Efficient Simulated Data in Electron Scattering Experiments: Event Generation, Reweighting, Likelihood Fits, and Machine-Learned Fast Simulation

Jefferson Lab / CLAS12 Analysis Note

November 18, 2026

Abstract

Modern electron scattering experiments, such as those conducted with the CLAS12 spectrometer at Jefferson Lab, rely on high-statistics Monte Carlo (MC) simulations to unfold detector acceptance, model resolution distortions, and extract physical observables via multidimensional fits. Because full detector simulation in GEANT4 and event reconstruction are computationally expensive, maximizing the statistical power generated per CPU hour is essential. This pedagogical note reviews the mathematical and practical principles of efficient MC sampling across the full analysis pipeline: adaptive rejection sampling in event generators (`elSpectro`), fiducial pre-filtering prior to `GEMC`, post-simulation multidimensional reweighting via density ratio estimation, normalization in unbinned extended maximum likelihood fits, kinematic handling of steep Q^2 and t dependencies, and factorized machine-learned fast simulation. Rigorous mathematical derivations and proofs are provided in the appendices.

Contents

1 Statistical Motivation for Matching Simulation to Data **2**

2 Event Generation and Accept/Reject Efficiency **3**

3 Pre-Simulation Fiducial Filtering and Acceptance Accounting **3**

4 Post-Simulation Event Reweighting and Density Ratios **4**

5 Application to Unbinned Extended Maximum Likelihood Fits **5**

6 Kinematic Dependencies in Electron Scattering: Q^2 and t **6**

7 Machine-Learned Fast Simulation **7**
- 7.1 Motivation and Factorized Architecture 7
- 7.2 Fast Acceptance Modeling via Calibrated Density Ratios 7
- 7.3 Fast Resolution Modeling via Over-Fitted Decision Tree Ensembles 8
- 7.4 Applications and the Combinatorial Statistical Advantage 9

A Please note appendices have not been checked in detail **11**

B Derivation of Importance Sampling Variance and Kish's Effective Sample Size **11**

C Proof of Accept/Reject Exactness and Grid Envelope Optimization **12**

D Fiducial Pre-Filtering and Cross-Section Normalization **12**

E Variational Proof of the Likelihood Ratio Trick and Cascaded Reweighting **13**
E.1 Validity of Empirical Reconstructed-Space Reweighting without Unfolding . . 14

F Extended Maximum Likelihood Normalization with Monte Carlo Events **14**

G Electroproduction Flux Factorization and t-Slope Weight Variance **15**
G.1 One-Photon Exchange Factorization 15
G.2 Analytical Weight Variance of Exponential t-Slope Reweighting 15

H Non-Parametric Conditional Resolution Sampling and Metric Definitions **16**
H.1 Conditional Sampling via Randomized Decision Tree Partitions 16
H.2 Quantitative Smoothness (S) and Least Squares Error (LSE) Metrics 17
H.3 Combinatorial Scaling in Factorized Multi-Particle Final States 17

---

1 Statistical Motivation for Matching Simulation to Data

In experimental nuclear and particle physics, the primary purpose of generating Monte Carlo (MC) simulated data is to evaluate detector-corrected physical observables and fit templates with minimal statistical uncertainty from a finite computational budget.

Suppose we wish to evaluate the expectation value of an observable (or a parametric fit projection) $\mathcal{O}(x)$ over the true physical phase-space probability density $p_{\text{data}}(x)$, where x denotes the vector of kinematic variables:

$$\langle \mathcal{O} \rangle = \int_{\Omega} \mathcal{O}(x)\, p_{\text{data}}(x)\, dx. \tag{1}$$

In practice, we do not know $p_{\text{data}}(x)$ analytically prior to the measurement, or it is too complex to sample from directly. Instead, we generate N_{MC} simulated events $\{x_1, x_2, \ldots, x_{N_{\text{MC}}}\}$ drawn from a known baseline generator distribution $p_{\text{gen}}(x)$ (with $p_{\text{gen}}(x) > 0$ wherever $p_{\text{data}}(x) > 0$). By applying *Importance Sampling* [1], we rewrite the integral as an expectation over $p_{\text{gen}}(x)$ and approximate it as a weighted sum over the simulated sample:

$$\langle \mathcal{O} \rangle = \int_{\Omega} \mathcal{O}(x) \frac{p_{\text{data}}(x)}{p_{\text{gen}}(x)} p_{\text{gen}}(x)\, dx \approx \frac{\sum_{i=1}^{N_{\text{MC}}} w_i\, \mathcal{O}(x_i)}{\sum_{i=1}^{N_{\text{MC}}} w_i}, \quad \text{where} \quad w_i \propto \frac{p_{\text{data}}(x_i)}{p_{\text{gen}}(x_i)}. \tag{2}$$

Using the self-normalized estimator in Eq. (2) guarantees that the observable is properly normalized even when $p_{\text{data}}(x)$ and $p_{\text{gen}}(x)$ are known only up to overall cross-section or luminosity scale factors.

While Eq. (2) is asymptotically unbiased for any valid baseline distribution $p_{\text{gen}}(x)$, its statistical precision depends sensitively on the spread of the event weights $\{w_i\}$. A sample dominated by a handful of events with huge weights has far less statistical resolving power than an unweighted sample of the same size. As derived in **Appendix B**, the statistical power of a weighted MC sample is quantified by *Kish's effective sample size*, N_{eff} [2]:

$$N_{\text{eff}} = \frac{\left(\sum_{i=1}^{N_{\text{MC}}} w_i\right)^2}{\sum_{i=1}^{N_{\text{MC}}} w_i^2} = \frac{N_{\text{MC}}}{1 + \text{Var}(w)/\bar{w}^2}, \tag{3}$$

where $\bar{w}$ and $\text{Var}(w)$ are the sample mean and variance of the weights. Essentially, a weighted sample of N_{MC} events carries the same statistical precision as an unweighted sample of size $N_{\text{eff}} \leq N_{\text{MC}}$.

If the generated distribution $p_{\text{gen}}(x)$ deviates substantially from the true physics distribution $p_{\text{data}}(x)$, the relative weight variance $\text{Var}(w)/\bar{w}^2$ becomes large, causing N_{eff} to collapse ($N_{\text{eff}} \ll N_{\text{MC}}$). Physically, this means CPU cycles were wasted tracking events through regions of phase space that rarely occur in nature, while high-probability physical regions were undersampled and subsequently assigned massive weights.

The theoretical maximum of statistical efficiency is achieved when the simulation matches the underlying physics distribution: $p_{\text{gen}}(x) = p_{\text{data}}(x)$. In this ideal limit, all weights are identical ($w_i = \text{const}$), $\text{Var}(w) = 0$, and $N_{\text{eff}} = N_{\text{MC}}$. Designing the simulation pipeline so that $p_{\text{gen}}(x)$ approximates $p_{\text{data}}(x)$ as closely as possible is therefore a strict mathematical requirement for minimizing MC statistical uncertainties.

2 Event Generation and Accept/Reject Efficiency

To generate unweighted events ($w_i = 1$) distributed according to a complex theoretical target distribution $f(x)$ (such as a coupled-channel partial-wave amplitude or parametrized electroproduction cross section), event generators employ von Neumann's *Accept/Reject* (rejection sampling) algorithm [3, 1].

Because sampling directly from a multidimensional, non-invertible distribution $f(x)$ is generally impossible, we introduce a proposal distribution (or *envelope*) $g(x)$ that is computationally trivial to sample from. We then scale $g(x)$ by a bounding constant $M \geq 1$ such that the envelope strictly majorizes the target density across all allowed kinematics $x \in \Omega$:

$$f(x) \leq M \cdot g(x) \quad \forall x \in \Omega \implies M = \sup_{x \in \Omega} \left(\frac{f(x)}{g(x)} \right). \tag{4}$$

The algorithm proceeds in two steps: (1) draw a candidate event x from $g(x)$ and a uniform random number $u \sim U(0,1)$; (2) accept the candidate if $u \leq \frac{f(x)}{Mg(x)}$, and reject it otherwise.

As proven in **Appendix C**, the accepted events are distributed *exact*ly according to $f(x)$, and the global generation efficiency (the fraction of candidate draws accepted) is given by:

$$\epsilon_{\text{gen}} = \frac{1}{M} \int_{\Omega} f(x)\, dx = \frac{1}{M}, \tag{5}$$

assuming $f(x)$ and $g(x)$ are normalized to unity. If $g(x)$ is chosen naively as uniform phase space while $f(x)$ contains sharp resonance peaks or steep kinematic poles, the local ratio $f(x)/g(x)$ at the peak forces $M \gg 1$, driving the generation efficiency $\epsilon_{\text{gen}} \to 0$.

Modern electron scattering generators such as `elSpectro` [4] overcome this bottleneck by constructing piecewise, factorized grid envelopes. Rather than using a flat proposal function, `elSpectro` pre-samples the integrand to build factorized 1D and 2D histogram envelopes (for example, factorizing the virtual photon flux in (W, Q^2) from conditional momentum-transfer envelopes $g(t \mid W)$ and resonance mass shapes). Because the step-function histogram envelope $g(x)$ tightly wraps the local shape of $f(x)$ across phase space, the supremum ratio M remains close to 1, maximizing generation speed while yielding strictly unweighted events.

