

Bayesian Monte Carlo Extraction of Sea Asymmetry with SeaQuest and STAR Data



Chris Cocuzza (Temple U.)
Andreas Metz (Temple U.)
Wally Melnitchouk (Jefferson Lab)
Nobuo Sato (Jefferson Lab)

May 4, 2022

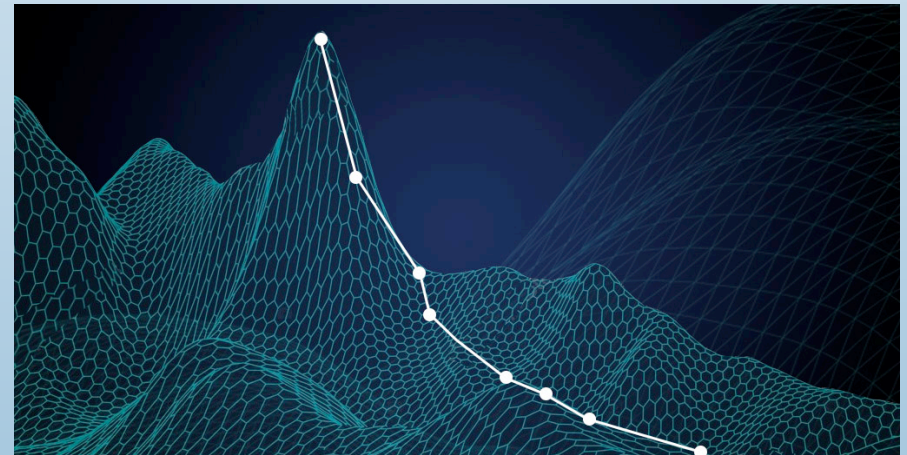
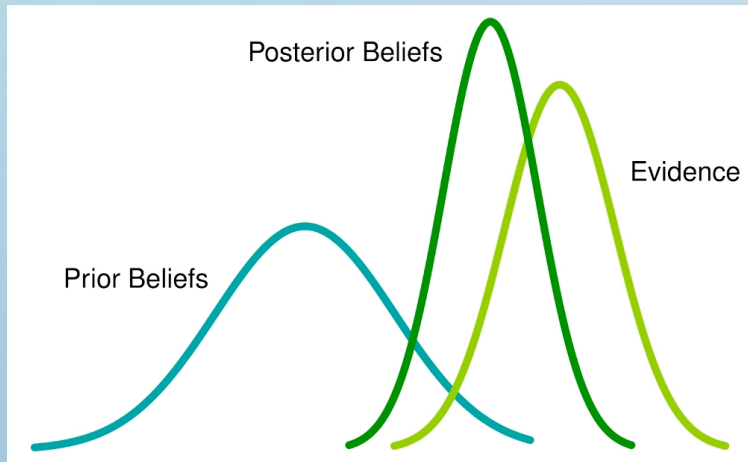


1. Introduction

2. Data and Fitting

3. Extracted Sea Asymmetry

4. Conclusions and Outlook



JAM Collaboration

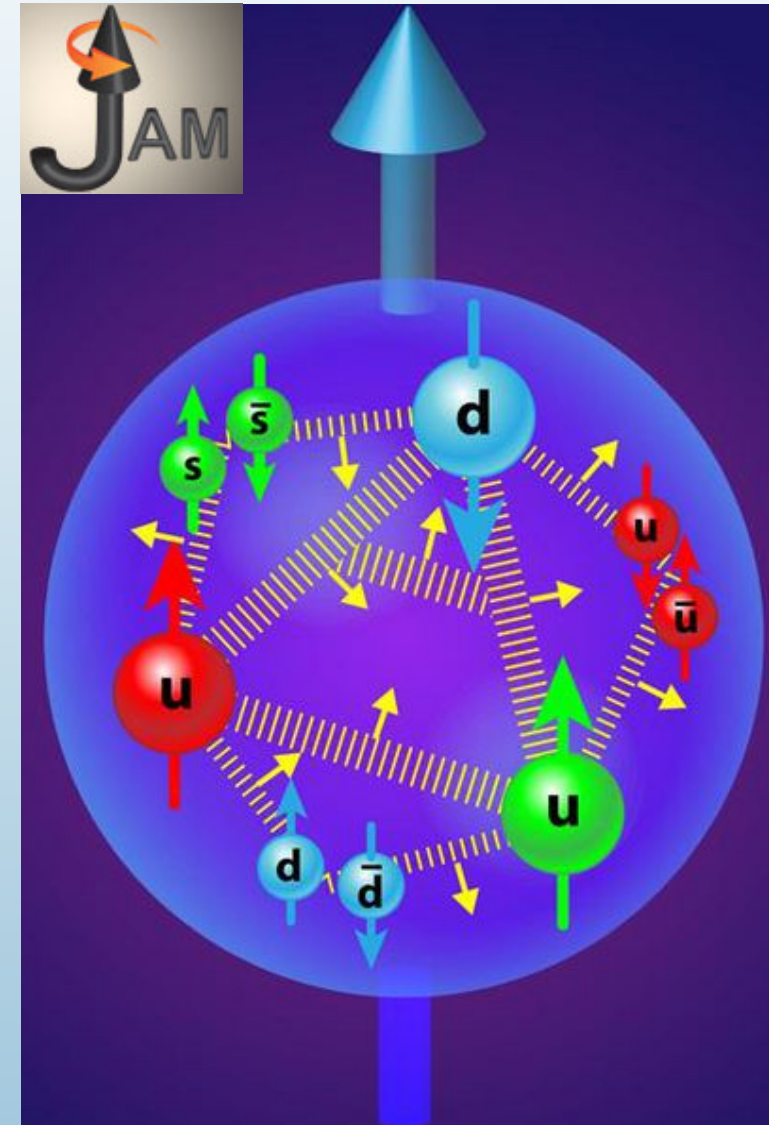
3-dimensional structure of nucleons:

- Parton distribution functions (PDFs)
- Fragmentation functions (FFs)
- Transverse momentum dependent distributions (TMDs)
- Generalized parton distributions (GPDs)

Collinear factorization in perturbative QCD

Simultaneous determinations of PDFs, FFs, etc.

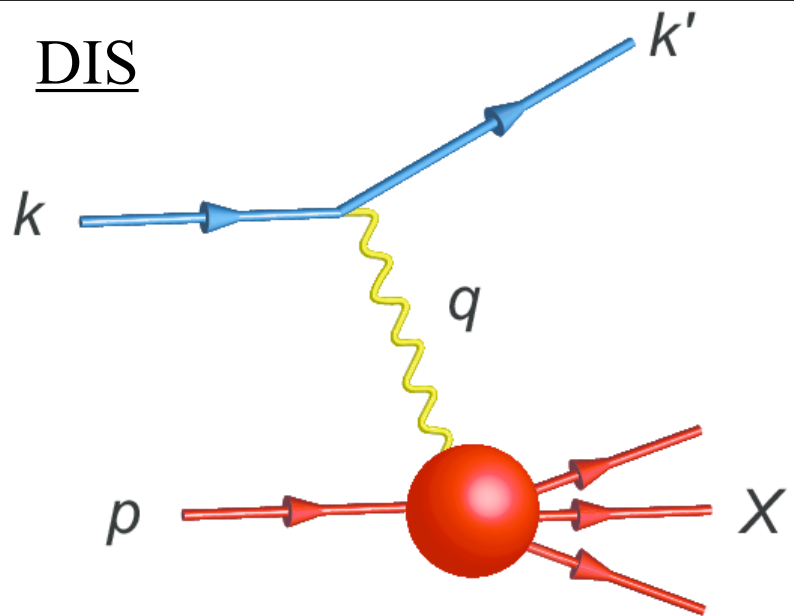
Monte Carlo methods for Bayesian inference



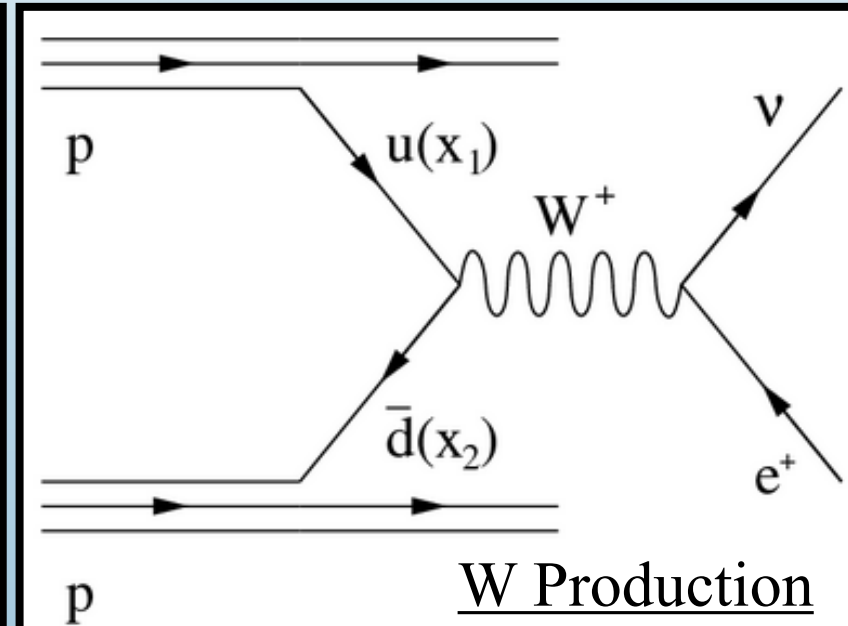
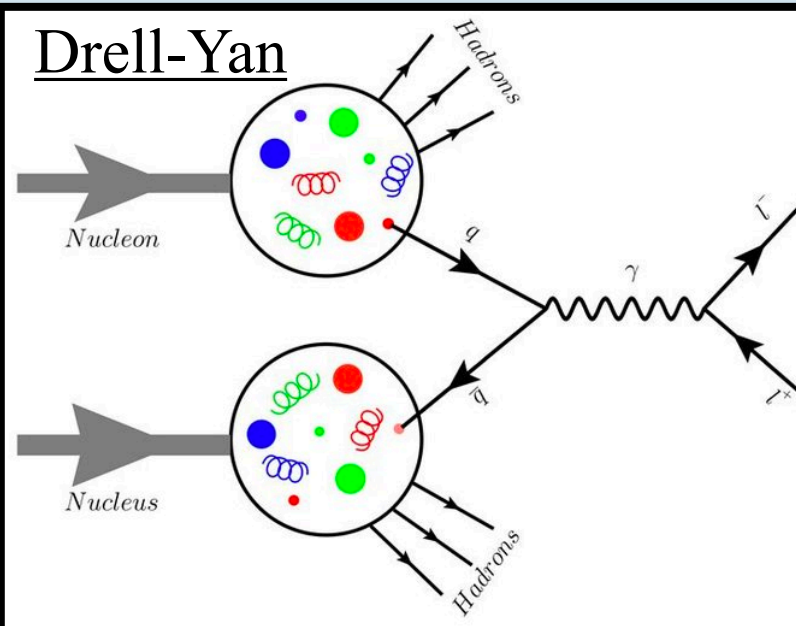
A Global Analysis

