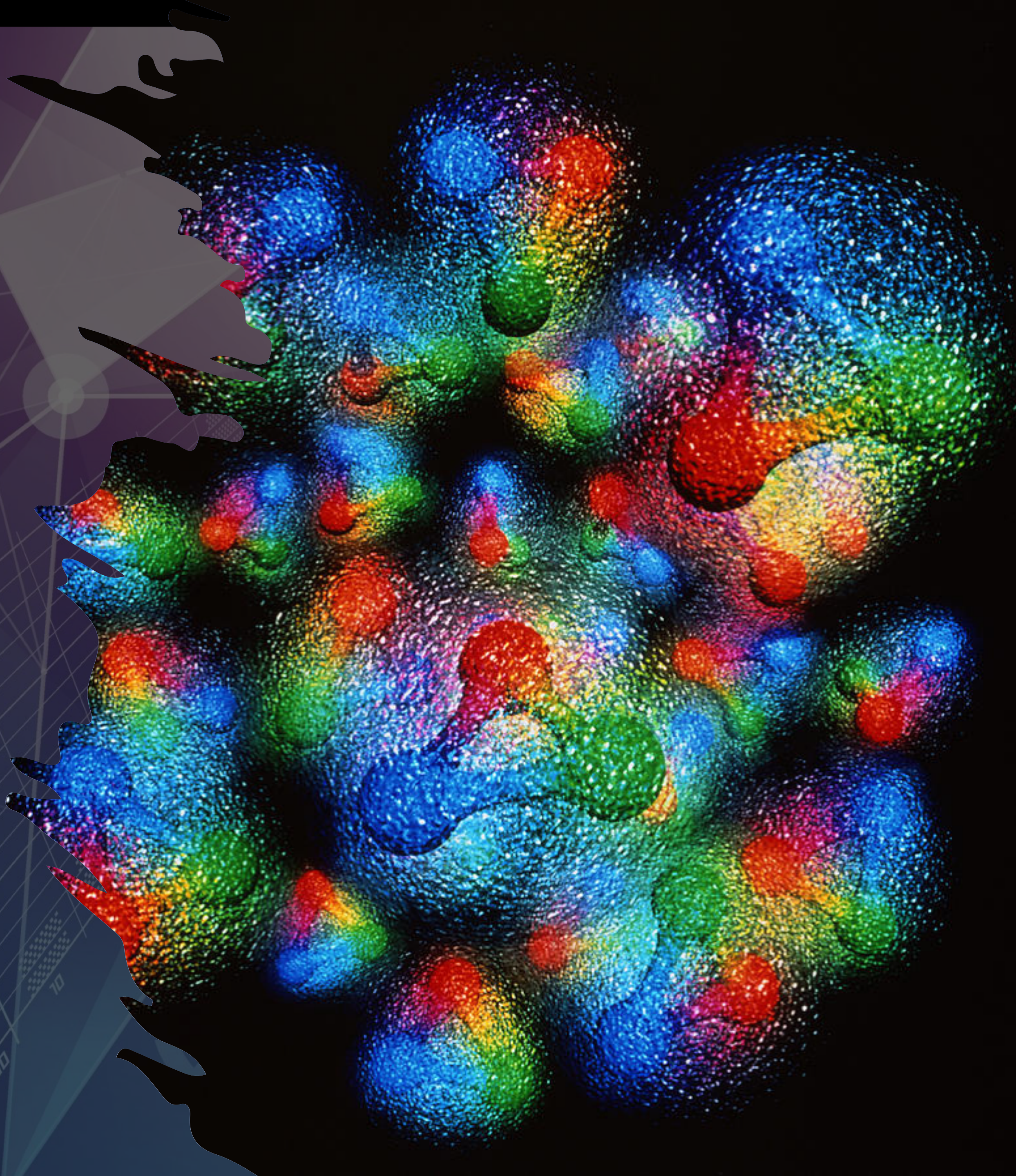
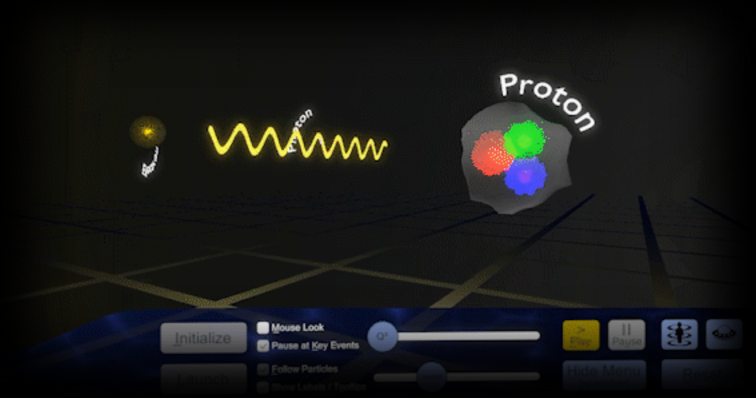
Highlights of the EXCLAIM collaboration

Simonetta Liuti



“The EIC will be a particle accelerator that collides electrons with protons and nuclei to produce snapshots of those particles’ internal structure—like a CT scanner for atoms. The electron beam will reveal the arrangement of the quarks and gluons that make up the protons and neutrons of nuclei.”

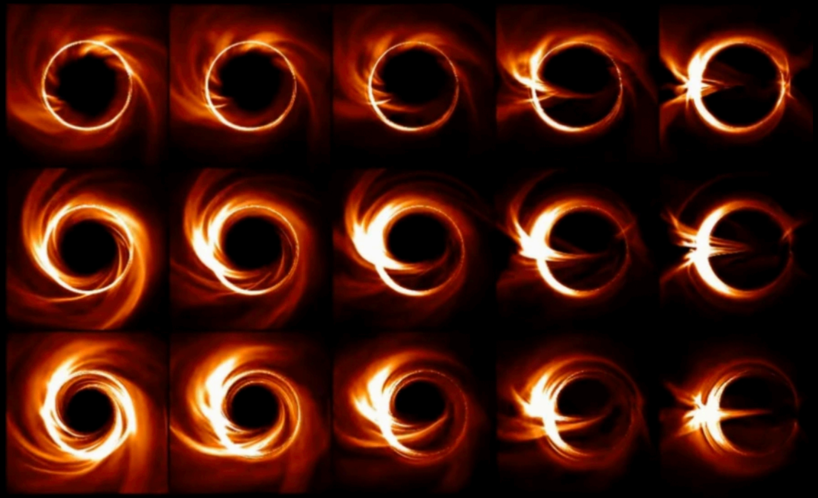
https://www.bnl.gov/eic/



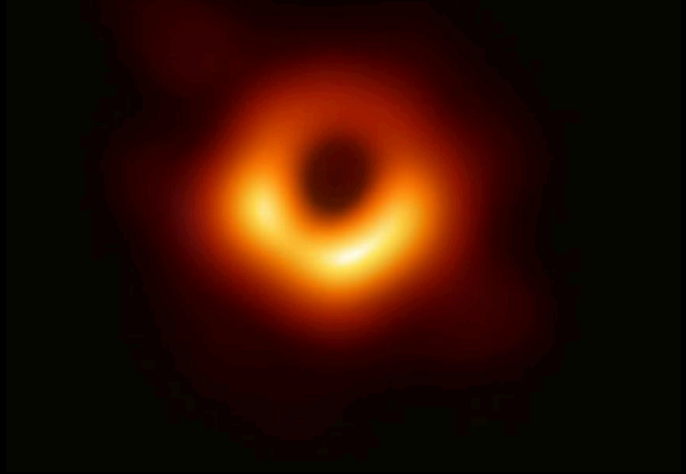
One prominent example of the importance of imaging: the Event Horizon Telescope (EHT) imaged an object as far as 55 M light-years = 5×10^{23} m

- But what is the science that goes into imaging the proton, observing its spatial structure at 10^{-15} m?

SgsA* (



M87* (first image, no ML, 2017)



The EXCLAIM collaboration

https://willuhmjs.github.io/exclaim/

PIs: Marie Boer, Gia-Wei Chern , Michael Engelhardt, Gary Goldstein, Yaohang Li, Huey-Wen Lin, SL, Matt Sievert, (Dennis Sivers)

Funded Postdocs: Douglas Adams, Marija Cuic, Liam Hockley, Saraswati Pandey, Emanuel Ortiz, Kemal Tezgin

Graduate Students: Andrew Dotson, Carter Gustin, Jang (Jason) Ho, Fayaz Hossen, Adil Khawaja, Zaki Panjsheeri, Anusha Singireddy, Jitao Xu

Undergraduate: Will Faircloth

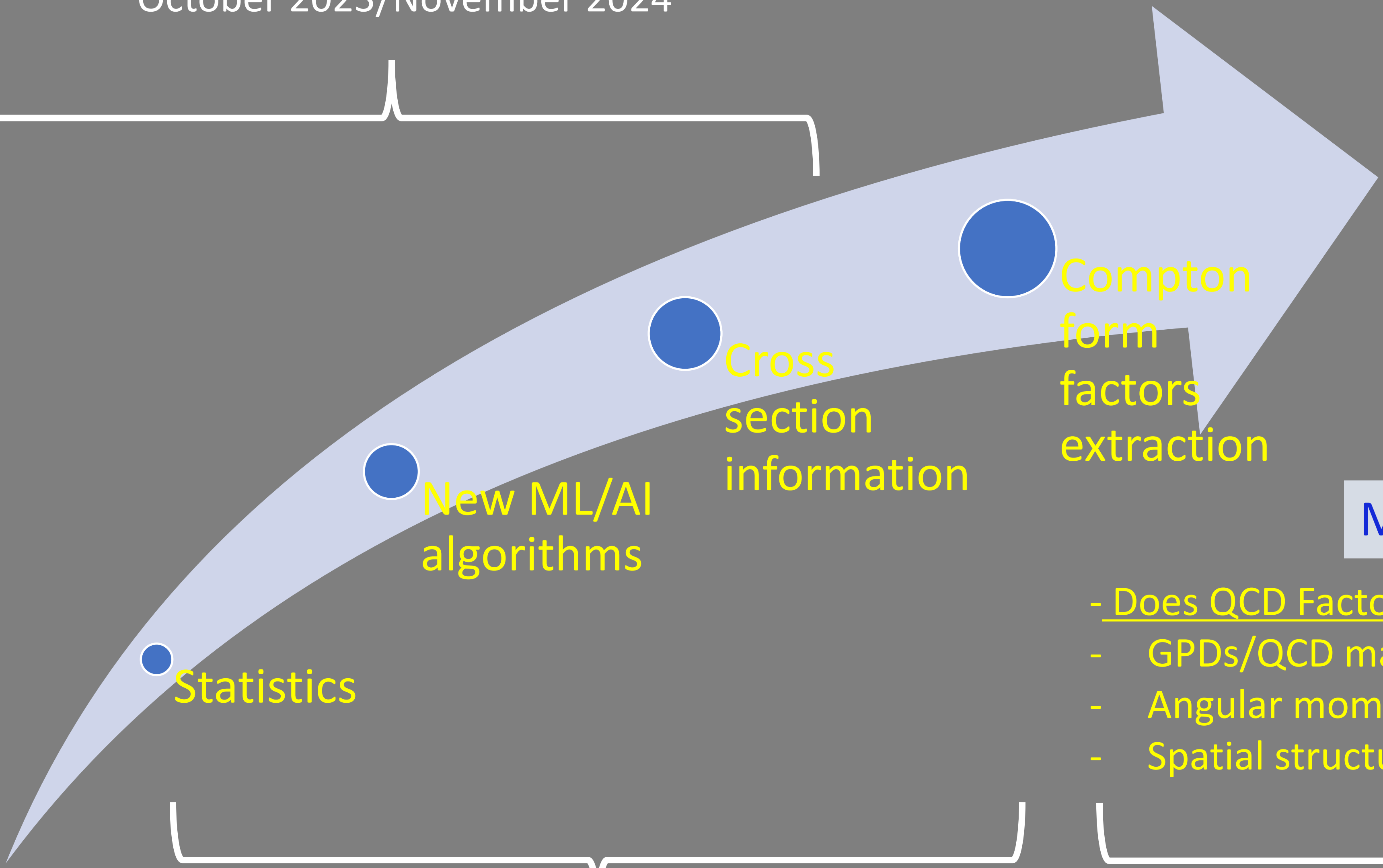
EXCLAIM Collaboration Budget by Institution