3 Pre-Simulation Fiducial Filtering and Acceptance Accounting

In large-acceptance fixed-target spectrometers such as CLAS12 at Jefferson Lab [5], event generation in `elSpectro` takes microseconds per event, whereas propagating the final-state particles through the toroidal and solenoidal magnetic fields and detector materi-

als in GEMC (the CLAS12 GEANT4 simulation package [6, 7]) and reconstructing tracks in CLARA [8] takes hundreds of milliseconds to seconds per event.

Consequently, it is standard practice to apply fast generator-level fiducial filters prior to passing events to GEMC, discarding events in which key particles fall entirely outside the active geometric acceptance of the spectrometer. For instance, in CLAS12 electron scattering analyses using `clas12-elSpectro` [4]:

- **Forward Tagger (FT) configurations** (quasi-real photoproduction at very low Q^2) require the scattered electron to lie within the FT polar angle window, typically $2^\circ \leq \theta_e \leq 5^\circ$ and 0.4 GeV $\leq p_e \leq 6.0$ GeV [9].
- **Forward Detector (FD) configurations** (electroproduction at moderate-to-high Q^2) require the scattered electron to clear the torus beampipe and enter the drift chambers, typically $\theta_e \geq 5^\circ$.

Events falling outside these kinematic windows have zero probability of triggering the relevant detector subsystem; filtering them out at the generator stage prevents wasting GEMC CPU time on unobservable events.

However, this pre-simulation filtering requires strict bookkeeping in downstream cross-section extractions. As derived in **Appendix D**, if the generator samples over the full 4π phase space $\Omega_{4\pi}$ (generating $N_{\text{trial}}^{4\pi}$ total events) and retains only the subset N_{filt} falling inside the fiducial volume Ω_{fid} for GEMC processing, one has two mathematically equivalent ways to normalize the cross section:

1. **Unconstrained 4π Normalization:** Define the global acceptance over the full phase space as $\epsilon_{4\pi} = N_{\text{reco}}/N_{\text{trial}}^{4\pi}$, using the pre-filter trial count $N_{\text{trial}}^{4\pi}$ in the denominator.
2. **Fiducial Volume Normalization:** Define the acceptance relative to the filtered sample as $\epsilon_{\text{fid}} = N_{\text{reco}}/N_{\text{filt}}$, and explicitly multiply the integrated generator cross section (or phase-space volume) by the fiducial branching fraction $f_{\text{fid}} = \sigma(\Omega_{\text{fid}})/\sigma(\Omega_{4\pi}) = N_{\text{filt}}/N_{\text{trial}}^{4\pi}$.

Using N_{filt} in the acceptance denominator while normalizing to the unconstrained 4π generator cross section omits the factor f_{fid} and artificially underestimates the measured physical cross section by f_{fid}.

4 Post-Simulation Event Reweighting and Density Ratios

When theoretical models are updated or baseline generator parameters require empirical tuning to match experimental data, re-running GEMC and CLARA for every iteration is computationally prohibitive. Instead, the existing baseline simulated sample $p_{\text{base}}(x)$ is reweighted to match the target distribution $p_{\text{target}}(x)$[cite: 2, 6]. As established in Section 1, maintaining statistical efficiency ($N_{\text{eff}} \approx N_{\text{MC}}$) requires the weights $w(x) = p_{\text{target}}(x)/p_{\text{base}}(x)$ to remain clustered as close to 1 as possible.

In experimental analyses where the simulated sample is tuned directly to real data, the true underlying kinematics are unobservable in the data; only the detector-reconstructed kinematics x_{reco} are available. Avoiding a full multidimensional unfolding procedure (which is beyond the scope of this work), the empirical density ratio must therefore be evaluated in *reconstructed space* by comparing the reconstructed real data distribution $p_{\text{data}}(x_{\text{reco}})$ to the reconstructed baseline MC distribution $p_{\text{base}}(x_{\text{reco}})$[cite: 2, 6]:

$$w(x_{\text{reco}}) = \frac{p_{\text{data}}(x_{\text{reco}})}{p_{\text{base}}(x_{\text{reco}})}. \tag{6}$$

When this empirically derived weight function $w(\cdot)$ is subsequently used to update the acceptance correction $\epsilon = \sum_{i\in\text{acc}} w_i / \sum_{j\in\text{gen}} w_j$, it must also be evaluated for the generated MC events in the denominator (where x_{reco} does not exist for rejected events). Applying the reconstructed weight parameterization $w(\cdot)$ to the generated kinematics x_{gen} relies on the implicit approximation that detector resolution smearing is small compared to the kinematic scale over which the cross section and acceptance vary, so that $p_{\text{data}}(x_{\text{reco}})/p_{\text{base}}(x_{\text{reco}}) \approx p_{\text{data}}(x_{\text{gen}})/p_{\text{base}}(x_{\text{gen}})$ (see **Appendix E**).

Depending on the dimensionality of the reconstructed phase space, the density ratio in Eq. (6) can be evaluated using two computational methods that operate on equal statistical footing:

- **Histograms:** In 1 or 2 dimensions where bins are well populated, the weight is evaluated directly by dividing the normalized reconstructed data histogram by the reconstructed base MC histogram[cite: 2, 6]. Each event is assigned a weight equal to the ratio of the respective bin contents.
- **Probabilistic Classifiers:** In higher-dimensional phase spaces (e.g., fully correlated 4D or 5D final states) where multidimensional histograms suffer from empty bins, Machine Learning classifiers (such as DNNs or GBDTs) can be trained to distinguish reconstructed target (data) events from reconstructed base MC events [10, 26, 12, 13][cite: 2, 6, 28]. A well-calibrated classifier outputting a posterior probability $s(x_{\text{reco}})$ can be algebraically rearranged to evaluate the exact same density ratio [10][cite: 2, 6, 28]:

$$w(x_{\text{reco}}) = \left(\frac{N_{\text{base}}}{N_{\text{data}}}\right) \frac{s(x_{\text{reco}})}{1 - s(x_{\text{reco}})}, \tag{7}$$

  reducing to $w(x_{\text{reco}}) = \frac{s(x_{\text{reco}})}{1-s(x_{\text{reco}})}$ when the training samples are normalized to equal prior weights[cite: 6].

Either method is valid: histograms are straightforward and robust for low-dimensional projections[cite: 2], whereas probabilistic classifiers should be deployed when preserving unbinned, high-dimensional kinematic correlations is necessary[cite: 2, 6].

5 Application to Unbinned Extended Maximum Likelihood Fits

In hadron spectroscopy and amplitude analyses at CLAS12 (often performed using toolkits such as `brufit`), physical parameters $\boldsymbol{\theta}$ (e.g., resonance masses, widths, electrocouplings, or moments of angular distributions) are extracted using unbinned *Extended Maximum Likelihood* (EML) fits.

In an EML fit, the probability density of observing N_{data} events at kinematics $\{x_1, \ldots, x_{N_{\text{data}}}\}$ depends both on the point-by-point theoretical intensity $\mathcal{I}(x \mid \boldsymbol{\theta})$ and on the total expected yield $N_{\text{exp}}(\boldsymbol{\theta})$ across the detector acceptance $\eta(x)$. As derived in **Appendix F**, the negative log-likelihood (up to parameter-independent constants) takes the form:

$$-\ln \mathcal{L}(\boldsymbol{\theta}) = N_{\text{exp}}(\boldsymbol{\theta}) - \sum_{i=1}^{N_{\text{data}}} \ln \mathcal{I}(x_i \mid \boldsymbol{\theta}). \tag{8}$$

Evaluating the normalization integral $N_{\text{exp}}(\boldsymbol{\theta}) = \mathcal{L}_{\text{int}} \int \mathcal{I}(x \mid \boldsymbol{\theta})\eta(x)\, dx$ analytically is impossible for a realistic detector like CLAS12. Instead, we evaluate it via Monte Carlo integration using the accepted events (N_{acc}) from our simulated sample. If the baseline simulation was generated with intensity $\mathcal{I}_{\text{gen}}(x)$ and optionally carries post-simulation tuning

weights w_j^{tune}, the expected yield is computed dynamically during likelihood minimization as:

$$N_{\text{exp}}(\boldsymbol{\theta}) = \mathcal{C} \sum_{j=1}^{N_{\text{acc}}} w_j \, \frac{\mathcal{I}(x_j \mid \boldsymbol{\theta})}{\mathcal{I}_{\text{gen}}(x_j)}, \tag{9}$$

where $\mathcal{C}$ is an overall luminosity normalization constant.

