Learning Proton Structure with Gaussian Processes

Joe Karpie (A206)

Jefferson Lab

Overview

- Ioffe time Distributions define Generalized Parton Distributions and Double Distributions

- They cannot be directly calculated on the lattice, but there are many methods to get by that

- Modeling lattice QCD data with polynomials for Mellin moments and with Gaussian Processes Regression for Double Distributions

- Results for the H and E GPD on from one ensemble (single lattice spacing, pion mass, and volume)

Two Faced Distribution

GPDs: x, ξ

Radon Transform ⟷

DDs: β, α

$$f(x,\xi) = \int_{-1}^{1} d\beta \int_{-1+|\beta|}^{1-|\beta|} d\alpha\, \delta(x-\beta-\xi\alpha)\tilde{f}(\alpha,\beta)$$

- Interpretation: average/change in parton momentum fraction

- Mellin moments give Form Factors and Angular Momentum decomposition

- Complex interrelation of variables
  - ERBL/DGLAP regions and polynomiality

Statue of Janus Bifrons (Wikipedia)

- Interpretation: Hybridize PDFs/DAs

- β acts like PDF x
- α acts like DA x

- GPD evolution and polynomiality arise naturally from parameterized DD

Generalized Parton Distributions

- Generalized Ioffe time distributions (GITD)

“Ioffe time distributions instead of parton momentum distributions in description of DIS”
V. Braun, P. Gornicki, L. Mankiewicz
Phys Rev D 51 (1995) 6036-6051

$$I^\mu(p',p,z=z^-,\mu^2) = \frac{1}{2P^\mu}\langle p' | \bar{q}\big(-\frac{z^-}{2}\big)\gamma^\mu W\big(-\frac{z^-}{2},\frac{z^-}{2}\big) q\big(\frac{z^-}{2}\big) | p\rangle_{\mu^2}$$

Ioffe Time

$$\nu = \frac{p+p'}{2}\cdot z = P\cdot z$$

Momentum Transfer

squared

$$t = (p'-p)^2 = q^2$$

Skewness

$$\xi = \frac{q\cdot z}{P\cdot z}$$

- Generalized Parton Distributions

X. Ji PRL 78 (1997) 610-613

$$z_\mu I^\mu(\nu,t,\xi,\mu^2) = \int dx\, e^{i\nu x}\left[H(x,t,\xi)\,\bar{u}' z_\mu \gamma^\mu u + E(x,t,\xi)\,\bar{u}' \frac{i z_\mu \sigma^{\mu\nu} q_\nu}{2m} u\right]$$

Taylor expand $e^{i\nu x}$ to see GITD as series of Mellin moments of GPD

Double Distributions

- Generalized Ioffe time distributions (GITD)

“Ioffe time distributions instead of parton momentum distributions in description of DIS”
V. Braun, P. Gornicki, L. Mankiewicz
Phys Rev D 51 (1995) 6036-6051

$$I^\mu(p',p,z=z^-,\mu^2) = \frac{1}{2P^\mu}\langle p' | \bar{q}\Big(-\frac{z^-}{2}\Big)\gamma^\mu W\Big(-\frac{z^-}{2},\frac{z^-}{2}\Big)q\Big(\frac{z^-}{2}\Big)|p\rangle_{\mu^2}$$

Ioffe Time

$$\nu = \frac{p+p'}{2}\cdot z = P\cdot z$$

Momentum Transfer

squared

$$t = (p'-p)^2 = q^2$$

Skewness

$$\xi = \frac{q\cdot z}{P\cdot z}$$

- Double Distributions

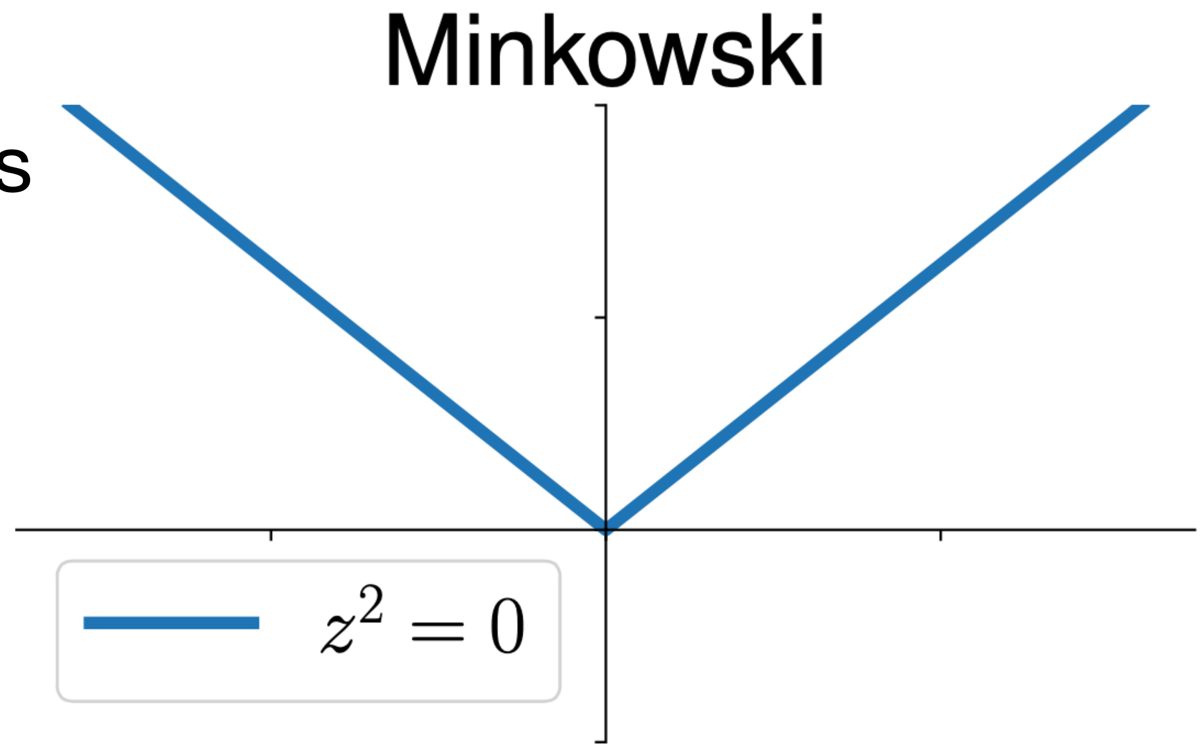
$$z_\mu I^\mu(\nu,t,\xi,\mu^2) = \int d\beta d\alpha e^{i(\nu\beta+\xi\nu\alpha)}\Bigg[h(\alpha,\beta,t)\,\bar{u}' z_\mu\gamma^\mu u + e(\alpha,\beta,t)\,\bar{u}'\frac{i z_\mu\sigma^{\mu\nu}q_\nu}{2m}u + \delta(\beta)D(\alpha,t)\,\xi\nu\,\bar{u}'u\Bigg]$$

β is conjugate to $\nu = P\cdot z$ and α is conjugate to $\xi\nu = q\cdot z$

A.V. Radyushkin Phys Lett B380 (1996) 417-425 M.V. Polyakov and C. Weiss PRD 60, 114017 (1999)

Parton Distributions and the Lattice

- Parton Distributions are defined by operators with light-like separations



- Use space-like separations

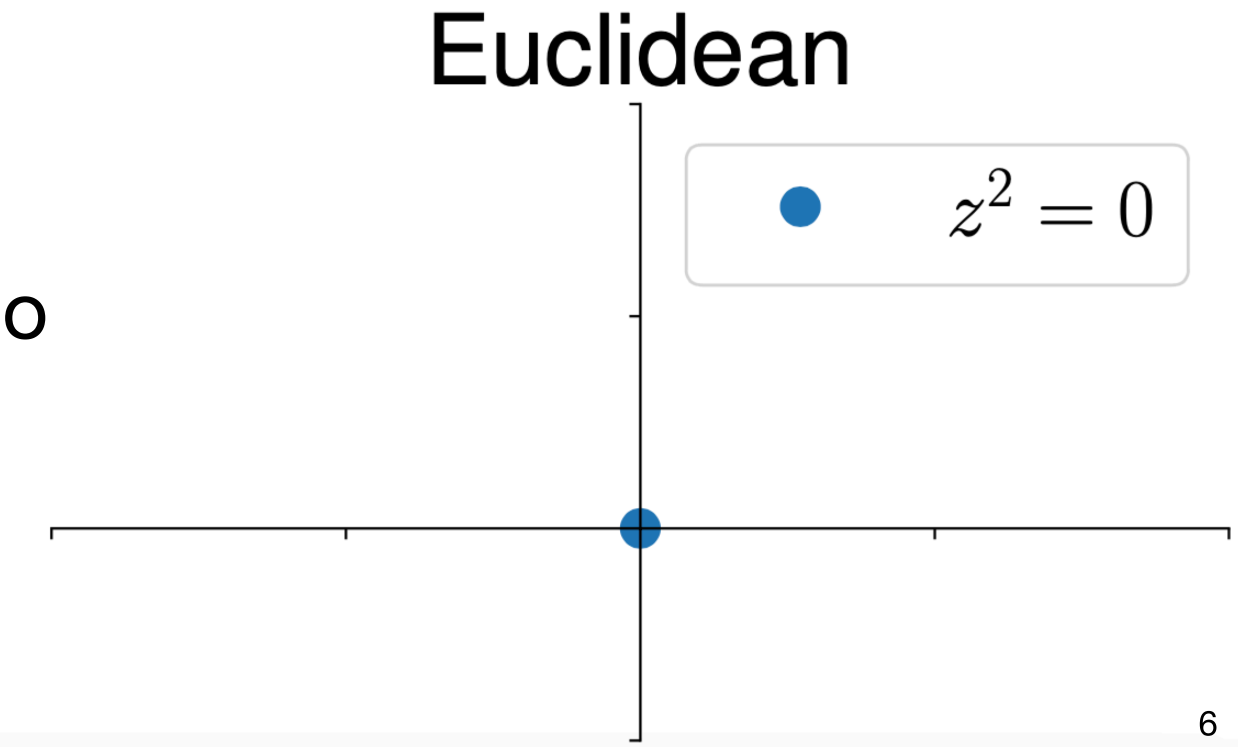
X. Ji *Phys Rev Lett* 110 (2013) 262002

  - Wilson line operators

$$O_\Gamma^{\mathrm{WL}}(z) = \bar{\psi}(z)\Gamma W(z;0)\psi(0)$$

$$z^2 \neq 0$$

  - Factorizations exist analogous to cross sections

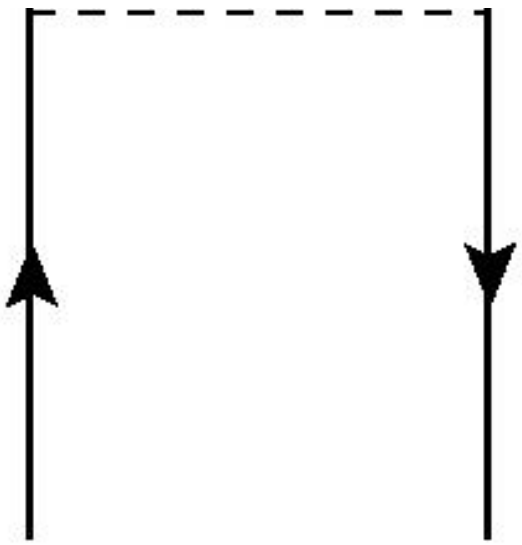


Many Good Lattice Cross Sections

GLCS are calculable matrix elements which factorize to light cone distributions

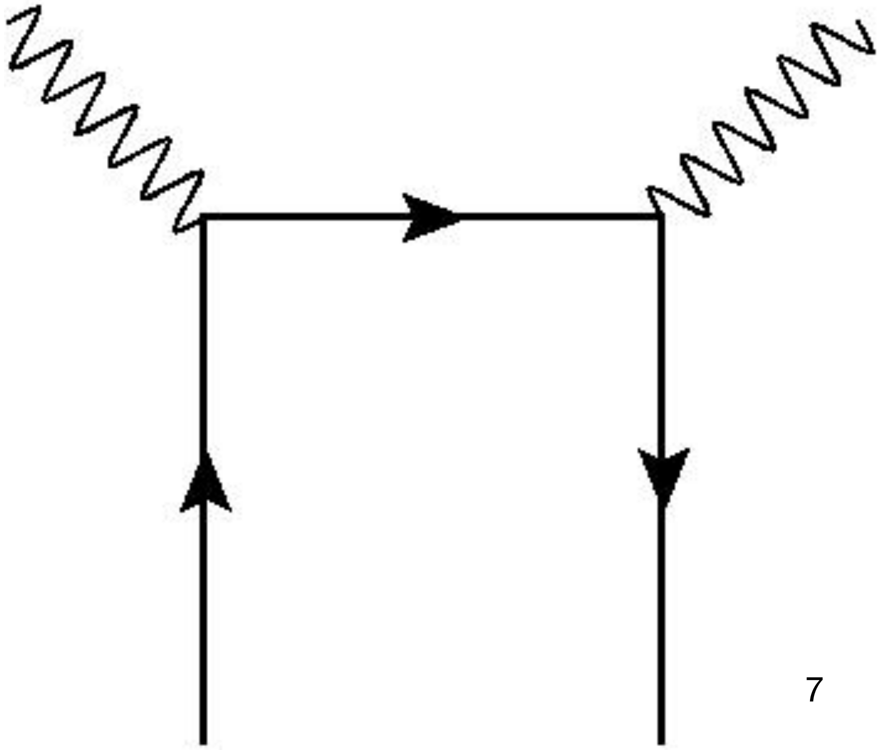
- **Non-local bilinears operators**

$$O_{WL}(x;z) = \bar{\psi}(x+z)\Gamma W(x+z;x)\psi(x)$$



  - LaMET/quasi-PDF X. Ji *Phys. Rev. Lett.* 110 (2013) 262002
  - **Pseudo-PDF** A. Radyushkin *Phys. Rev. D* 96 (2017) 3, 034025
  - Coulomb Gauge Fixed Wilson Line-less
    X. Gao et al *Phys. Rev. D* 109 (2024) 9, 094506
- Two current correlators

$$O_{CC}(x,y) = J_{\Gamma}(x)J_{\Gamma'}(y)$$



  - Hadronic Tensor
    K.-F. Liu et al *Phys. Rev. Lett.* 72 1790 (1994)
    Phys. Rev. D 62 (2000) 074501
  - HOPE
    W. Detmold and C.-J. D. Lin, *Phys. Rev. D* 73 (2006) 014501
  - Short distance OPE
    V. Braun and D. Muller *Eur. Phys. J. C* 55 (2008) 349
    Y.-Q. Ma and J.-W. Qiu *Phys. Rev. Lett.* 120 (2018) 2, 022003
  - OPE-without-OPE
    A. Chambers et al, *Phys. Rev. Lett.* 118 (2017) 242001

The Role of Separation and Momentum

- In Structure Functions, quasi-PDF, and pseudo-PDF, variables have common roles

Scale:

$Q^2 \,/\, p_z^2 \,/\, z^2$

- Scale for factorization to light cone
- Scale in power expansion
- Keep away from Λ_{QCD}^2
- Technically only requires single value, use many to study systematics

Dynamical variable:

$x_B \,/\, z \,/\, p_z$, or $\nu = p \cdot z$

- Variable describes non-perturbative dynamics
- Can take large or small value
- Want as many as are available
- Wider range improves the inverse problem

There are matrix elements that are interesting

How do we learn them from the lattice

Lattice QCD systematics

All systematics are improvable, but at what cost?

