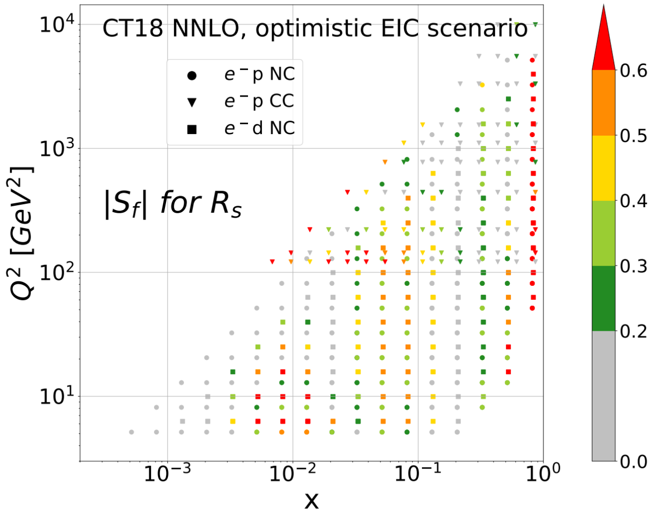
Artificial Intelligence in the EIC era at the BSM-PDF frontier

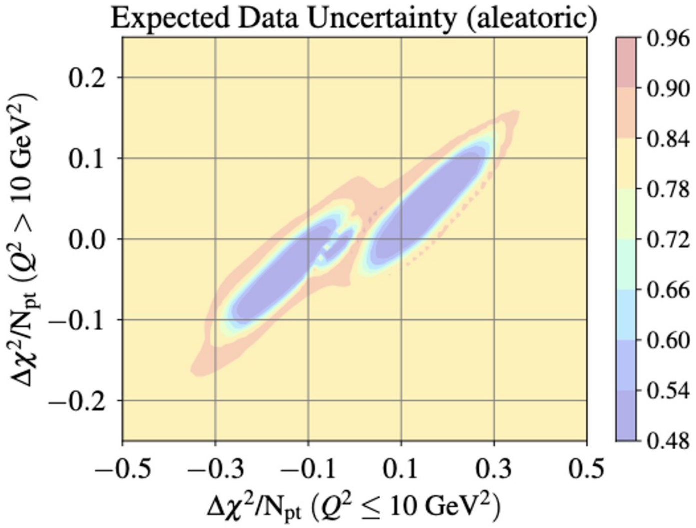
Tim Hobbs

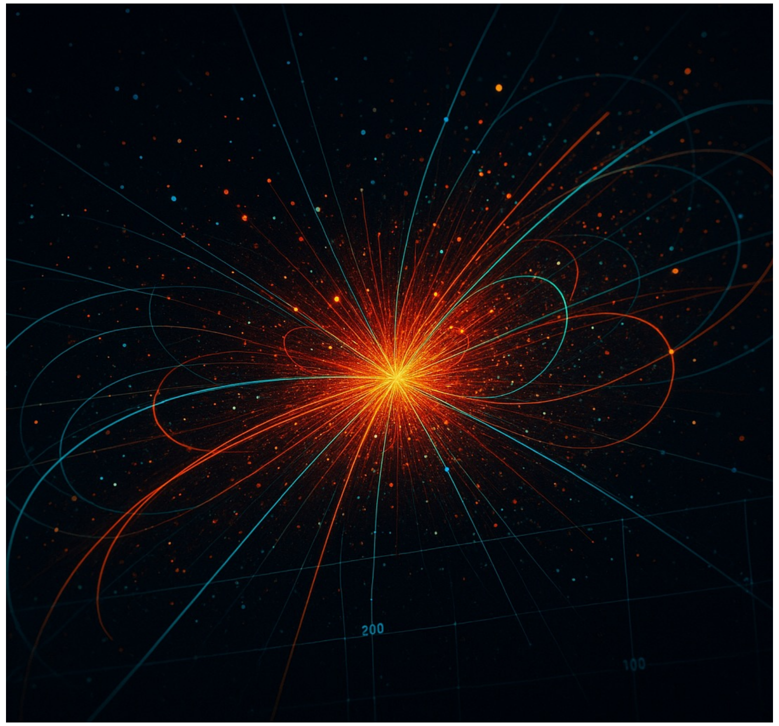
Argonne National Laboratory, USA

28 October 2025

Artificial Intelligence for the EIC





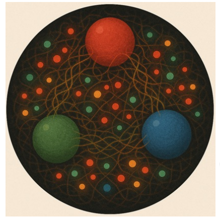


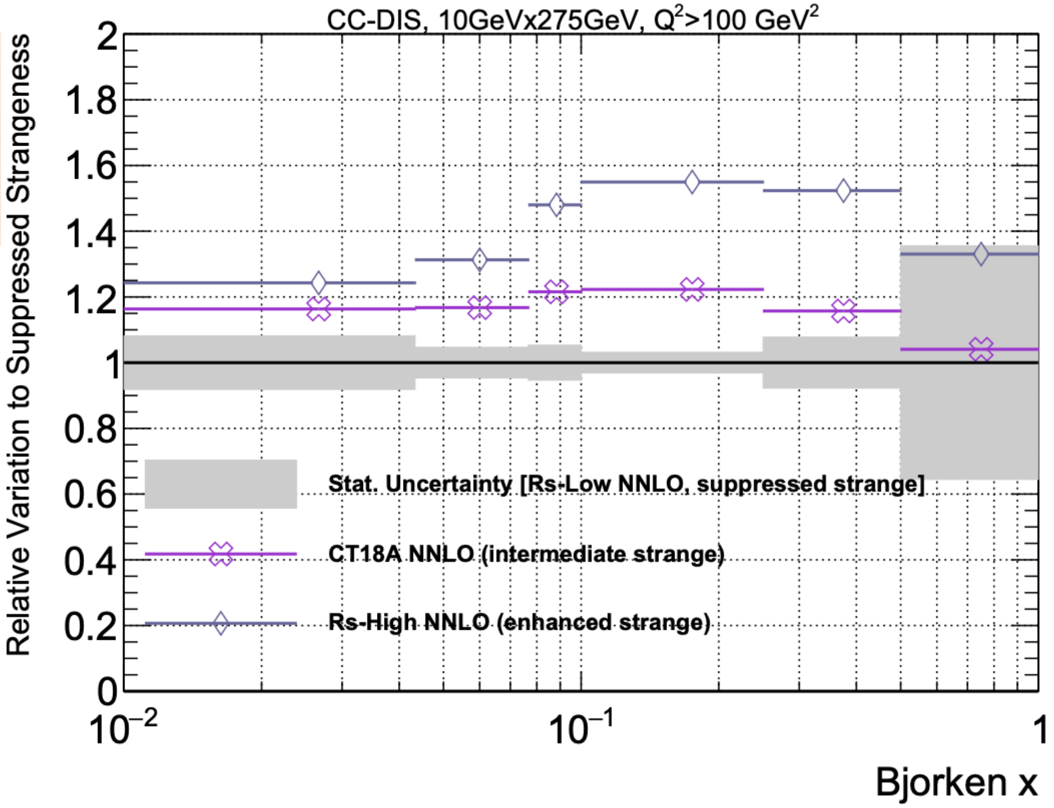
HEP Division, Particle Theory

with: **Brandon Kriesten**; Adriana Baniecki, Jonathan Gomprecht

EIC physics: must navigate **precision QCD**, **EW/BSM**, **big data**

- **HEP quandary**: tests of SM (BSM searches), involve [large] data sets from overlapping hadronic experiments
- **EIC program** (including CC DIS): novel probes of flavor currents (left) *AND* possible constraining power over BSM scenarios (right)

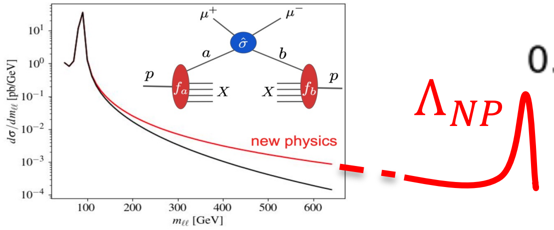


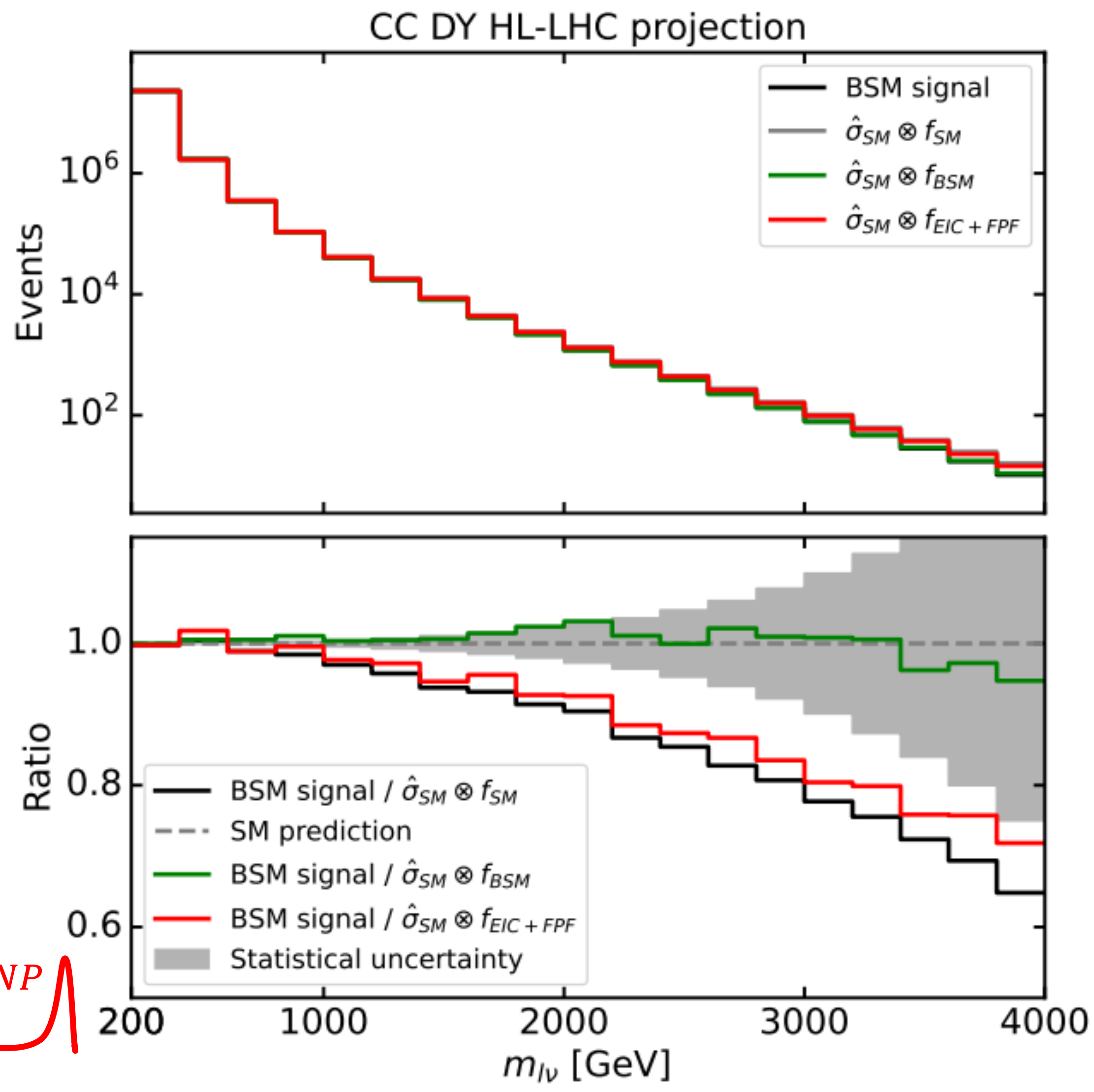


Arratia, Furletova TJH, Olness, Sekula; PRD103 (2021) 7, 074023

AI/ML: essential role in model exploration

Das Bakshi, TJH, Kriesten; in prep.