Equation (9) highlights two key practical lessons:

1. **Dynamic Model Updates:** By reweighting the accepted MC events inside the normalization sum, the fit dynamically accounts for how the detector acceptance $\eta(x)$ sculpts any trial physics hypothesis $\boldsymbol{\theta}$, without generating new GEANT4 samples.
2. **Finite-MC Variance in Fits:** Because $N_{\text{exp}}(\boldsymbol{\theta})$ is itself a stochastic MC estimate, large weight variances (Var(w)) introduce statistical noise into the normalization integral. To prevent the MC statistical uncertainty from dominating the parameter errors on $\boldsymbol{\theta}$, the effective sample size N_{eff} of the normalization sample should be much larger than N_{data} (typically $N_{\text{eff}} \gtrsim 10\, N_{\text{data}}$), and finite-MC error propagation (analogous to the Barlow-Beeston formalism [14]) should be included in the covariance matrix.

6 Kinematic Dependencies in Electron Scattering: Q^2 and t

Exclusive and semi-inclusive electron scattering reactions, $ep \to e'X$, exhibit extreme kinematic variations across phase space that make naive flat generation impossible:

- **Virtual Photon Propagator (Q^2 and W):** In the one-photon exchange (OPE) approximation, the electroproduction cross section factorizes into a virtual photon flux factor $\Gamma_v(E_b, Q^2, W)$ and a virtual-photon–nucleon cross section. Because the photon propagator scales as $1/Q^4$ (and approximately as $1/E_{\gamma^*}$ in photon energy), the cross section spans several orders of magnitude between the Forward Tagger ($Q^2 \sim 10^{-2}$ GeV^2) and high-Q^2 Forward Detector regimes ($Q^2 \sim$ 1–10 GeV^2). Generators must therefore sample the scattered electron directly from a nominal virtual photon flux enriched with an empirical form-factor damping $F(Q^2)$.
- **Four-Momentum Transfer to the Target (t):** Peripheral and diffractive meson electroproduction reactions are governed by t-channel Regge or meson exchange, resulting in a cross section that peaks sharply in the forward direction and falls roughly exponentially with the squared four-momentum transfer: $d\sigma/dt \propto \exp(-b|t|)$, where $b \sim 3$–8 GeV^{-2} is the characteristic t-slope.

Because both the physical cross section and the CLAS12 detector acceptance vary strongly with t, choosing how to handle the t-dependence during observable extraction is critical. Two primary strategies are used:

Method 1: Global Integration over t via Reweighting. The MC sample is generated with a nominal exponential slope b_{gen}, iteratively reweighted so that the reconstructed simulated t-distribution matches the experimental data across the full t range, and integrated over t to extract a t-integrated cross section or angular moments. While this accounts for the t-dependence of the acceptance, it relies heavily on the global fidelity of the reweighting model. As proven analytically in **Appendix G**, if the nominal generator slope b_{gen} differs from the true data slope b_{data} by $\Delta b = b_{\text{data}} - b_{\text{gen}}$, the weight variance grows exponentially with the integration width Δt. Worse still, if the generated distribution falls off more than twice as fast as the data ($b_{\text{gen}} \geq 2b_{\text{data}}$), the variance of the weights diverges (Var$(w) \to \infty$ as $|t|_{\text{max}} \to \infty$) because rare events in the high-$|t|$ tail receive exponentially exploding weights!

Method 2: Binning in t and Independent Extraction. The experimental data and MC sample are partitioned into narrow, discrete bins of width δt, and the maximum likelihood fit is performed independently within each t-bin. This strategy provides two major advantages:

1. It directly resolves the physical t-evolution of the cross section and production amplitudes.
2. Because both the detector acceptance and the true cross section vary modestly across a narrow interval δt, the intra-bin variation of the weights is bounded by $\mathrm{Var}(w)/\bar{w}^2 \approx \frac{1}{12}(\Delta b\, \delta t)^2 \ll 1$ (see **Appendix G**). Consequently, $N_{\mathrm{eff}} \approx N_{\mathrm{MC}}$ is preserved inside every bin, making the extraction nearly immune to mismodeling of the global generator t-slope.

7 Machine-Learned Fast Simulation

7.1 Motivation and Factorized Architecture

Even with optimized generators and fiducial filtering, producing massive GEANT4 samples for systematic Toy-MC parameter-extraction studies or high-statistics multi-channel background templates remains a severe computational bottleneck [15]. This overhead comes from two stages: (1) step-by-step Monte Carlo tracking of particles through complex detector materials and magnetic fields in GEANT4, and (2) pattern recognition and Kalman-filter track reconstruction from raw detector hits [15].

To bypass these overheads and achieve generation rates exceeding 10 kHz to 100 kHz [15], Darulis et al. [15] developed a machine-learned (ML) fast simulation framework built on a *factorized, per-particle architecture*. Rather than training a monolithic generative network (such as an event-level GAN [16, 17, 18]) to hallucinate entire multi-particle final states at once, the framework decouples the physics event generator from the detector response and processes each generated particle species independently through two sequential stages [15]:

1. **Fast Acceptance Model:** Determines the probability that a particle of a given species with true three-momentum $\mathbf{p}_{\mathrm{true}} = (p, \theta, \phi)$ is successfully detected and reconstructed [15].
2. **Fast Resolution Model:** For accepted particles, samples a stochastic distortion vector $\delta\mathbf{p} = (\delta p, \delta\theta, \delta\phi)$ conditional on $\mathbf{p}_{\mathrm{true}}$ to yield the reconstructed momentum $\mathbf{p}_{\mathrm{reco}} = \mathbf{p}_{\mathrm{true}} + \delta\mathbf{p}$ [15].

In practice, the training data for each particle species can be obtained either (I) via a single-particle "particle gun" covering the full kinematic range of the spectrometer (maximizing generality across reactions) or (II) by separating particles from an existing full simulation of a specific reaction (concentrating training statistics in the most relevant phase-space regions) [15]. Computationally, it is most efficient to organize the kinematic data in contiguous memory columns grouped by particle species, evaluating the acceptance and resolution models for one particle species at a time across the entire dataset rather than loading multiple models event-by-event [15].

7.2 Fast Acceptance Modeling via Calibrated Density Ratios

Detector acceptance $A(p, \theta, \phi)$ is a complex multidimensional function shaped by active detector geometry (e.g., CLAS12 torus coils and polar angle holes), magnetic field bending, trigger thresholds, and reconstruction efficiencies [15]. Following Section 4, the fast

acceptance stage estimates the multi-dimensional detection probability by framing acceptance as a binary classification density ratio problem between all generated truth particles (class $i = 0$) and accepted reconstructed particles (class $i = 1$) [15]. From a calibrated classifier outputting $c(\mathbf{p}_{\text{true}}) \approx P(i = 1 \mid \mathbf{p}_{\text{true}})$, the per-particle acceptance probability weight is given by [15]:

$$w(\mathbf{p}_{\text{true}}) = \frac{p_{\text{acc}}(\mathbf{p}_{\text{true}})}{p_{\text{gen}}(\mathbf{p}_{\text{true}})} \approx \frac{c(\mathbf{p}_{\text{true}})}{1 - c(\mathbf{p}_{\text{true}})}, \tag{10}$$

which can either be applied directly as an event weight or used in an accept/reject pass ($u \leq w$) to filter particles [15].

Systematic comparisons of classifier architectures in Ref. [15] revealed complementary strengths and weaknesses between tree-based and neural models:

- **Standalone GBDTs vs. Deep Neural Networks (DNNs):** Because standard probabilistic classifiers tend to avoid predicting strictly zero probability [19, 15], standalone GBDTs (e.g., 100 trees of maximum depth 10) accurately capture 1D projections and sharp polar boundaries (such as $0°$–$5°$ and $175°$–$180°$ beamline holes) but fail to reproduce correlated 2D (θ, ϕ) zero-acceptance gaps, bleeding false-positive events into dead zones [15]. Conversely, fully connected ReLU DNNs with sufficient capacity (requiring wide hidden layers with at least ~ 500 nodes, such as a 4-layer 512–256–128–16 or 512–384–256–16 architecture trained with the ADAM optimizer [20]) capture smooth 2D angular correlations well, but struggle with sharp 1D edge cutoffs on raw inputs [15].

- **Gaussian Quantile Preprocessing:** The inability of DNNs to model sharp boundary gaps is resolved by applying a marginal *Gaussian quantile transformation* to each input variable prior to training [15]:

$$\tilde{x}_k = G^{-1}(F_k(x_k)) \,, \tag{11}$$

  where F_k is the empirical cumulative distribution function of kinematic component x_k and G^{-1} is the standard normal quantile (probit) function [15]. This transformation maps each 1D momentum component onto a smooth Gaussian and merges discontinuous 2D angular regions into a single connected topology that neural networks learn with high fidelity [15].

- **Cascaded GBDT Reweighting Correction:** To eliminate minor residual 1D discrepancies (e.g., in ϕ) and stabilize against stochastic neural weight initialization, a secondary GBDT classifier is trained to distinguish the DNN-weighted output from the true accepted sample [15]. The GBDT provides a multiplicative correction weight w_{corr}, yielding a combined acceptance probability $w_{\text{new}} = w_{\text{DNN}} \cdot w_{\text{corr}}$ that combines the 2D topological fidelity of the DNN with the sharp 1D local calibration of the GBDT [15].