- Finite lattice spacing $a \sim 0.045 - 0.1$ fm

Polynomial of a to model

- Finite volumes $L \sim 3 - 5$ fm and $m_\pi L \sim 4 - 6$

Single hadron: Exponential decay in $m_\pi L$ to model

Multi-hadron: Polynomial in L^{-1} which Luscher method removes

- Heavy quarks / pions $m_\pi \sim 140 - 600$ MeV

Chiral PT gives polynomials and logs of m_π to model

- Excited state control $\Delta \sim 140 - 500$ MeV

Use larger T and do better fits

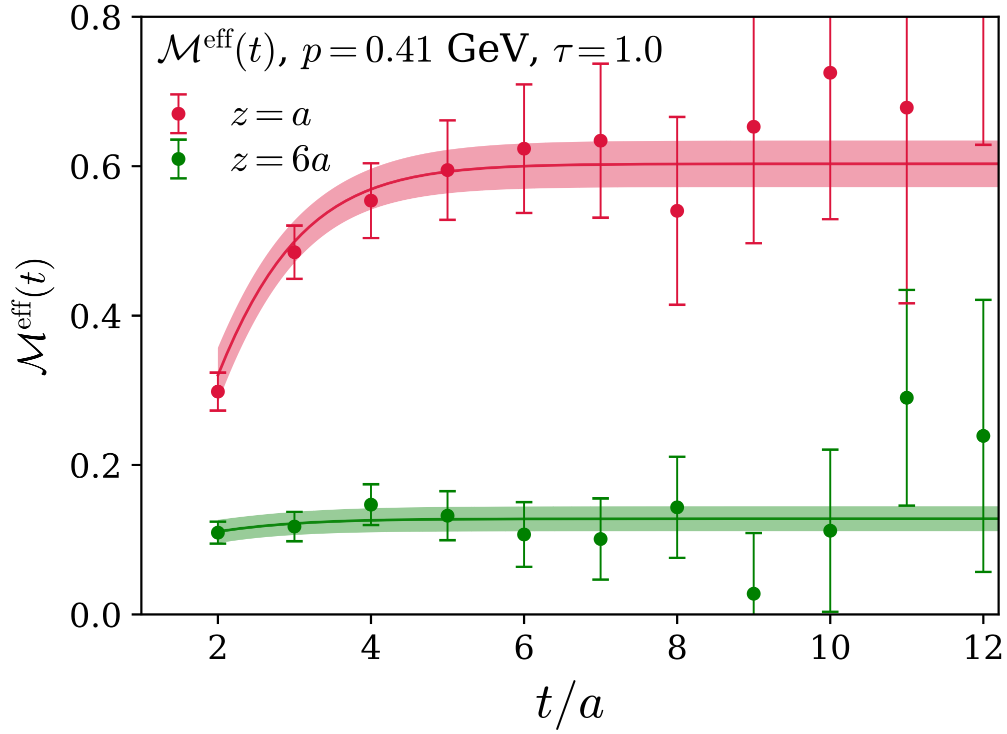
Variational can separate lowest states

- Statistics Always there

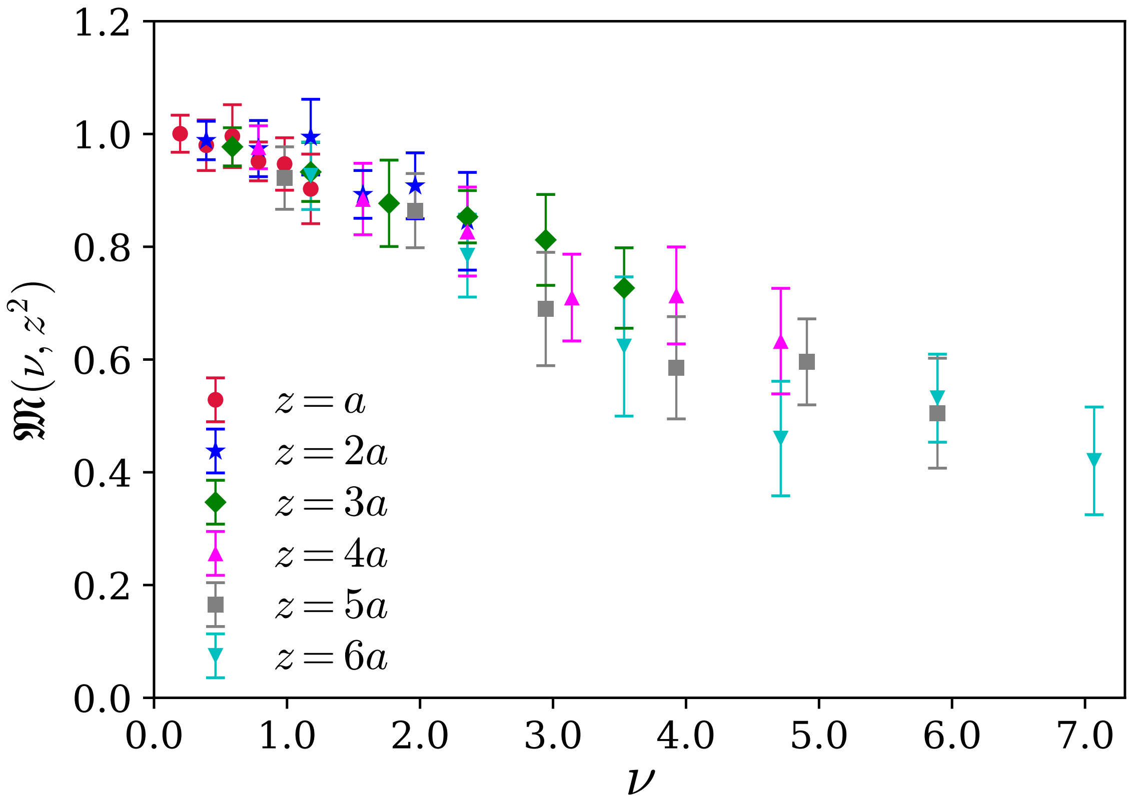
Beg the DOE for bigger computer

From Lattice QCD to PDFs

Lattice Correlation Functions



Hadron Matrix Elements



- Correlators (vacuum expectation values of time separated operators) are described as sums over an exponential for each energy eigenstate.

- Coefficients are matrix elements and exponential rates are energy levels

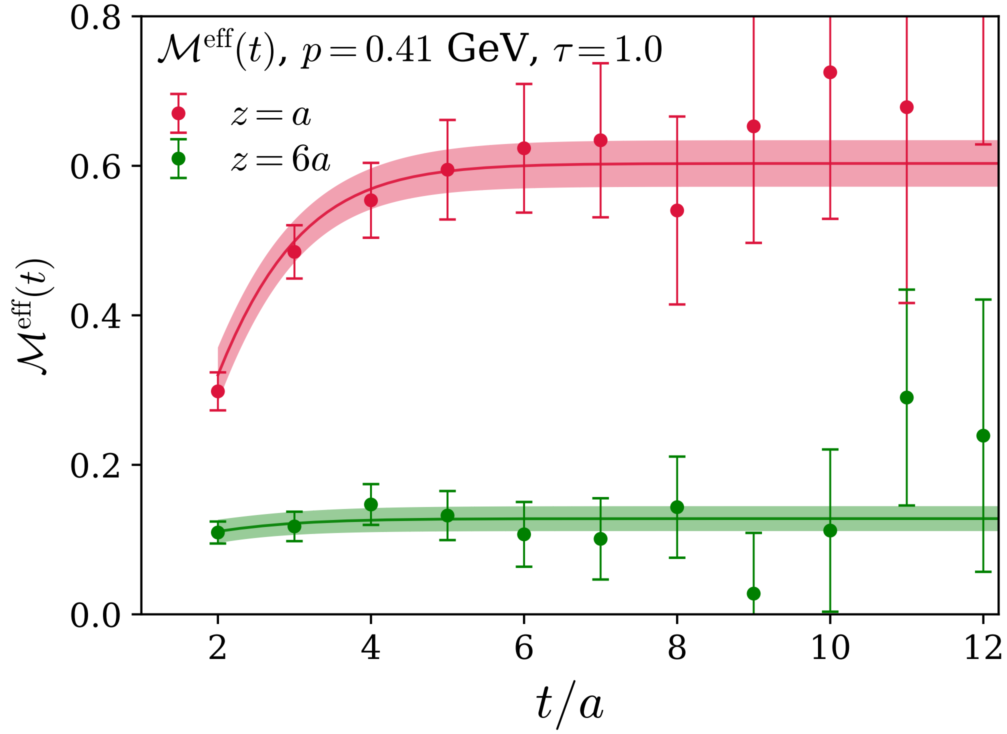
- Model and/or remove subdominant states by using large time but noise grows exponentially

Unpolarized Gluon PDF

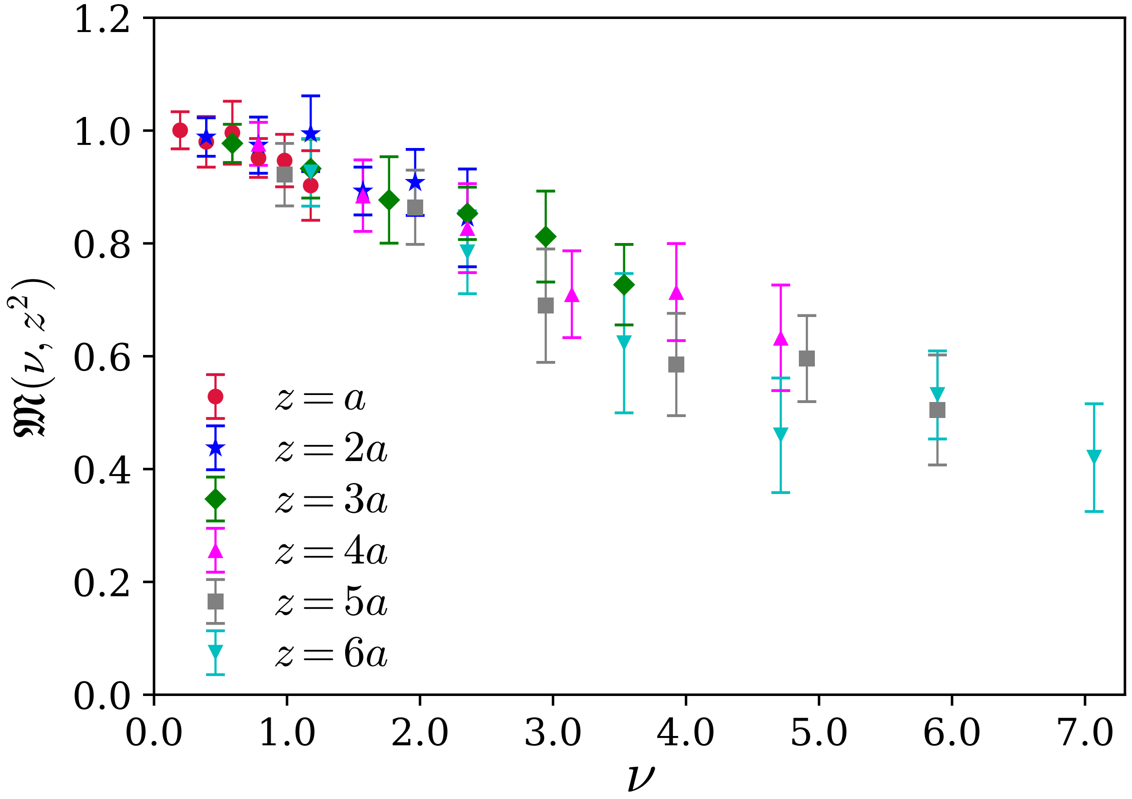
T. Khan, R. Sufian, JK, C. Monahan, C. Egerer, B. Joo, W. Morris, K. Orginos, A. Radyushkin, D. Richards, E. Romero, S. Zafeiropoulos
PRD 104 (2021) 9, 094516

From Lattice QCD to PDFs

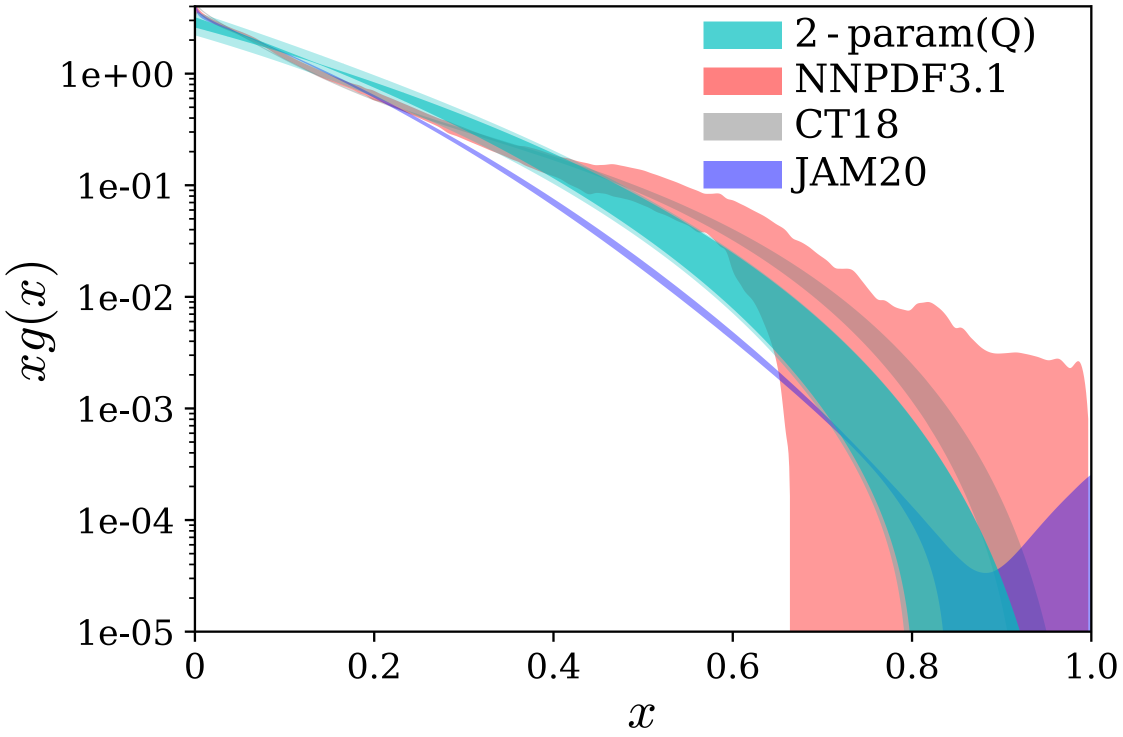
Lattice Correlation Functions



Hadron Matrix Elements



Parton Distributions



- Incomplete information gives integral inverse problem

$$M(\nu) = \int dx C(x\nu) xg(x)$$

$$xg(x) = x^a(1-x)^b / B(a+1, b+1)$$

- To more accurately infer PDF, we need larger ν

Unpolarized Gluon PDF

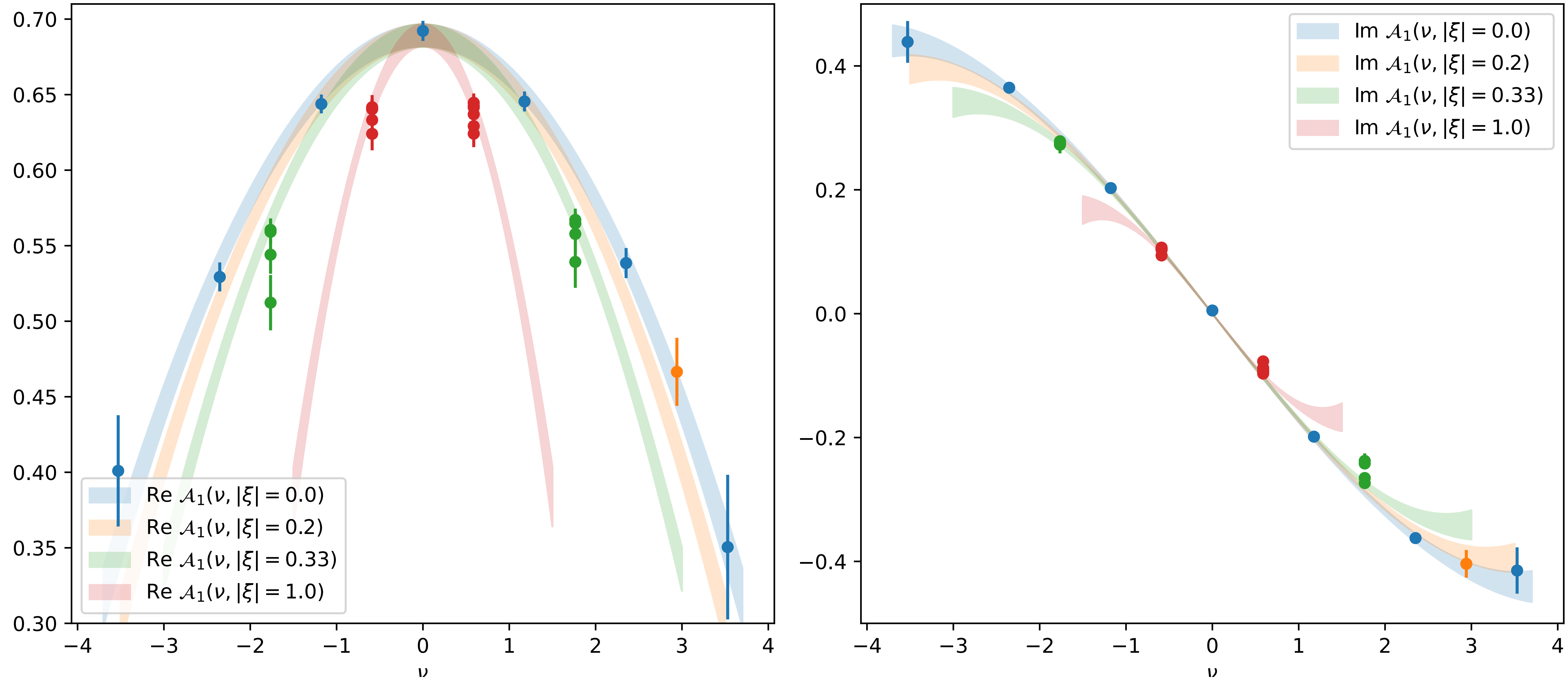
T. Khan, R. Sufian, JK, C. Monahan, C. Egerer, B. Joo, W. Morris, K. Orginos, A. Radyushkin, D. Richards, E. Romero, S. Zafeiropoulos PRD 104 (2021) 9, 094516

