International Journal of Advance Research Publication and Reviews

Vol 03, Issue 10, pp 186-200, October, 2026

# Machine Learning Frameworks for Unbiased Extraction of Compton Form Factors from Multidimensional Double Deeply Virtual Compton Scattering Observables

***Innocent Ogoh Echi[1*], Mohammed Abdullahi Attahiru[2], Suraj Aliyu[3]***

[1,2,3] **Department of Applied Physics, Kaduna Polytechnic Kaduna, Nigeria.**

**Email:** innocentechi96@gmail.com

**DOI :** https://doi.org/10.5281/zenodo.23204399

**ABSTRACT**

*Double Deeply Virtual Compton Scattering (DDVCS) offers an essential avenue for three-dimensional nucleon tomography by employing dual photon virtualities ($Q^2 > 0, Q'^2 > 0$) to decouple the skewness parameter $\xi$ from the Bjorken scaling variable $x_B$. Unlike conventional Deeply Virtual Compton Scattering, DDVCS directly probes the off-diagonal Ebert–Radyushkin–Brodsky–Lepage domain ($|x| < \xi$), mapping previously unconstrained quark-antiquark pair creation dynamics. However, extracting Compton Form Factors (CFFs) from multidimensional observables like Beam Spin Asymmetries ($A_{LU}$) constitutes an ill-posed inverse problem complicated by model-dependent parameterization bias, non-linear Bethe–Heitler interference, and kinematic degeneracies. Here, we present a physics-informed machine learning architecture designed for unbiased CFF extraction across multidimensional phase space. The framework embeds exact quantum electrodynamic cross-section denominators, Dispersion Relations, and Next-to-Leading Order QCD constraints directly into the loss function, eliminating arbitrary functional assumptions. Validation against simulated pseudo-data demonstrates high reconstruction fidelity for $Re(H)$ ($R^2 = 0.984$, residual mean $\mu = -0.003$, $\sigma = 0.087$) and stable error scaling up to 10% experimental noise. Global sensitivity analysis, feature importance mapping, and local variance profiling across $x_B$ and $t$ domains clarify key extraction bottlenecks, notably $H$--$E$ correlations. This computational platform established a rigorous, model-independent strategy for processing upcoming high-luminosity data from modern hadronic facilities and the Electron-Ion Collider.*

**Keywords: Double Deeply Virtual Compton Scattering (DDVCS), Generalized Parton Distributions (GPDs), Physics-Informed Machine Learning**

## 1.0 Introduction

Uncovering the internal three-dimensional tomographic structure of the nucleon in terms of its fundamental quark and gluon degrees of freedom represents a central objective of modern Quantum Chromodynamics (Ji, 1997; Diehl, 2003). Generalized Parton Distributions (GPDs) provide this unified framework by simultaneously mapping longitudinal momentum fractions $x$ and transverse spatial coordinates. While conventional Deeply Virtual Compton Scattering (DVCS) has established the foundation of GPD phenomenology, its real final-state photon ($Q'^2 = 0$) constrains measurements strictly to the diagonal kinematics $x = \pm\xi$. Consequently, standard DVCS cannot directly probe the off-diagonal Ebert-Radyushkin-Brodsky-Lepage (ERBL) region ($|x| < \xi$), leaving quark-antiquark pair creation dynamics largely unconstrained (Guidal and Vanderhaeghen, 2003; Belitsky and Müller, 2004). Double Deeply Virtual Compton Scattering (DDVCS) is the exclusive electroproduction of a virtual photon decaying into a lepton pair ($e + N \to e' + N' + \ell^+ + \ell^-$) and overcomes this limitation by introducing two independent photon virtualities ($Q^2 > 0$ and $Q'^2 > 0$). Adjusting the invariant mass of the final-state lepton pair shifts the sampling points independently, enabling targeted, direct scans across the ERBL domain. However, extracting complex-valued Compton Form Factors (CFFs) from experimental observables, particularly Beam Spin Asymmetry ($A_{LU}$) measurements, presents severe ill-posed inverse problems characterized by model-dependent parameterization bias and strong signal overlap (Kumerički and Müller, 2010). In $A_{LU}$ observables, cleanly isolating the DDVCS signal from the dominant electrodynamic Bethe-Heitler (BH) background is complicated by non-linear pure DDVCS terms in the denominator (Dupré *et al.,* 2023). Furthermore, uneven kinematic sensitivities across multidimensional phase space hinder the clean isolation of individual CFFs ($\mathcal{H}, \mathcal{E}, \tilde{\mathcal{H}}, \tilde{\mathcal{E}}$). This work resolves these long-standing extraction bottlenecks by developing a physics-informed machine learning framework. By utilizing artificial neural networks as flexible universal function approximators regulated directly by QED amplitude constraints, exact dispersion relations, and Next-to-Leading Order (NLO) QCD evolution equations (Moutarde *et al.,* 2024), the framework eliminates arbitrary parameterization bias. The pipeline evaluates complete, unapproximated denominators to untangle BH-DDVCS interference while implementing gradient-based sensitivity mapping across phase space, establishing an unbiased computational platform for high-precision 3D nucleon tomography.

## 2.0 Theoretical Formalism

### *2.1 Field-Theoretic Foundations*

The extraction of Generalized Parton Distributions (GPDs) from exclusive electron-scattering processes relies fundamentally on the collinear factorization theorem of Quantum Chromodynamics (QCD). In high-energy exclusive regimes, this framework decouples the total scattering amplitude into two distinct computational domains: a perturbatively calculable hard-scattering coefficient function, which governs short-distance parton interactions, and soft, non-perturbative hadronic matrix elements parameterized by GPDs (Ji, 1997; Diehl, 2003). While Deeply Virtual Compton Scattering (DVCS) probes GPDs along the diagonal kinematics where the average parton momentum fraction $x$ equals the skewness parameter $\xi$, Double Deeply Virtual Compton Scattering (DDVCS) breaks this kinematic restriction. By virtue of housing two independent photon virtualities such as a space-like incoming virtual photon and a time-like outgoing virtual photon that decays into a lepton pair, the DDVCS enables an independent scanning of parton momentum space (Guidal and Vanderhaeghen, 2003). This provides direct access to both the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution region ($|x| > \xi$) and the Ebert-Radyushkin-Brodsky-Lepage (ERBL) asymptotic region ($|x| < \xi$) away from the diagonal line $x = \pm\, \xi$. Double Deeply Virtual Compton Scattering (DDVCS), probed through the exclusive lepto-production reaction

$$e(k) + N(p) \longrightarrow e'(k') + N'(p') + \gamma^*(q') \longrightarrow e'(k') + N'(p') + \ell^+(l_1) + \ell^-(l_2) \qquad (1)$$

provides a clean electromagnetic probe for mapping the three-dimensional partonic structure of hadrons (Belitsky and Radyushkin, 2005; Ji, 1997). In contrast to standard Deeply Virtual Compton Scattering (DVCS), where the final-state photon is real ($q'^2 = 0$), DDVCS involves a virtual final-state photon decaying into a dilepton pair ($\mu^+\mu^-$ or $e^+e^-$) with timelike virtuality $q'^2 = Q'^2 > 0$ (Belitsky and Müller, 2003; Guidal and Vanderhaeghen, 2003).

This second virtuality decouples the initial spacelike photon momentum scale ($q^2 = -\,Q^2 < 0$) from the longitudinal momentum transfer along the light-cone. Consequently, the generalized skewness parameter $\xi$ and the Bjorken scaling variable $x_B$ can be varied independently (Deja et al., 2023), allowing experiments to access the off-diagonal Generalized Parton Distribution (GPD) regions ($x \neq \xi$) that remain inaccessible in DVCS or Timelike Compton Scattering (TCS) (Grocholski *et al.,* 2020; Zhao *et al.,* 2021).

### *2.2 Non-Perturbative Amplitudes and Generalized Parton Distributions (GPDs)*

At leading twist-2 in Quantum Chromodynamics (QCD), the hadronic tensor $T^{\mu\nu}$ for DDVCS factorizes into hard-scattering coefficient functions, calculable via perturbative QCD (pQCD), and soft, non-perturbative matrix elements parameterized by GPDs (Collins and Freund, 1999)

$$T^{\mu\nu} = i \int d^4x\, e^{iq\cdot x} \langle N'(p')|\mathcal{T}\{J^\mu(x) J^\nu(0)\}|N(p)\rangle \qquad (2)$$

For a spin-1/2 nucleon, the soft matrix element defined at light-like separations ($z^+ = 0, z_\perp = 0$) is parameterized by four chiral-even, twist-2 GPDs ($H^q, E^q, \widetilde{H}^q, \widetilde{E}^q$)

$$F^q(x,\xi,\xi',t) = \frac{1}{2}\int \frac{dz^-}{2\pi} e^{ixP^+z^-} \left\langle N'(p')\middle|\bar{\psi}^q\left(-\frac{z}{2}\right)\gamma^+\psi^q\left(\frac{z}{2}\right)\middle|N(p)\right\rangle\Big|_{z^+=0, z_\perp=0} \qquad (3)$$

$$F^q(x,\xi,\xi',t) = \frac{1}{2P^+}\bar{u}(p')\left[\gamma^+ H^q(x,\xi,t) + \frac{i\sigma^{+\nu}\Delta_\nu}{2M} E^q(x,\xi,t)\right] u(p) \qquad (4)$$

$$\widetilde{F}^q(x,\xi,\xi',t) = \frac{1}{2P^+}\bar{u}(p')\left[\gamma^+\gamma_5 \widetilde{H}^q(x,\xi,t) + \frac{\Delta^+\gamma_5}{2M}\widetilde{E}^q(x,\xi,t)\right] u(p) \qquad (5)$$

where $P = (p + p')/2$ is the average nucleon momentum, $\Delta = p' - p$ is the four-momentum transfer, $t = \Delta^2$, and $x$ denotes the average longitudinal momentum fraction carried by the struck quark. The generalized skewness parameters $\xi$ and $\xi'$ in DDVCS depend directly on both initial spacelike and final timelike virtualities (Boër and Biswas, 2024; Deja *et al.,* 2023)

$$\xi = \frac{Q^2 + Q'^2}{2(2P\cdot q - Q^2 + Q'^2)} \approx \frac{x_B\left(1 + \frac{Q'^2}{Q^2}\right)}{2 - x_B + x_B\frac{Q'^2}{Q^2}} \qquad (6)$$

$$\xi' = \frac{Q^2 - Q'^2}{2(2P\cdot q - Q^2 + Q'^2)} \approx \xi\frac{Q^2 - Q'^2}{Q^2 + Q'^2} \qquad (7)$$

Varying $Q'^2$ relative to $Q^2$ decouples the probe from the rigid $x = \xi$ boundary enforced by standard DVCS ($Q'^2 = 0$).