Spin-averaged PDFs are extracted using many different processes

DIS



Drell-Yan



W Production

Parameters to Observables

Parameterize PDFs at input scale $Q_0^2 = m_c^2$

$$f_i(x) = Nx^\alpha(1-x)^\beta(1+\gamma\sqrt{x}+\eta x)$$

Evolve PDFs using DGLAP

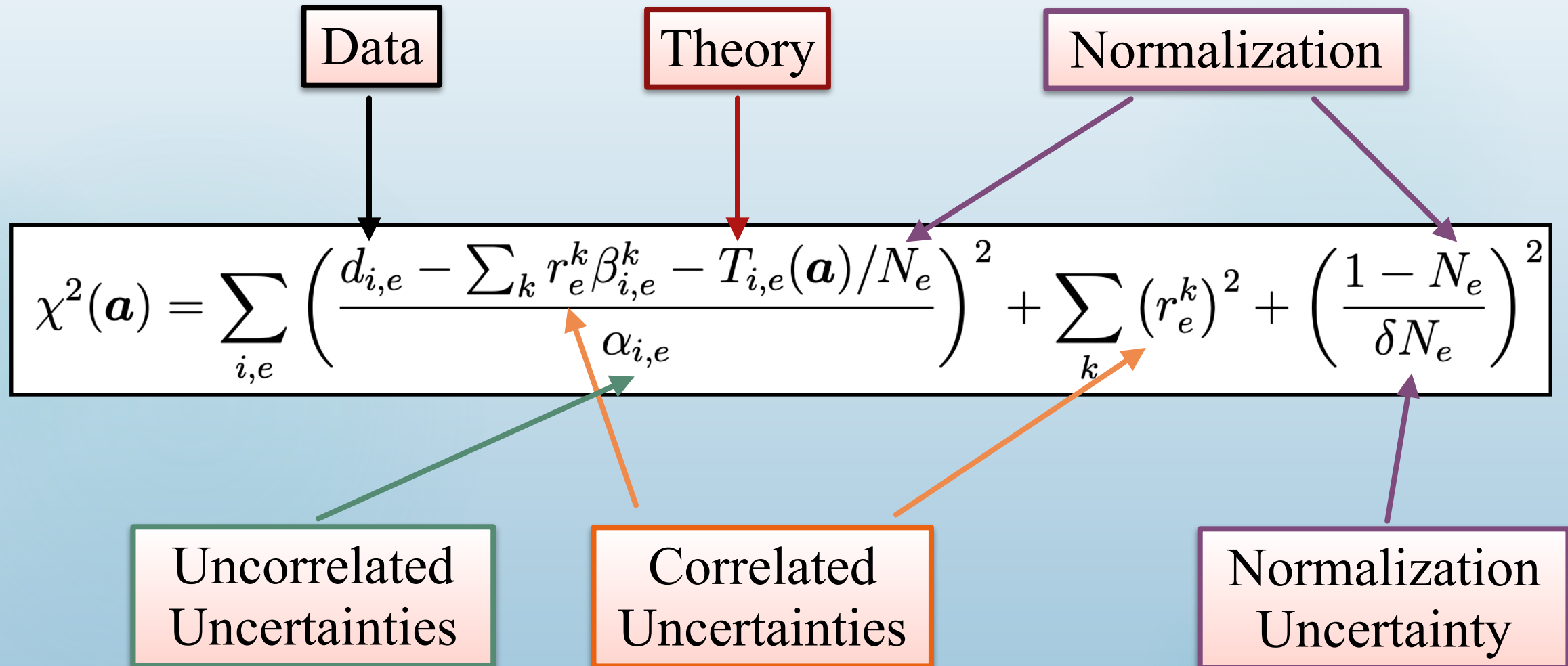
$$\frac{d}{d \ln(\mu^2)} f_i(x, \mu) = \sum_j \int_x^1 \frac{dz}{z} P_{ij}(z, \mu) f_j\left(\frac{x}{z}, \mu\right)$$

Calculate Observables

$$d\sigma_{\text{DY}} = \sum_{i,j} H_{ij}^{\text{DY}} \otimes f_i \otimes f_j$$

The χ^2 function

Now that the observables have been calculated...

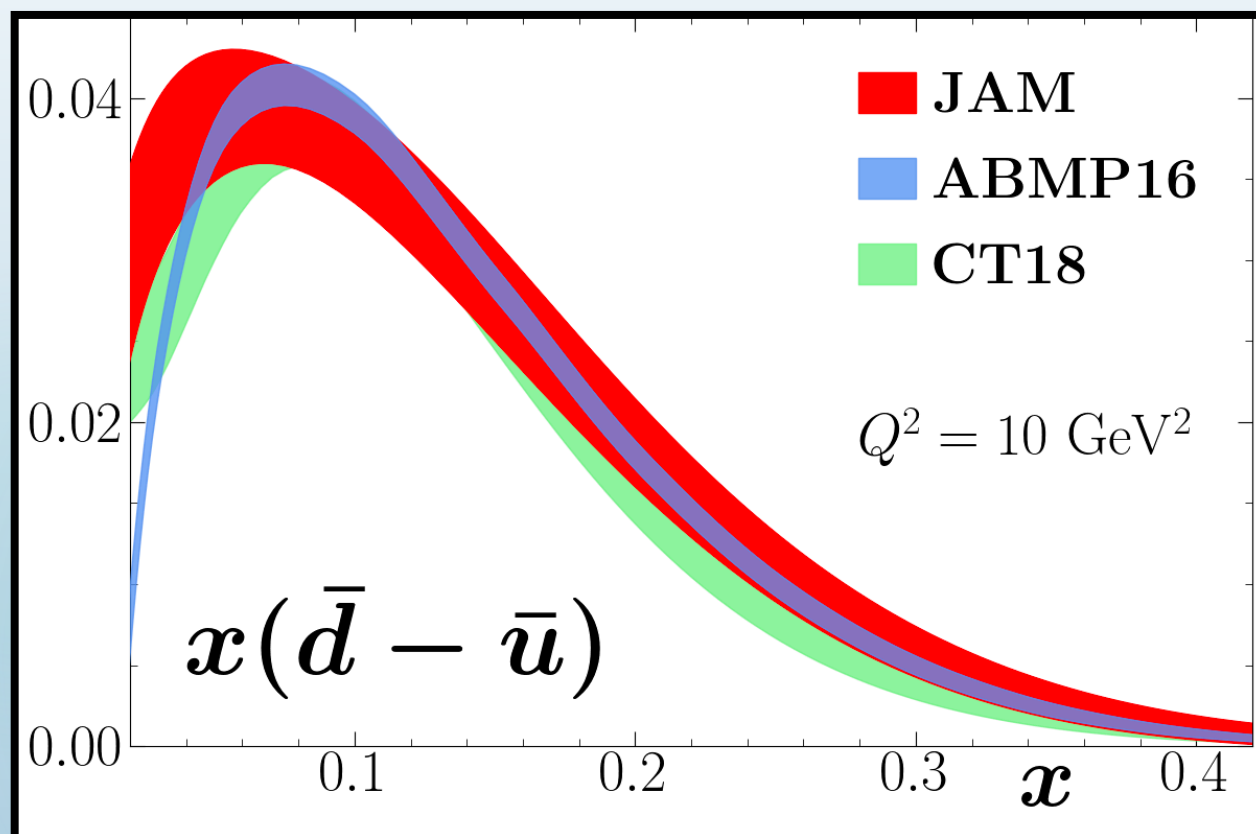


$$\chi^2(\mathbf{a}) = \sum_{i,e} \left(\frac{d_{i,e} - \sum_k r_e^k \beta_{i,e}^k - T_{i,e}(\mathbf{a})/N_e}{\alpha_{i,e}} \right)^2 + \sum_k (r_e^k)^2 + \left(\frac{1 - N_e}{\delta N_e} \right)^2$$

Diagram illustrating the components of the χ^2 function:

- Data** (points to $d_{i,e}$)
- Theory** (points to $T_{i,e}(\mathbf{a})/N_e$)
- Normalization** (points to N_e)
- Uncorrelated Uncertainties** (points to $\alpha_{i,e}$)
- Correlated Uncertainties** (points to $\alpha_{i,e}$)
- Normalization Uncertainty** (points to δN_e)