V. Braun and D. Muller *Eur. Phys. J. C* 55 (2008) 349
JK et al JHEP 11 (2018) 178
HadStruc arXiv:2405.10304

Fits to Moments

Mellin moments form polynomials in $\nu, \xi\nu$

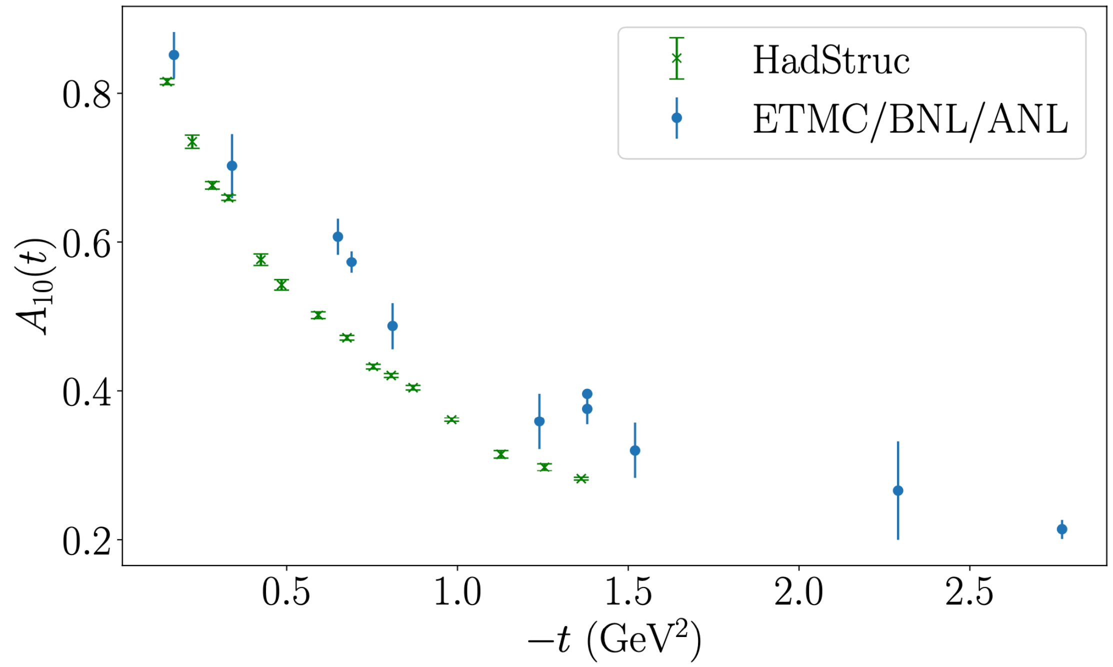
$$A_1(\nu, \xi\nu, t) = A_{10}(t) - i\nu A_{20}(t) - \frac{\nu^2}{2}\left(A_{30}(t) + \xi^2 A_{32}(t)\right) + i\frac{\nu^3}{6}\left(A_{40}(t) + \xi^2 A_{42}(t)\right) + \ldots$$

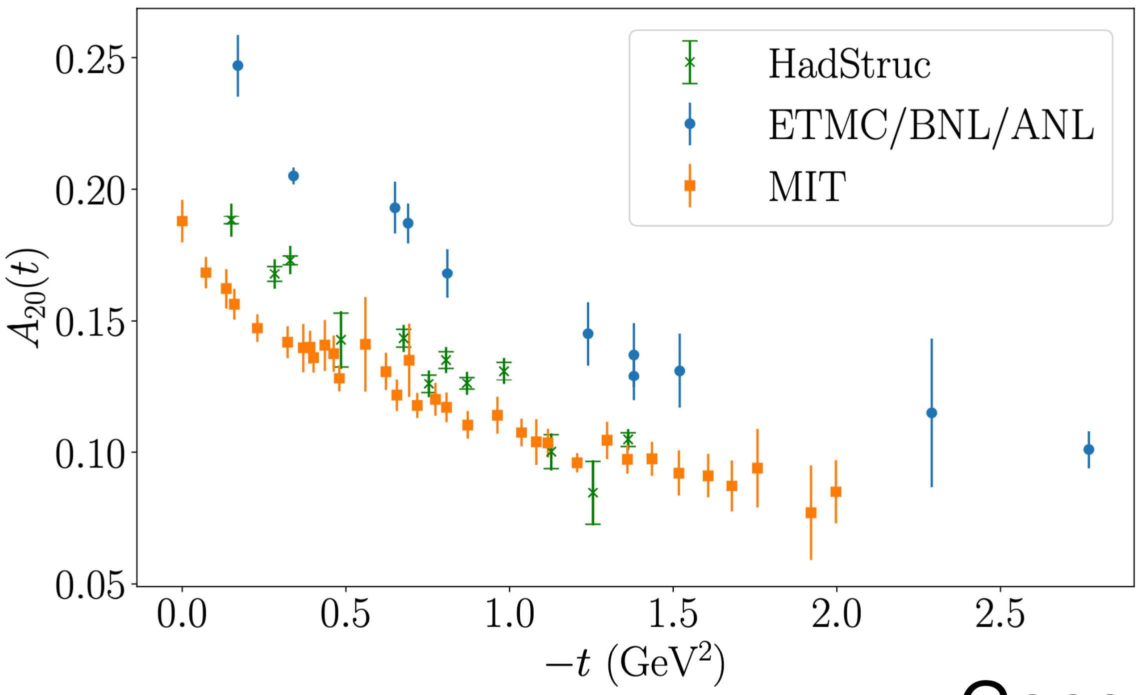


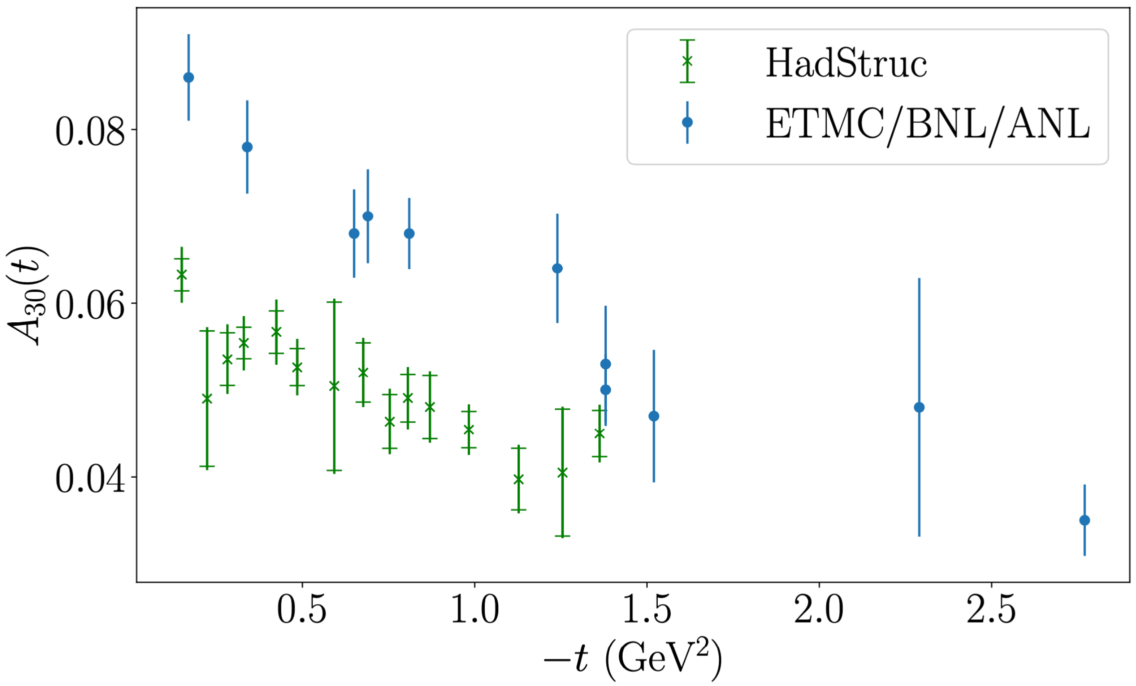
ID	a (fm)	m_π 3(MeV)	β	$m_\pi L$	$L^3 \times N_T$	N_{cfg}	N_{srcs}	rk ($\mathcal{D}$)
`a094m358`	0.094(1)	358(3)	6.3	5.4	$32^3 \times 64$	348 ([44])	4	64

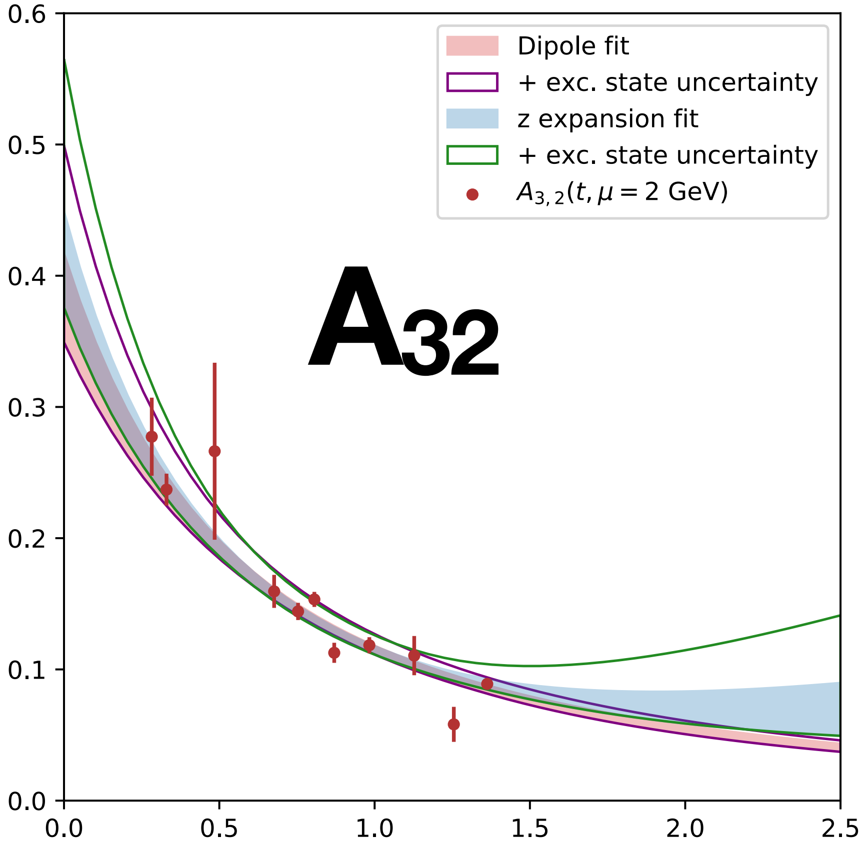
Moments of H

S. Bhattacharya et al (ETMC/BNL/ANL) arXiv:2305.11117
D. Hackett, D. Pefkou, P. Shanahan (MIT) arXiv:2301.08484
HadStruc arXiv:2405.10304









- General agreement with other calculations
- First access to skewness component of moments
- Model by dipole and z expansion

Summary of Dipole fits

$$F(t) = F(0)/(1 - t/m^2)^2$$

HadStruc arXiv:2405.10304

Value at t = 0

GPD H^{u-d}		GPD E^{u-d}	
$A_{1,0}$ 0.97(2)		$B_{1,0}$ 3.44(4)	
$A_{2,0}$ 0.204(4)		$B_{2,0}$ 0.36(2)	
$A_{3,0}$ 0.062(4)	$A_{3,2}$ 0.42(7)	$B_{3,0}$ 0.07(2)	$B_{3,2}$ 0.9(6)
$A_{4,0}$ 0.06(1)	$A_{4,2}$ 0.5(2)	$B_{4,0}$ 0.06(4)	$B_{4,2}$ 1.2(9)

Dipole mass (GeV)

GPD H^{u-d}		GPD E^{u-d}	
$A_{1,0}$ 1.25(2)		$B_{1,0}$ 0.982(6)	
$A_{2,0}$ 1.86(6)		$B_{2,0}$ 1.41(8)	
$A_{3,0}$ 2.2(4)	$A_{3,2}$ 1.07(9)	$B_{3,0}$ 2.4(9)	$B_{3,2}$ 1.0(3)
$A_{4,0}$ Unreliable	$A_{4,2}$ 1.2(2)	$B_{4,0}$ Unreliable	$B_{4,2}$ 1.1(2)

D-term^{u-d}

	Value at t = 0	Dipole mass (GeV)
C_2	0.025^{+8}_{-8}	> 2.2

We have nice matrix elements which can be fit with physics informed models.

How much do we really trust the model does not bias the result from the truth?

Approaches to inverse problem

- **Parametric**
  - Fit a phenomenologically motivated function
    - Method used by global fits
    - Potentially significant, but controllable model bias
  - Fit to a neural network

S. Forte, L. Garrido, J. Latorre, A. Piccione (2002) 0204232
K. Cichy, L. Del Debbio, T. Giani (2019) 1907.06037
L. Del Debbio, T. Giani, JK, K. Orginos, A. Radyushkin, S. Zafeiropoulos (2020) 2010.03996

- **Non-Parametric**

For NN/BG/MEM/BR: JK, K. Orginos, A. Rothkopf, S. Zafeiropoulos (2019) 1901.05408

  - Backus Gilbert

J. Liang, K-F. Liu, Y-B. Yang (2017) 1710.11145

  - Maximum Entropy Method / Bayesian Reconstruction

Y. Burnier and A. Rothkopf (2013) 1307.6106, J. Liang et al (2019) 1906.05312

  - Bayes-Gauss-Fourier transfrom

C. Alexandrou, G. Iannelli, K. Jansen, F. Manigrasso (2020) 2007.13800

  - **Gaussian Process Regression**

A. Candido, L. Del Debbio, T. Giani, G. Petrillo (2024) 2404.07573
H. Dutrieux, JK, K. Orginos, S. Zafeiropoulos (2024) 2412.05227

Gaussian Process Priors

ETMC PRD 102 (2020) 9, 094508
A. Candido et al Eur. Phys. J. C 84 (2024) 7, 716
H. Dutrieux et al PRD 111 (2025) 3, 034515

Bayes's Theorem

$$P[q|M,I] = \frac{P[M|q]\,P[q|I]}{P[M|I]}$$

- Given a theory: $M(\nu) = \int dx B(\nu, x) q(x)$
- Assume data-likelihood is Gaussian $P[M|q] \propto \exp[-\frac{1}{2}\chi^2]$
- Assume the PDF has a Gaussian prior with given mean $g(x)$ and covariance kernel $K(x,x')$

$$P[q|I] \propto \exp\left[-\frac{1}{2}\int dx dx' \big(q(x) - g(x)\big) K^{-1}(x,x') \big(q(x') - g(x')\big)\right]$$

- The posterior is Gaussian

$$P[q|M,I] \propto \exp\left[-\frac{1}{2}\int dx dx' \big(q(x) - \bar{q}(x)\big) H^{-1}(x,x') \big(q(x') - \bar{q}(x')\big) + \dots\right]$$

