ELUQuant: Event-Level Uncertainty Quantification using Physics-Informed Bayesian Neural Networks with Flow approximated Posteriors - A DIS Study

Cristiano Fanelli, James Giroux
Department of Data Science
William & Mary
Williamsburg, VA 23185
{cfanelli,jgiroux}@wm.edu

Abstract

We present a Bayesian deep learning architecture with multiplicative normalizing flows for precise uncertainty quantification (UQ) at the physics event level. This method distinguishes both types of uncertainties, aleatoric and epistemic, offering nuanced insights. When applied to Deep Inelastic Scattering (DIS) events, the model extracts kinematic variables effectively, paralleling the efficacy of recent techniques, but with event-level UQ. This UQ proves essential for tasks like event filtering and can rectify errors without the ground truth. Tests using the H1 detector at HERA suggest potential applications for the future EIC, including data monitoring and anomaly detection. Notably, our model efficiently handles large samples with low inference time.

1 Introduction

In the realms of experimental nuclear (NP) and high-energy physics (HEP), data analyses predominantly involve regression of core quantities from observables acquired during experiments. Essential to this is event-level uncertainty quantification (UQ). Our method, ELUQuant (ELUQ), stands out as a trailblazer in this context within NP/HEP, integrating concepts from computer vision Louizos and Welling [2017] and Bayesian Neural Networks (BNN) with Multiplicative Normalizing Flows (MNF) Kendall and Gal [2017]. This allows the accurate capture of both heteroskedastic aleatoric and epistemic uncertainties influencing the regression process. Epistemic uncertainty generally refers to a systematic uncertainty characterized by the analysis tool. In our case, epistemic uncertainty refers directly to the uncertainty in our model. On the other hand, aleatoric uncertainty is the inherent uncertainty in the data (stochastic). While deep inelastic scattering (DIS) has already embraced deep learning advancements, previous works like Diefenthaler et al. [2022] and Arratia et al. [2022] primarily focused on leveraging deep neural networks (DNN) for kinematic variable inference but did not venture into UQ. Our work with ELUQ fills this gap, underscoring the importance and novelty of comprehensive event-level UQ in the context of the referenced studies. This approach is anticipated to be pivotal for any physics analysis requiring intricate UQ.

This paper is organized as: Sec. 2 outlines DIS kinematics and the types of uncertainties addressed; Sec. 3 explores the mechanics of ELUQ, including its architecture and performance metrics; Sec. 4 presents findings using the H1 neutral current DIS Monte Carlo dataset, also discussed in Arratia et al. [2022]. The concluding section, Sec. 5, ponders the broader implications of ELUQ, spotlighting its proficiency in UQ and potential wider applications.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

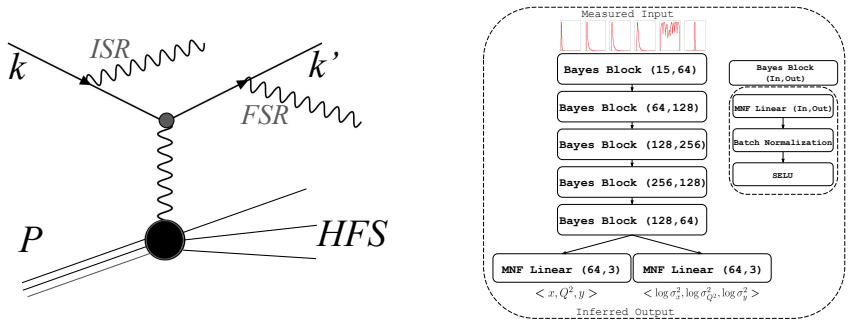

Figure 1: (left) Neutral current DIS diagram showcasing potential initial and final state radiation; additional effects may arise from higher-order QED and QCD corrections Diefenthaler et al. [2022], Arratia et al. [2022]; (right) ELUQuant: a dual-headed regression network using Bayesian blocks and MNF modules for event-level UQ.