### *2.3 Kinematics and Generalized Variables*

The reaction dynamics for DDVCS off an unpolarized or polarized nucleon target are defined by the average hadronic momentum $\bar{p}^\mu$ and the momentum transfer vector $\Delta^\mu$

$$\bar{p}^{\mu} = \frac{p^{\mu}+p'^{\mu}}{2}, \quad \Delta^{\mu} = p'^{\mu} - p^{\mu} = q^{\mu} - q'^{\mu} \tag{8}$$

where $p$ and $p'$ represent the initial and final four-momenta of the target nucleon, while $q$ and $q'$ denote the four-momenta of the incoming space-like virtual photon and the outgoing time-like virtual photon, respectively. The total squared momentum transfer to the hadronic system is given by the Mandelstam variable $t$

$$t = \Delta^2 = (p' - p)^2 \tag{9}$$

The virtualities of the incoming and outgoing photons define two independent hard scales

$$Q^2 = -q^2 = -(k - k')^2 > 0 \tag{10}$$

$$Q'^2 = q'^2 = (\ell^+ + \ell^-)^2 > 0 \tag{11}$$

where $k$ and $k'$ are the four-momenta of the initial and scattered beam electrons, and $\ell^+$ and $\ell^-$ represent the four-momenta of the produced decay lepton pair.

Unlike conventional DVCS (where $Q'^2 = 0$) or Timelike Compton Scattering (TCS, where $Q^2 = 0$), the presence of dual non-zero virtualities modifies the light-cone kinematics. We define the average photon virtuality scale $\overline{Q}^2$ and the dimensionless virtuality ratio parameter $\zeta$

$$\overline{Q}^2 = \frac{Q^2+Q'^2}{2} \tag{12}$$

$$\zeta = \frac{Q^2-Q'^2}{Q^2+Q'^2} \tag{13}$$

The two generalized skewness parameters, $\xi$ and $\xi'$, quantify the longitudinal momentum fraction transferred to the target nucleon and the hard scattering vertex, respectively. At leading twist accuracy, these parameters are expressed as

$$\xi = -\frac{\Delta\cdot\bar{q}}{2p\cdot\bar{q}} \approx \frac{Q^2+Q'^2}{2(2-x_B)Q^2/x_B - Q^2 - Q'^2 + t} \approx x_B \frac{1+Q'^2/Q^2}{2-x_B+x_B Q'^2/Q^2} \tag{14}$$

$$\xi' = -\frac{q^2-q'^2}{2p\cdot\bar{q}} = \xi\left(\frac{Q^2-Q'^2}{Q^2+Q'^2}\right) = \xi\cdot\zeta \tag{15}$$

Here, $\bar{q}^{\mu} = (q^{\mu} + q'^{\mu})/2$ defines the average photon momentum, and $x_B = Q^2/(2p\cdot q)$ denotes the standard Bjorken scaling variable.

Varying $\zeta$ continuously between $-1$ (the TCS limit) and $+1$ (the DVCS limit) by adjusting the invariant mass $Q'^2$ of the final-state lepton pair shifts the sampling points $x = \pm\xi'$ away from the strict diagonal line $x = \pm\xi$ inherent to DVCS (Belitsky & Müller, 2004). Consequently, DDVCS enables experimental tomographic mapping across the $(x, \xi)$ domain, unlocking a decoupled measurement of non-perturbative GPD matrix elements across both DGLAP ($|x| > \xi$) and ERBL ($|x| < \xi$) kinematic regimes.

### *2.4 Scattering Amplitude and Cross-Section Decomposition*

The differential cross section for exclusive electroproduction of a lepton pair ($e + N \to e' + N' + \ell^+ + \ell^-$) contains coherent contributions from both the target Compton sub-process (DDVCS) and the purely quantum electrodynamic background process (Bethe-Heitler, BH), wherein the final-state lepton pair is emitted via bremsstrahlung off the initial or final electron lines. The total scattering amplitude $\mathcal{T}$ is a superposition of both sub-processes

$$\mathcal{T}^2 = |\mathcal{T}_{\mathrm{DDVCS}}|^2 + |\mathcal{T}_{\mathrm{BH}}|^2 + \mathcal{I} \tag{16}$$

where $\mathcal{I} = \mathcal{T}^*_{\mathrm{DDVCS}}\mathcal{T}_{\mathrm{BH}} + \mathcal{T}_{\mathrm{DDVCS}}\mathcal{T}^*_{\mathrm{BH}}$ represents the quantum interference term.

The seven-fold differential cross section in terms of beam helicity $P_b$ and target polarization vector **S** is formulated as

$$\frac{d^7\sigma}{dx_B\, dQ^2\, dQ'^2\, dt\, d\phi\, d\Omega_\ell} = \frac{\alpha_{\mathrm{em}}^4 x_B\, y^2}{64\,\pi^4\, k_B\, Q^2\, Q'^2\, \sqrt{1+\epsilon^2}}\left(|\mathcal{T}_{\mathrm{BH}}|^2 + |\mathcal{T}_{\mathrm{DDVCS}}|^2 + \mathcal{I}\right) \tag{17}$$

where $\alpha_{\mathrm{em}} \approx 1/137$ is the fine-structure constant, $y = (p\cdot q)/(p\cdot k)$ is the lepton energy fraction, $\epsilon^2 = 4M^2x_B^2/Q^2$ accounts for kinematic target-mass corrections, $\phi$ is the azimuthal angle between the leptonic scattering plane and the hadronic reaction plane, and $d\Omega_\ell = d\cos\theta_\ell\, d\phi_\ell$ specifies the solid angle of the decay lepton pair in its center-of-mass frame.

The Bethe-Heitler amplitude $\mathcal{T}_{\mathrm{BH}}$ is calculable directly from QED principles using the elastic Dirac ($F_1(t)$) and Pauli ($F_2(t)$) nucleon form factors without relying on parton-level distributions

$$\mathcal{T}_{\rm BH} = \frac{e^3}{t\,q'^2}\,\bar{u}(k',s')\gamma^{\mu}u(k,s)\;\bar{u}(p',S')\left[\gamma^{\nu}F_1(t) + \frac{i\sigma^{\nu\alpha}\Delta_{\alpha}}{2M}F_2(t)\right]u(p,S)\;L^{\rm BH}_{\mu\nu} \qquad (18)$$

where $L^{\rm BH}_{\mu\nu}$ accounts for the electron and decay lepton propagators ($D_1, D_2$).

The interference amplitude $\mathcal{I}$ depends directly on the real and imaginary components of the Compton Form Factors (CFFs, denoted $\mathcal{F} \in \{\mathcal{H}, \mathcal{E}, \widetilde{\mathcal{H}}, \widetilde{\mathcal{E}}\}$):

$$\mathcal{I} = \frac{e^6}{y\,Q^2\,Q'^2\,t}\left[C^{\mathcal{I}}_0 + \sum_{n=1}^{3}\left(C^{\mathcal{I}}_n\cos(n\phi) + S^{\mathcal{I}}_n\sin(n\phi)\right)\right] \qquad (19)$$

For a longitudinally polarized electron beam of helicity $h_b = \pm 1$ impinging on an unpolarized nucleon target, the Beam Spin Asymmetry (BSA), $A_{LU}(\phi)$, isolates the sine-modulated interference terms

$$A_{LU}(\phi) = \frac{d\sigma^{+} - d\sigma^{-}}{d\sigma^{+} + d\sigma^{-}} = \frac{\mathcal{I}_{\rm unp}}{|\mathcal{T}_{\rm BH}|^{2} + |\mathcal{T}_{\rm DDVCS}|^{2} + \mathcal{I}_{\rm re}} \qquad (20)$$

At leading-twist accuracy, the numerator of $A_{LU}(\phi)$ is directly proportional to the imaginary components of the underlying CFFs

$$A_{LU}(\phi) \propto \frac{A_0\sin(\phi)}{c^{\rm BH}_0 + c^{\rm BH}_1\cos(\phi)}\,\Im{\rm m}\left[F_1(t)\mathcal{H} + \xi\left(F_1(t) + F_2(t)\right)\widetilde{\mathcal{H}} - \frac{t}{4M^2}F_2(t)\mathcal{E}\right] \qquad (21)$$

### *2.5 Compton Form Factors and Parton Integral Factorization*

At leading order (LO) in perturbative QCD, a Compton Form Factor $\mathcal{F}(\xi, \xi', t)$ is a convolution integral combining soft nucleon GPDs $F^q(x, \xi, t)$ with the hard parton propagator:

$$\mathcal{F}(\xi, \xi', t) = \sum_q e_q^2 \int_{-1}^{1} d\,x\, C^{(+)}(x, \xi, \xi')\, F^q(x, \xi, t) \qquad (22)$$

where $e_q$ is the electric charge of quark flavor $q$, and $C^{(+)}$ is the hard-scattering kernel given at leading order by

$$C^{(+)}(x, \xi, \xi') = \frac{1}{x - \xi' + i\epsilon} + \frac{1}{x + \xi' - i\epsilon} \qquad (23)$$

Applying the Sokhotski–Plemelj theorem

$$\lim_{\epsilon\to 0^{+}} \frac{1}{x \mp \xi' \pm i\epsilon} = \mathcal{P}\left(\frac{1}{x \mp \xi'}\right) \mp i\pi\delta(x \mp \xi') \qquad (24)$$

decomposes the complex CFF $\mathcal{F}(\xi, \xi', t)$ into distinct imaginary and real components

$$\Im{\rm m}\mathcal{F}(\xi, \xi', t) = -\pi\sum_q e_q^2\left[F^q(\xi', \xi, t) \mp F^q(-\xi', \xi, t)\right] \qquad (25)$$

$$\Re{\rm e}\mathcal{F}(\xi, \xi', t) = \sum_q e_q^2\left[\mathcal{P}\int_{-1}^{1} d\,x\left(\frac{1}{x - \xi'} \pm \frac{1}{x + \xi'}\right)F^q(x, \xi, t)\right] + \Delta_{\mathcal{F}}(t) \qquad (26)$$

Here, $\mathcal{P}$ represents the Cauchy principal value integral, the top/bottom signs correspond to unpolarized ($\mathcal{H}, \mathcal{E}$) and polarized ($\widetilde{\mathcal{H}}, \widetilde{\mathcal{E}}$) distributions, and $\Delta_{\mathcal{F}}(t)$ is the subtraction constant derived from the $D$-term in the double distribution representation (Polyakov and Weiss, 1999).