Final Mean

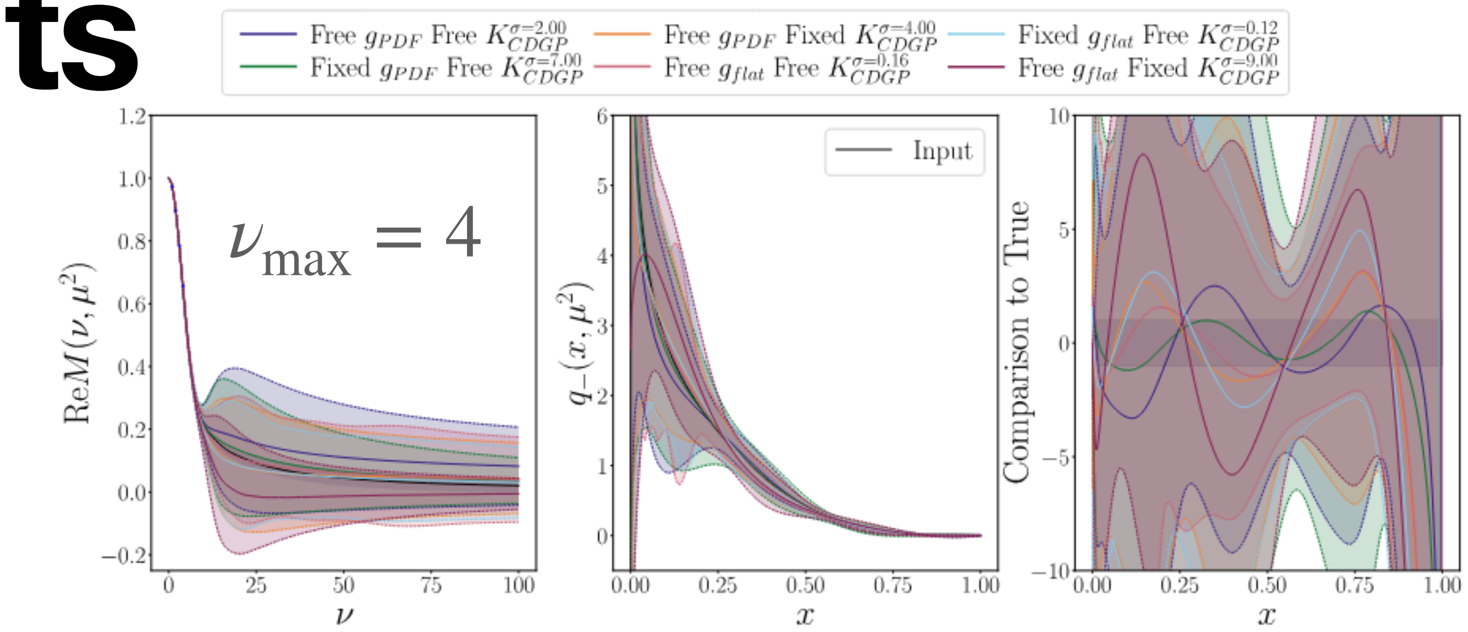
Final Covariance

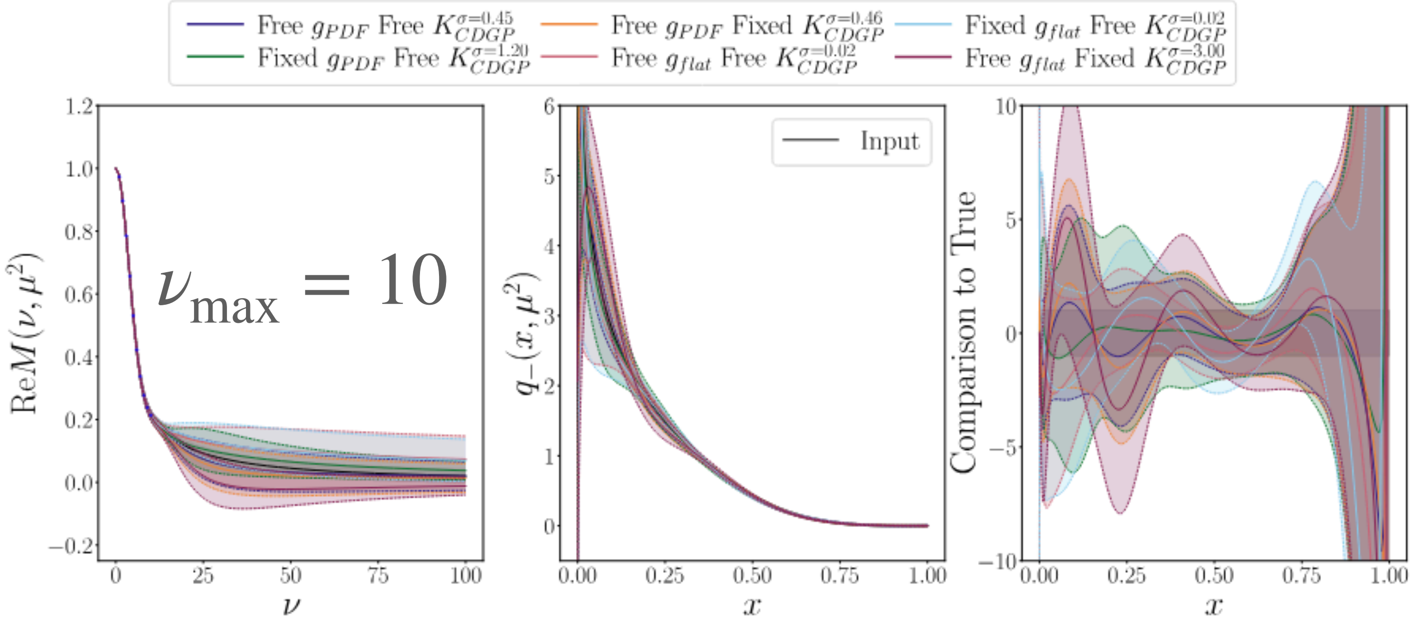
$q(x)$ independent terms

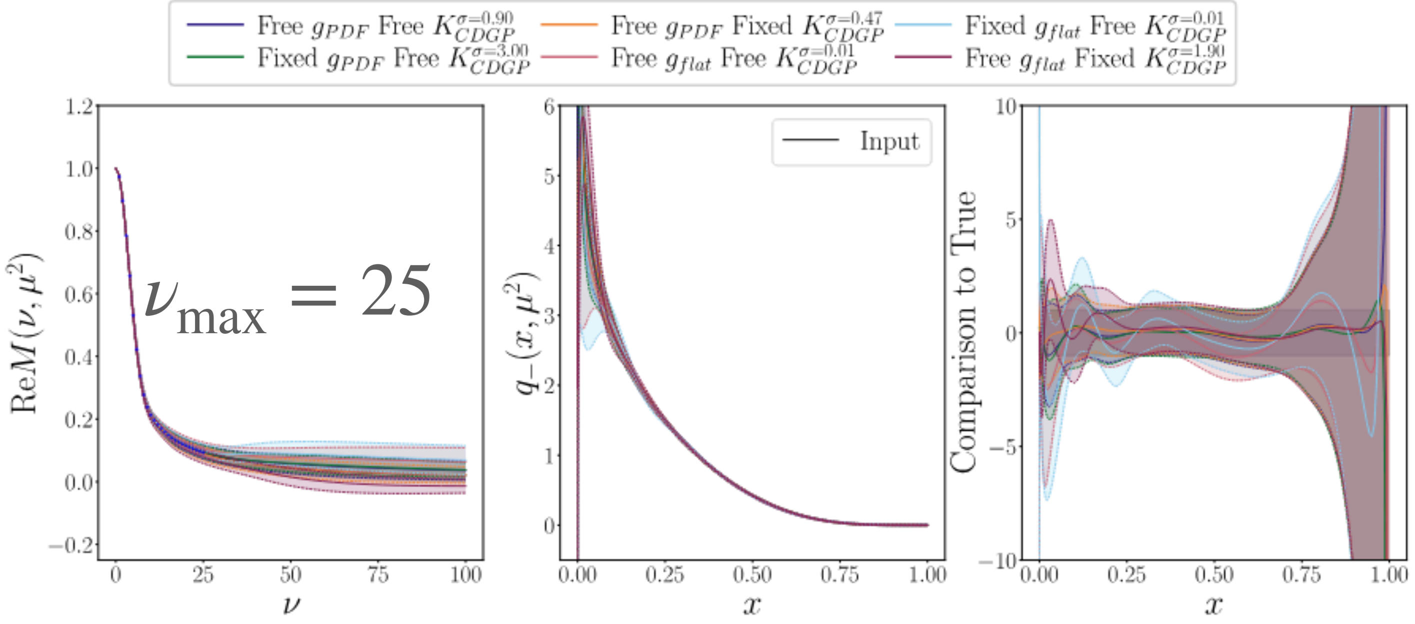
- $K(x,x')$ and $g(x)$ may have hyperparameters to tune or integrate over

Mock data results

- Many options of prior covariances/means and ways to treat their parameters
- Closure test to study the systematic improvement over different ν range
- All give highly consistent results in agreement with data
- Use information criteria to select models or take weighted average



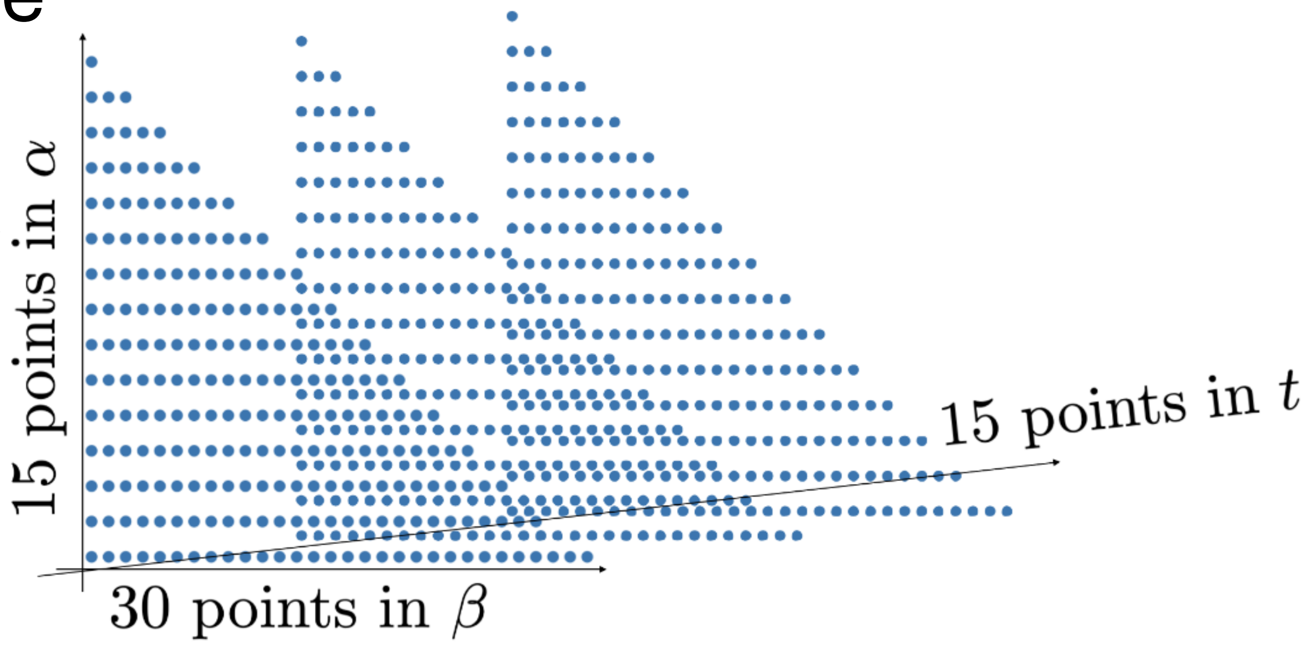




Y Cahuana Medrano, H Dutrieux, JK,
K Orginos, S Zafeiropoulos arXiv:2510.21041

Three Dimensional GPR

- CP Symmetry to bring DD into one quadrant
- Integrals approximated with Finite Element Method on grid of points interpolated by cubic splines
- Mean function is 0 and the 3D kernel to allow certain features
- Future work will study new kernels and model averaging like PDFs



Preliminary!

Converge as power of $(1-\beta)$ as $\beta \to 1$

$$
\begin{aligned}
K_{ij} &= K[(\beta_i,\alpha_i,t_i),(\beta_j,\alpha_j,t_j)]\,, \\
&= \sigma^2(\beta_i\beta_j)^{-\gamma}((1-\beta_i)(1-\beta_j))^{\delta}((1-\alpha_i-\beta_i)(1-\alpha_j-\beta_j))^{\eta} \\
&\quad \times \exp\left(-\frac{(\ln(\beta_i)-\ln(\beta_j))^2}{2l_\beta^2}\right)\exp\left(-\frac{(\beta_i\beta_j)^{-\gamma}(\alpha_i-\alpha_j)^2}{2l_\alpha^2}\right)\exp\left(-\frac{(t_i-t_j)^2}{2l_t^2}\right)
\end{aligned}
$$

Converge on boundary

Diverge as $\beta \to 0$

Decorrelate small and large β

Less correlation in α when β is small

Going beyond GPD moments requires larger momenta.

Sneak peak at upcoming paper

Kinematical coverage

249 initial and final momenta configurations

ID	a (fm)	m_π 3(MeV)	β	$m_\pi L$	$L^3 \times N_T$	$N_{\rm cfg}$	$N_{\rm srcs}$	rk $(\mathcal{D})$
`a094m358`	0.094(1)	358(3)	6.3	5.4	$32^3 \times 64$	348 ([44]) 1490 (new)	4	64

