Pixel-Based Imaging of TMDs:
A novel approach to Distribution Functions

Marco Zaccheddu- Jefferson Lab

BNL-INT Workshop: Bridging Theory and Experiment at the Electron-Ion Collider

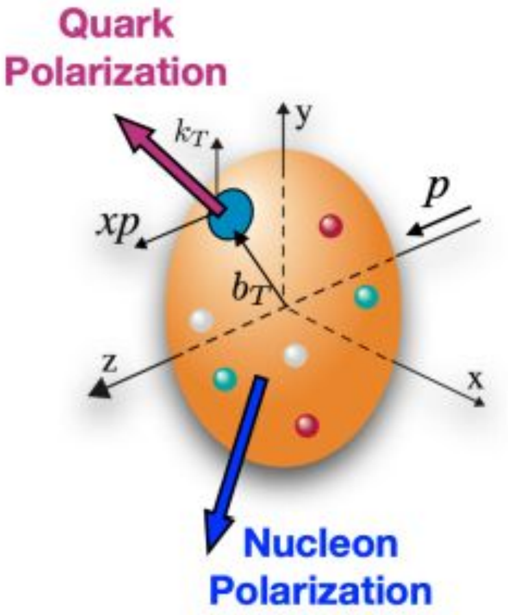
Jefferson Lab

Outline

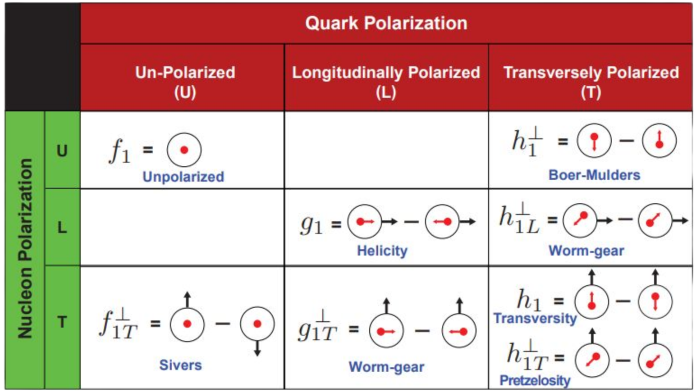
- **TMDs:** 3D Nucleon Structure, Definition, Importance, Categories
- **Extraction:** From Data, Inverse Problem, UQ
- **Pixel Approach:** Discretization, Advantages, Generative AI
- **Normalizing Flows:** Core Concept & Training
- **Proof of Concept:** Closure Test, Pseudo-Data
- **Results:** Pixel vs. Traditional, Uncertainty Handling
- **Conclusions:** Benefits of Pixel-Based Method

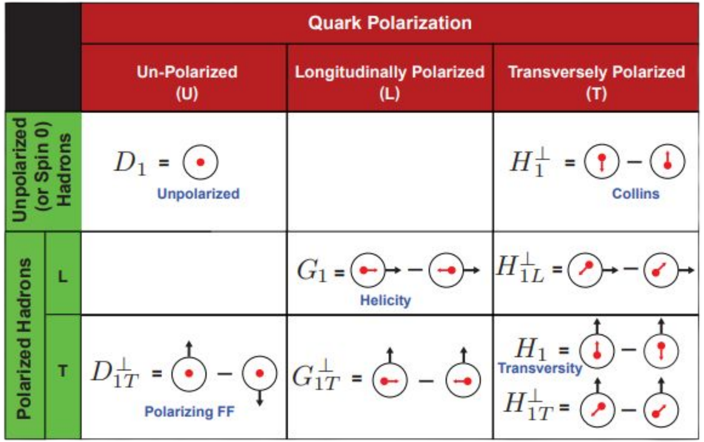
TMD Distributions

- **Transverse Momentum Dependent** Parton Distributions reveal the **3D structure** of protons and neutrons.
- Unlike standard PDFs, TMDs account for a parton's **transverse momentum (kT)**, not just its collinear motion.
- This kT information offers a more complete, **spatial view** of parton dynamics within the nucleon.
- **Fragmentation Functions (FFs)** are also crucial, describing how quarks and gluons **hadronize** into observable particles.



TMD Distributions





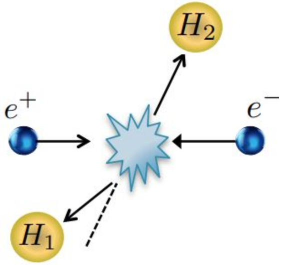
- TMDs are classified into **two categories**, each containing **8 independent functions.**
- They describe the **correlation** between parton **transverse momentum** and the **spin** of both partons and the transverse hadron.
- Different TMDs give rise to various **hadronic asymmetries** observed in experiments.

TMDs, as **continuous functions**, inherently carry rich information.

Hadron **tomography** in transverse momentum space.

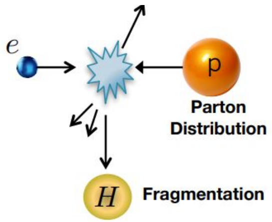
TMD Handbook: arXiv:2304.03302

Where do we extract TMDs?

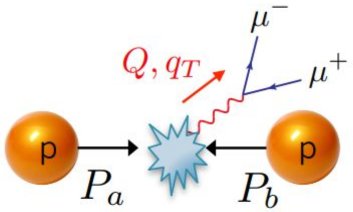


e^+e^-

SIDIS



DY



- TMDs are extracted from diverse data, thanks to **factorization.**
- Reconstructing TMDs from finite data is an **inverse problem.**
- Robust **Uncertainty Quantification** (UQ) is critical for reliable TMD reconstructions.

TMDs extraction

$$\tilde{f}^{[\Gamma]0(u)}_{i/p_s}(x, \mathbf{b}_T, \epsilon, \tau, xP^+) \quad \tilde{\Delta}^{[\Gamma]0(u)}_{h;i}(z, \mathbf{b}_T, \epsilon, \tau, P^+/z)$$

Continuous Function
=
“Infinite Information”

