" enables full reporting of correlated uncertainties
  - Make our hard work even more informative
- **MicroBooNE ν_μ CC0πNp** results (arXiv:2403.19574) provide detailed demonstration
  - Overall goodness-of-fit reveals interesting tensions



- **CCBC** provides new way of refining background predictions with data
  - Basic idea has existed for some time, now applied to cross-section extraction