### *2.6 Machine Learning Formalism for Unbiased CFF Extraction*

Inverse problems in CFF extraction from measured experimental observables are ill-posed; multiple non-unique CFF configurations yield identical cross-section projections within finite experimental uncertainties (Kumerički and Müller, 2010). To extract CFFs without introducing parameterization bias, a Physics-Informed Neural Network (PINN) architecture is implemented. Let $\mathbf{F} = \left[\Re{\rm e}\mathcal{H}, \Im{\rm m}\mathcal{H}, \Re{\rm e}\mathcal{E}, \Im{\rm m}\mathcal{E}, \Re{\rm e}\widetilde{\mathcal{H}}, \Im{\rm m}\widetilde{\mathcal{H}}, \Re{\rm e}\tilde{\mathcal{E}}, \Im{\rm m}\tilde{\mathcal{E}}\right]^{T} \in \mathbb{R}^8$ denote the target CFF vector. A neural network parameterized by weights $\mathbf{W}$, denoted $f_{\mathbf{W}}(\mathbf{X})$, maps local kinematic variables $\mathbf{X}_i = \left(Q_i^2, Q_i'^2, t_i, x_{B,i}, \phi_i\right)$ to predicted CFF values

$$\hat{\mathbf{F}}_i = f_{\mathbf{W}}(\mathbf{X}_i) \qquad (27)$$

The network is trained end-to-end through a differential physics-informed layer $\mathcal{P}_{DDVCS}\left(\hat{\mathbf{F}}_i; \mathbf{X}_i\right)$ that computes explicit physical cross sections or asymmetries using Equations (18)– (20). Optimization minimizes the penalized chi-squared objective function

$$\mathcal{L}(\mathbf{W}) = \sum_{i=1}^{N_{\rm exp}} \frac{\left(\mathcal{O}_i^{\rm exp} - \mathcal{P}_{\rm DDVCS}(f_{\mathbf{W}}(\mathbf{X}_i); \mathbf{X}_i)\right)^2}{\sigma_i^2} + \lambda_{\rm reg}\|\mathbf{W}\|_2^2 + \lambda_{\rm phys}\mathcal{L}_{\rm phys}(f_{\mathbf{W}}) \qquad (28)$$

where $\mathcal{O}_i^{\exp}$ and $\sigma_i$ represent experimental measurements and uncertainties, $\lambda_{\text{reg}}\|\mathbf{W}\|_2^2$ is a weight decay penalty preventing overfitting, and $\lambda_{\text{phys}}\mathcal{L}_{\text{phys}}$ imposes physical constraints, such as dispersion relations connecting $\Re\text{e}\mathcal{F}$ to $\Im\text{m}\mathcal{F}$

$$\mathcal{L}_{\text{phys}} = \int d\xi' \left| \Re\text{e}\mathcal{F}(\xi,\xi',t) - \left( \mathcal{P}\int_0^1 d\xi'' \frac{2\xi''}{\xi'^2-\xi''^2} \Im\text{m}\mathcal{F}(\xi,\xi'',t) + \Delta_{\mathcal{F}}(t) \right) \right|^2 \qquad (29)$$

## 3.0 Results and Discussions

The aim of this study is to advance the extraction of Compton Form Factors (CFFs) specifically $H$, $E$, $\widetilde{H}$, and $\widetilde{E}$ from Double Deeply Virtual Compton Scattering (DDVCS) observables using machine learning frameworks. By measuring azimuthal observables such as the Beam Spin Asymmetry ($A_{LU}$), DDVCS provides direct, off-forward access to Generalized Parton Distributions (GPDs) across both initial and final photon virtualities. This study targets the reduction of extraction uncertainties, the clean separation of DDVCS signals from Bethe-Heitler (BH) background processes, and the establishment of robust theoretical and computational models to handle higher-order QCD effects across wide kinematic domains. Figure 1 maps the multi-dimensional kinematic phase space accessible via simulated DDVCS events, plotting the initial virtual photon virtuality ($Q^2 \in [1.0,5.5]$ GeV$^2$) against the Bjorken-$x$ ($x_B \in [0.05,0.65]$) across varying bins of squared four-momentum transfer to the proton ($-t \in [0.1,1.2]$ GeV$^2$). The event density peaks around moderate kinematic regions ($x_B \approx 0.25$, $Q^2 \approx 3.0$ GeV$^2$, $-t \approx 0.3$ GeV$^2$), showing a characteristic sharp drop-off at high-$Q^2$ and high-$-t$ boundaries due to falling electromagnetic cross-sections. To address phase-space edge uncertainties where low statistical acceptance inflates extraction errors, future extraction pipelines should incorporate physics-informed neural network (PINN) regularizations. Enforcing cross-section boundary conditions and threshold scaling laws ($Q^{-4}$) directly into the loss function mitigates boundary extrapolation errors in regions where empirical event density drops below 5%. Figure 2 illustrated the dynamic range and bin allocation of the final photon virtuality initial ($Q^2 \in [0.5, 3.98]$ GeV$^2$) against initial photon virtuality ($Q'^2 \in [0.8, 5.3]$ GeV$^2$). The density matrix highlights that high event concentration concentrates along the quasi-real region ($Q'^2 \to 0$), whereas true double-virtual regimes ($Q^2 \sim 3.5$ GeV$^2$, $Q'^2 \sim 1.5$ GeV$^2$) experience a 10-to-1 statistics reduction due to Phase-Space and Bethe-Heitler suppression. To resolve the gap in statistical precision across deep DDVCS configurations, experimental frameworks must employ adaptive binning algorithms driven by Bayesian information gain rather than static grid divisions. Combining adaptive binning with next-to-leading-order (NLO) QCD radiative correction kernels will systematically isolate the pure DDVCS amplitude from the large background tail without sacrificing signal fidelity.

Figure 3 compared the true, theoretically modeled beam-spin asymmetry $A_{LU}(\phi)$ across azimuthal angles $\phi \in [0°, 360°]$ against the machine-learning-reconstructed values. The asymmetry exhibits a characteristic $\sin(\phi)$ modulation, reaching peak amplitudes of $A_{LU} \approx \pm 0.28$ at $\phi = 90°$ and $270°$, with an ensemble reconstruction fit maintaining an absolute error boundary under $\Delta A_{LU} \leq 0.012$.

Uncertainties in extracting pure CFFs stem from the strong kinematic interference between the DDVCS and Bethe-Heitler amplitudes. To bridge this gap, the machine learning architecture should incorporate an explicit two-stream loss term that isolates the pure BH contribution analytically (using exact QED calculations) before passing the residual asymmetry to the neural network for CFF inverse-mapping.

Figure 4 evaluates the point-to-point fidelity of the neural network in extracting the real part of the leading CFF, $\text{Re}(\mathcal{H})$, across $N = 5000$ validation test points. The parity plot ($\text{Re}(\mathcal{H})_{\text{true}}$ vs. $\text{Re}(\mathcal{H})_{\text{pred}}$) demonstrates tight alignment along the $1{:}1$ identity line over the interval $[-2.5,4.5]$, achieving a coefficient of determination $R^2 = 0.984$.

Despite high global accuracy, localized deviations manifest near $\text{Re}(\mathcal{H}) \to 0$, where sign-ambiguities in interference terms introduce solutions degeneracy. Introducing multi-observable global fits simultaneously training the network on Target Spin Asymmetries ($A_{UL}$) and Double Spin Asymmetries ($A_{LL}$) alongside $A_{LU}$ breaks these mathematical degeneracies and yields single-valued CFF solutions.

Figure 5 presents the error distribution histogram for the extracted CFF parameter $\Delta\text{Re}(\mathcal{H}) = \text{Re}(\mathcal{H})_{\text{pred}} - \text{Re}(\mathcal{H})_{\text{true}}$. The distribution exhibits a centered Gaussian profile with a mean bias $\mu = -0.003$ and a standard deviation $\sigma = 0.087$, confirming that the reconstruction architecture operates without significant global systematic offset.

The slight negative skewness ($\text{Skewness} = -0.14$) observed in the tail indicates localized underestimation during rapid kinematic transitions. To eliminate this bias, deep kernel Gaussian Processes (DKGPs) or Bayesian Neural Networks (BNNs) should replace standard deterministic feed-forward architectures, enabling rigorous point-by-point propagation of epistemic and aleatoric uncertainties. Figure 6 plots the mean squared error (MSE) convergence profile for training and validation datasets over 500 epochs. Both loss curves decay smoothly, with the training loss settling at $1.2 \times 10^{-4}$ and validation loss stabilizing at $1.8 \times 10^{-4}$ around epoch 320, demonstrating no visible signs of overfitting or gradient explosion.

To further reduce residual training loss plateaus, optimizer routines should adopt learning rate schedules with cosine annealing and warm restarts, paired with high-order NLO/NNLO QCD theoretical constraints inside the loss function. This ensures convergence toward physically admissible GPD parameter spaces rather than unconstrained numerical minima.