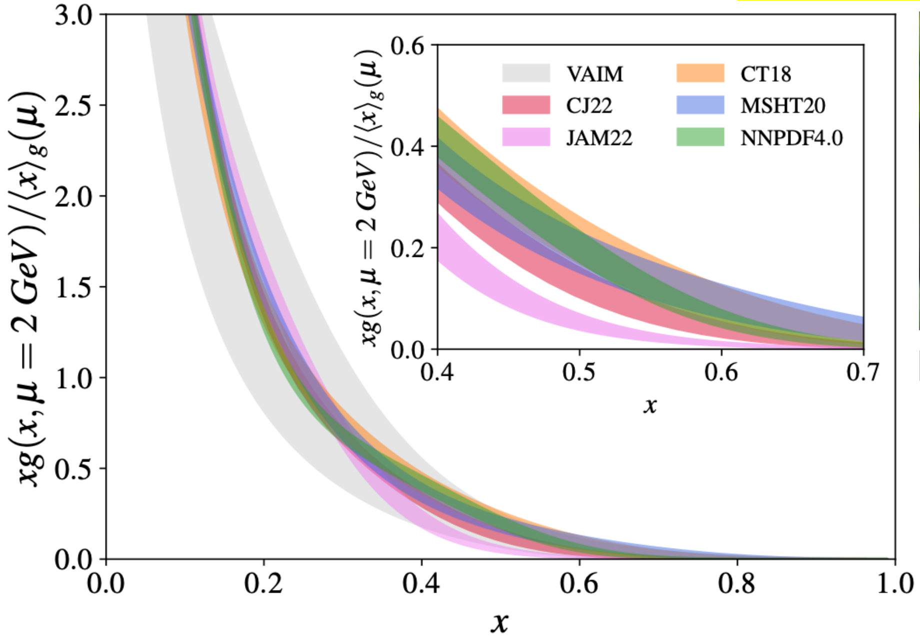
Hammou and Ubiali; PRD111 (2025) 9, 095028

broader context: ML for PDF and related theory

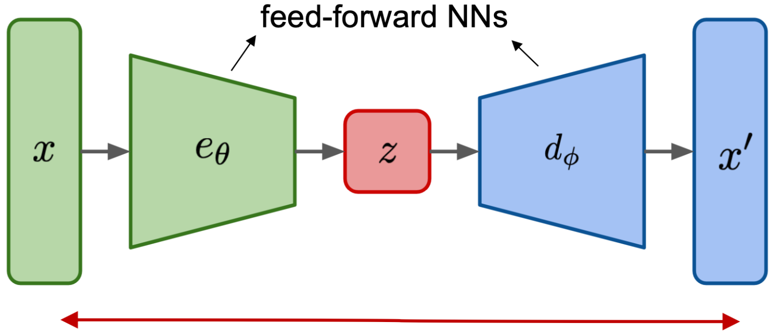
- ❑ ML models can learn/emulate PDFs; this talk extends to related HEP theory --- **distinguishing BSM scenarios**

- reconstruct PDFs from Mellin moments; RpITDs

Kriesten, NieMiera, Good, TJH, Lin; 2507.17810



Brandon Kriesten, ANL HEP Theory

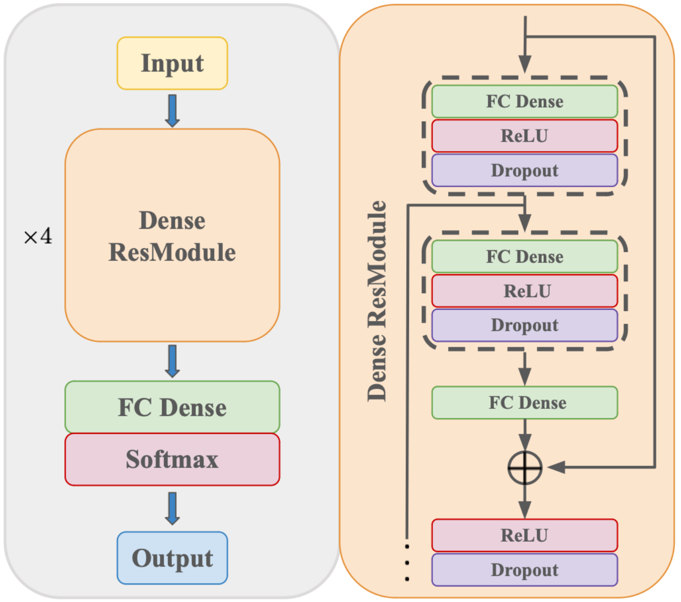


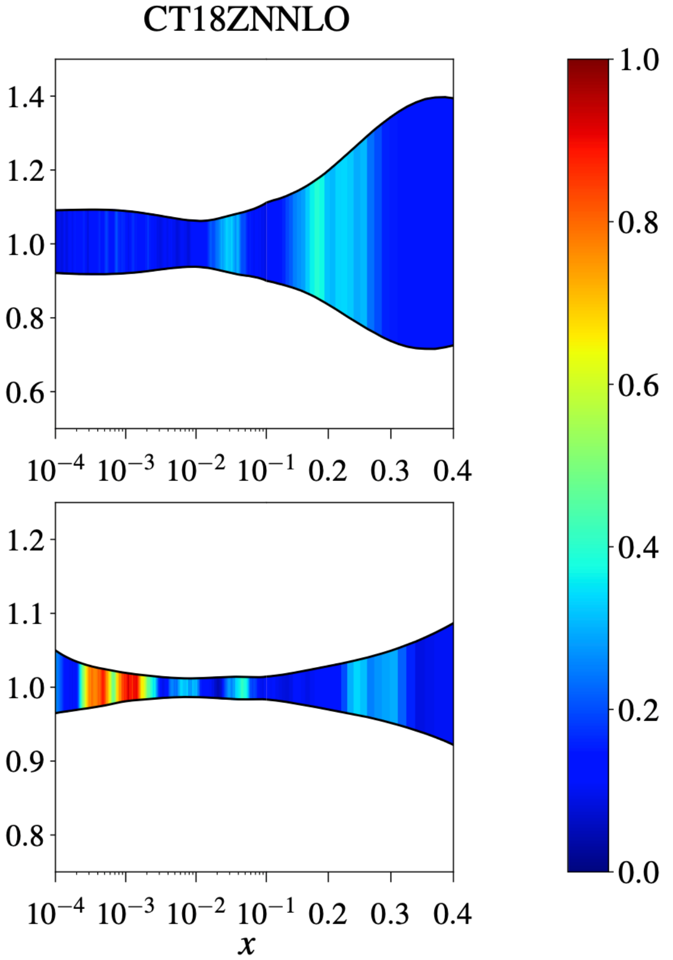
$$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2$$

Kriesten, TJH, PRD111 (2025) 1, 014028

- guided backprop: link *x*-dependent PDF features to theory assumptions in parent global fits

Kriesten, Gomprecht, TJH, JHEP11 (2024) 007





`PDFdecoder`

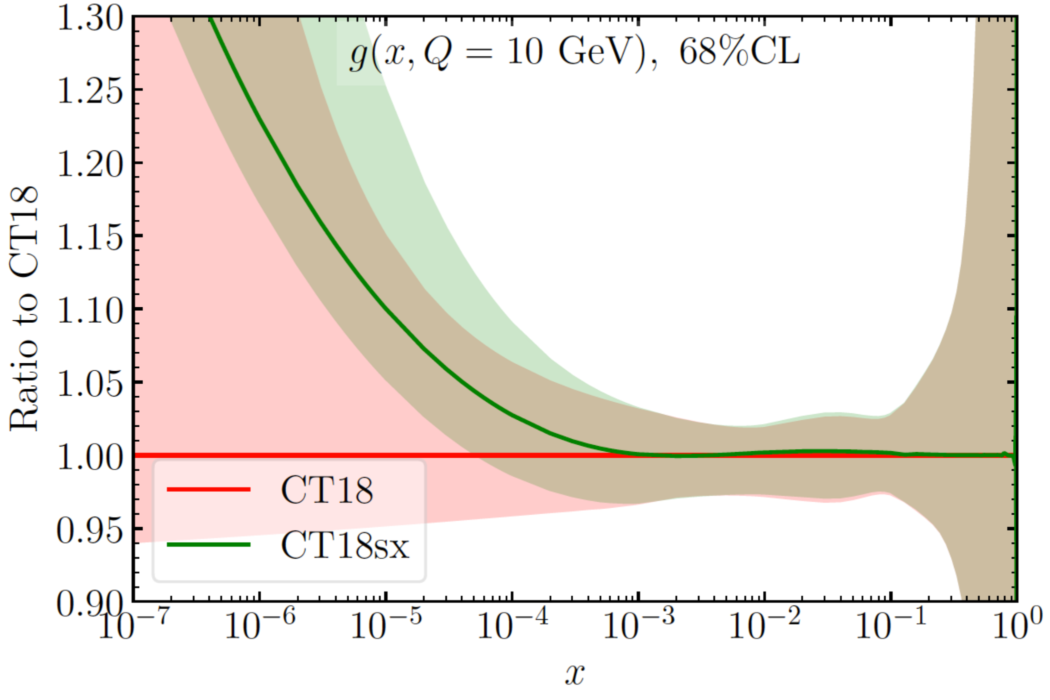
`XAIPDF` *(packages available)*

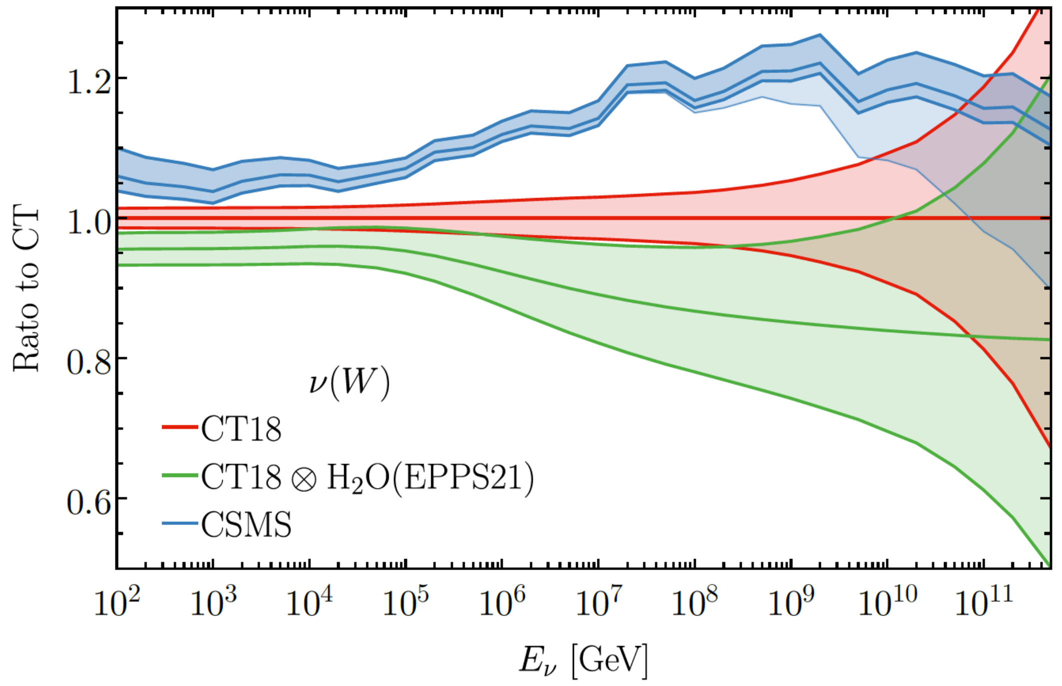
<u>this talk</u>: CC DIS (**neutrino scattering**) demonstration

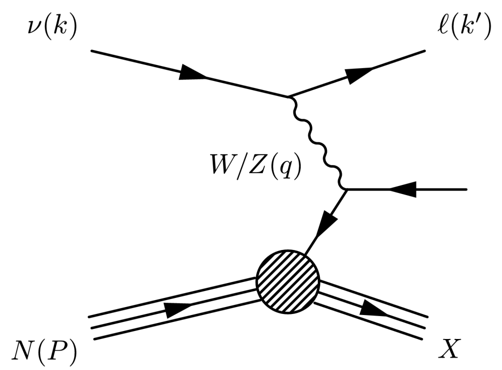
- neutrino DIS has a rich phenomenology from high to low energies; **complementary to CC DIS at EIC**

$$(\text{EIC: } e+p \to \nu_e + X)$$

$$\frac{d^2\sigma^{\nu(\bar{\nu})}}{dxdy} = \frac{G_F^2 m_N E_\nu}{\pi(1+Q^2/M_{W,Z}^2)^2}\left[\frac{y^2}{2}2xF_1 + \left(1-y-\frac{m_N xy}{2E_\nu}\right)F_2 \pm y\left(1-\frac{y}{2}\right)xF_3\right]$$







[neutrino DIS]

[Xie, Gao, TJH, Stump, Yuan; PRD109, 113001]

- recent calculations (*e.g.*, above) have focused on high-energy scattering; probe extra-dimensional theories
- potential sensitivity to BSM relevant at lower (GeV) energies as well;   (experiments: CDHSW, NuTeV, DUNE, …)

neutrino DIS involves interplay of EW interaction and QCD structure

- BSM at large scales can be parametrized by EFTs like SMEFT, matched to a **lower-energy EFT** (WEFT):
  - resulting nonstandard interaction might simply appear as shifted parton-level neutrino couplings

$$\mathcal{L}_{\mathrm{WEFT}} \supset -\frac{2V_{ij}}{v^2}\left\{\left[1+\epsilon_L^{ij}\right]_{ab}\left(\bar{u}^i\gamma^\mu P_L d^j\right)\left(\bar{\ell}_a\gamma_\mu P_L \nu_b\right) + \mathrm{h.c.}\right\}$$

$$V_{ij} \to V'_{ij} = \Delta_{ij}\cdot V_{ij}$$

- EW shifts (anomalous EW interactions, AEWIs) can become entangled with PDF dependence in DIS

$$F_2^{W^+}(x,Q^2) = 2x\left\{\sum_{j=d,s,b}|V'_{uj}|^2\bar{u}(x,Q^2) + \sum_{i=u,c,t}|V'_{id}|^2 d(x,Q^2) + \cdots\right\}$$

  - can an ML classifier distinguish among scenarios producing small (sub-percent) structure function shifts?