2 Kinematic Reconstruction of DIS

Deep Inelastic Scattering DIS is a reaction used to probe the internal structure of hadrons. In this process, high-energy leptons are scattered off hadrons, revealing intricate details about quarks and gluons Devenish and Cooper-Sarkar [2004]. Pioneering DIS investigations were led by experiments at the HERA collider, the sole electron-proton collider ever built Abt et al. [1997], Abramowicz et al. [2015]. The upcoming Electron Ion Collider (EIC) Khalek et al. [2022] is poised to further our understanding by exploring new facets of the DIS kinematic domain. Fig. 1 (left) depicts the DIS process, where k, k' and P are the four-momenta of the electrons and proton, respectively, and HFS is the hadronic final state. Key kinematic terms for DIS are: squared four-momentum transfer Q^2, inelasticity y, and Bjorken scaling x, interconnected by $Q^2 = sxy$, where s denotes squared center-of-mass energy. Various methods, such as Electron (EL), Double Angle (DA), and Jacquet Blondel (JB), enable reconstruction of these kinematics from measurements, each having its own advantages and sensitivities Bassler and Bernardi [1995], Diefenthaler et al. [2022], Arratia et al. [2022]. The DIS process can be influenced by several factors, such as initial-state and final-state radiation (ISR, FSR). Moreover, higher order QED and QCD corrections can also manifest in the process. Hence, regression linking measured quantities to true kinematics is crucial. The true values of the DIS kinematic observables x, Q^2, and y in our data are derived from generator-level particle 4-vectors, considering effects like ISR and FSR radiation. Our work employs a synthetic dataset from the H1 experiment[1], consisting of 8.7×10^6 training, 1.9×10^6 validation, and 1.9×10^6 testing events. We incorporate 15 measured input features, outlined in Arratia et al. [2022], which includes 7 features sensitive to QED radiation and other 8 related to scattered electron and HFS features.

3 ELUQuant Architecture

In this work, ELUQ is applied to the H1's DIS simulated dataset, extracting kinematic variables x, Q^2, y and their respective uncertainties from 15 input features. Epistemic uncertainty improves with refined models, whereas aleatoric uncertainty typically remains consistent despite more data. ELUQ, as depicted in Fig. 1 (right), is a Bayesian network designed to form weight posteriors at each layer utilizing MNF Kendall and Gal [2017] linear layers. The MNF layers improve upon traditional BNN techniques, which involve directly optimizing over a weight distribution at each layer, by sampling neuron pre-activations to obtain different weights for each data point in a mini-batch. These pre-activations are linear combinations of Gaussian weights. Doing so provides lower variance in the gradients, resulting in a more stable and effective optimization. Kendall and Gal [2017] show their method results in improvements in both performance and reliability of predictive uncertainty. After training, each forward pass samples a unique set of weights from this posterior. The total loss function comprises various terms:

$$\mathcal{L}_{Tot.} = \mathcal{L}_{Reg.} + \alpha\mathcal{L}_{Phys.} + \beta\mathcal{L}_{MNF.} \tag{1}$$

The regression loss, Eq. 2, predicts the kinematic vector $\mathbf{v} = (x, Q^2, y)$ and its heteroskedastic aleatoric term. Its logarithmic formulation, based on a multivariate normal distribution, is numerically stable

[1]Data sourced from tutorial by Acosta and Mikuni [2022] at AI4EIC 2022, also used in Arratia et al. [2022].

Kendall and Gal [2017]:

$$\mathcal{L}_{Reg.} = \frac{1}{N}\sum_i \sum_j \frac{1}{2}(e^{-\mathbf{s_j}}\|\mathbf{v}_j - \hat{\mathbf{v}}_j\|^2 + \mathbf{s}_j),\ \mathbf{s}_j = \log \boldsymbol{\sigma}_j^2 \tag{2}$$

We adopt the SELU Klambauer et al. [2017] activation functions, known for their self-normalizing properties, in tandem with batch normalization to enhance convergence. The physics-informed term ensures that regressed observables align with truth values:

$$\mathcal{L}_{Phys.} = \frac{1}{N}\sum_i \log \hat{Q}_i^2 - (\log s_i + \log \hat{x}_i + \log \hat{y}_i), \tag{3}$$

where the Mandelstam s is calculated at the ground truth level. The Kullback Leibler term, inspired by Louizos and Welling [2017], refines the posterior approximation using Gaussian priors and a mixing density:

$$\begin{aligned}\mathcal{L}_{MNF.} &= -KL(q(\mathbf{W})\|p(\mathbf{W})) \\ &= \mathbb{E}_{q(\mathbf{W},\mathbf{z}_T)}[-KL(q(\mathbf{W}|\mathbf{z}_{T_f})\|p(\mathbf{W})) + \log r(\mathbf{z}_{T_f}|\mathbf{W}) - \log q(\mathbf{z}_{T_f})]\end{aligned} \tag{4}$$

