

Learning Particle Physics - From Simulation to Inference with uncertainty aware Neural Networks

Anja Butter

LPNHE (Paris), ITP (Heidelberg)



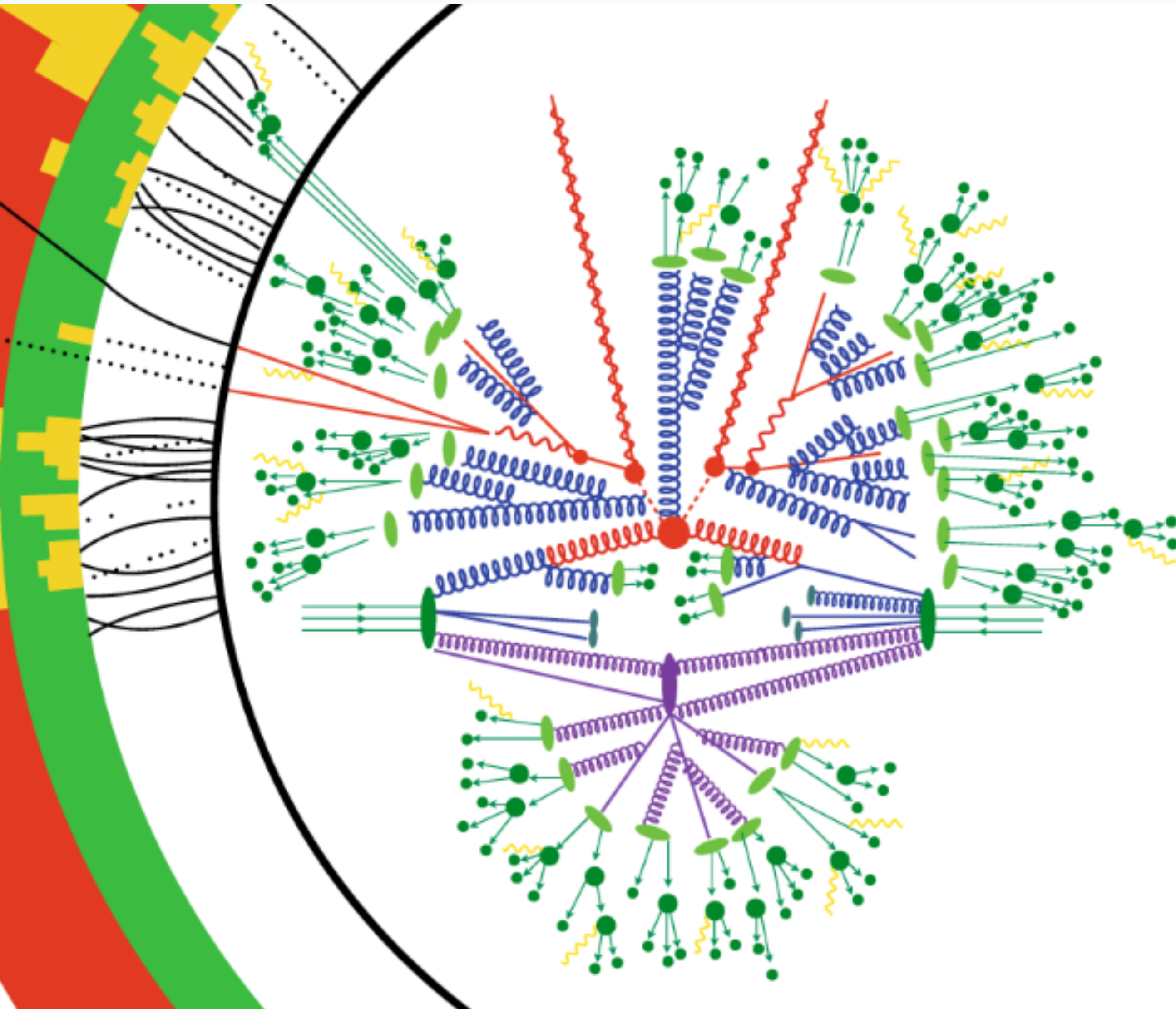
Why use machine learning in particle physics?

Curiosity

Pile up event in ATLAS



Control from theory & phenomenology



LHC analyses require for each analysis:

- * High statistics - millions of events
- * High precision - N^x LO process, off-shell,...

State of the art runtime for individual diagrams

Integrator \ Accuracy		10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}
banana_3mass	DISTEVAL	2.1 s	2.1 s	2.4 s	2.6 s	2.6 s	2.9 s	3.6 s
	INTLIB	5.0 s	4.9 s	6.4 s	7.2 s	8.5 s	8.5 s	13.8 s
	Ratio	2.3	2.3	2.7	2.7	3.2	3.0	3.9
bubble6L	DISTEVAL	1.8 m	1.8 m	1.8 m	2.1 m	3.8 m	10.2 m	1.2 h
	INTLIB	39.5 m	38.8 m	39.6 m	43.8 m	85.1 m	170.7 m	11.6 h
	Ratio	22	22	22	21	22	17	10
formfactor4L	DISTEVAL	4.1 m	4.1 m	4.1 m	4.4 m	7.7 m	14.6 m	0.96 h
	INTLIB	74 m	73 m	73 m	74 m	136 m	246 m	10.9 h
	Ratio	18	18	18	17	18	17	11
elliptic2L_physical	DISTEVAL	1.6 s	1.5 s	1.7 s	1.9 s	4.0 s	19 s	7.6 m
	INTLIB	3.1 s	4.8 s	4.9 s	7.3 s	13.8 s	53 s	4.3 m
	Ratio	1.9	3.1	2.8	3.9	3.4	2.9	0.6
hz2L_nonplanar	DISTEVAL	2.1 s	2.6 s	4.6 s	30.4 s	2.2 m	5.1 m	27.1 m
	INTLIB	9 s	17 s	41 s	163 s	9.6 m	16.0 m	27.3 m
	Ratio	1.8	3.4	4.6	4.4	4.2	3.0	1.0
Nbox2L_split_b	DISTEVAL	2.7 s	9.8 s	16.8 s	0.58 m	2.4 m	9.1 m	20 m
	INTLIB	24 s	73 s	223 s	6.6 m	26 m	43 m	93 m
	Ratio	3.0	4.6	9.7	9.9	10.5	4.8	4.7
pentabox_fin	DISTEVAL	5 s	8 s	11 s	0.71 m	3.7 m	18.5 m	1.1 h
	INTLIB	45 s	65 s	88 s	3.2 m	11.3 m	74.8 m	4.6 h
	Ratio	8.6	7.9	7.7	4.5	3.1	4.0	4.2

Table 4: Integration timings on a GPU for different examples using the INTLIB qmc integrator and the new DISTEVAL integrator. All timings are using the CBC generating vectors from the previous release, meaning the ratios between INTLIB and DISTEVAL are purely due to the improvements described in [Section 2.1](#). The significantly improved timings achieved by using the new median generating vectors are shown in [Table 1](#).

Why use machine learning in particle physics?