Tractable inference via prior

TMD model parametrization

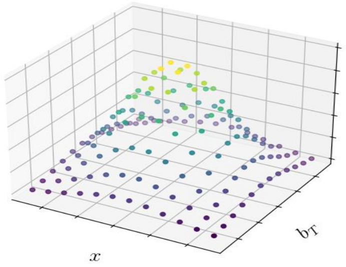
Classical models

MAPTMD 2024 arXiv:2405.13833
ART25 arXiv:2503.11201

NN representations

MAPTMDNN 2025
arXiv:2502.04166

Pixel-Based
approach

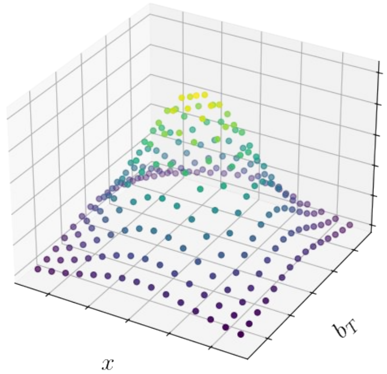


Pixel
Optimization
&
UQ

Reducing Continuum Information

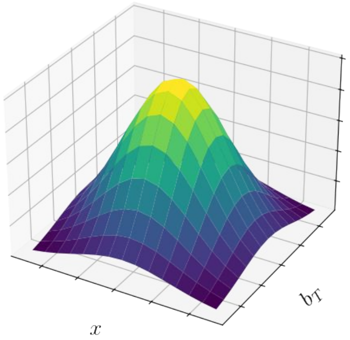
Pixel based Reconstruction of TMDs

Discretized TMD in x and bT space

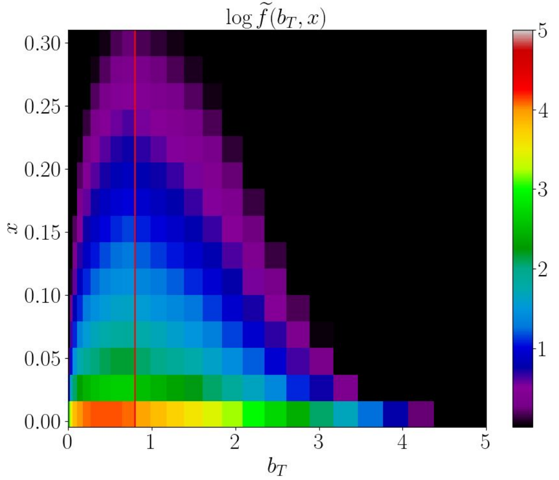


Linear Interpolation

Continuous Function



Pixel-based:
because results can be represented as an images!



- Explicit control of ***expressivity*** and TMD structure via discretization (i.e. # pixels).
- Quantifying ***local resolution*** from data.
- Improved ***Uncertainty Quantification***.
- **Model Independent Ansatz**

Image Reconstruction using Generative AI

Generative AI

Understanding, Generating, Reconstructing Images

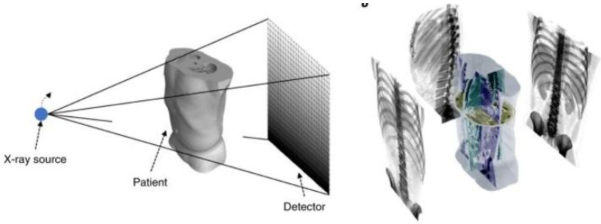
Generative AI Examples:

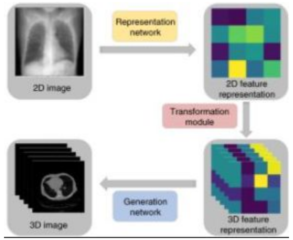
Generation of highly realistic human faces

Architectural visualizations and design

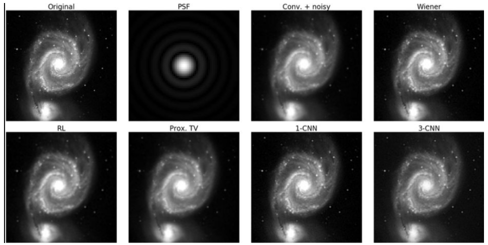
Used in Scientific Fields for Reconstruction:

