

Reconstruction of gravitational form factors using generative machine learning

Herzallah Alharazin[✉] and Julia Yu. Panteleeva[✉]

Institut für Theoretische Physik II, Ruhr-Universität Bochum, D-44780 Bochum, Germany



(Received 8 March 2026; accepted 28 April 2026; published 2 June 2026)

We develop a generative framework based on denoising diffusion for the model-independent reconstruction of hadronic form factors from sparse and noisy data. The generative prior is built from a large ensemble of synthetic curves drawn from ten distinct functional classes rooted in different theoretical approaches to hadron structure. Applied to the proton gravitational form factors $A(t)$, $J(t)$, and $D(t)$, the framework yields nonparametric reconstructions consistent with lattice QCD across the full kinematic range $0 \leq -t \leq 2 \text{ GeV}^2$, remaining robust even when only one or two conditioning points are retained. The densely sampled output enables a direct extraction of the chiral low-energy constants $c_8 = -4.6 \pm 0.8 \text{ GeV}^{-1}$ and $c_9 = -0.61 \pm 0.19 \text{ GeV}^{-1}$. Using these values at the physical pion mass, we obtain $D(0) = -4.3 \pm 0.8$ for the nucleon D -term.

DOI: [10.1103/glj2-2ww2](https://doi.org/10.1103/glj2-2ww2)

I. INTRODUCTION

The hadronic matrix elements of the energy-momentum tensor (EMT) have been studied intensively over the past decades, both theoretically and experimentally. While much attention has been devoted to the physical interpretation of the Fourier transforms of the gravitational form factors (GFFs) in terms of spatial density distributions of energy (mass), spin, pressure, and shear forces inside hadrons, see, e.g., Refs. [1–4], the GFFs themselves, as functions of momentum transfer, are interesting in their own right. In particular, they can be extracted from generalized parton distributions (GPDs) [5–8], which are experimentally accessible through exclusive processes such as deeply virtual Compton scattering (DVCS) [6,9–11] and hard exclusive meson production [12]. Knowing the GFFs of a system that can be probed experimentally, such as the nucleon, establishes a direct connection between theoretical predictions and experiment, thereby testing the validity and consistency of the underlying theory.

On the theory side, there exist two main model-independent approaches to obtain the nucleon GFFs, each with its own limitations. The first is chiral perturbation theory (ChPT), which is valid at low energies but suffers from the presence of unknown low-energy constants (LECs). While it can predict the leading nonanalytic behavior of the GFFs at low energies, and hence the

long-range behavior of their Fourier transforms in the chiral limit, it cannot provide definitive statements about the values of the GFFs near $t \approx 0$ [13–15].

The second approach is lattice QCD. It performs well for the form factors $A(t)$ and $J(t)$, whose values at $t = 0$ are constrained to 1 and $1/2$, respectively, by the operator structure, and it covers a wide range of momentum transfer t . However, current lattice QCD calculations of the GFFs suffer from several systematic limitations: most rely on a single gauge ensemble at a single lattice spacing and volume, preventing a controlled continuum and infinite-volume extrapolation. Moreover, most calculations employ unphysical pion masses. These limitations, combined with the intrinsically noisy signal, yield large uncertainties near $t = 0$ for $D(t)$, whose extrapolation from the smallest accessible $|t|$ to $t = 0$ introduces a model dependence [16,17].

Unlike $A(t)$ and $J(t)$, the value of $D(t)$ at $t = 0$ is not constrained by the operator structure. This value, the so-called D -term, represents a fundamental global property of the nucleon, considered as important as its mass and charge [18–21]. These limitations motivate the development of complementary frameworks that can provide reliable, parametrization-independent predictions for the full set of nucleon GFFs, in particular for the D -term. Ideally, such a framework should possess two key properties. First, it should be capable of producing robust reconstructions from a small number of precisely determined data points, making optimal use of their accuracy rather than relying on dense kinematic coverage. As will be demonstrated in this work, the constraining power of a few well-placed, high-precision measurements can rival or exceed that of a larger but noisier dataset. Second, it should be able to combine the sparse first-principles data from

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

lattice QCD with the analytic constraints from ChPT, without committing to a specific functional form for the t -dependence.

Denoising diffusion probabilistic models (DDPMs) [22–24] have emerged as the leading class of deep generative models for high-resolution image synthesis [25,26], owing to their ability to learn complex, high-dimensional probability distributions and to generate diverse samples from conditional posteriors. These same properties make them natural candidates for inverse problems in the natural sciences, where a small number of noisy observations must be completed into a full function or field configuration. In fundamental physics, diffusion models have recently entered the domain of lattice field theory, where the formal connection between the denoising process and stochastic quantization has been established for scalar and Abelian gauge fields and subsequently exploited for the generation of non-Abelian lattice gauge configurations [27–30].

In this work, we develop a physics-guided DDPM framework for the model-independent reconstruction of hadronic form factors from sparse and noisy data. We stress that *model independent* refers to the reconstruction of the form factors: no functional ansatz is imposed on their t -dependence, and the result does not rely on any particular dynamical model of hadron structure. The word *model* in *diffusion model*, by contrast, refers solely to the neural-network architecture that performs the generation. The method treats the form factor as a function discretized on a dense grid and casts the reconstruction as a conditional generation problem, analogous to image inpainting with only a small fraction of observed pixels. The central idea is to construct the training prior from a large and diverse ensemble of synthetic form-factor curves drawn from multiple distinct functional families, each grounded in a different theoretical or phenomenological approach to hadron structure. The resulting framework is general and can, in principle, be applied to any form factor for which such a physics-motivated prior can be assembled. As a first and stringent test, we apply it to the three gravitational form factors of the proton, $A(t)$, $J(t)$, and $D(t)$, conditioning on lattice QCD data from Ref. [17] and, where appropriate, on constraints from the Poincaré algebra and ChPT [13,14].

Our main results are as follows: first, the diffusion model provides a fully nonparametric reconstruction of all three GFFs at the unphysical pion mass of the lattice ensemble, consistent with the lattice data and with standard parametric fits across the full kinematic range $0 \leq -t \leq 2 \text{ GeV}^2$. A systematic data-ablation protocol demonstrates that the reconstruction remains stable even when only one or two conditioning points are retained, revealing the strong constraining power of the physically informed training prior. Second, the continuous, covariance-resolved output of the diffusion model enables a direct extraction of the leading unknown low-energy constants of the chiral Lagrangian, c_8 and c_9 , by matching to chiral perturbation theory in the

low- $|t|$ region. The resulting values are in agreement with the model-independent dispersive determination of Cao *et al.* [31], despite the two approaches being methodologically entirely distinct. Third, using the extracted LECs to evaluate the chiral expressions at the physical pion mass, we obtain a prediction for the nucleon D -term, $D(0) = -4.3 \pm 0.8$, consistent with the ChPT bound of Ref. [32] and with existing lattice and dispersive extractions.

The remainder of this paper is organized as follows: in Sec. II we provide a brief background on the gravitational form factors of hadrons. In Sec. III we describe the construction of the training dataset from ten classes of physically motivated functional forms and their convex combinations. Section IV presents the DDPM architecture, including the v -prediction formulation, the dual conditioning mechanism, and the error-propagation scheme. In Sec. V we demonstrate the model-independent reconstruction of the proton GFFs at unphysical pion mass and extract the LECs c_8 and c_9 . In Sec. VI we reconstruct the GFFs at physical pion mass. Network architecture, training procedure, ablation studies on synthetic test curves are collected in Appendixes A and B.

II. GRAVITATIONAL FORM FACTORS OF HADRONS

The matrix element of the conserved and symmetric EMT, $T^{\mu\nu}$, between hadronic states of the same species can be parametrized in terms of independent, conserved, Lorentz-invariant structures. The corresponding coefficients are the GFFs. A comprehensive and systematic analysis of the parametrization of EMT matrix elements for arbitrary spin in terms of GFFs is provided in Ref. [33].

As mentioned in the introduction, the GFFs are of broad theoretical and phenomenological interest. Their Fourier transforms admit an interpretation in terms of spatial density distributions inside hadrons. Moreover, they are directly connected to GPDs and are, in low-energy regimes, accessible within ChPT when applicable to the hadronic system under consideration. These features enable applications that extend beyond hadronic physics, including astrophysical phenomena such as neutron star mergers [34], connections to fundamental aspects of general relativity (see, e.g., Ref. [35]), and applications to the theory of hard exclusive processes and physics of exotic hadro-charmonia [36–38].

To illustrate some of these features we consider the matrix element of the EMT between spin- $\frac{1}{2}$ one-particle states, such as the nucleon, which can be decomposed as [6,18,19]

$$\begin{aligned} \langle p' | T^{\mu\nu}(0) | p \rangle = & \bar{u}(p') \left[A(t) \frac{P^\mu P^\nu}{m_N} + J(t) \frac{i(P^\mu \sigma^{\nu\alpha} + P^\nu \sigma^{\mu\alpha}) \Delta_\alpha}{2m_N} \right. \\ & \left. + D(t) \frac{\Delta^\mu \Delta^\nu - g^{\mu\nu} \Delta^2}{4m_N} \right] u(p), \end{aligned} \quad (1)$$

where $P = (p + p')/2$ is the average momentum, $\Delta = p' - p$ is the momentum transfer with $t = \Delta^2 < 0$ in the spacelike region, m_N is the nucleon mass, and $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$. The three form factors $A(t)$, $J(t)$, and $D(t)$ are scalar functions of the Lorentz invariant t and carry direct physical meaning:

- (i) $A(t)$ is related to the distribution of energy and momentum inside the nucleon. Conservation of the total four-momentum of the nucleon requires $A(0) = 1$.
- (ii) $J(t)$ encodes the distribution of total angular momentum. For a spin- $\frac{1}{2}$ state, the Poincaré algebra fixes $J(0) = \frac{1}{2}$.
- (iii) $D(t)$ governs the spatial distribution of internal forces acting inside the hadron [20,21]. Its forward-limit value $D(0)$, the so-called D -term, is not constrained by any spacetime symmetry and must be determined dynamically. It is considered a fundamental mechanical property of the nucleon, on par with its mass and charge.

Various approaches and choices of coordinate frames exist for defining spatial density distributions from these GFFs; see, e.g., Refs. [1,2,21,39–47].

The above GFFs are not directly measurable in scattering off gravitons but can be extracted indirectly through their connection to GPDs via the second Mellin moments [5–7] as follows:

$$\begin{aligned} \int_{-1}^1 dx x H(x, \xi, t) &= A(t) + \xi^2 D(t), \\ \int_{-1}^1 dx x E(x, \xi, t) &= 2J(t) - A(t) - \xi^2 D(t), \end{aligned} \quad (2)$$

where H and E are the unpolarized GPDs, x is the average longitudinal momentum fraction, and ξ is the skewness variable. Determining the GFFs, and in particular

the D -term, from either experiment or first-principles calculations remains challenging: current experimental extractions, whether from DVCS [11] or from elastic scattering data (see, e.g., Ref. [48]), rely on specific GPD parametrizations, while lattice QCD computations yield results at discrete values of t with sizable uncertainties near $t = 0$. Reconstructing the full t -dependence from such sparse and noisy data without committing to a specific functional ansatz is the central goal of the present work. To this end, we construct a diverse training prior from multiple families of physically motivated functions, described in the next section.