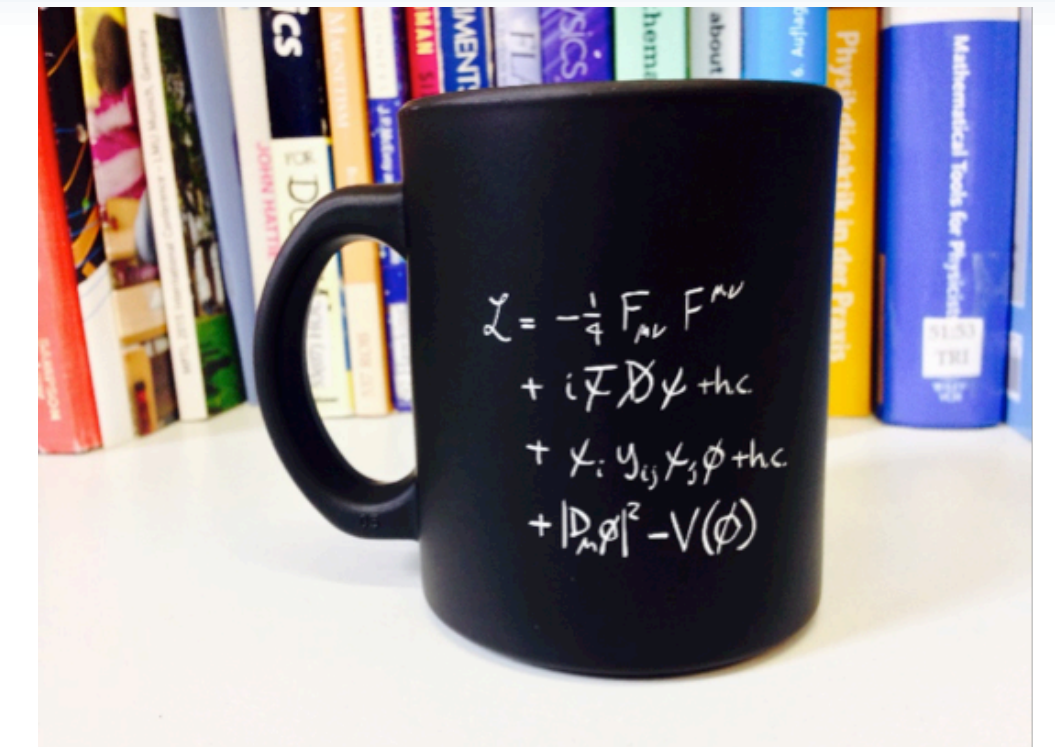
1. Curiosity

2. Perfect conditions

- Big data
- Full control through simulations
- Interesting Problem

3. Large gain

- Improve efficiency & precision -> better reach of analyses
- Free capacity through agentification
- New ideas (anomaly detection, unbinned unfolding, ...)
- New dynamic from fast moving developments in ML



Why use machine learning in particle physics?

1. Curiosity

2. Perfect conditions

- Big data
- Full control through
- Interesting Problem

3. Large gain

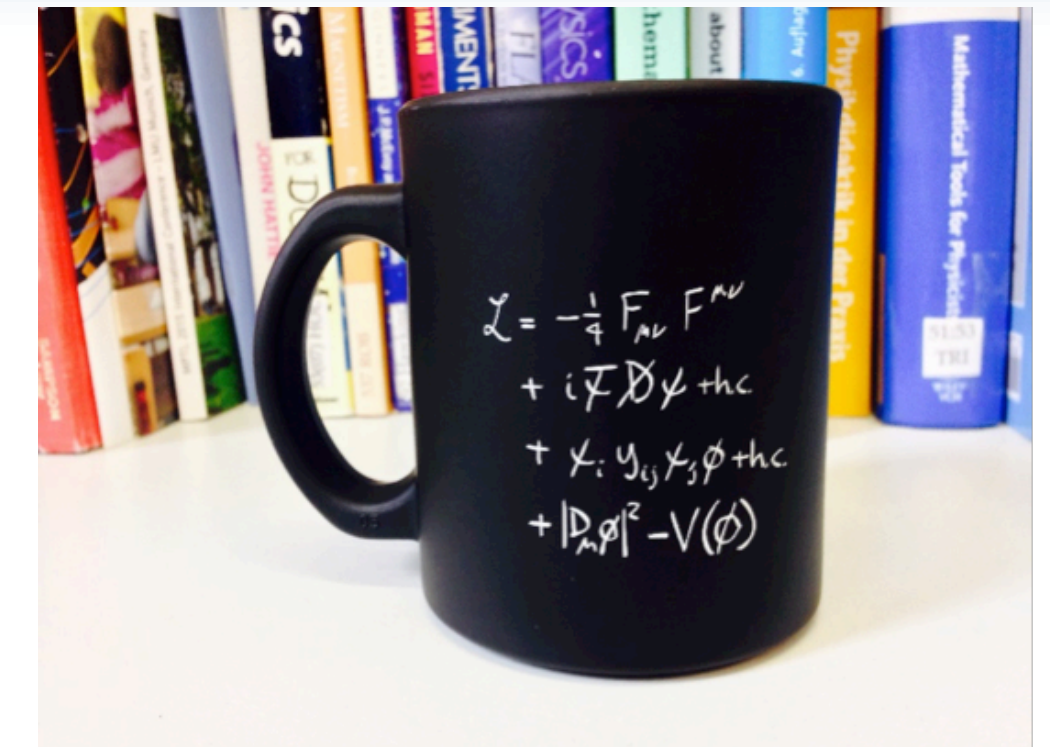
- Improve efficiency &
- Free capacity through
- New ideas (anomaly)
- New dynamic from fast moving developments in ML

Develop & adapt ML techniques
→ right balance between ML & theory

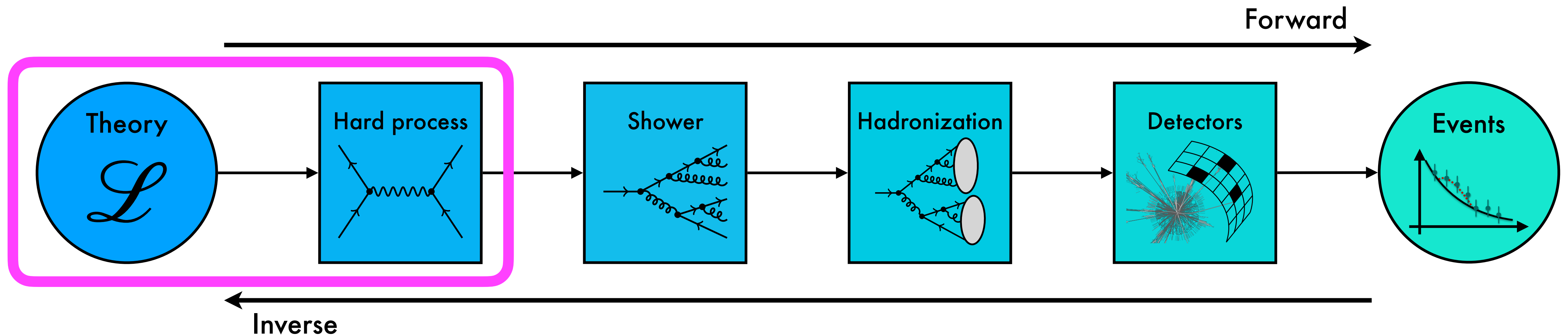
- Fast precision calculations & simulations -