Medical Imaging





Astrophysics

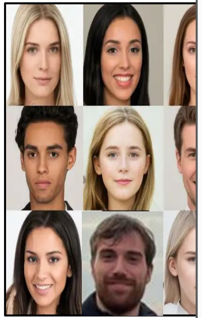


Art Restoration

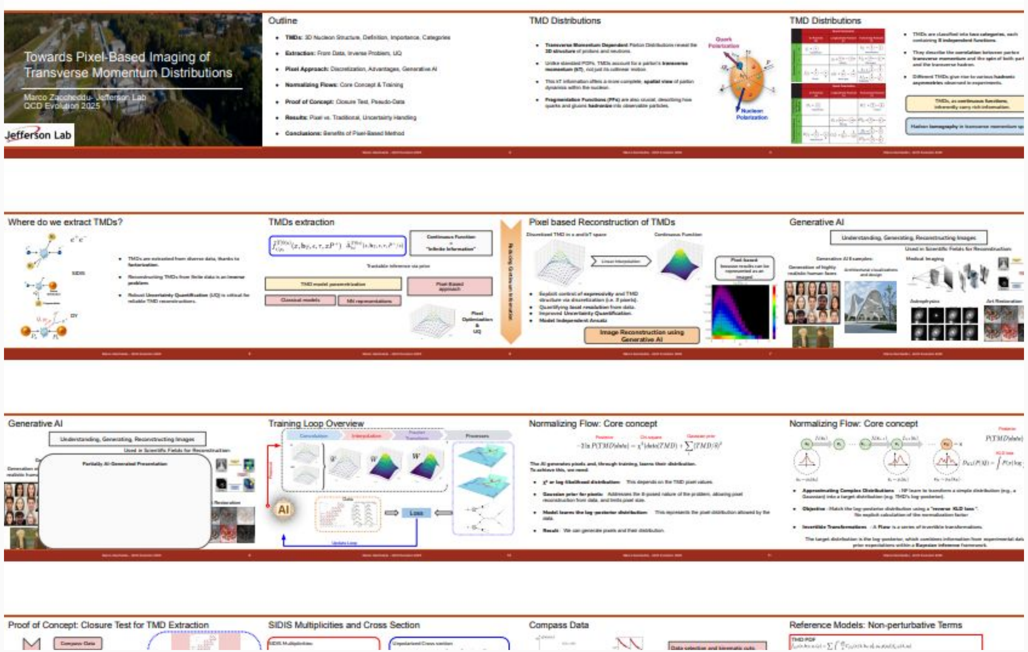
Generative AI

Understanding, Generating, Reconstructing Images

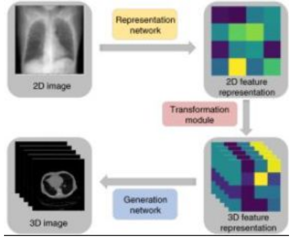
Generatio
realistic hu



Partially AI-Generated Presentation

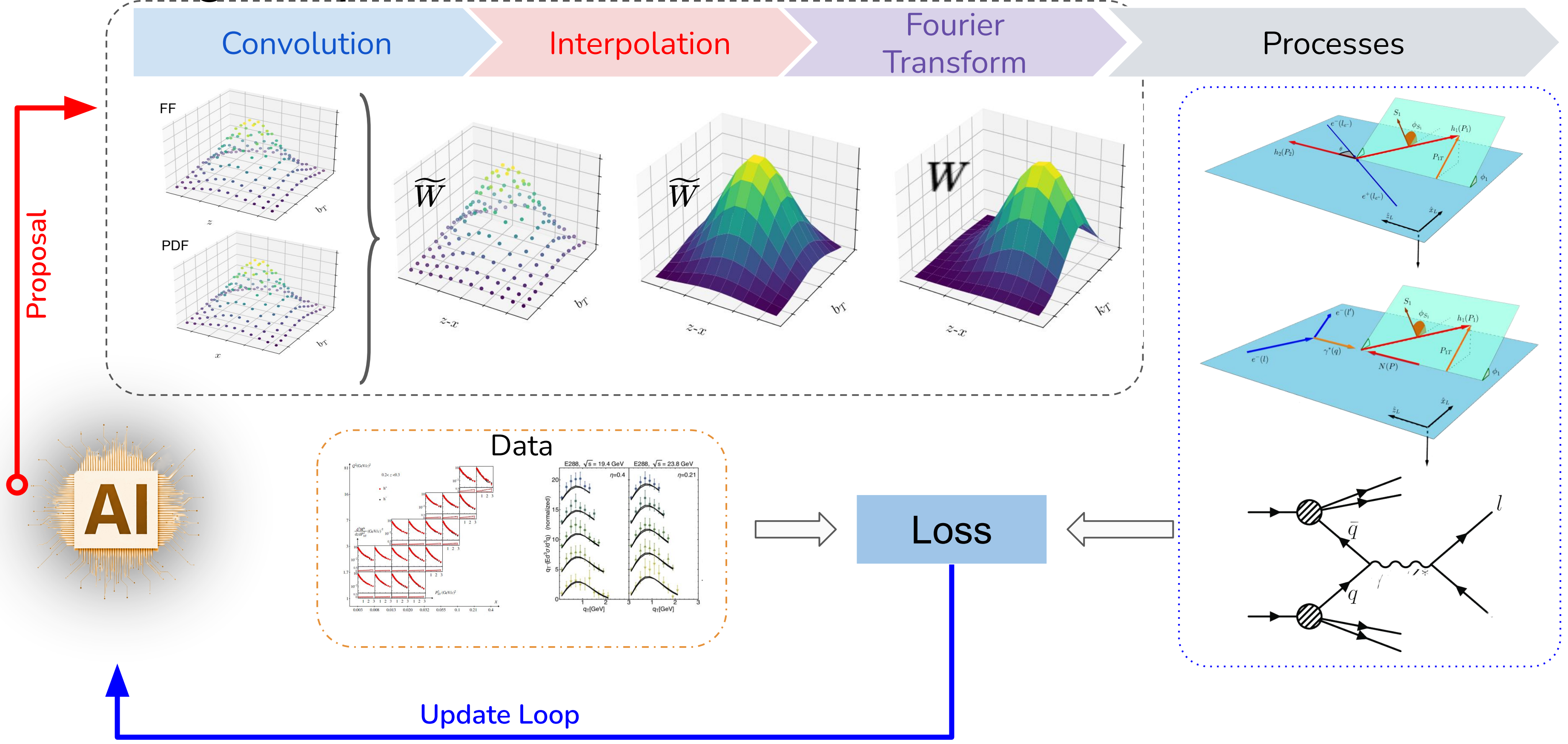


r Reconstruction:



Art Restoration

Training Loop Overview



Normalizing Flow: Core concept

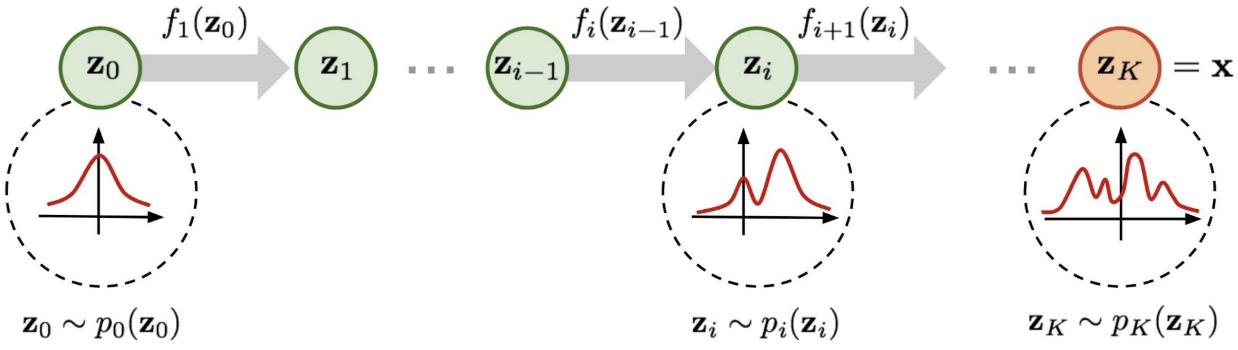
Posterior Chi-square Gaussian prior

$$-2\ln P(TMD|data) = \chi^2(data|TMD) + \sum_i (TMD/\delta)^2$$

The AI generates pixels and, through training, learns their distribution.
To achieve this, we need:

- **χ^2 or log-likelihood distribution:** This depends on the TMD pixel values.
- **Gaussian prior for pixels:** Addresses the ill-posed nature of the problem, allowing pixel reconstruction from data, and limits pixel size.
- **Model learns the log-posterior distribution:** This represents the pixel distribution allowed by the data.
- **Result**: We can generate pixels and their distribution.