Institution	Original Request (k $)	Adjusted Request (k $)
UVA	531.0	442
MSU	233.3	191
NMSU	333.0	280
ODU	344.4	238
Tufts	157	148
VT	230.9	201
Total	1829.6	1500

Institution	Budget Received/Allocated FY 2024 (k$)
UVA	221
MSU	95.5
NMSU	140
ODU	119
Tufts	124
VT	100.5
Total	750

Scope of project/Deliverables timeline

October 2023/November 2024



More theoretical questions

- Does QCD Factorization work at the scale of data?
- GPDs/QCD matrix elements
- Angular momentum
- Spatial structure

1st set of inverse problem techniques

2nd inverse problem set

Authors	Papers	Submission Status	Dates
Anusha Singireddy, Andrew Dotson et al. (EXCLAIM Collaboration)	Generalized Parton Distributions from Symbolic Regression	In preparation, arXiv:2412.XXXXX	
Douglas Q. Adams et al. (EXCLAIM Collaboration)	Likelihood and Correlation Analysis of Compton Form Factors for Deeply Virtual Exclusive Scattering on the Nucleon	Submitted to PRD arXiv:2410.23469v1	30 Oct 2024
Fayaz Hossen et al. (EXCLAIM Collaboration)	Variational autoencoder inverse mapper for extraction of Compton form factors: Benchmarks and conditional learning	Submitted to PRX arXiv:2408.11681v1	21 Aug 2024
Almaeen et al. (EXCLAIM Collaboration)	VAIM-CFF: A variational autoencoder inverse mapper solution to Compton form factor extraction from deeply virtual exclusive reactions	Submitted to EPJC arXiv:2405.05826v2	10 Aug 2024
Almaeen et al. (EXCLAIM Collaboration)	Benchmarks for a Global Extraction of Information from Deeply Virtual Exclusive Scattering	arXiv:2207.10766v1	21 Jul 2022

Authors	Proceedings	Venue	Dates
Simonetta Liuti et al. (EXCLAIM Collaboration)	AI for Nuclear Physics: the EXCLAIM project (arXiv:2408.00163v2, submitted to JINST)	AI4EIC	31 July 2024
Simonetta Liuti et al. (EXCLAIM Collaboration)	Defining the Benchmarks for the Extraction of Information from Polarized Deep Exclusive Scattering	PoS SPIN2023	30 July 2024

Speakers	Talks, Seminars, and Colloquia	Venue	Dates
Simonetta Liuti	Highlights from the EXCLAIM collaboration	PDFLattice Workshop at Jefferson Lab	19 November 2024
Andrew Dotson	Learning Hadron Structure	Florida International University	8 November 2024
Simonetta Liuti	Extracting information on the nucleon 3D structure from data: likelihood analysis of deeply virtual Compton scattering as a case study	REF 2024 Institut de Physique Theorique, Saclay	16 October 2024
1. Marija Cuic 2. Simonetta Liuti	1. GPD analysis 2. GPD extraction and EXCLAIM project	International School of Hadron Femtography at Jefferson Lab	16-17, 23 September 2024
1. Douglas Q. Adams 2. Marija Cuic	1. Maximum Likelihood and MCMC combined analysis of DVCS from the nucleon 2. Symbolic Regression as a tool for studying GPDs	Towards improved hadron tomography with hard exclusive reactions at ECT* Trento	8-9 August 2024
Simonetta Liuti	The EXCLAIM collaboration Highlights and Recent Results on Deeply Virtual Exclusive Processes	Inverse Problems and Uncertainty Quantification in Nuclear Physics at the INT	12 July 2024
Marija Cuic	Towards benchmarking the partonic structure of the proton by the EXCLAIM collaboration	EIC User Group Early Careeer Workshop, Lehigh University	22 July 2024
Marija Cuic	First Results on Deeply Virtual Exclusive Experiments from the EXCLAIM collaboration	QCD Evolution 2024	31 May 2024
Yaohang Li	Exclusives with Artificial Intelligence and Machine learning	DIS2024 at Grenoble	9 April 2024
Simonetta Liuti	AI for Nuclear Physics	Florida International University	3 November 2023

Presenters	Posters	Venue	Dates
Andrew Dotson	Symbolic Regression of Generalized Parton Distributions using PySR*	PDFLattice Workshop at Jefferson Lab	18 November 2024
Ho Jang	End to End Problem		
Adil Umar Khawaja	Flexible global parametrization of generalized parton distributions from deep inelastic scattering data and lattice QCD moments		
Saraswati Pandey	Likelihood and Markov Chain Monte Carlo combined Analysis of spin structure function g_1^p		
Zaki Panjsheeri	Neural network GPDs (NNGPDs)		
Emmanuel Ortiz-Pacheco, Liam Hockley	Machine Learning for Gluon Parton Distribution Functions		

Participants	Activities	Venue	Dates
Simonetta Liuti Zaki Panjsheeri	Institute for Artificial Intelligence and Fundamental Interactions (IAIFI) Workshop IAIFI Summer School	Massachusetts Institute of Technology	5-16 August 2024
Saraswati Pandey	2024 CFNS EIC Summer School	Stony Brook University	2-15 June 2024

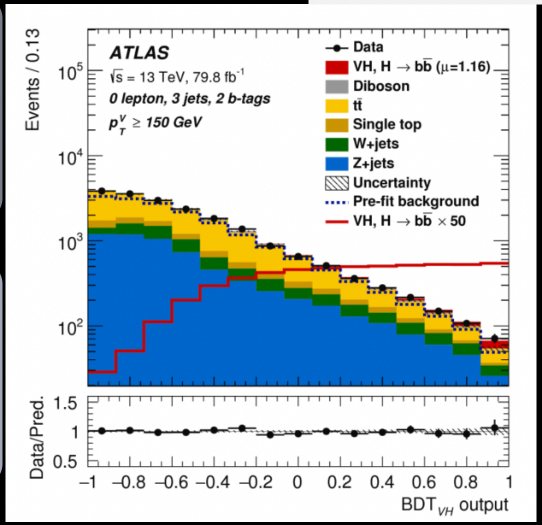
EXCLAIM Collaboration Meetings	Dates
Southeastern Universities Research Association (SURA) in Washington, DC	21-22 October 2023
Southeastern Universities Research Association (SURA) in Washington, DC	20-21 May 2024
Old Dominion University	16-17 November 2024

Standard Approach to ML/AI in Physics

Standard Approaches: Industrial Machine Learning (ML) tools are used/adapted to aid computation in Nuclear/Particle Physics

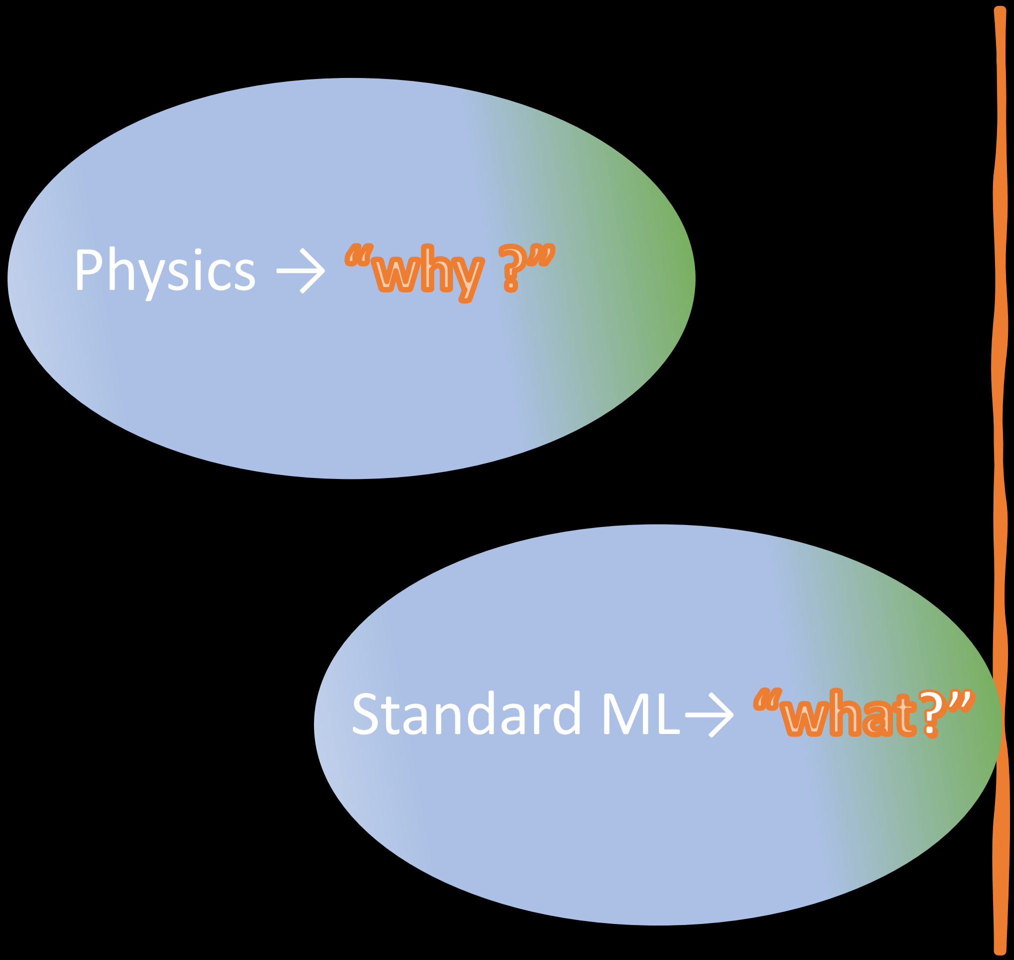
Example: Ensemble learning methods such as Boosted Decision Trees (BDT) invented for image recognition/object detection used in self-driving cars are used to identify b-hadrons

$H \to b\bar{b}$



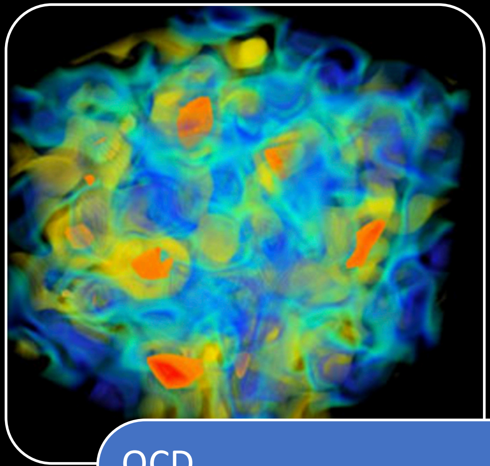
https://atlas.cern/

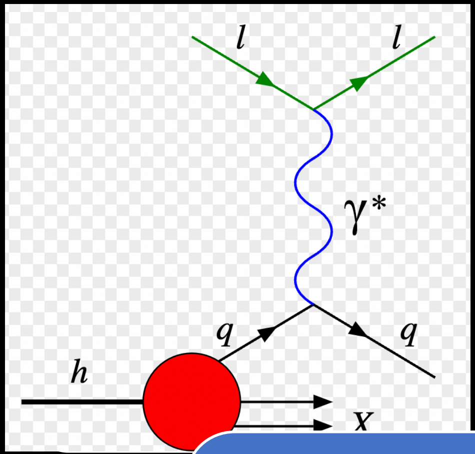
To address the "why"?