- Optimized data analysis -

- Calibrated uncertainties -

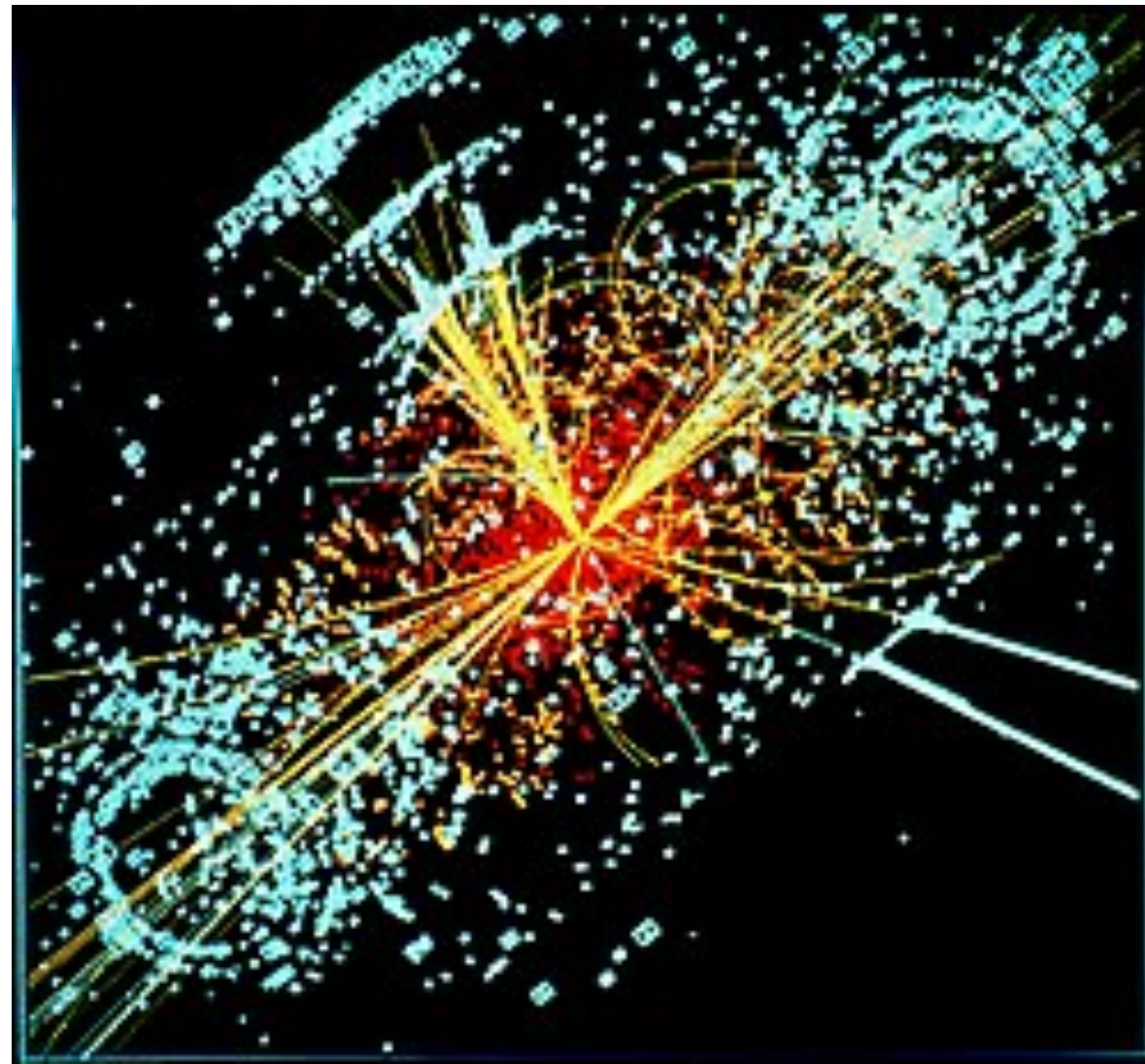


Monte Carlo event generation

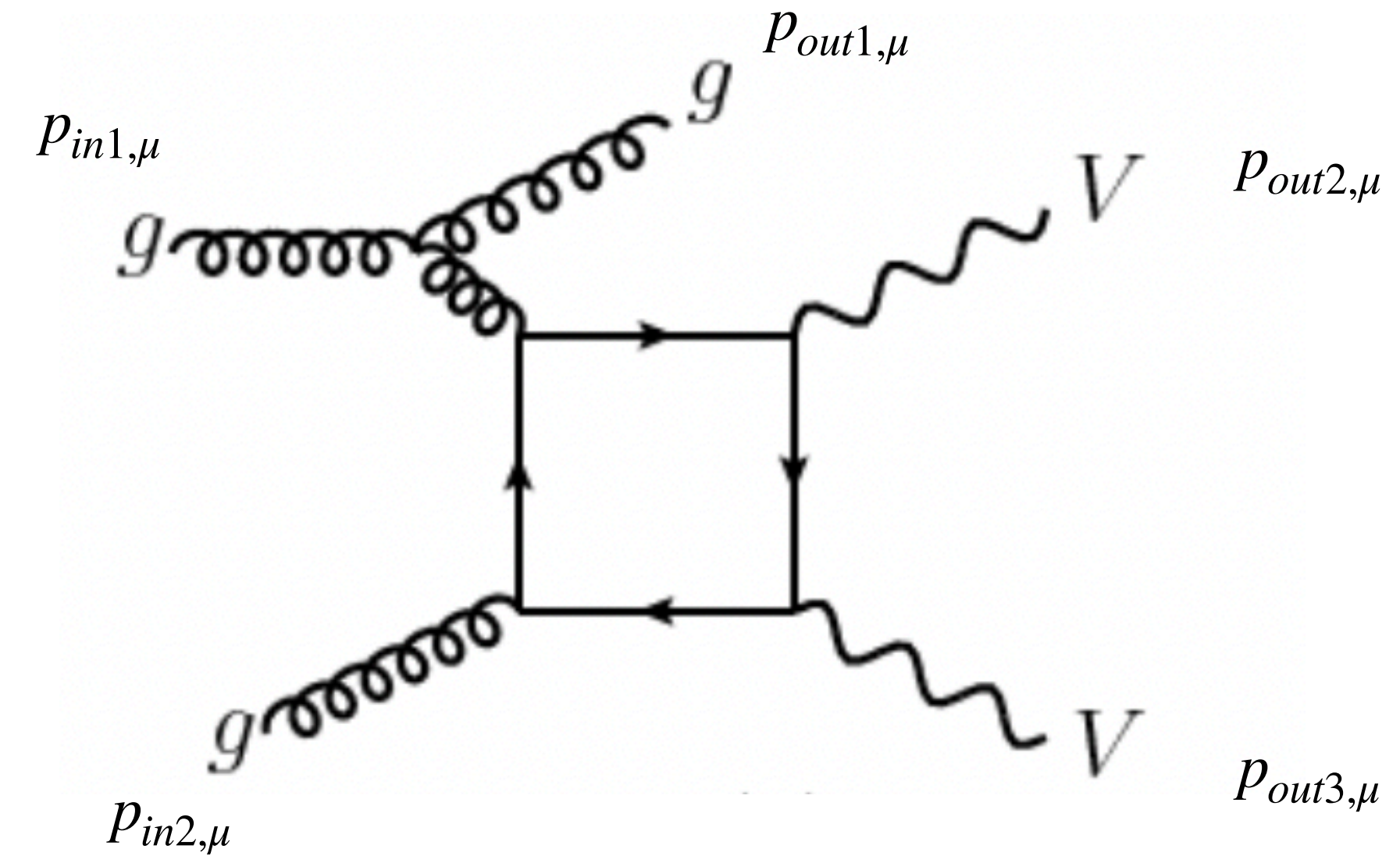


1. Precision amplitudes or Quantifying uncertainties in regression problems

Learning Amplitudes - A Regression Problem



Experimentalist's view



Theorist's view

- Amplitudes are @ the heart of LHC physics - probability of particle interactions
- Compute $A(\{p_{i,\mu}\} \mid i = 1, \dots, N_{particles})$ as function of particle momenta p_μ
- Spans multiple orders of magnitude
- Very expensive for higher order processes and large $N_{particles}$

→ **Use NN to interpolate between known amplitudes**

Machine Learning for theorists

Machine Learning

Physics

Problem setting

Loss function $\mathcal{L}$

Likelihood $\mathcal{L}$

Learnable function

Neural network f_θ

Physics knowledge $f_{physics}$

Data

Sampling $\frac{1}{N} \sum_{i=1}^N \mathcal{L}(f_\theta(x_i); x_i)$

Phase space integration $\int dx p(x) \dots$

Solving the problem

Gradient descent $\theta^i \rightarrow \theta^{i+1} = \theta^i - \nabla_{\theta_i} \mathcal{L}$

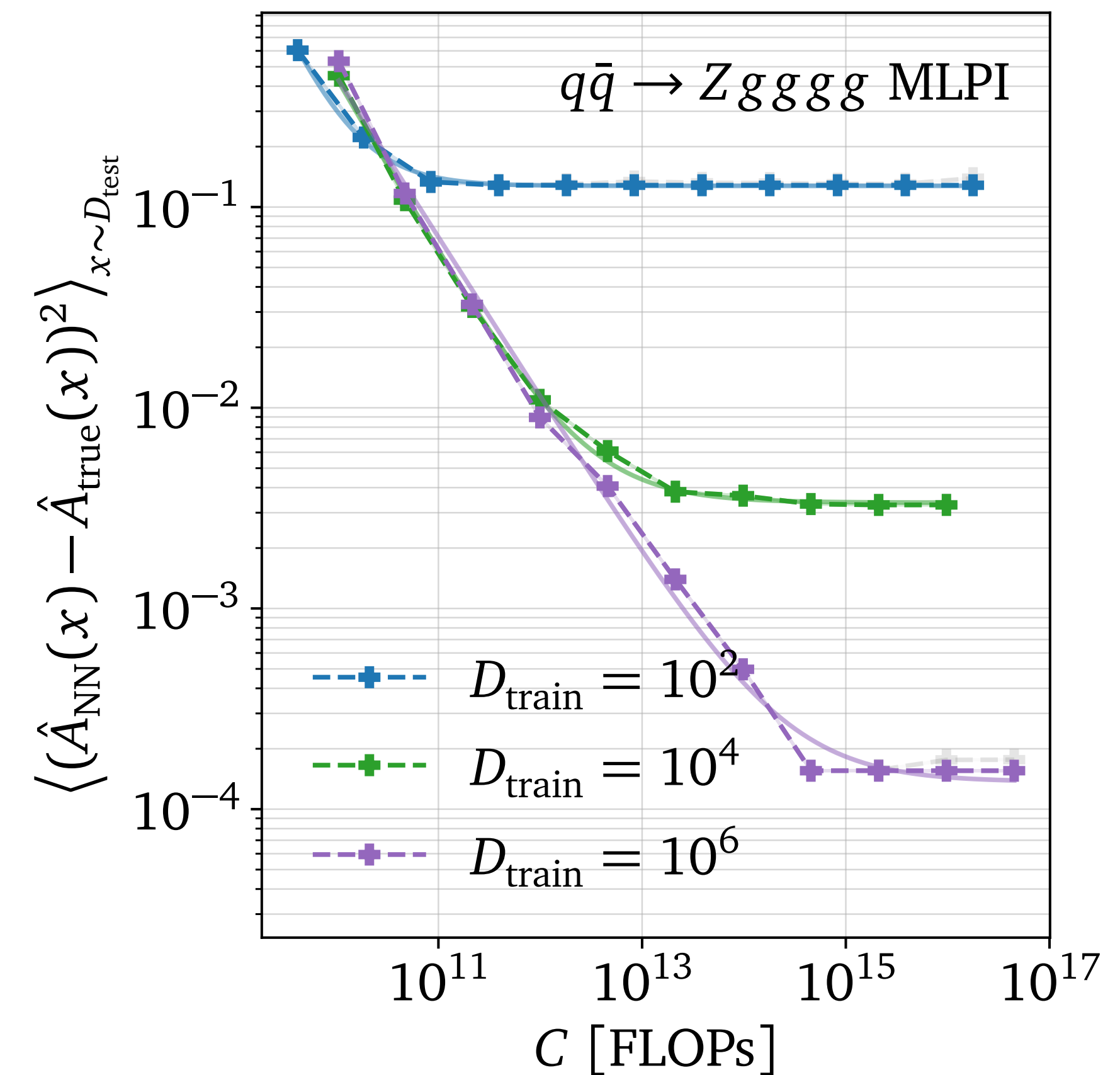
Functional derivative $\frac{\delta \mathcal{L}}{\delta f} \stackrel{!}{=} 0$