Normalizing Flow: Core concept



Posterior

$$P(TMD|data)$$

KLD loss

$$D_{KL}(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx$$

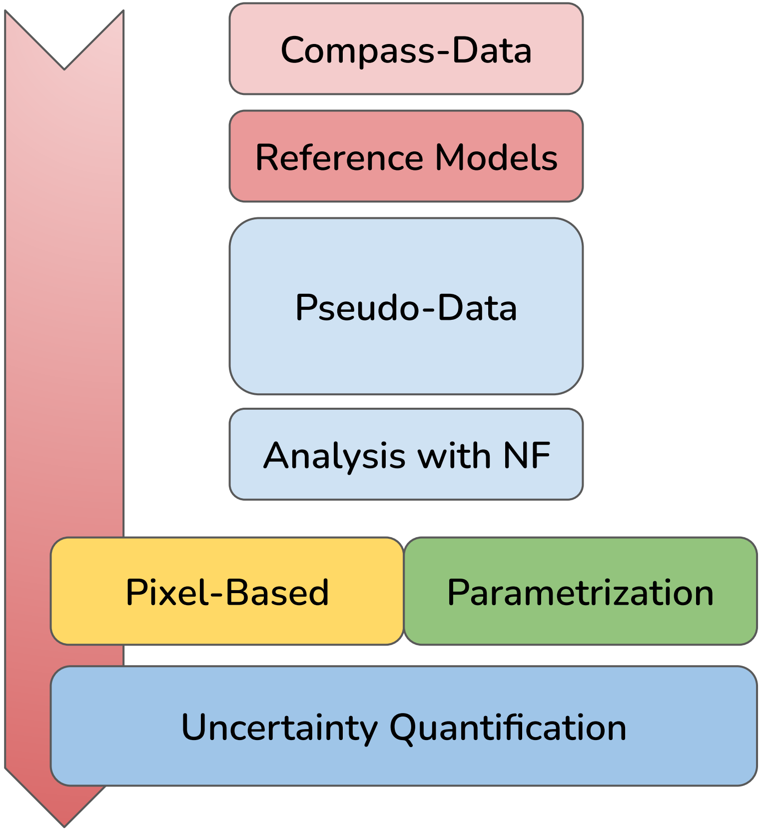
- **Approximating Complex Distributions** : NF learn to transform a simple distribution (e.g., a Gaussian) into a target distribution (e.g. TMD's log-posterior).

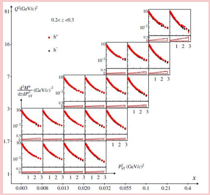
- **Objective** : Match the log-posterior distribution using a "**reverse KLD loss** ".
  No explicit calculation of the normalization factor

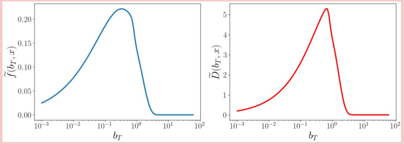
- **Invertible Transformations** : A **Flow** is a series of invertible transformations.

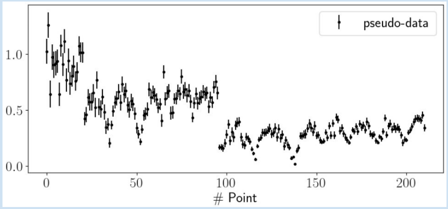
The target distribution is the log-posterior, which combines information from experimental data and prior expectations within a **Bayesian inference** framework.

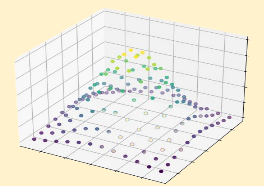
Proof of Concept: Closure Test for TMD Extraction











$$M_f(b_T, x; b_{max}) = \exp\left(-\alpha_f b_T^2\, R(b_T; b_{max})/4\right)$$

$$g_K(b_T; b_{max}) = -g_2 b_T^2\, R(b_T; b_{max})/2$$

SIDIS Multiplicities and Cross Section

SIDIS Multiplicities:

$$\frac{d^2M^h(x,z,P_{h\perp}^2,Q^2)}{dzdP_{h\perp}^2} = \left(\frac{d^4\sigma^h}{dxdQ^2dzdP_{h\perp}^2}\right) \Big/ \left(\frac{d^2\sigma^{DIS}}{dxdQ^2}\right)$$

Unpolarized Cross section

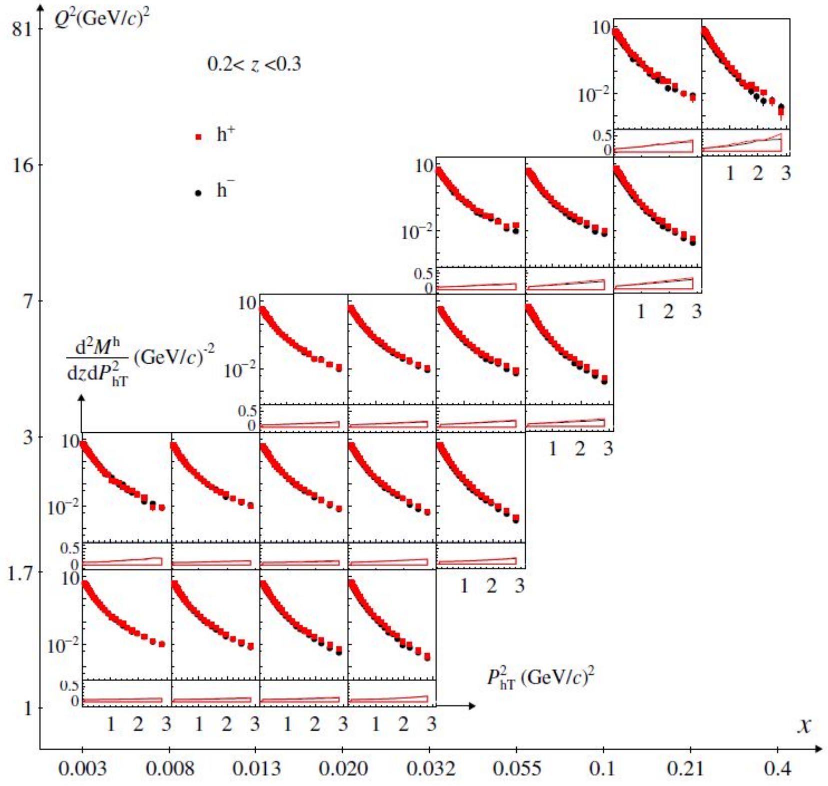
$$\frac{d\sigma}{dxdydzd\phi_S d\phi_h dP_{h\perp}^2} = \frac{\alpha_{em}^2}{xyQ^2}\frac{y^2}{2(1-\epsilon)}\left(1+\frac{\gamma^2}{2x}\right)F_{UU,T}$$

$$\frac{d\sigma^{DIS}}{dxdQ^2} = \frac{4\pi\alpha_{em}^2}{xQ^2}\mathcal{Y}_+F_2(x,Q^2)$$

Form Factor

$$\begin{aligned}
F_{UU}(x,z,q_T^2,Q^2) &= \mathcal{C}[f_1D_1] \\
&= x\,H(Q^2)\sum_a e_a^2\int d^2p_Td^2k_T\delta^{(2)}(\mathbf{p}_T-\mathbf{k}_T-\mathbf{q}_T)f_1^a(x,p_T^2)D_1^a(z,k_T^2) \\
&= x\,H(Q^2)\sum_a e_a^2\frac{1}{2\pi}\int db_Tb_TJ_0(q_Tb_T)\tilde{f}_1^a(x,b_T)\tilde{D}_1^a(z,b_T)
\end{aligned}$$