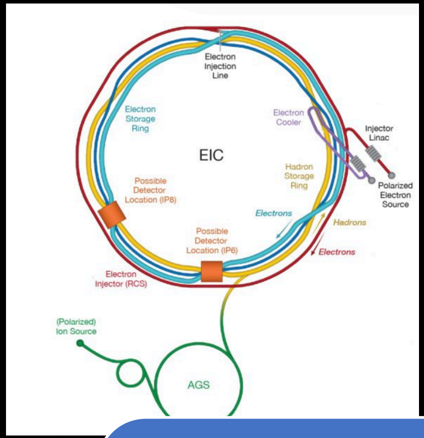


- We introduce physics aware NNs as explainable ML models: C-VAIM
- Symbolic Regression: ML algorithm where data are modeled directly with analytic expressions. Direct interpretability
- Explainable/interpretable models are necessary for the 3D nuclear problem directly enabling discovery of laws in Nuclear and Particle Physics
- Not just a set of advanced computational tools: It is about finding a common language between physics and AI

Forward Problem







QCD
- quark field
- gluon field
- correlation functions

Observables
- Form Factors
- Structure Functions
- Compton Form Factors
- Fragmentation functions
- …

Data/Measurements

Inverse Problem

- **Interpretability + predictions**

The goal in physics is to extract information as accurately as possible from data/concise understandable equations/generalizable

- **Only Predictivity**

The goal of ML is to obtain statistical models that can make predictions from the data/no interpretability

- **Inverse problem**

To extract model information from data we need to define a bridge between CS experts and physicists that is centered on how we define and treat the respective data uncertainty and correlations while being aware that their objective functions differ

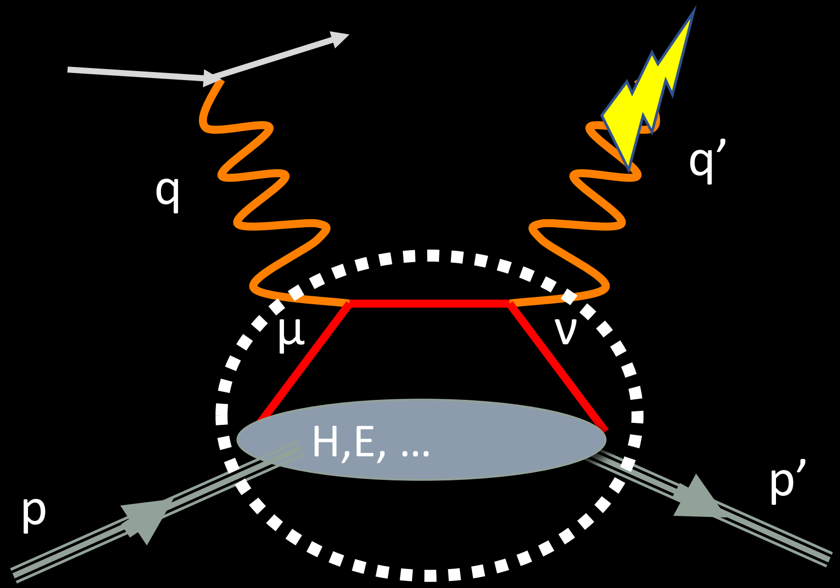
An immense potential!

Through ML we will be able to uncover **new physics relations/laws**

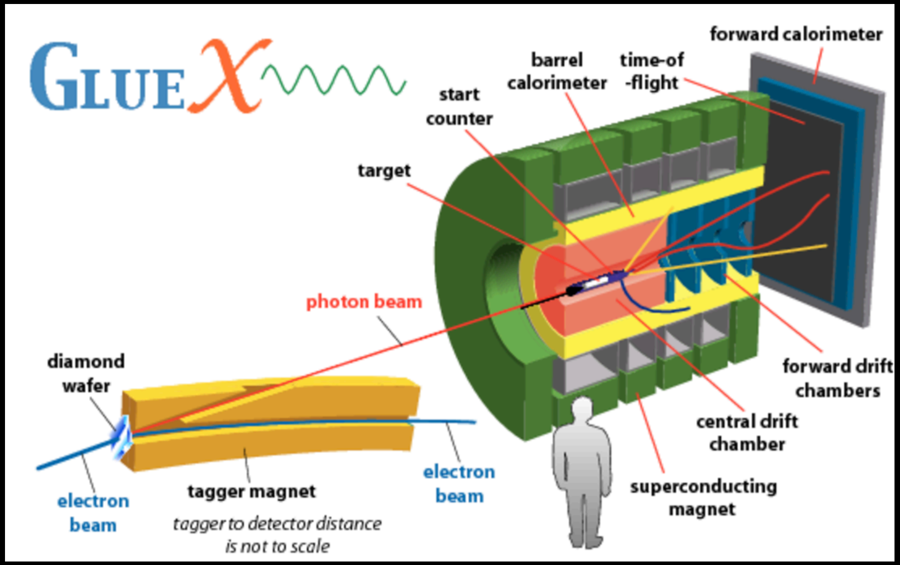
ML emergent behavior is decades from being understood!

Physics Case: Extracting information from exclusive deeply virtual scattering

QCD matrix element



Experiment



Generalized Parton Distributions

Moments → Energy Momentum Tensor Form factors

Angular momentum, Mass

Twist three GPD Physical interpretation at the core of spin puzzle

$$
\begin{array}{ccccc}
J_L & = & L_L & + & S_L \\
\frac{1}{2}\int dx\, x(H+E) & = & \boxed{\int dx\, x(\widetilde{E}_{2T}+H+E)} & + & \frac{1}{2}\int dx\, \widetilde{H} \\
 & = & \boxed{-\int dx\, F_{14}^{(1)}} & + & \frac{1}{2}\int dx\, \widetilde{H}
\end{array}
$$

A. Rajan, M. Engelhardt and S. Liuti, Phys. Rev. D98, 074022 (2018)
A. Rajan, A. Courtoy, M. Engelhardt and S. Liuti, Phys. Rev. D94, 034041 (2016)
M. Rodekamp, M. Engelhardt, J.R. Green, S. Krieg, S. Liuti, S. Meinel, J.W. Negele, A. Pochinsky and S. Syritsyn, Phys. Rev. D 109, 074508 (2024)

*Twist 3 GPD notation from Meissner, Metz and Schlegel, JHEP(2009)

Transverse Angular Momentum Sum Rule

O. Alkassasbeh, M. Engelhardt, SL and A. Rajan, https://arxiv.org/abs/2410.21604

$$\underbrace{\frac{1}{2}\int dx\, x(H+E) - \frac{1}{2}\int dx\, \mathcal{M}_T}_{J_T} = \underbrace{\int dx\, x\big(\widetilde{E}_{2T} + H + E + \frac{H_{2T}}{\xi}\big)}_{L_T} + \underbrace{\frac{1}{2}\int dx\, g_T - \frac{1}{2}\int dx\, x\mathcal{A}_T}_{S_T}$$

Extract Compton Form Factors from Leading order parametrization of DVCS cross section

Compton Form factors

$$|T_{UU}^{BH}|^2 = \frac{\Gamma}{t}\left[A_{UU}^{BH}(F_1^2+\tau F_2^2) + B_{UU}^{BH}\tau G_M^2(t)\right]$$

$$|T_{UU}^{\mathcal{I}}|^2 = \frac{\Gamma}{Q^2 t}\left[A_{UU}^{\mathcal{I}}\Re e\,(F_1\mathcal{H}+\tau F_2\mathcal{E}) + B_{UU}^{\mathcal{I}}G_M\Re e\,(\mathcal{H}+\mathcal{E}) + C_{UU}^{\mathcal{I}}G_M\Re e\widetilde{\mathcal{H}}\right]$$

$$|T_{LU}^{\mathcal{I}}|^2 = \frac{\Gamma}{Q^2 t}\left[A_{LU}^{\mathcal{I}}\Im m\,(F_1\mathcal{H}+\tau F_2\mathcal{E}) + B_{LU}^{\mathcal{I}}G_M\Im m\,(\mathcal{H}+\mathcal{E}) + C_{LU}^{\mathcal{I}}G_M\Im m\widetilde{\mathcal{H}}\right]$$

$$\begin{aligned}|T_{UU}^{DVCS}|^2 = \frac{\Gamma}{Q^2}\frac{2}{1-\epsilon}\Bigg[&(1-\xi^2)\left[(\Re e\mathcal{H})^2+(\Im m\mathcal{H})^2+(\Re e\widetilde{\mathcal{H}})^2+(\Im m\widetilde{\mathcal{H}})^2\right]\\ &+\frac{t_o-t}{4M^2}\left[(\Re e\mathcal{E})^2+(\Im m\mathcal{E})^2+\xi^2(\Re e\widetilde{\mathcal{E}})^2+\xi^2(\Im m\widetilde{\mathcal{E}})^2\right]\\ &-2\xi^2\left(\Re e\mathcal{H}\,\Re e\mathcal{E}+\Im m\mathcal{H}\Im m\mathcal{E}+\Re e\widetilde{\mathcal{H}}\,\Re e\widetilde{\mathcal{E}}+\Im m\widetilde{\mathcal{H}}\Im m\widetilde{\mathcal{E}}\right)\Bigg]\end{aligned}$$

Azimuthal angle ϕ dependent coefficients

spin non-flip

spin flip

mixed

- B. Kriesten et al, *Phys.Rev. D* 101 (2020)
- B. Kriesten and S. Liuti, *Phys.Rev. D105 (2022),* arXiv 2004.08890
- B. Kriesten and S. Liuti, Phys. Lett. B829 (2022), arXiv:2011.04484

First Inverse Problem!

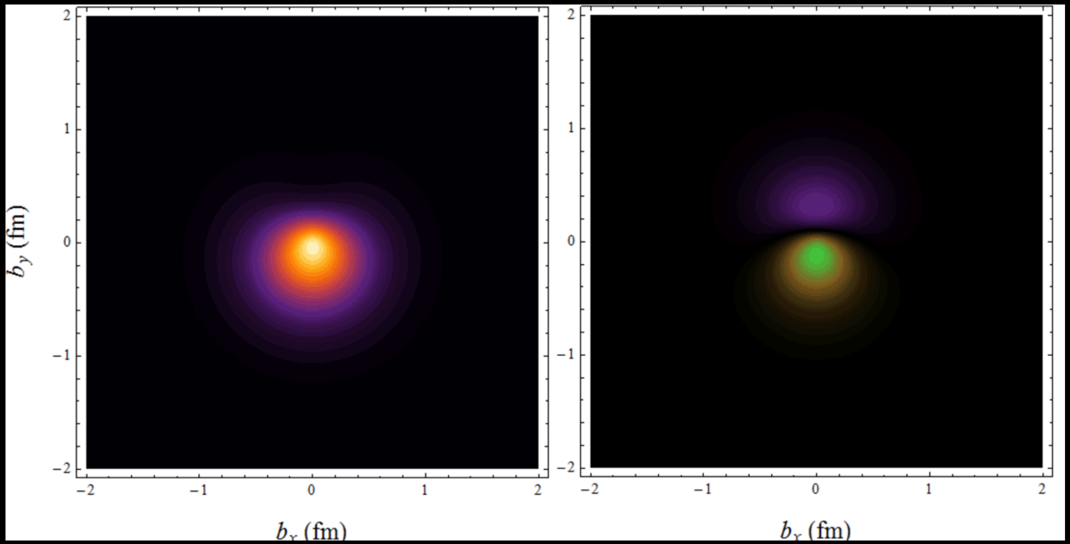
$$\underbrace{\mathrm{CFF}}_{\text{Assume known}} = \int (\mathrm{QCDKernel}) \times \mathrm{GPD}$$

Second Inverse Problem!