Ioffe Time

$$\nu = \frac{p + p'}{2} \cdot z = P \cdot z$$

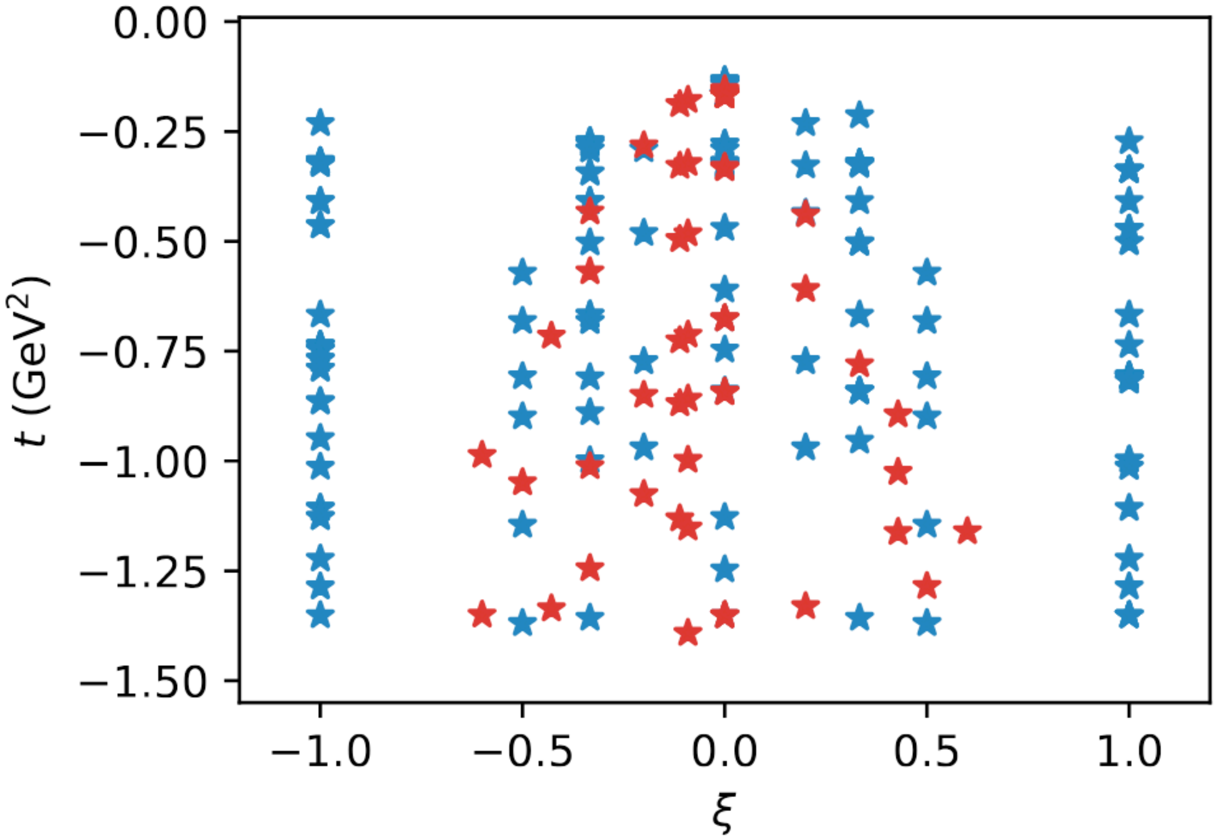
Momentum Transfer

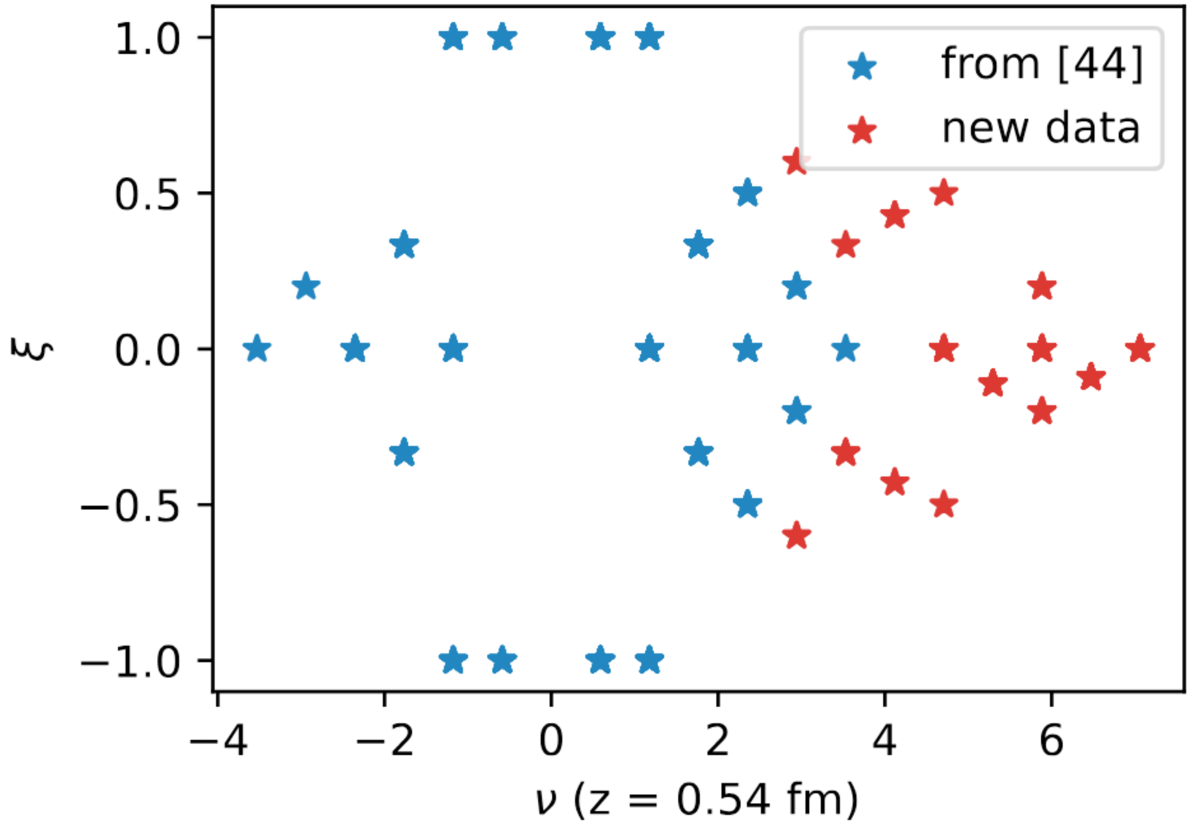
squared

$$t = (p' - p)^2 = q^2$$

Skewness

$$\xi = \frac{q \cdot z}{P \cdot z}$$

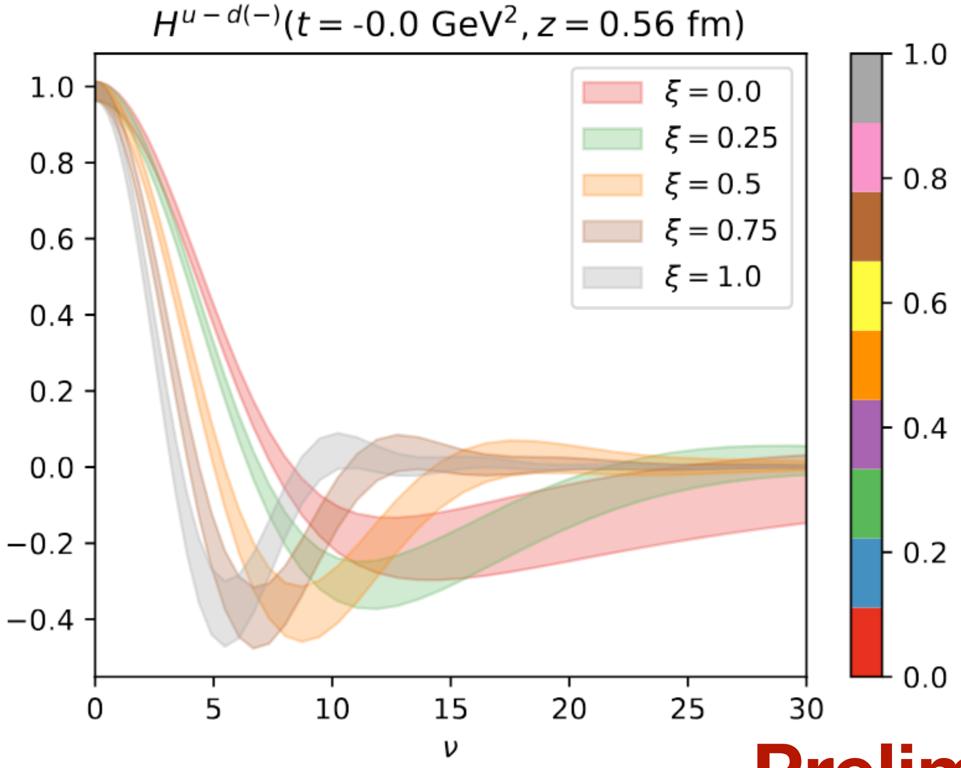


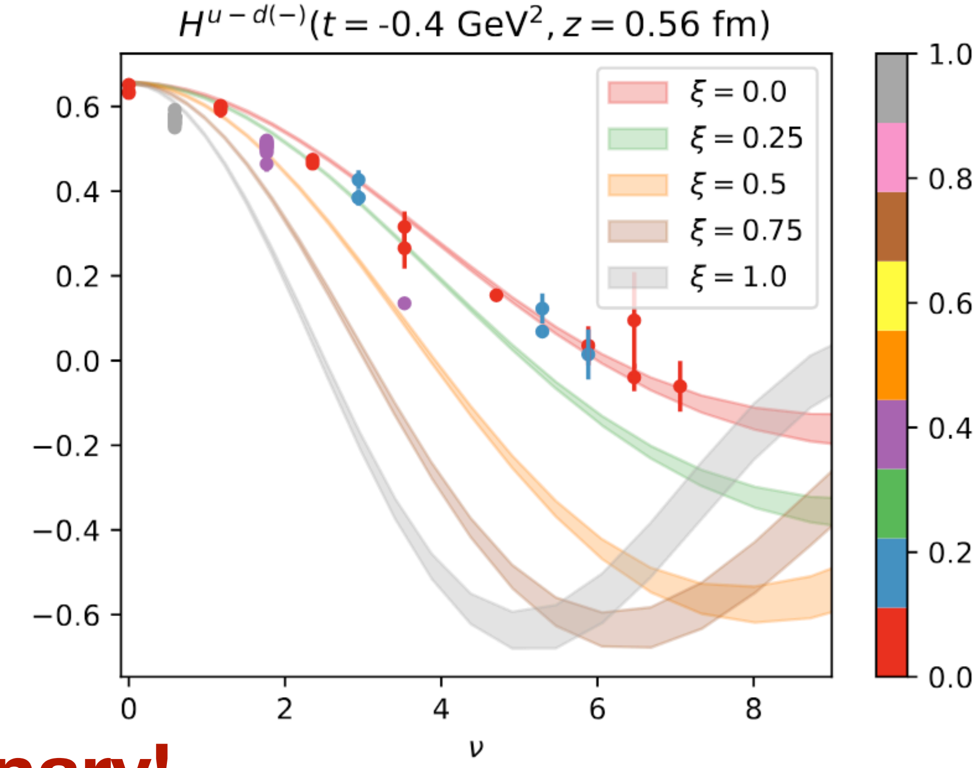


Ioffe time pseudo-distributions

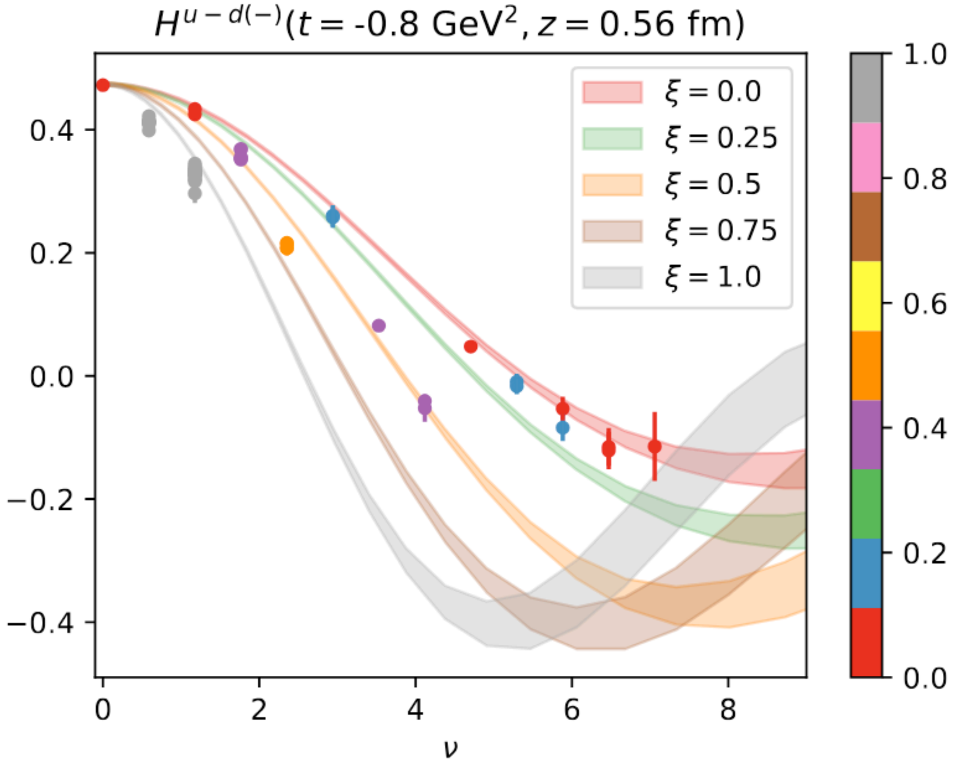
Iso-vector valence H pseudo-GITD

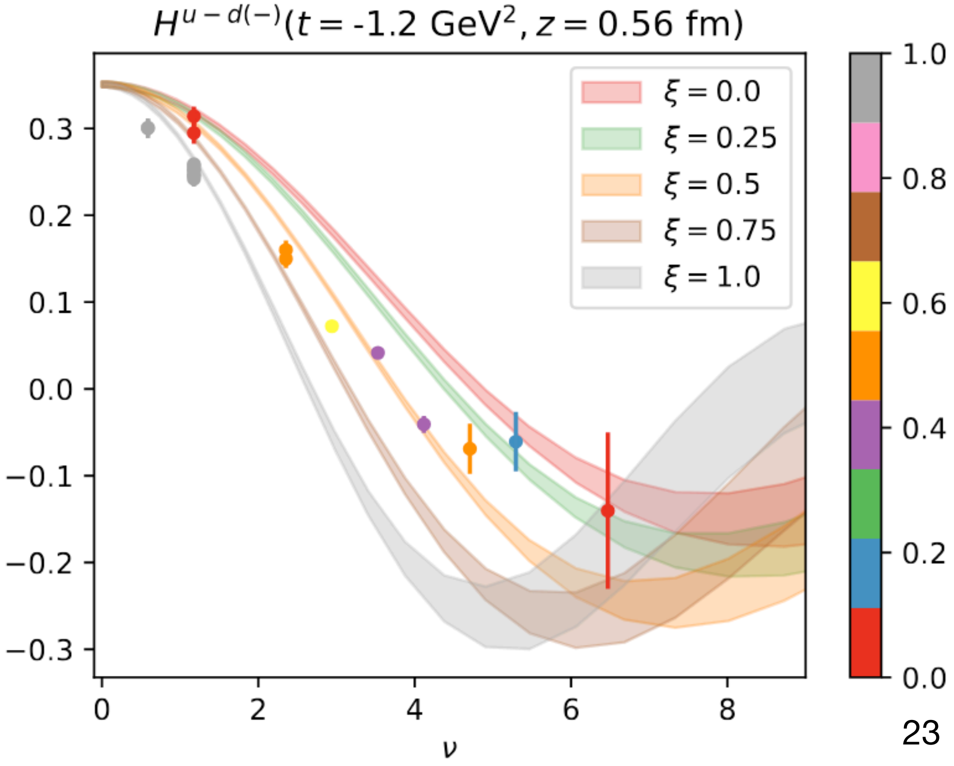
- Model DD and reconstruct GITD at wanted t, ξ
- Skewness dependence extrapolated to $t = 0$





Preliminary!

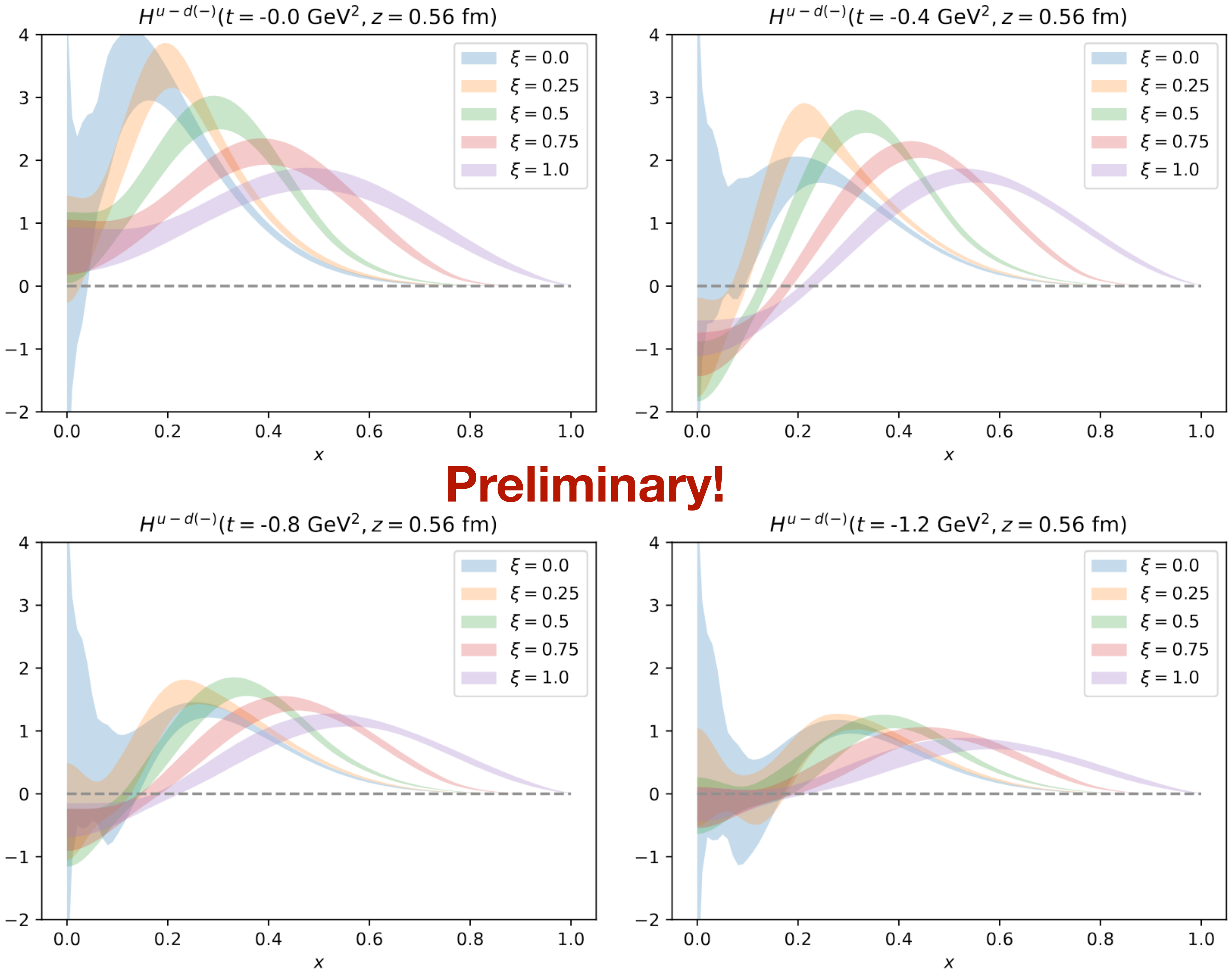




Generalized pseudo-distributions

Iso-vector valence H pseudo-GPD

- Model DD and reconstruct GPD at wanted t, ξ
- Skewness dependence extrapolated to $t = 0$
- Low β, α are not constrained by the data which leads to uncertainty at low x

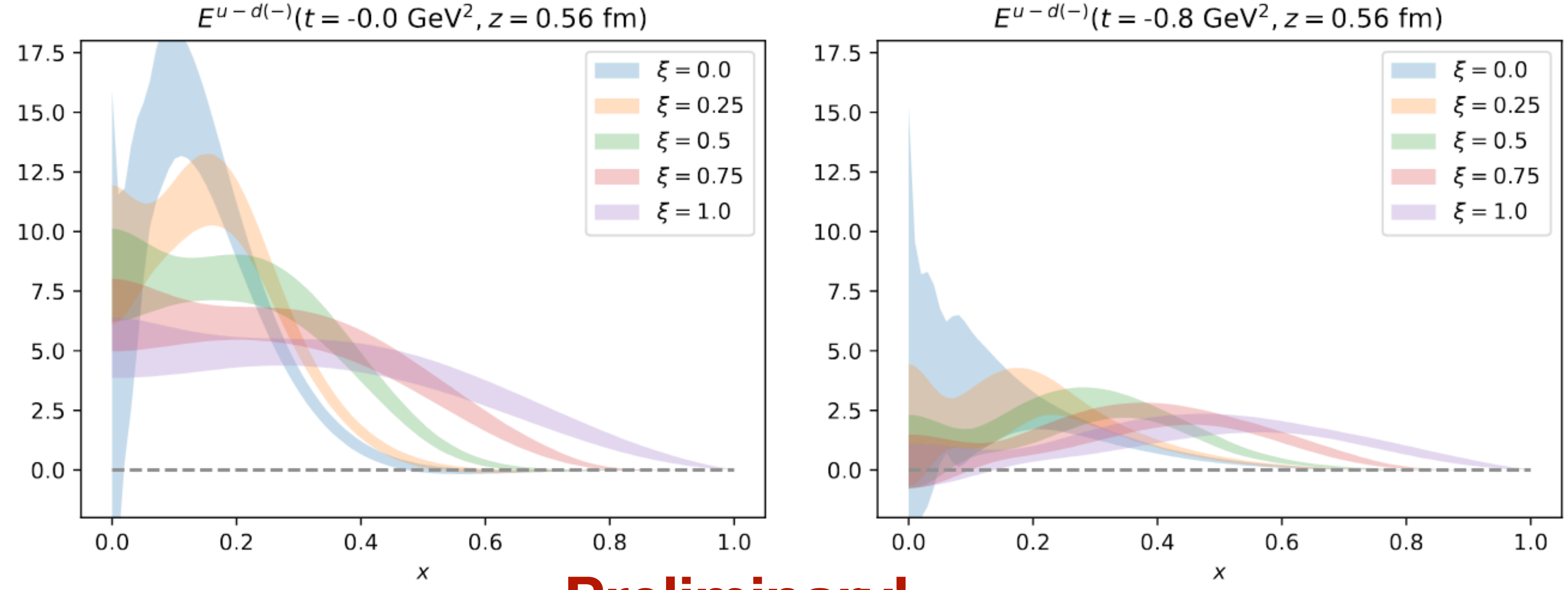


Preliminary!