Compass Data



PHYS. REV. D 97, 032006 (2018)

Data selection and kinematic cuts:

$Q^2 > 5\, GeV^2$

$q_T/Q < 0.5$

$z < 0.6$

$had = h^+$

211 points

Extract Reference Models

Pseudo-data:
Same statistics and kinematic cuts

Reference Models: Non-perturbative Terms

TMD PDF

$$\tilde{f}_{j/P}(x, \mathbf{b}_T; \mu, \zeta_F) = \sum_j \int_x^1 \frac{d\hat{x}}{\hat{x}} C_{f/j}(x/\hat{x}, \mathbf{b}_*; \mu_b^2, \mu_b, g(\mu_b)) f_{j/P}(\hat{x}, \mu_b)$$

$$\times \exp\left\{ \ln \frac{\sqrt{\zeta_F}}{\mu_b} \tilde{K}(b_*; \mu_b) + \int_{\mu_b}^{\mu} \frac{d\mu'}{\mu'} \left[\gamma_F(g(\mu'); 1) - \ln \frac{\sqrt{\zeta_F}}{\mu'} \gamma_K(g(\mu')) \right] \right\} \underbrace{\boxed{\exp\left\{ g_{j/P}(x, b_T) \right.}}_{M_f(b_T, x; b_{max})} \left. + g_K(b_T) \ln \frac{\sqrt{\zeta_F}}{\sqrt{\zeta_{F,0}}} \right\}$$

Non-perturbative Models:

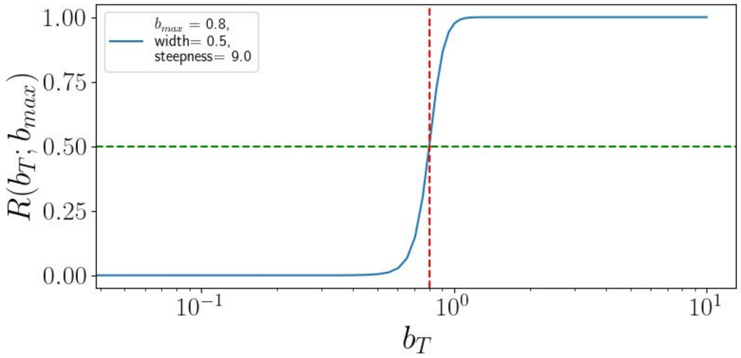
$$M_f(b_T, x; b_{max}) = \exp\left(-\alpha_f b_T^2 R(b_T; b_{max})/4\right) \qquad \alpha_f = 0.84$$

$$M_D(b_T, z; b_{max}) = \exp\left(-\alpha_D b_T^2 R(b_T; b_{max})/4z^2\right) \qquad \alpha_D = 0.24$$

$$g_K(b_T; b_{max}) = -g_2 b_T^2 R(b_T; b_{max})/2 \qquad g_2 = 0.29$$

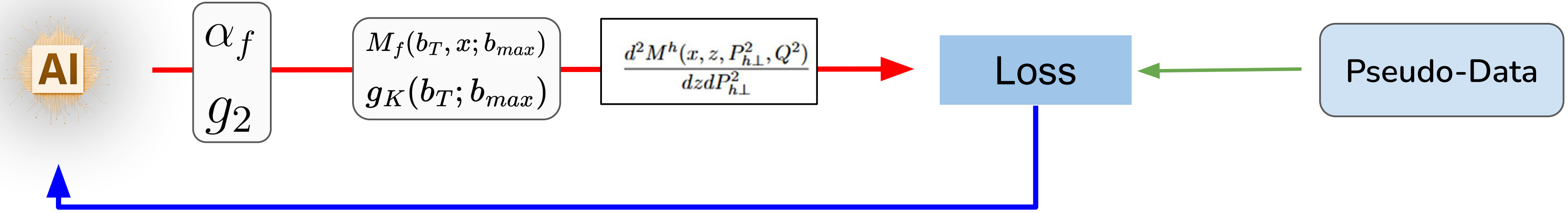
Transition Function

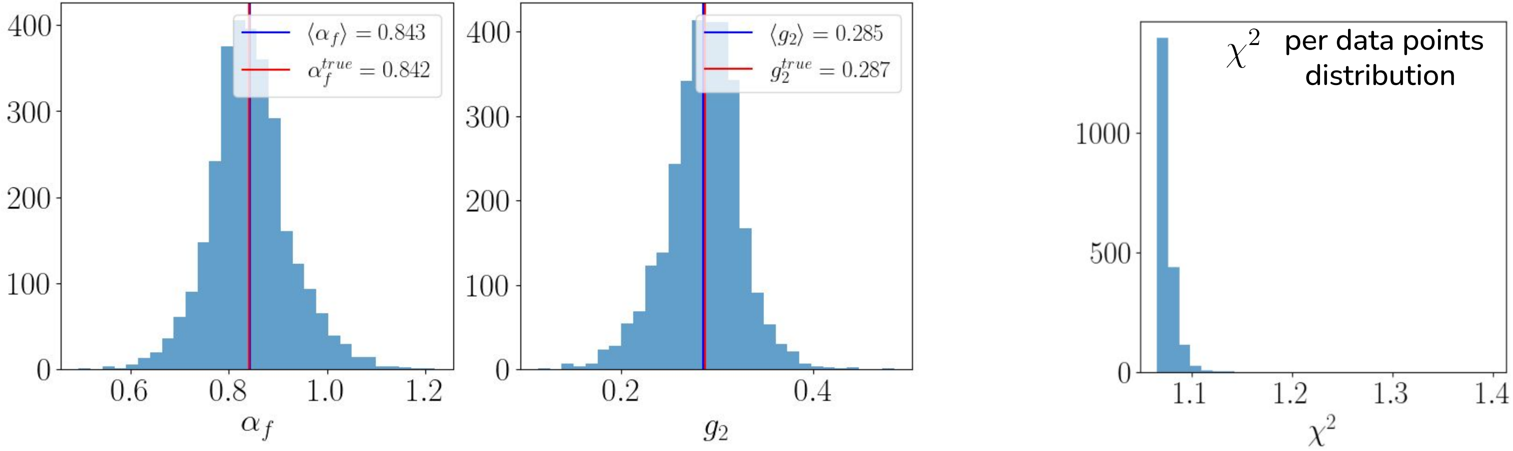
$$R(b_T; b_{max}) = 1 - 1/(1 + \exp(stps \cdot (b_T - b_{max})/wdt))$$



Closure Test: Parametrization Based

PDF parameters fitted. FF is fixed.