Training and Inference Specifications The ELUQuant model was trained using PyTorch 1.12.1 in a Python 3.9.12 environment with CUDA 11.3. The training protocol employed the *Adam* optimizer, an initial learning rate of $5e^{-4}$, and a stepped decay rate (γ) of 0.1 every 50 epochs, for a maximum of 100 epochs, using a batch size of 1028. Through grid search, optimal values of $\alpha = 1.0$ (for physics loss scaling) and $\beta = 0.01$ (for KL scale) were determined. The total dataset comprised $\sim$ 12M events, partitioned into a standard $70\%, 15\%, 15\%$ for training, validation, and testing—equating to $\sim$ 2M test events. Inputs and targets were scaled between (-1,1). Table 1 summarizes the specs for training.

Training Parameter	**value**
Max Epochs	100
Batch Size	1024
Learning Rate	$5e^{-4}$
Decay Steps	50
Decay Factor (γ)	0.1
Physics Loss Scale (α)	1.0
KL Scale (β)	0.01
Training GPU Memory	$\sim$ 1GB
Network memory on local storage	$\sim$ 7MB
Trainable parameters	611,247
Wall Time	$\sim$ 1 Day

Table 1: Training Specs of the ELUQuant architecture: training was performed with an Intel i7-12700k 12 Core CPU, Nvidia RTX 3090 24GB GPU, and 64GB memory.

During inference, each event was sampled 10k times, with the final prediction being the mean value. The epistemic uncertainty component is given by the standard deviation in these predictions. Each individual inference will also provide the corresponding aleatoric uncertainty, in which we again take the mean as our final aleatoric component. For inference, the architecture processed batches of 100 samples, equating to a total size of 1 million. Each inference consumed around 24GB of GPU memory, with an inference time per event approximately equating to 20ms. The model's training utilized 1 GB of GPU memory and took $\sim$24 hours, while the trained architecture occupied 7MB, consisting of 611,247 parameters. The system specs included an Intel i7-12700k CPU, Nvidia RTX 3090 GPU, and 64GB RAM.

Limitations Selecting an appropriate architecture for a BNN is challenging, as it requires optimizing a weight distribution for each layer. Moreover, the optimization of a BNN becomes increasingly harder with an increased number of parameters due to inherent stochasticity in the training process. To mitigate this, we adopted the structure from its deterministic counterpart (a DNN) that achieves satisfactory performance, and integrated it into ELUQ. Specifically, the pipeline for designing ELUQ-style networks should first involve minimizing the complexity of the deterministic counterpart. This

y bin	RMS(x_{DA})	RMS(x_{ele})	RMS(x_{DNN})	σ(x)	RMS(Q^2_{DA})	RMS(Q^2_{ele})	RMS(Q^2_{DNN})	$\sigma(Q^2)$	RMS(y_{DA})	RMS(y_{ele})	RMS(y_{DNN})	σ(y)
(0.5, 0.8)	0.15	0.079	0.062	**0.058**	0.095	0.057	0.044	**0.041**	0.061	0.041	0.031	**0.035**
(0.2, 0.5)	0.13	0.14	0.075	**0.062**	0.068	0.056	0.038	**0.032**	0.082	0.100	0.053	**0.044**
(0.1, 0.2)	0.15	0.25	0.098	**0.071**	0.060	0.054	0.033	**0.030**	0.099	0.18	0.078	**0.062**
(0.05, 0.1)	0.18	0.36	0.13	**0.083**	0.059	0.053	0.033	**0.029**	0.13	0.25	0.11	**0.078**
(0.01, 0.05)	0.25	0.43	0.18	**0.12**	0.059	0.053	0.032	**0.029**	0.16	0.28	0.15	**0.12**

Table 2: Comparisons between the aleatoric uncertainty of ELUQ with the RMS of other methods, for the DIS kinematic variables x, Q^2, y.