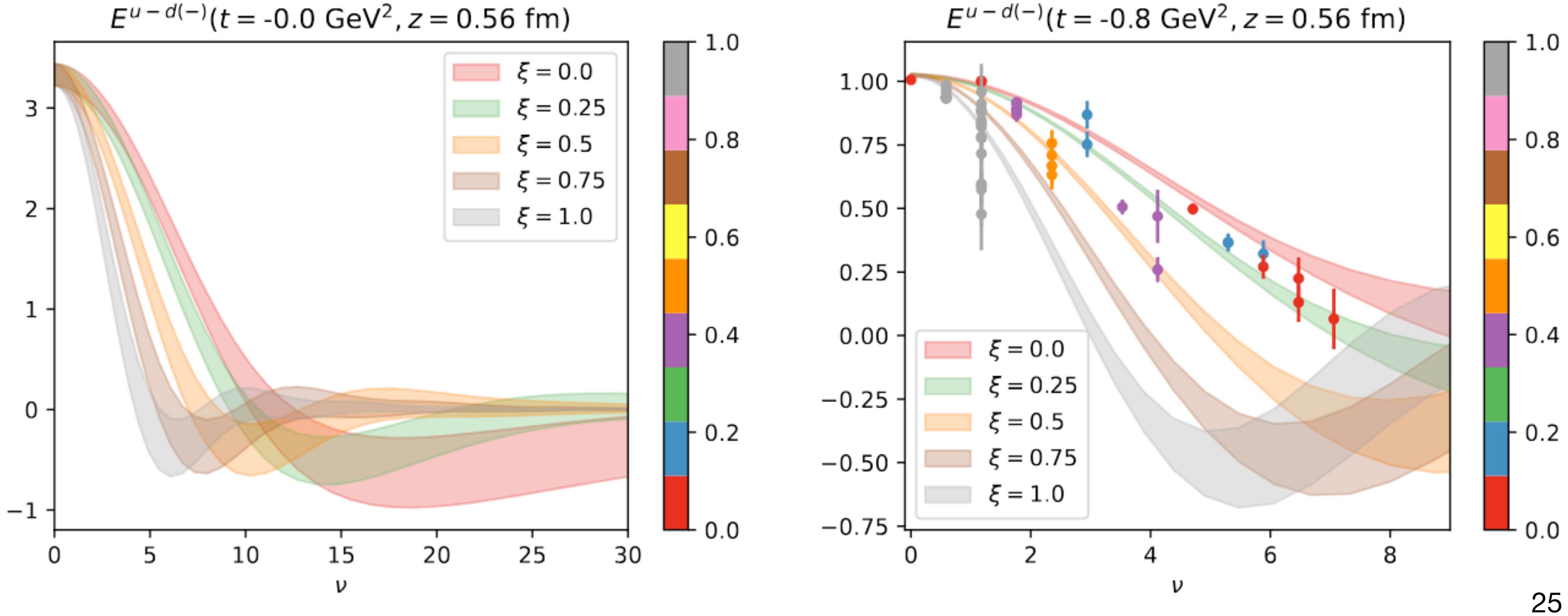
Ioffe time pseudo-distributions

Iso-vector valence E pseudo-GITD

- Model DD and reconstruct GITD/GPD at wanted t, ξ
- Skewness dependence extrapolated to $t = 0$
- Low β, α are not constrained by the data which leads to uncertainty at low x



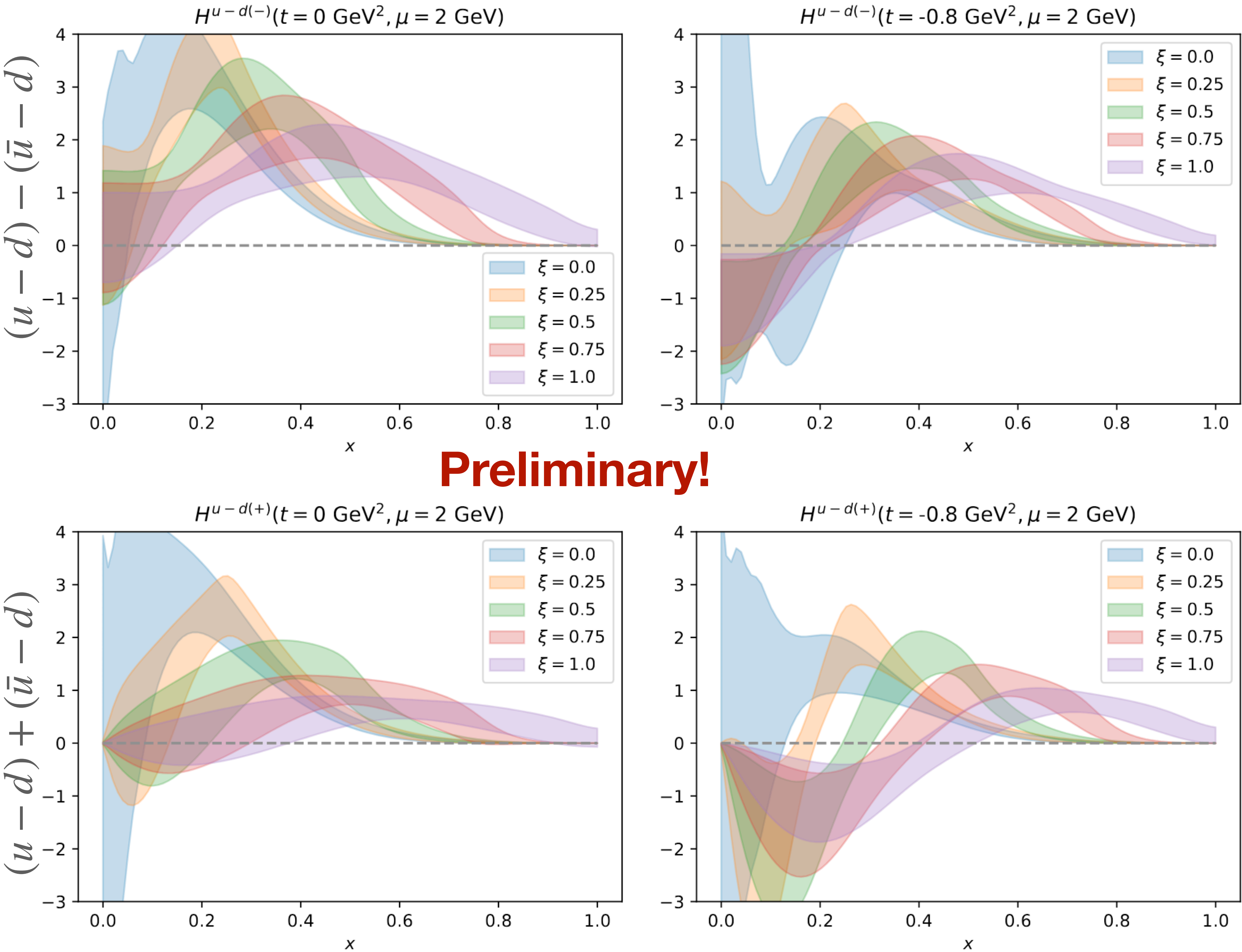
Preliminary!



Matching to the $\overline{\mathrm{MS}}$ scheme

LO (1 loop) Matching to H GPD

- Matched at $\mu_0^2 = 1/z^2$ then evolved to 2 GeV
- Uncertainty is statistical and excited state contamination
- Varying matching scale by $\sqrt{2}$ and changing hyperparameters were smaller errors

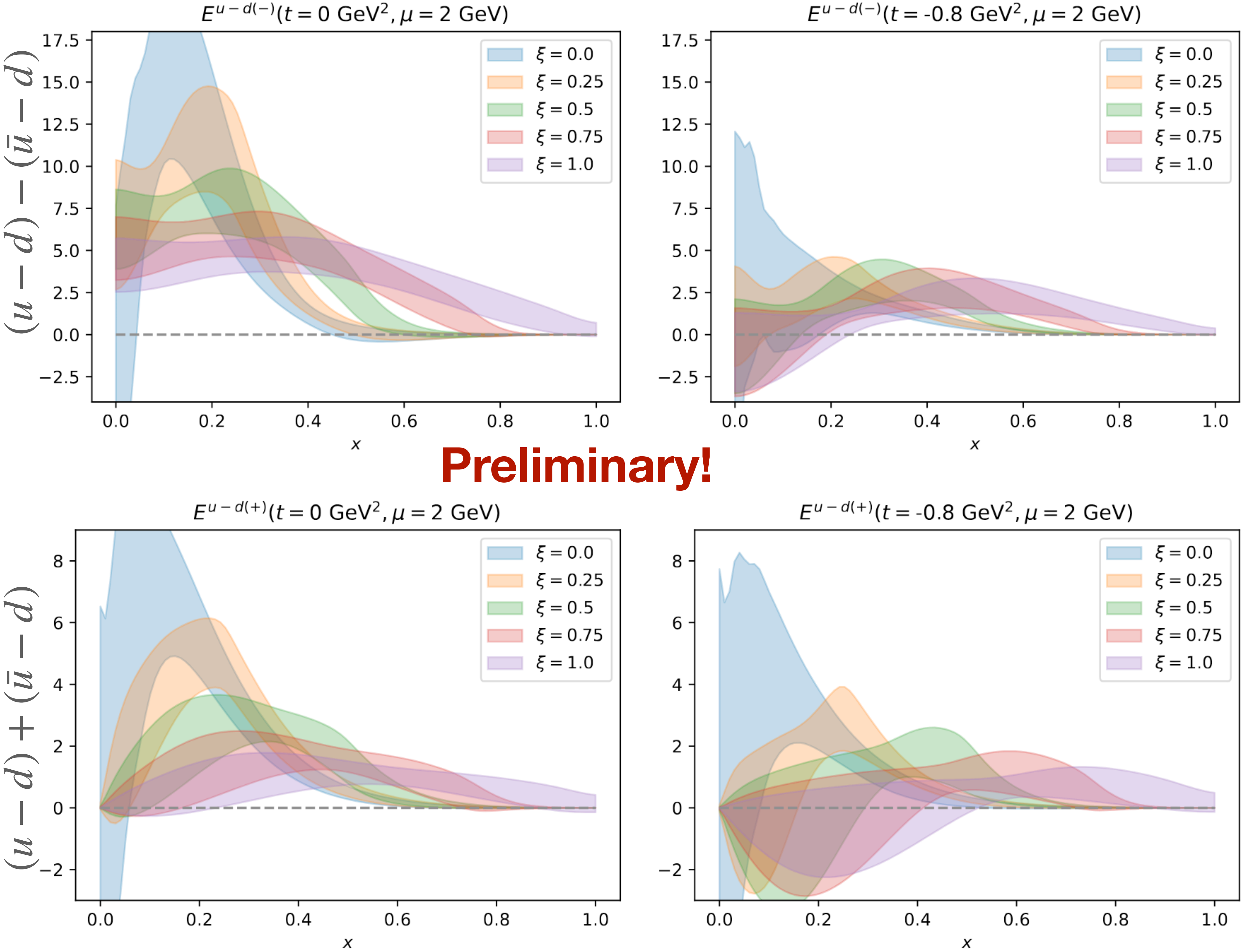


Preliminary!

Matching to the $\overline{\mathrm{MS}}$ scheme

LO (1 loop) Matching of E GPD

- Matched at $\mu_0^2 = 1/z^2$ then evolved to 2 GeV
- Uncertainty is statistical and excited state contamination
- Varying matching scale by $\sqrt{2}$ and changing hyperparameters were smaller errors



Preliminary!

Future Outlook

Calculation of GPDs full 3D structure from lattice QCD is possible!

Lattice Direction

- Lattice systematics of volume, lattice spacing, and pion mass must be varied
- Higher order perturbation theory and control over higher twist needs precision study
- Extension to gluon and singlet underway

Compare and Confront Phenomenology

- Pseudo-GPDs have finite power corrections allowing comparison at $x = \xi$
- Combined analysis may give information at ξ where lattice does not constrain and $x \neq \xi$ which many experiments do not constrain

Back up slides

Wait, didn’t you say Lattice have many errors to control?

How do we know when lattice data are ready?

Window Observables instead of x space comparison

- Lattice QCD data is more precise in the low ν (wide, slow Fourier modes) and lack any information at high ν (narrow, fast Fourier modes)
- The lattice results have this precision hiding in the covariance of the PDF

Window Moments

$$a_n(x_-, x_+) = \int_{x_-}^{x_+} dx\, x^n f(x)$$

``Truncated moments" ($x_+ = 1$)
have known evolution

Gaussian windows

$$g_n(x_-, x_+) = \int_{-1}^{1} dx \sqrt{\frac{n^2}{2\pi x_d^2}} \exp\left[-n^2 \frac{(x - x_0)^2}{2x_d^2}\right] f(x)$$

$$x_d = x_+ - x_- \qquad x_0 = \frac{x_+ + x_-}{2}$$

JK et al arXiv:2505.22795

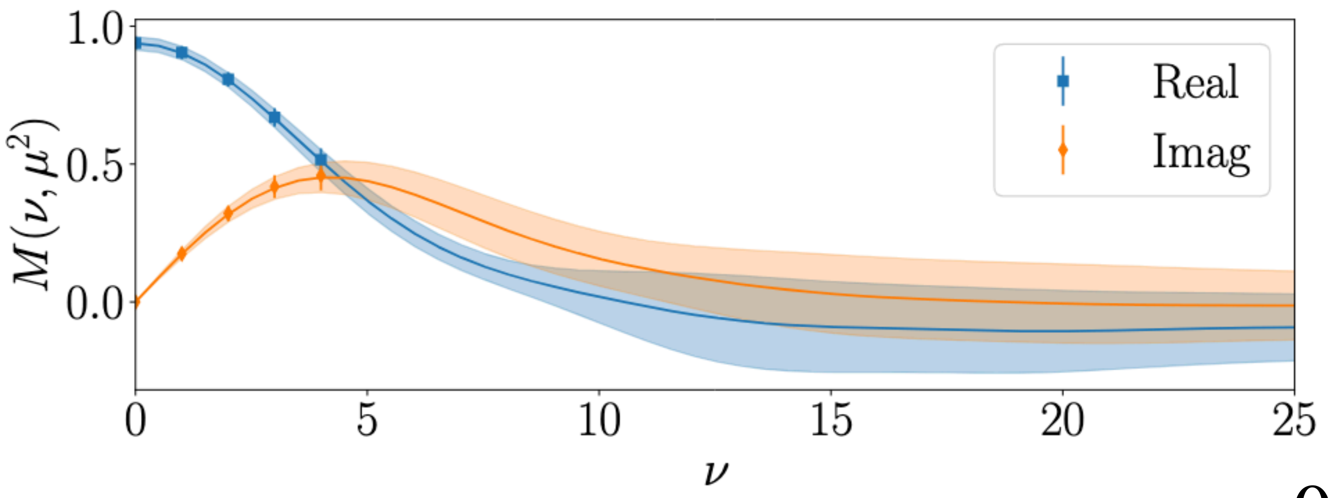
Window Observables instead of x space comparison