How to achieve high accuracy amplitudes?

[2601.13308] H. Bahl, V. Bréso, AB, J. Ramirez

Example $q\bar{q} \rightarrow Zgggg$

- Reminder:
 - Amplitudes are known exactly
 - Numerical noise negligible
- Setup:
 - Standard NN (MLP)
 - Use invariants (s, t, u) as input
 - Evaluate MSE over test data set



How to achieve high accuracy amplitudes?

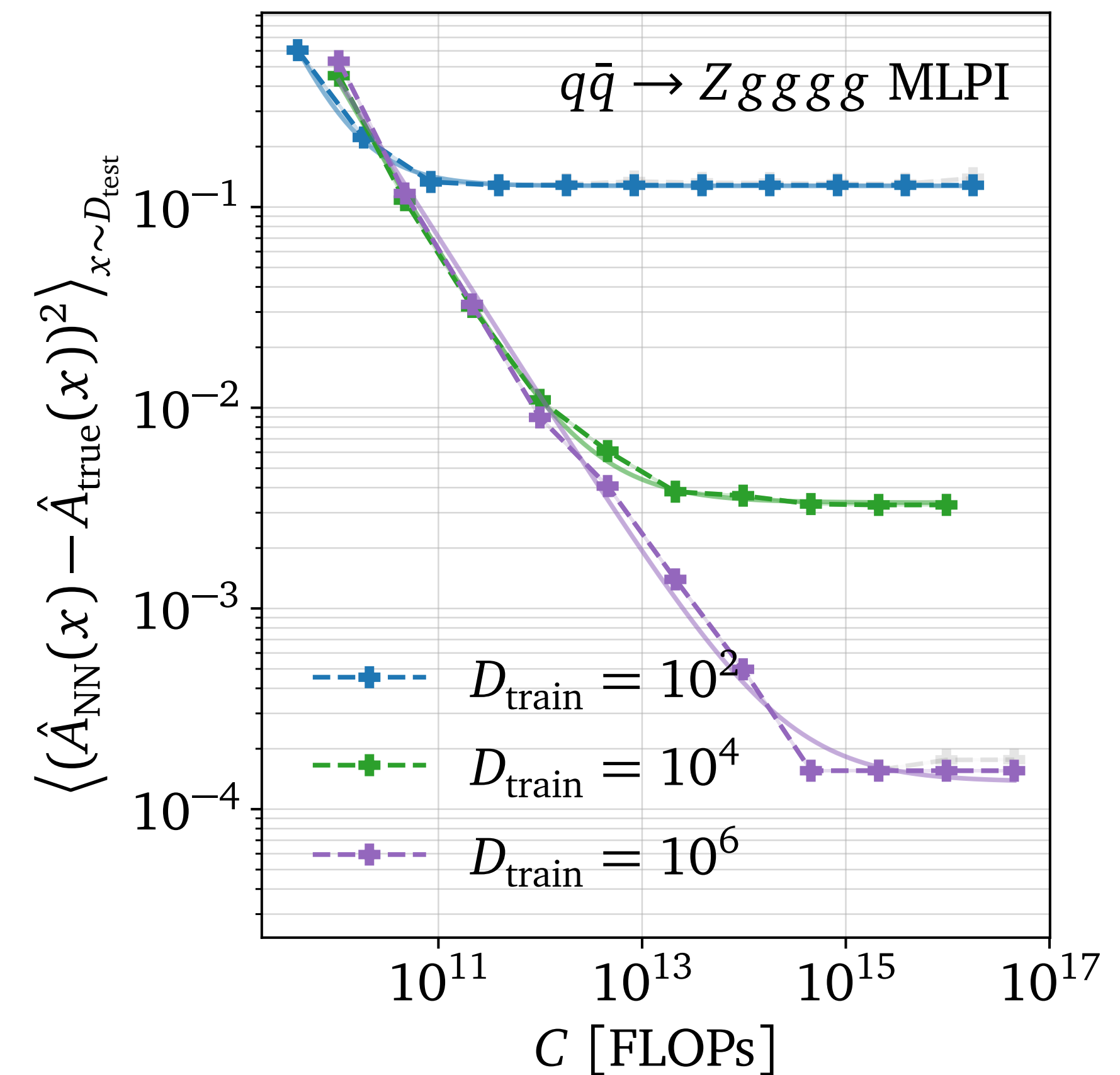
[2601.13308] H. Bahl, V. Bréso, AB, J. Ramirez

Example $q\bar{q} \rightarrow Zgggg$

- Reminder:
 - Amplitudes are known exactly
 - Numerical noise negligible
- Setup:
 - Standard NN (MLP)
 - Use invariants (s, t, u)
 - Evaluate MSE over test