3D Coordinate Space Representation requires a Fourier transform

Observables from DVES matrix elements can be Fourier transformed from momentum space into coordinate space, providing insight into the spatial distributions of quarks and gluons inside the proton, besides matter and charge distributions.

$$\underbrace{\boxed{\mathcal{H}^q(X,0,b_T)}}_{\text{Wigner phase space distribution}} = \int \frac{d^2\Delta_T}{(2\pi)^2} \underbrace{\boxed{H^q(X,0,\Delta_T)}}_{\text{GPD}} e^{-i\Delta_T\cdot b_T}$$



H-W Lin,

https://web.pa.msu.edu/people/hwlin/research.html

EXCLAIM Research Accomplishments

1. Fully constraining Likelihood analysis

2. Inverse Problem techniques: Variational Autoencoder Inverse Mapper (VAIM)

3. Symbolic Regression for Partonic Observables

All these methods share the common goal of going beyond simple regression by understanding the underlying correlations of the system

Graduate Students:

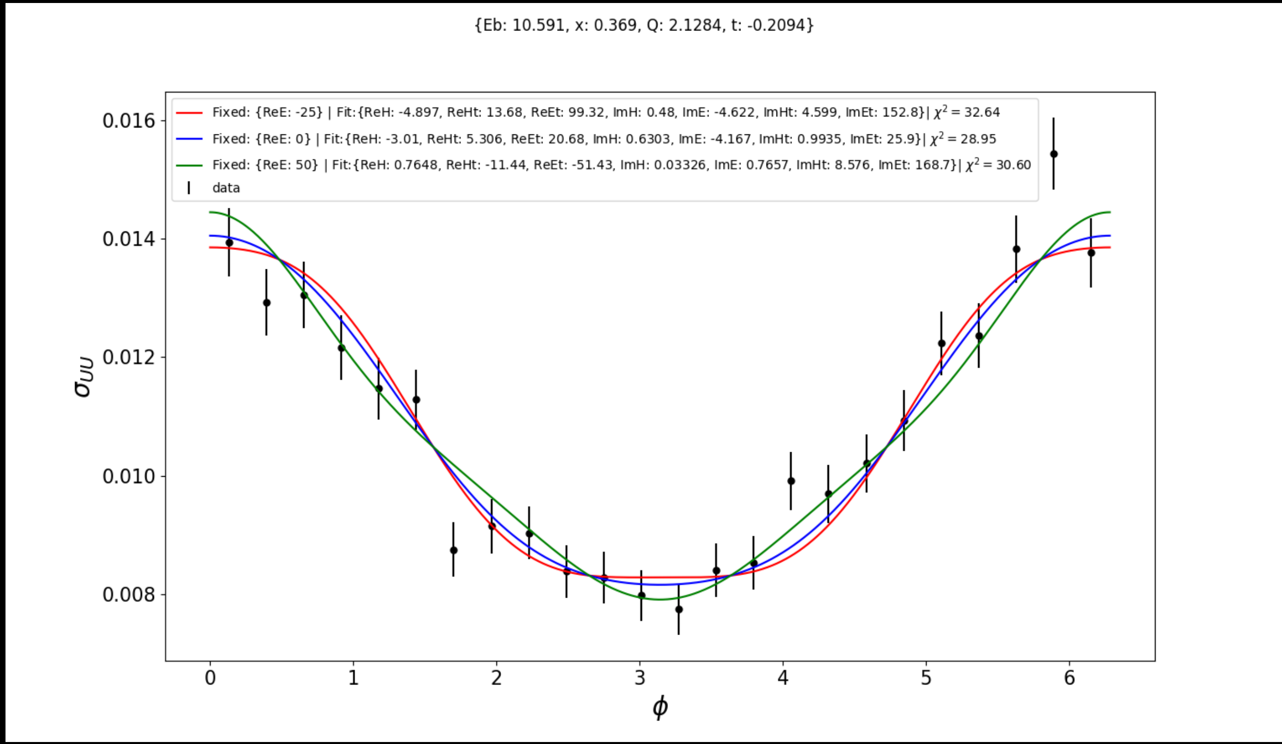
Joshua Bautista, Adil Khawaja, Zaki Panjsheeri

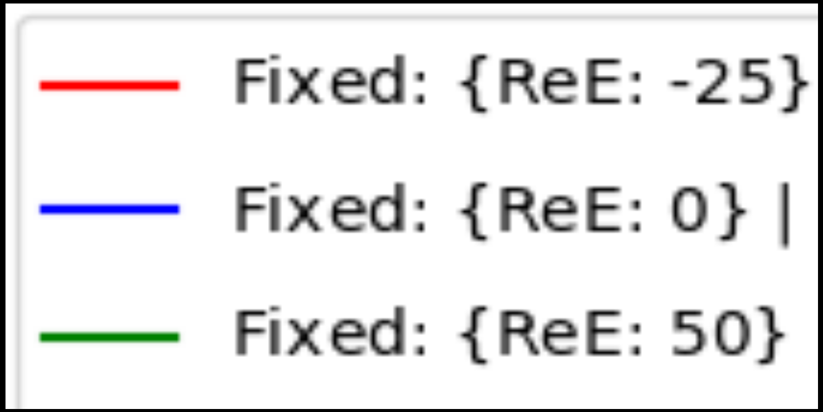
1. Fully Constraining CFFs : Likelihood Analysis

GOAL: Use DVCS data and comparing to cross section model to find CFFs

- Curve fit: A bad result: Encounter a problem 1!
- Definition of the likelihood: Try to fix the problem
- Canonical Likelihood: Reproduces the problem in explainable way
    - Difference method Likelihood
    - Canonical Likelihood
- Encounter a problem 2!: covariance
- Results: Table of CFFs and errors

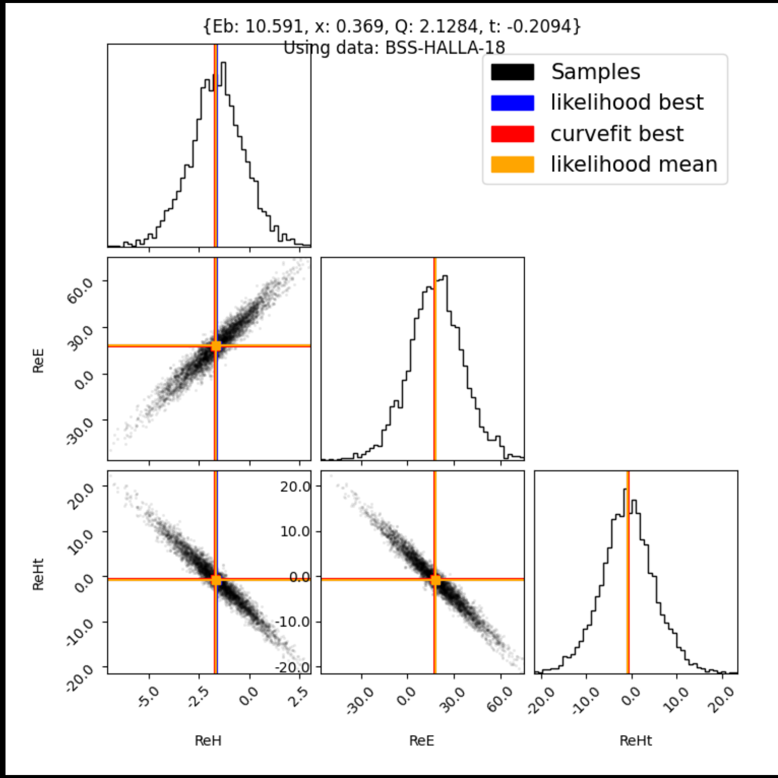
Degeneracy of Curve Fit Results

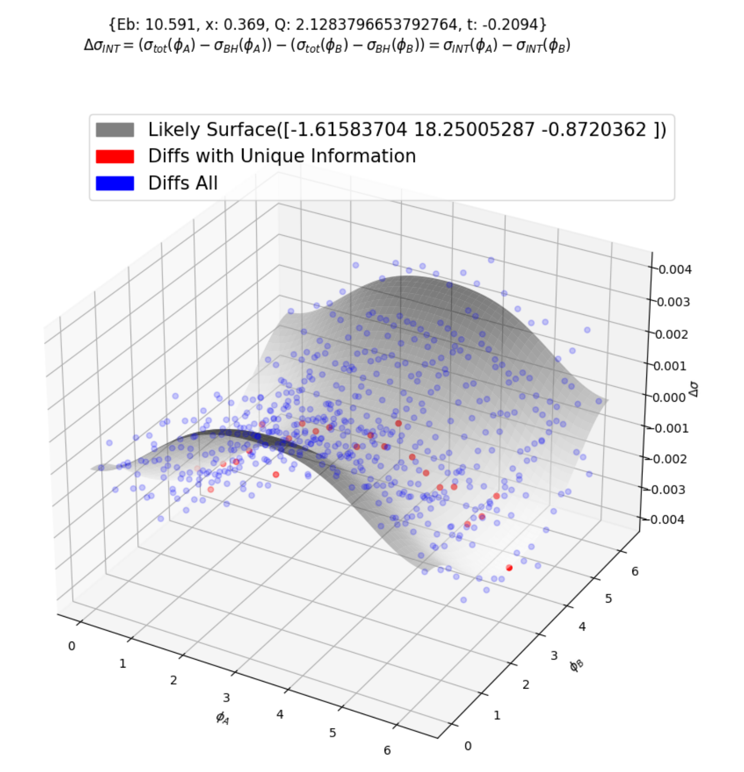




Covariance of CFF Results

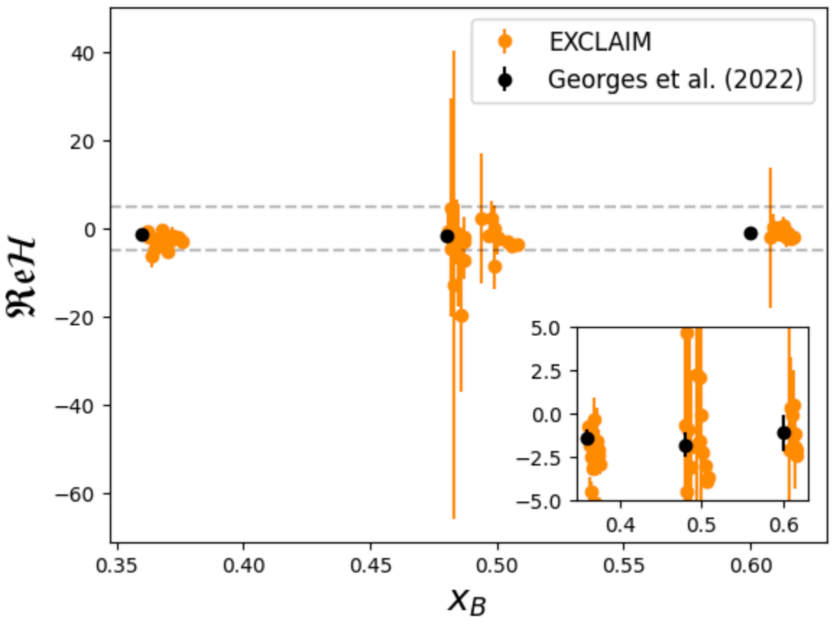
- Here the maximum likelihood is achieved allowing 3 CFFs to vary.
- Only 23 combinations of 2 angles are used.