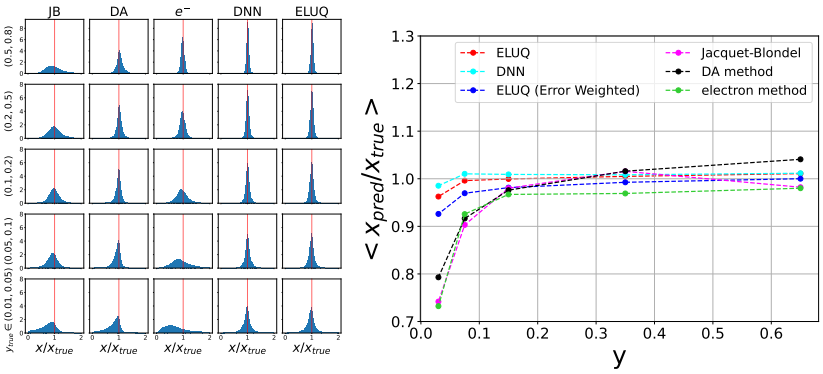

Figure 2: (left) Resolutions for x in bins of y and comparison among reconstruction methods; (right) The average predicted-to-truth ratio of the DIS kinematic variables for ELUQ and DNN compared to the classical methods, in bins of y. ELUQ and DNN outperform the other methods.

can allow easier generalization of the network to other datasets. An initial training caveat is that a misconfigured model might yield poor regression targets with small aleatoric components, causing erratic fluctuations. To enhance training stability, we imposed bounds on the minimum and maximum variances. These bounds were chosen to ensure they don't bias the learned aleatoric component, preventing the network from defaulting to these extremities.

4 Analysis Results

We trained a simplified DNN, achieving comparable performance to Arratia et al. [2022] despite having fewer parameters (150k vs 1.2M). While both ELUQ and the DNN have similar architecture, ELUQ offers advanced UQ not feasible with a basic DNN. Using ELUQ, we predicted the DIS kinematic variables along with their aleatoric and epistemic uncertainties. Comparisons were made with traditional methods like EL, JB, and DA, as described in Sec. 2. We incorporated DNN into our visual comparisons. Fig. 2 (left) displays the resolutions for x (similar plots can be produced for Q^2 and y), showing similarities between DNN and ELUQ over all y ranges. These methods outperform traditional ones, which can have varying performances based on y. With ELUQ, we determine event-level uncertainties. The weighted average of an observable $\hat{O}_k$ over the dataset is $\sum_{k=1}^{N} \frac{\hat{O}_k}{\sigma_k^2} / \sum_{k=1}^{N} \frac{1}{\sigma_k^2}$ and its uncertainty is $1/\sqrt{\sum_{k=1}^{N} \frac{1}{\sigma_k^2}}$. Other methods do not offer direct event-level uncertainty, but we estimated it using RMS as in Arratia et al. [2022]. Fig. 2 (right) depicts the average ratio of predicted to true observables in bins of y. There's a noticeable drop in the $< R_x >$ at low y, which may be due to detector effects that deteriorates the measurement of the HFS observables, see Arratia et al. [2022]. The total uncertainty in ELUQ is $\sigma_{tot} = \sqrt{\sigma_{ale}^2 + \sigma_{epi}^2}$. Fig. 3 (left) compares σ_{ale}, σ_{epi}, and the DNN's RMS. Furthermore, numerical closure tests validated the aleatoric uncertainties obtained: as depicted in Table 2, the aleatoric uncertainties coincide with RMS when inaccuracy is negligible and distributions are approximately Gaussian. The epistemic component is validated via increases with larger inaccuracy as shown in Fig. 3 (middle). Other studies also showed that the physics-informed term contributed to reduce the Q^2 inaccuracy compared to the case trained without this term. Event-level UQ can control data quality. Fig. 3 (right) demonstrates the effect of excluding events with significant relative uncertainty. This improves the ratio of predicted-to-true for x, but reduces overall statistics ($\gtrsim$40% in the first bin in y).