Power law scaling
$$L(X, Y, Z) = (X_c/X)^{\alpha_x} + K(Y, Z)$$

for
Model size
Dataset size
Compute

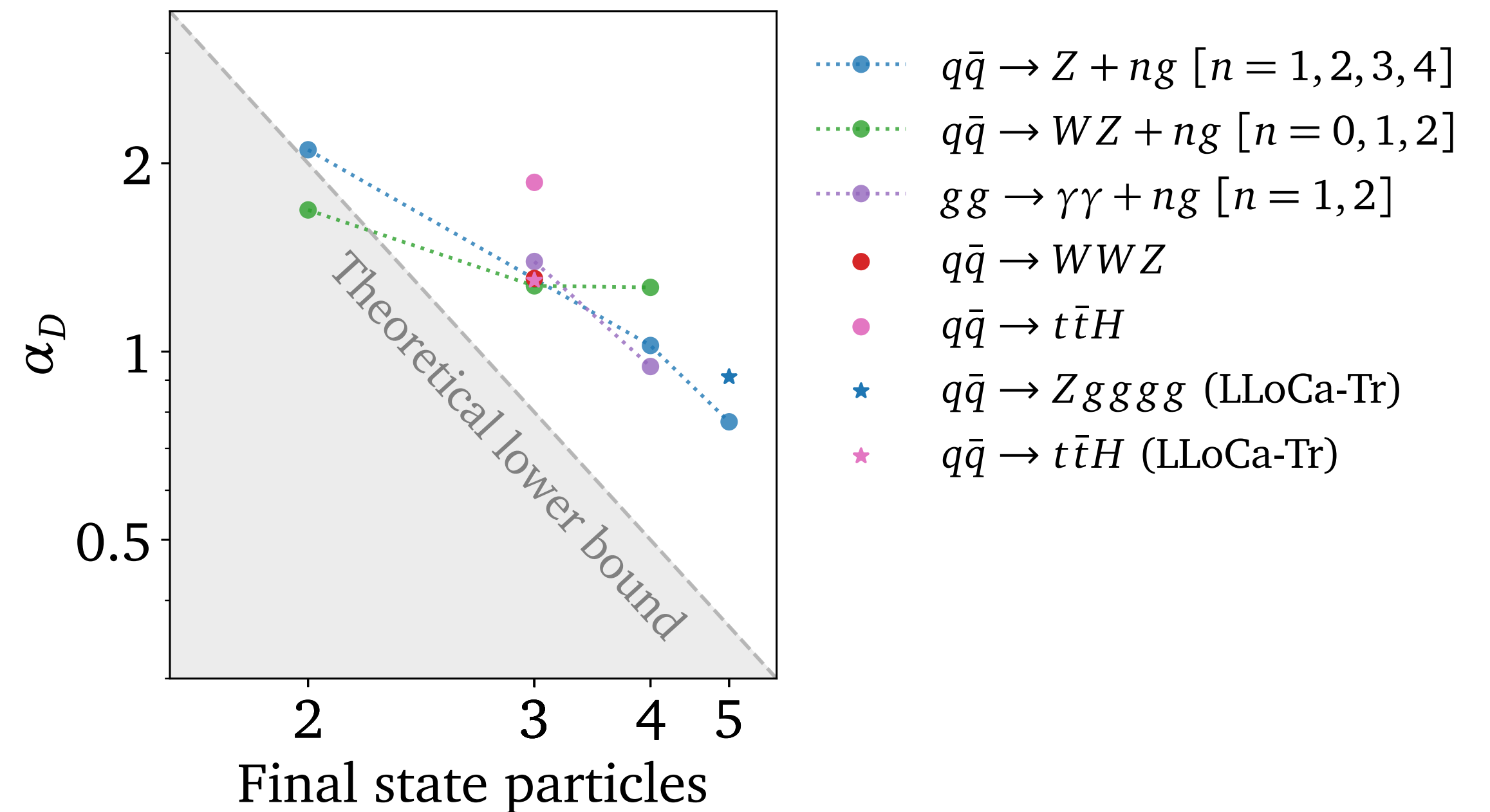
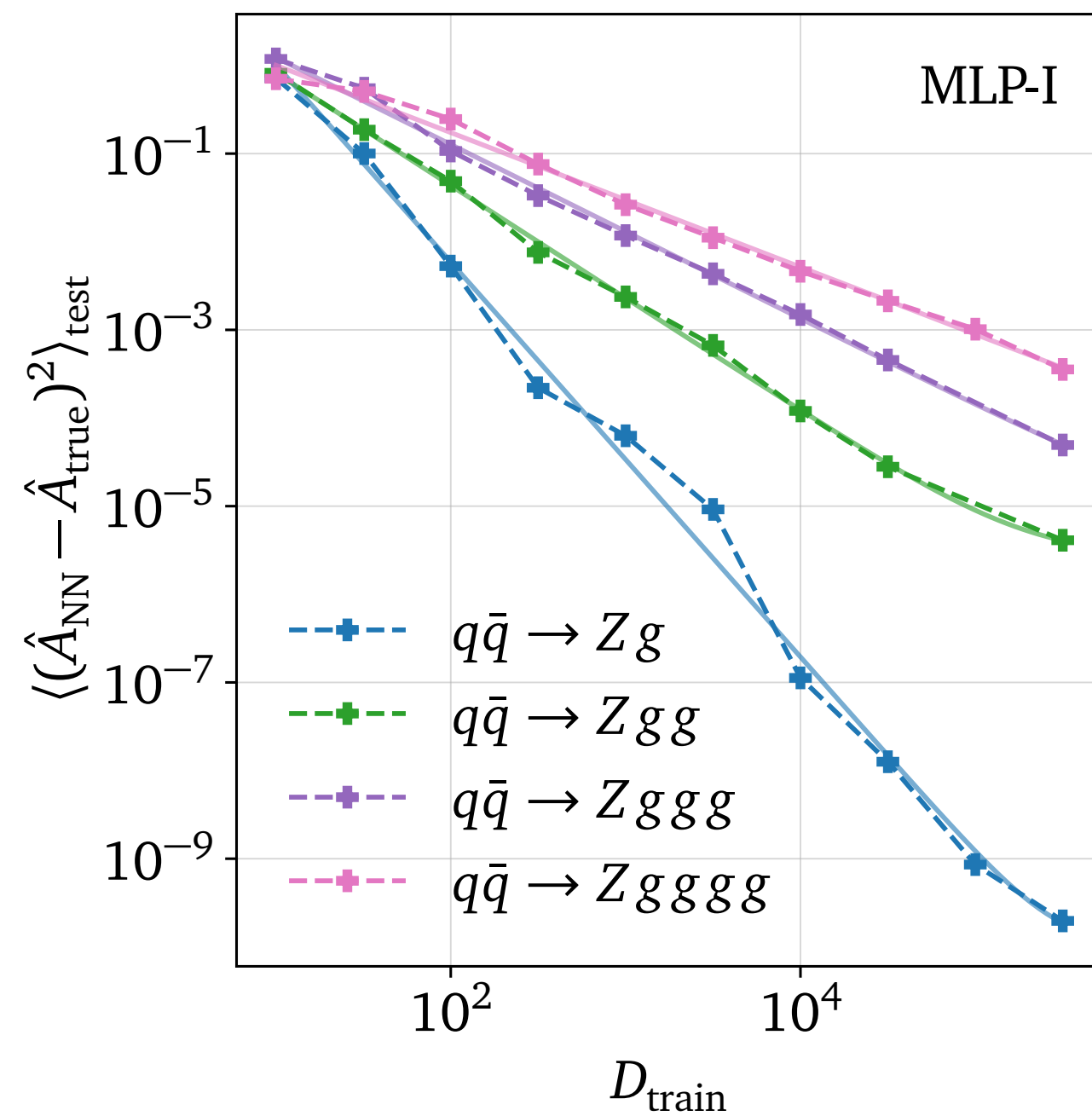


Can we predict the scaling?

[2601.13308] H. Bahl, V. Bréso, AB, J. Ramirez

Example $q\bar{q} \rightarrow Z + ng$

- Scaling depends on degrees of freedom



Including symmetries

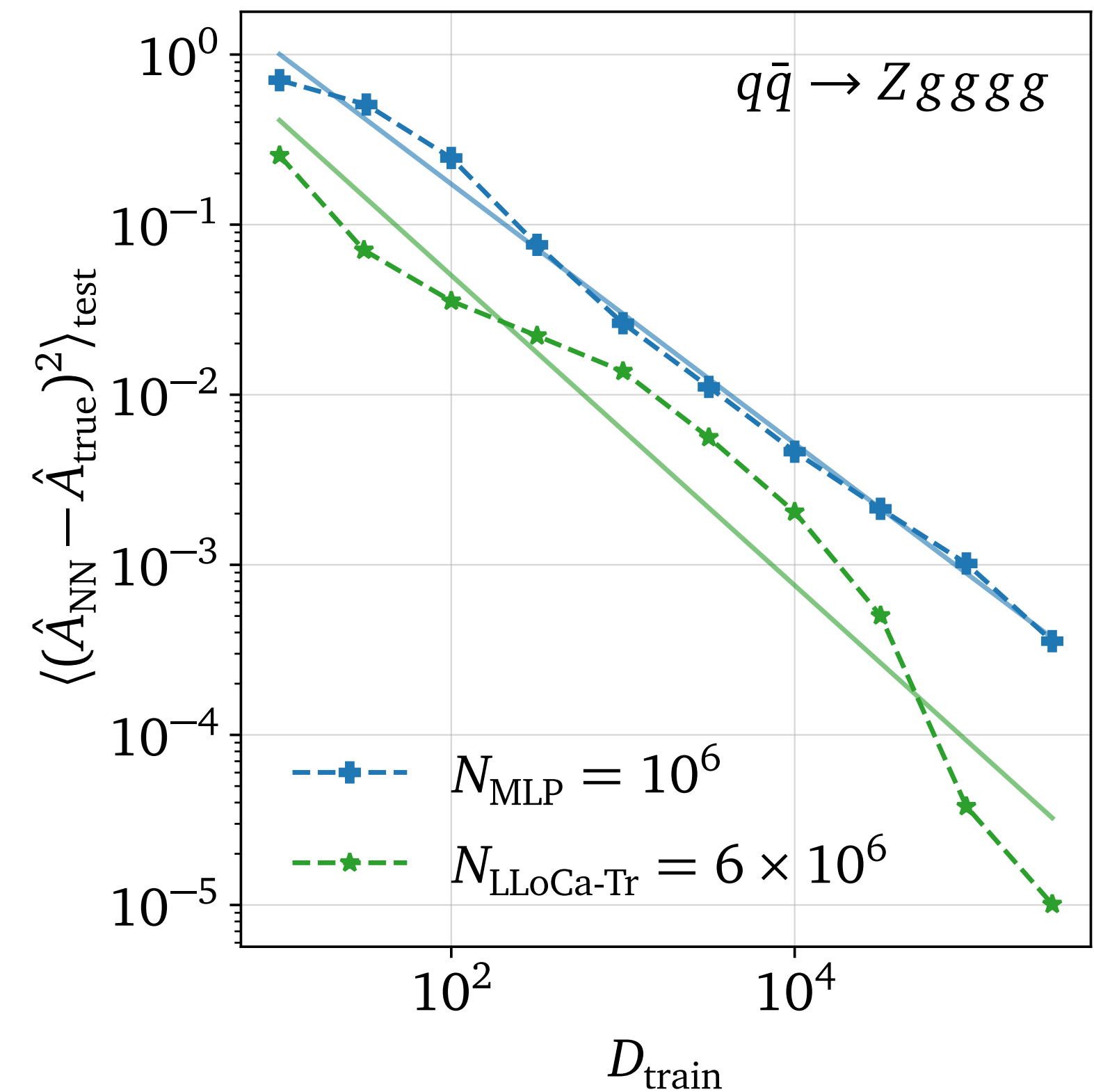
[2601.13308] H. Bahl, V. Bréso, AB, J. Ramirez