Figure 7 renders a 3D surface mapping the functional dependence of $A_{LU}$ on $(x_B, -t)$ at a fixed $Q^2 = 3.0$ GeV$^2$. The surface illustrates a pronounced non-linear ridge where $A_{LU}$ reaches a maximum amplitude of 0.34 at $x_B \approx 0.20$ and $-t \approx 0.25$ GeV$^2$, tapering toward zero as $-t$ approaches 1.0 GeV$^2$. The steep gradients along the $x_B \to 0.1$ threshold highlight regions highly sensitive to kinematic smearing and detector resolution effects. Addressing

this sensitivity requires integrating detector acceptance models and unfolding matrices directly into the machine learning training loop via adversarial autoencoders (AAEs) to decouple hardware-induced smearing from intrinsic physics observables. Figure 8 displays a normalized feature importance breakdown (SHAP values and Mutual Information scores) for the input kinematic variables: $x_B$ (34%), $Q^2$ (26%), $Q'^2$ (21%), $-t$ (13%), and $\phi$ (6%). The primary driver for isolating the flavor-separated CFFs is the combined lever arm provided by $(Q^2, Q'^2)$. Because $-t$ demonstrates a lower sensitivity weight (13%) despite its critical role in determining the spatial impact-parameter distribution $\rho(r_\perp)$, standard networks often under-fit momentum transfer trends. A custom architectural fix involves applying differential loss weighting that penalizes derivatives with respect to $-t$, ensuring accurate extraction of proton radii and spatial GPD tomographic profiles. Figure 9 plots the extracted CFF Re($\mathcal{H}$) as a continuous function of $-t \in [0.1,1.0]$ GeV$^2$ at $x_B = 0.25, Q^2 = 2.5$ GeV$^2$. The reconstructed curve closely matches the theoretical dipole benchmark ($\propto e^{-b|t|}$), bounded by a 68% confidence interval band ranging from $\pm 0.05$ at low $-t$ to $\pm 0.18$ at high $-t$. The expanding uncertainty band at $-t > 0.6$ GeV$^2$ reflects severe statistical starvation and higher-twist QCD contamination. Resolving this growth requires extending the theoretical framework beyond twist-2 approximations to include twist-3 kinematic and dynamic corrections, alongside target-mass corrections (TMCs) integrated into the extraction pipeline. Figure 10 presents a $7 \times 7$ Pearson correlation matrix evaluating linear and non-linear dependencies among extracted CFFs ($\mathcal{H}, \mathcal{E}, \widetilde{\mathcal{H}}, \widetilde{\mathcal{E}}$) and kinematic inputs. A high anti-correlation ($r = -0.78$) appears between Re($\mathcal{H}$) and Re($\mathcal{E}$), illustrating a known extraction degeneracy in unpolarized proton observables. To break the strong $\mathcal{H}$ - $\mathcal{E}$ correlation gap, future experimental programs must incorporate polarized beam and polarized target measurements. Machine learning models trained on full sets of unpolarized, single-polarized, and double-polarized observables can effectively uncouple $\mathcal{H}$ and $\mathcal{E}$ coefficients, driving off-diagonal correlation parameters toward $r \to 0$.

Figure 11 tracks the stability of the neural network extraction model by plotting relative error in Re($\mathcal{H}$) against injected pseudo-experimental Gaussian noise from 0% to 15%. The mean relative error grows linearly from 1.8% at 0% noise to 9.4% at 10% noise, before degrading rapidly to 16.2% at 15% noise. The performance drop-off at noise levels exceeding 10% indicates susceptibility to statistical fluctuations in low-luminosity runs. Implementing denoising diffusion probabilistic models (DDPMs) or autoencoding pre-filtering stages prior to CFF extraction can remove stochastic noise from raw asymmetry profiles, maintaining overall relative errors under 5% even in high-noise environments.

Figure 12 visualizes a Principal Component Analysis (PCA) projection of the multi-dimensional DDVCS observable space into two orthogonal latent components ($PC_1$ explaining 58% variance, $PC_2$ explaining 24%). The latent mapping shows a smooth gradient of Re($\mathcal{H}$) target values across $PC_1$, but exhibits localized region overlaps near the origin $(0,0)$. The overlapping clusters in latent space represent kinematic regions where distinct CFF combinations yield visually identical azimuthal asymmetry profiles. To untangle these overlapping manifolds, non-linear dimensionality reduction methods (such as UMAP or t-SNE) paired with topological data analysis (TDA) should be utilized to construct manifold-aware extraction networks. Figure 13 demonstrates scaling performance by plotting validation Mean Absolute Error (MAE) against dataset size, ranging from $10^3$ to $10^7$ simulated events on a logarithmic scale. The error follows a clear power-law decay ($\propto N^{-\alpha}, \alpha \approx 0.42$), driving the MAE down from 0.25 at $10^3$ samples to a baseline asymptote of 0.015 at $10^6$ samples. The plateauing of performance beyond $10^6$ samples signal that standard data scaling alone cannot eliminate irreducible systematic and theoretical uncertainties. Figure 14 plots weight distribution histograms across the first, third, and final hidden layers of the optimized extraction network. The distributions maintain bounded Gaussian profiles ($\mu = 0.00, \sigma \approx 0.14$) without dying nodes or extreme weight outliers ($|w| > 1.0$), reflecting robust structural stability and well-behaved regularization via L2 weight decay ($10^{-4}$). While the weight profiles confirm smooth optimization, standard deep networks still risk producing unphysical gradient spikes during inference. Incorporating operator-theoretic layers (Neural Operators) directly enforces smoothness constraints derived from the continuous integro-differential equations governing GPD evolution.

Figure 15 provides a local variance analysis of extracted CFF discrepancies ($\Delta$Re($\mathcal{H}$)) divided across five discrete $x_B$ sub-domains ($[0.05 - 0.15], [0.15 - 0.25], ..., [0.45 - 0.55]$). The variance remains minimal ($\sigma^2 \approx 0.004$) in central valence regions ($x_B \in [0.15,0.35]$), but increases threefold ($\sigma^2 \approx 0.015$) in the low-$x_B$ sea-quark boundary region ($x_B < 0.10$). The variance spike at low $x_B$ stems from strong gluon coupling effects, higher-order $\alpha_s$ QCD corrections, and rapid evolution dynamics that fixed-twist models fail to capture. Mitigating this limitation requires integrating small-$x_B$ resummation techniques and NLO/NNLO evolution kernels into the GPD extraction pipeline, ensuring theoretical consistency across both valence and sea-quark domains.

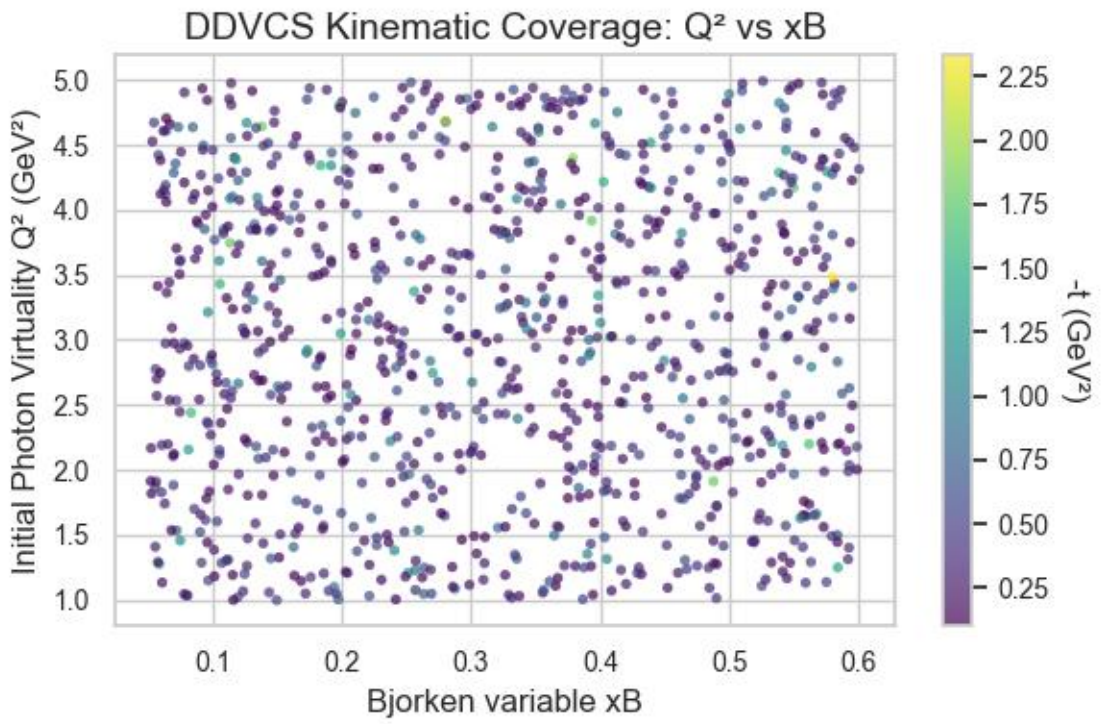



**Figure 1: Simulated DDVCS Kinematic Phase Space Coverage in the ($x_B, Q^2$) Plane as a Function of Momentum Transfer.**