Introduction to Sea Asymmetry



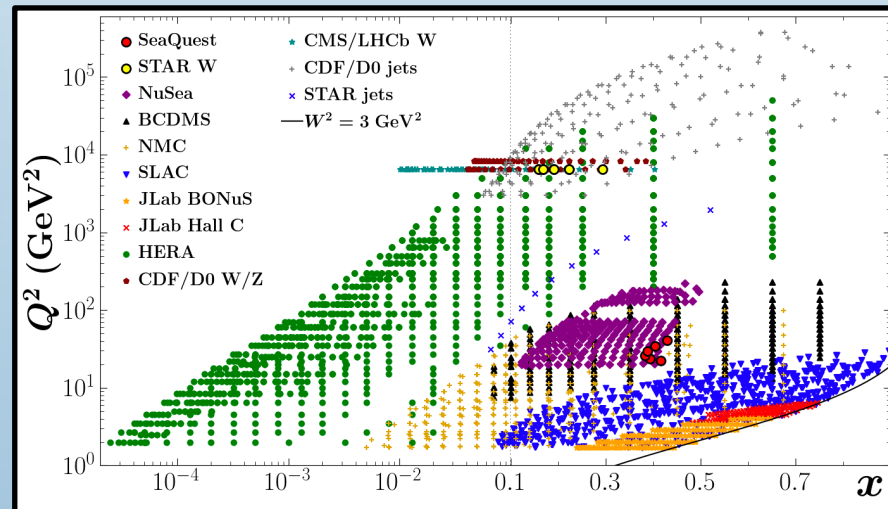
Unpolarized

Cannot be explained from gluons splitting into quark-antiquark pairs

Meson Cloud Models
 Chiral Soliton Models
 Statistical Models

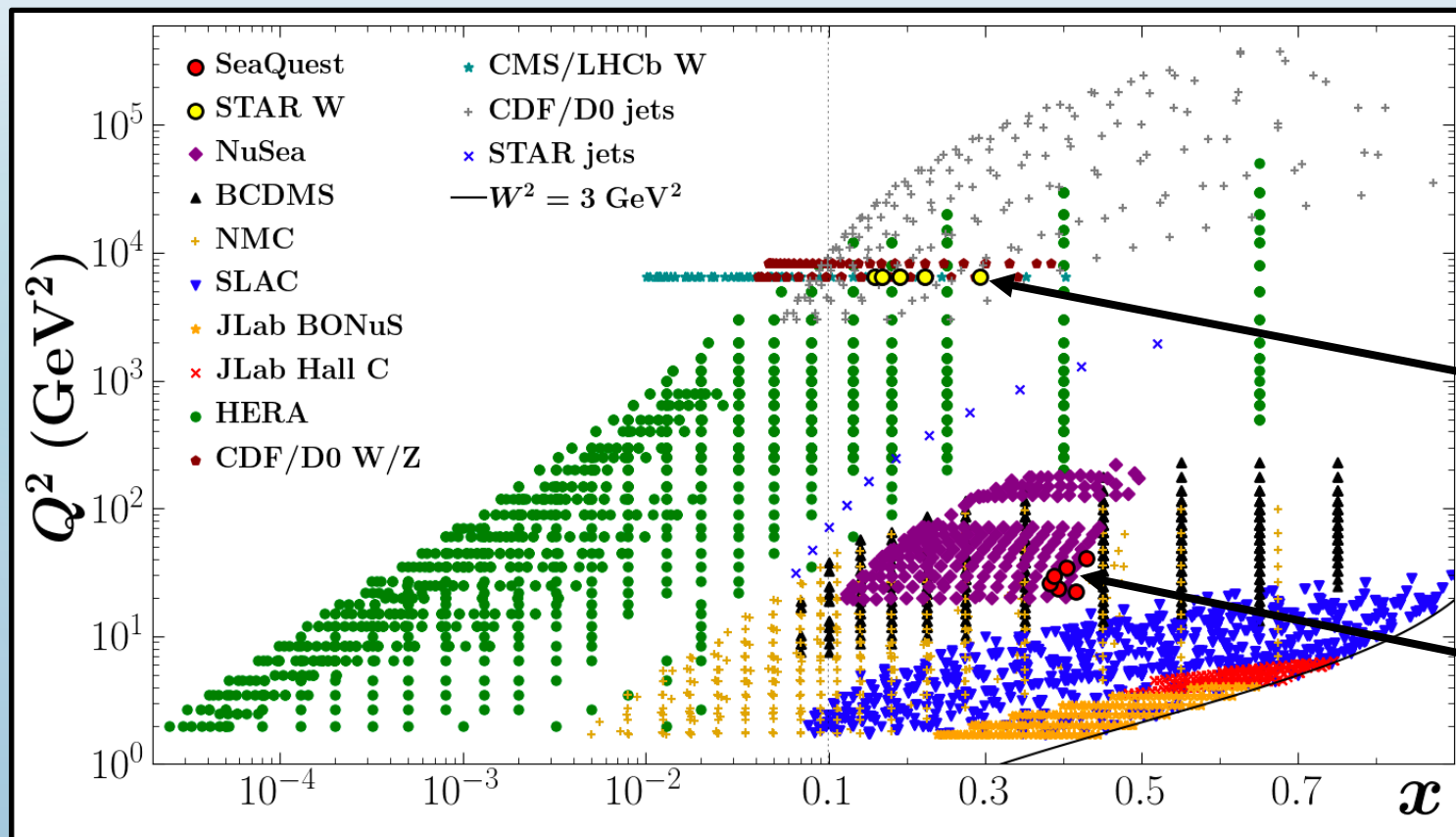
Still questions at high $x > 0.2$ and for helicity asymmetry

1. Introduction
2. Data and Fitting
3. Extracted Sea Asymmetry
4. Conclusions and Outlook



Kinematic Coverage (Spin-Averaged)

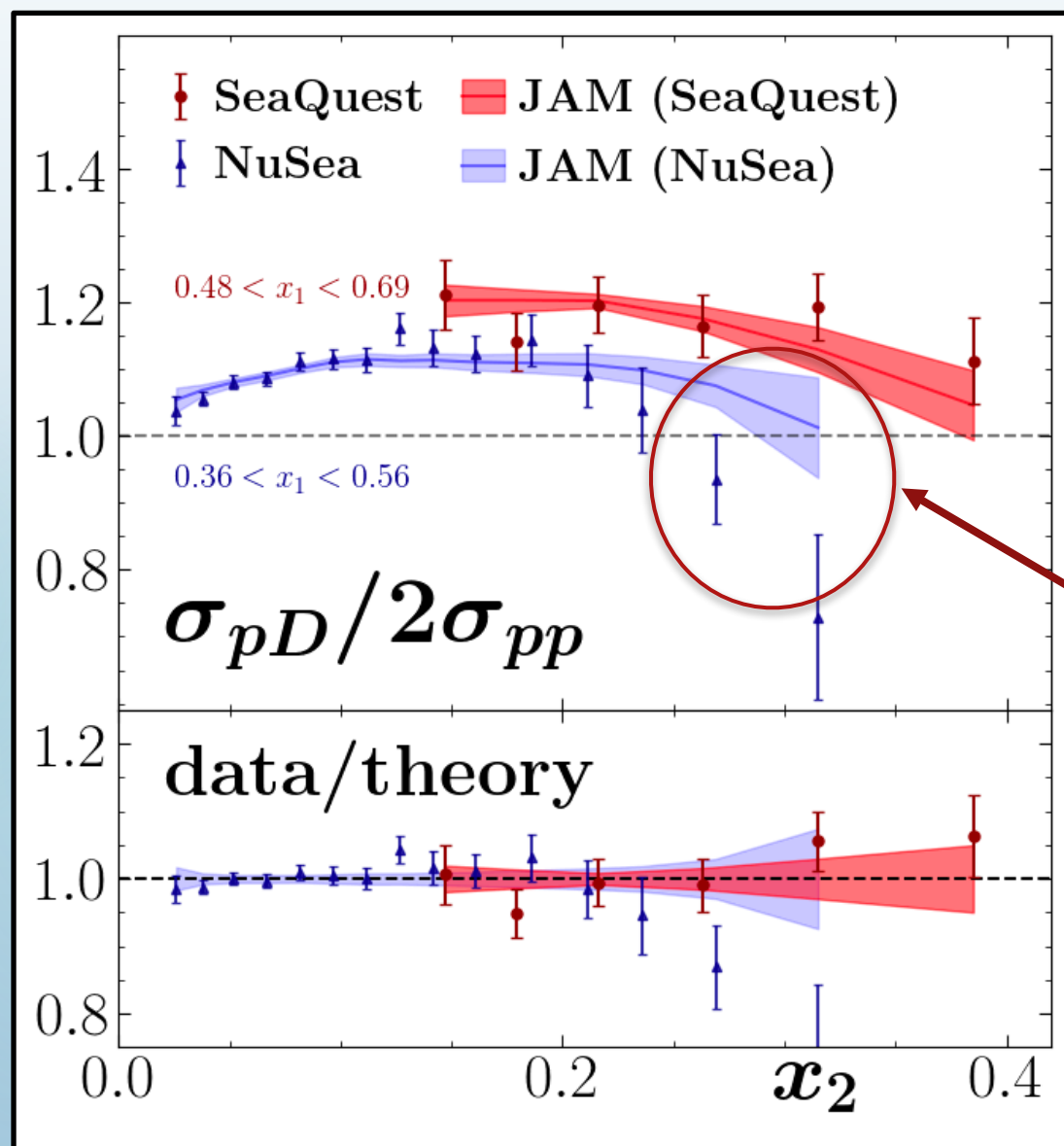
Deep Inelastic Scattering	BCDMS, NMC, SLAC, HERA, Jefferson Lab	3863 points
Drell-Yan	Fermilab E866, E906	205 points
W/Z Boson Production	CDF/D0, STAR, LHCb, CMS	153 points
Jets	CDF/D0, STAR	200 points



New STAR data

New SeaQuest data

SeaQuest and NuSea Quality of Fit

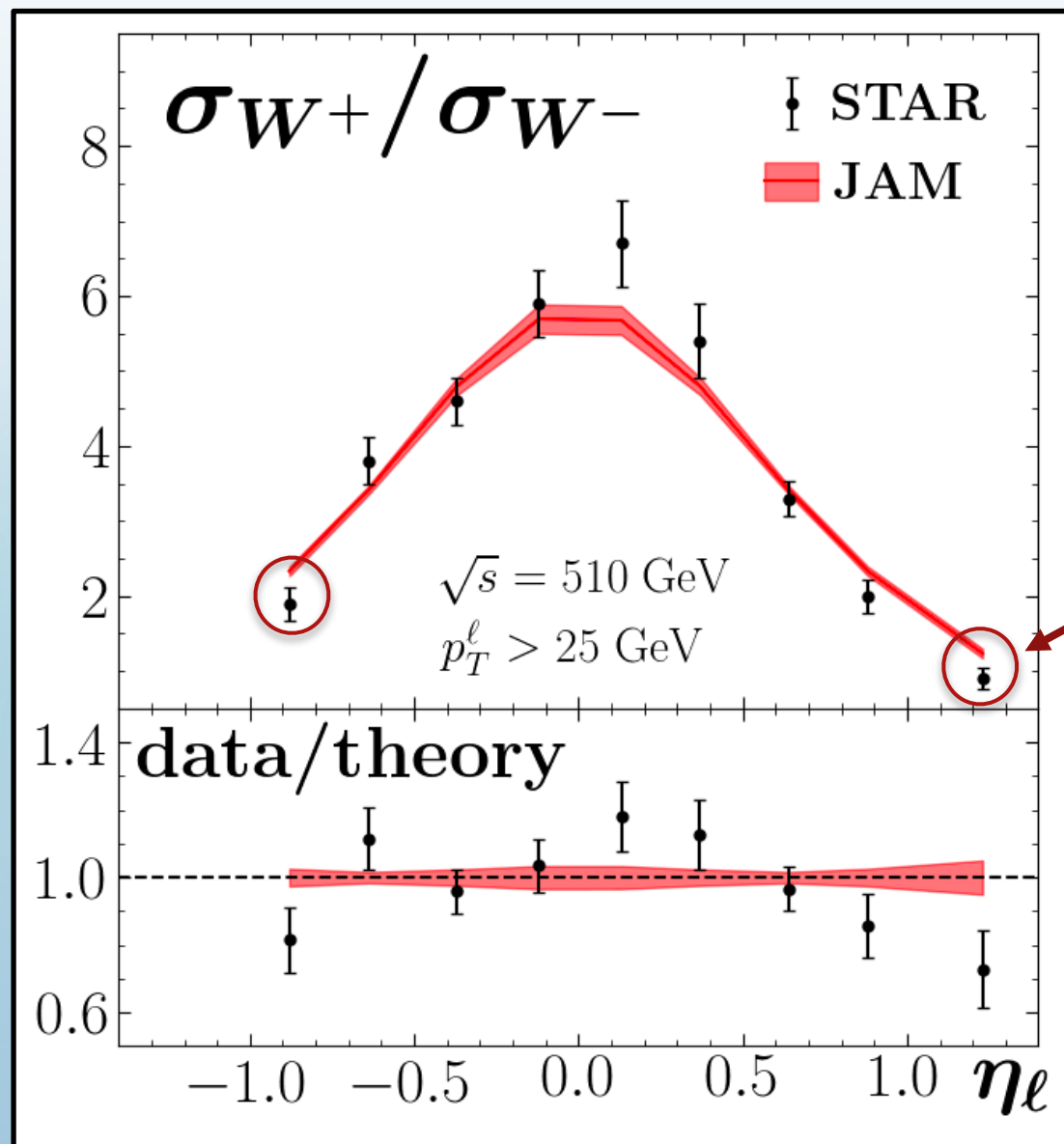


process		N_{dat}	χ^2/N_{dat}
Drell-Yan			
NuSea	pp	184	1.21
NuSea	$pD/2pp$	15	1.30
SeaQuest	$pD/2pp$	6	0.82

$$\left. \frac{\sigma_{pD}}{2\sigma_{pp}} \right|_{x_1 \gg x_2} \approx \frac{1}{2} \left[1 + \frac{\bar{d}(x_2)}{\bar{u}(x_2)} \right]$$