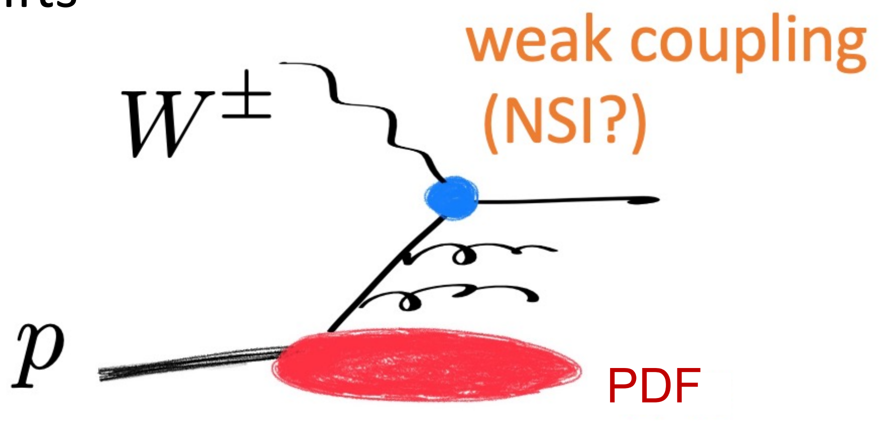
generate simple anomalous electroweak interaction scenarios

- motivation: distinguishing BSM signatures from PDF-related effects remains challenging

Hammou et al., JHEP11 (2023) 090
Gao et al., JHEP05 (2023) 003

- in lieu of a specific, UV-complete BSM model, generate randomized CKM shifts

$$V'_{ij} = \begin{bmatrix} \Delta_{ud} \cdot V_{ud} & \Delta_{us} \cdot V_{us} & \Delta_{ub} \cdot V_{ub} \\ \Delta_{cd} \cdot V_{cd} & \Delta_{cs} \cdot V_{cs} & \Delta_{cb} \cdot V_{cb} \\ \Delta_{td} \cdot V_{td} & \Delta_{ts} \cdot V_{ts} & \Delta_{tb} \cdot V_{tb} \end{bmatrix}$$



  - for given scenario of the EW couplings, compute ensemble of structure function predictions from CT18 MC replicas

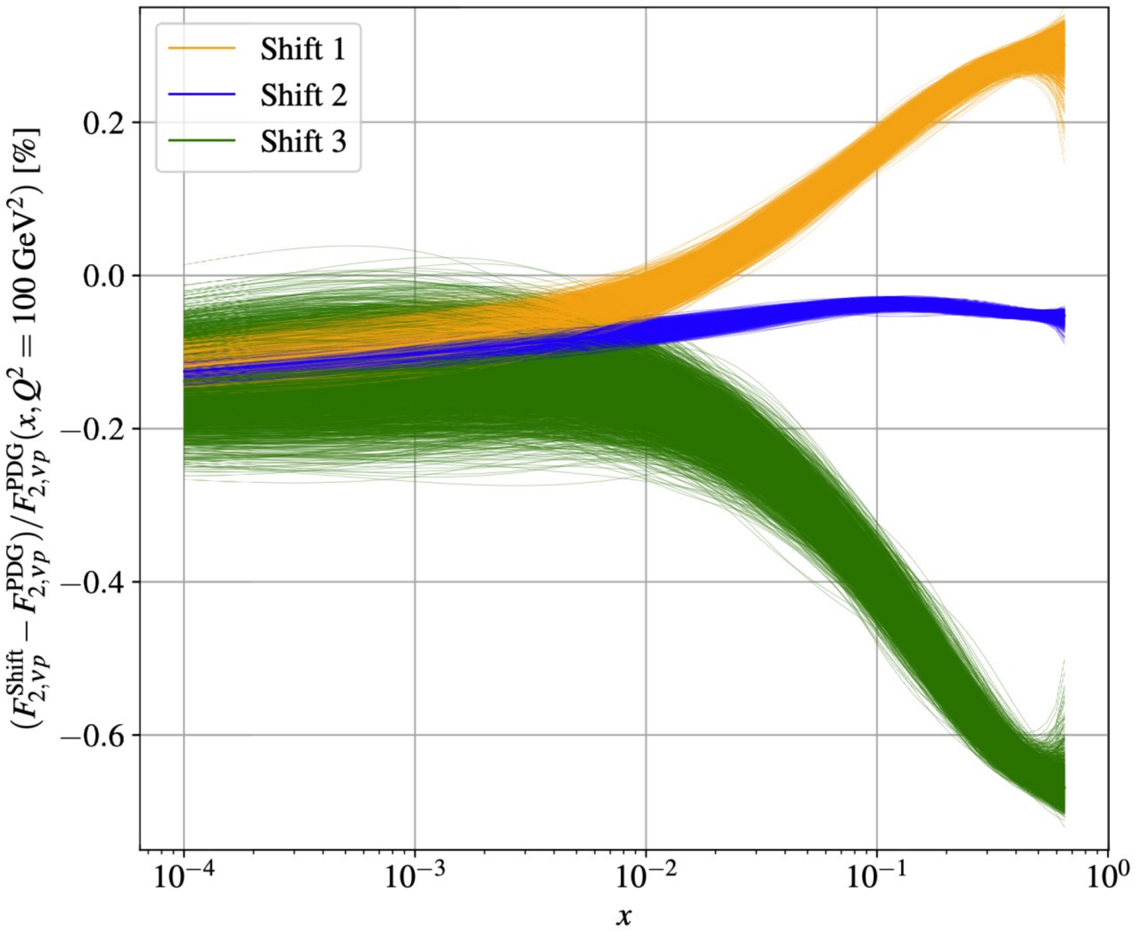
`[yadism]`

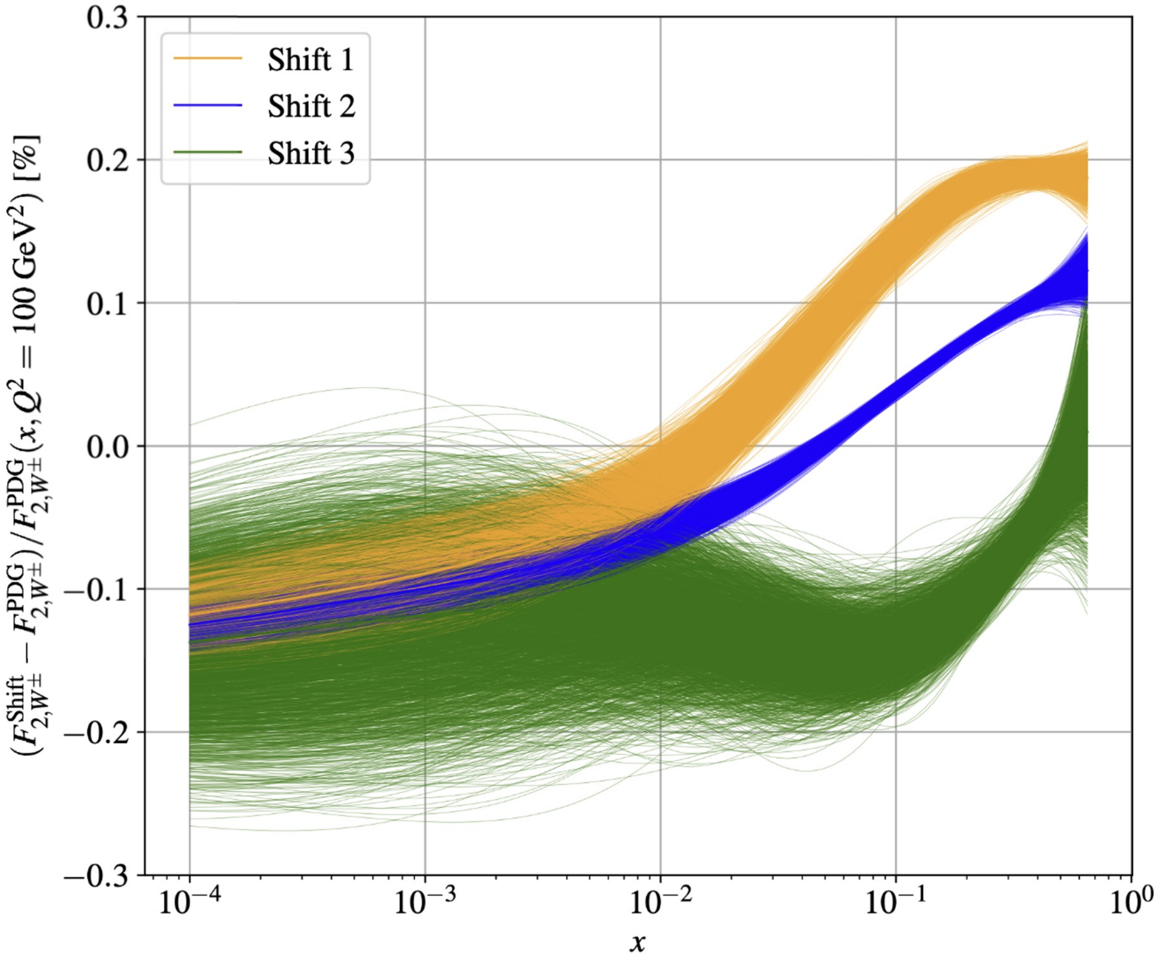
- curate randomized CKM shifts to produce ~multi-sigma 'discovery-level' shifts
  - can we distinguish these effects through evidential learning techniques (ML)?

shifts in electroweak couplings are imprinted on high-x behavior of structure functions

$$F_2^{W^+}(x,Q^2) = 2x\left\{\sum_{j=d,s,b}|V'_{uj}|^2\bar{u}(x,Q^2) + \sum_{i=u,c,t}|V'_{id}|^2 d(x,Q^2) + \cdots\right\}$$

$$F_2^{W^\pm}(x,Q^2) \equiv \frac{1}{2}\left(F_2^{W^+}(x,Q^2) + F_2^{W^-}(x,Q^2)\right)$$





- EW coupling shifts yield ~percent-level spread in neutrino-nucleon (proton) charge-current SFs [left]
- reduced in isoscalar combinations relevant for fixed-target experiments, *e.g.*, CDHSW [right]

❑ how might these differences be learned? What are the associated uncertainties?

ML models might distinguish among BSM effects above as a classification task

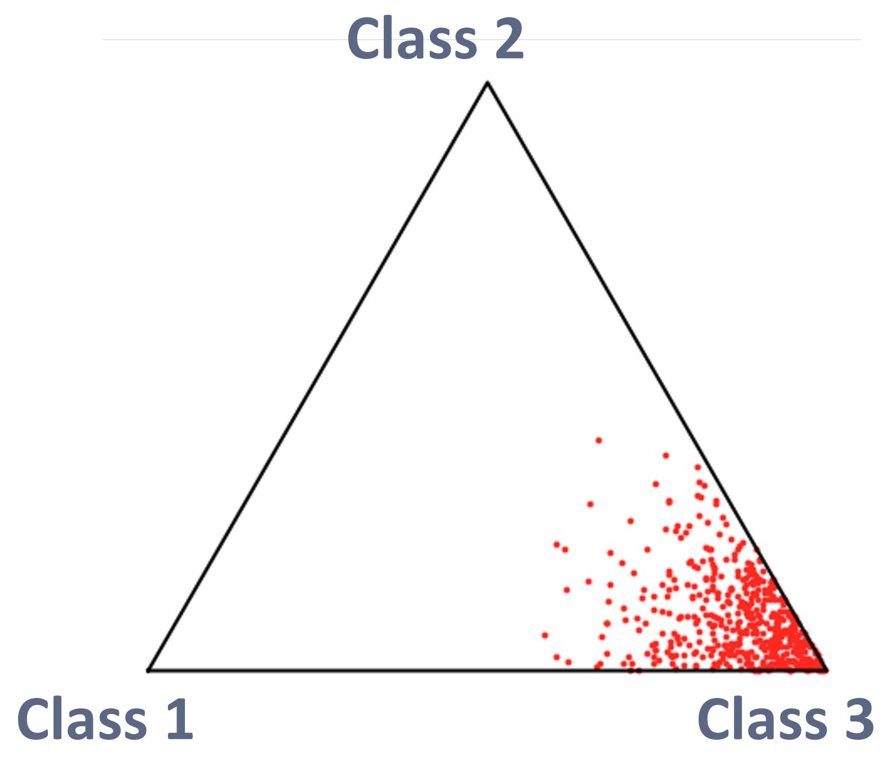
Bayesian Neural Networks

$$p(w_c \mid x^*, \mathcal{D}) = \int p(w_c \mid x^*, \theta) p(\theta \mid \mathcal{D})$$

What if instead we predicted the parameters of the ensemble --- the conjugate prior to the categorical distribution … a Dirichlet!