LLoCa network

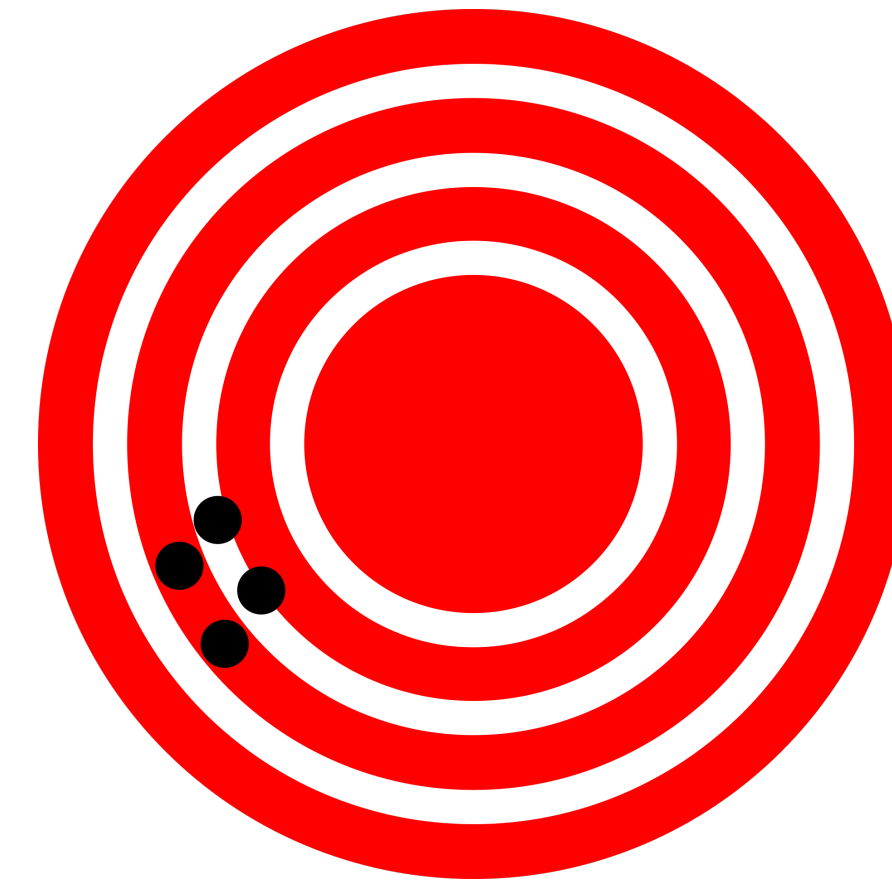
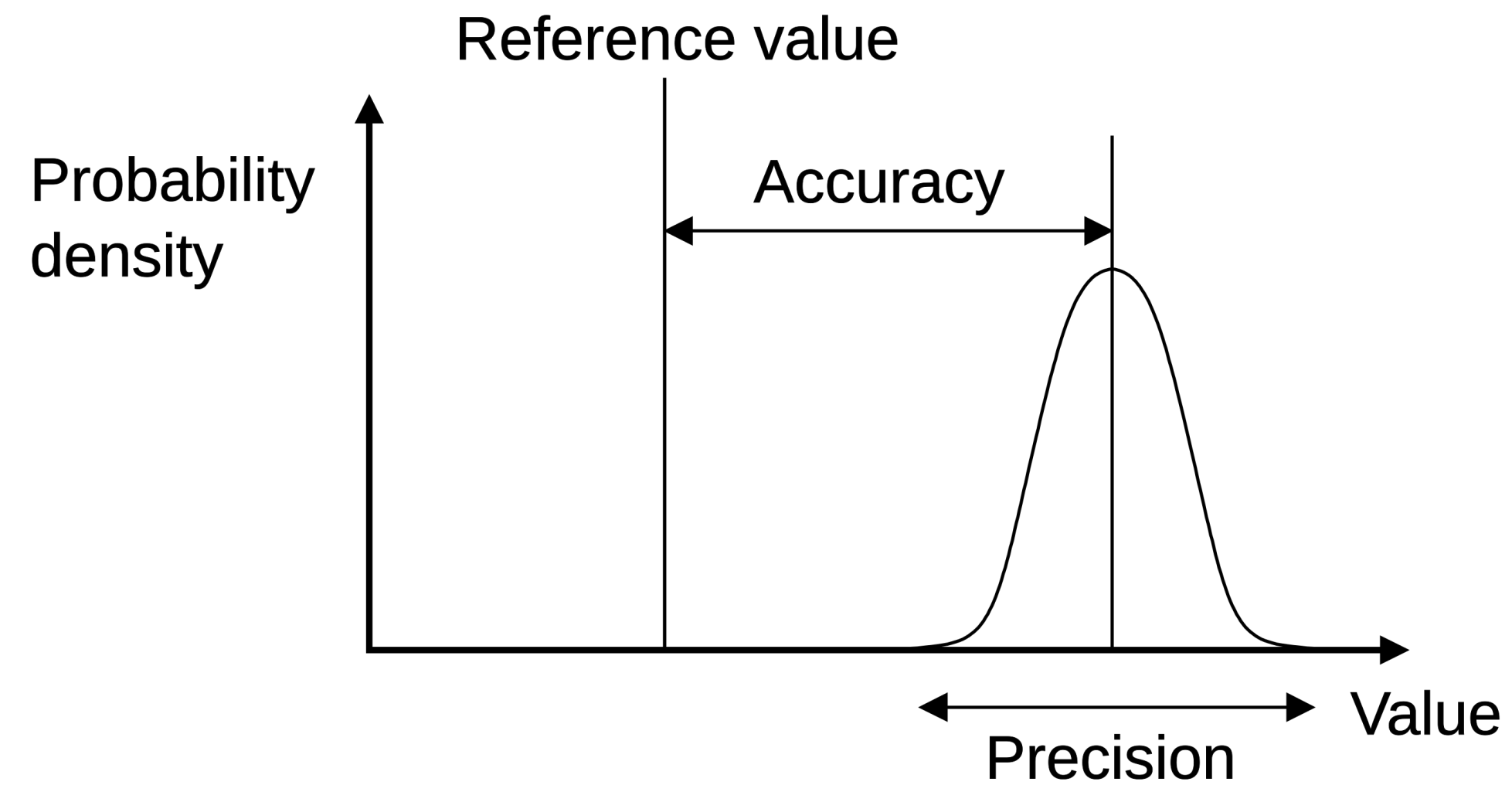
Lorentz Local Canonicalization [2505.20280]