III. TRAINING DATA: CLASSES OF FUNCTIONAL FORMS

The diffusion model learns the manifold of physically plausible GFF curves from a training set of 6×10^5 synthetic form factors, sampled on a uniform grid of $N_t = 200$ points in $0 \leq -t \leq 2 \text{ GeV}^2$. The training curves are drawn from ten classes of functional forms: eight rooted in distinct theoretical or phenomenological approaches to hadron form factors, and two built from convex combinations that populate the gaps in shape space between the theory-motivated families. For each class, the free parameters are sampled from multiple randomized strategies, typically five to nine per parameter, covering uniform, log-uniform, Gaussian, physically motivated, and correlated joint distributions, ensuring dense coverage of the accessible shape space without overpopulating any single region. The forward-limit value $\text{GF}(0)$ is drawn from a broad mixture of distributions spanning $[-10, 10]$. All curves are subject to the acceptance criteria $|\text{GF}(t)| \leq 10$, absence of numerical singularities, and a minimum variation threshold that removes effectively constant functions.

The eight base functional forms are

$$\text{Multipole: } \text{GF}(t) = \frac{\text{GF}(0)}{(1 - t/M^2)^n}, \quad z\text{-exp.: } \text{GF}(t) = \sum_{k=0}^K a_k z(t)^k, \quad (3a)$$

$$\text{Meson dom.: } \text{GF}(t) = \sum_{i=1}^N \frac{c_i}{m_i^2 - t}, \quad \text{Mod.exp.: } \text{GF}(t) = \text{GF}(0)(1 - t/\Lambda^2)^\beta e^{\gamma t/\Lambda^2}, \quad (3b)$$

$$\text{Padé: } \text{GF}(t) = \frac{a_0 + \sum_{i=1}^N a_i (-t)^i}{1 + \sum_{j=1}^M |b_j| (-t)^j}, \quad \text{Dispersive: } \text{GF}(t) = \frac{1}{\pi} \int_{t_{\text{cut}}}^{s_{\text{max}}} ds \frac{\rho(s)}{s - t}, \quad (3c)$$

$$\text{Log-mod.: } \text{GF}(t) = \frac{\text{GF}(0)}{(1 - t/M^2)^n} [1 + c \ln(1 - t/M^2)]^\delta, \quad \text{Bag: } \text{GF}(t) = \text{GF}(0) \frac{3j_1(R\sqrt{-t})}{R\sqrt{-t}} e^{-\beta(-t)}. \quad (3d)$$

In Eq. (3a), $z(t) = (\sqrt{t_{\text{cut}} - t} - \sqrt{t_{\text{cut}} - t_0})/(\sqrt{t_{\text{cut}} - t} + \sqrt{t_{\text{cut}} - t_0})$ is the conformal variable that maps the cut t -plane onto the unit disk. In Eq. (3c), the spectral function is modeled as $\rho(s) = (s - t_{\text{cut}})^p \sum_i c_i \Gamma_i / [(s - m_i^2)^2 + (\Gamma_i/2)^2] e^{-s/s_{\text{max}}}$, combining a threshold factor $(s - t_{\text{cut}})^p$, Breit-Wigner peaks Γ_i , and a high-energy damping s_{max} .

The motivation for including each class is as follows: the *multipole* (3a) is the standard parametrization in lattice QCD extractions of GFFs [17,49–51] and therefore must be well represented in the prior. The *z-expansion* (3a) is the most general parametrization consistent with the analyticity of the form factor [52–54] and is used alongside the multipole in the lattice analysis of Ref. [17]; including it ensures that the prior covers the full space of analytic functions, not only the pole-dominated subset. Because $|z_{\max}| \approx 0.39$ on the physical grid, naïve coefficient sampling produces nearly linear curves; we compensate it by rescaling $a_k \rightarrow a_k/(|z_{\max}|^k)^\alpha$, $\alpha \in [0.3, 1]$, so that each order contributes $\mathcal{O}(1)$ to the shape variation. *Meson dominance* (3b) represents the GFF as a sum of *t*-channel meson exchanges with the appropriate quantum numbers [55–57]; for $\approx 70\%$ of the $N \geq 3$ samples the residues satisfy the superconvergence sum rule $\sum_i c_i = 0$ [55,58], guaranteeing the correct perturbative QCD (pQCD) falloff $\text{GF}(t) \sim 1/(-t)^2$. The *modified exponential* (3b) captures the Gaussian-like shapes that arise in constituent-quark models and soft-wall anti-de Sitter (AdS)/QCD [59,60], with the polynomial prefactor $(1 - t/\Lambda^2)^\beta$ resolving the vanishing-slope problem of the pure exponential at $t = 0$. *Padé approximants* (3c) provide a systematic rational-function generalization that encompasses multipoles and the Kelly parametrization [61–64]; the use of $|b_j|$ in the denominator guarantees the absence of spacelike poles by construction, and the condition $M \geq N + 2$ (imposed for $\approx 75\%$ of samples) enforces the correct large- $(-t)$ asymptotics. The *dispersive representation* (3c) exploits analyticity through an unsubtracted dispersion relation, following the approach of Refs. [65–68]; its inclusion ensures that the prior contains curves shaped by the two-pion threshold and resonance interference patterns that are absent from the purely algebraic families. The *log-modified multipole* (3d) incorporates the logarithmic corrections from the running of α_s at large $-t$ [58,69,70], recovering the pure multipole in the limit $c \rightarrow 0$ or $\delta \rightarrow 0$. Finally, the *bag-model/Bessel-type* form (3d) is motivated by the MIT bag model [71–74], whose GFFs arise from Fourier transforms of step-functionlike spatial distributions; its characteristic mild shoulder near $-t \sim 1/R^2$ introduces a qualitative feature absent from all smooth monotone families.

To fill the remaining gaps in shape space between the eight base classes, we generate two additional families of convex combinations. Class 9 consists of pairwise combinations $\lambda f_A + (1 - \lambda)f_B$ with $\lambda \sim \text{U}(0, 1)$, drawn from all $\binom{8}{2} = 28$ pairs. Class 10 uses Dirichlet-weighted mixtures of M randomly chosen parent families, with the concentration parameter itself randomized to interpolate between near-uniform and peaked weightings. In total, 6×10^4 accepted curves are generated per class, yielding 6×10^5 training samples; a separate validation set of 6×10^3 curves

is held out for model selection. Each curve is normalized to zero mean and unit variance per grid point before training.

The complete generation code, including all parameter sampling strategies, acceptance criteria and analysis of the final results, is provided in Ref. [75] in fully reproducible form.

IV. DENOISING DIFFUSION PROBABILISTIC MODEL

We employ a denoising diffusion probabilistic model [23] with *v*-prediction [76] to reconstruct GFF curves from sparse data. This section summarizes the formalism; for a comprehensive treatment, see Refs. [23,26,76].

A. Forward process and *v*-prediction

Let $x_0 \in \mathbb{R}^L$ ($L = 200$) denote a GFF curve sampled on a uniform grid. The forward (noising) process produces a noisy version at diffusion step $\tau \in \{1, \dots, T\}$, with $T = 1000$, via

$$x_\tau = \sqrt{\bar{\alpha}_\tau} x_0 + \sqrt{1 - \bar{\alpha}_\tau} \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}),$$

$$\bar{\alpha}_\tau = \prod_{s=1}^{\tau} \alpha_s, \quad \alpha_s = 1 - \beta_s, \quad (4)$$

where ε denotes the pure noise drawn from a standard normal distribution $\mathcal{N}(0, \mathbf{I})$, $\bar{\alpha}_\tau$ is the signal preservation coefficient and $\beta_s \in (0, 1)$ is the noise variance at step s . We adopt the cosine noise schedule of Ref. [26],

$$\bar{\alpha}_\tau = \frac{f(\tau)}{f(0)}, \quad f(\tau) = \cos^2\left(\frac{\tau/T + s\pi}{1 + s} \frac{\pi}{2}\right), \quad (5)$$

with offset $s = 0.008$, which ensures a smooth transition from $\bar{\alpha}_0 \approx 1$ (pure signal) to $\bar{\alpha}_T \approx 0$ (pure noise). Rather than predicting the noise ε or the clean data x_0 directly, our network predicts the *velocity*,

$$v_\tau = \sqrt{\bar{\alpha}_\tau} \varepsilon - \sqrt{1 - \bar{\alpha}_\tau} x_0, \quad (6)$$

which has uniform variance across all noise levels, avoiding the signal-to-noise imbalance that afflicts ε - and x_0 -prediction at opposite ends of the diffusion schedule. From a predicted $\hat{v}_\tau$ one obtains the predicted

$$\hat{x}_0 = \sqrt{\bar{\alpha}_\tau} x_\tau - \sqrt{1 - \bar{\alpha}_\tau} \hat{v}_\tau, \quad (7)$$

$$\hat{\varepsilon} = \sqrt{1 - \bar{\alpha}_\tau} x_\tau + \sqrt{\bar{\alpha}_\tau} \hat{v}_\tau, \quad (8)$$

which can be verified by substituting Eqs. (4)–(6). The relations in Eq. (7) are exact identities: given the true v_τ , one recovers x_0 and ε without approximation. At inference time the network provides an approximation $\hat{v}_\tau$, so the recovered quantities are likewise approximated.

B. Reverse sampling

To generate a GFF curve one must run the noising process in reverse, stepping from pure noise x_T back to clean data x_0 . The one-step reverse distribution $q(x_{\tau-1}|x_\tau)$, the probability of the less noisy state given the current noisy state, is intractable, since it depends on the unknown data distribution. However, if one additionally conditions on the clean data x_0 , Bayes' theorem yields a closed-form Gaussian (see, e.g., Ref. [23]), as follows:

$$q(x_{\tau-1}|x_\tau, x_0) = \mathcal{N}(x_{\tau-1}; \tilde{\mu}_\tau, \tilde{\beta}_\tau \mathbf{I}), \quad (9)$$

referred to as the *posterior*,¹ with mean and variance

$$\tilde{\mu}_\tau = \frac{\sqrt{\tilde{\alpha}_{\tau-1}}\beta_\tau}{1 - \tilde{\alpha}_\tau} x_0 + \frac{\sqrt{\tilde{\alpha}_\tau}(1 - \tilde{\alpha}_{\tau-1})}{1 - \tilde{\alpha}_\tau} x_\tau, \quad (10)$$

$$\tilde{\beta}_\tau = \frac{\beta_\tau(1 - \tilde{\alpha}_{\tau-1})}{1 - \tilde{\alpha}_\tau}. \quad (11)$$

Of course x_0 is unknown at inference time, it is precisely what we wish to reconstruct. The neural network f_θ fills this gap: at each step it predicts $\hat{v}_\tau$, from which an estimate $\hat{x}_0$ is obtained via Eq. (7) and substituted for x_0 in Eq. (10). Starting from $x_T \sim \mathcal{N}(0, \mathbf{I})$, each reverse step reads²

$$x_{\tau-1} = \tilde{\mu}_\tau + \sqrt{\tilde{\beta}_\tau} z, \quad z \sim \mathcal{N}(0, \mathbf{I}), \quad \tau = T, \dots, 2, \quad (12)$$

with $x_0 = \tilde{\mu}_1$ at the final step (no noise added).