By Monte Carlo sampling the model parameters, we create an ensemble of categorical distributions

$$\left\{ p(w_c \mid x^*, \theta^{(i)}) \right\}_{i=1}^{M}$$



Computationally expensive! Alternative approaches?

Dirichlet Prior Networks (DPNs)

- categorical distributions model individual probabilities of specific BSM scenarios; Dirichlet distribution models prior beliefs over each scenario probability in DPN parameters

$$\text{BSM/scenario probabilities: } (p_1, p_2, \cdots, p_k)$$

$$\text{DPN parameters: } (\alpha_1, \alpha_2, \cdots, \alpha_k)$$

$$p(\mu;\alpha) = \frac{\Gamma(\alpha_0)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \prod_{k=1}^{K} p_k^{\alpha_k - 1} \qquad \alpha_0 = \sum_{k=1}^{K} \alpha_k$$

- By exponentiating our model outputs to the Dirichlet parameters, we recover the Softmax function for training a multi-class classification scheme:

$$\alpha_k = e^{f_k(x^*, \hat{\theta})} \qquad \mathbb{E}[p(\mu;\alpha)] = \frac{\alpha_k}{\alpha_0} = \left[\frac{e^{f_{k=1}(x^*,\hat{\theta})}}{\sum_{k=1}^{K} e^{f_k(x^*,\hat{\theta})}}, \cdots, \frac{e^{f_{k=K}(x^*,\hat{\theta})}}{\sum_{k=1}^{K} e^{f_k(x^*,\hat{\theta})}} \right]$$

uncertainty decomposition(s) in evidential learning

Malinin and Gales,1802.10501

Through training with maximum likelihood estimation (MLE) we can rework the expression to factorize into two distinct components. One is epistemic --- the KL divergence term; and the other is aleatoric --- the entropy term.

$$\mathbb{E}_{p_{\mathrm{true}}(x,y)}\left[\mathcal{L}^{\mathrm{NLL}}(y,x,\theta)\right] = \mathbb{E}_{p_{\mathrm{true}}(x)}\left[D_{\mathrm{KL}}\left(p_{\mathrm{true}}(y|x)\,\|\,p(y|x,\theta)\right)\right] - \mathbb{H}\left(p_{\mathrm{true}}(y|x)\right)$$

Epistemic

reducible by training procedures or algorithmic development; can be influenced by choice of distribution to model the data.

Aleatoric

not reducible because it is related to the true underlying data distribution.

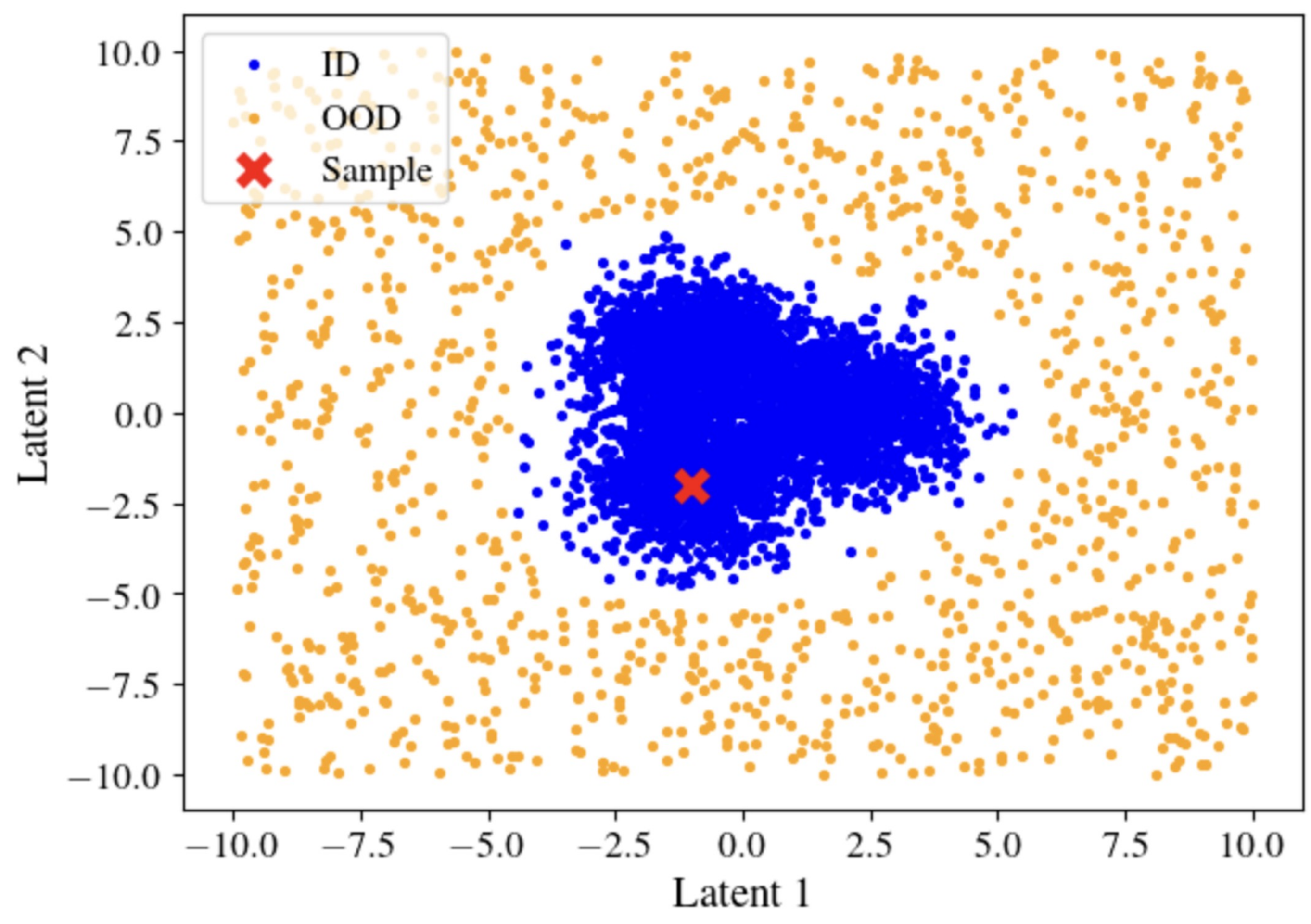
dissection of contributions to ML model uncertainties; **large ecology of errors**, **mapping to PDF/BSM uncertainty sources**

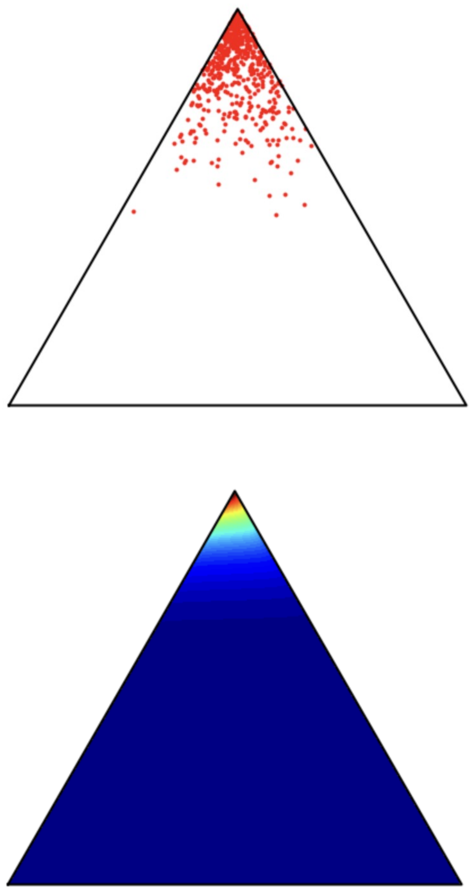
$$p(w_c \mid x^*, \mathcal{D}) = \int\int p(w_c \mid \mu)\, p(\mu \mid x^*, \theta)\, p(\theta \mid \mathcal{D})\, d\theta\, d\mu$$

Kotz, Courtoy, Nadolsky, Olness, Ponce-Chavez: 2311.08447

Dirichlet Prior Networks (DPNs) in practice (*i*)

- ❑ as a concrete example, demonstrate evidential learning on a 3-class toy problem

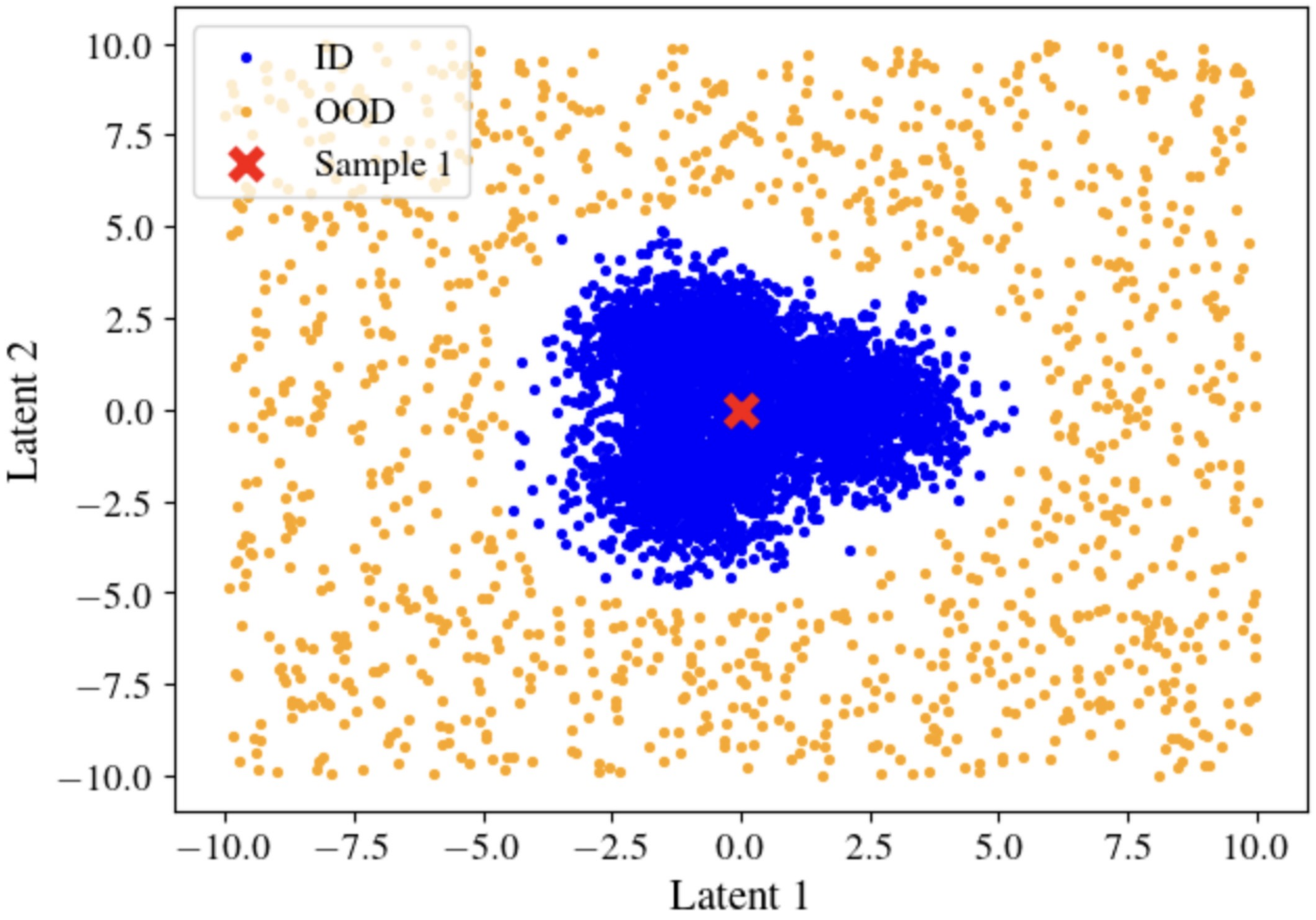


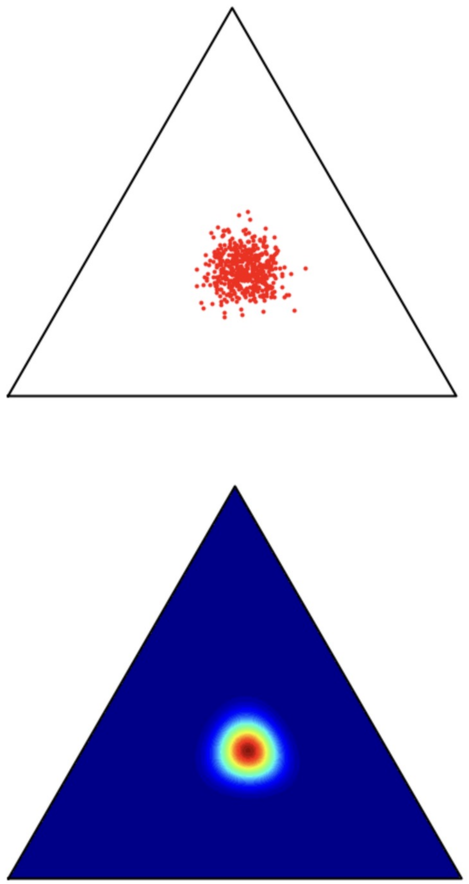


in-distribution sampling with low data uncertainty (samples are located at a specific corner) and low knowledge uncertainty (high sample density).