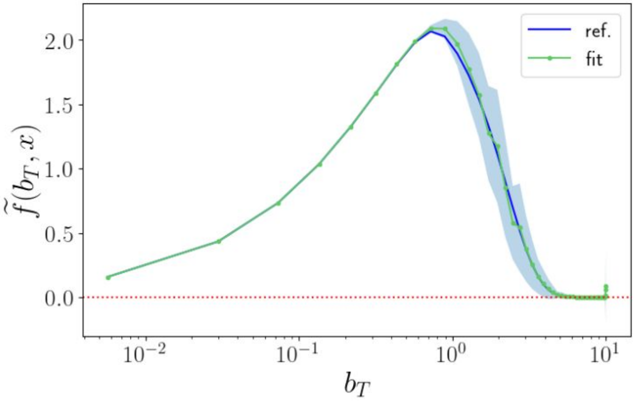
$$\chi^2_{pts} = 1.07$$

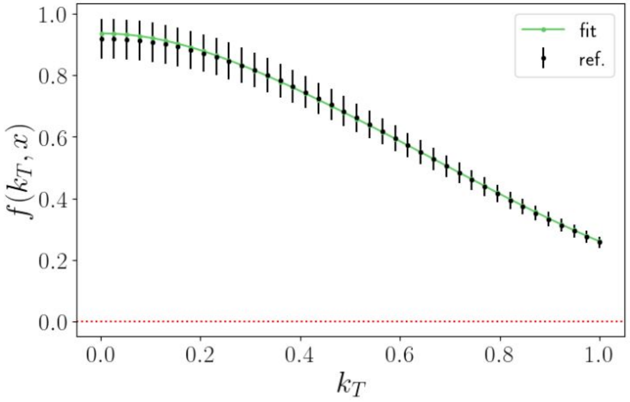
Closure Test: Pixel-based

First check: can we invert the Fourier Transform?

Fourier Transform inversion is unique,
assuming the entire frequency spectrum is known.

$$f(x, k_T) = \frac{1}{2\pi} \int db_T\, b_T J_0(b_T k_T) \widetilde{f}(x, b_T)$$





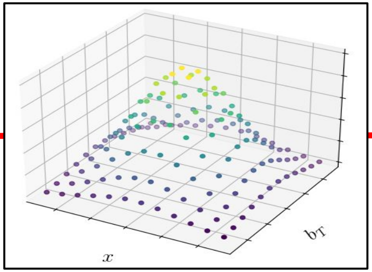
We tested varying numbers of kT points to assess inversion feasibility:

- The final band width depends on the number of available kT points and their associated errors.
- The prior influences the final width of the band.

Closure Test: Pixel-based

PDF pixels reconstructed. FF is fixed.