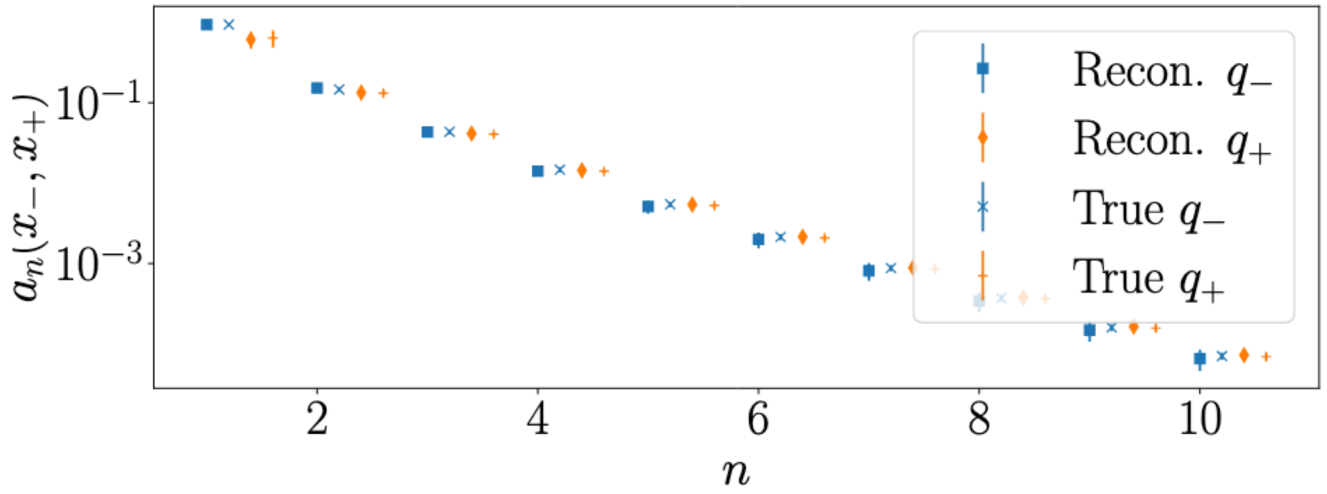
- GP reconstruction of mock data from JAM3D* transversity u-d PDFs

L. Gamberg et al arXiv:2205.00999
C. Cocuzza et al arXiv:2306.12998

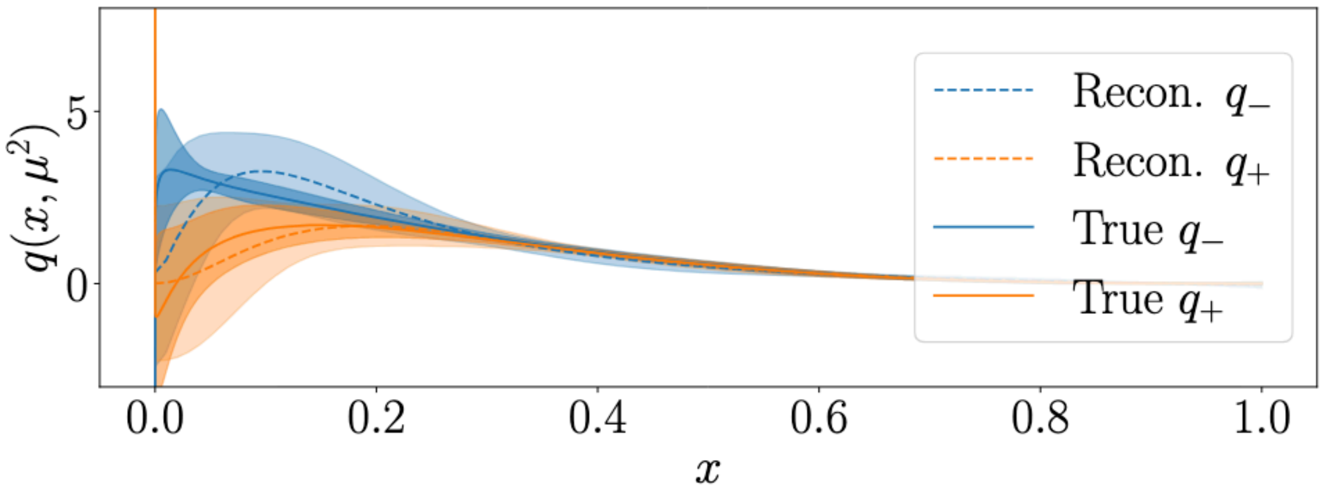
- Precision of the low n (widest windows) is hiding in the covariance of the data

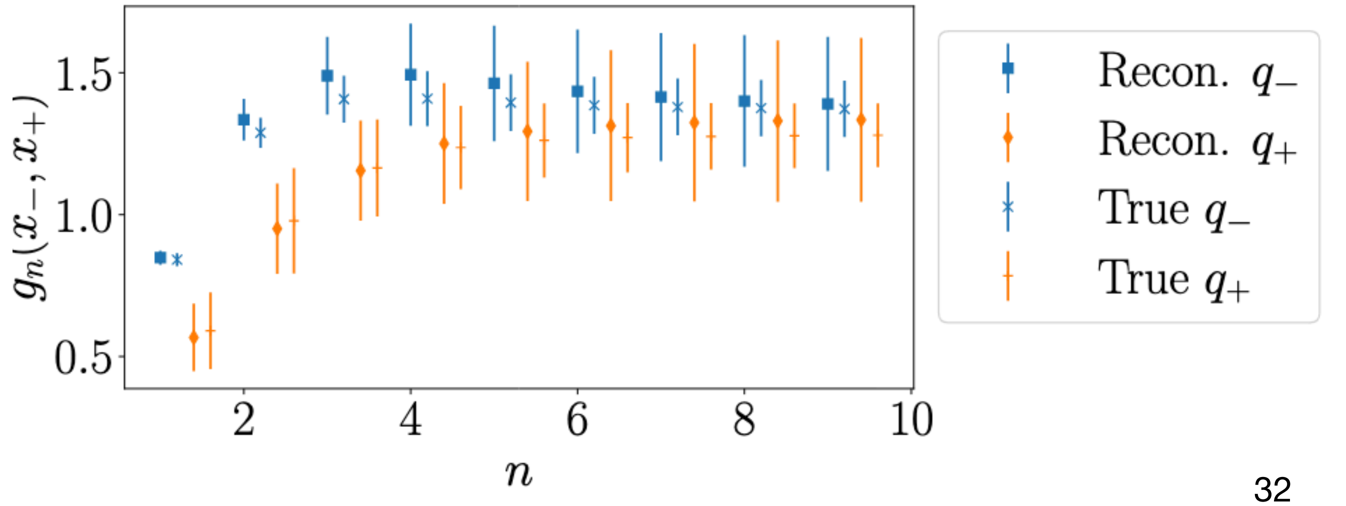
JK et al arXiv:2505.22795





$x_- = 0.1 \quad x_+ = 0.5$



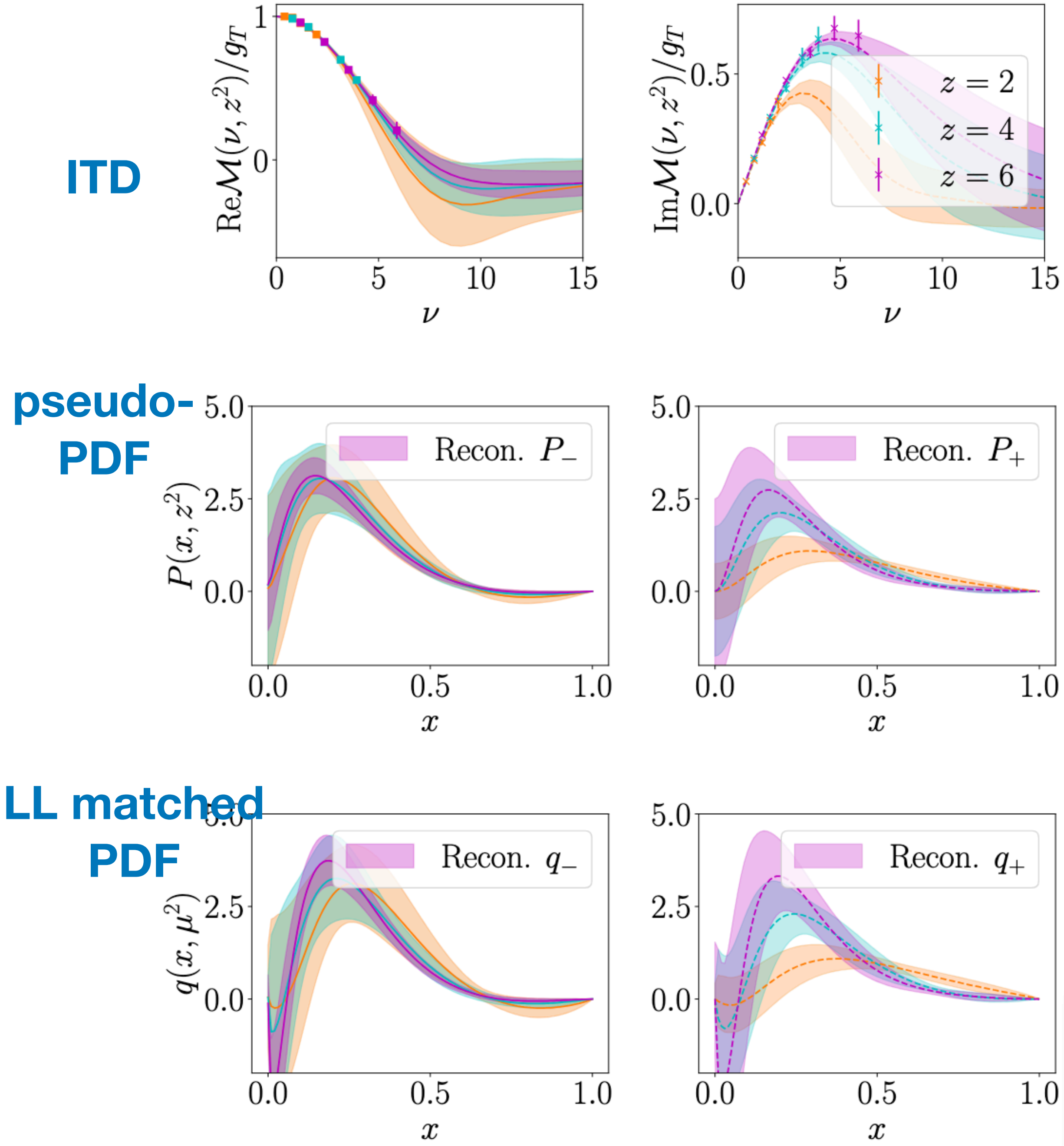


Window Observables instead of x space comparison

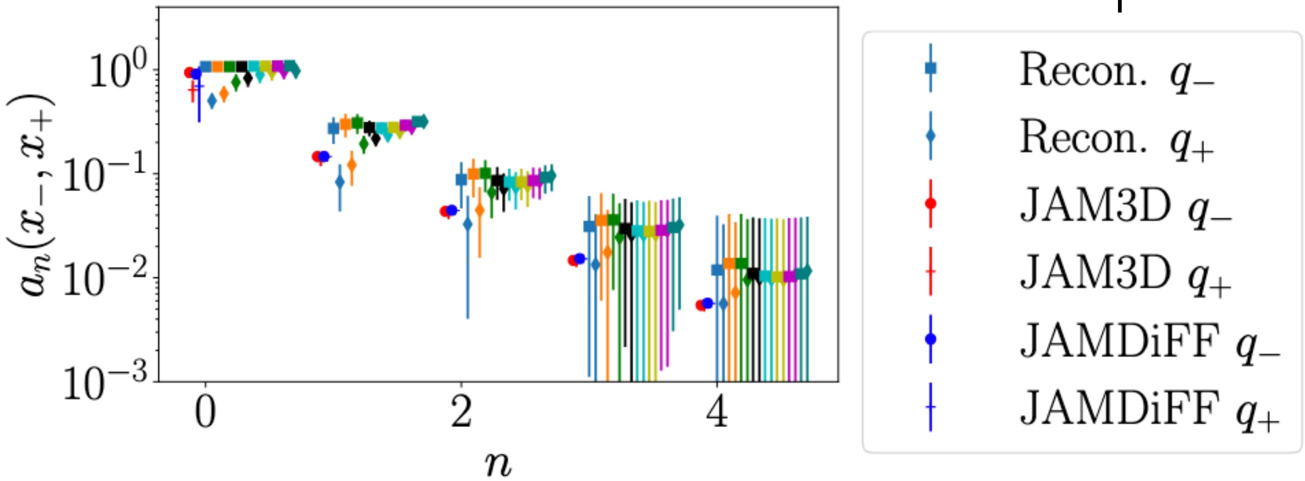
- GP reconstruction of HadStruc transversity u-d PDFs

C. Egerer et al arXiv:2111.01808

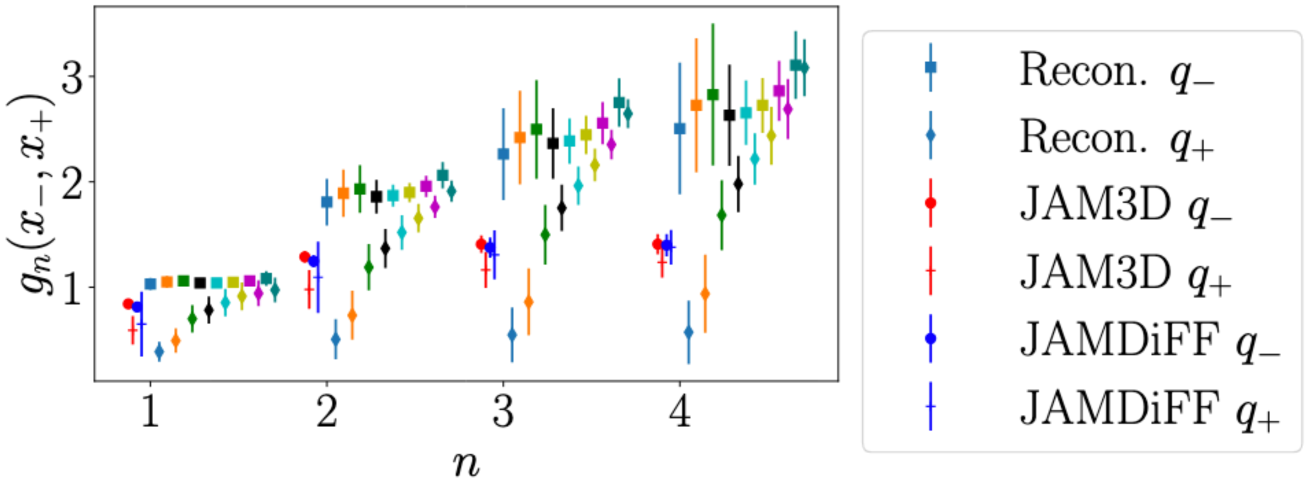
- Precision of the low n (widest windows) is hiding in the covariance of the data



$x_- = 0.1 \quad x_+ = 0.5$



$a = 0.094\mathrm{fm} \;\; m_\pi = 358\mathrm{MeV}$

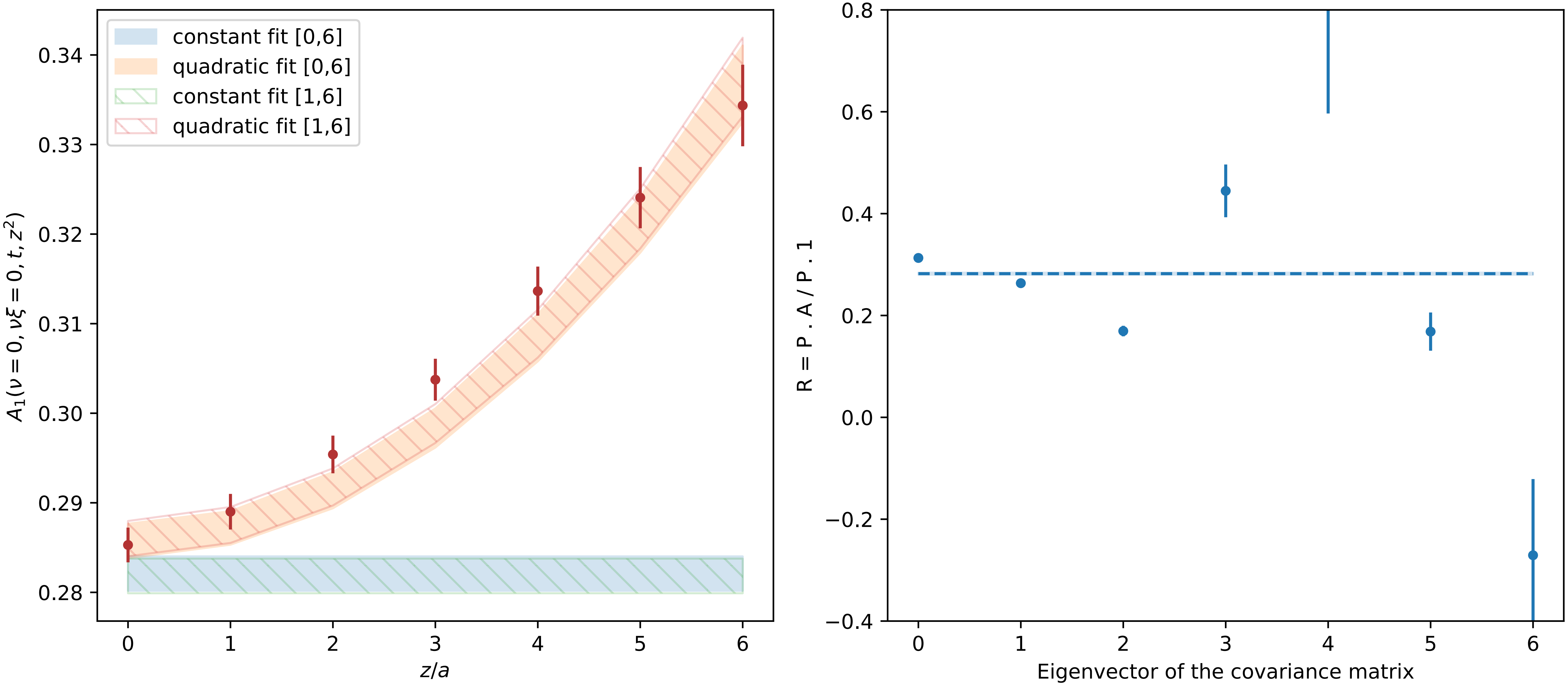


JK et al arXiv:2505.22795

- Lowest zs give bad reconstruction, but errors are promising

There is less than you think

- Lattice data are highly correlated and you have less information than you think
- What is the average of these 7 data points?