Dirichlet Prior Networks (DPNs) in practice (*ii*)

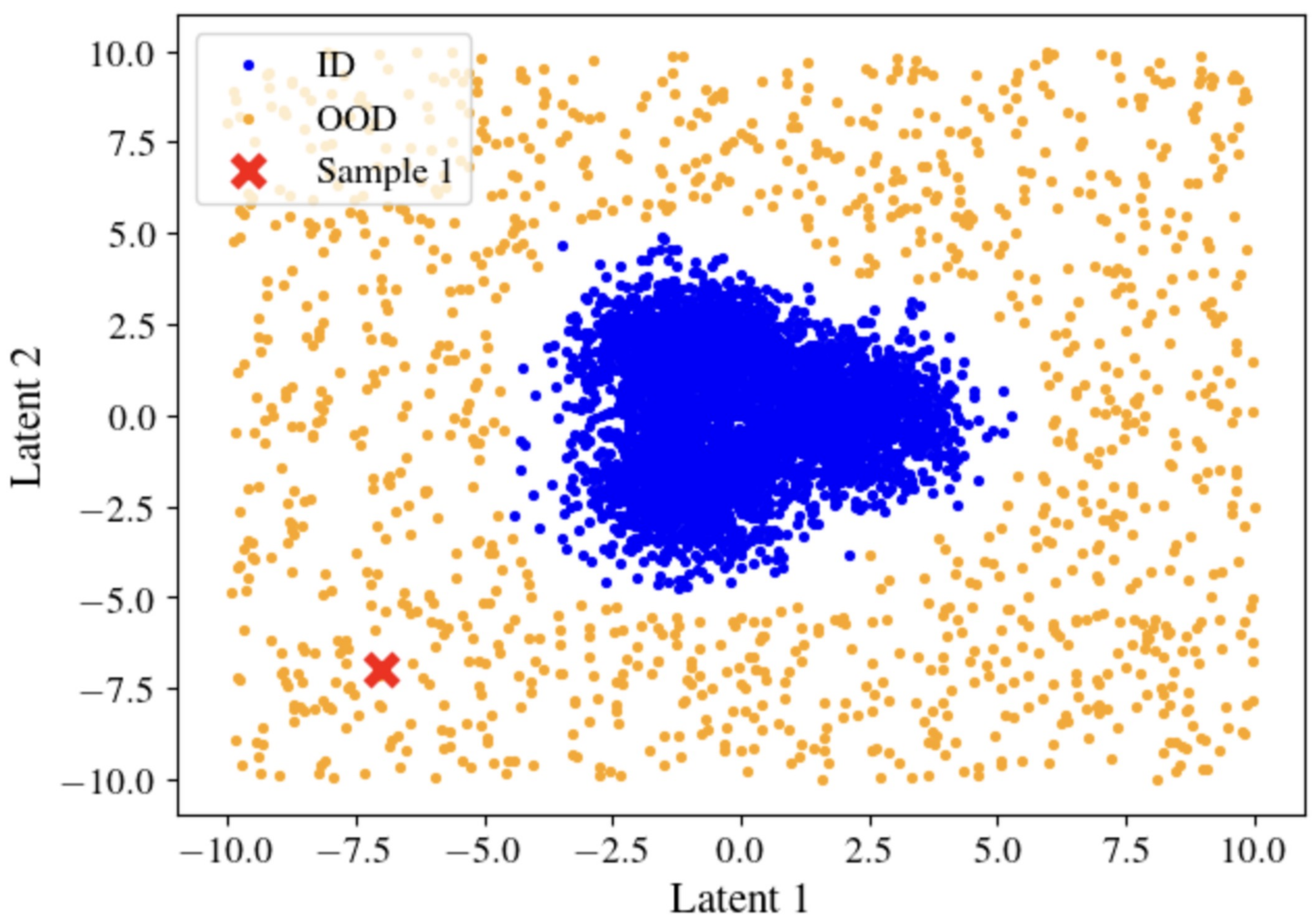
- ❑ as a concrete example, demonstrate evidential learning on a 3-class toy problem

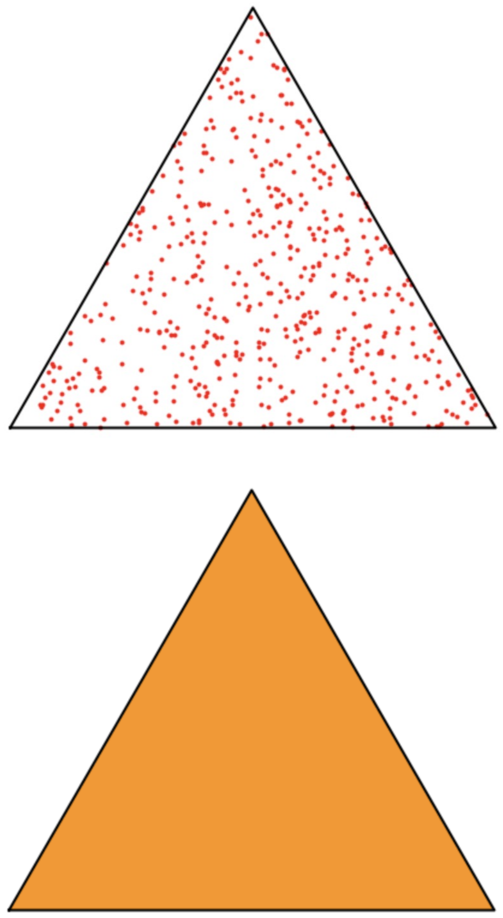




In-distribution sampling with high data uncertainty (samples are squeezed to the center of the simplex) with low knowledge uncertainty (high sample density).

- as a concrete example, demonstrate evidential learning on a 3-class toy problem





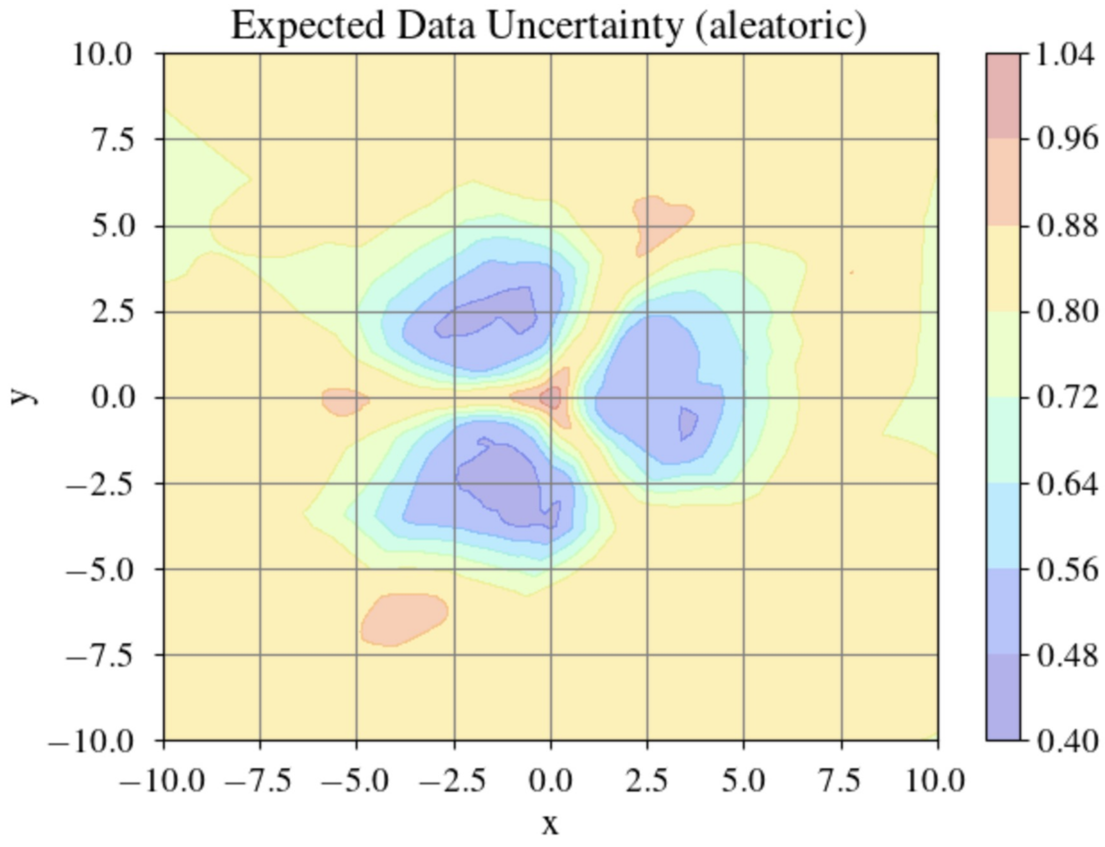
Out-of-distribution sampling with high data uncertainty (samples are not located at a specific corner) and high knowledge uncertainty (samples are diffuse).

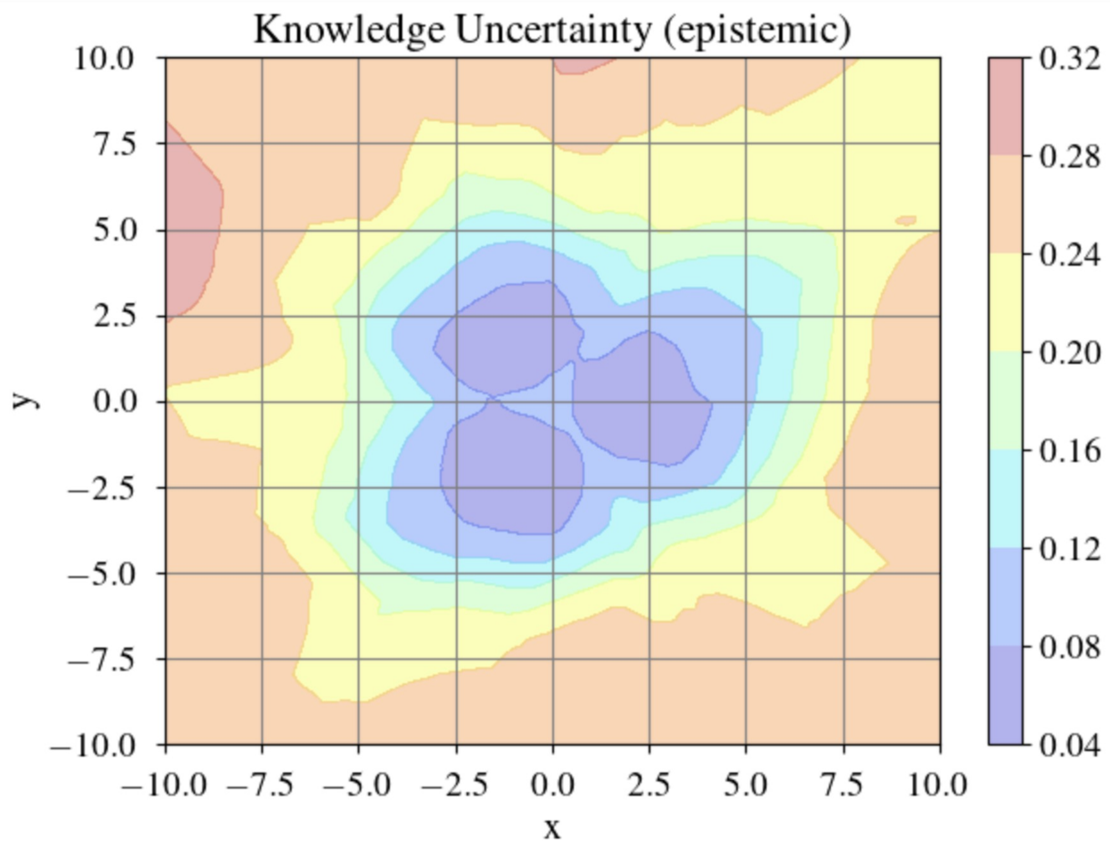
information theory-based uncertainty metrics

- We can separate the total classification predictive uncertainty into contributions from aleatoric and epistemic through analytic expressions of the Dirichlet parameters.

$$\mathbb{E}_{p(\mu|x^*,\hat{\theta})}\left[\mathbb{H}[p(y \mid \mu)]\right] = \sum_{k=1}^{K} -\frac{\alpha_k}{\alpha_0}\left(\psi(\alpha_k+1) - \psi(\alpha_0+1)\right)$$