C. Conditioning and error propagation

Our goal is to reconstruct the full GFF curve given K known data points³ $\{(t_k, F_k \pm \sigma_k)\}_{k=1}^K$ from, e.g., lattice QCD or ChPT. The model conditions on this information through two complementary mechanisms. First, we have the concatenation conditioning: each known value is mapped to the nearest grid index j_k and normalized to the model's internal representation: $\tilde{F}_k = (F_k - \mu_{j_k})/\sigma_{j_k}$, where μ and σ are the per-grid-point mean and standard deviation of the training set. The network receives three input channels at every reverse step: the current noisy estimate x_τ , a binary mask $m \in \{0, 1\}^L$ with $m_{j_k} = 1$ at known positions, and a condition vector $c \in \mathbb{R}^L$ with $c_{j_k} = \tilde{F}_k$ and zero elsewhere. During training, random masks (50% uniformly random, 30% clustered in the low- $|t|$ region, 20% unconditional) teach the network to

reconstruct missing values conditioned on an arbitrary subset of known points. To propagate the uncertainties of the known data into the reconstructed uncertainty band, each of the N generated samples receives its own realization of the conditioning values,⁴

$$c_n^{(j_k)} = \tilde{F}_k + \xi_n, \quad \xi_n \sim \mathcal{N}(0, \tilde{\sigma}_k^2), \quad \tilde{\sigma}_k = \sigma_k/\sigma_{j_k}, \quad (13)$$

where σ_k is the error of the known data and $\tilde{\sigma}_k$ its normalized counterpart. This noise injection is drawn once per sample and held fixed throughout the reverse chain: both the concatenation input c and the replacement values use the same c_n . Then we have the replacement part. At each intermediate reverse step $\tau \rightarrow \tau - 1$, the known grid points in $x_{\tau-1}$ are overwritten by a forward-noised version of the conditioning values at the appropriate noise level:

$$x_{\tau-1}^{(j)} \leftarrow \begin{cases} \sqrt{\tilde{\alpha}_{\tau-1}} c_n^{(j)} + \sqrt{1 - \tilde{\alpha}_{\tau-1}} \eta_\tau, & j \in \{j_k\}, \\ x_{\tau-1}^{(j)}, & \text{otherwise,} \end{cases} \quad (14)$$

where $\eta_\tau \sim \mathcal{N}(0, \mathbf{I})$ is drawn independently at each reverse step. At the final step ($\tau = 1$) the clean conditioning values are inserted directly: $x_0^{(j_k)} \leftarrow c_n^{(j_k)}$.

It is important to emphasize that the two conditioning channels, concatenation and replacement, play complementary roles. The concatenation input (m, c_n) informs the network *globally* at every denoising step, allowing it to shape the predicted $\hat{v}_\tau$ over the entire grid in light of the known data. The replacement step enforces *local* consistency by anchoring the known grid points to their correct marginal distribution $q(x_{\tau-1}^{(j_k)}|c_n^{(j_k)})$ at each noise level, eliminating any residual drift at those positions. Because every sample n uses a different noised realization c_n , the empirical spread of the N generated curves encodes both the model's learned prior uncertainty and the propagated experimental errors.

Our network f_θ was trained to minimize the mean-squared error on the v -target,

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0, \varepsilon, \tau, m} [\|f_\theta(x_\tau, \tau, m, c) - v_\tau\|^2]. \quad (15)$$

Further details about the network architecture and training are given in Appendixes A and B. The trained model, sampling scripts, and all configuration details required to reproduce the results presented in this work are publicly available in Ref. [75].

¹In the Bayesian sense: q is the distribution of $x_{\tau-1}$ after conditioning on the observed quantities x_τ and x_0 . This is standard terminology in the generative-modeling literature [23,26,76].

²The fresh noise z injected at each step is the source of sample diversity: different random seeds yield different plausible GFF curves whose spread constitutes the model's uncertainty band.

³Here F_k refers to a (gravitational) form factor $F(t_k)$.

⁴It is worth mentioning that throughout the lattice reconstructions presented in the next section, the uncertainties of the conditioning points are propagated independently via Eq. (13), as the full covariance matrix of Ref. [17] is publicly not available; the impact of off-diagonal elements is expected to be subdominant, since the reconstruction remains stable under data reduction, as we will see below.

For each reconstruction presented below, $N = 10^4$ curves are generated from the reverse process. The point-wise 68% credible interval is defined by the 16th and 84th percentiles of the ensemble at each grid point. A jackknife resampling of the ensemble mean yields standard errors that are indistinguishable from the median curve at the scale of all figures shown; we therefore report the median as the central estimate throughout.

V. MODEL-INDEPENDENT RECONSTRUCTION OF THE PROTON GFFs AT UNPHYSICAL PION MASS

In this section, we demonstrate that the diffusion-based generative framework yields a fully *model-independent* reconstruction of the total proton GFFs $A(t)$, $J(t)$, and $D(t)$, consistent with the lattice QCD results of Ref. [17] across the full kinematic range $0 \leq -t \leq 2 \text{ GeV}^2$, without imposing any functional form on the t -dependence. The only inputs to the reconstruction are:

- (1) A limited small number of lattice data points from the total GFF results of Ref. [17].
- (2) Exact physical constraints that follow from the symmetry structure of the EMT, or, where no such constraint exists, well-motivated sign conditions supported by theoretical arguments and the data themselves.

Having learned the space of physically admissible form-factor shapes from the training prior (Sec. III), the diffusion model samples from the posterior distribution consistent with the supplied conditioning input, yielding a nonparametric estimate of both the central value and the pointwise uncertainty of the reconstructed form factor.

To probe the robustness of the reconstruction, we adopt a systematic *data-ablation protocol*: starting from the most informative conditioning set, we progressively remove lattice data points until only a single point and the relevant physical constraint remain (for $D(t)$ we retain two points). This protocol reveals how much information each data point contributes, exposes the regions of t where the reconstruction is driven by the physical prior rather than the data, and provides a transparent, data-driven quantification of the extrapolation uncertainty that parametric fits absorb into the assumed functional form.

We emphasize that for each available lattice dataset we select as conditioning input those points with only modest statistical uncertainties compared with the others, in other words, the reconstruction is not aided by large error bars that would trivially accommodate many functional shapes. The diffusion model must reproduce sharply defined values around the conditioning points while simultaneously generating physically sensible interpolations and extrapolations throughout the kinematic range. That it succeeds with as few as one or two data points is a direct consequence of the physically informed prior built into the training data, and constitutes a stringent validation of the approach.

A. Reconstruction of symmetry-constrained form factors $A(t)$ and $J(t)$

The form factors $A(t)$ and $J(t)$ benefit from exact forward-limit constraints that follow from the Poincaré algebra, as follows:

$$A(0) = 1, \quad J(0) = \frac{1}{2}. \quad (16)$$

The first one expresses conservation of total four-momentum, and $J(0) = 1/2$ follows from the proton being a spin- $\frac{1}{2}$ state whose total angular momentum is encoded in the forward matrix element of the EMT. That is, we impose them as exact conditioning points alongside a small number of lattice data points. The results of the data-ablation study are presented in Fig. 1, we observe:

Full conditioning: when conditioned on four lattice data points plus the forward-limit constraint (top panels), the diffusion-model median is virtually indistinguishable from both the dipole and z-expansion fits of Ref. [17] across the entire kinematic range. The 68% credible band tightly envelopes both parametrizations and, consequently, also encompasses all remaining lattice data points for $A(t)$ and $J(t)$ reported in Fig. 1 of Ref. [17] that were *not* used as conditioning input, confirming that the model has captured the correct shape of both form factors without being informed of any functional form.

Stability under data reduction: as conditioning points are progressively removed (second through third panels), the median remains remarkably stable and continues to track the parametric fits. The credible interval broadens precisely in the regions from which data have been removed, predominantly at large $-t$, providing a faithful, data-driven measure of the growing extrapolation uncertainty.

Reconstruction from minimal input: the most notable result appears in the bottom panels, where only a single lattice data point at intermediate $-t$ and the forward-limit constraint are supplied. Even in this extreme scenario the reconstruction yields a smoothly decaying curve whose median remains compatible with the full parametric fits. At $-t = 2 \text{ GeV}^2$ the credible band remains moderate for both form factors. The notably narrow uncertainty obtained with only two conditioning inputs, despite the wide range of mathematically allowed curves through those points, is a direct consequence of the physical prior encoded in the training data.

B. Reconstruction of D form factor

The form factor $D(t)$ presents the most stringent test of the reconstruction framework, for two reasons. First, unlike $A(t)$ and $J(t)$, the forward limit $D(0)$ is not fixed by any spacetime symmetry, i.e., it is a quantity that must be determined from experiment or first-principles calculation. Second, in the low-energy region relevant for the extrapolation to $D(0)$, the lattice data for $D(t)$ exhibit substantially

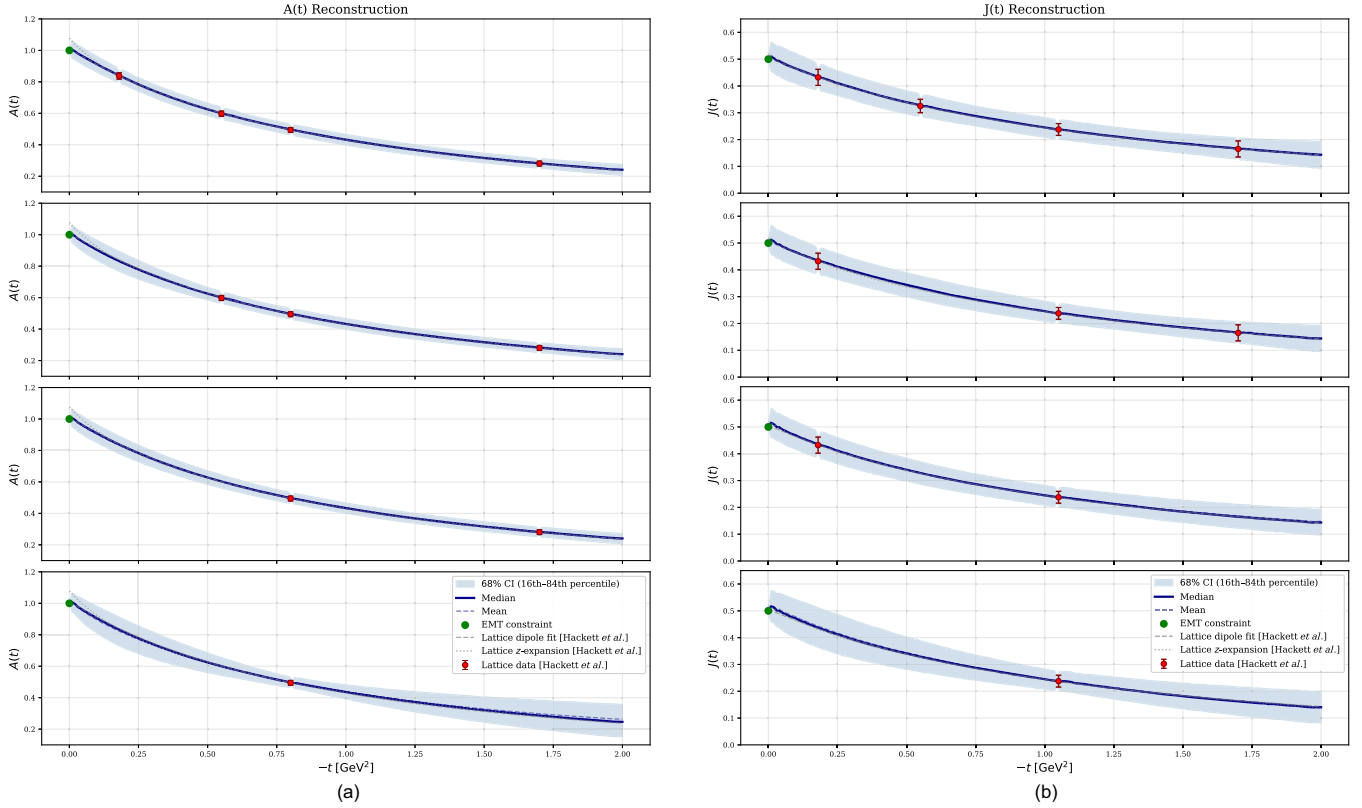


FIG. 1. Model-independent reconstruction of $A(t)$ and $J(t)$: (a) reconstruction of $A(t)$; (b) reconstruction of $J(t)$. In each column, the four panels correspond to progressively sparser conditioning: the top panel uses the largest subset of lattice data points (red markers), and successive panels remove points until only a single datum remains. The green marker denotes the constraint imposed by the Poincaré algebra. Gray dashed and dotted curves are the dipole and z -expansion fits of Ref. [17].

larger statistical uncertainties than those for $A(t)$ and $J(t)$. In combination, these features lead to a strong *sensitivity* of $D(0)$ to the choice of fitting ansatz.