J. Spinner, L. Favaro, P. Lippmann, S. Pitz,
G. Gerhartz, T. Plehn, F. A. Hamprecht

- Equivariance wrt. Lorentz transformations
 1. Transforms each particle into local reference frame
 2. Apply arbitrary backbone architecture
 3. Transform back to initial frame work



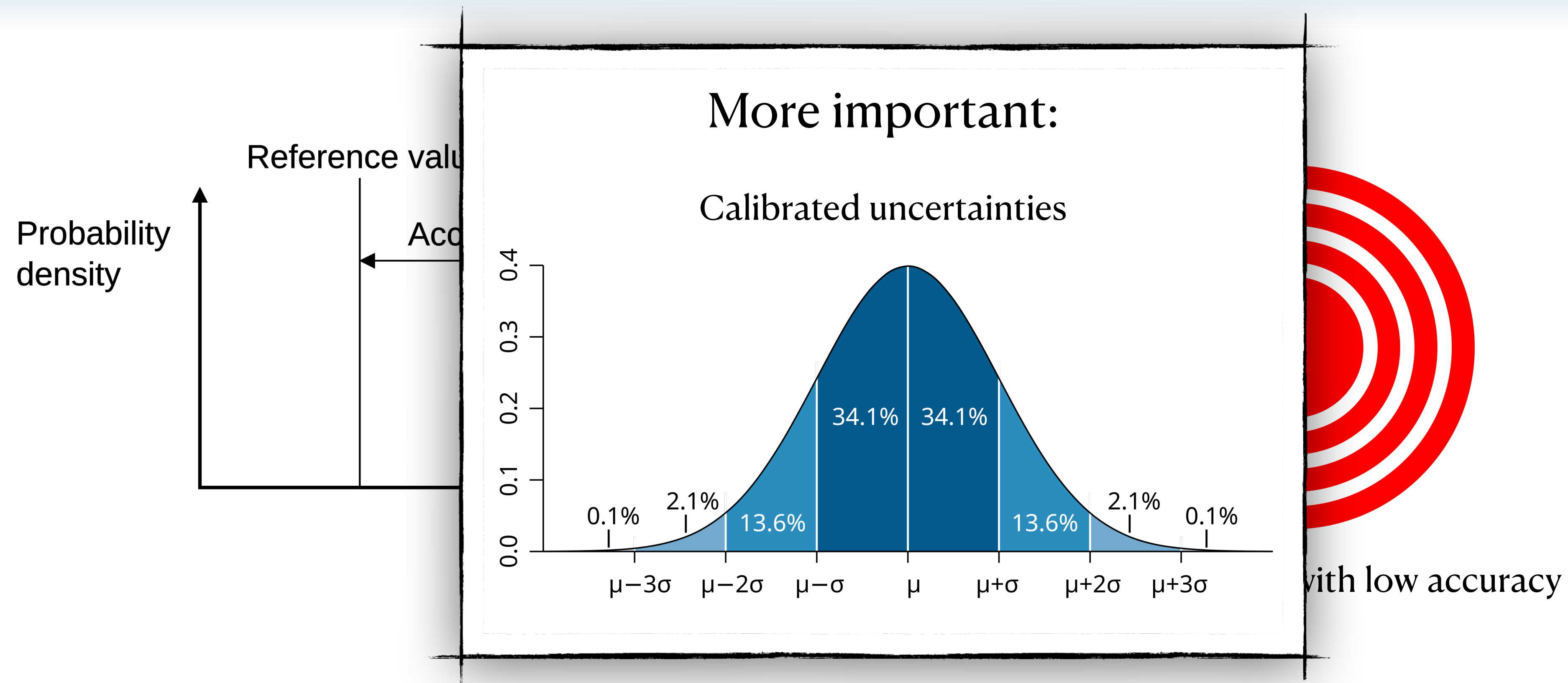
Accuracy vs precision



High precision with low accuracy

Best of all worlds: high accuracy & high precision

Accuracy vs precision

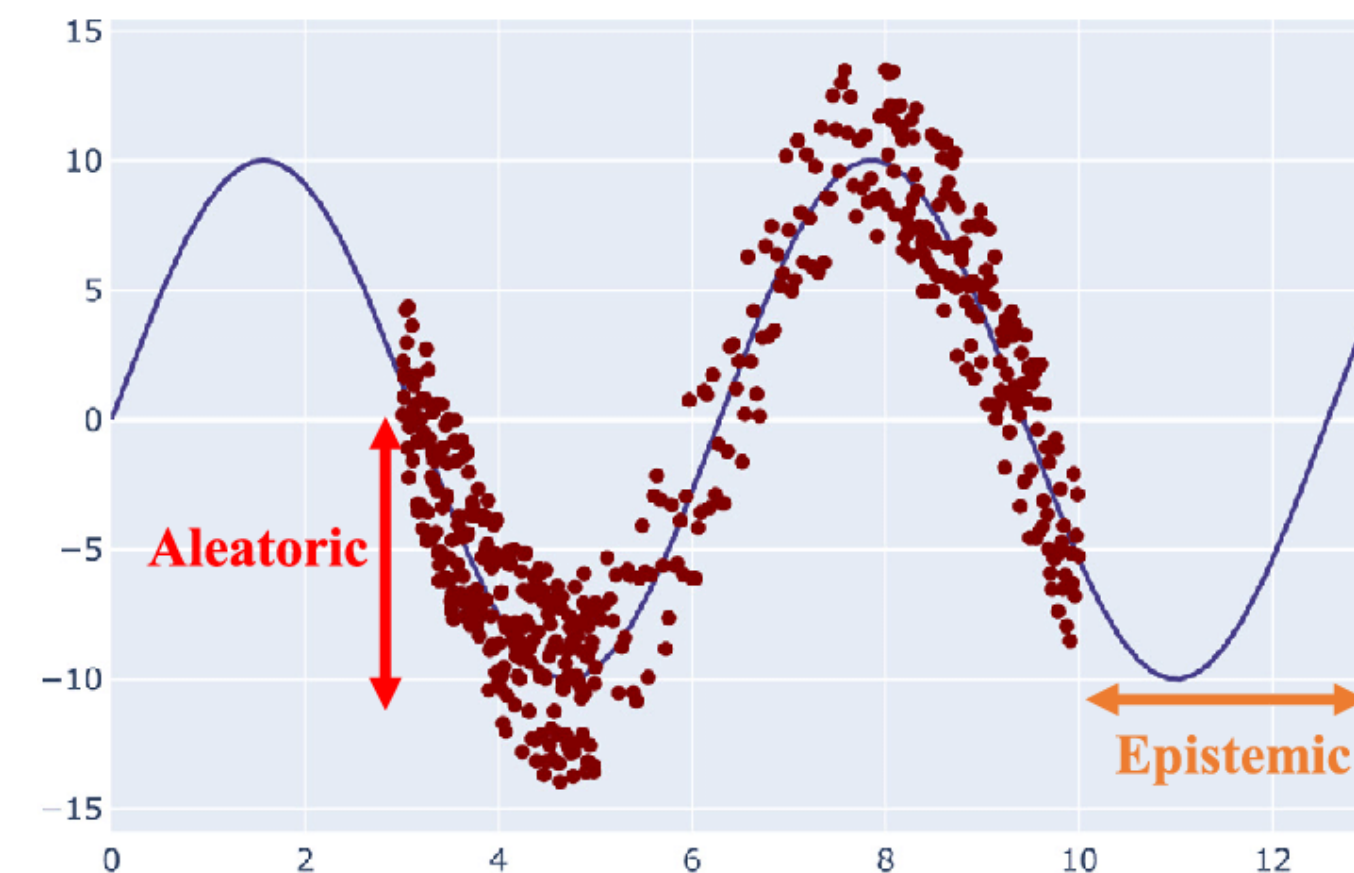


Best of all worlds: high accuracy & high precision

Who is talking about uncertainties?

Physicist	Why ?	How ?	ML researcher	Meaning
Statistical	Limited data	Poisson	 Aleatoric /Stochastic Epistemic/Systematic	Intrinsic noise, irreducible with more data
Systematic	Jet energy resolution	Gaussian		
Theorie	Higher order calculation	Flat?		Reducible with more knowledge/ data

... what you say vs what they hear ...