When deploying this acceptance model in practice, two experimental subtleties must be considered [15]: first, if particle identification (PID) cuts are applied prior to defining the "accepted" training class, PID efficiency is automatically folded into $w(\mathbf{p}_{\text{true}})$; second, multi-particle correlations induced by hardware data-acquisition triggers are not captured by independent single-particle models and, where significant, require an auxiliary event-level classification stage taking all final-state momenta simultaneously as input [15].

7.3 Fast Resolution Modeling via Over-Fitted Decision Tree Ensembles

Once a particle passes the acceptance stage, its true kinematics $\mathbf{p}_{\text{true}} = (p, \theta, \phi)$ must be smeared by a conditional resolution distribution $p(\delta\mathbf{p} \mid \mathbf{p}_{\text{true}})$ that exhibits non-Gaussian

radiative/energy-loss tails (such as Crystal Ball momentum shapes), momentum-dependent widths, and spherical coordinate correlations (such as $\delta\phi' = \delta\phi/\sin\theta$ and $(\delta\theta, \delta\phi)$ covariance) [15].

Rather than training complex conditional GANs [21, 22] or assuming simplified Gaussian parameterizations [23], Ref. [15] introduces a fast non-parametric sampling algorithm based on *over-fitted Decision Tree (DT) regressors* [24, 15]. By growing a regression decision tree without depth limits until every terminal leaf contains a single training event, the DT acts as an ultra-fast binary-search nearest-neighbor lookup table ($\sim 10^5$–10^6 predictions per second, an order of magnitude faster than a $k = 1$ k-Nearest Neighbour search at $\sim 10^4$ s^{-1}) [15].

Because a deterministic nearest-neighbor lookup on (p, θ, ϕ) alone would return a single fixed residual $\delta\mathbf{p}$ when queried repeatedly at the same kinematics, stochasticity and smoothness are achieved through three architectural design choices (formally analyzed in **Appendix H**) [15]:

1. **Auxiliary Random Input Injection:** The input feature vector is augmented with K independent uniform random numbers $\mathbf{r} = (r_1, \ldots, r_K) \sim U(0,1)^K$ alongside the normalized true kinematics (p, θ, ϕ) [15]. During inference at fixed (p, θ, ϕ), drawing fresh random vectors $\mathbf{r}$ routes the query to different nearby leaves in the training set, sampling from the local conditional distribution $p(\delta\mathbf{p} \mid \mathbf{p}_{\text{true}})$ [15]. Unlike kNN (where changing the numerical range of r_k alters its weight in the Euclidean distance metric), DT binary split thresholds are scale-invariant with respect to the range of $\mathbf{r}$ [15]. Increasing K improves resolution smoothness (S metric) at the cost of gradually broadening the local kinematic neighborhood and increasing the Least Squares Error (LSE) of the fitted widths, with $K = 5$ providing an optimal balance [15].

2. **Ensemble Randomization Across Multiple Trees:** Training an ensemble of N_{DT} decision trees on the same 10^6 simulated events—each initialized with an independent realization of the random inputs $\mathbf{r}$—and randomly selecting one tree per particle during inference dramatically smooths the predicted resolution distribution without degrading the width accuracy (LSE) or inference speed [15]. Smoothness plateaus at $N_{\text{DT}} = 50$ trees [15]. Because each unpruned tree trained on 10^6 events occupies ~ 150 MB on disk (7.5 GB total for 50 trees), memory usage during evaluation is kept minimal by loading one tree at a time and applying it to a randomly masked $1/N_{\text{DT}}$ subset of the dataset [15].

3. **Uniform Output Scaling and Low-Level Extension:** Because multi-output DT splits minimize the combined variance across all target variables simultaneously, leaving one target variable with a larger numerical range than the others biases the split selection and degrades the LSE [15]. All non-random output residuals $(\delta p, \delta\theta, \delta\phi)$ must therefore be scaled to an identical numerical interval (e.g., $[0, 100]$) prior to training and inverted after prediction [15]. Furthermore, adding extra target outputs does not degrade the S or LSE metrics [15], allowing the exact same DT regressor architecture to simultaneously predict lower-level detector quantities—such as time-of-flight, path length, and calorimeter energy deposits—to support realistic downstream PID algorithms [15].

7.4 Applications and the Combinatorial Statistical Advantage

This factorized fast simulation framework supports three primary workflows in electron scattering analyses [15]:

1. **Large-Scale Toy-MC Parameter Extraction Studies:** Rapidly generating hundreds of pseudo-datasets across a grid of input physics parameters to verify estimator unbiasedness, pull distributions, and fitting systematics in the presence of realistic acceptance holes and non-Gaussian resolution tails [15].

2. **High-Statistics Background Templates:** Simulating vast samples of multi-channel physical backgrounds where only a tiny kinematic tail leaks into the signal selection region, eliminating template bin fluctuations in yield extraction fits [15].

3. **Pre-Simulation Importance Filtering for GEANT4:** Using the fast acceptance weight $w(\mathbf{p}_{\text{true}})$ in an accept/reject filter prior to `GEMC` so that full GEANT4 tracking is only executed on events likely to be detected, then correcting the post-simulation distributions with the known acceptance weight [15].

Finally, it is important to address the statistical validity of machine-learned datasets. Matchev et al. [25] demonstrated that when a monolithic generative model (such as an end-to-end event-level GAN) is trained on N_{train} fully simulated events to replace the entire Monte Carlo pipeline, the statistical information content of the generated synthetic dataset is fundamentally capped by N_{train}, regardless of how many synthetic events are drawn [25, 15].

Crucially, this statistical limitation is substantially mitigated by the factorized per-particle architecture [15]. First, the physics reaction kinematics are sampled fresh for every event from an exact theoretical generator rather than interpolated from a fixed training set [15]. Second, for an M-body exclusive reaction (such as $\gamma^{(*)}p \to p'\pi^+\pi^-$ tested in Ref. [15], where 5×10^6 reaction events were simulated from models trained on 10^6 single-particle events, achieving an overall normalization agreement within 1% [15]), each final-state particle is smeared independently by its species-specific acceptance and resolution ensemble [15]. As shown in **Appendix H**, combining M independently perturbed particle four-vectors produces up to $\sim (N_{\text{train}})^M$ distinct combinatorial configurations in derived multi-particle observables—such as invariant masses W, $M(\pi^+\pi^-)$, and Gottfried-Jackson or center-of-mass decay angles $(\cos\theta_{\text{GJ}}, \phi_{\text{GJ}}, \cos\theta_{\text{CM}}, \phi_{\text{CM}})$ [15]. This random combinatorial mixing naturally smooths out residual single-particle decision-tree discretization artifacts and provides vastly greater statistical power than monolithic event-level generative models [15].

A Please note appendices have not been checked in detail

B Derivation of Importance Sampling Variance and Kish's Effective Sample Size

Consider estimating the expectation of an observable $\mathcal{O}(x)$ under a target physical probability density $p_{\text{data}}(x)$ using events $\{x_1, \ldots, x_{N_{\text{MC}}}\}$ drawn independently from a baseline generator density $p_{\text{gen}}(x)$. When $p_{\text{data}}(x)$ or $p_{\text{gen}}(x)$ is known up to a normalization constant, we define the unnormalized weight function $w(x) = c\,\frac{p_{\text{data}}(x)}{p_{\text{gen}}(x)}$ (where $c > 0$ is an arbitrary constant) and construct the self-normalized importance sampling estimator [1]:

$$\hat{\mathcal{O}} = \frac{\frac{1}{N_{\text{MC}}}\sum_{i=1}^{N_{\text{MC}}} w_i \mathcal{O}(x_i)}{\frac{1}{N_{\text{MC}}}\sum_{i=1}^{N_{\text{MC}}} w_i} \equiv \frac{\overline{w\mathcal{O}}}{\bar{w}}, \tag{12}$$

where $w_i = w(x_i)$. Taking expectations with respect to the sampling distribution $p_{\text{gen}}(x)$, the mean weight is:

$$\mu_w \equiv \mathbb{E}_{\text{gen}}[w] = \int c\frac{p_{\text{data}}(x)}{p_{\text{gen}}(x)} p_{\text{gen}}(x)\, dx = c, \tag{13}$$

and the mean of the numerator is $\mu_{w\mathcal{O}} \equiv \mathbb{E}_{\text{gen}}[w\mathcal{O}] = c\int \mathcal{O}(x) p_{\text{data}}(x)\, dx = c\langle\mathcal{O}\rangle$.