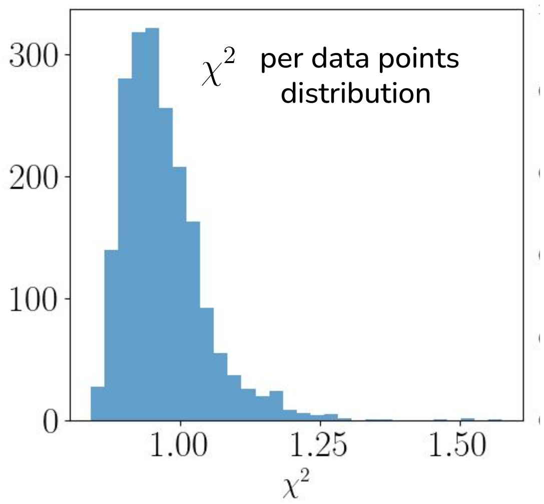
AI



$$\frac{d^2 M^h(x, z, P_{h\perp}^2, Q^2)}{dz dP_{h\perp}^2}$$

Loss

Pseudo-Data



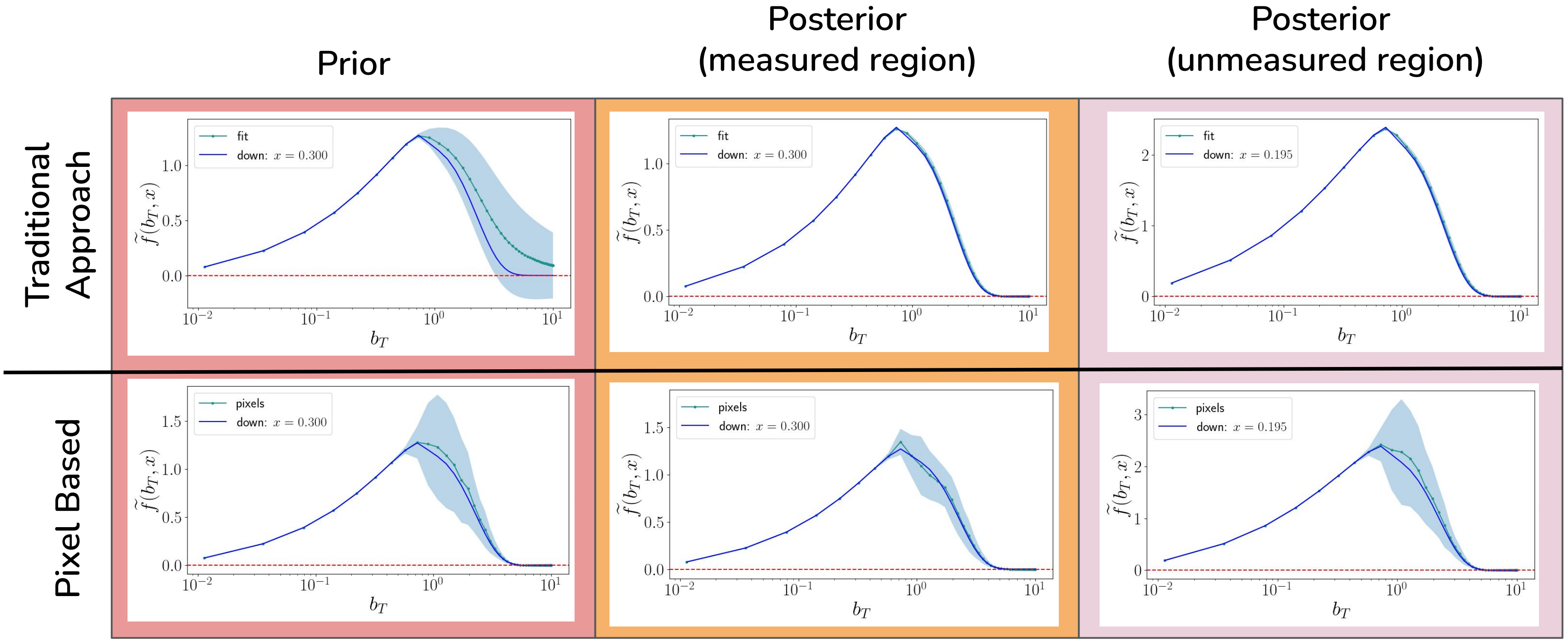
Number of pixels: 750

$n_x = 15$

$n_{b_T} = 50$

5 Flows

$\chi^2_{pts} = 0.84$



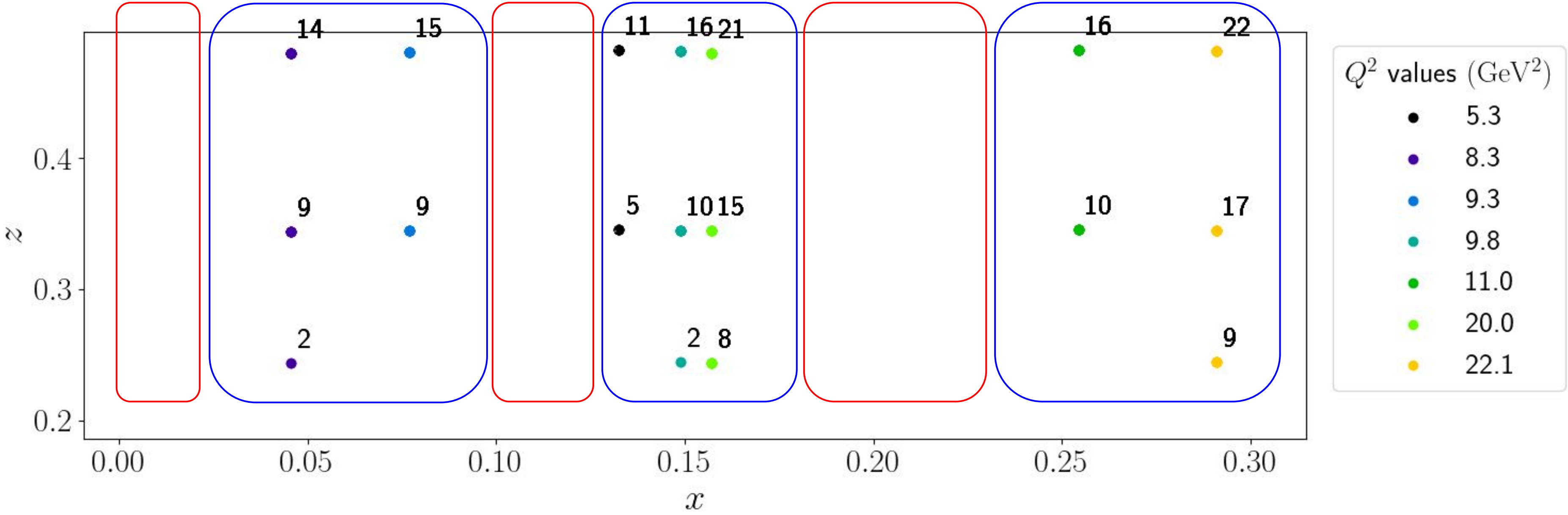
Traditional Approach:

- Smaller uncertainties
- Artificial constraints on extrapolated region

Pixel Based:

- Error band shrinks in measured region
- Band remains constant in unmeasured region (like Prior)

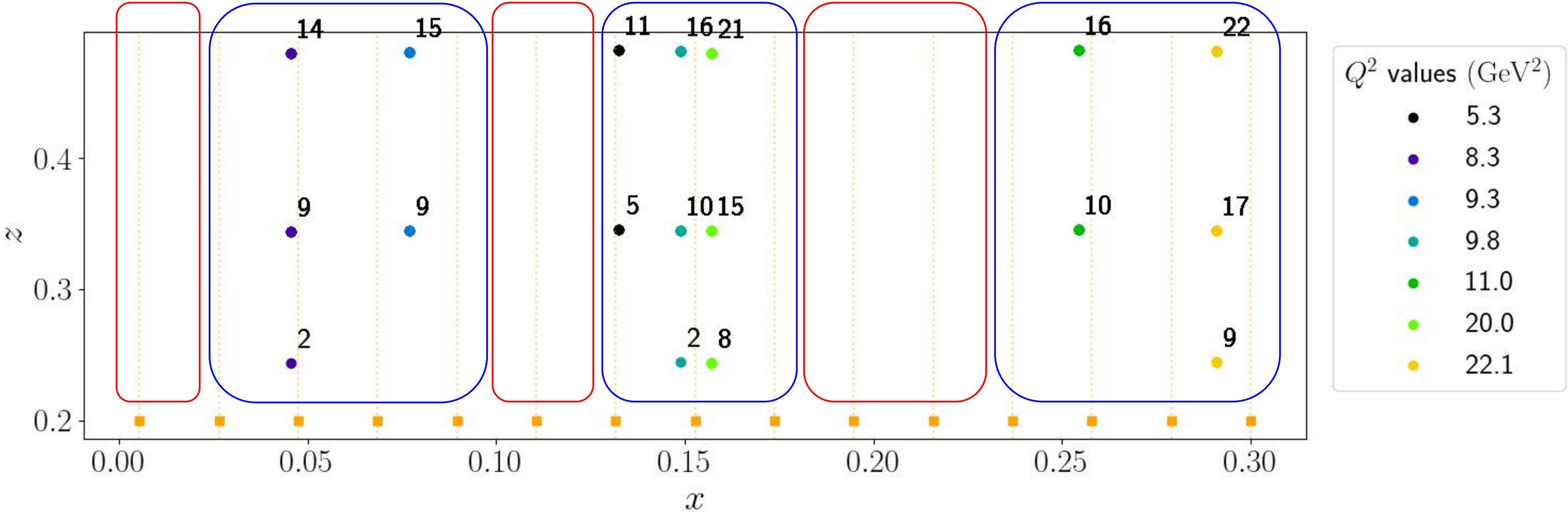
Pseudo Data



Measured regions

Unmeasured regions

Pseudo Data



Measured regions

Unmeasured regions

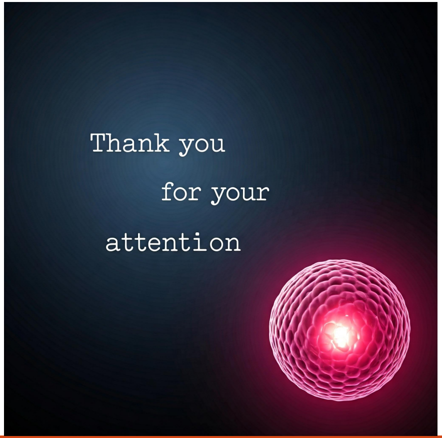

Pixels

Conclusions

- **Transverse Momentum Distributions** (TMDs) provide a **3D view** of nucleon structure.
- A **novel pixel-based** approach reconstructs TMDs using **Generative AI**, specifically Normalizing Flows.
- This pixel method allows explicit **control of expressivity**, improved **uncertainty quantification**, and **model independence.**
- Normalizing Flows learn the **log-posterior distribution of TMD pixels**, matching experimental data with prior expectations.
- **Closure tests** with pseudo-data demonstrate that the pixel-based approach offers more **realistic uncertainty handling** compared to traditional methods.