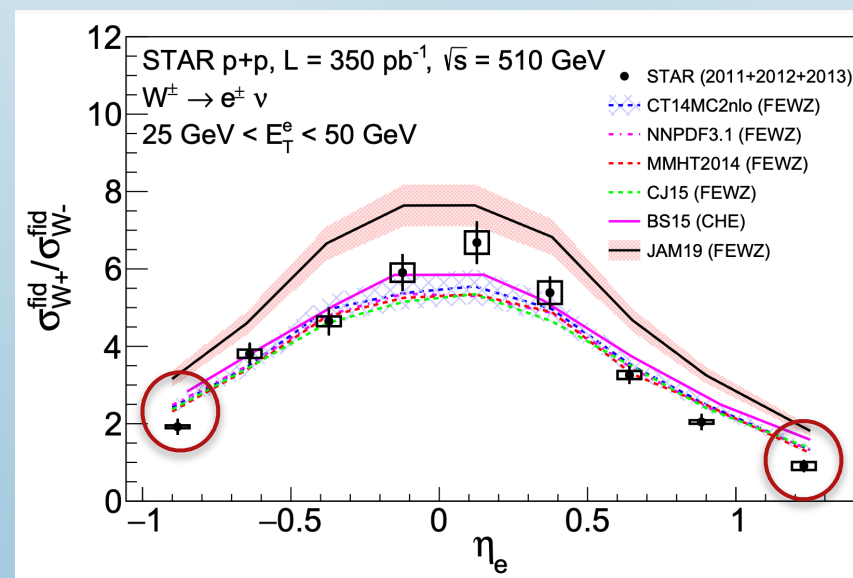
Well-known tension
between NuSea and
SeaQuest

STAR Quality of Fit

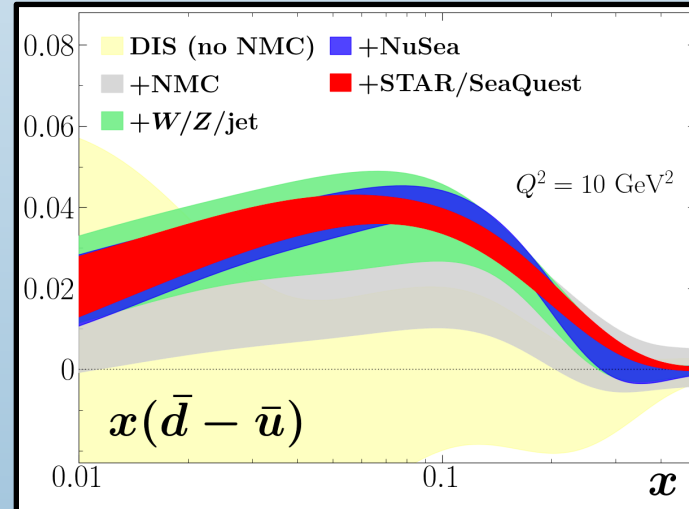


process	N_{dat}	χ^2/N_{dat}
W-lepton		
STAR W^+/W^-	9	2.02

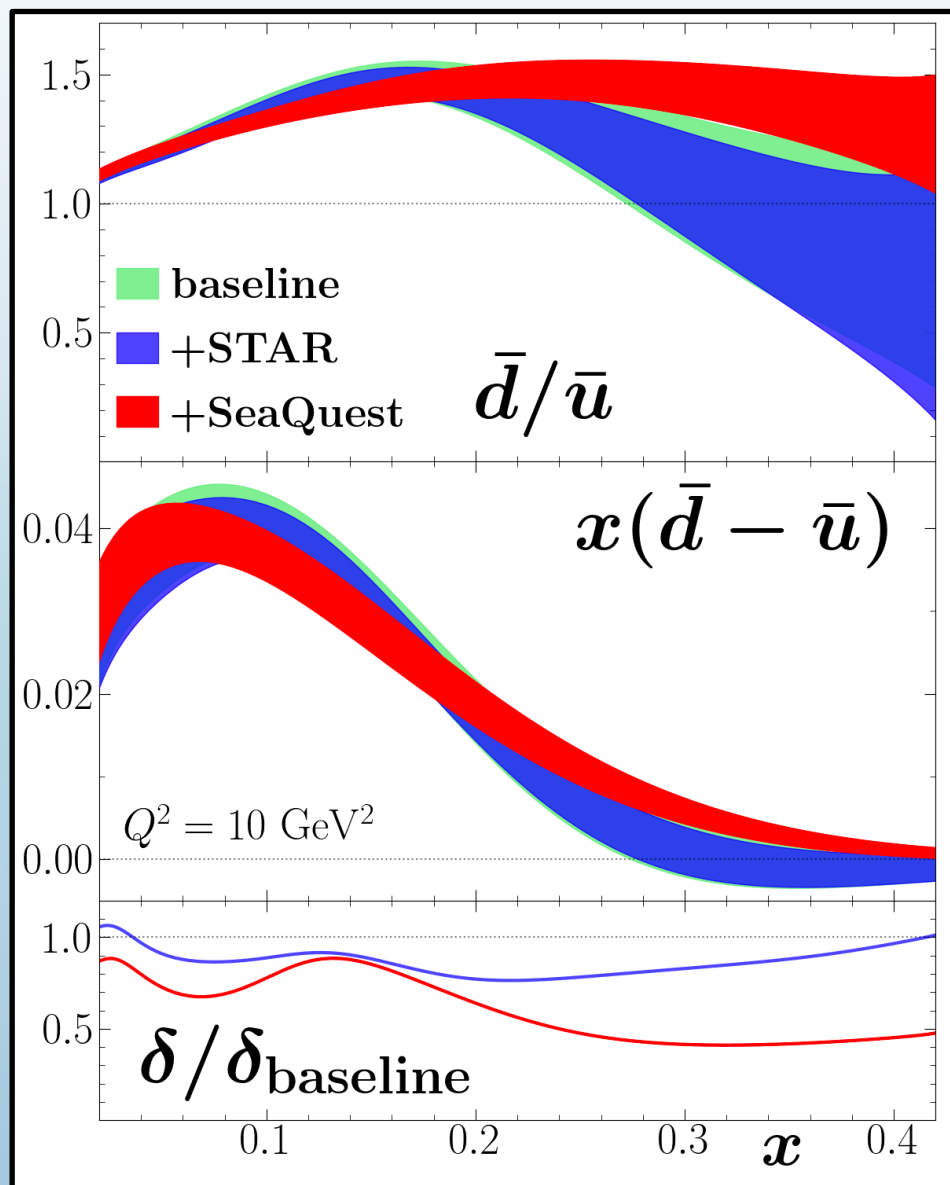
Difficult to describe at extreme rapidity



1. Introduction
2. Data and Fitting
- 3. Extracted Sea Asymmetry**
4. Conclusions and Outlook



Impact from STAR and SeaQuest

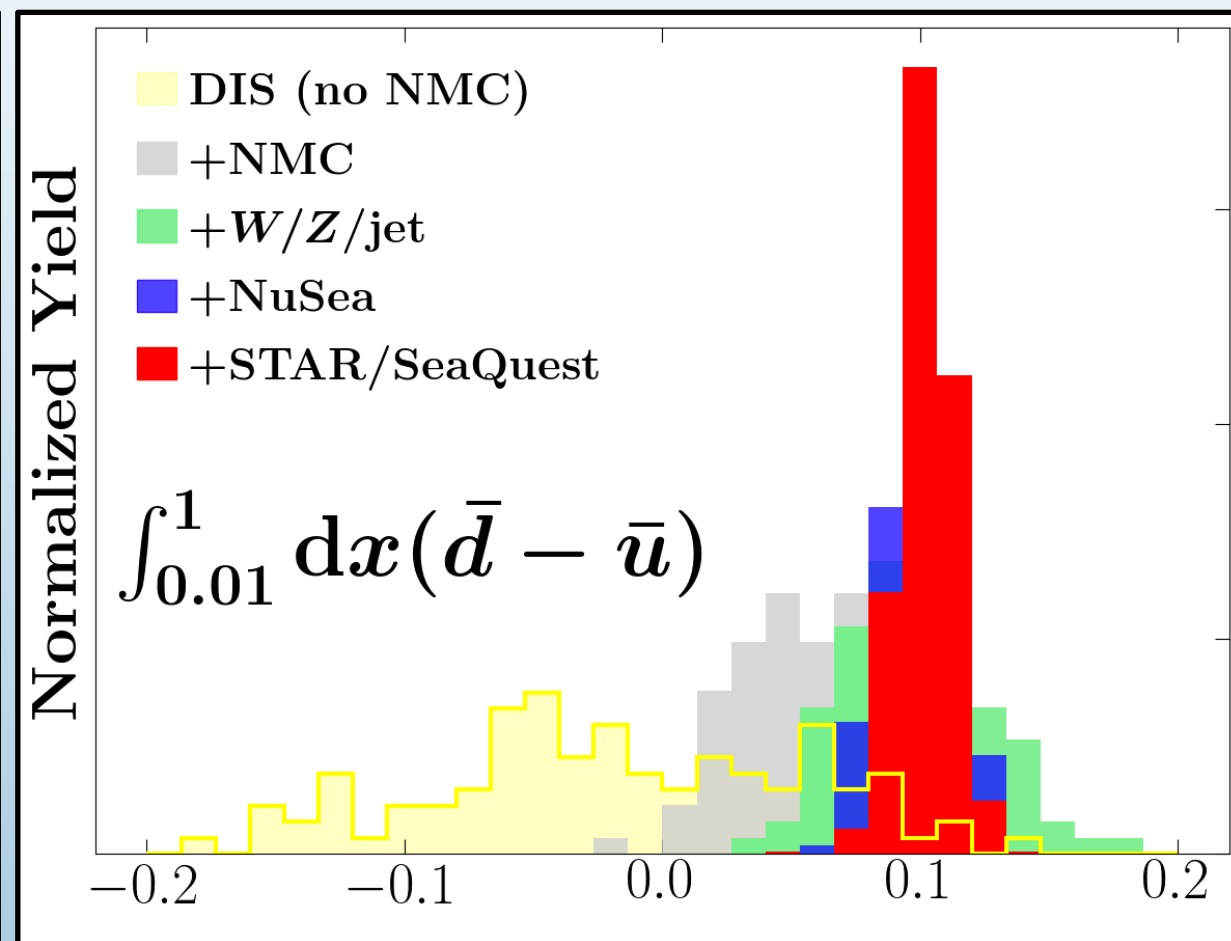
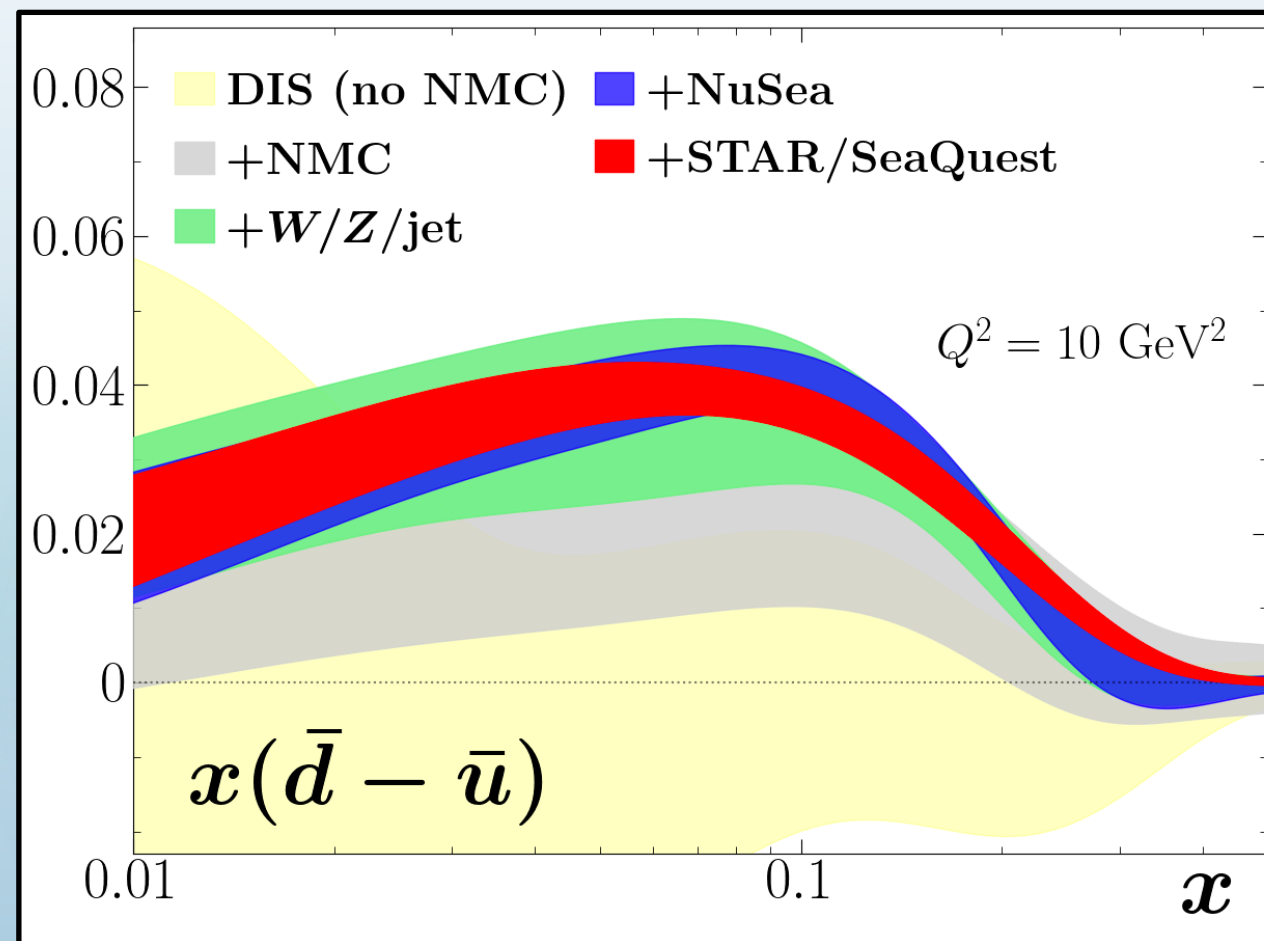