To find the variance of the ratio $\hat{\mathcal{O}} = \overline{w\mathcal{O}}/\bar{w}$ for large N_{MC}, we perform a first-order Taylor expansion (the statistical delta method) around the true means $(\mu_{w\mathcal{O}}, \mu_w)$:

$$\hat{\mathcal{O}} - \langle\mathcal{O}\rangle \approx \frac{1}{\mu_w}(\overline{w\mathcal{O}} - \mu_{w\mathcal{O}}) - \frac{\mu_{w\mathcal{O}}}{\mu_w^2}(\bar{w} - \mu_w) = \frac{1}{N_{\text{MC}}\mu_w}\sum_{i=1}^{N_{\text{MC}}} w_i \left(\mathcal{O}(x_i) - \langle\mathcal{O}\rangle\right). \tag{14}$$

Since the events x_i are independent and identically distributed under $p_{\text{gen}}(x)$ and $\mathbb{E}_{\text{gen}}[w(\mathcal{O} - \langle\mathcal{O}\rangle)] = 0$, the asymptotic variance of $\hat{\mathcal{O}}$ is:

$$\text{Var}(\hat{\mathcal{O}}) \approx \frac{1}{N_{\text{MC}}\mu_w^2}\mathbb{E}_{\text{gen}}\left[w(x)^2\left(\mathcal{O}(x) - \langle\mathcal{O}\rangle\right)^2\right]. \tag{15}$$

To obtain a general figure of merit for the MC sample itself (independent of a specific observable $\mathcal{O}$), Kish [2] considered the approximation in which the local fluctuations of $(\mathcal{O}(x) - \langle\mathcal{O}\rangle)^2$ are uncorrelated with $w(x)^2$, allowing the expectation in Eq. (15) to factorize into the intrinsic physical variance $\sigma_{\mathcal{O}}^2 = \text{Var}_{\text{data}}(\mathcal{O})$ and the second moment of the weights:

$$\text{Var}(\hat{\mathcal{O}}) \approx \frac{\sigma_{\mathcal{O}}^2}{N_{\text{MC}}}\frac{\mathbb{E}_{\text{gen}}[w^2]}{\mu_w^2} = \frac{\sigma_{\mathcal{O}}^2}{N_{\text{MC}}}\left(1 + \frac{\text{Var}_{\text{gen}}(w)}{\mu_w^2}\right). \tag{16}$$

For an ideal unweighted sample drawn directly from $p_{\text{data}}(x)$ of size N_{eff}, the variance of the sample mean is simply $\text{Var}(\hat{\mathcal{O}}_{\text{ideal}}) = \sigma_{\mathcal{O}}^2/N_{\text{eff}}$. Equating $\text{Var}(\hat{\mathcal{O}}) = \sigma_{\mathcal{O}}^2/N_{\text{eff}}$ and replacing population moments with sample moments ($\mu_w \approx \bar{w} = \frac{1}{N_{\text{MC}}}\sum w_i$ and $\mathbb{E}_{\text{gen}}[w^2] \approx \frac{1}{N_{\text{MC}}}\sum w_i^2$) yields Kish's effective sample size:

$$N_{\text{eff}} = \frac{N_{\text{MC}}}{\mathbb{E}_{\text{gen}}[w^2]/\mu_w^2} \approx \frac{\left(\sum_{i=1}^{N_{\text{MC}}} w_i\right)^2}{\sum_{i=1}^{N_{\text{MC}}} w_i^2} = \frac{N_{\text{MC}}}{1 + \text{Var}(w)/\bar{w}^2}. \tag{17}$$

By the Cauchy-Schwarz inequality, $\left(\sum_{i=1}^{N_{\text{MC}}} w_i \cdot 1\right)^2 \le \left(\sum_{i=1}^{N_{\text{MC}}} w_i^2\right)\left(\sum_{i=1}^{N_{\text{MC}}} 1^2\right) = N_{\text{MC}}\sum_{i=1}^{N_{\text{MC}}} w_i^2$, proving that $N_{\text{eff}} \le N_{\text{MC}}$, with equality if and only if $w_1 = w_2 = \cdots = w_{N_{\text{MC}}}$.

C Proof of Accept/Reject Exactness and Grid Envelope Optimization

Let $f(x)$ be a target probability density and $g(x)$ be a proposal density on Ω satisfying $f(x) \le Mg(x)$ for all $x \in \Omega$, where $M = \sup_{x\in\Omega}[f(x)/g(x)] \ge 1$. In each trial, we draw $X \sim g(x)$ and $U \sim U(0,1)$ independently, accepting X if the event $A = \left\{U \le \frac{f(X)}{Mg(X)}\right\}$ occurs.

The conditional probability of acceptance given $X = x$ is simply the length of the uniform interval:

$$P(A \mid X = x) = P\left(U \le \frac{f(x)}{Mg(x)}\right) = \frac{f(x)}{Mg(x)}, \tag{18}$$

since $0 \le \frac{f(x)}{Mg(x)} \le 1$ by construction. The marginal probability that a single trial is accepted (the generation efficiency ϵ_{gen}) is obtained by integrating over all possible draws x:

$$\epsilon_{\text{gen}} \equiv P(A) = \int_\Omega P(A \mid X = x)\, g(x)\, dx = \int_\Omega \frac{f(x)}{Mg(x)} g(x)\, dx = \frac{1}{M}\int_\Omega f(x)\, dx = \frac{1}{M}. \tag{19}$$

By Bayes' theorem, the probability density of the accepted events, $p(x \mid A)$, is:

$$p(x \mid A) = \frac{P(A \mid X = x)\, g(x)}{P(A)} = \frac{\left(\frac{f(x)}{Mg(x)}\right) g(x)}{1/M} = f(x). \tag{20}$$

Thus, accepted events follow $f(x)$ exactly without approximation.

In a grid-based generator such as `elSpectro` [4], the phase space Ω is partitioned into B hyper-rectangular cells $\{\Omega_b\}_{b=1}^B$. A piecewise constant proposal distribution is constructed by setting $g(x) = c_b/Z$ for $x \in \Omega_b$, where $c_b = \sup_{x\in\Omega_b} f(x)$ is the local maximum in cell b and $Z = \sum_{b=1}^B c_b \text{Vol}(\Omega_b)$ normalizes $g(x)$. Then, for any $x \in \Omega_b$:

$$\frac{f(x)}{g(x)} = Z\frac{f(x)}{c_b} \le Z \implies M_{\text{grid}} = Z = \sum_{b=1}^{B}\left(\sup_{x\in\Omega_b} f(x)\right)\text{Vol}(\Omega_b). \tag{21}$$

As the grid resolution increases ($B \to \infty$, $\text{Vol}(\Omega_b) \to 0$), the Riemann upper sum converges to the exact integral: $M_{\text{grid}} \to \int_\Omega f(x)\, dx = 1$, driving the generation efficiency $\epsilon_{\text{gen}} = 1/M_{\text{grid}} \to 1$.

D Fiducial Pre-Filtering and Cross-Section Normalization

The physical differential cross section in a kinematic bin Δx is related to the experimentally observed background-subtracted yield $N_{\text{data}}(\Delta x)$ and integrated luminosity $\mathcal{L}_{\text{int}}$ by:

$$\frac{d\sigma}{dx}(\Delta x) = \frac{N_{\text{data}}(\Delta x)}{\mathcal{L}_{\text{int}}\, \Delta x\, \epsilon_{4\pi}(\Delta x)}, \tag{22}$$

where $\epsilon_{4\pi}(\Delta x)$ is the true fraction of all physical events produced in Δx (over the full 4π decay/scattering solid angle $\Omega_{4\pi}$) that are reconstructed in the detector.

Suppose the event generator generates $N^{4\pi}_{\text{trial}}(\Delta x)$ events over $\Omega_{4\pi}$, applies a fast geometric pre-filter $\Omega_{\text{fid}} \subset \Omega_{4\pi}$ (such as the CLAS12 FT or FD electron angle cut), passes only the $N_{\text{filt}}(\Delta x)$ surviving events into `GEMC`, and reconstructs $N_{\text{reco}}(\Delta x)$ events. Because any

event outside Ω_{fid} has strictly zero probability of being reconstructed in the target detector configuration ($P(\text{reco} \mid x \notin \Omega_{\text{fid}}) = 0$), the Law of Total Probability gives:

$$\begin{aligned}\epsilon_{4\pi}(\Delta x) &= P(\text{reco} \mid x \in \Omega_{4\pi}) \\ &= P(\text{reco} \mid x \in \Omega_{\text{fid}})\, P(x \in \Omega_{\text{fid}} \mid x \in \Omega_{4\pi}) + \underbrace{P(\text{reco} \mid x \notin \Omega_{\text{fid}})}_{=0}\, P(x \notin \Omega_{\text{fid}} \mid x \in \Omega_{4\pi}) \\ &= \left(\frac{N_{\text{reco}}(\Delta x)}{N_{\text{filt}}(\Delta x)}\right)\left(\frac{N_{\text{filt}}(\Delta x)}{N^{4\pi}_{\text{trial}}}(\Delta x)\right) \equiv \epsilon_{\text{fid}}(\Delta x) \cdot f_{\text{fid}}(\Delta x). \end{aligned} \tag{23}$$

Equivalently, in in-situ generator normalization where the total simulated integrated luminosity is $\mathcal{L}_{\text{MC}} = N^{4\pi}_{\text{trial}}/\sigma_{\text{gen}}(\Omega_{4\pi})$, restricting generation to Ω_{fid} replaces $\sigma_{\text{gen}}(\Omega_{4\pi})$ with the fiducial cross section $\sigma_{\text{gen}}(\Omega_{\text{fid}}) = f_{\text{fid}}\, \sigma_{\text{gen}}(\Omega_{4\pi})$, so that:

$$\mathcal{L}_{\text{MC}} = \frac{N^{4\pi}_{\text{trial}}}{\sigma_{\text{gen}}(\Omega_{4\pi})} = \frac{N_{\text{filt}}}{\sigma_{\text{gen}}(\Omega_{\text{fid}})}. \tag{24}$$

E Variational Proof of the Likelihood Ratio Trick and Cascaded Reweighting

Let $p_0(x)$ (e.g., baseline or generated events, label $y = 0$) and $p_1(x)$ (e.g., target or accepted events, label $y = 1$) be two probability densities on Ω, represented by finite training samples of sizes N_0 and N_1 with prior class probabilities $\pi_0 = \frac{N_0}{N_0+N_1}$ and $\pi_1 = \frac{N_1}{N_0+N_1}$ [10, 15].