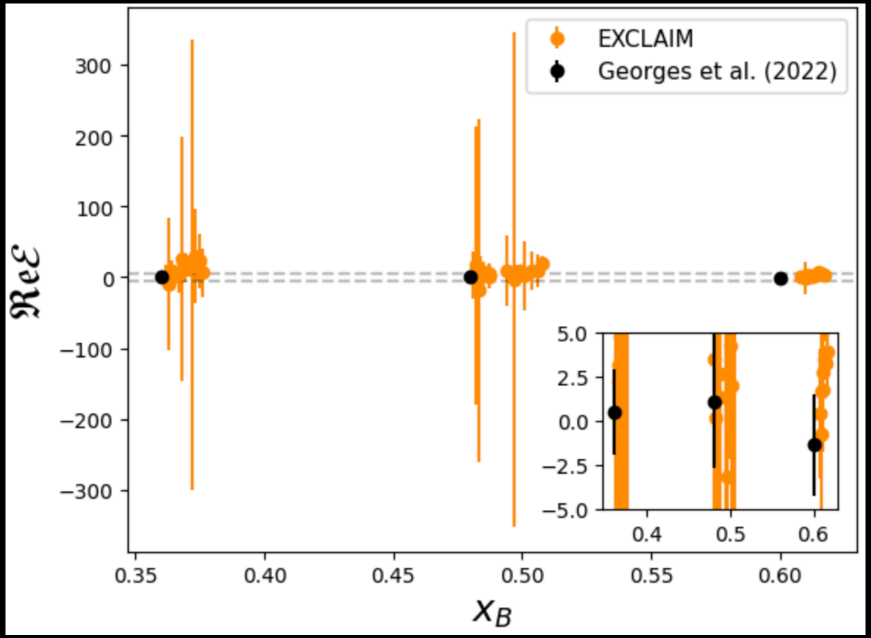


Marginal Posterior CFF Results

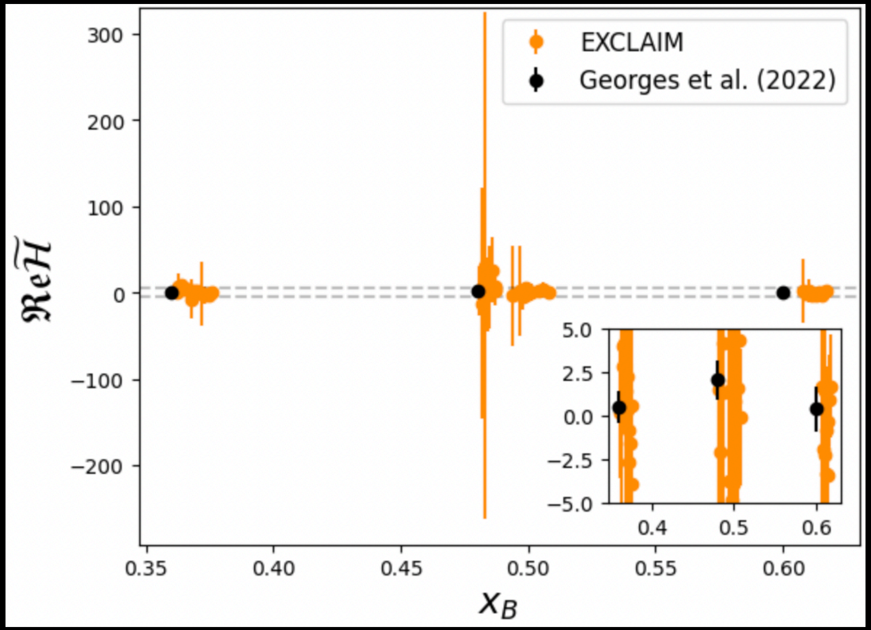
$\mathcal{H}$



$\mathcal{E}$



$\widetilde{\mathcal{H}}$

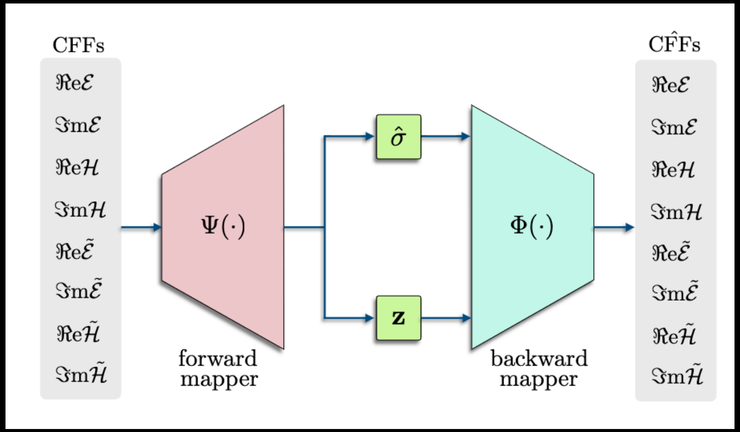


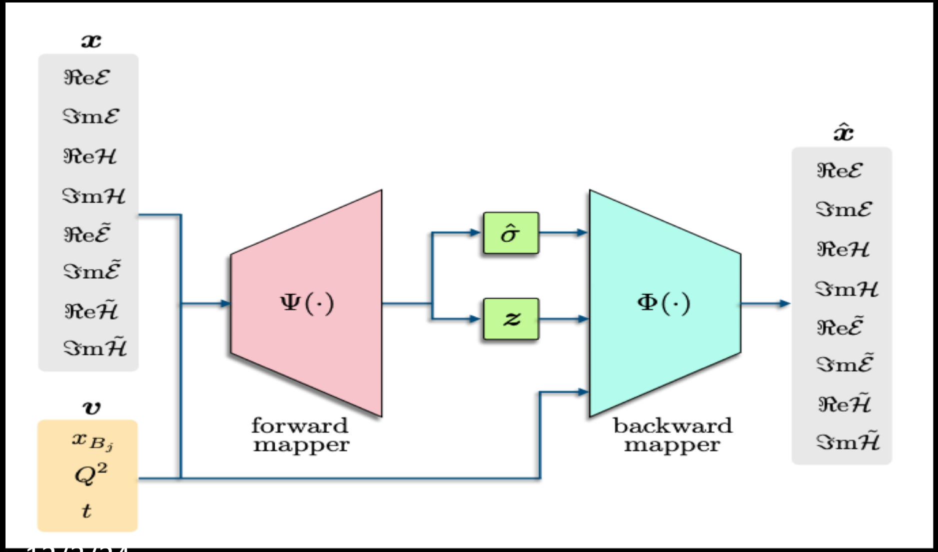
2. Inverse Problem Techniques: VAIM, C-VAIM, MCMC

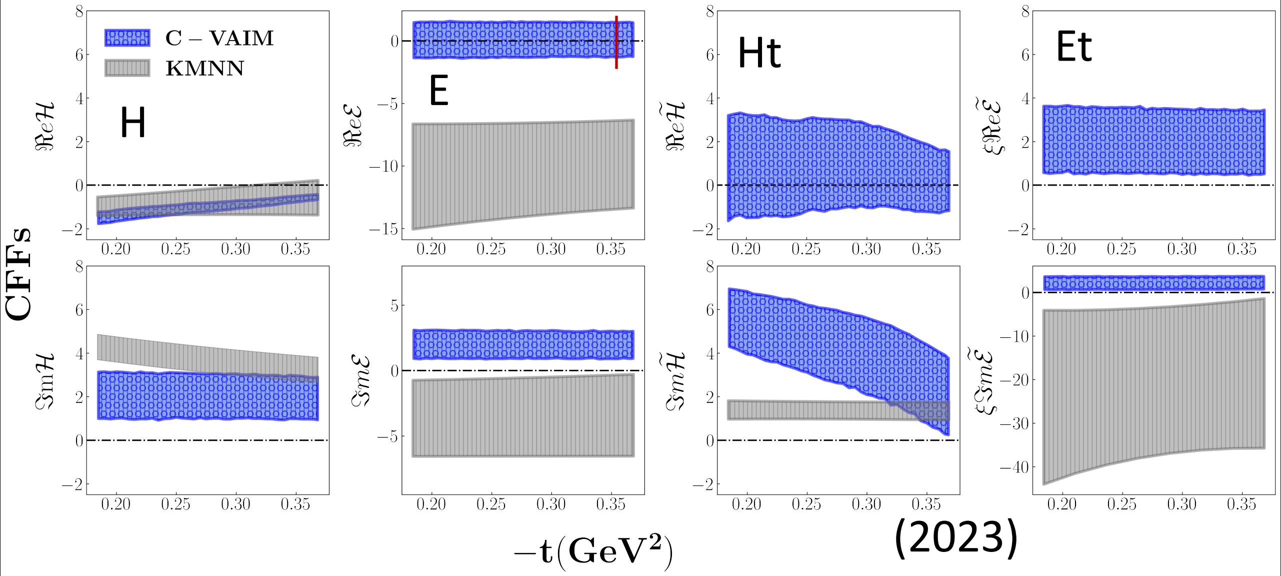
Approaches to find parameters statistically in an underdetermined system:

- Can quantify parameter uncertainty when more parameters than data
- Techniques highly dependent on bounded parameter priors
- These methods give us an initial way to perceive:
  - the correlation between parameters on a complicated model
  - what information is missing (latent space)

A variational autoencoder inverse mapper solution to Compton form factor extraction from deeply virtual exclusive reaction arXiv: 2405.05826