In the absence of a sum rule, we impose physically motivated sign constraints. From the raw diffusion-model ensemble we retain only those samples satisfying the prescribed sign condition, without modifying the generative process. Two variants of constraints are considered:

- (i) Minimal constraint ($D(0) < 0$): the requirement that $D(0)$ be negative follows from the mechanical stability of the proton [21,38]: a positive D -term would imply a repulsive normal force and the system would collapse. This negativity is also supported by ChPT [32], chiral quark-soliton models [77,78], and the lattice data themselves. This weakest condition allows the form factor to fluctuate mildly at large $-t$.
- (ii) Full negativity ($D(t) < 0$): Negativity is enforced at every point on the spacelike grid. This stronger condition is motivated by all known model calculations, see, e.g., Refs. [31,65,78–80].

Figure 2(a) shows the reconstruction with the condition $D(0) < 0$ only. In the most densely conditioned scenario (top panel), in the region $-t > 0.3 \text{ GeV}^2$, where the lattice data are less noisy, the median closely tracks both

parametric fits and the credible band encompasses both across the entire $-t$ range. Requiring only $D(0) < 0$ restricts the reconstructed ensemble to negative values over nearly the entire kinematic range. A small positive deviation of the upper edge of the credible band appears only in the far high- $|t|$ tail, $-t \gtrsim 1.65 \text{ GeV}^2$, where the form factor is close to zero and the statistical spread is largest. However, the ensemble median and mean remain strictly negative at every value of t in all four conditioning scenarios. Imposing the stronger condition $D(t) < 0$ at every grid point [Fig. 2(b)] removes this residual fraction of samples and yields a modest tightening of the band for $-t \gtrsim 1.65 \text{ GeV}^2$. Away from this narrow corner of the kinematic range the two conditioning schemes are essentially indistinguishable. In particular, the median, mean, and credible interval in the vicinity of $t = 0$, where the physical interest in $D(0)$ resides, change only marginally between Figs. 2(a) and 2(b). This near-equivalence is nontrivial: it demonstrates that the diffusion-model prior, shaped by the physically motivated training data, already strongly disfavors zero crossings of $D(t)$ in the spacelike region. Requiring $D(0) < 0$ alone is therefore, within the resolution of the present analysis, sufficient to enforce negativity across the full t -range.

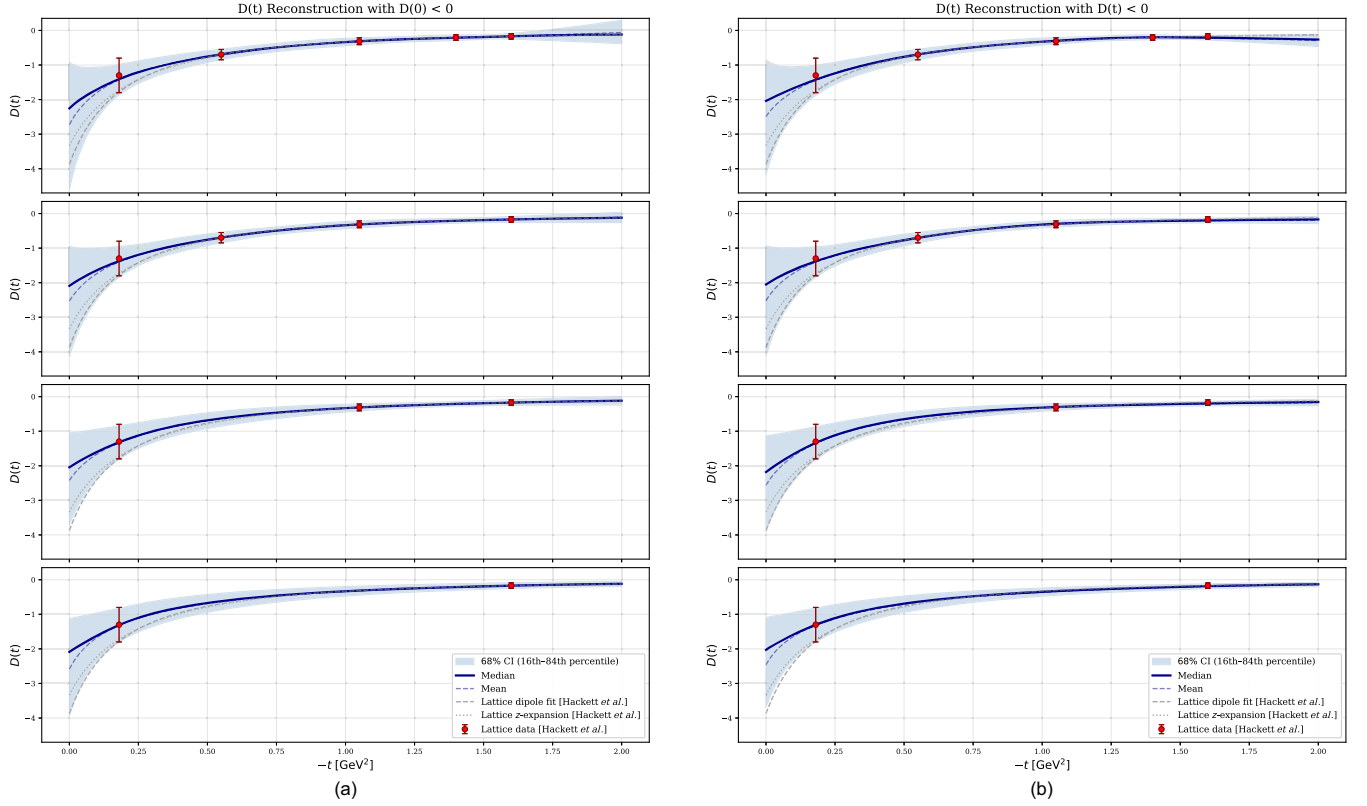


FIG. 2. Reconstruction of the $D(t)$ under two physics-informed criteria. (a) Only the D -term is negative. (b) Negativity is enforced at every point on the spacelike grid: $D(t) < 0$. Panel layout, markers, and bands are as in Fig. 1.

Across all four conditioning scenarios, the width of the credible band at small $-t$ remains essentially unchanged, reflecting the presence of the lowest- $|t|$ point in every reconstruction. This behavior has a clear physical origin: the form factor attains its largest absolute value in the forward region and must vanish at large $|t|$, so the bulk of the dynamic range is concentrated at low momentum transfer, and a single well-placed conditioning point there exerts maximal leverage on the reconstruction. By contrast, at large $|t|$ the form factor flattens toward zero and becomes comparatively insensitive to individual data points. This is also evident in the final panels of Fig. 1, where retaining only the $t = 0$ constraint and a single intermediate- $|t|$ lattice point already suffices to constrain the forward behavior.

These observations are corroborated by the ablation study on synthetic test curves presented in Appendix C. That analysis demonstrates that a reliable model-independent reconstruction requires, at minimum, two well-determined points: one in the forward region $-t \lesssim 0.25 \text{ GeV}^2$, which anchors the form-factor normalization and slope, and a second at intermediate momentum transfer $-t \sim 0.25\text{--}0.5 \text{ GeV}^2$, which stabilizes the interpolation into the mid- $|t|$ range. The inclusion of a third conditioning point at higher $|t|$ further tightens the credible band but yields a comparatively smaller marginal improvement, consistent with the flattening of the form factor in

that region. Conversely, when only high- $|t|$ data are available, the reconstruction in the physically critical forward region degrades substantially (see the bottom row of Fig. 8).

These findings may also be of relevance when planning future lattice QCD calculations of the gravitational form factors. Since the model-independent reconstruction is most sensitive to the precision of data in the low- and intermediate- $|t|$ region, investments in reducing statistical uncertainties at $-t \lesssim 0.5 \text{ GeV}^2$, for instance through increased statistics at small momentum transfer, would be particularly valuable for tightening the extrapolation to $t = 0$. Improvements at large $|t|$, while beneficial for the overall characterization of the form factor, have a more modest impact on the quantities of primary physical interest.

As a further application of our approach, we reconstruct the quark and gluon contributions to the $D(t)$ form factor using two lattice data points from Ref. [17] and compare our result with the model prediction of Ref. [81]. As can be seen from Figs. 3(a) and 3(b), the model-independent reconstruction from only two lattice data points is consistent with the tripole parametrization of Ref. [81]. This is again noteworthy because our method does not assume a specific functional form of the GFF.

The reconstructions obtained above provide a continuous and parametrization-independent estimate of the proton

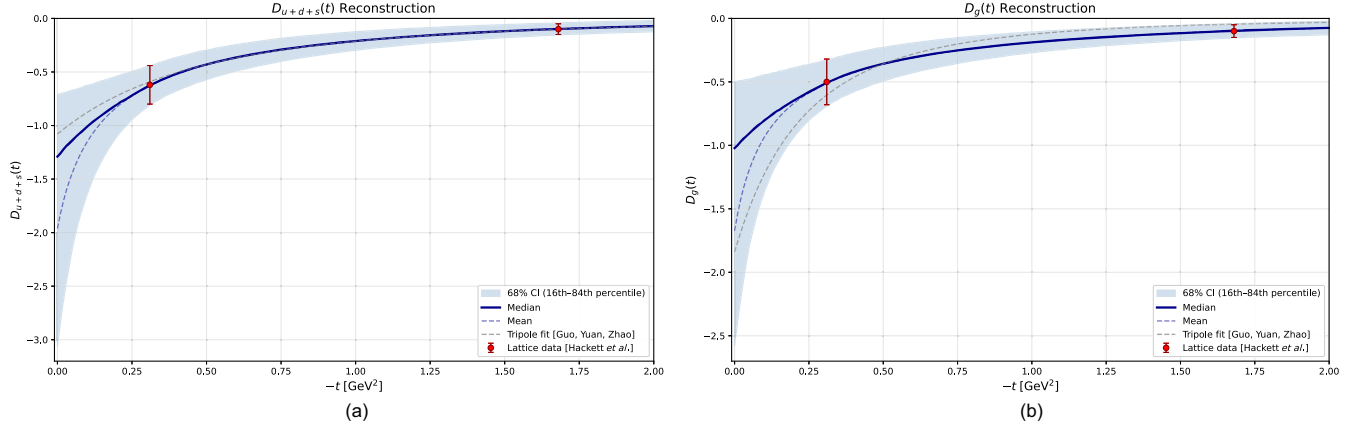


FIG. 3. Diffusion-model reconstruction of $D(t)$ at the unphysical pion mass, conditioned on the lattice data from Ref. [17], compared with the model prediction from Ref. [81]: (a) quark contribution to $D(t)$; (b) gluon contribution to $D(t)$.