A probabilistic classifier $c(x) \in (0, 1)$ is trained to minimize the population binary cross-entropy (log-loss) functional [10, 15]:

$$\mathcal{J}[c] = -\int_{\Omega} \Big[\pi_1 p_1(x) \ln c(x) + \pi_0 p_0(x) \ln\big(1 - c(x)\big)\Big] dx. \tag{25}$$

To find the optimal function $c^*(x)$ that minimizes $\mathcal{J}[c]$, we set the functional derivative $\frac{\delta \mathcal{J}}{\delta c(x)} = 0$ pointwise for each x:

$$\frac{\delta \mathcal{J}}{\delta c(x)} = -\frac{\pi_1 p_1(x)}{c^*(x)} + \frac{\pi_0 p_0(x)}{1 - c^*(x)} = 0 \implies \pi_1 p_1(x)\big(1 - c^*(x)\big) = \pi_0 p_0(x) c^*(x). \tag{26}$$

Solving for $c^*(x)$ yields the Bayes-optimal posterior probability:

$$c^*(x) = \frac{\pi_1 p_1(x)}{\pi_0 p_0(x) + \pi_1 p_1(x)} = P(y = 1 \mid x), \tag{27}$$

and the second functional derivative $\frac{\delta^2 \mathcal{J}}{\delta c(x)^2} = \frac{\pi_1 p_1}{c^2} + \frac{\pi_0 p_0}{(1-c)^2} > 0$ confirms that $c^*(x)$ is a strict global minimum. Taking the odds ratio of $c^*(x)$ directly recovers the exact multidimensional density ratio [10, 13, 15]:

$$\frac{c^*(x)}{1 - c^*(x)} = \frac{\pi_1 p_1(x)}{\pi_0 p_0(x)} = \left(\frac{N_1}{N_0}\right)\frac{p_1(x)}{p_0(x)} \implies w(x) \equiv \frac{p_1(x)}{p_0(x)} = \left(\frac{N_0}{N_1}\right)\frac{c^*(x)}{1 - c^*(x)}. \tag{28}$$

Note that in the acceptance problem of Section 7.2, if the classifier is trained directly on the unnormalized empirical event sets where $N_1 = N_{\text{acc}}$ and $N_0 = N_{\text{gen}}$, the raw odds ratio $\frac{c(x)}{1-c(x)}$ automatically equals $\frac{N_{\text{acc}}\, p_{\text{acc}}(x)}{N_{\text{gen}}\, p_{\text{gen}}(x)} = A(x)$, the absolute per-event detection probability in $[0, 1]$ [15]!

Furthermore, suppose an initial classifier (such as a DNN) produces an approximate weight $w_{\text{DNN}}(x)$, defining an intermediate weighted density $p_{\text{int}}(x) = w_{\text{DNN}}(x)\, p_0(x)$ [15].

Training a second classifier (such as a GBDT) to discriminate $p_1(x)$ from $p_{\text{int}}(x)$ yields a residual correction weight $w_{\text{corr}}(x) = \frac{p_1(x)}{p_{\text{int}}(x)}$ [15]. Multiplying the two weights cancels the intermediate distribution identically [15]:

$$w_{\text{new}}(x) = w_{\text{DNN}}(x) \cdot w_{\text{corr}}(x) = \frac{p_{\text{int}}(x)}{p_0(x)} \cdot \frac{p_1(x)}{p_{\text{int}}(x)} = \frac{p_1(x)}{p_0(x)}. \tag{29}$$

E.1 Validity of Empirical Reconstructed-Space Reweighting without Unfolding

In data-driven reweighting, the density ratio is learned from reconstructed distributions $p_{\text{data}}(x_{\text{reco}})$ and $p_{\text{base}}(x_{\text{reco}})$[cite: 2, 6], which are related to the true underlying densities $p(x_{\text{true}})$ through the detector acceptance $\eta(x_{\text{true}})$ and resolution response kernel $R(x_{\text{reco}} \mid x_{\text{true}})$:

$$w(x_{\text{reco}}) = \frac{p_{\text{data}}(x_{\text{reco}})}{p_{\text{base}}(x_{\text{reco}})} = \frac{\int_\Omega R(x_{\text{reco}} \mid x_{\text{true}})\, \eta(x_{\text{true}})\, p_{\text{data}}(x_{\text{true}})\, dx_{\text{true}}}{\int_\Omega R(x_{\text{reco}} \mid x_{\text{true}})\, \eta(x_{\text{true}})\, p_{\text{base}}(x_{\text{true}})\, dx_{\text{true}}}. \tag{30}$$

If the detector resolution kernel $R(x_{\text{reco}} \mid x_{\text{true}})$ is localized with characteristic width σ_x around $x_{\text{true}} \approx x_{\text{reco}}$, and the true ratio $r(x_{\text{true}}) \equiv p_{\text{data}}(x_{\text{true}})/p_{\text{base}}(x_{\text{true}})$ varies smoothly on the scale of σ_x ($|\nabla \ln r|\, \sigma_x \ll 1$), we can Taylor-expand $p_{\text{data}}(x_{\text{true}}) = r(x_{\text{true}})\, p_{\text{base}}(x_{\text{true}}) \approx r(x_{\text{reco}})\, p_{\text{base}}(x_{\text{true}})$ inside the integral. The common factor $\int R(x_{\text{reco}} \mid x_{\text{true}})\, \eta(x_{\text{true}})\, p_{\text{base}}(x_{\text{true}})\, dx_{\text{true}}$ then cancels between numerator and denominator, yielding:

$$w(x_{\text{reco}}) = r(x_{\text{reco}}) + \mathcal{O}\big(\sigma_x^2 \nabla^2 r\big) \approx \left.\frac{p_{\text{data}}(x_{\text{true}})}{p_{\text{base}}(x_{\text{true}})}\right|_{x_{\text{true}} = x_{\text{reco}}}. \tag{31}$$

Under this small-smearing condition, both the detector acceptance $\eta(x)$ and resolution kernel R cancel to leading order in the reconstructed ratio, justifying the direct evaluation of the learned weight function $w(\cdot)$ on the generator-level kinematics x_{gen} without full unfolding.

F Extended Maximum Likelihood Normalization with Monte Carlo Events

In an unbinned Extended Maximum Likelihood (EML) fit, the total number of observed events N_{data} is treated as a Poisson random variable with expectation $N_{\text{exp}}(\boldsymbol{\theta})$, while each event x_i follows the normalized probability density $P_{\text{obs}}(x \mid \boldsymbol{\theta})$ of reconstructed events:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{e^{-N_{\text{exp}}(\boldsymbol{\theta})} \big[N_{\text{exp}}(\boldsymbol{\theta})\big]^{N_{\text{data}}}}{N_{\text{data}}!} \prod_{i=1}^{N_{\text{data}}} P_{\text{obs}}(x_i \mid \boldsymbol{\theta}). \tag{32}$$

Given a theoretical differential cross section or intensity $\mathcal{I}(x \mid \boldsymbol{\theta})$, an integrated luminosity $\mathcal{L}_{\text{int}}$, and a detector acceptance efficiency $\eta(x)$, the expected yield is:

$$N_{\text{exp}}(\boldsymbol{\theta}) = \mathcal{L}_{\text{int}} \int_\Omega \mathcal{I}(x \mid \boldsymbol{\theta})\, \eta(x)\, dx, \tag{33}$$

and the normalized probability density of observed events is:

$$P_{\text{obs}}(x \mid \boldsymbol{\theta}) = \frac{\mathcal{L}_{\text{int}}\, \mathcal{I}(x \mid \boldsymbol{\theta})\, \eta(x)}{N_{\text{exp}}(\boldsymbol{\theta})}. \tag{34}$$

Substituting $P_{\text{obs}}(x \mid \boldsymbol{\theta})$ into Eq. (32) cancels $\big[N_{\text{exp}}(\boldsymbol{\theta})\big]^{N_{\text{data}}}$, and taking the negative logarithm gives:

$$-\ln \mathcal{L}(\boldsymbol{\theta}) = N_{\text{exp}}(\boldsymbol{\theta}) - \sum_{i=1}^{N_{\text{data}}} \ln \mathcal{I}(x_i \mid \boldsymbol{\theta}) - \underbrace{\sum_{i=1}^{N_{\text{data}}} \ln\big[\mathcal{L}_{\text{int}}\, \eta(x_i)\big] + \ln(N_{\text{data}}!)}_{\text{independent of } \boldsymbol{\theta}}. \tag{35}$$

Notice that the point-by-point detector acceptance $\eta(x_i)$ drops out of the data sum because it does not depend on the physics parameters $\boldsymbol{\theta}$; the detector acceptance enters the optimization exclusively through the normalization integral $N_{\text{exp}}(\boldsymbol{\theta})$.