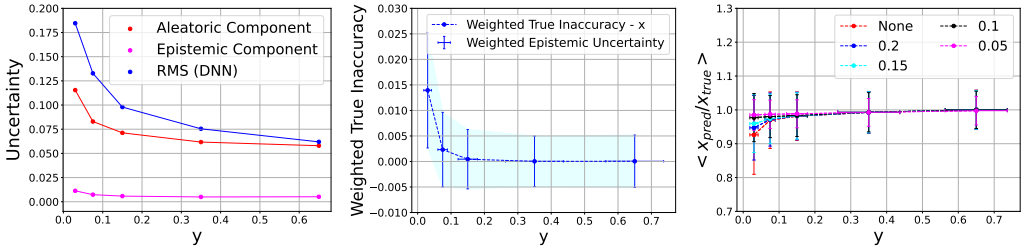

Figure 3: (left) Average event-level aleatoric and epistemic uncertainty on the reconstruction-to-true ratio of x for ELUQ, compared to the RMS obtained from the DNN model; (middle) Average event-level epistemic vs weighted true inaccuracy for the DIS kinematics in bins of y. The epistemic increases with the inaccuracy; (right) Weighted average predicted-to-true ratio in bins of y for ELUQ after applying cuts at various thresholds on the relative uncertainty of x, Q^2, y to reject events.

5 Future Directions an Broader Impacts

ELUQuant, combining Physics-Informed Bayesian Neural Networks with Flow-approximated Posterior, offers a novel approach to analyzing heteroskedastic aleatoric and epistemic uncertainties event-wise. Validated with H1 detector's DIS data from HERA, its potential extends to the upcoming EIC, highlighting its capability in examining kinematic observables possibly influenced by radiation effects. Closure tests confirmed the reliable extraction of both aleatoric and epistemic uncertainties. ELUQuant's adaptability holds promise for a wider spectrum of event-level physics analyses, with the potential to pinpoint and mitigate inaccuracies, suggesting notable utility in data quality monitoring and anomaly detection. ELUQuant offers adaptability for diverse event-level physics analyses, showing potential in mitigating inaccuracies, suggesting notable utility in data quality monitoring and anomaly detection. Its performance—10,000 samples per event in just 20 ms on an RTX 3090—reiterates its significance in real-world applications. ELUQuant sets a paradigm shift in uncertainty quantification (UQ) for nuclear and particle physics, enabling in-depth, event-level analyses.

Acknowledgments

We thank the H1 Collaboration for allowing us to use the simulated MC event samples.

References

Halina Abramowicz et al. Combination of measurements of inclusive deep inelastic eˆ ± p e±p scattering cross sections and QCD analysis of HERA data: H1 and ZEUS Collaborations. *The European Physical Journal C*, 75:1–98, 2015. doi: 10.1140/epjc/s10052-015-3710-4.

I. Abt et al. The H1 detector at HERA. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 386(2):310–347, 1997. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(96)00893-5. URL `https://www.sciencedirect.com/science/article/pii/S0168900296008935`.

F. Torales Acosta and V. Mikuni. Unfolding with ML : OmniFold, 2022. URL `https://indico.bnl.gov/event/16586/page/426-tutorials`.

Miguel Arratia, Daniel Britzger, Owen Long, and Benjamin Nachman. Reconstructing the kinematics of deep inelastic scattering with deep learning. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1025: 166164, 2022. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2021.166164. URL `https://www.sciencedirect.com/science/article/pii/S0168900221010445`.

Ursula Bassler and Gregorio Bernardi. On the kinematic reconstruction of deep inelastic scattering at HERA. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 361(1-2):197–208, 1995. doi: https://doi.org/10.1016/0168-9002(95)00173-5.

Robin Devenish and Amanda Cooper-Sarkar. *Deep inelastic scattering*. OUP Oxford, 2004. doi: 10.1093/acprof:oso/9780198506713.001.0001.

Markus Diefenthaler, Abdullah Farhat, Andrii Verbytskyi, and Yuesheng Xu. Deeply learning deep inelastic scattering kinematics. *The European Physical Journal C*, 82(11):1064, 2022. doi: https://doi.org/10.1140/epjc/s10052-022-10964-z.

Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf.

R Abdul Khalek et al. Science requirements and detector concepts for the electron-ion collider: EIC yellow report. *Nuclear Physics A*, 1026:122447, 2022. doi: 10.1016/j.nuclphysa.2022.122447.

Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/5d44ee6f2c3f71b73125876103c8f6c4-Paper.pdf.

Christos Louizos and Max Welling. Multiplicative Normalizing Flows for Variational Bayesian Neural Networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2218–2227. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/louizos17a.html.