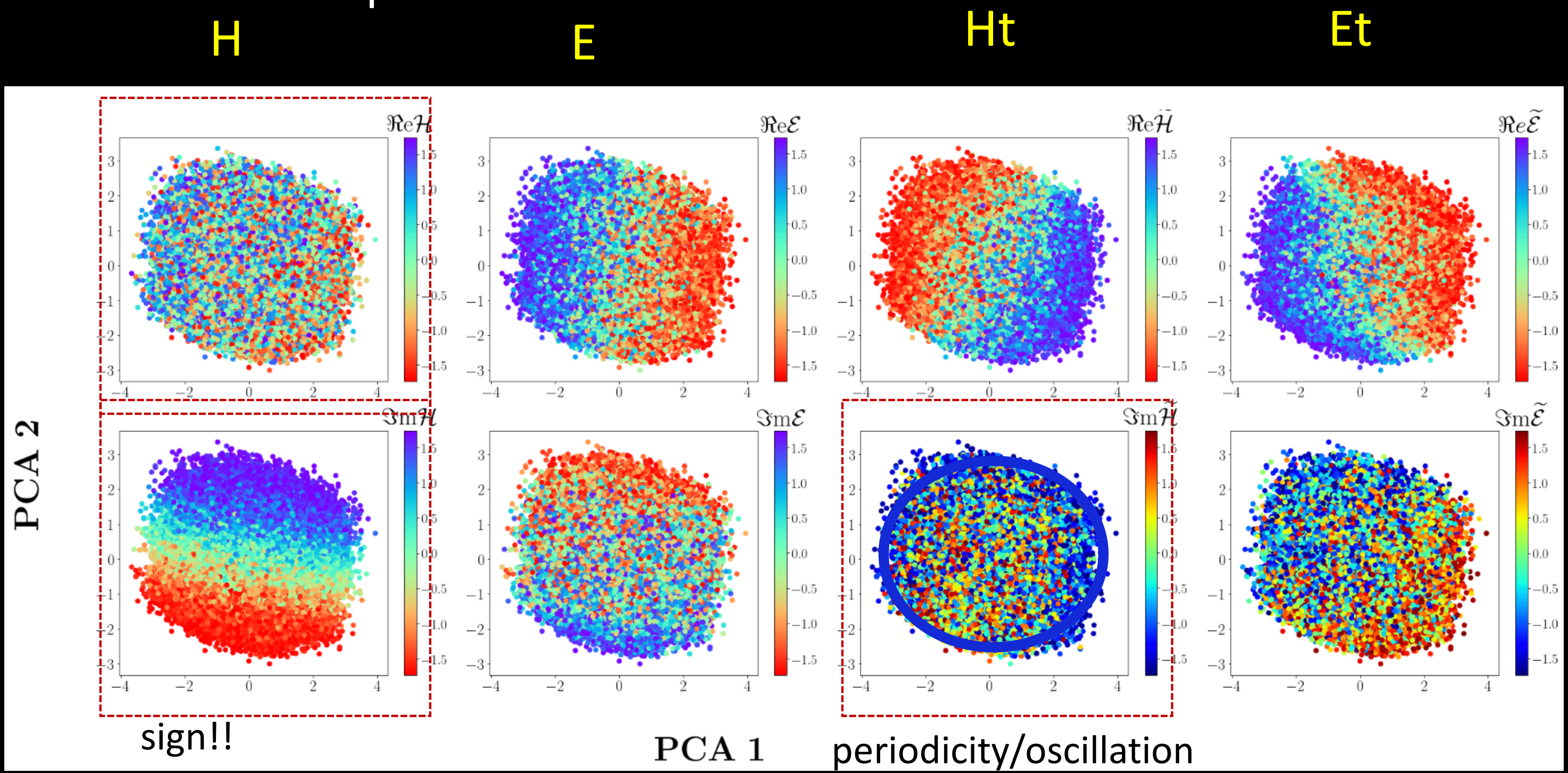






- KMNN, https://arxiv.org/abs/2007.00029

CFFs Analysis of Latent Space



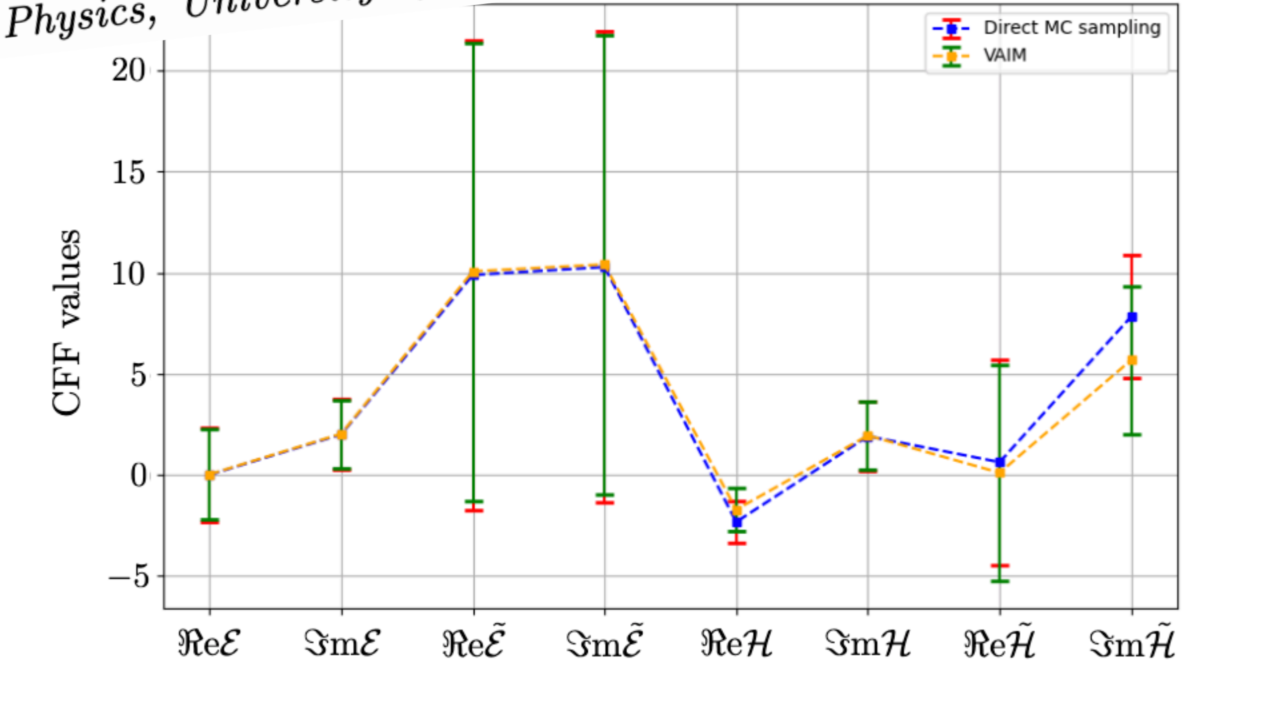
Comparison of MCMC sampling/inverse mapper and VAIM

Variational autoencoder inverse mapper for extraction of Compton form factors: Benchmarks and conditional learning

Fayaz Hossen,[1] Joshua Bautista,[2] Douglas Adams,[2] Yaohang Li,[1,*] Gia-Wei Chern,[2,†] and Simonetta Liuti[2,‡]

[1] *Department of Computer Science, Old Dominion University, Norfolk, VA 23529, USA.*

[2] *Department of Physics, University of Virginia, Charlottesville, VA 22904, USA.*



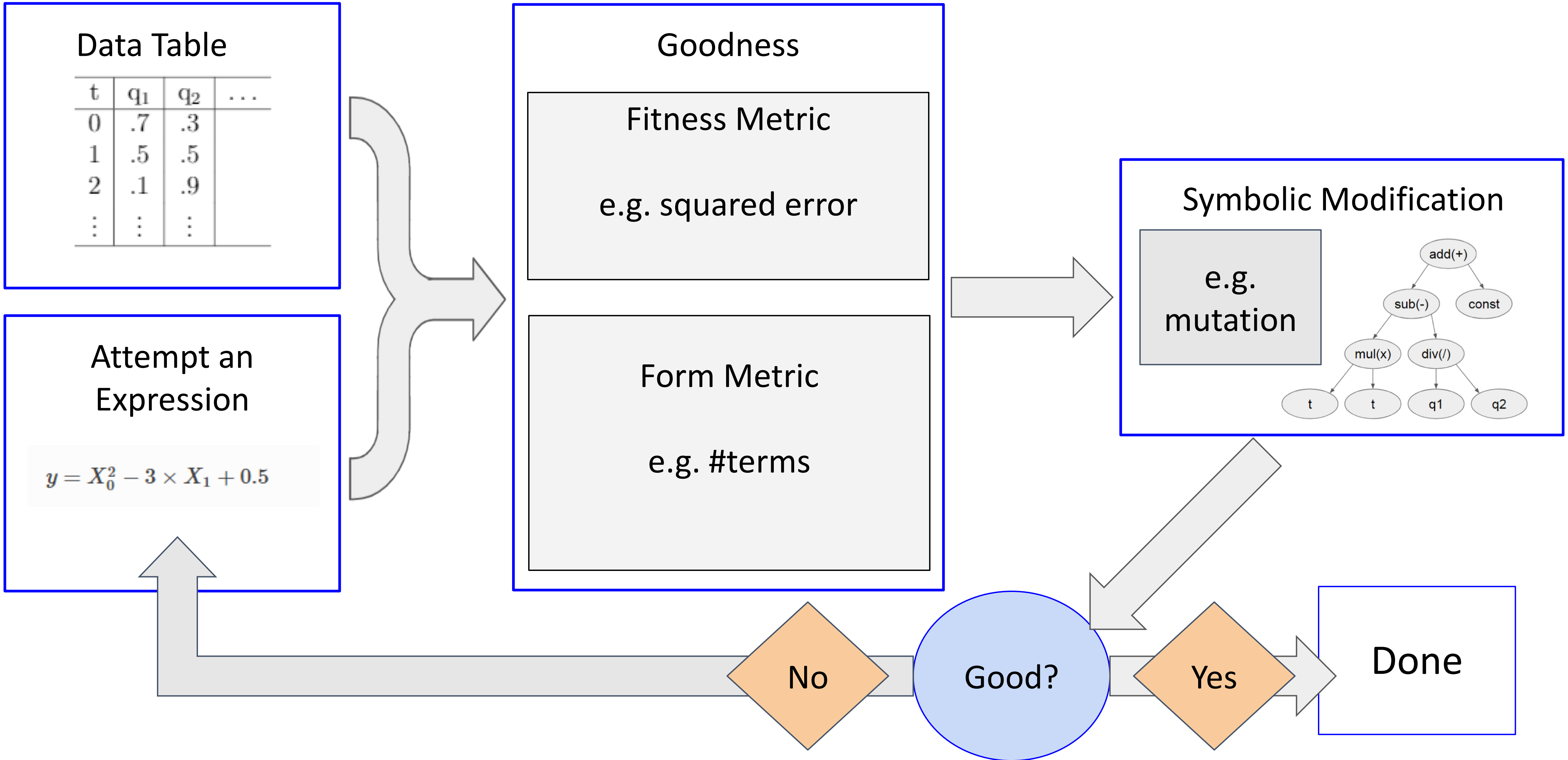
3. Symbolic Regression for Parton Research

1) What is symbolic regression and why do we care?
   a) Spoiler: because then humans can read the answer

2) What are the existing tools out there?
   a) Eureka
   b) Gplearn
   c) AI Feynman
   d) PySr
   e) RL-SR
   f) *Meijer-G-Function (very preliminary)

Graduate Students:
Andrew Dotson (NMSU)
Anusha SingiReddy (ODU)
Zaki Panjsheeri (UVA)

What is symbolic regression (SR)?



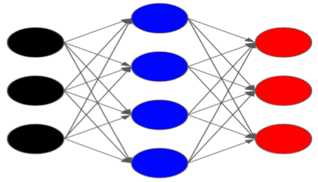
Why bother with SR when we have neural networks?

Which is easier to read? (a.k.a interpretability of AI)

$$y = X_0^2 - 3 \times X_1 + 0.5$$

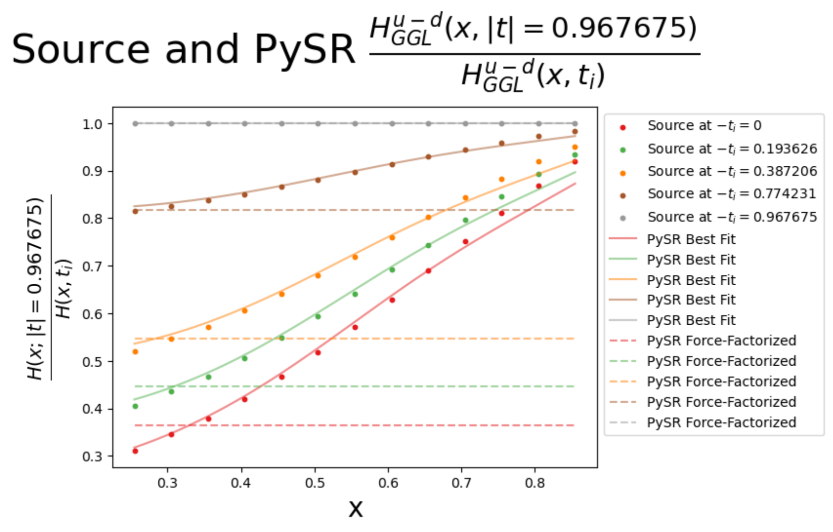
vs

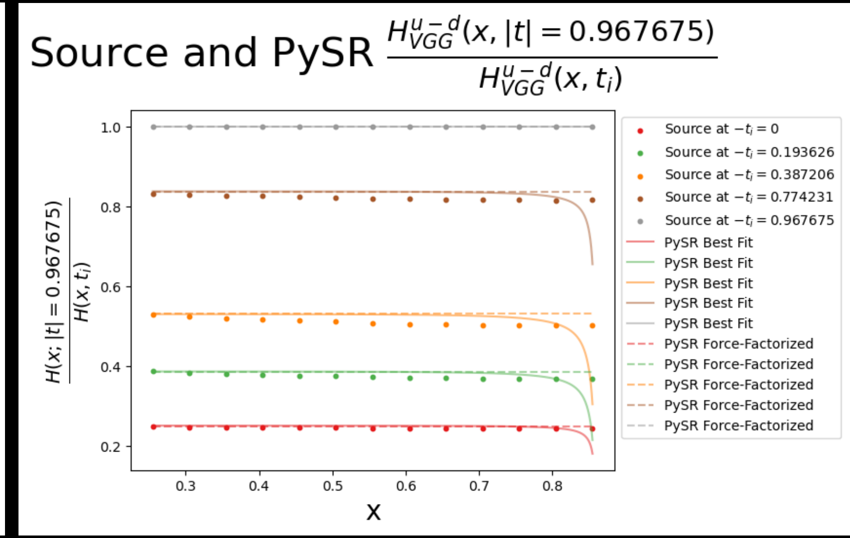
Black Box Regression

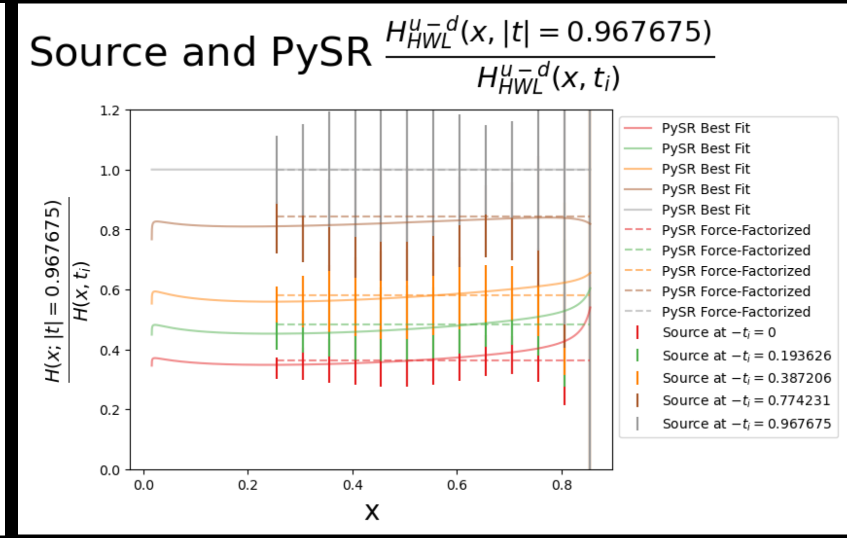


$$q_{i,BB}(t) \approx \ldots$$

Testing x and t factorization (important for spatial configurations!)