To evaluate $N_{\text{exp}}(\boldsymbol{\theta})$ via Monte Carlo, suppose N_{gen} events are drawn from a generator distribution $p_{\text{gen}}(x) = \mathcal{I}_{\text{gen}}(x)/\sigma_{\text{gen}}$ and passed through the detector simulation, which accepts each event with probability $\eta(x)$, leaving N_{acc} accepted MC events. The joint density of accepted MC events is $p_{\text{acc}}(x)\,dx = \eta(x) p_{\text{gen}}(x)\,dx/\bar{\eta}$, where $\bar{\eta} \approx N_{\text{acc}}/N_{\text{gen}}$. Rewriting $N_{\text{exp}}(\boldsymbol{\theta})$ via importance sampling over $p_{\text{gen}}(x)$ yields:

$$\begin{aligned} N_{\text{exp}}(\boldsymbol{\theta}) &= \mathcal{L}_{\text{int}} \sigma_{\text{gen}} \int_{\Omega} \frac{\mathcal{I}(x \mid \boldsymbol{\theta})}{\mathcal{I}_{\text{gen}}(x)} \eta(x) p_{\text{gen}}(x)\, dx \\ &\approx \frac{\mathcal{L}_{\text{int}} \sigma_{\text{gen}}}{N_{\text{gen}}} \sum_{j=1}^{N_{\text{acc}}} w_j^{\text{corr}} \frac{\mathcal{I}(x_j \mid \boldsymbol{\theta})}{\mathcal{I}_{\text{gen}}(x_j)}, \end{aligned} \tag{36}$$

where w_j^{corr} incorporates any post-simulation tuning or tracking efficiency weights.

G Electroproduction Flux Factorization and t-Slope Weight Variance

G.1 One-Photon Exchange Factorization

In exclusive electroproduction $e(k)+p(P) \to e'(k')+X(P')$, the invariant kinematic variables are the photon virtuality $Q^2 = -(k-k')^2$, the center-of-mass energy squared $W^2 = (q+P)^2 = M_p^2 + 2M_p\nu - Q^2$ (with $q = k - k'$ and $\nu = E_b - E_{e'}$), and the squared four-momentum transfer to the target nucleon $t = (P - P_{p'})^2$. In the one-photon exchange approximation, the unpolarized triple-differential cross section factorizes as:

$$\frac{d^3\sigma}{dQ^2\, dW\, dt} = \Gamma_v(E_b, Q^2, W)\, \frac{d\sigma_{\gamma^* p}}{dt}(W, Q^2, t), \tag{37}$$

where the virtual photon flux factor Γ_v (in the Hand convention) is:

$$\Gamma_v(E_b, Q^2, W) = \frac{\alpha_{\text{em}}}{2\pi^2} \frac{W(W^2 - M_p^2)}{2M_p^2 E_b^2 Q^2 (1-\epsilon)}, \quad \text{with} \quad \epsilon = \left[1 + 2\left(1 + \frac{\nu^2}{Q^2}\right) \tan^2 \frac{\theta_e}{2}\right]^{-1}. \tag{38}$$

Combined with the transition form factor $F(Q^2) \sim (1 + Q^2/\Lambda^2)^{-2}$ inside $d\sigma_{\gamma^* p}/dt$, the cross section scales as $\sim 1/Q^4$ at low Q^2 and even more steeply at large Q^2.

G.2 Analytical Weight Variance of Exponential t-Slope Reweighting

Let $\tau \equiv |t| - |t|_{\text{min}} \in [0, \Delta t]$ denote the shifted momentum transfer over an interval of width Δt. Suppose the baseline MC is generated with an exponential slope $b_{\text{gen}} > 0$ and reweighted to match a true physical slope $b_{\text{data}} > 0$:

$$p_{\text{gen}}(\tau) = \frac{b_{\text{gen}} e^{-b_{\text{gen}}\tau}}{1 - e^{-b_{\text{gen}}\Delta t}}, \qquad p_{\text{data}}(\tau) = \frac{b_{\text{data}} e^{-b_{\text{data}}\tau}}{1 - e^{-b_{\text{data}}\Delta t}}. \tag{39}$$

Defining the slope mismatch $\Delta b \equiv b_{\text{data}} - b_{\text{gen}}$, the importance weight is $w(\tau) = \frac{p_{\text{data}}(\tau)}{p_{\text{gen}}(\tau)} \propto e^{-\Delta b\,\tau}$ with $\mathbb{E}_{\text{gen}}[w] = 1$. The second moment of the weights under $p_{\text{gen}}(\tau)$ is:

$$\begin{aligned}\mathbb{E}_{\text{gen}}[w^2] &= \int_0^{\Delta t} \frac{p_{\text{data}}(\tau)^2}{p_{\text{gen}}(\tau)}\, d\tau = \frac{b_{\text{data}}^2}{b_{\text{gen}}} \frac{1 - e^{-b_{\text{gen}}\Delta t}}{\left(1 - e^{-b_{\text{data}}\Delta t}\right)^2} \int_0^{\Delta t} e^{-(2b_{\text{data}} - b_{\text{gen}})\tau}\, d\tau \\ &= \frac{b_{\text{data}}^2}{b_{\text{gen}}(2b_{\text{data}} - b_{\text{gen}})} \left(\frac{1 - e^{-b_{\text{gen}}\Delta t}}{\left(1 - e^{-b_{\text{data}}\Delta t}\right)^2}\right) \left(1 - e^{-(2b_{\text{data}} - b_{\text{gen}})\Delta t}\right). \end{aligned} \tag{40}$$

Equation (40) reveals two critical regimes corresponding to Methods 1 and 2 in Section 6:

1. **Global Integration Limit ($\Delta t \to \infty$, Method 1):** If $b_{\text{gen}} < 2b_{\text{data}}$, the exponentials vanish for wide t ranges and the relative weight variance approaches:

$$\frac{\text{Var}(w)}{\bar{w}^2} = \mathbb{E}_{\text{gen}}[w^2] - 1 \xrightarrow{\Delta t \to \infty} \frac{b_{\text{data}}^2}{b_{\text{gen}}(2b_{\text{data}} - b_{\text{gen}})} - 1 = \frac{(\Delta b)^2}{b_{\text{gen}}(2b_{\text{data}} - b_{\text{gen}})}. \tag{41}$$

   However, if the generator slope is too steep ($b_{\text{gen}} \geq 2b_{\text{data}}$), the exponent $-(2b_{\text{data}} - b_{\text{gen}})\Delta t \geq 0$ in Eq. (40) grows exponentially with Δt, causing $\text{Var}(w) \to \infty$ and $N_{\text{eff}} \to 0$!

2. **Narrow Binning Limit ($|\Delta b|\,\delta t \ll 1$, Method 2):** Expanding Eq. (40) in powers of a narrow bin width δt (for *any* values of b_{gen} and b_{data}) gives:

$$\frac{\text{Var}(w)}{\bar{w}^2} = \frac{1}{12}(\Delta b\,\delta t)^2 + \mathcal{O}\big((\delta t)^4\big) \implies \frac{N_{\text{eff}}}{N_{\text{MC}}} \approx \frac{1}{1 + \frac{1}{12}(\Delta b\,\delta t)^2}. \tag{42}$$

   For example, even with a substantial slope mismatch of $|\Delta b| = 2$ GeV^{-2} across a bin of width $\delta t = 0.1$ GeV^2, the relative weight variance is only $\frac{1}{12}(0.2)^2 \approx 0.0033$, preserving $N_{\text{eff}} = 99.7\%\, N_{\text{MC}}$.

H Non-Parametric Conditional Resolution Sampling and Metric Definitions

H.1 Conditional Sampling via Randomized Decision Tree Partitions

Let $\mathcal{D} = \{(\mathbf{p}_i, \delta\mathbf{p}_i)\}_{i=1}^{N_{\text{train}}}$ be a training sample drawn from the true joint detector distribution $p(\mathbf{p}_{\text{true}}, \delta\mathbf{p}) = p(\mathbf{p}_{\text{true}})\, p(\delta\mathbf{p} \mid \mathbf{p}_{\text{true}})$ [15]. For each tree $m \in \{1, \ldots, N_{\text{DT}}\}$, we augment each training event i with an independent K-dimensional uniform random vector $\mathbf{r}_i^{(m)} \sim U(0,1)^K$, forming the extended input space $\mathbf{z}_i^{(m)} = (\mathbf{p}_i, \mathbf{r}_i^{(m)}) \in [0,1]^{3+K}$ [15].