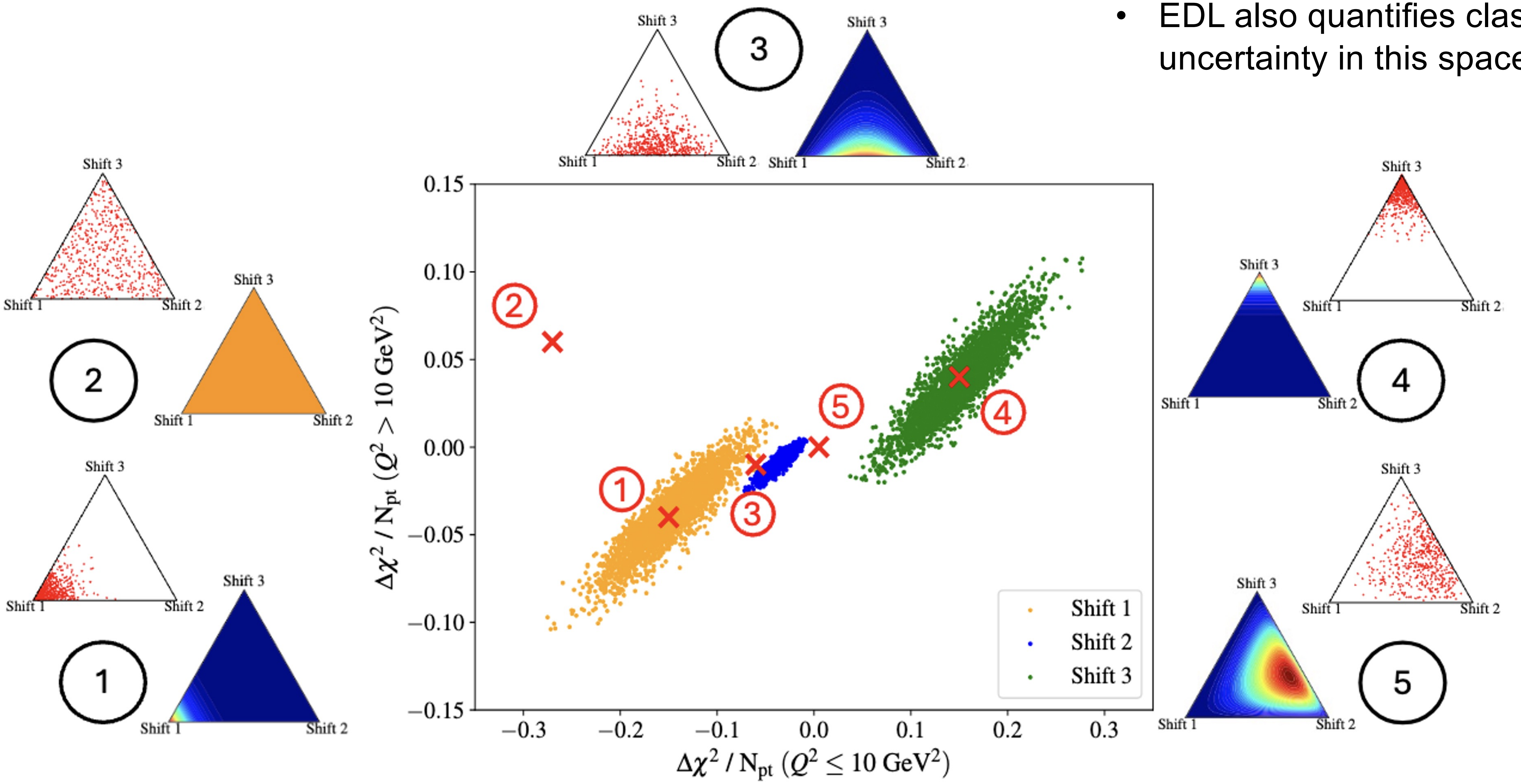
$$\mathcal{I}[y, \mu \mid x^*, \mathcal{D}] = \mathbb{H}\left[\mathbb{E}_{p(\mu|x^*,\mathcal{D})}[p(y \mid \mu)]\right] - \mathbb{E}_{p(\mu|x^*,\mathcal{D})}\left[\mathbb{H}[p(y \mid \mu)]\right]$$





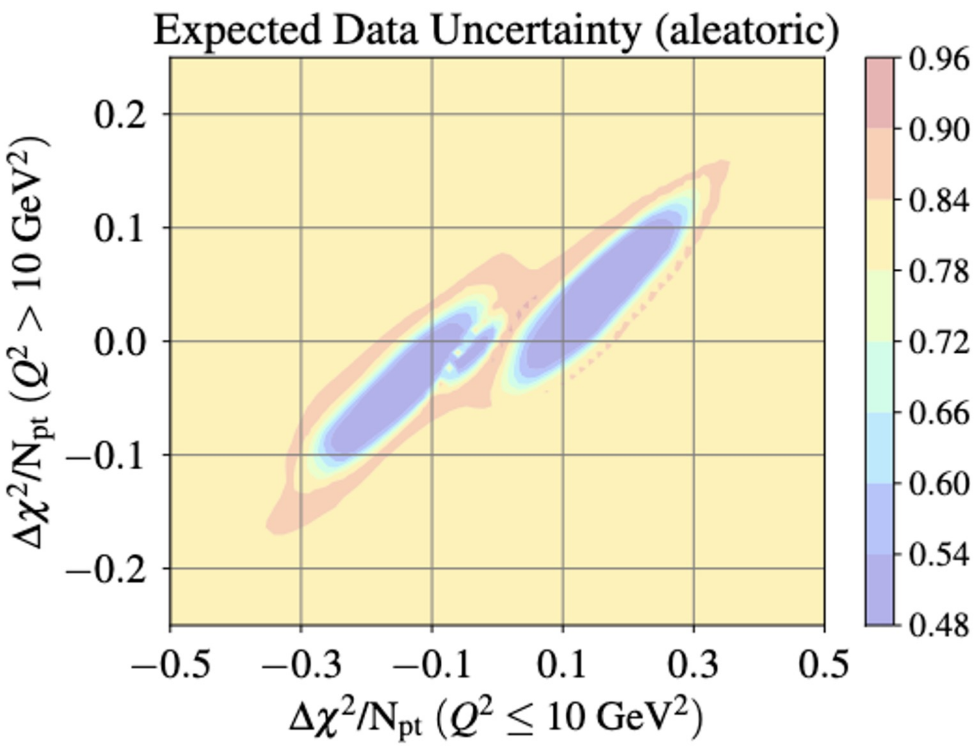
DPN discriminates among anomalous EW scenarios

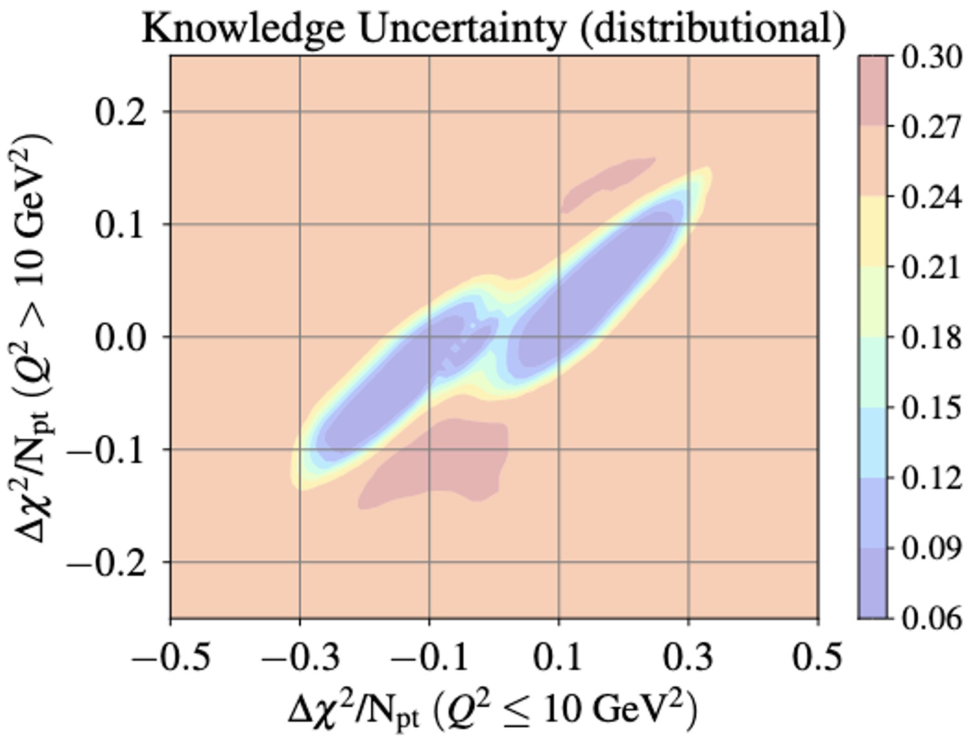
- DPN classifies EW scenarios in space of $\Delta\chi^2/N_{pt}$ ($Q^2 > 10$ GeV^2 and $Q^2 \leq 10$ GeV^2) for CDHSW
  - EDL also quantifies classification uncertainty in this space



can map ML model uncertainties in anomalous coupling scenarios

- the mapped space above can decompose EW coupling classifications into ML uncertainty sources



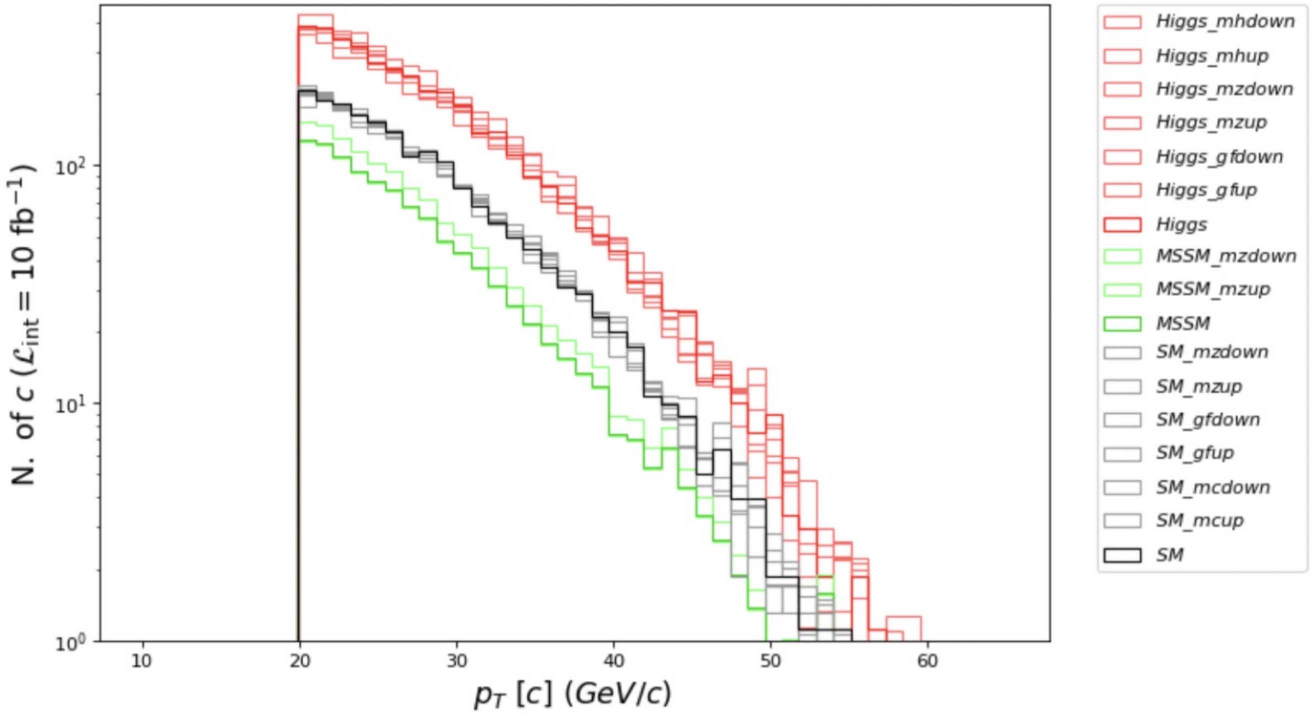


The information theory metrics can be used for high dimensional inputs where it may not be obvious where such overlaps occur.