phenomenology

lattice

J. Holligan and H-W Lin, Phys.Rev. D110, 034503 (2023)
H-W Lin, Phys. Lett. B824, 136821 (2022)

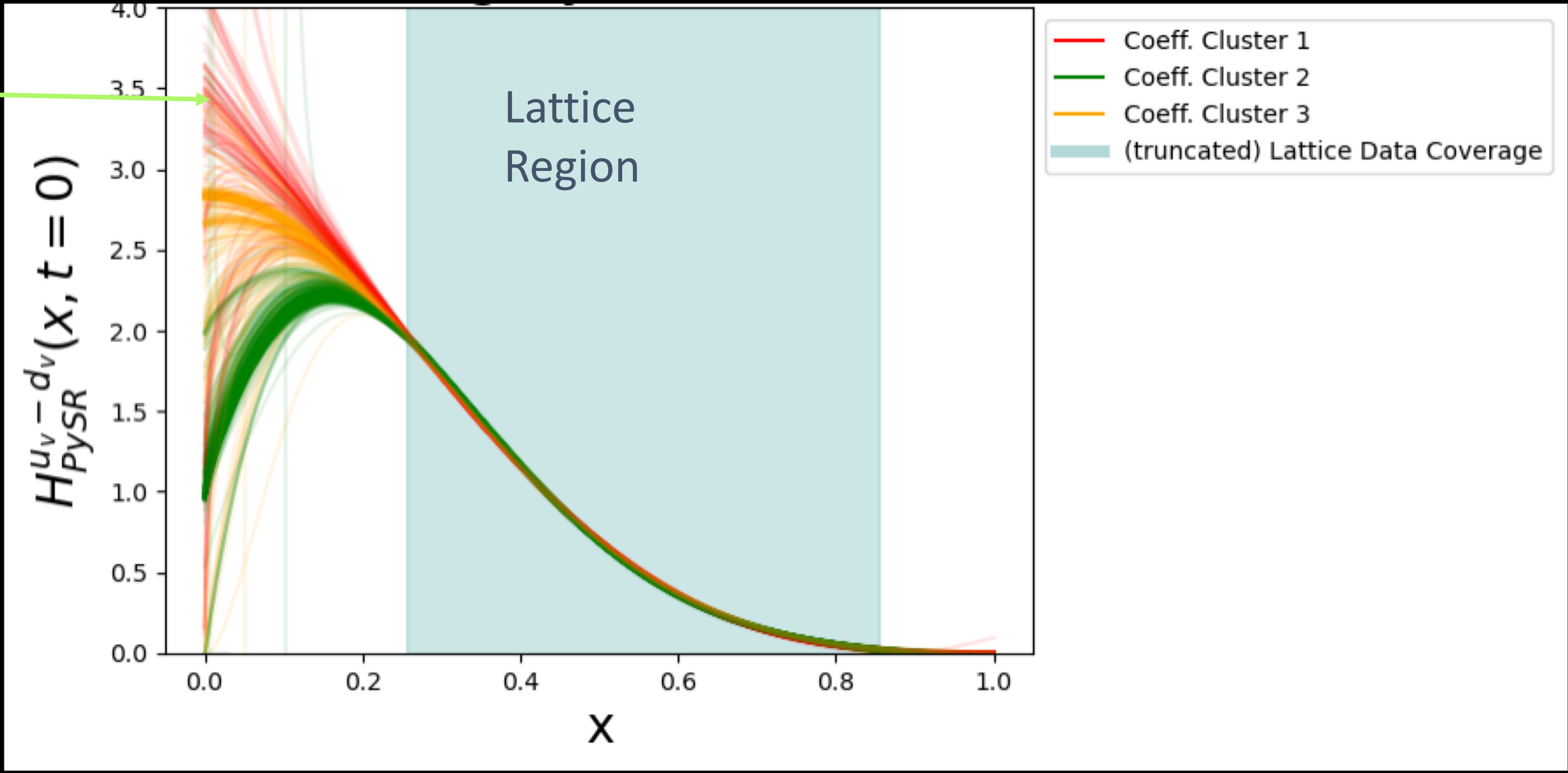
Example of a factorized in x and t form

(MSE constant power)

$$\frac{1.07 \cdot \left(2.78\,(1 - 0.766x)^{3.83} - 0.0437\right)}{-t + 0.603}$$

Novel SR Convergence Clustering

Extrapolation



Why is it important to check x and t factorization?

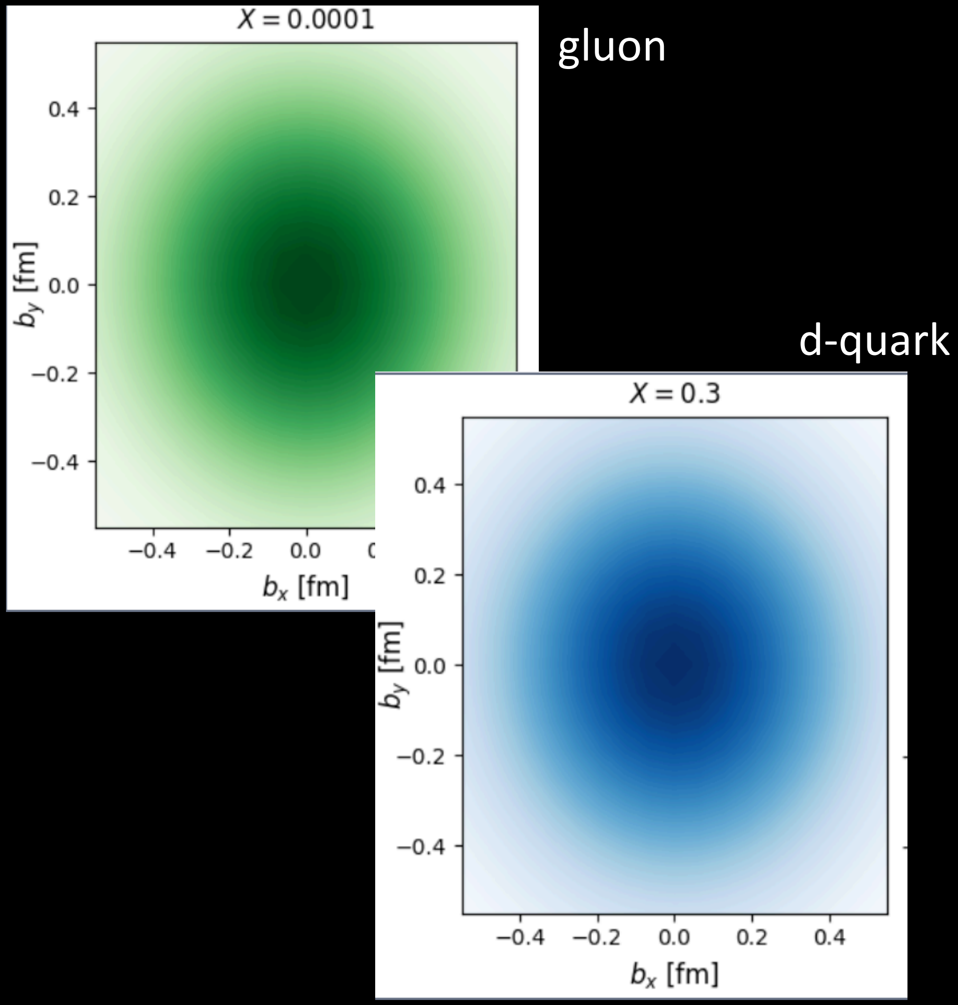
Consequences on the 3D Coordinate Space picture

GPDs can be Fourier transformed from momentum space into coordinate space, providing insight into the spatial distributions of quarks and gluons inside the proton, besides matter and charge distributions.

Slice of Wigner phase space distribution

$$\underbrace{\mathcal{H}^q(X,0,b_T)}_{\text{Slice of Wigner phase space distribution}} = \int \frac{d^2\Delta_T}{(2\pi)^2} \underbrace{H^q(X,0,\Delta_T)}_{\text{GPD}} e^{-i\Delta_T\cdot b_T}$$

GPD



With Z. Panjsheeri and J. Bautista

Gluon and quark matter density radius

$$< b_T^2 >^q (X) = \frac{\int_0^\infty d^2 b_T b_T^2 \mathcal{H}^q(X,0,b_T)}{\int_0^\infty d^2 b_T \mathcal{H}^q(X,0,b_T)}$$

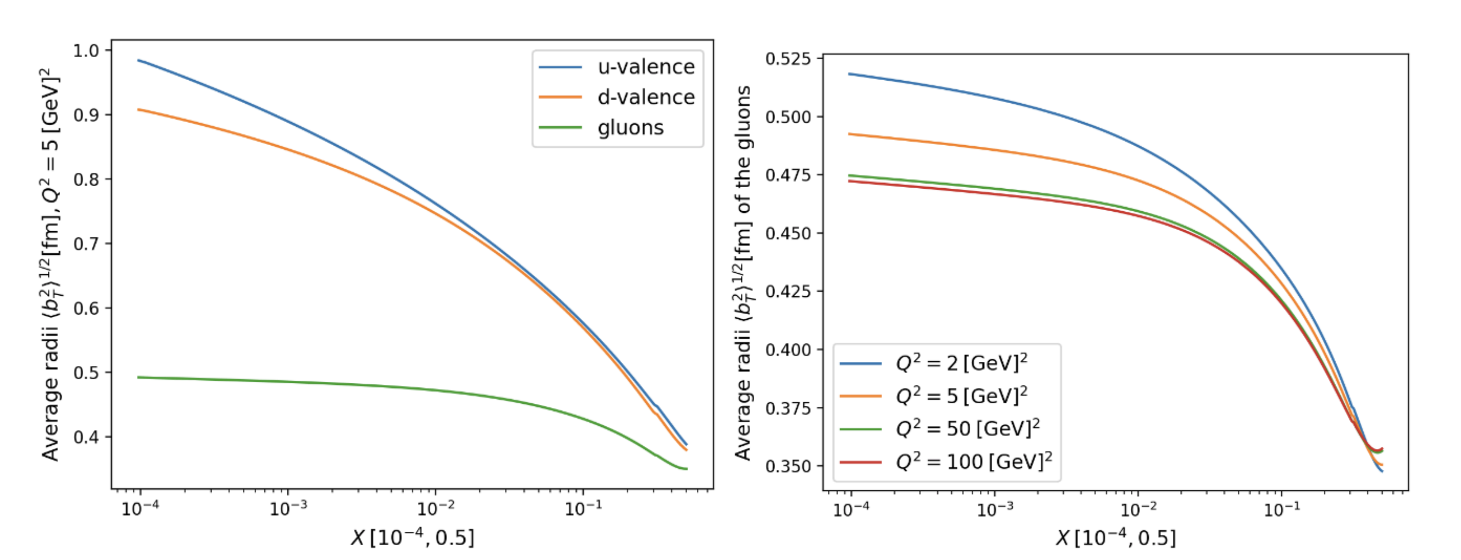
Bautista, Panjsheeri, SL (2024)

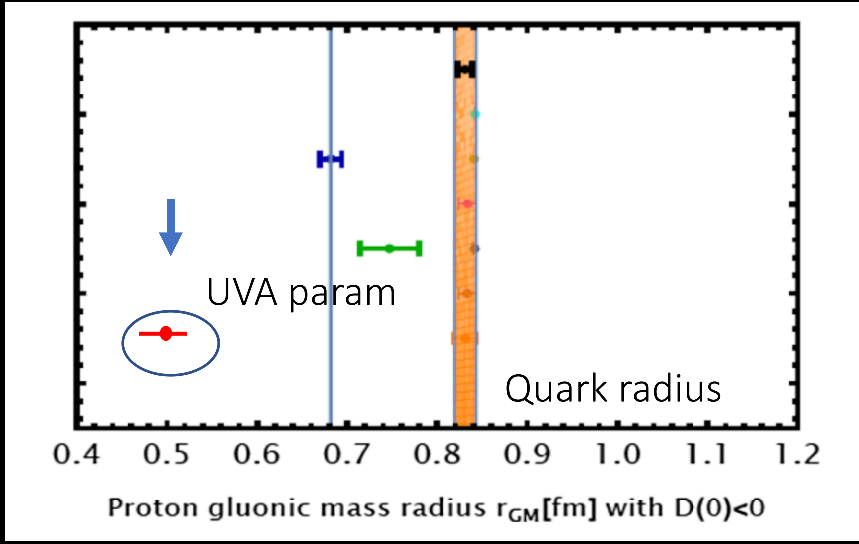
Compare to lattice and AdS/CFT integrated value