GFFs together with their full covariance structure across the low- $|t|$ region where ChPT is applicable. This makes it possible to extract the leading unknown LECs of the effective chiral Lagrangian directly from the diffusion-model posterior, by evaluating the ChPT expressions at the unphysical pion mass of the lattice ensemble, and without relying on any specific parametric fit.

C. Extracting the unknown LECs c_8 and c_9

The low-energy constant c_9 is extracted by matching the diffusion-model reconstruction of $A(t)$ to our $\mathcal{O}(p^4)$ ChPT expression $A(t) = a(t) + c_9 b(t)$ as derived in Refs. [13,15], using generalized least squares (GLS) at five points in $-t \in [0.02, 0.10]$ GeV^2 , where the chiral truncation error remains below 1.3%. Using the diffusion-model covariance as the GLS weight matrix, we obtain

$$c_9 = -0.61^{+0.17}_{-0.17} (\text{stat.}) \pm 0.05 (\text{ChPT trunc.}) \pm 0.06 (\text{fit range}) \text{ GeV}^{-1}. \quad (17)$$

A cross-check using $J(t)$ at $\mathcal{O}(p^3)$ yields $c_9|_J = -0.40^{+0.41}_{-0.41} \text{ GeV}^{-1}$, consistent within uncertainties. For the physical-pion-mass analysis we combine the uncertainties in quadrature and adopt $c_9 = -0.61 \pm 0.19 \text{ GeV}^{-1}$. Similarly, c_8 is extracted from $D(t)$ at $\mathcal{O}(p^2)$ via $D(t) = a(t) + c_8 b(t)$ at four points in $-t \in [0.02, 0.08]$ GeV^2 , as follows:

$$c_8 = -4.6^{+0.6}_{-0.6} (\text{stat.}) \pm 0.3 (\text{ChPT trunc.}) \pm 0.4 (\text{fit range}) \text{ GeV}^{-1}. \quad (18)$$

This yields $c_8 = -4.6 \pm 0.8 \text{ GeV}^{-1}$ (all uncertainties added in quadrature). The larger reconstruction uncertainty near $t = 0$ ($\sigma_{\text{diff}} \sim 0.8\text{--}1.4$) reflects the absence of a sum-rule constraint on $D(0)$ and the large errors existing in the lattice QCD at the unphysical pion mass around $t \approx 0$.

A sharper determination of $D(t)$ with the present lattice data would require including the full set of generally covariant chiral Lagrangian contributions up to $\mathcal{O}(p^4)$, incorporating explicit Δ -resonance degrees of freedom, and performing a dedicated analysis to determine the additional low-energy constants that enter at this order [15].

It is instructive to compare these values with the model-independent determination by Cao *et al.* [31], who employ a data-driven dispersive framework in which the nucleon GFF spectral functions are constructed from $\pi\pi$ and $K\bar{K}$ intermediate states using Roy-Steiner equation solutions for the $\pi\pi/K\bar{K} \rightarrow N\bar{N}$ partial-wave amplitudes, matched to the ChPT expressions with external gravitational source. They obtain $c_8 = -4.28^{+0.37}_{-0.38} \text{ GeV}^{-1}$ and $c_9 = -0.68^{+0.06}_{-0.05} \text{ GeV}^{-1}$ [31]. Both values are in excellent agreement with our extractions, $c_8 = -4.6 \pm 0.8 \text{ GeV}^{-1}$ and $c_9 = -0.61 \pm 0.19 \text{ GeV}^{-1}$, despite the two approaches being methodologically entirely distinct: theirs reconstructs spectral functions from scattering data at the physical pion mass, while ours conditions a diffusion-model posterior on lattice QCD results at an unphysical pion mass. That the central values agree well within 1 standard deviation for both LECs provides a nontrivial cross-check for both determinations. In the next section, the extracted values of c_8 and c_9 are substituted into the ChPT expressions for the nucleon GFFs, which are then evaluated at the physical pion mass to provide approximate estimates, with uncertainties reflecting the limited high- $|t|$ information.

VI. THE PROTON GFFS AT PHYSICAL PION MASS

A. Reconstruction of GFFs from ChPT

The low-energy constants c_8 and c_9 , extracted by matching the $\mathcal{O}(p^2)$ – $\mathcal{O}(p^4)$ ChPT expressions for $A(t)$, $J(t)$, and $D(t)$ to the diffusion-model reconstruction at the lattice pion mass (see Sec. V C), are used to evaluate the proton GFFs at the physical point ($m_\pi \approx 139 \text{ MeV}$). The strategy proceeds in two steps. First, the ChPT

expressions evaluated at physical pion mass with the extracted LECs provide conditioning data concentrated in the low- $|t|$ region where the chiral expansion still has a converging behavior. Second, these points, together with the Poincaré constraints of Eq. (16) for $A(t)$ and $J(t)$, are fed into the diffusion model to reconstruct the full t -dependence over the range $0 \leq -t \leq 2 \text{ GeV}^2$.

Figures 4 and 5 present the reconstruction of all three GFFs at the physical pion mass. For each form factor, the three panels show the effect of progressively reducing the number of ChPT conditioning points, from the full set (top) to a minimal subset (bottom). For $A(t)$ and $J(t)$ [Figs. 4(a) and 4(b)], the Poincaré constraints at $t = 0$ anchor the normalization, and the reconstruction remains stable even as conditioning points are removed.

In all three form factors the credible band widens appreciably at larger $|t|$. This behavior is expected and can be traced to two factors. First, the most reliable conditioning data, i.e., those with the smallest uncertainties, are confined to the region $-t \lesssim 0.15 \text{ GeV}^2$; beyond this range the uncertainties on the conditioning points grow rapidly and their constraining power diminishes. Second, the cluster of points below $-t \approx 0.15 \text{ GeV}^2$ alone does not provide sufficient leverage for the model to extrapolate into the intermediate- $|t|$ region with high confidence. Indeed, the ablation study presented in Appendix C indicates that at

least one additional well-constrained point at $-t \gtrsim 0.25 \text{ GeV}^2$ would be required to significantly reduce the reconstruction uncertainty at intermediate momentum transfers.

The form factor $D(t)$ [Fig. 5(a)] presents a qualitatively different situation. Unlike $A(t)$ and $J(t)$, no sum-rule constraint fixes $D(0)$, so the reconstruction relies entirely on the ChPT conditioning data. Moreover, the available ChPT expression for $D(t)$ is of lower chiral order than those for $A(t)$ and $J(t)$, so that the higher- $|t|$ conditioning points probe a regime where the truncated chiral expansion is no longer reliable and the quoted uncertainties underestimate the true error. The data-ablation sequence reveals that removing these higher- $|t|$ points actually *improves* the reconstruction: the diffusion-model median is not pulled toward the potentially unreliable high- $|t|$ ChPT values but instead follows a trajectory shaped by the training prior, effectively filtering out the incomplete chiral information.

While the reconstruction of the full t -dependence of $D(t)$ is sensitive to the limitations discussed above, the value at $t = 0$ can be obtained directly from the extracted LEC without relying on the diffusion-model extrapolation. Taking the physical nucleon mass $m_N = 0.938 \text{ GeV}$ gives us the following value of the D -term of the nucleon:

$$D(0) = c_8 m_N = -4.3 \pm 0.8, \quad (19)$$

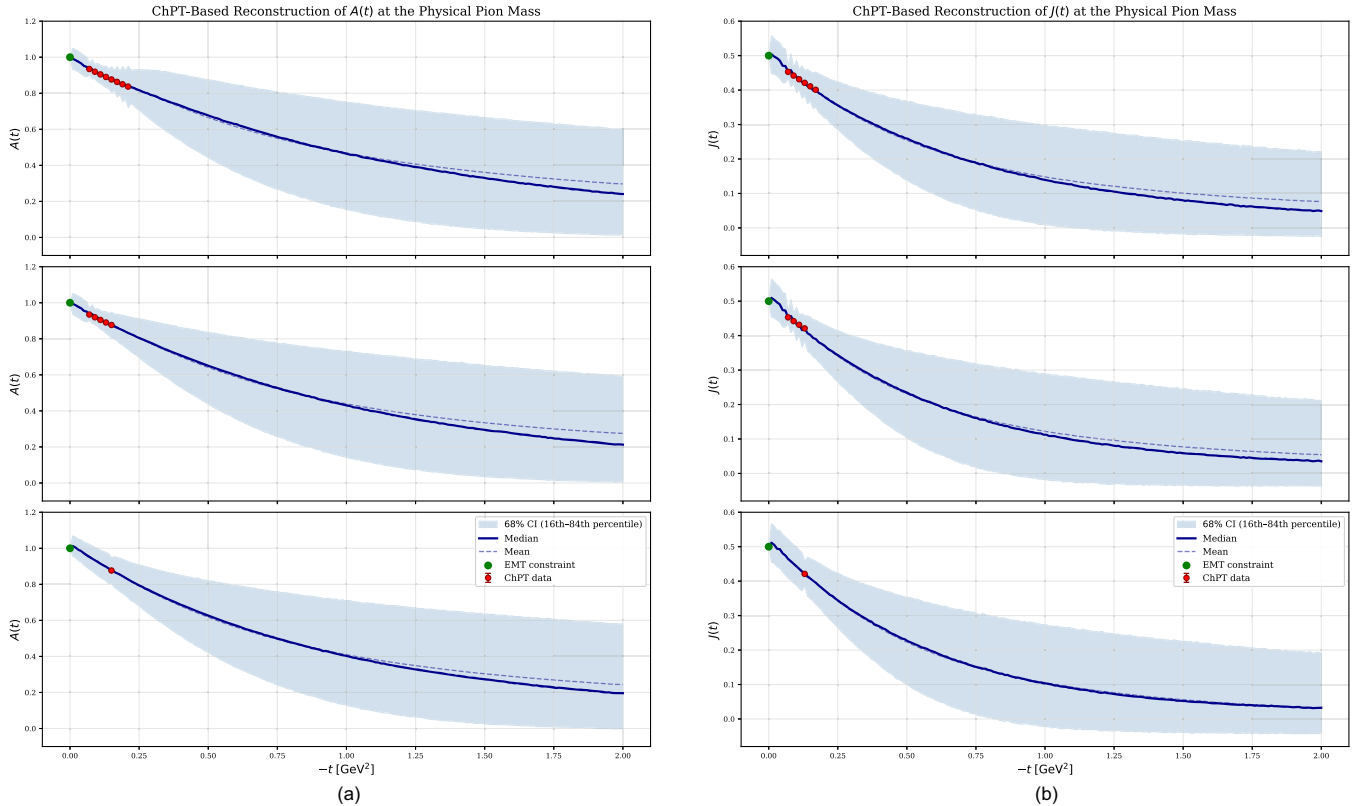


FIG. 4. Diffusion-model reconstruction of $A(t)$ and $J(t)$ at the physical pion mass, conditioned on ChPT data points. Panel layout follows the data-ablation protocol, with fewer conditioning points from top to bottom.