STAR: Moderate reduction of uncertainties

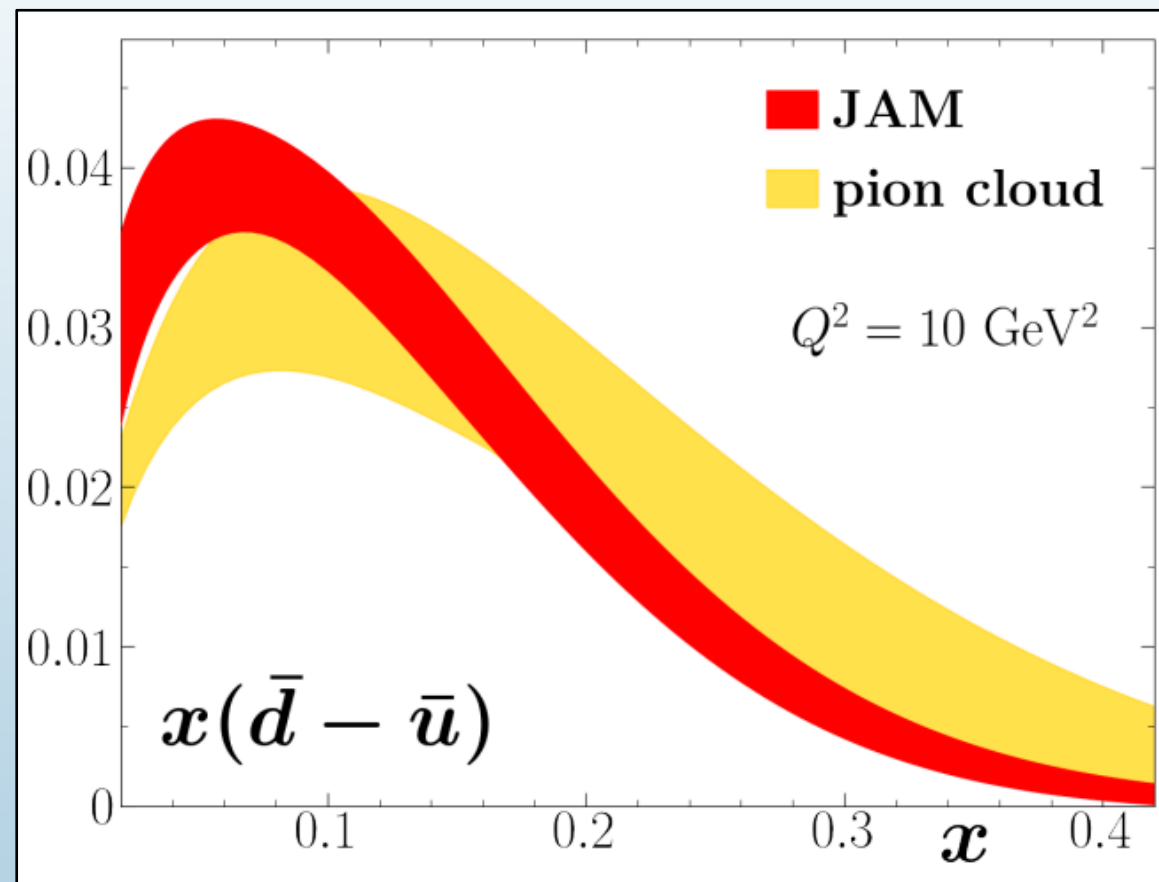
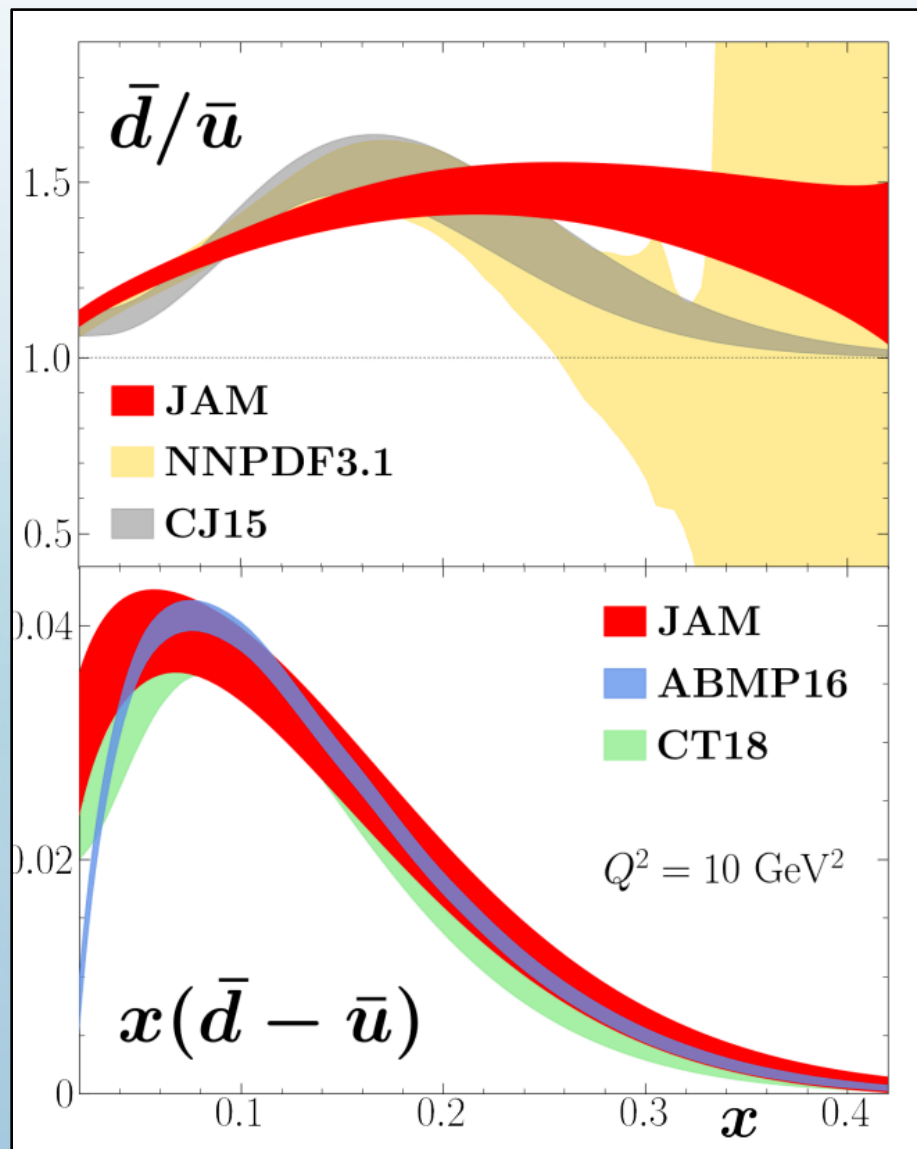
SeaQuest: Large reduction of uncertainties, especially at $x > 0.2$.