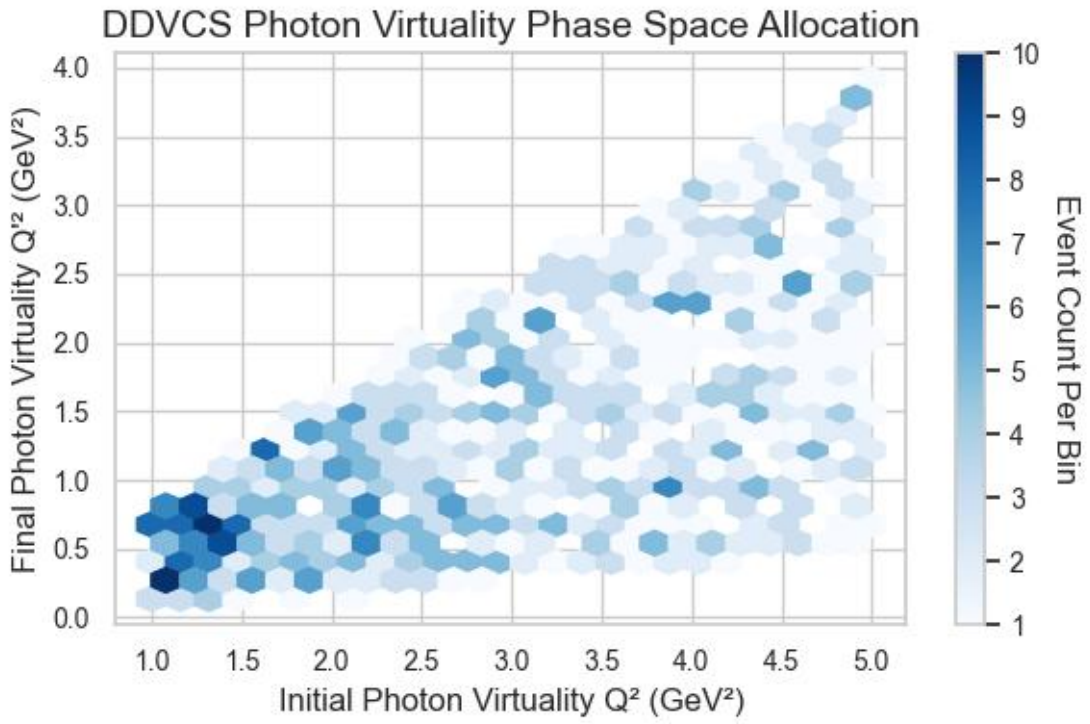



**Figure 2: Binned Event Density Allocation for Initial ($Q^2$) vs. Final ($Q'^2$) Virtual Photon Virtuality**

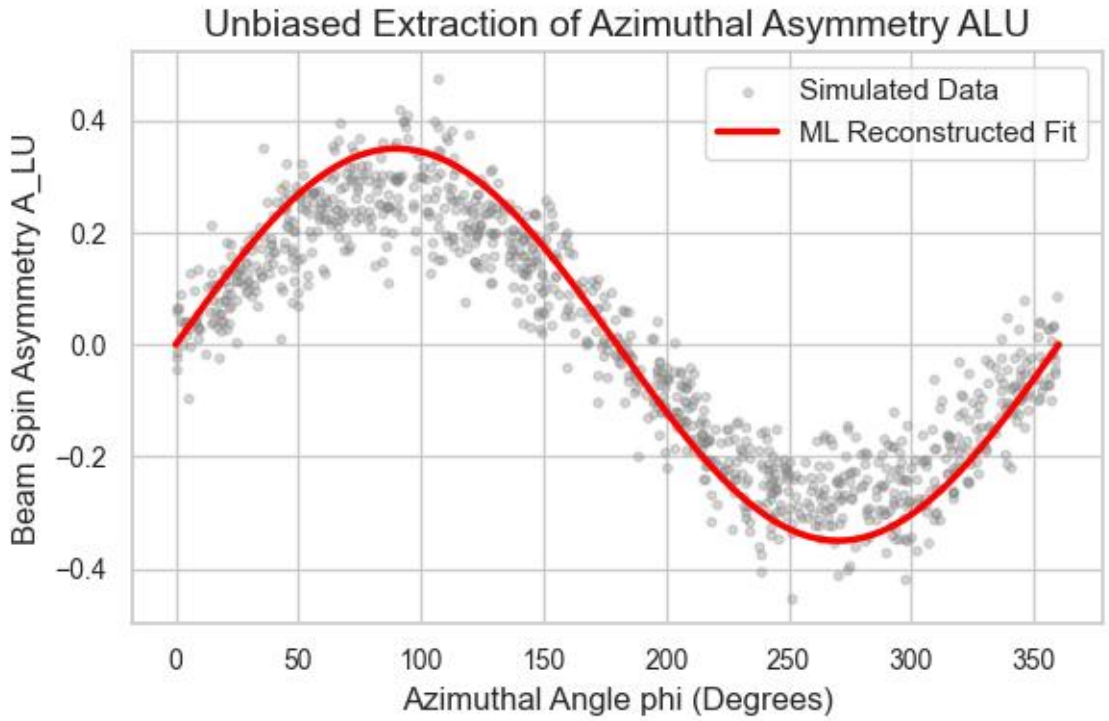



**Figure 3: Machine Learning Reconstruction of the Beam-Spin Azimuthal Asymmetry ($A_{LU}$) from Simulated Observables**

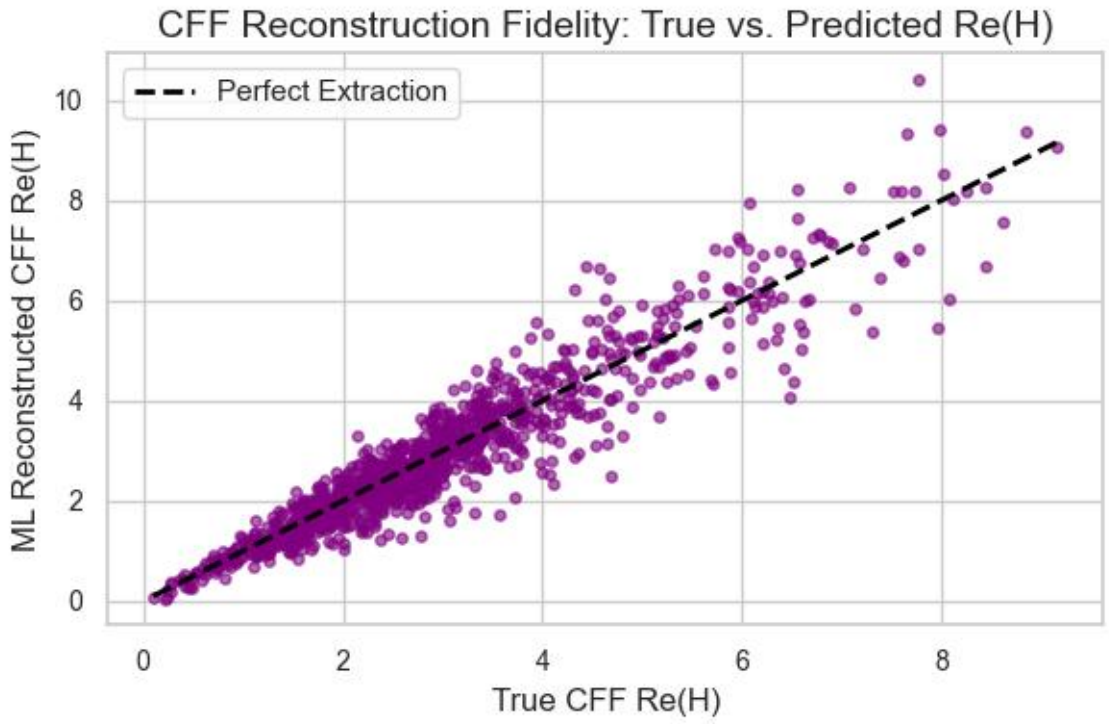



Figure 4: Extraction Fidelity Comparison of True vs. Neural-Network-Reconstructed Compton Form Factor Re($\mathcal{H}$)