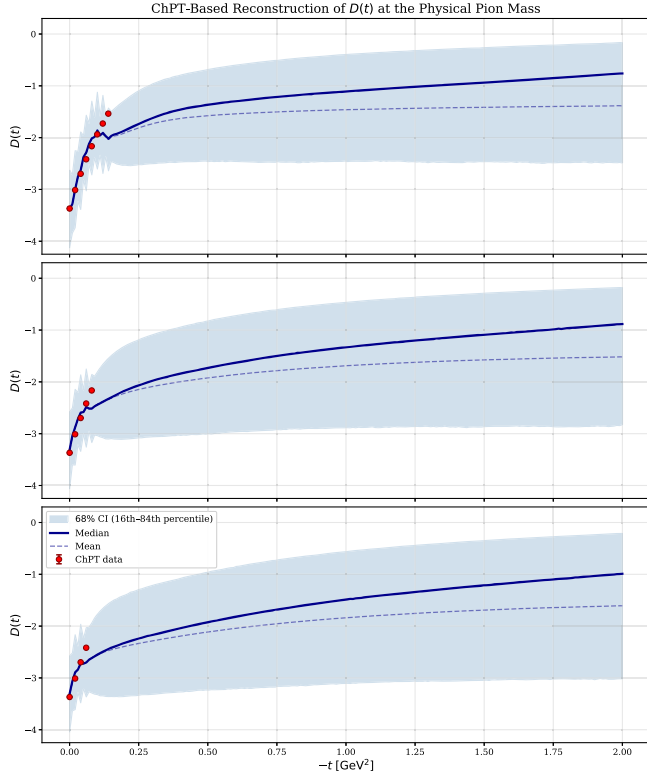


FIG. 5. Diffusion-model reconstruction of $D(t)$ at the physical pion mass, conditioned on ChPT data points. Panel layout follows the data-ablation protocol, with fewer conditioning points from top to bottom.

where the uncertainty is propagated linearly from δc_8 . This value is consistent with the dipole-fit result $D(0) = -3.87 \pm 0.97$ and the z -expansion result $D(0) = -3.35 \pm 0.58$ reported by Ref. [17] at the unphysical pion mass, with the value $D(0) = -3.38^{+0.34}_{-0.35}$ obtained in Ref. [31] using the dispersive techniques.⁵ It also comfortably satisfies the chiral perturbation theory upper bound $D(0)/m_N \leq -1.1(1) \text{ GeV}^{-1}$ [32].

It is worth noting that for the form factor $D(t)$ plotted in Fig. 5(a), we employed the unexpanded ChPT results of Refs. [13,15]. However, these results do not yet include all contributions to $D(t)$ arising from the third- and fourth-order chiral Lagrangians. In particular, the fourth-order Lagrangian in covariant form has yet to be constructed in its most minimal and general form. For the determination of the D -term, we therefore restricted ourselves to the low-energy region and retained only the lowest-order terms from the small-quantity expansion of the results in Refs. [13,15]. This also explains the difference between

⁵It is interesting to notice that the quark contribution alone to the D -term reported in Ref. [82], $D^Q(0) = -3.37 \pm 0.17$, obtained from a χ^2 analysis of Compton form factor data using skewness-dependent GPDs based on the double-distribution representation, is still compatible with our result.

Fig. 6, where we used the complete expanded ChPT results at lowest order, and Fig. 5(a), where we used the unexpanded but incomplete ChPT contributions.

B. Reconstruction of GFFs from DVCS data

In Ref. [11], the quark part of $D(t)$ form factor was extracted from deeply virtual Compton scattering data measured with the CEBAF Large Acceptance Spectrometer (CLAS) at Jefferson Lab. It should be noted, however, that this extraction relies on specific parametrizations of the Compton form factors and the assumption of GPD H dominance, rendering the resulting $D(t)$ values inherently model dependent [83,84]. Nevertheless, these data points constitute the only currently available experimental determination of the gravitational form factor $D(t)$ of the proton. In the following, we reconstruct the quark part of $D(t)$ form factor from the extracted values reported in Ref. [11] in two scenarios: first, using the data points alone, and second, incorporating the forward-limit constraint obtained in the previous section in order to illustrate the constraining power that a known forward limit provides, noting that this value refers to the total form factor and not to the quark contribution alone. The results of the two reconstruction scenarios are shown in Fig. 6. When only the DVCS data points are used the extrapolation toward the forward limit is largely unconstrained, as reflected by the rapidly widening 68% confidence interval and the significant deviation between mean and median at low $-t$. Incorporating the leading-order ChPT constraint dramatically reduces the uncertainty in the forward region.

VII. SUMMARY AND OUTLOOK

We have developed a generative framework based on denoising diffusion for the model-independent reconstruction of hadronic form factors from sparse and noisy data. The training prior, built from ten distinct functional classes rooted in different theoretical approaches to hadron structure, encodes a broad manifold of physically plausible shapes without privileging any single parametrization. Applied to the proton gravitational form factors $A(t)$, $J(t)$, and $D(t)$, the framework yields nonparametric reconstructions consistent with lattice QCD and with standard parametric fits across the full kinematic range $0 \leq -t \leq 2 \text{ GeV}^2$. A systematic data-ablation protocol demonstrates that the reconstruction remains robust even when only one or two conditioning points are retained, confirming the strong constraining power of the physically informed prior. The densely sampled output of the diffusion model enables a direct extraction of the chiral low-energy constants $c_8 = -4.6 \pm 0.8 \text{ GeV}^{-1}$ and $c_9 = -0.61 \pm 0.19 \text{ GeV}^{-1}$, in agreement with the independent dispersive determination of Ref. [31], and yields the nucleon D -term $D(0) = -4.3 \pm 0.8$ at the physical pion mass.

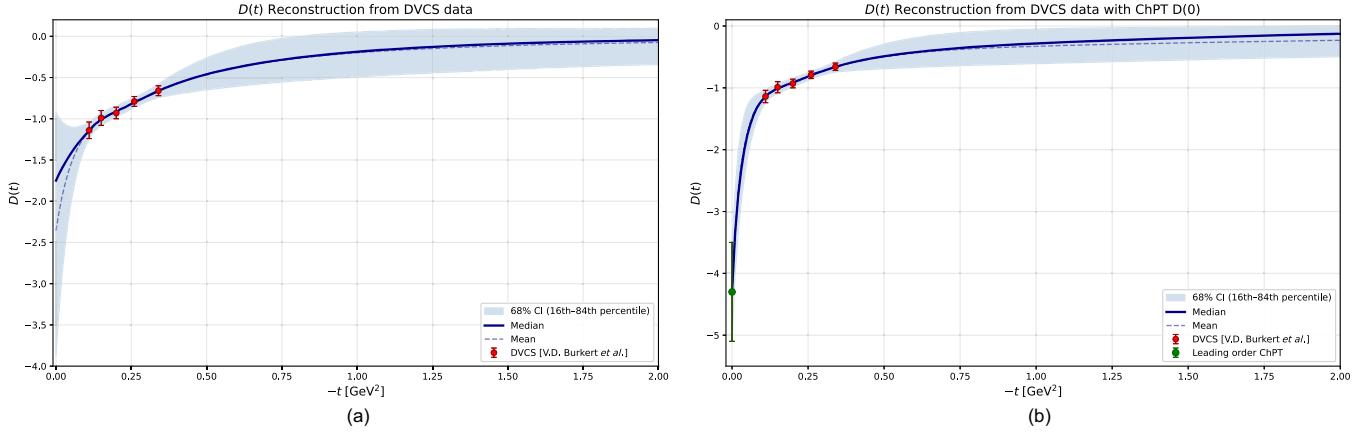


FIG. 6. Diffusion-model reconstruction of the quark part of $D(t)$ at the physical pion mass, conditioned on (a) the DVCS data of Ref. [11] alone, and (b) the DVCS data combined with the ChPT forward-limit constraint obtained in Sec. VI A.

The precision of the present extraction is limited primarily by the use of the ChPT expression for $D(t)$ at $\mathcal{O}(p^2)$. A natural next step is the construction of the complete generally covariant chiral Lagrangian up to and including $\mathcal{O}(p^4)$ with explicit $\Delta(1232)$ degrees of freedom [15]. Matching the diffusion-model reconstruction to these higher-order expressions is expected to sharpen the determination of c_8 and c_9 , reduce the uncertainty on the physical-mass extrapolation, and, in addition, provide access to the unknown LECs (h_i) of the Δ -resonance Lagrangian appearing in Ref. [14]. This would yield the first model-independent, ChPT-based estimates of the GFFs and D -term of the Δ baryon.

As already mentioned, this approach can be applied to reconstruct the GFFs of various hadrons, such as Δ -resonances, ρ -mesons, or even nuclear systems, including electromagnetic and transition form factors. Furthermore, our framework is naturally extensible to the reconstruction of GPDs, where the generative prior can be built from synthetic ensembles constructed via double distributions, ensuring that fundamental constraints such as polynomiality and positivity are satisfied by construction. Experimental Compton form factors, extracted from DVCS measurements, then serve as the conditioning data, while GFF sum rules provide additional moment constraints to further regularize the reconstruction. Work along these lines is currently underway.

ACKNOWLEDGMENTS

The authors thank Jambul Gegelia and Davide Laudicina for useful discussions. We are also grateful to Xiong-Hui Cao for clarifying the chiral order at which the low-energy constants in Ref. [31] were determined, and thank Xiangdong Ji and Latifa Elouadrhiri for helpful comments. The calculations of this work were performed on the HPC cluster Elysium of the Ruhr University Bochum, subsidized by the DFG (INST 213/1055-1).

DATA AVAILABILITY

The data that support the findings of this article are openly available [75].

APPENDIX A: NETWORK ARCHITECTURE

The denoising network f_θ is a one-dimensional ResNet-Attention hybrid that operates at constant spatial resolution $L = 200$ throughout, with no downsampling or upsampling layers. This design choice reflects the fact that all grid points carry comparable physical significance and avoids information loss from pooling. The network has the following properties:

- (i) It receives as input three channels $(x_\tau, m, c) \in \mathbb{R}^{3 \times L}$ that are mapped to a hidden representation $h \in \mathbb{R}^{C \times L}$ ($C = 256$) by a single convolutional layer with kernel size $K = 7$. A symmetric output head, Group-Norm, SiLU activation, and a 1×7 convolution, projects back to a single channel, yielding the predicted $\hat{v}_\tau \in \mathbb{R}^L$.
- (ii) The network comprises $N_{\text{res}} = 12$ residual blocks organized into $G = 3$ groups of four, each group followed by a multihead self-attention layer ($H = 4$ heads). Every residual block consists of two 1×7 convolutions interleaved with GroupNorm and SiLU activations, with a dropout rate of 0.1 before the second convolution. The diffusion timestep τ is injected via Feature-wise Linear Modulation (FiLM) [85]: a sinusoidal positional encoding of τ is passed through a two-layer multilayer perceptron (MLP) to produce per-channel scale and shift parameters applied after the first convolution. The self-attention layers provide a global receptive field, enabling the network to capture long-range correlations across the full grid, essential when enforcing physical constraints.
- (iii) The second convolution in each residual block and the output projection of each attention layer are zero-initialized, so that every block acts as an

identity map at the start of training. This stabilizes the initial gradient flow through the deep residual stack [86].

The resulting network has approximately 19×10^6 trainable parameters.