H. Dutrieux et al (HadStruc) JHEP 08 (2024) 162

Other Faces of WL Matrix Element

- Some are Lorentz invariant interpretations
- These interpretations nor the functions' bounds require small z^2 , only relation to light cone PDF with $z^2 = 0$ and some other regulation

VDF: A. Radyushkin (2016) PRD 93 056002

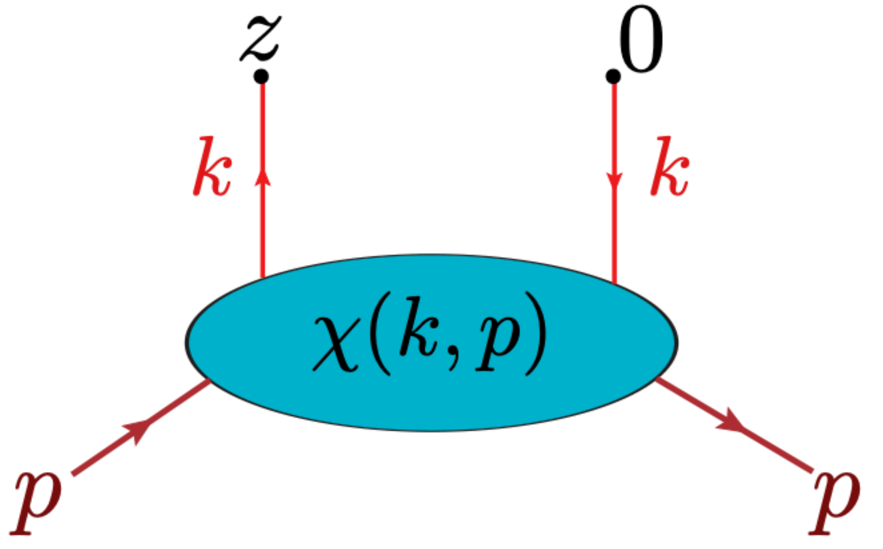
$$i\chi_{d_i}(k,p) = i^l \frac{P(\text{c.c.})}{(4\pi i)^{2L}} \int_0^\infty \prod_{j=1}^{l} d\alpha_j [D(\alpha)]^{-2}$$

$$\times \exp\left\{ik^2\frac{A(\alpha)}{D(\alpha)} + i\frac{(p-k)^2B_s(\alpha)+(p+k)^2B_u(\alpha)}{D(\alpha)}\right\}$$

$$\times \exp\left\{ip^2\frac{C(\alpha)}{D(\alpha)} - i\sum_j \alpha_j(m_j^2 - i\epsilon)\right\} ,$$

$$\sigma_{d_i} = \frac{A_{d_i}(\alpha) + B_{s_{d_i}}(\alpha) + B_{u_{d_i}}(\alpha)}{D_{d_i}(\alpha)}$$

$$x_{d_i} = \frac{B_{s_{d_i}}(\alpha) - B_{u_{d_i}}(\alpha)}{A_{d_i}(\alpha) + B_{s_{d_i}}(\alpha) + B_{u_{d_i}}(\alpha)}$$



$$i\chi(k,p) = \int_0^\infty d\sigma \int_{-1}^{1} dx e^{i\sigma\left[k^2 - 2x(k\cdot p) + i\epsilon\right]} V(x,\sigma)$$

α_j are positive numbers and A, B_u, B_s, C, D are sums of products of α_j

Fourier transform to position space

$$\mathscr{M}(\nu, z^2) = \int \frac{d^4k}{(2\pi)^4} e^{ik\cdot z}\chi(k,p) = \int_{-1}^{1} dx e^{i\nu x} \int_0^\infty e^{-i\sigma(z^2-\epsilon)} V(x,\sigma)$$

Other Faces of WL Matrix Element

- Some are Lorentz invariant interpretations
- These interpretations nor the functions' bounds require small z^2 , only relation to light cone PDF with $z^2 = 0$ and some other regulation

VDF: A. Radyushkin (2016) PRD 93 056002

Virtuality Distribution Function
Lorentz invariant picture
σ^{-1} pole gives log z^2
Limits from nature of Feynman diagrams

$$\mathscr{M}(\nu, z^2) = \int_{-1}^{1} dx e^{i\nu x} \int_0^{\infty} d\sigma e^{-i\sigma(z^2-\epsilon)} V(x,\sigma)$$

pseudo-PDF
Lorentz invariant picture
log z^2 divergence from poles of TMD/VDF

$$\mathscr{M}(\nu, z^2) = \int_{-1}^{1} dx e^{i\nu x} P(x, z^2)$$

$$\tilde{q}(y, p_z^2) = \int dz \int_{-1}^{1} dx e^{ip_z z(x-y)} P(x, z^2)$$

Musch, Hagler, Negele, Schafer PRD 83 (2011) 094507

Straight Link / Primordial TMD
Frame dependent picture with nice interpretation
$1/k_T^2$ pole gives log z^2

$$z = (0, z^-, z_T) \qquad p = (p^+, \frac{m^2}{p^+}, 0_T)$$

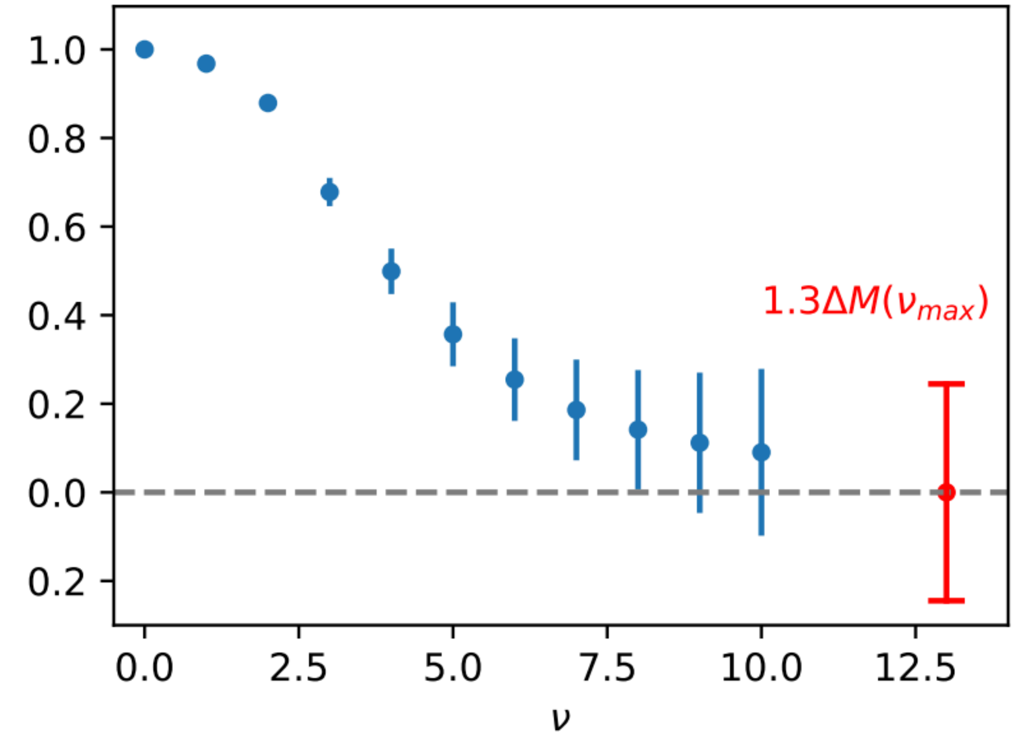
$$\mathscr{M}(\nu, z^2) = \int_{-1}^{1} dx e^{i\nu x} \int d^2k_T e^{ik_T \cdot z_T} F(x, k_T^2)$$

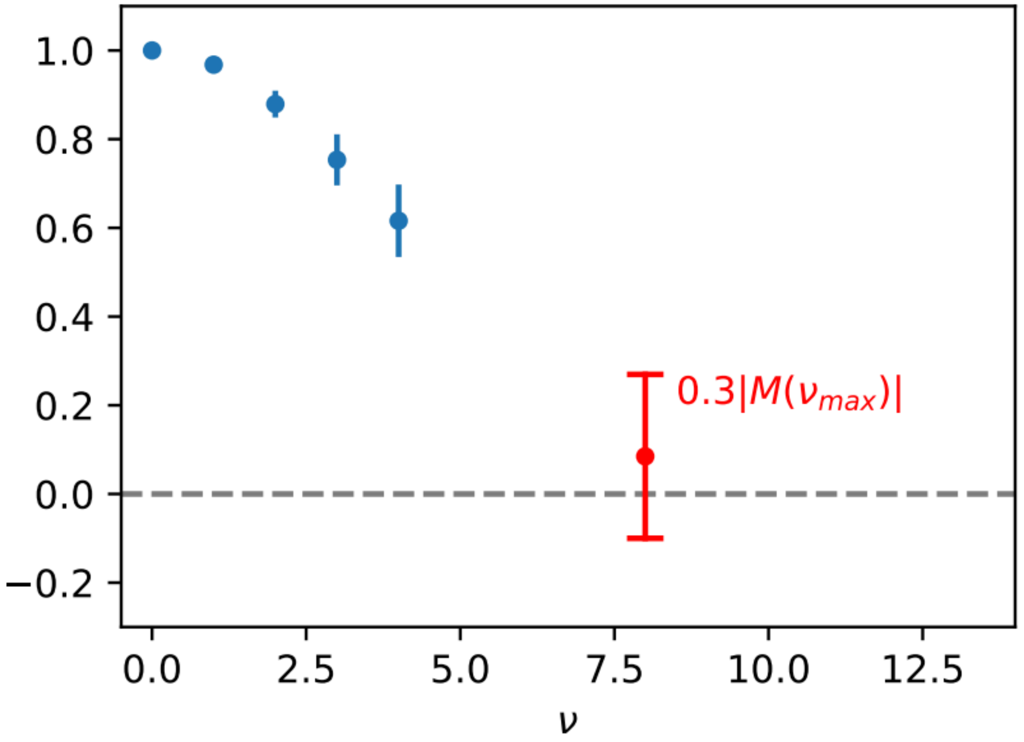
Light cone PDF from regulated integral of TMD
Relate to the Lorentz invariant VDF
$1/k_T^2$ or σ^{-1} poles generate $\log \mu^2$ divergence

$$f(x, \mu^2) = \int^{\mu^2} d^2k_T F(x, k_T^2) = \int_0^{\infty} d\sigma \left[1 - e^{-\frac{i}{\sigma}(\mu^2 - i\epsilon)}\right] V(x, \sigma)$$

Choosing GPR Hyperparameters

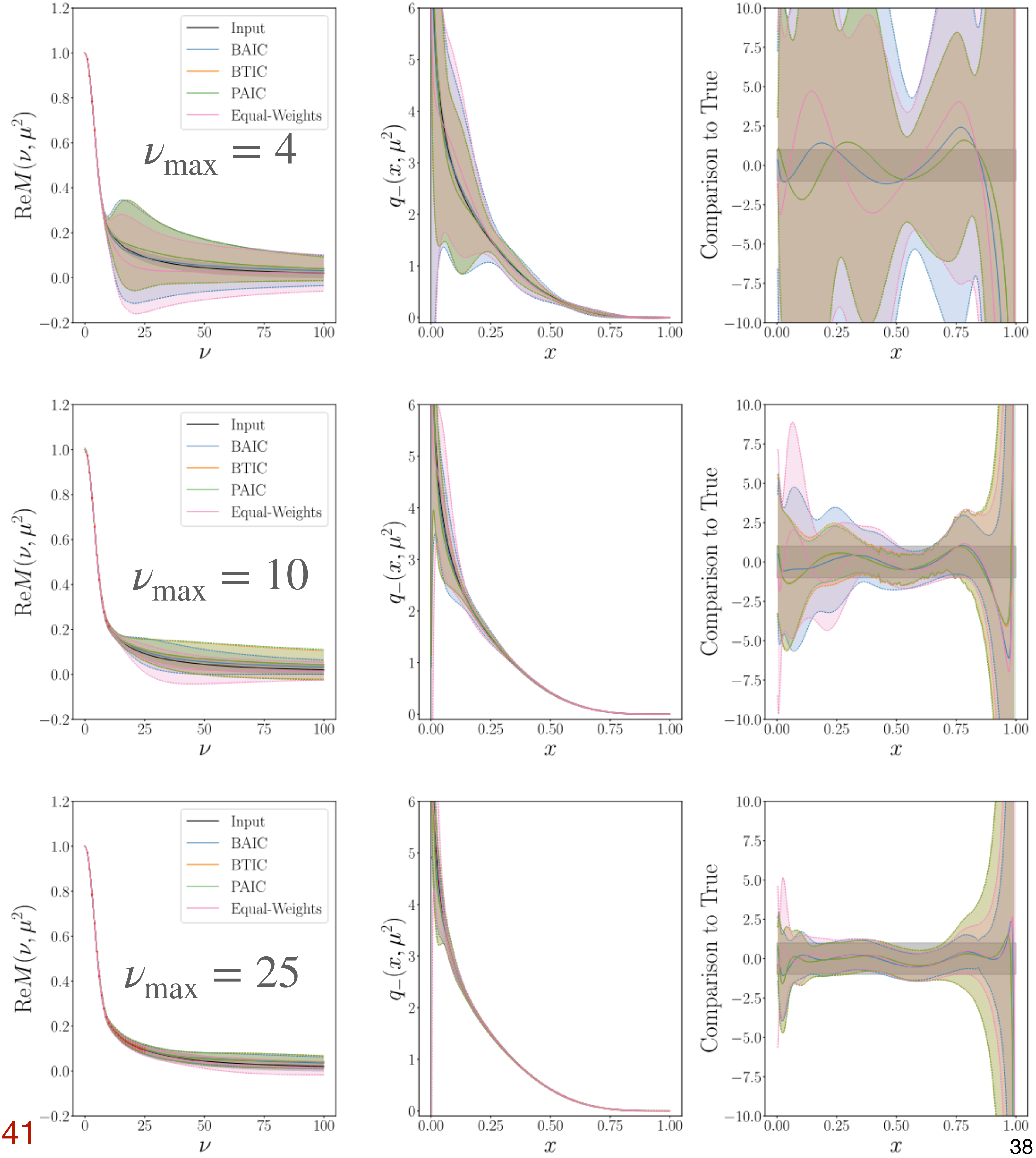
Choice	Motivation
$K(x, x')$ eq. (11)	Enforce smoothness in x while decorrelating small x from large x
$l = \ln(2) \approx 0.693$	Set the flexibility of the x-dependence
$g(x) = \sigma$	Set the uncertainty to 100% at $x = 0$ and loosely enforce positivity
σ^2 such that the worst uncertainty in Fourier space is $\max(1.3\Delta M(\nu_{max}), 0.3\lvert M(\nu_{max})\rvert)$	Control the size of uncertainty in the extrapolation region in Fourier space





Which Model is the best?

- "Information criteria" are different approximations of difference between your posterior and the true posterior
- IC define weights to average different models
- Largely give consistent central values with slightly different error estimates



Y Cahuana Medrano, H Dutrieux, JK,
K Orginos, S Zafeiropoulos arXiv:2510.21041