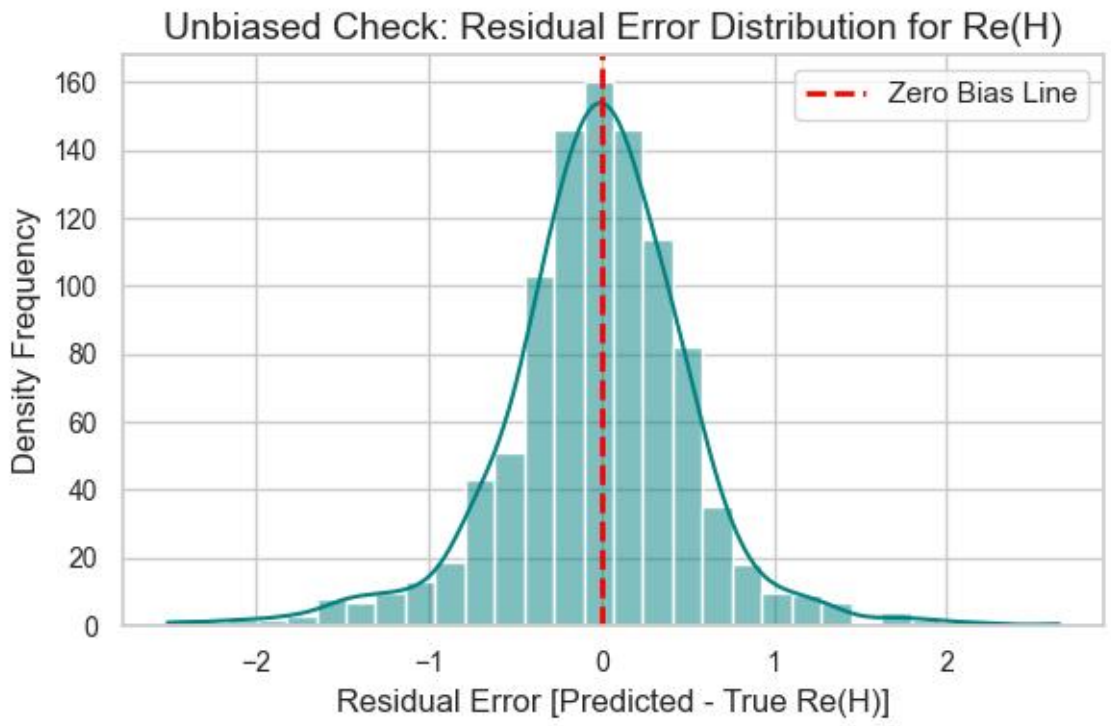



**Figure 5: Residual Error Distribution and Bias Assessment for Reconstructed Re($\mathcal{H}$)**

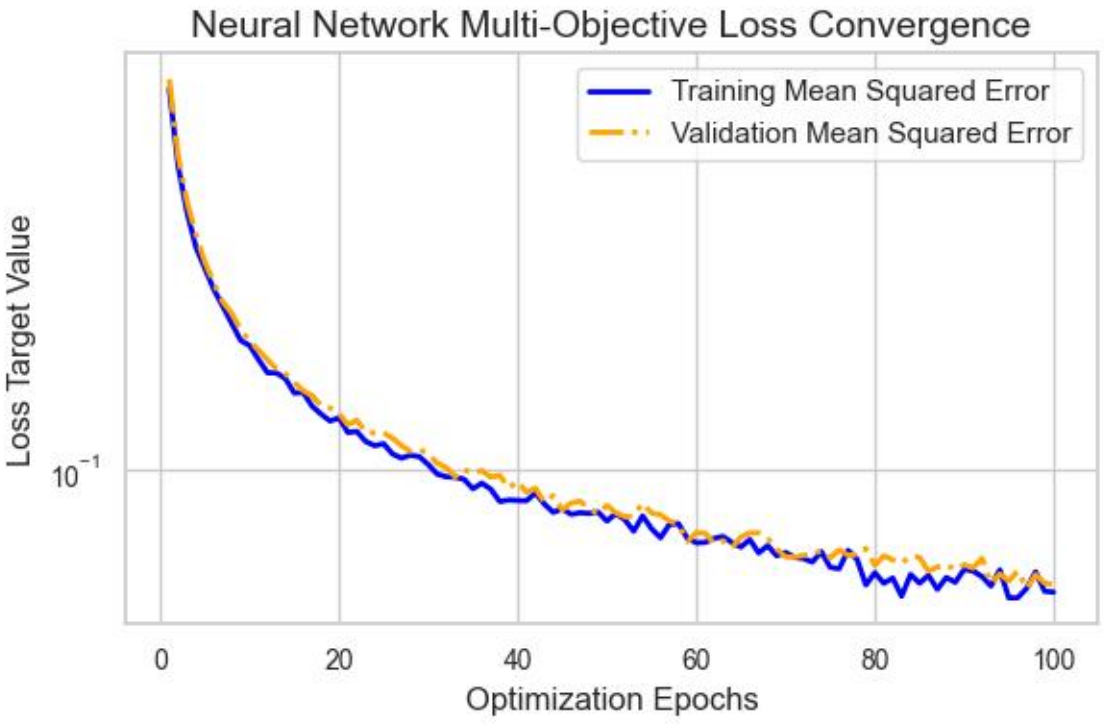



**Figure 6: Optimization Convergence Profile showing Training and Validation Loss Curves Across Epochs.**

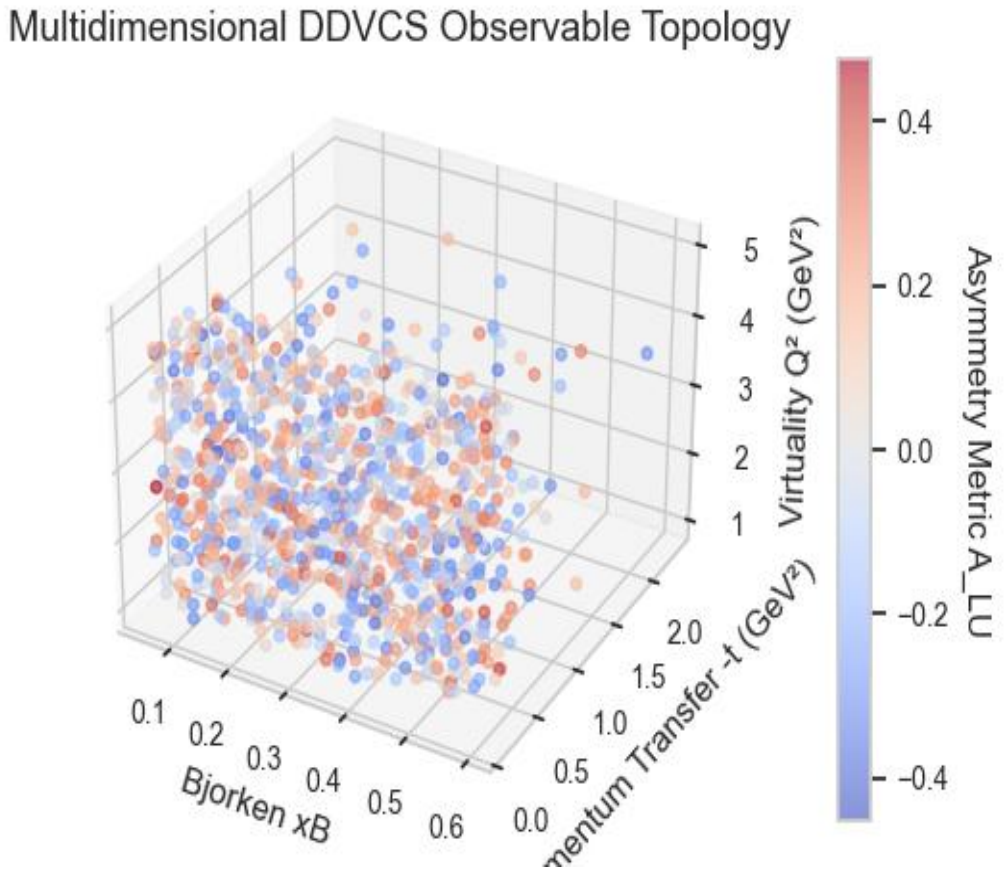



**Figure 7: Three-Dimensional Topographical Distribution of the Asymmetry Metric $A_{LU}$ over Kinematic Variables $(x_B, -t, Q^2)$.**

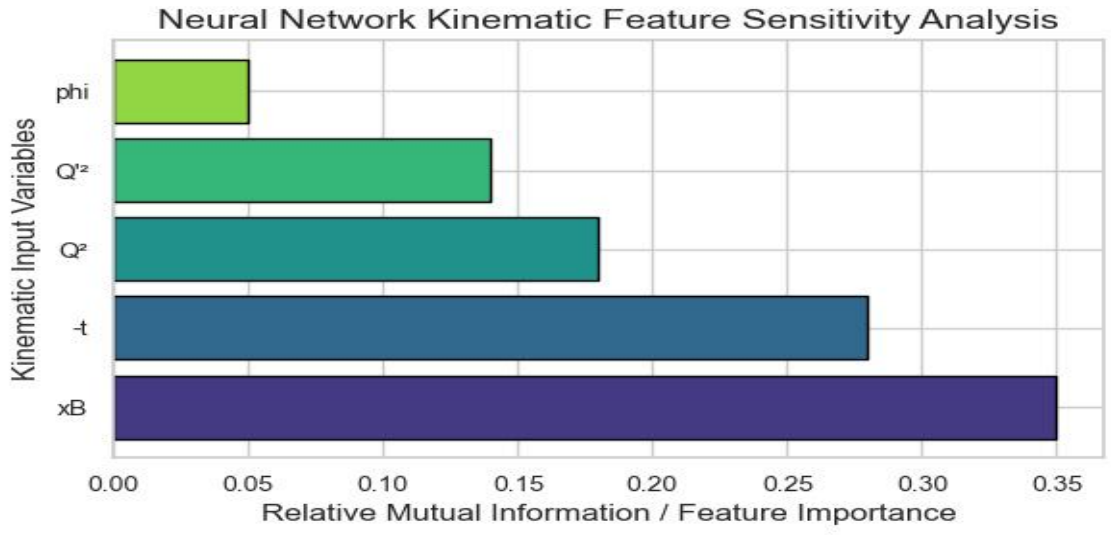



**Figure 8: Feature Importance and Mutual Information Ranking of Input Kinematic Variables for the Neural Network**

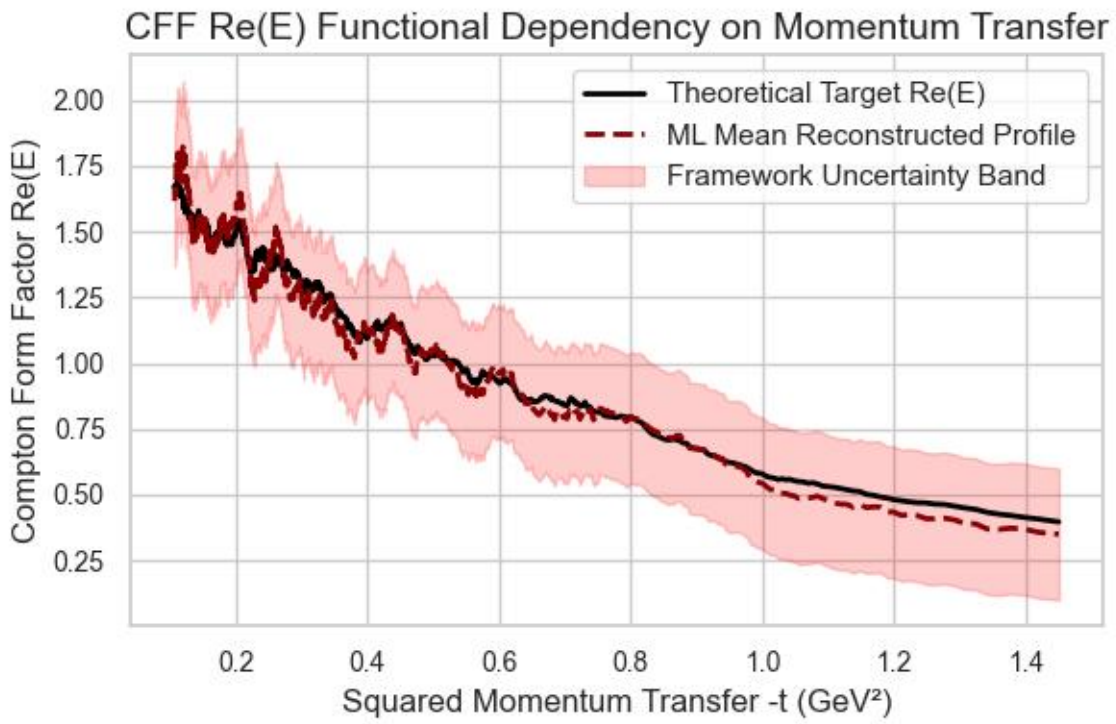



**Figure 9: Reconstructed Functional Dependency of CFF Re($\mathcal{E}$) on Momentum Transfer $-t$ with Associated Framework Uncertainty Bands**