$\bar{d}/\bar{u} > 1$ up to $x \approx 0.4$, in agreement with models

Sources of Asymmetry

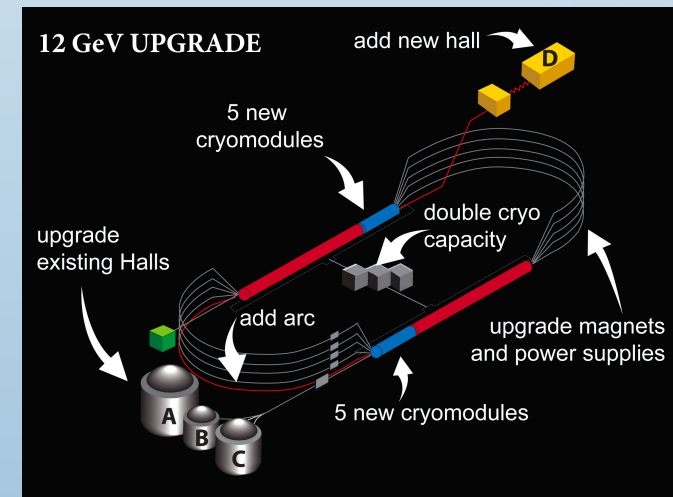
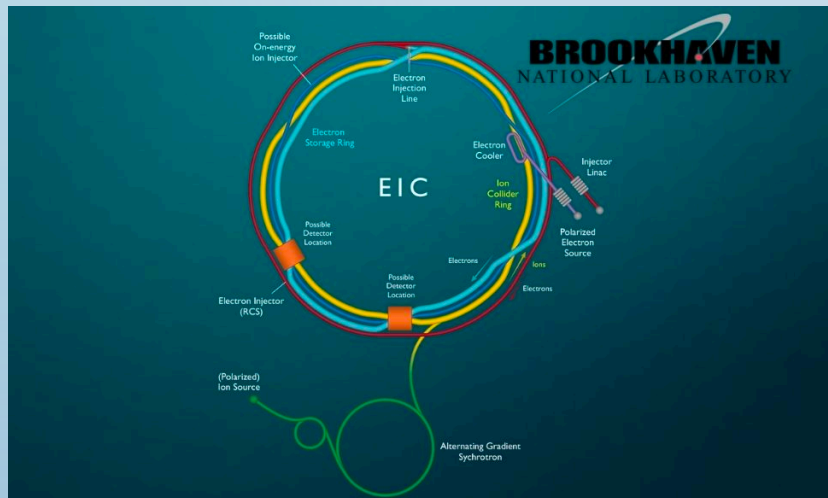


Comparison to other fits and pion cloud model



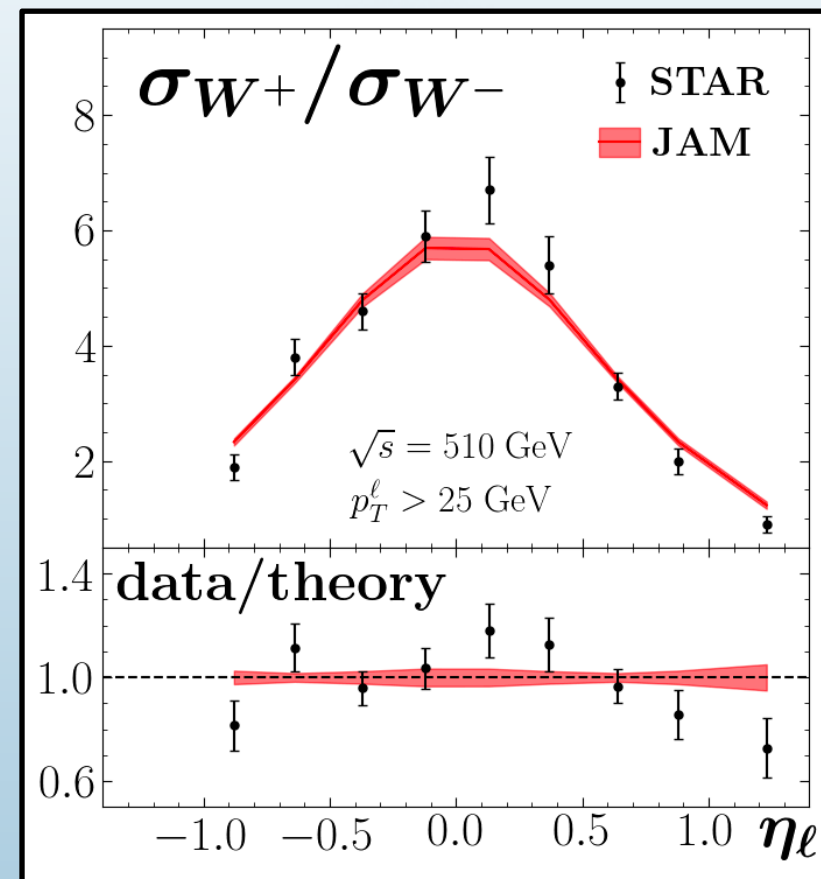
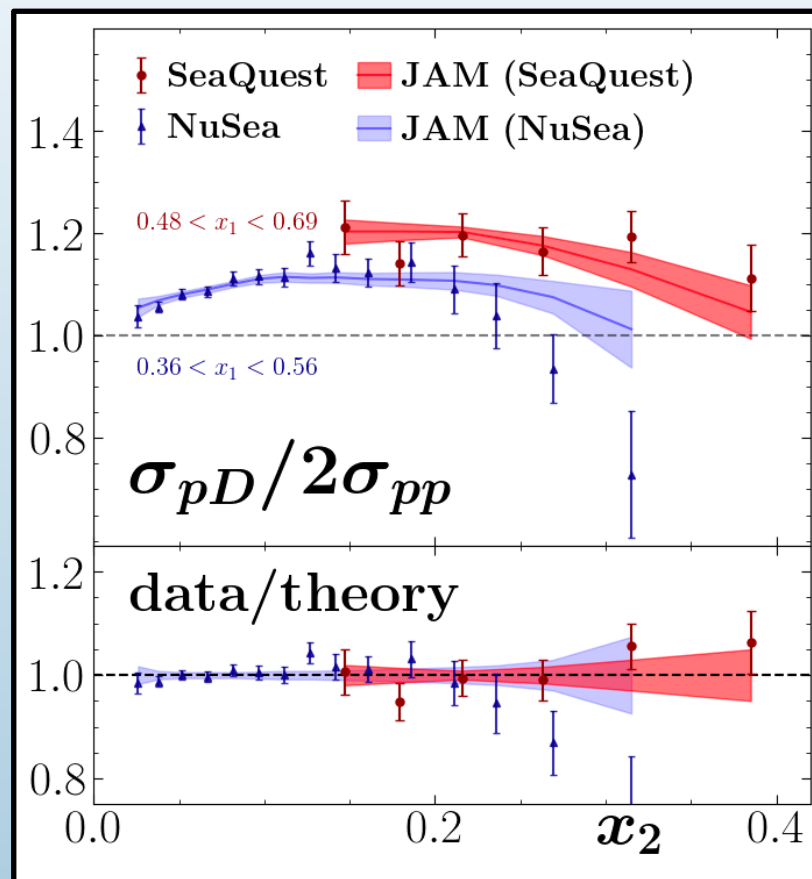
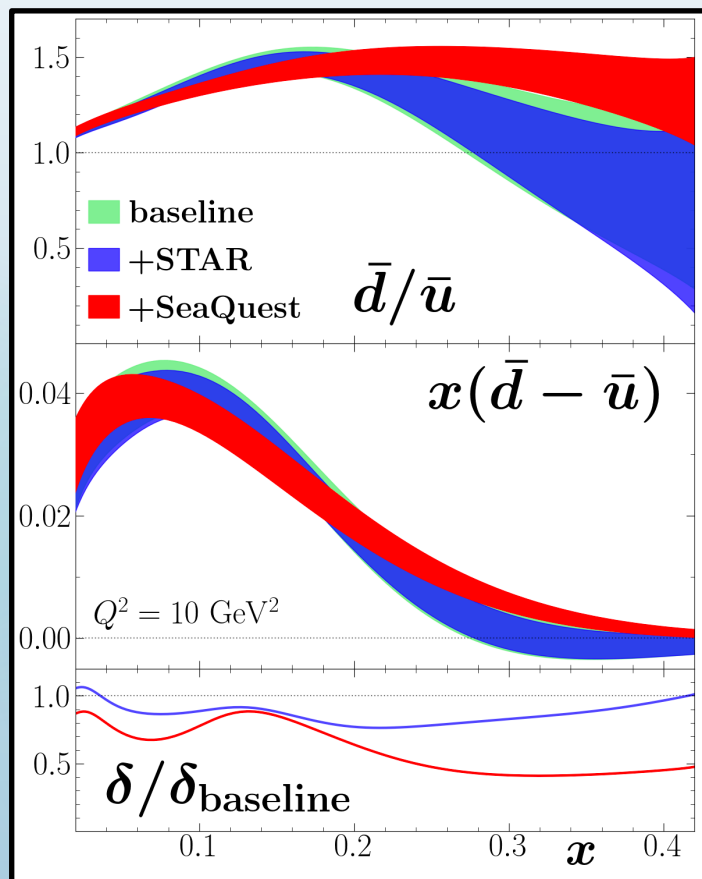
Good agreement with
pion cloud model

1. Introduction
2. Data and Fitting
3. Extracted Sea Asymmetry
4. Conclusions and Outlook



Results Summary

First global QCD analysis of SeaQuest and STAR data



<https://arxiv.org/abs/2109.00677>

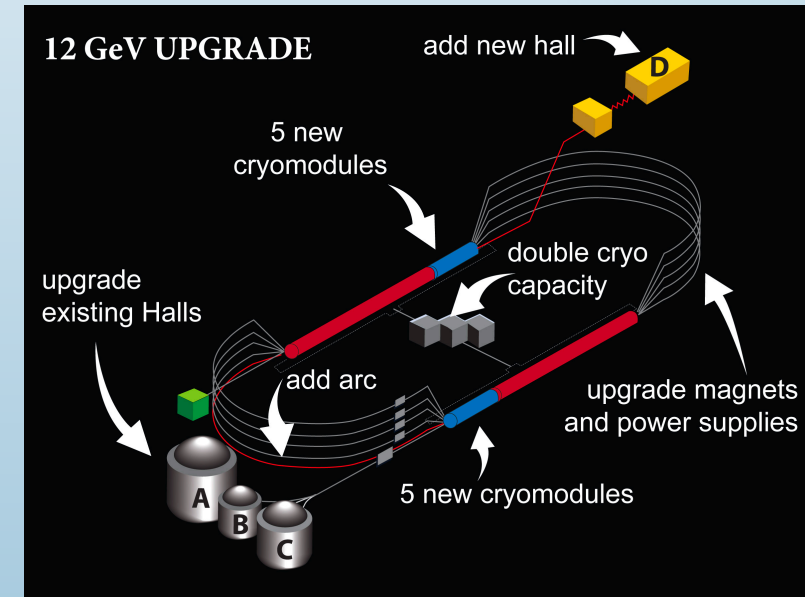
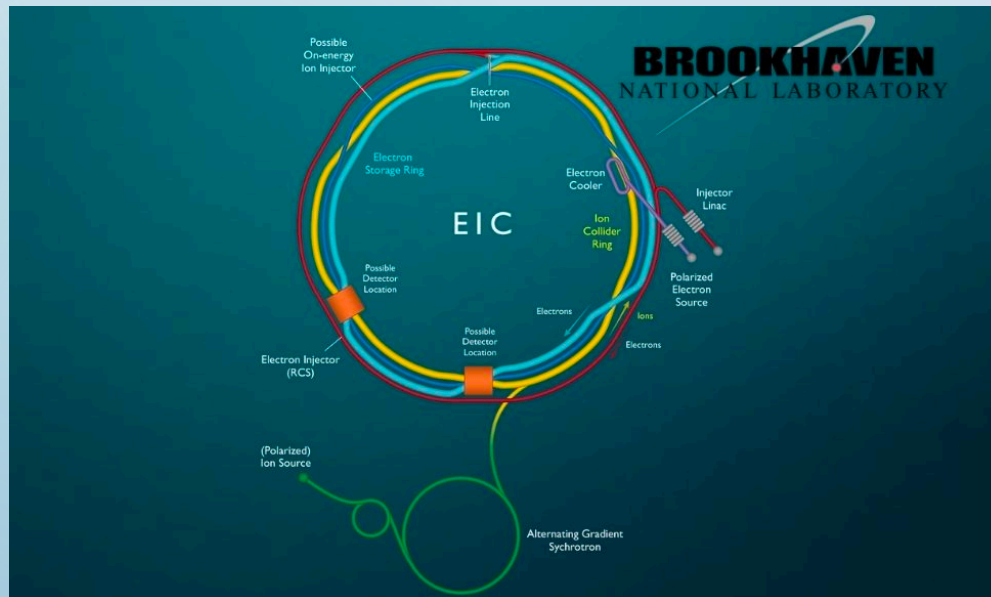
C. Cocuzza *et al.*, Phys. Rev. D. **104**, no. 7, 074031 (2021).