APPENDIX B: NETWORK TRAINING

As mentioned in Sec. III, the model was trained on 6×10^5 GFF curves, each sampled on a grid of $L = 200$ points and normalized to zero mean and unit variance per grid point. A separate validation set of 6×10^3 curves was used for model selection. Training minimizes the v -prediction loss of Eq. (15) with the AdamW optimizer [87] (learning rate $\eta = 10^{-4}$, weight decay 10^{-4}) and a batch size of 256. The learning-rate schedule consists of a linear warmup over the first two epochs followed by cosine decay to zero over the remaining 98 epochs, applied per optimizer step (2320 steps per epoch, 232 000 total). Gradients are clipped at unit norm to stabilize training in the presence of the attention layers. Moreover, an exponential moving average (EMA) of the network weights with decay $\lambda = 0.9999$ is maintained throughout training; all reported validation losses and all inference results use the EMA weights exclusively. Checkpoints are saved every 10 epochs, and the model with the lowest EMA validation loss is retained separately. Figure 7(a) shows the training and validation loss curves together with the learning-rate schedule. Both losses decrease rapidly during the first ~ 15 epochs, after which the training loss settles into a smooth descent from ≈ 0.022 to ≈ 0.019 . The validation

loss, evaluated with freshly drawn random masks at each epoch, exhibits larger fluctuations but remains within the 0.015–0.023 band from epoch 20 onward, indicating that the model neither overfits nor underfits. The best validation loss of 0.0153 is reached at epoch 65, whose EMA weights are used for all inference results reported in this work.

To verify robustness against the choice of conditioning strategy, we fine-tuned the converged model for 50 additional epochs at a flat learning rate of 2×10^{-5} ($5\times$ below the initial peak, to preserve the learned representations) and a broader five-type mask distribution. In addition to the three original mask types (see Sec. IV C), the expanded distribution introduces a high- $|t|$ cluster (5–15 points in the last 40% of the grid) and a sparse-spread mode (8–12 quasiuniformly spaced points across the full grid). As shown in Fig. 7(b), the validation loss, evaluated with the original mask distribution for comparability, remains within the 0.017–0.021 band throughout, consistent with the initial convergence plateau. This confirms that the learned GFF manifold is stable and that reconstruction quality is insensitive to the specific mask distribution. We have further verified that varying the network architecture and training strategy leaves the validation loss within the same fluctuation band, indicating that systematic errors from these design choices are subdominant to the statistical noise floor.

APPENDIX C: ABLATION STUDIES

To validate the reconstruction capability of the diffusion model of this work, we perform a systematic ablation study

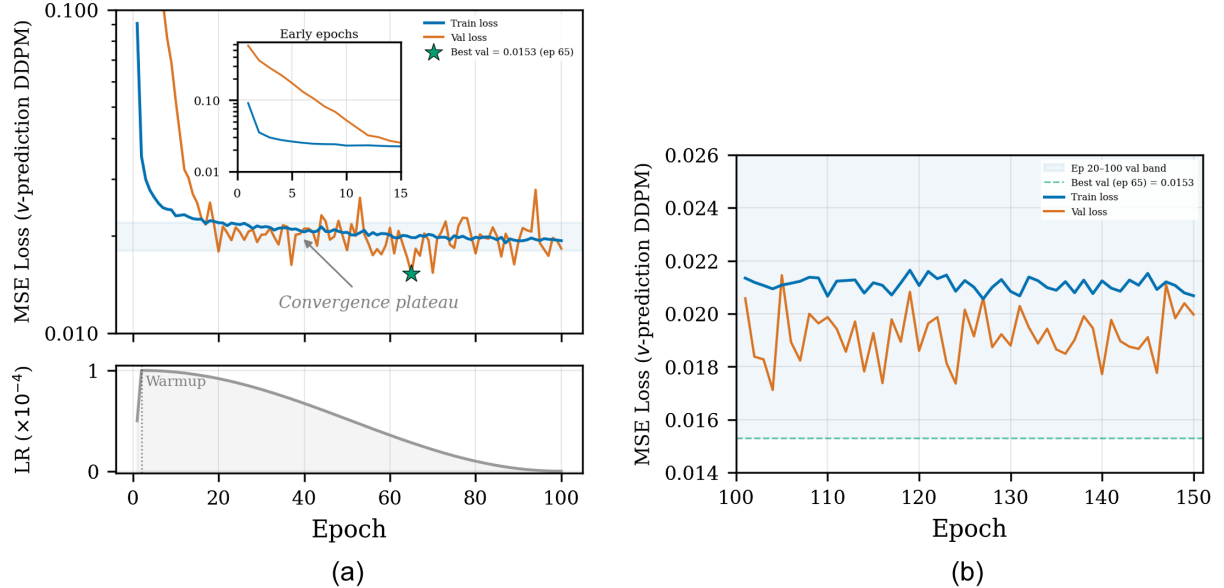


FIG. 7. Training diagnostics for the v -prediction DDPM. (a) Training and validation loss (upper panel, log scale) and learning-rate schedule (lower panel). The inset magnifies the first 15 epochs; the star marks the best EMA validation loss. (b) Fine-tuning with an expanded five-type mask distribution. The shaded band and dashed line show the validation range and best validation loss from the original training.

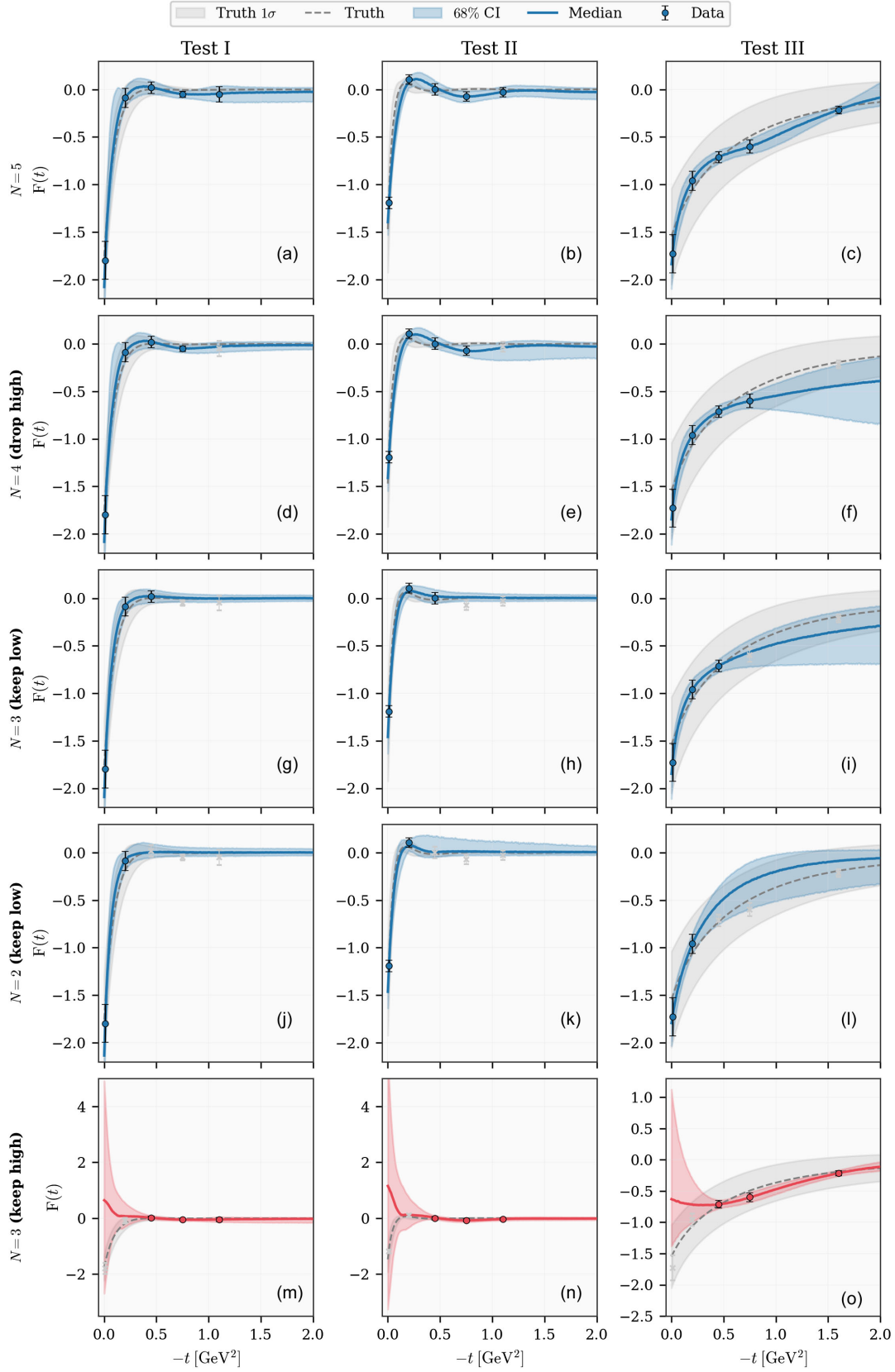


FIG. 8. Ablation study on three synthetic test curves (Test I, Test II, Test III), with N denoting the number of conditioning points: panels (a)–(c), $N = 5$; (d)–(f), $N = 4$ (drop high); (g)–(i), $N = 3$ (keep low); (j)–(l), $N = 2$ (keep low); (m)–(o), $N = 3$ (keep high). Rows 1–4 progressively remove high- $|t|$ conditioning points; row 5 keeps only high- $|t|$ points for contrast. The bottom row (red) uses the same number of points as the third row but retains only high- $|t|$ measurements, demonstrating the disproportionate importance of forward-region data.

on three synthetic test curves, shown in Fig. 8. The three scenarios are chosen to probe different form-factor behaviors: test curves I and II both exhibit a large negative value at low $|t|$ with a sharp rise toward zero, followed by a small local maximum. The two cases differ in the width and height of this local maximum. Test curve III, by contrast, displays a broad, monotonically decreasing shape with a correspondingly wider truth band.

For each scenario, a pool of candidate curves is drawn from the eight classes used in the training-data generation (Sec. III). An ensemble of similar-shape curves is selected via nearest neighbors in the mean squared distance, yielding a central curve $\bar{F}(t)$ (ensemble mean) and a pointwise uncertainty $\sigma_{\text{truth}}(t)$ (ensemble standard deviation). Pseudo-data are then generated at N chosen values of $-t$ by evaluating the central curve and adding Gaussian noise,

$$F_i^{\text{data}} = \bar{F}(t_i) + z_i \sigma_i, \quad z_i \sim \mathcal{N}(0, 1), \quad (\text{C1})$$

where σ_i represents the experimental uncertainty at each kinematic point. This procedure mimics realistic experimental conditions in which the true form factor is unknown and only noisy measurements at discrete momentum transfers are available. The diffusion model is then conditioned on the tuples $\{(t_i, F_i^{\text{data}}, \sigma_i)\}$ and tasked with reconstructing the full t -dependence, including extrapolation into unconstrained regions. Starting from the full set of $N = 5$ conditioning points, we progressively remove measurements from the high- $|t|$ end (rows 1–4 of Fig. 8) and compare the resulting 68% credible intervals against the known truth band. The bottom row provides the key comparison: three points retained at high $|t|$ instead of low $|t|$, with the same total number of constraints as row 3. Across all tests, the reconstruction degrades significantly when low- $|t|$ data, which anchor the form-factor normalization, are removed, while high- $|t|$ points can be dropped with comparatively little loss of accuracy.