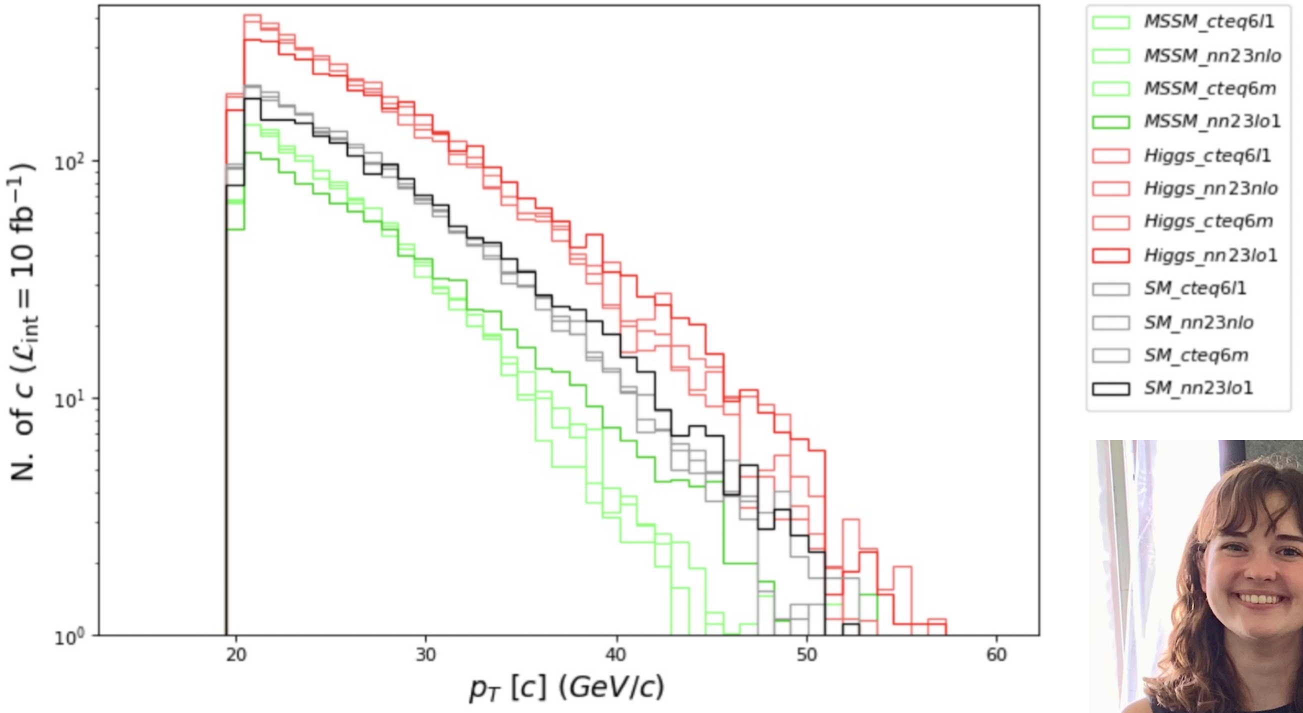
EIC-LHC interplay: UV-complete BSM scenarios

- foregoing demonstration: few-parameter variations in neutrino-quark effective couplings
  - → specific BSM theories can richly, nontrivially influence simultaneous descriptions of DIS/DY (*!*)

Baniecki, Das Bakshi, TJH, Kriesten; 25xx.xxxx



$$e^- + p \to \nu_e + \bar{c} + X$$



Adriana Baniecki,
Notre Dame (SULI)

- heavy-quark-tagged DIS: p_T spectra normalizations, shapes vary over (B)SM theories [2HDM, MSSM, SM] (left); unique interferences with PDF uncertainties (right)

SMEFT theory collider foundation models (FMs)

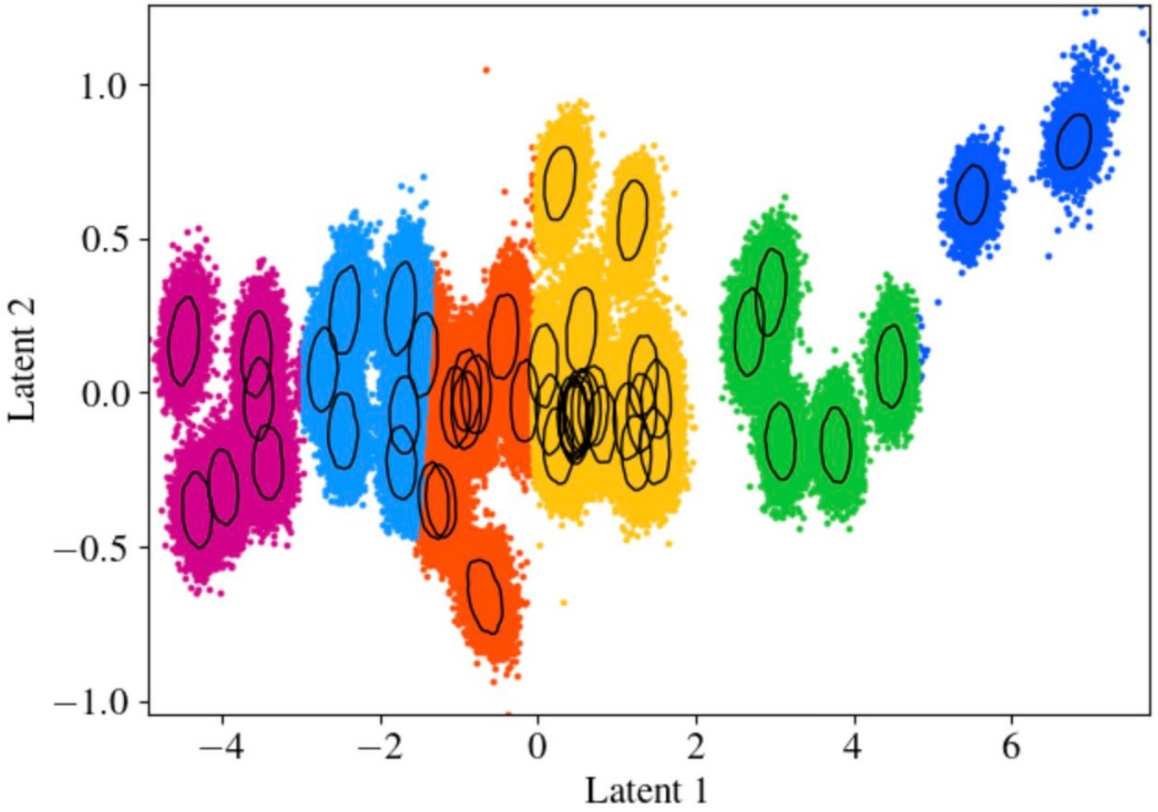
- can dramatically extend scope of BSM separations with SMEFT-driven collider FMs

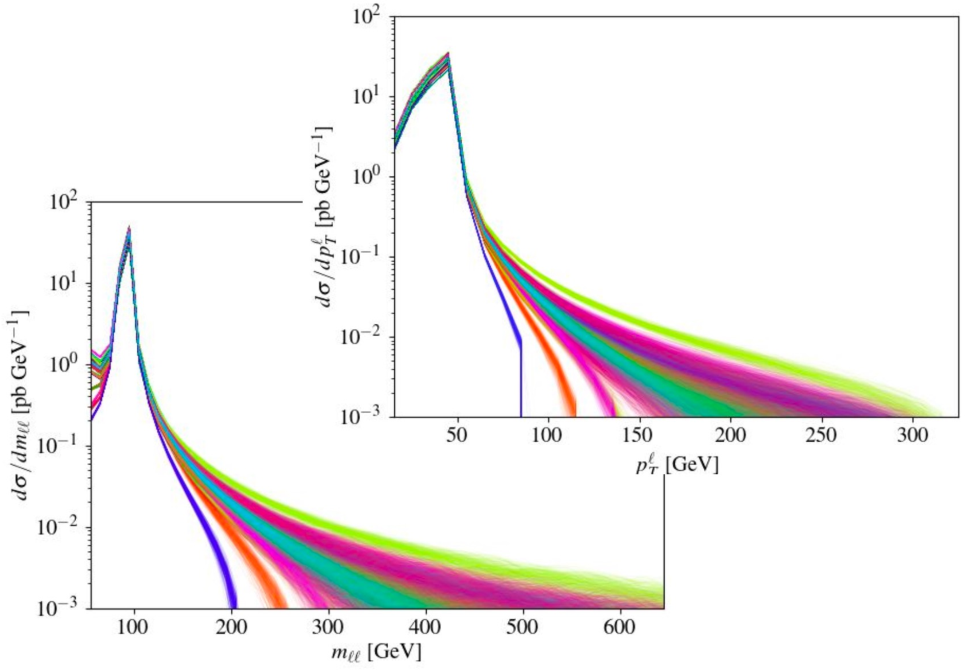
- here, large sampling of dim-6 operators at LHC (NC DY)

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{d>4} \frac{1}{\Lambda^{d-4}} \sum_i C_i^{(d)}\, \mathcal{O}_i^{(d)}$$

- ➢ separate, classify, cluster (right) many BSM scenarios (at EIC, LHC, …)

Das Bakshi, TJH, Kriesten; 25xx.xxxx





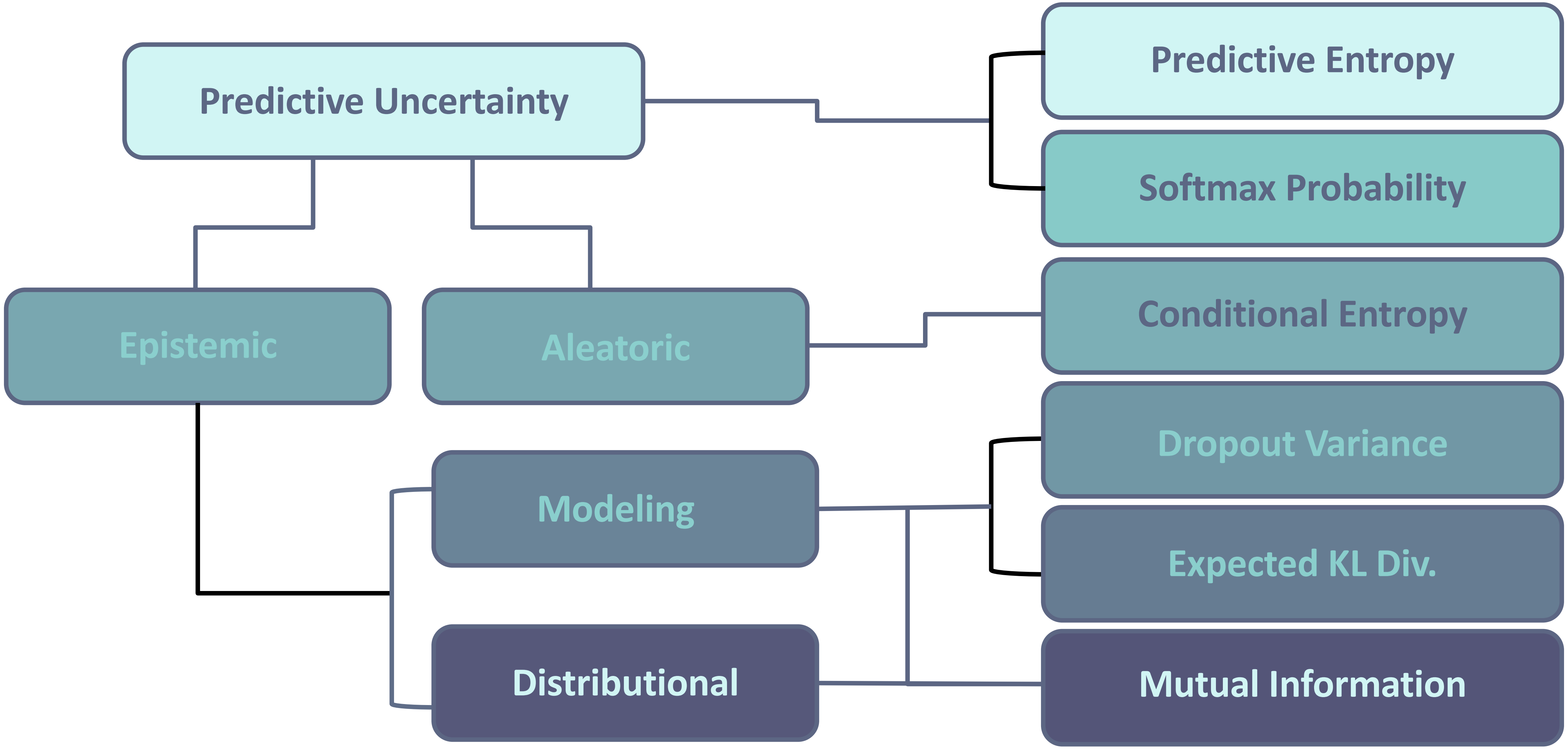
Collider Process	$\mathcal{O}(\Lambda^{-2})$ Dim-6 SMEFT Operators
$pp \to \mu^+\mu^-$	$C_{\varphi\Box}, C_{\varphi D}, C_{\varphi G}, C_{\varphi WB}, C_{e\varphi}, C_{u\varphi}$ $C_{d\varphi}, C_{eW}, C_{eB}, C_{uW}, C_{uB}, C_{dW}, C_{dB}$ $C_{\varphi\ell}^{(1)}, C_{\varphi\ell}^{(3)}, C_{\varphi e}, C_{\varphi q}^{(1)}, C_{\varphi q}^{(3)}, C_{\varphi u}, C_{\varphi d}$ $C_{\ell\ell}^{(1)}, C_{\ell q}^{(1)}, C_{\ell q}^{(3)}, C_{eu}, C_{ed}, C_{\ell u}, C_{\ell d}$ $C_{qe}, C_{\ell eqd}, C_{\ell equ}^{(1)}, C_{\ell equ}^{(3)}$

outlook

- illustrated novel ML approach for BSM discrimination, **suited to EIC-like analyses**
  - method capable of classifying potentially BSM-driven EW coupling shifts
  - demonstrated on typical neutrino DIS; extends to charge-current DIS at EIC (other processes)
  - generalizable to more complex BSM theories and parametrizations; larger analyses
- approach can extend to phenomenologically relevant (vague) classification problems
  - relevant to BSM searches and fits; SMEFT analyses, treatment of BSM theory classification
- exploring interface with PDF uncertainty definitions; BSM (EFT) foundation models