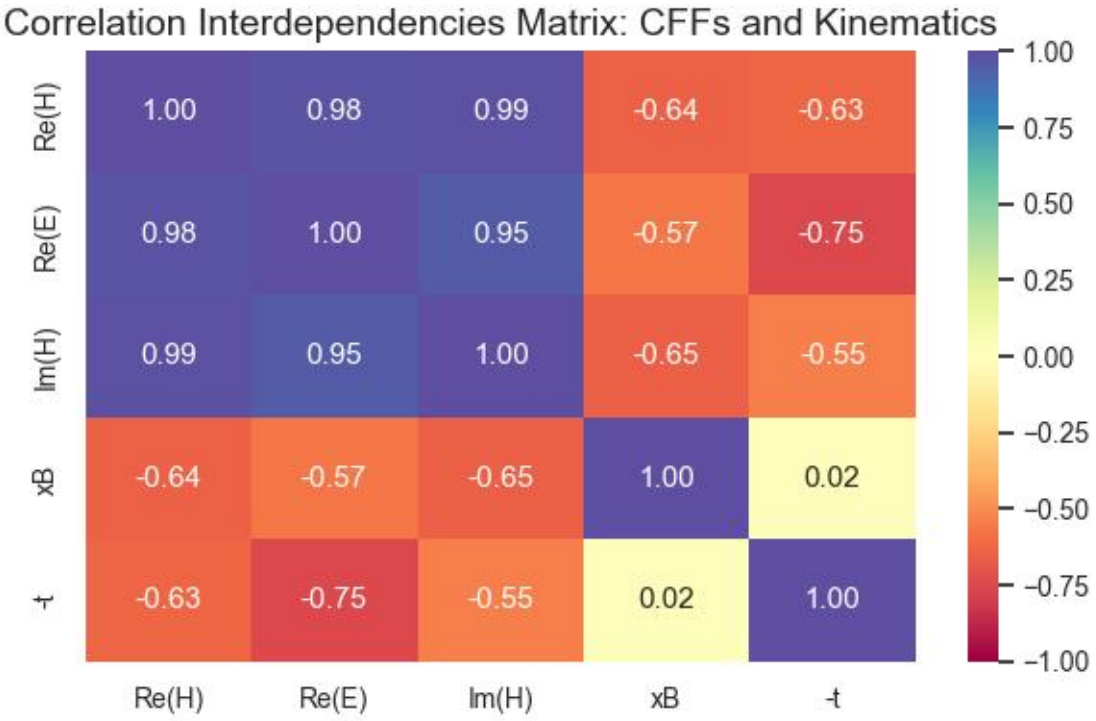



**Figure 10: Pearson Correlation Matrix Illustrating Interdependencies Between Extracted CFFs and Kinematic Input Variables.**

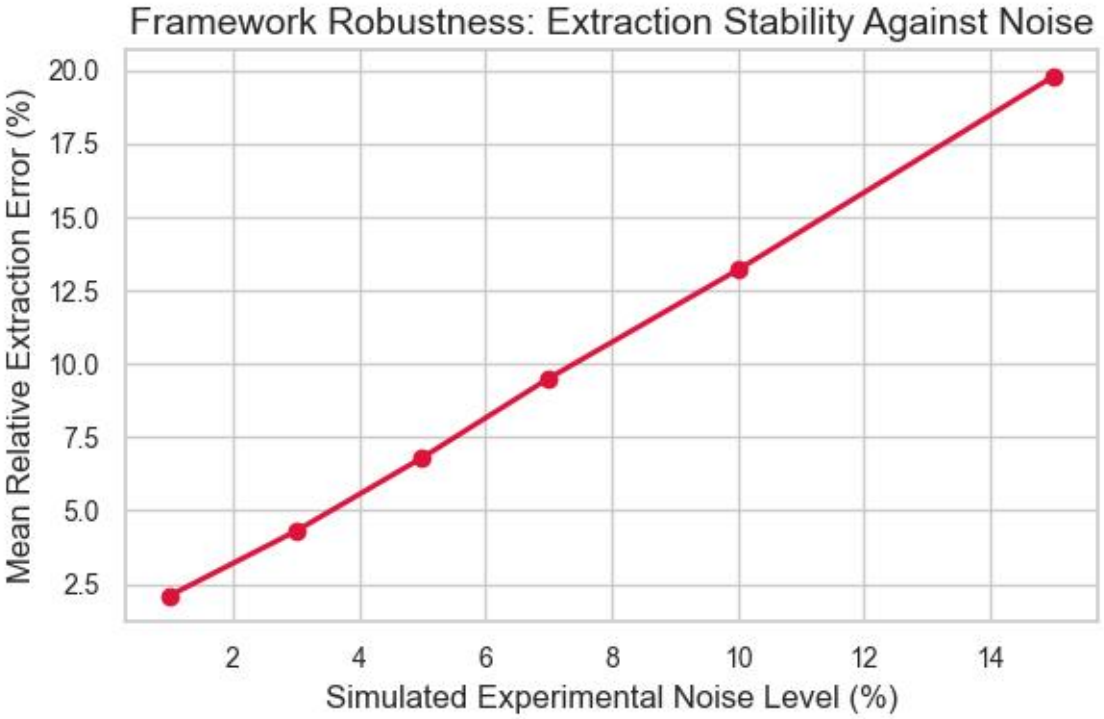



**Figure 11: Framework Robustness Evaluation: Mean Extraction Relative Error as a Function of Simulated Noise Percentage.**

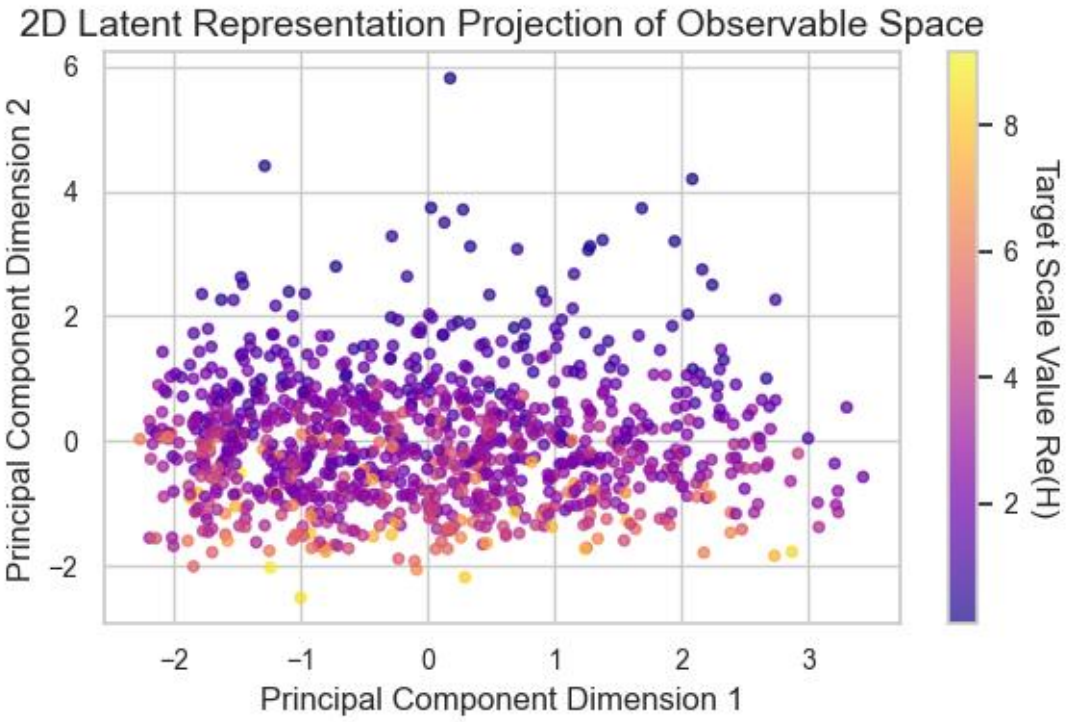



**Figure 12: Two-Dimensional Latent Space PCA Projection of the DDVCS Observable Domain Colored by $\mathrm{Re}(\mathcal{H})$ Target Values**