-
- [1] R. L. Jaffe, *Phys. Rev. D* **103**, 016017 (2021).
 - [2] J. Y. Panteleeva, E. Epelbaum, J. Gegelia, and U.-G. Meißner, *Eur. Phys. J. C* **83**, 617 (2023).
 - [3] A. Freese and G. A. Miller, *Phys. Rev. D* **105**, 014003 (2022).
 - [4] X. Ji and C. Yang, *Nucl. Phys.* **B1024**, 117342 (2026).
 - [5] D. Müller, D. Robaschik, B. Geyer, F. M. Dittes, and J. Hořejši, *Fortschr. Phys.* **42**, 101 (1994).
 - [6] X. D. Ji, *Phys. Rev. Lett.* **78**, 610 (1997).
 - [7] A. V. Radyushkin, *Phys. Lett. B* **380**, 417 (1996).
 - [8] K. Goetze, M. V. Polyakov, and M. Vanderhaeghen, *Prog. Part. Nucl. Phys.* **47**, 401 (2001).
 - [9] A. V. Radyushkin, *Phys. Rev. D* **56**, 5524 (1997).
 - [10] V. D. Burkert, L. Elouadrhiri, and F. X. Girod, *Nature (London)* **557**, 396 (2018).
 - [11] V. D. Burkert, L. Elouadrhiri, and F. X. Girod, *arXiv*: 2104.02031.
 - [12] J. C. Collins, L. Frankfurt, and M. Strikman, *Phys. Rev. D* **56**, 2982 (1997).
 - [13] H. Alharazin, D. Djukanovic, J. Gegelia, and M. V. Polyakov, *Phys. Rev. D* **102**, 076023 (2020).
 - [14] H. Alharazin, E. Epelbaum, J. Gegelia, U.-G. Meißner, and B.-D. Sun, *Eur. Phys. J. C* **82**, 907 (2022).
 - [15] H. Alharazin, *Phys. Rev. D* **109**, 016009 (2024).
 - [16] P. E. Shanahan and W. Detmold, *Phys. Rev. D* **99**, 014511 (2019).
 - [17] D. C. Hackett, D. A. Pefkou, and P. E. Shanahan, *Phys. Rev. Lett.* **132**, 251904 (2024).
 - [18] I. Y. Kobzarev and L. B. Okun, *Zh. Eksp. Teor. Fiz.* **43**, 1904 (1962) [*Sov. Phys. JETP* **16**, 1343 (1963)].
 - [19] H. Pagels, *Phys. Rev.* **144**, 1250 (1966).
 - [20] M. V. Polyakov, *Phys. Lett. B* **555**, 57 (2003).
 - [21] M. V. Polyakov and P. Schweitzer, *Int. J. Mod. Phys. A* **33**, 1830025 (2018).
 - [22] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, in *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 (PMLR, 2015), p. 2256.
 - [23] J. Ho, A. Jain, and P. Abbeel, in *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Associates, Red Hook, NY, 2020), p. 6840.
 - [24] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *arXiv*:2011.13456.
 - [25] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Piscataway, NJ, 2022), p. 10684.
 - [26] A. Q. Nichol and P. Dhariwal, in *Proc. 38th Int. Conf. on Machine Learning*, Vol. 139 (PMLR, 2021), p. 8162.
 - [27] L. Wang, G. Aarts, and K. Zhou, *J. High Energy Phys.* **05** (2024) 060.
 - [28] Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang, *J. High Energy Phys.* **05** (2025) 043.
 - [29] G. Aarts, D. E. Habibi, A. Ipp, D. I. Müller, T. R. Ranner, L. Wang, W. Wang, and Q. Zhu, *arXiv*:2601.19552.
 - [30] H. Alharazin, J. Yu. Panteleeva, and B.-D. Sun, *arXiv*: 2602.09045.
 - [31] X. H. Cao, F. K. Guo, Q. Z. Li, B. W. Wu, and D. L. Yao, *arXiv*:2507.05375.
 - [32] J. Gegelia and M. V. Polyakov, *Phys. Lett. B* **820**, 136572 (2021).
 - [33] S. Cotogno, C. Lorcé, P. Lowdon, and M. Morales, *Phys. Rev. D* **101**, 056016 (2020).
 - [34] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Phys. Rev. Lett.* **119**, 161101 (2017).
 - [35] P. P. Avelino, *Phys. Lett. B* **795**, 627 (2019).
 - [36] S. Dubynskiy and M. B. Voloshin, *Phys. Lett. B* **666**, 344 (2008).

- [37] M. I. Eides, V. Y. Petrov, and M. V. Polyakov, *Phys. Rev. D* **93**, 054039 (2016).
- [38] I. A. Perevalova, M. V. Polyakov, and P. Schweitzer, *Phys. Rev. D* **94**, 054024 (2016).
- [39] F. J. Ernst, R. G. Sachs, and K. C. Wali, *Phys. Rev.* **119**, 1105 (1960).
- [40] R. G. Sachs, *Phys. Rev.* **126**, 2256 (1962).
- [41] G. A. Miller, *Phys. Rev. C* **99**, 035202 (2019).
- [42] G. A. Miller, *Phys. Rev. Lett.* **99**, 112001 (2007).
- [43] C. Lorcé, *Phys. Rev. Lett.* **125**, 232002 (2020).
- [44] E. Epelbaum, J. Gegelia, N. Lange, U. G. Meißner, and M. V. Polyakov, *Phys. Rev. Lett.* **129**, 012001 (2022).
- [45] A. Freese and G. A. Miller, *Phys. Rev. D* **108**, 034008 (2023).
- [46] J. Y. Panteleeva, E. Epelbaum, J. Gegelia, and U. G. Meißner, *Phys. Rev. D* **106**, 056019 (2022).
- [47] C. Lorcé, H. Moutarde, and A. P. Trawiński, *Eur. Phys. J. C* **79**, 89 (2019).
- [48] H. Hashamipour, M. Goharipour, K. Azizi, and S. V. Goloskokov, *Phys. Rev. D* **107**, 096005 (2023).
- [49] D. A. Pefkou, D. C. Hackett, and P. E. Shanahan, *Phys. Rev. D* **105**, 054509 (2022).
- [50] C. Alexandrou *et al.*, *Phys. Rev. D* **101**, 094513 (2020).
- [51] G. S. Bali *et al.* (RQCD Collaboration), *Phys. Rev. D* **100**, 014507 (2019).
- [52] R. J. Hill and G. Paz, *Phys. Rev. D* **82**, 113005 (2010).
- [53] B. Bhattacharya, R. J. Hill, and G. Paz, *Phys. Rev. D* **84**, 073006 (2011).
- [54] G. Lee, J. R. Arrington, and R. J. Hill, *Phys. Rev. D* **92**, 013013 (2015).
- [55] W. Broniowski and E. Ruiz Arriola, *Phys. Rev. D* **111**, 074017 (2025).
- [56] P. Masjuan and S. Peris, *Phys. Lett. B* **686**, 307 (2010).
- [57] R. Bijker and F. Iachello, *Phys. Rev. C* **69**, 068201 (2004).
- [58] H. Tong, J.-W. Ma, and F. Yuan, *J. High Energy Phys.* **10** (2022) 046.
- [59] S. J. Brodsky and G. F. de Téramond, *Phys. Rev. D* **78**, 025032 (2008).
- [60] Z. Abidin and C. E. Carlson, *Phys. Rev. D* **79**, 115003 (2009).
- [61] J. J. Kelly, *Phys. Rev. C* **70**, 068202 (2004).
- [62] Z. Ye, J. Arrington, R. J. Hill, and G. Lee, *Phys. Lett. B* **777**, 8 (2018).
- [63] L. Schlessinger, *Phys. Rev.* **167**, 1411 (1968).
- [64] Z.-F. Cui, D. Binosi, C. D. Roberts, and S. M. Schmidt, *Phys. Rev. Lett.* **127**, 092001 (2021).
- [65] X.-H. Cao, F.-K. Guo, Q.-Z. Li, and D.-L. Yao, *Nat. Commun.* **16**, 6979 (2025).
- [66] H.-W. Hammer and U.-G. Meißner, *Eur. Phys. J. A* **20**, 469 (2004).
- [67] G. Höhler, E. Pietarinen, I. Sabba-Stefanescu, F. Borkowski, G. G. Simon, V. H. Walther, and R. D. Wendling, *Nucl. Phys. B* **114**, 505 (1976).
- [68] M. A. Belushkin, H.-W. Hammer, and U.-G. Meißner, *Phys. Rev. C* **75**, 035202 (2007).
- [69] S. J. Brodsky and G. R. Farrar, *Phys. Rev. Lett.* **31**, 1153 (1973).
- [70] G. P. Lepage and S. J. Brodsky, *Phys. Rev. D* **22**, 2157 (1980).
- [71] A. Chodos, R. L. Jaffe, K. Johnson, C. B. Thorn, and V. F. Weisskopf, *Phys. Rev. D* **9**, 3471 (1974).
- [72] X. Ji, *J. Phys. G* **24**, 1181 (1998).
- [73] M. J. Neubelt, A. Sampino, and P. Schweitzer, *Phys. Rev. D* **101**, 034013 (2020).
- [74] T. Tezgin, P. Schweitzer, and N. Kersting, *Phys. Rev. D* **109**, 014010 (2024).
- [75] H. Alharazin and J. Yu. Panteleeva, <https://github.com/Herzallah15/diffusion-gff-reconstruction> (2026); H. Alharazin and J. Yu. Panteleeva, https://github.com/JuliaYu24/GFF_Diffusion_Analysis (2026).
- [76] T. Salimans and J. Ho, [arXiv:2202.00512](https://arxiv.org/abs/2202.00512).
- [77] K. Goeke, J. Grabis, J. Ossmann, P. Schweitzer, A. Silva, and D. Urbano, *Phys. Rev. C* **75**, 055207 (2007).
- [78] C. Cebulla, K. Goeke, J. Ossmann, and P. Schweitzer, *Nucl. Phys. A* **794**, 87 (2007).
- [79] J. Deng and D. Hou, [arXiv:2512.17554](https://arxiv.org/abs/2512.17554).
- [80] Z. Q. Yao, Y. Z. Xu, D. Binosi, Z. F. Cui, M. Ding, K. Raya, C. D. Roberts, J. Rodríguez-Quintero, and S. M. Schmidt, *Eur. Phys. J. A* **61**, 92 (2025).
- [81] Y. Guo, F. Yuan, and W. Zhao, *Phys. Rev. Lett.* **135**, 111902 (2025).
- [82] M. Goharipour, H. Hashamipour, H. Fatehi, F. Irani, K. Azizi, and S. V. Goloskokov (MMGPDs Collaboration), *Phys. Rev. D* **112**, 014016 (2025).
- [83] K. Kumerički, *Nature (London)* **570**, E1 (2019).
- [84] H. Dutrieux, C. Lorcé, H. Moutarde, P. Sznajder, A. Trawiński, and J. Wagner, *Eur. Phys. J. C* **81**, 300 (2021).
- [85] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, FiLM: Visual Reasoning with a General Conditioning Layer in *Proceedings of the AAAI Conference on Artificial Intelligence* Vol. 32 (2018).
- [86] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He, [arXiv:1706.02677](https://arxiv.org/abs/1706.02677).
- [87] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, in *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)* (2019).