K. Mamo and I. Zaeed

PRD106, 086004 (2022)

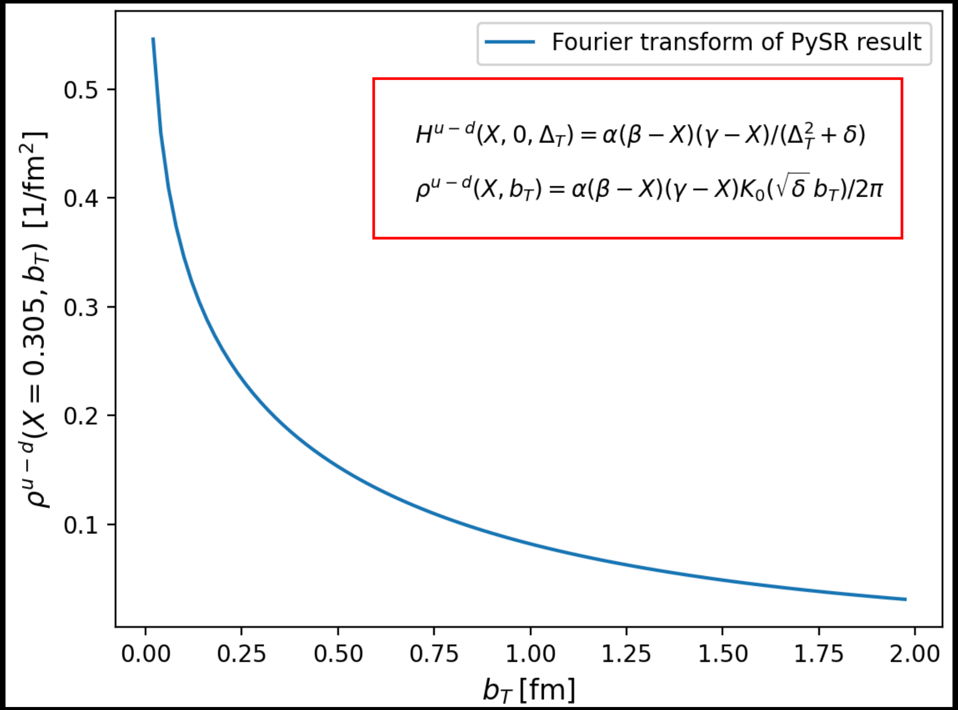
LQCD: Detmold and Shanahan

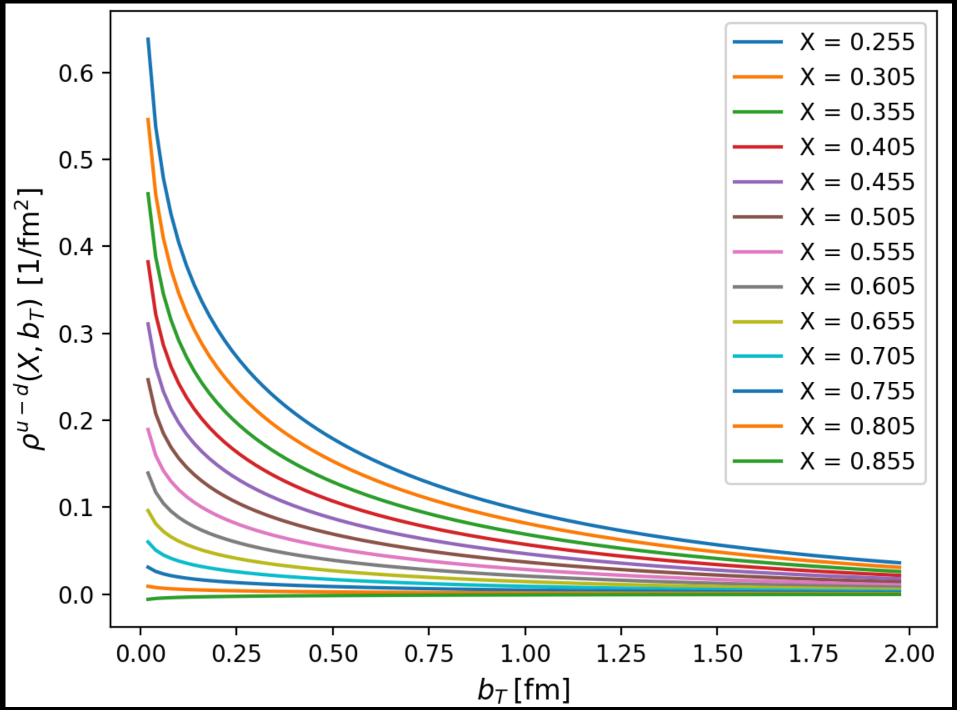




arXiv:2405.05842

From SR Analysis





Conclusions

1. A successful reconstruction of the **spatial structure of the proton** (and all of its mechanical properties) relies on our ability to understand the **cross section** for **all the various DVES processes**

2. This implies solving **multiple inverse problems**

3. We have defined a path to extract the **observables** from experiment that allows us to fully take into account UQ from data and ab initio QCD calculations

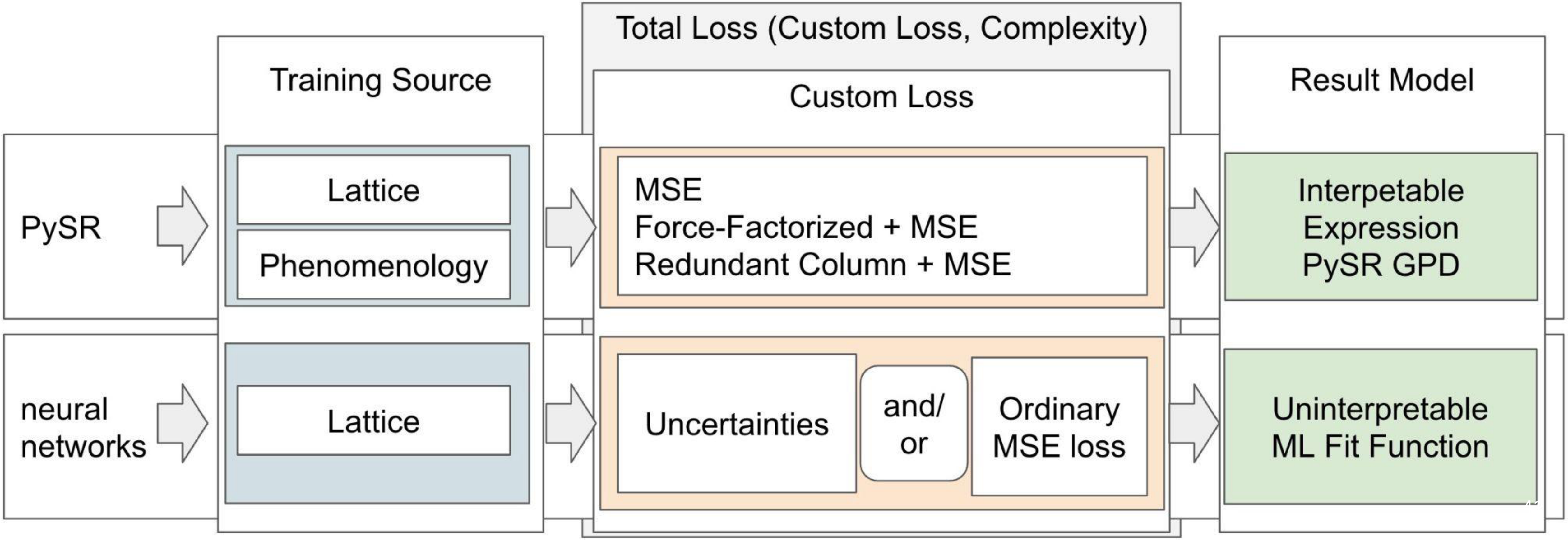
4. Bringing interpretability and benchmarking to AI tools is a necessity for us to progress faster towards understanding the 3D picture of the proton

5. Obtaining spatial images of the proton including UQ is feasible using AI/ML to extend the momentum transfer reach for an accurate Fourier transformation

Back up

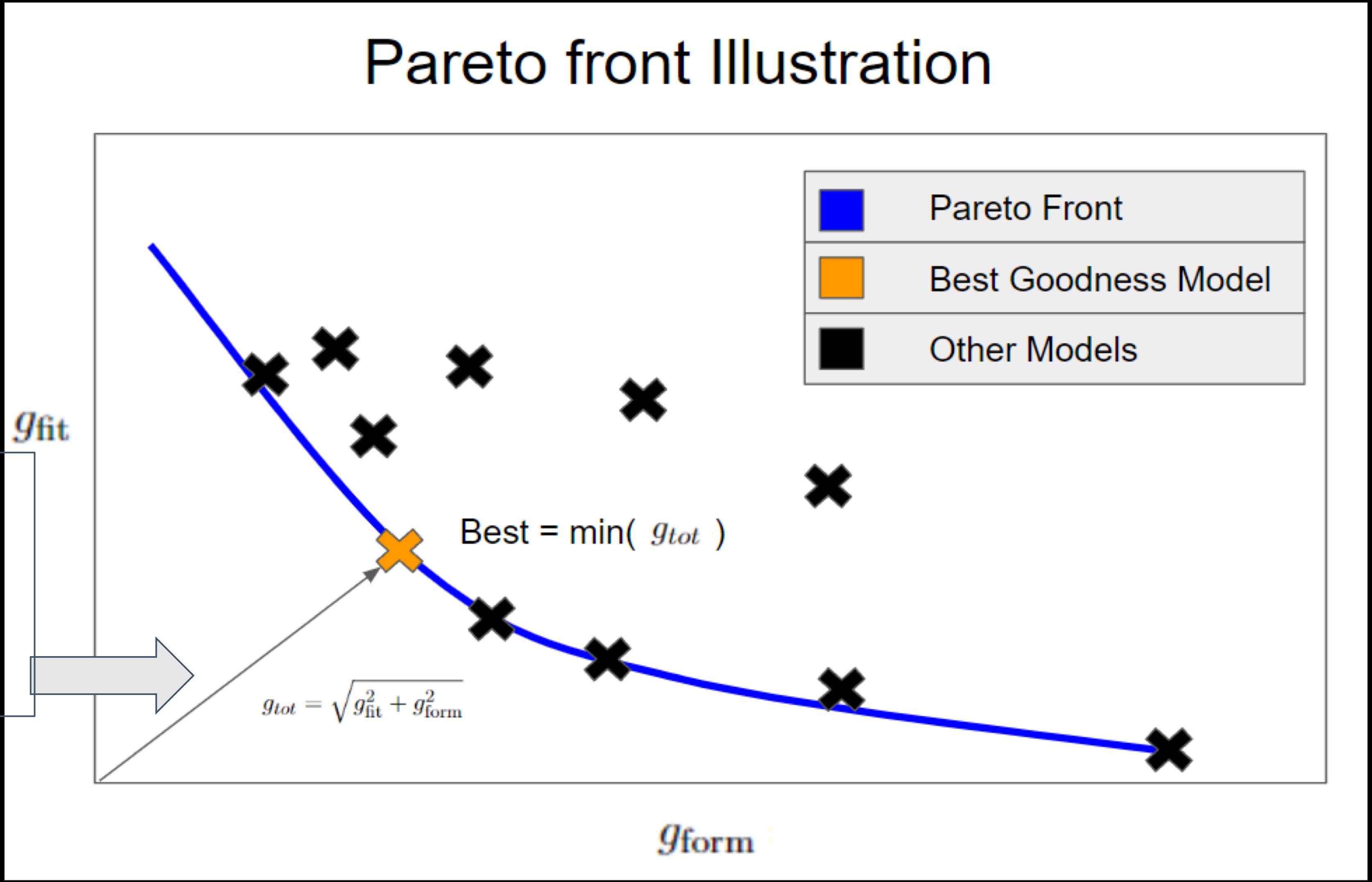
Generalized Parton Distributions from Symbolic Regression

We have a lattice simulation of a GPD as a function of x, t, Q^2
The goal is to find a closed form expression for that GPD



Using a pareto front to choose amongst forms

One must define a "metric" which balances fit vs. form to choose a best



PySR convergence

Andrew Dotson