An unpruned regression tree T_m partitions $[0,1]^{3+K}$ into N_{train} disjoint hyper-rectangular leaf cells $\{R_{m,i}\}_{i=1}^{N_{\text{train}}}$, such that $\mathbf{z}_i^{(m)} \in R_{m,i}$ and the tree outputs $\delta\mathbf{p}_i$ whenever a query $\mathbf{z} = (\mathbf{p}, \mathbf{r})$ falls inside $R_{m,i}$ [15]. At fixed kinematics $\mathbf{p}$, let $V_{m,i}(\mathbf{p}) = \int_{[0,1]^K} \mathbb{I}((\mathbf{p}, \mathbf{r}) \in R_{m,i})\, d^K\mathbf{r}$ be the K-dimensional volume of leaf $R_{m,i}$ sliced at $\mathbf{p}$. Averaging over uniform draws of the tree index $m \in \{1, \ldots, N_{\text{DT}}\}$ and random inputs $\mathbf{r} \sim U(0,1)^K$, the fast resolution model defines the empirical conditional density:

$$\hat{p}(\delta\mathbf{p} \mid \mathbf{p}) = \frac{1}{N_{\text{DT}}} \sum_{m=1}^{N_{\text{DT}}} \sum_{i=1}^{N_{\text{train}}} V_{m,i}(\mathbf{p})\, \delta^{(3)}(\delta\mathbf{p} - \delta\mathbf{p}_i). \tag{43}$$

Because splits along the kinematic coordinates $\mathbf{p}$ localize $V_{m,i}(\mathbf{p})$ to training events i within a small neighborhood $\|\mathbf{p}_i - \mathbf{p}\| \leq \epsilon$, while splits along the auxiliary coordinates $\mathbf{r}$ randomly subdivide that local neighborhood among nearby training events, Eq. (43) acts as a localized kernel conditional density estimator that preserves exact 3D residual correlations $(\delta p_i, \delta\theta_i, \delta\phi_i)$ by returning complete training triplets $\delta\mathbf{p}_i$ [15].

H.2 Quantitative Smoothness (S) and Least Squares Error (LSE) Metrics

To optimize the hyperparameters (N_{DT}, K) quantitatively, Darulis et al. [15] define two complementary metrics:

1. **Smoothness Metric (S):** When a fixed kinematic vector $\mathbf{p}_{\text{true}} = (p, \theta, \phi)$ is evaluated N times (e.g., $N = 10^5$) and the predicted residuals are histogrammed into bins with counts $\{b_n\}$, the smoothness metric is defined as the normalized total variation across adjacent bins [15]:
$$S = \sum_n \frac{|b_n - b_{n+1}|}{2N}. \tag{44}$$
A completely degenerate prediction where all N evaluations fall into a single bin gives $S = 1$, whereas a smooth, well-populated continuous distribution approaches $S \to 0$ (achieving $S = 0.09$ for $N_{\text{DT}} = 50, K = 5$ at $N_{\text{train}} = 10^6$) [15].

2. **Least Squares Error (LSE):** To quantify the accuracy of the kinematic dependence of the resolution widths, the phase space (p, θ, ϕ) is divided into bins, and the resolution width σ_{obs} of the fast simulation is compared against the true toy/GEANT4 width σ_{exp} across all bins and all three momentum components [15]:
$$\text{LSE} = \sum_{\text{bins, } (\delta p, \delta\theta, \delta\phi)} \frac{(\sigma_{\text{obs}} - \sigma_{\text{exp}})^2}{\sigma_{\text{exp}}}. \tag{45}$$
An ideal fast simulation that reproduces the true conditional resolution widths everywhere yields LSE $= 0$ (achieving LSE $= 0.2$ for the baseline configuration in Ref. [15]) [15].

H.3 Combinatorial Scaling in Factorized Multi-Particle Final States

In an M-body exclusive reaction $ep \to e'h_1h_2 \dots h_M$, a monolithic event-level generative model trained on N_{train} events learns a direct mapping from latent noise to the $(3M)$-dimensional joint space, bounding the statistical precision of any derived observable by the Fisher information of the N_{train} training events [25, 15].

By contrast, in the factorized per-particle framework [15], the conditional distortion density for the full event factorizes as a product of M independent single-particle species densities:
$$\hat{p}(\delta\mathbf{p}_1, \dots, \delta\mathbf{p}_M \mid \mathbf{p}_1, \dots, \mathbf{p}_M) = \prod_{a=1}^{M} \hat{p}_a(\delta\mathbf{p}_a \mid \mathbf{p}_a). \tag{46}$$

Even if each single-particle model $\hat{p}_a$ draws residuals from a finite pool of $N_{\text{train}}^{(a)}$ training leaves, any multi-particle kinematic observable $Y = \mathcal{F}(\mathbf{p}_1 + \delta\mathbf{p}_1, \dots, \mathbf{p}_M + \delta\mathbf{p}_M)$ (such as a multi-hadron invariant mass W, two-pion mass $M(\pi^+\pi^-)$, or Gottfried-Jackson decay angle [15]) is a convolution over M independent leaf selections combined with continuous generator truth momenta $(\mathbf{p}_1, \dots, \mathbf{p}_M)$ [15]. The total number of distinct residual combinations available to populate the distribution of Y scales combinatorially as $\prod_{a=1}^{M} N_{\text{train}}^{(a)} \sim (N_{\text{train}})^M$, suppressing single-particle discretization variance by orders of magnitude [15].

References

[1] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, 3rd ed. (John Wiley & Sons, Hoboken, NJ, 2016).

[2] L. Kish, *Survey Sampling* (John Wiley & Sons, New York, 1965).

[3] J. von Neumann, "Various techniques used in connection with random digits," Nat. Bur. Stand. Appl. Math. Ser. **12**, 36–38 (1951).

[4] D. I. Glazier, "`elSpectro` and `clas12-elSpectro`: An Event Generator Framework for Incorporating Spectroscopy into Electro/Photoproduction Reactions," `https://github.com/dglazier/elSpectro` and `https://github.com/dglazier/clas12-elSpectro` (2022).

[5] V. D. Burkert *et al.* (CLAS Collaboration), "The CLAS12 Spectrometer at Jefferson Laboratory," Nucl. Instrum. Meth. A **959**, 163419 (2020).

[6] S. Agostinelli *et al.* (GEANT4 Collaboration), "GEANT4—a simulation toolkit," Nucl. Instrum. Meth. A **506**, 250–303 (2003).

[7] M. Ungaro *et al.*, "The CLAS12 Geant4 simulation," Nucl. Instrum. Meth. A **959**, 163422 (2020).

[8] N. Baltzell *et al.*, "The CLAS12 software framework and event reconstruction," Nucl. Instrum. Meth. A **959**, 163472 (2020).

[9] M. Battaglieri *et al.*, "The CLAS12 Forward Tagger," Nucl. Instrum. Meth. A **959**, 163441 (2020).

[10] M. Sugiyama, T. Suzuki, and T. Kanamori, "Probabilistic classification," in *Density Ratio Estimation in Machine Learning* (Cambridge University Press, Cambridge, 2012), pp. 47–55.

[11] K. Cranmer, J. Pavez, and G. Louppe, "Approximating Likelihood Ratios with Calibrated Discriminative Classifiers," arXiv:1506.02169 [stat.AP] (2015).

[12] A. Rogozhnikov, "Reweighting with Boosted Decision Trees," J. Phys. Conf. Ser. **762**, 012036 (2016).

[13] A. Andreassen and B. Nachman, "Neural Networks for Full Phase-Space Reweighting and Parameter Tuning," Phys. Rev. D **101**, 091901 (2020).

[14] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," Comput. Phys. Commun. **77**, 219–228 (1993).

[15] D. Darulis, R. Tyson, D. G. Ireland, D. I. Glazier, B. McKinnon, and P. Pauli, "Machine Learned Particle Detector Simulations," arXiv:2207.11254 [physics.data-an] (2022).

[16] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," arXiv:1406.2661 [stat.ML] (2014).

[17] B. Hashemi, N. Amin, K. Datta, D. Olivito, and M. Pierini, "LHC analysis-specific datasets with Generative Adversarial Networks," arXiv:1901.05282 [hep-ex] (2019).

[18] Y. Alanazi *et al.*, "Machine learning-based event generator for electron-proton scattering," arXiv:2008.03151 [hep-ph] (2020).

[19] A. Niculescu-Mizil and R. Caruana, "Predicting good probabilities with supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning (ICML '05)* (ACM, New York, 2005), pp. 625–632.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 [cs.LG] (2014).

[21] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," arXiv:1411.1784 [cs.LG] (2014).

[22] S. Diefenbacher, E. Eren, G. Kasieczka, A. Korol, B. Nachman, and D. Shih, "DCTRGAN: Improving the Precision of Generative Models with Reweighting," JINST **15**, P11004 (2020).

[23] C. Chen, O. Cerri, T. Q. Nguyen, J. R. Vlimant, and M. Pierini, "Data Augmentation at the LHC through Analysis-specific Fast Simulation with Deep Learning," Comput. Softw. Big Sci. **5**, 15 (2021).

[24] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011).

[25] K. T. Matchev, A. Roman, and P. Shyamsundar, "Uncertainties associated with GAN-generated datasets in high energy physics," SciPost Phys. **12**, 104 (2022).

[26] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr, "Advanced event reweighting using multivariate analysis," J. Phys. Conf. Ser. **368**, 012028 (2012).