Outlook

Combine analysis with semi-inclusive DIS data from COMPASS.

Jefferson Lab 12 GeV upgrade: Precise high- x DIS data

EIC: First polarized electron-ion collider



Collaboration

Andreas Metz



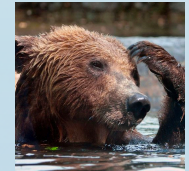
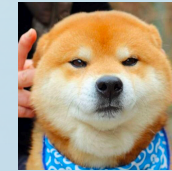
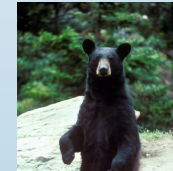
Wally Melnitchouk



Nobuo Sato



Thank you to Jacob Ethier, Yiyu Zhou, and Patrick Barry for helpful discussions



Backup

Bayes' Theorem

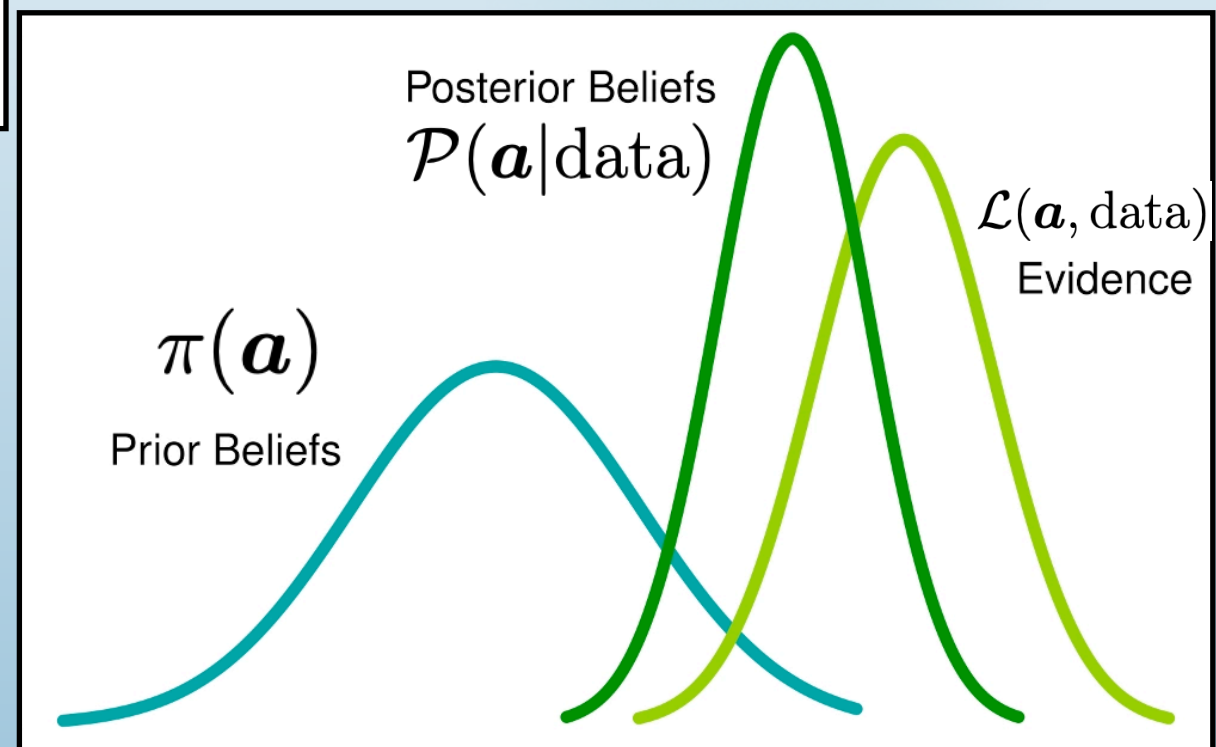
Now that we have calculated $\chi^2(\mathbf{a}, \text{data}) \dots$

Likelihood Function

$$\mathcal{L}(\mathbf{a}, \text{data}) = \exp \left(-\frac{1}{2} \chi^2(\mathbf{a}, \text{data}) \right)$$

Bayes' Theorem

$$\mathcal{P}(\mathbf{a}|\text{data}) \sim \mathcal{L}(\mathbf{a}, \text{data}) \pi(\mathbf{a})$$



Error Quantification

For a quantity $O(\mathbf{a})$: (for example, a PDF at a given value of (x, Q^2))

$$E[O] = \int d^n \mathbf{a} \, \rho(\mathbf{a} | \text{data}) \, O(\mathbf{a})$$

$$V[O] = \int d^n \mathbf{a} \, \rho(\mathbf{a} | \text{data}) \, [O(\mathbf{a}) - E[O]]^2$$

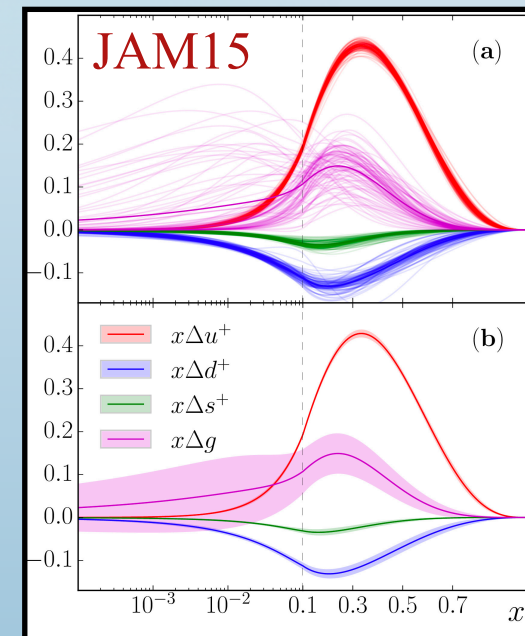
Build an MC ensemble

$$E[O] \approx \frac{1}{N} \sum_k O(\mathbf{a}_k)$$

$$V[O] \approx \frac{1}{N} \sum_k [O(\mathbf{a}_k) - E[O]]^2$$

Exact, but
 $n = \mathcal{O}(100)$!

Average over k sets
of the parameters
(replicas)