Outlooks:

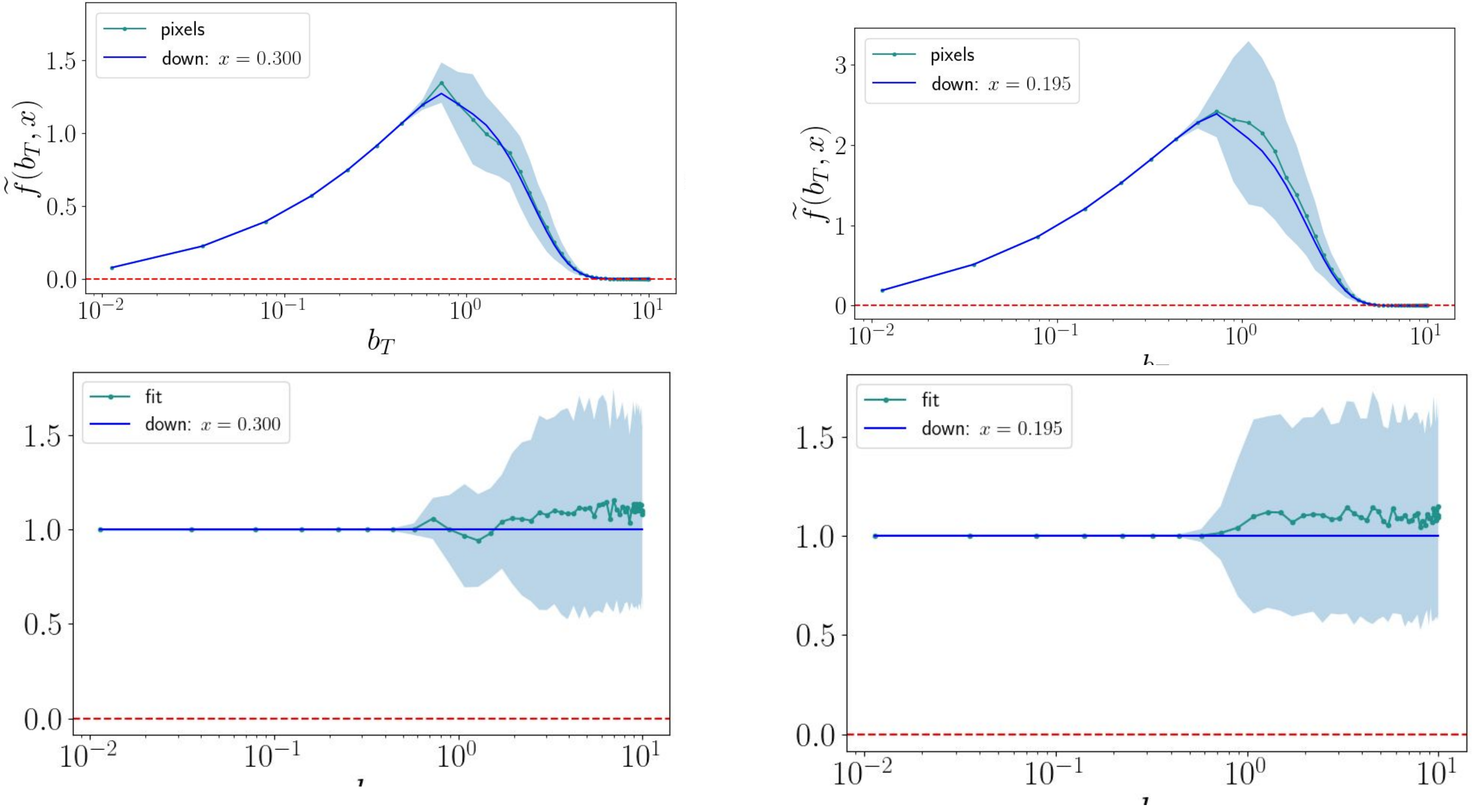
- **Simultaneous PDF and FF Reconstruction**
- **Compass Real Data Analysis**



Backup

$$\tilde{f} = OPE \cdot SDK \cdot \overbrace{\boxed{e^{-\alpha_f b_T^2 R/4 - g_2 b_T^2 R/2 \ln(Q/Q_0)}}}^{\text{reference model}} \overbrace{\boxed{e^{-\epsilon_j R - \epsilon_K R \ln(Q/Q_0)}}}^{\text{pixels}}$$

$$\tilde{f} = OPE \cdot SDK \cdot \underbrace{\boxed{e^{-\alpha_f b_T^2 R/4 - g_2 b_T^2 R/2 \ln(Q/Q_0)}}}_{\text{reference model}} \underbrace{\boxed{e^{-\epsilon_j R - \epsilon_K R \ln(Q/Q_0)}}}_{\text{pixels}}$$



$$\tilde{f}_{i/p_s}^{[\Gamma]0(u)}(x,\mathbf{b}_T,\epsilon,\tau,xP^+) = \int \frac{db^-}{2\pi} e^{-ib^-(xP^+)} \langle p(P,S)|[\bar{\psi}^i(b^\mu)W(b^\mu,0)\frac{\Gamma}{2}\psi^i(0)]_\tau|p(P,S)\rangle$$

$$\tilde{\Delta}_{h;i}^{[\Gamma]0(u)}(z,\mathbf{b}_T,\epsilon,\tau,P^+/z) = \frac{1}{4N_c z}\mathrm{Tr}\int \frac{db^-}{2\pi}\sum_X e^{ib^-(P^+/z)}\Gamma^+_{\alpha\alpha'}\langle 0|[W\psi_i^{0\alpha}(b)]_\tau|h(P,S),X\rangle\langle h(P,S),X|[W\bar{\psi}_i^{0\alpha'}(0)]_\tau|0\rangle$$