Backup

ecology of ML uncertainties and metrics

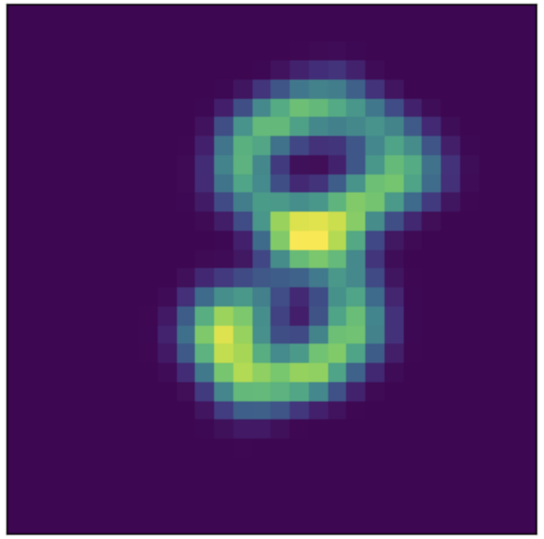


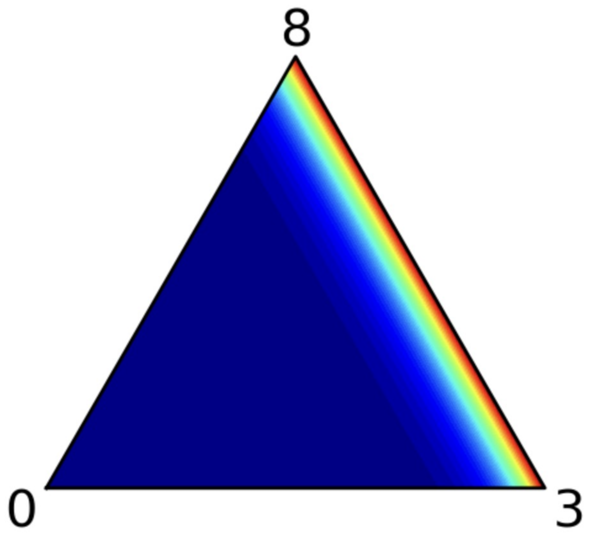
- ❑ these uncertainties apply to the ML modeling, but might ultimately be related to PDF uncertainties (in progress)

Kotz, Courtoy, Nadolsky, Olness, Ponce-Chavez: 2311.08447

in progress (i): hyper opinions and uncertain ground truth

- We may extend to scenarios in which uncertainties exist on the ground truth label itself; in that context, is it possible to exclude classification labels?





(Above) A toy illustration with MNIST: uncertain whether the ground truth is 3 or 8, but it certainly isn't 0

Grouped Dirichlet distribution models not only the distribution over the priors of the categorical, but also the distribution over composite labels

$$\mathrm{GDD}(p \mid \alpha, c) = Z^{-1} \prod_{k=1}^{K} p_k^{\alpha_k - 1} \prod_{j=1}^{\eta} \left(\sum_{l \in S_j} p_l \right)^{c_j}$$

Extends theory applications to exclusions of possible model or model combinations from classification

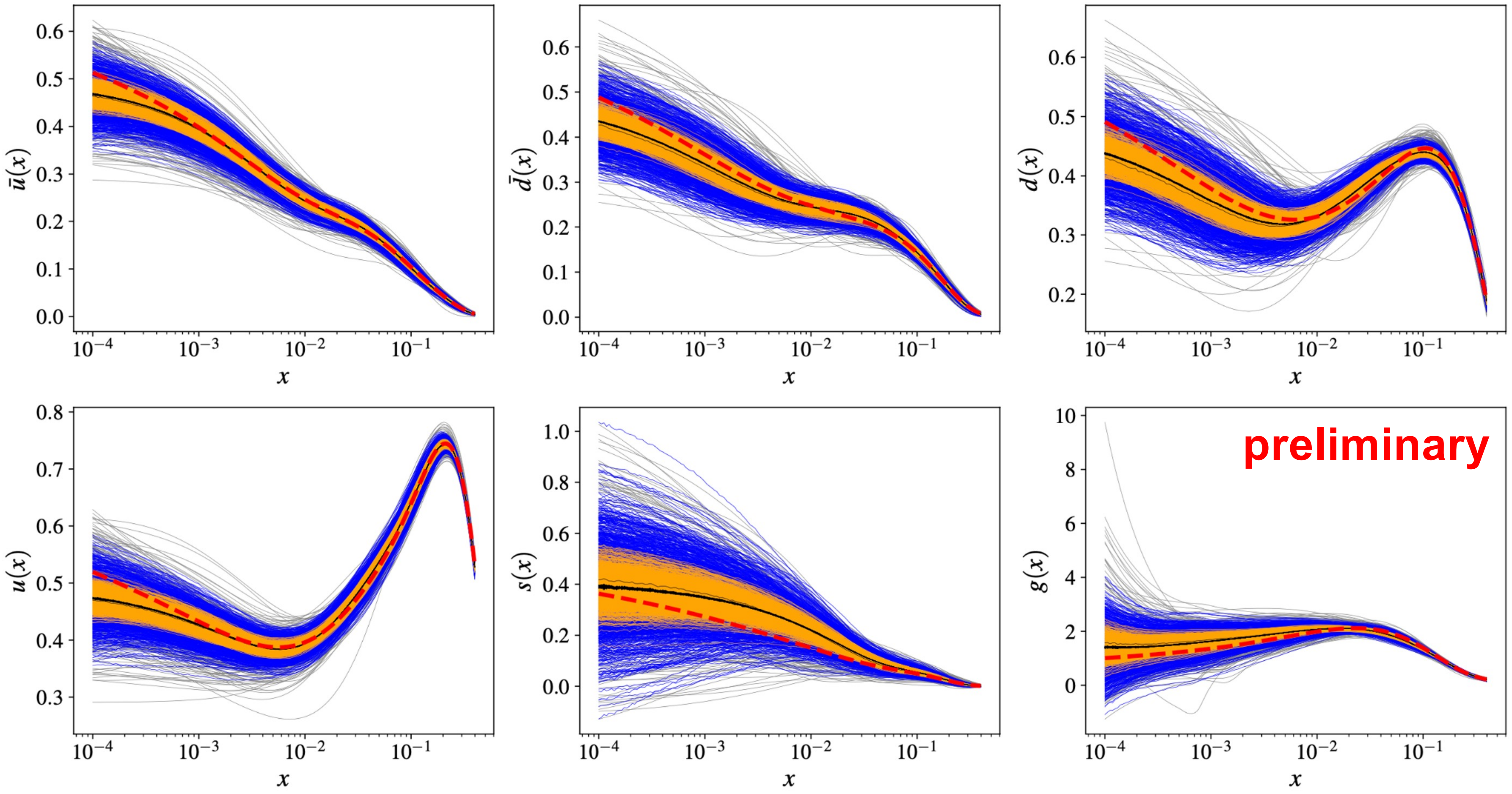
in progress (ii): connections to decoded (VAIM) PDFs; relation to PDF uncertainties

Kriesten, TJH PRD111 (2025) 1, 014028

- possible to apply these ML uncertainty methods to PDFs decoded from VAE inverse-mapper methods

Using MC error ensembles generated from global fits, can we extract new ensembles of solutions constrained to fit singular experiments directly from the data?

ensembles generated from CC νDIS data CDHSW



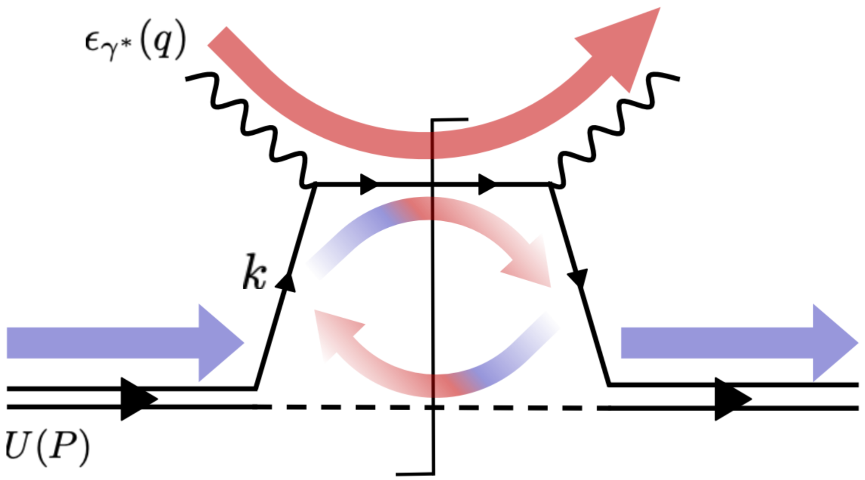
coda: testing factorization with quantum information-theoretic metrics

Bloss, Kriesten, TJH, 2503.15603

Aidala and Rogers, 2108.12319

- recent study --- factorization theorems may be recast in terms of entanglement and decoherence

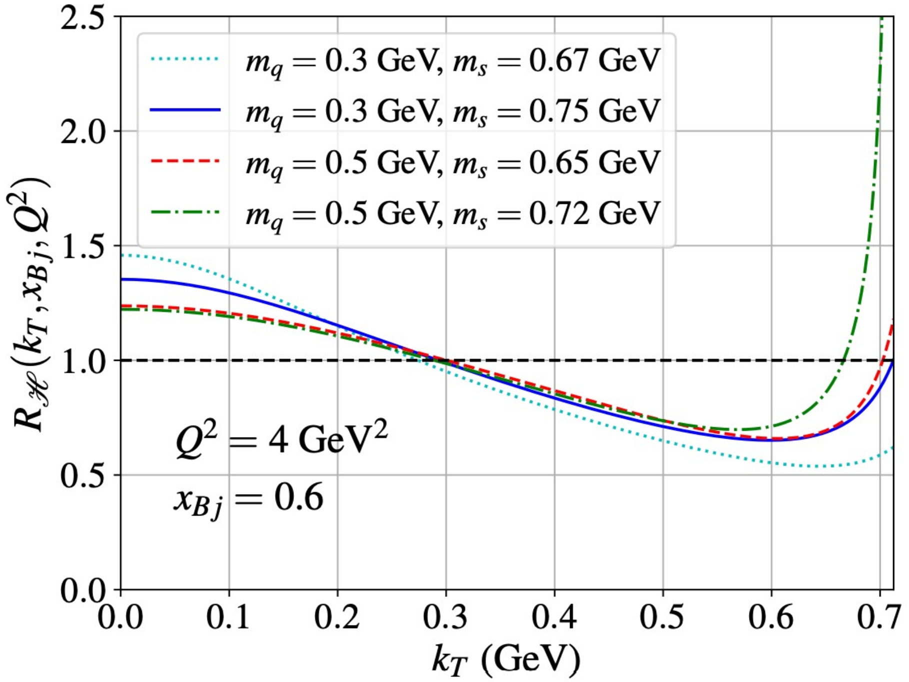
$$F(x,Q^2) = \sum_{i,j} \{C_{i,j} \otimes f_j\}(x,Q^2) + \mathcal{O}\left(\frac{M^2}{Q^2}\right)$$



- can quantify in a generic model of the nucleon (exact *vs.* factorized DIS calculation)

Moffat et al., 1702.03955

$$\mathcal{H}_{F_1}(x,k_T,Q^2) \equiv -k_T\, f_1(x,k_T,Q^2) \ln\left[f_1(x,k_T,Q^2)\right]$$



- power-suppressed corrections inhibit factorizability; realized as long-distance entanglement which prevent decoherence of hard scatter from binding parton-nucleon interaction

- simulatable quantum entropies have information-theoretic analogues (differential entropy; KL divergence; …)

A new uncertainty emerges … distributional uncertainty

The epistemic uncertainty seems to naturally separate as well:

- a contribution arising from modelling by the AI / ML algorithm
- a contribution arising from the choice of underlying distribution to describe the data

$$p(w_c|x^*, \mathcal{D}) = \int\int p(w_c|\mu)\boxed{p(\mu|x^*, \theta)p(\theta|\mathcal{D})}d\theta d\mu$$

In theory, we can model all three at the same time (aleatoric, distributional, epistemic) - creating an ecology of uncertainties!

Malinin and Gales arXiv:1802.10501