Data Resampling

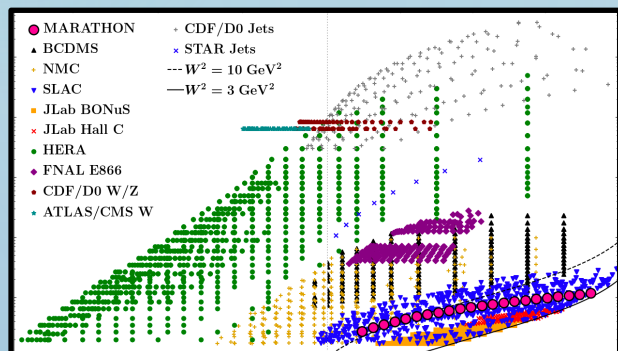
Pseudo-Data

$$\tilde{\sigma} = \sigma + N(0,1) \alpha$$

Uncorrelated
Uncertainties

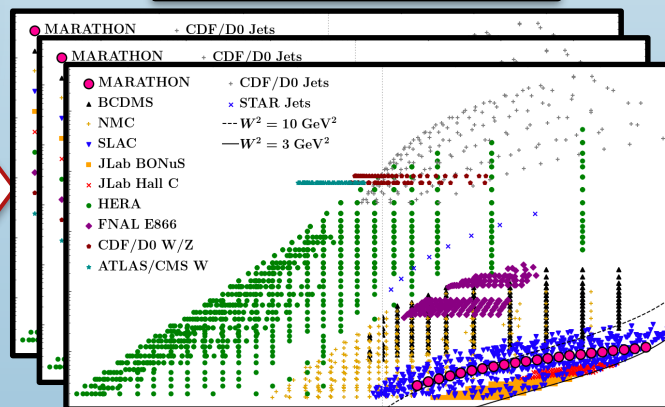
Data

Original Data



DR

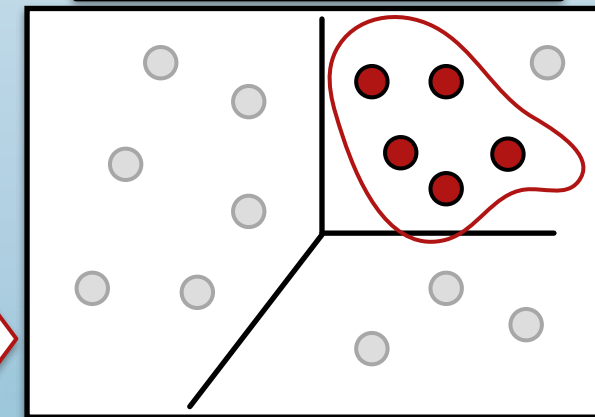
Replica Data


Maximum
Likelihood

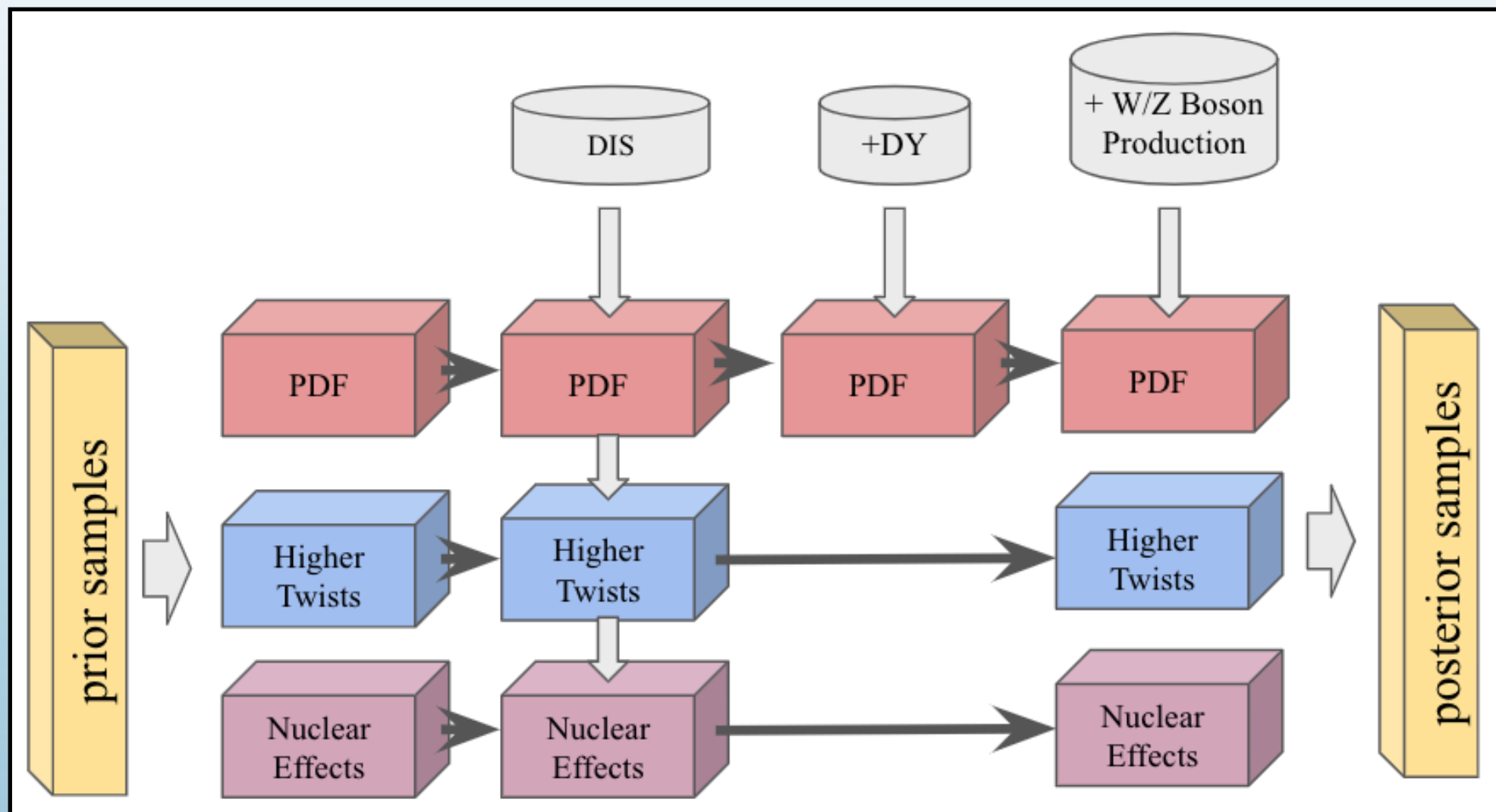
Maximum
Likelihood

Maximum
Likelihood

Parameter Space



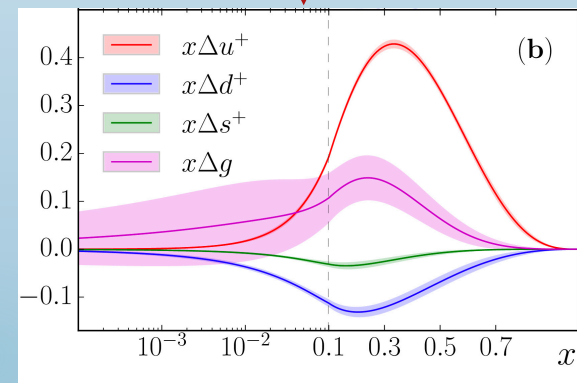
Multi-Step Strategy



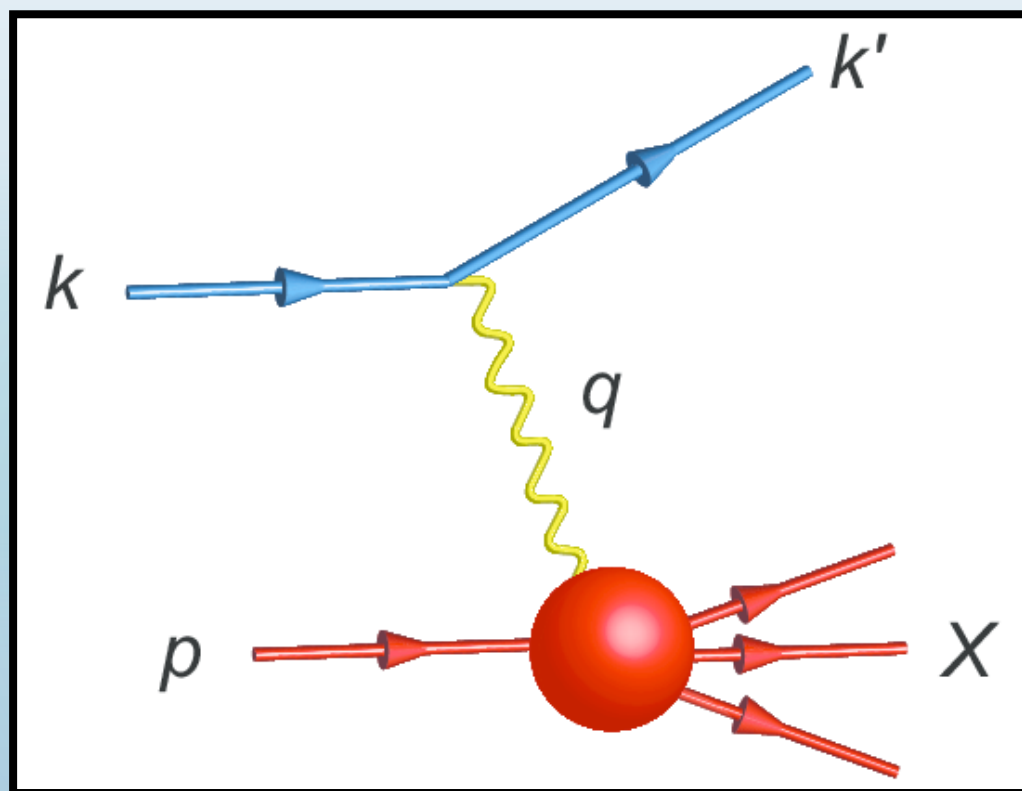
The flowchart illustrates the ensemble learning process for uncertainty quantification. It starts with 'data' (green rounded rectangle) and 'priors' (pink rounded rectangle). The 'data' is split into 'pseudo data₁', 'pseudo data₂', and 'pseudo data_K' (all green rounded rectangles). Each 'pseudo data_i' is further split into 'T data_i' (blue rounded rectangle) and 'V data_i' (pink rounded rectangle). 'T data_i' is used for fitting a model (fit_i, yellow oval) to produce a set of posterior probabilities $\{\bar{p}^{(j)}\}$ (blue rounded rectangle). 'V data_i' is used for validation (pink oval) to produce a set of predicted probabilities $\bar{a}^{(i)}$ (blue rounded rectangle). The 'priors' are used to initialize the process. The 'ensemble' (white rounded rectangle) is formed by combining the predicted probabilities $\bar{a}^{(1)}$, $\bar{a}^{(2)}$, and $\bar{a}^{(K)}$. The 'ensemble' produces 'new priors' (white rounded rectangle), which are then used to update the 'priors' (pink rounded rectangle) for the next iteration. A red curved arrow points from the 'ensemble' to the 'new priors'.







Deep Inelastic Scattering



Virtuality:

$$Q^2 = -q^2$$

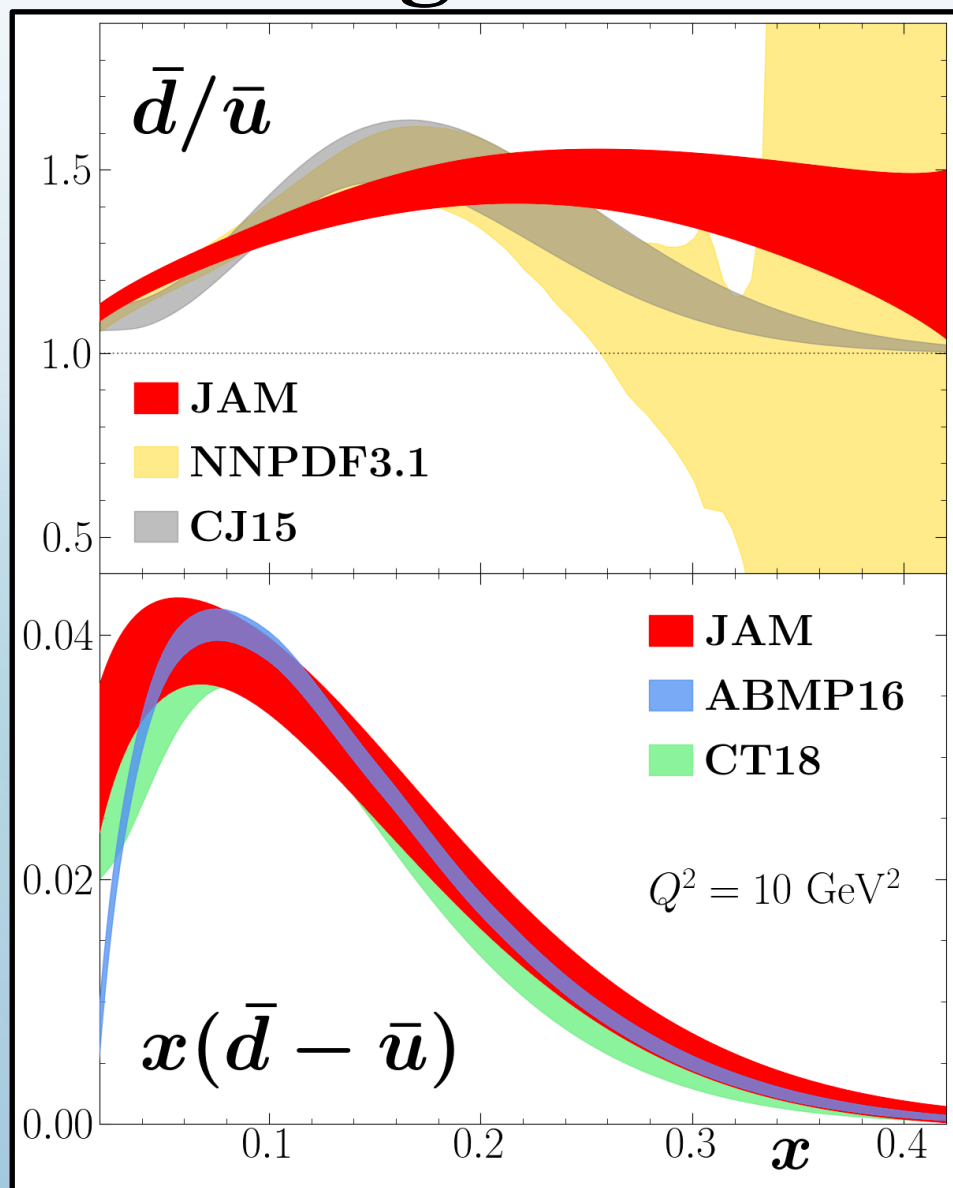
Bjorken x :

$$x = \frac{Q^2}{2p \cdot q}$$

Invariant mass of
outgoing particles:

$$W^2 = (p + q)^2$$

Resulting PDFs



Results for asymmetry largely agree with
 ABMP16, CT18;
 disagree with NNPDF3.1, CJ15 at high x .