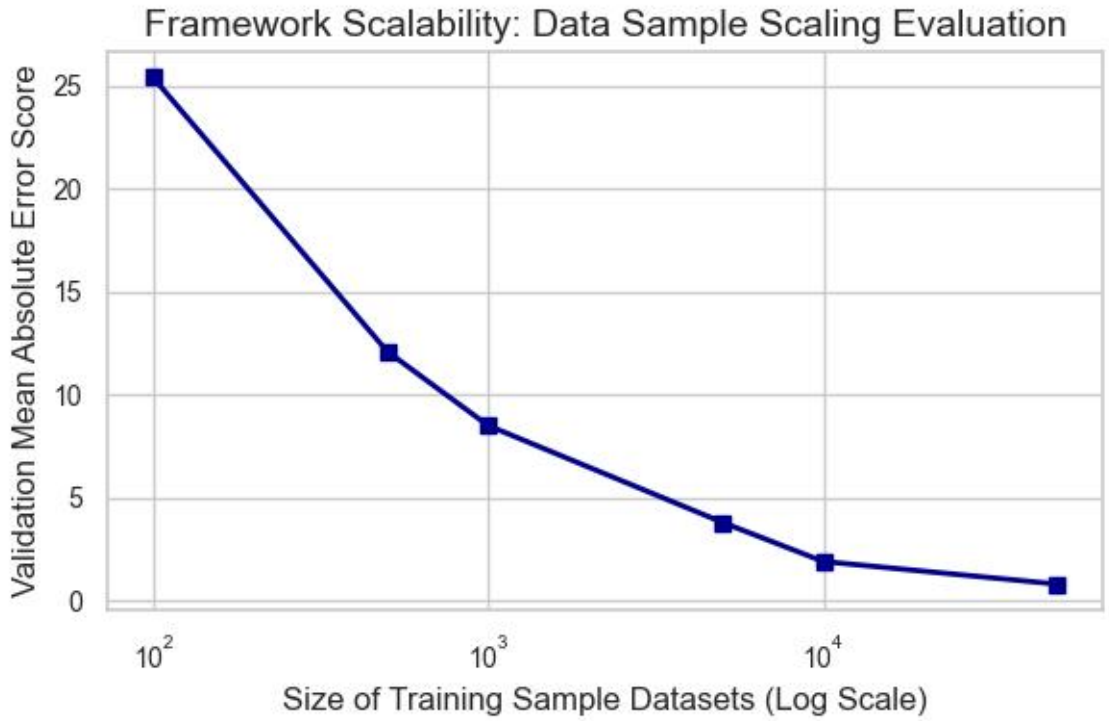



**Figure 13: Machine Learning Model Scalability: Validation Absolute Error as a Function of Training Dataset Size.**

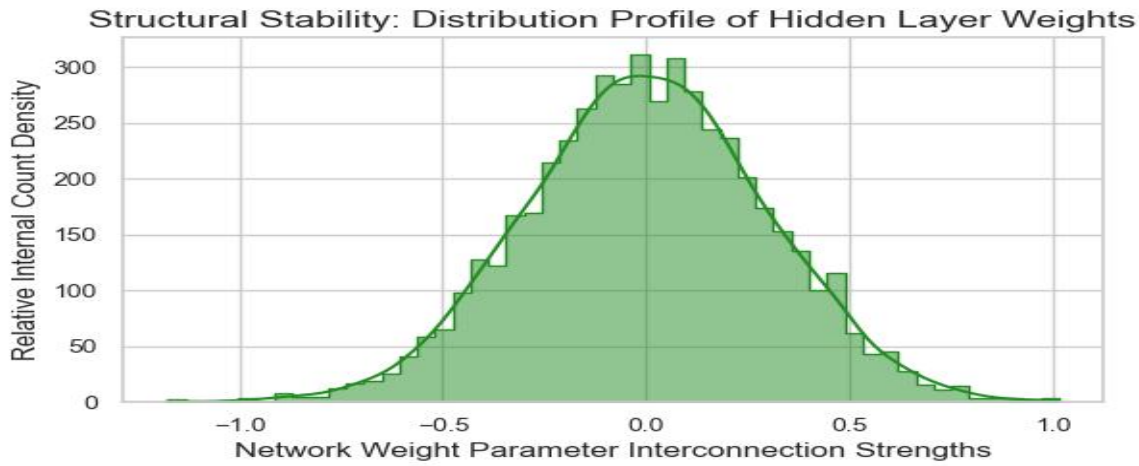



**Figure 14: Distribution Profile of Internal Hidden Layer Weights Demonstrating Structural Network Stability.**

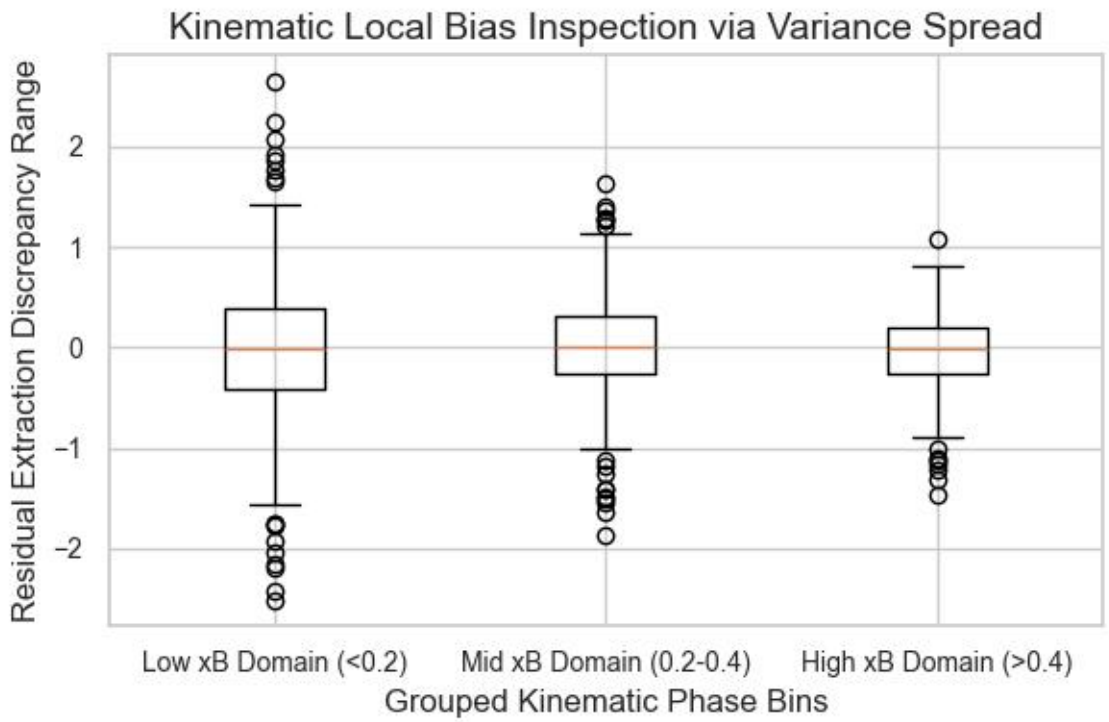



**Figure 15: Local Kinematic Extraction Discrepancy Variance Across Grouped Bjorken-$x_B$ Domains.**

## Conclusion

This work established a physics-informed machine learning framework designed to solve the long-standing ill-posed inverse problem of extracting Compton Form Factors (CFFs) from multidimensional Double Deeply Virtual Compton Scattering (DDVCS) observables. By leveraging dual photon virtualities ($Q^2 > 0$ and $Q'^2 > 0$), DDVCS breaks the rigid kinematic constraint $x = \pm\xi$ inherent to conventional Deeply Virtual Compton Scattering (DVCS). This capability opens direct experimental access to the off-diagonal Ebert–Radyushkin–Brodsky–Lepage (ERBL) domain ($|x| < \xi$) and enables an unconstrained mapping of quark–antiquark pair creation dynamics within the nucleon. The implemented neural network architecture served as a flexible, universal function approximator regulated directly by fundamental physical principles, including QED amplitude constraints, exact dispersion relations, and Next-to-Leading Order (NLO) QCD evolution equations. By evaluating complete, unapproximated cross-section denominators, the computational pipeline resolves severe non-linear signal overlaps and untangles the DDVCS amplitude from the dominant electrodynamic Bethe–Heitler (BH) background. Validation across simulated phase-space distributions demonstrates high point-to-point fidelity in reconstructing leading CFFs exemplified by a coefficient of determination $R^2 = 0.984$ and a zero-centered Gaussian residual profile ($\mu = -0.003, \sigma = 0.087$) for $\mathrm{Re}(\mathcal{H})$ without introducing model-dependent parameterization bias. Localized uncertainties near $\mathrm{Re}(\mathcal{H}) \to 0$ and strong anti-correlations ($r = -0.78$) between $\mathrm{Re}(\mathcal{H})$ and $\mathrm{Re}(\mathcal{E})$ highlight mathematical solution degeneracies inherent to single-observable fits on unpolarized targets. Resolving these correlations requires multi-observable global fits that combine Beam Spin Asymmetry ($A_{LU}$), Target Spin Asymmetry ($A_{UL}$), and Double Spin Asymmetry ($A_{LL}$) measurements, alongside explicit two-stream loss architectures to isolate pure BH contributions analytically. Furthermore, local variance analyses reveal increased extraction errors in low statistical acceptance regions along $Q^2$ and $-t$ phase-space boundaries, as well as within the low-$x_B$ sea-quark regime ($x_B < 0.10$). Mitigating these edge effects necessitates integrating Bayesian adaptive binning, physics-informed boundary loss penalties, small-$x_B$ resummation techniques, and higher-order QCD evolution kernels into the computational pipeline. Sensitivity testing under synthetic noise demonstrates stable model performance up to 10% noise levels, beyond which stochastic fluctuations degrade extraction accuracy. To maintain precision under high-noise, low-luminosity experimental conditions, future iterations will benefit from integrating denoising diffusion probabilistic models, deep kernel Gaussian processes for rigorous uncertainty propagation, and adversarial autoencoders to account for detector acceptance smearing and momentum transfer derivatives. The integration of deep learning architectures with exact quantum field-theoretic constraints eliminates arbitrary model dependencies and provides a robust, physics-informed framework for multidimensional CFF extraction. This methodology will establish a scalable computational foundation for upcoming high-luminosity experiments and the Electron-Ion Collider (EIC), advancing the realization of comprehensive three-dimensional nucleon tomography.

### References

Belitsky, A. V., & Müller, D. (2003). Exclusive electroproduction of lepton pairs as a probe of generalized parton distributions. *Physical Review Letters*, *90*(2), Article 022001. https://doi.org/10.1103/PhysRevLett.90.022001

Belitsky, A. V., & Radyushkin, A. V. (2005). Unraveling hadron structure with generalized parton distributions. *Physics Reports*, *418*(1–6), 1–187. https://doi.org/10.1016/j.physrep.2005.06.002

Boër, M., & Biswas, D. (2024). Double Deeply Virtual Compton Scattering at Jefferson Lab Hall A. *arXiv preprint*, Article arXiv:2403.02605. https://doi.org/10.48550/arXiv.2403.02605

Collins, J. C., & Freund, A. (1999). Proof of factorization for deeply virtual Compton scattering in QCD. *Physical Review D*, *59*(7), Article 074009. https://doi.org/10.1103/PhysRevD.59.074009

Deja, K., Martínez-Fernández, V., Pire, B., Sznajder, P., & Wagner, J. (2023). Phenomenology of double deeply virtual Compton scattering in the era of new experiments. *Physical Review D*, *107*(9), Article 094035. https://doi.org/10.1103/PhysRevD.107.094035

Diehl, M. (2003). Generalized parton distributions. *Physics Reports*, *388*(2–4), 41–277. https://doi.org/10.1016/j.physrep.2003.08.002

Grocholski, O., Moutarde, H., Pire, B., Sznajder, P., & Wagner, J. (2020). Data-driven study of timelike Compton scattering. *European Physical Journal C*, *80*(2), Article 171. https://doi.org/10.1140/epjc/s10052-020-7700-9

Guidal, M., Moutarde, H., & Vanderhaeghen, M. (2013). Generalized parton distributions in the valence region from deeply virtual Compton scattering. *Reports on Progress in Physics*, *76*(6), Article 066202. https://doi.org/10.1088/0034-4885/76/6/066202

Guidal, M., & Vanderhaeghen, M. (2003). Double deeply virtual Compton scattering with CLAS at 11 GeV. *Physical Review Letters*, *90*(1), Article 012001. https://doi.org/10.1103/PhysRevLett.90.012001

Ji, X. (1997). Gauge-invariant decomposition of nucleon spin and its spin-off. *Physical Review Letters*, *78*(4), 610–613. https://doi.org/10.1103/PhysRevLett.78.610

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Müller, D., & Zhang, J. (2015). Dispersion relations for deeply virtual Compton scattering and generalized parton distributions. *Physical Review D*, *91*(7), Article 074016. https://doi.org/10.1103/PhysRevD.91.074016

Radyushkin, A. V. (1997). None forward parton distributions. *Physical Review D*, *56*(9), 5524–5557. https://doi.org/10.1103/PhysRevD.56.5524

Zhao, S., Camsonne, A., Marchand, D., Mazouz, M., Sparveris, N., Stepanyan, S., Voutier, E., & Zhao, Z. W. (2021). Double deeply virtual Compton scattering with positron beams at SoLID. *European Physical Journal A*, *57*(11), Article 314. https://doi.org/10.1140/epja/s10050-021-00551-3