Heteroscedastic loss

Heteroscedastic = variance depends on input parameter

Traditional approach:

Fit with **known** uncertainties

New:

Learn **unknown** uncertainties

Direct consequence of likelihood approach

Maximize:

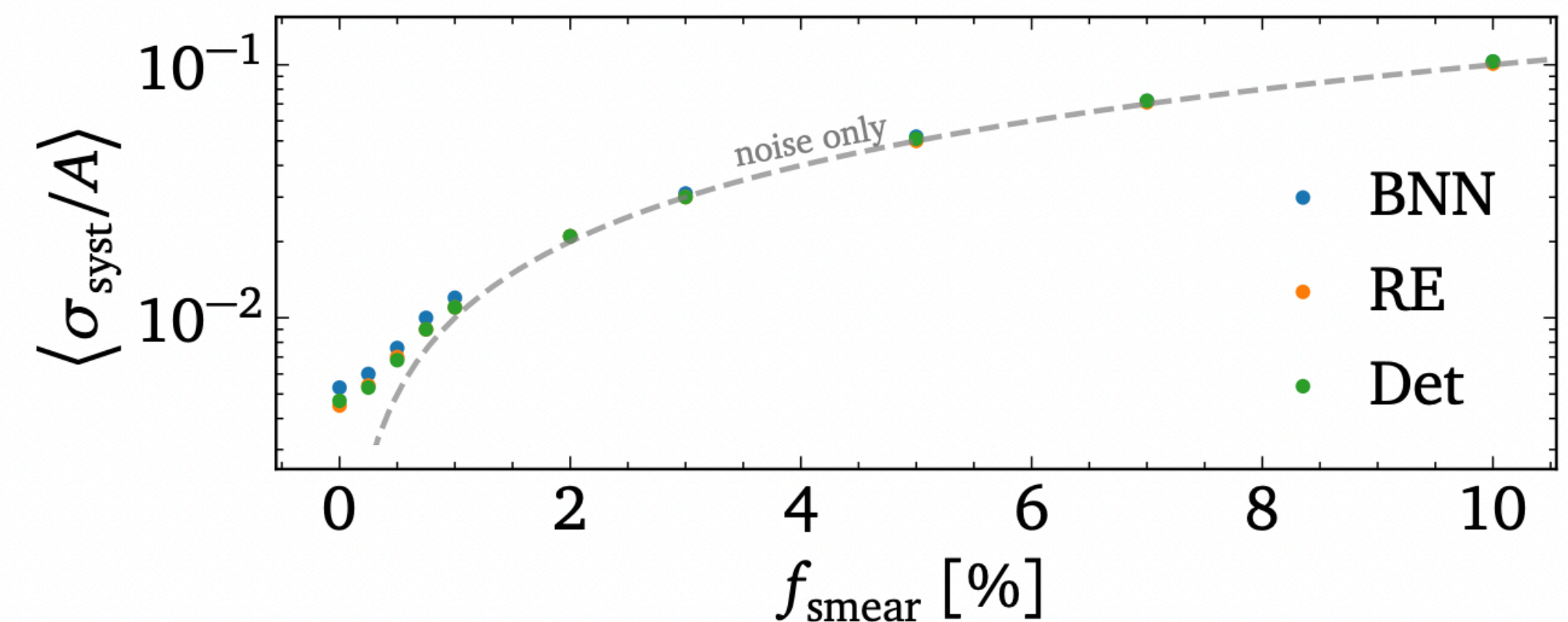
$$p(\theta | x_{\text{train}}) = p(x_{\text{train}} | \theta) \cdot p(\theta) \cdot \text{const}$$

Assume Gaussian likelihood

$$p(x_{\text{train}} | \theta) = \mathcal{N}(y_{\text{true}} | y_{\theta}, \sigma_{\theta})$$

$$-\log p(x_{\text{train}} | \theta) = \frac{|y_j - y_{\theta}(x_j)|^2}{2\sigma_{\theta}^2} + \log \sigma_{\theta}(x_j) + \text{const}$$

Add artificial noise to check calibration



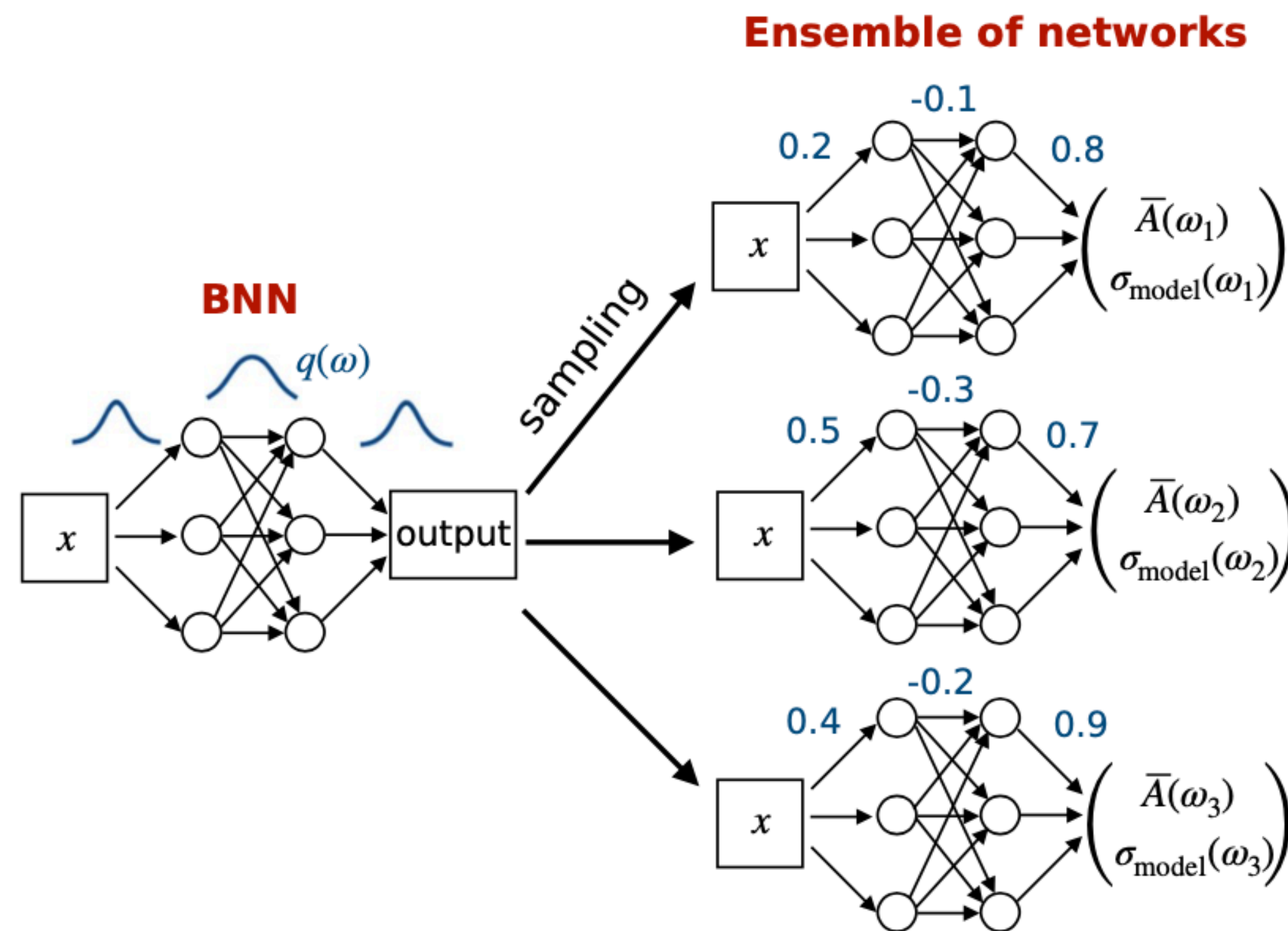
[2412.12069](#) H. Bahl, N. Elmer, L. Favaro, et al.

Det: correctly calibrated as long as noise dominates
Can not cover uncertainty on network parameters

Separating uncertainties with Bayesian networks

$$p(A) = \int d\omega \, p(A | \omega) p(\omega | T) \approx \int d\omega \, p(A | \omega) q(\omega)$$

$$\langle A \rangle(x) = \int d\theta \, q(\theta) A(x, \theta)$$



$$\begin{aligned} \sigma_{\text{sys}}^2 &= \int d\theta \, q(\theta) [\bar{A}^2(x, \theta) - \bar{A}(x, \theta)^2] \\ &= \int d\theta \, q(\theta) \sigma_{\text{model}}(\theta)^2 \end{aligned}$$

→ Captures noise, etc.

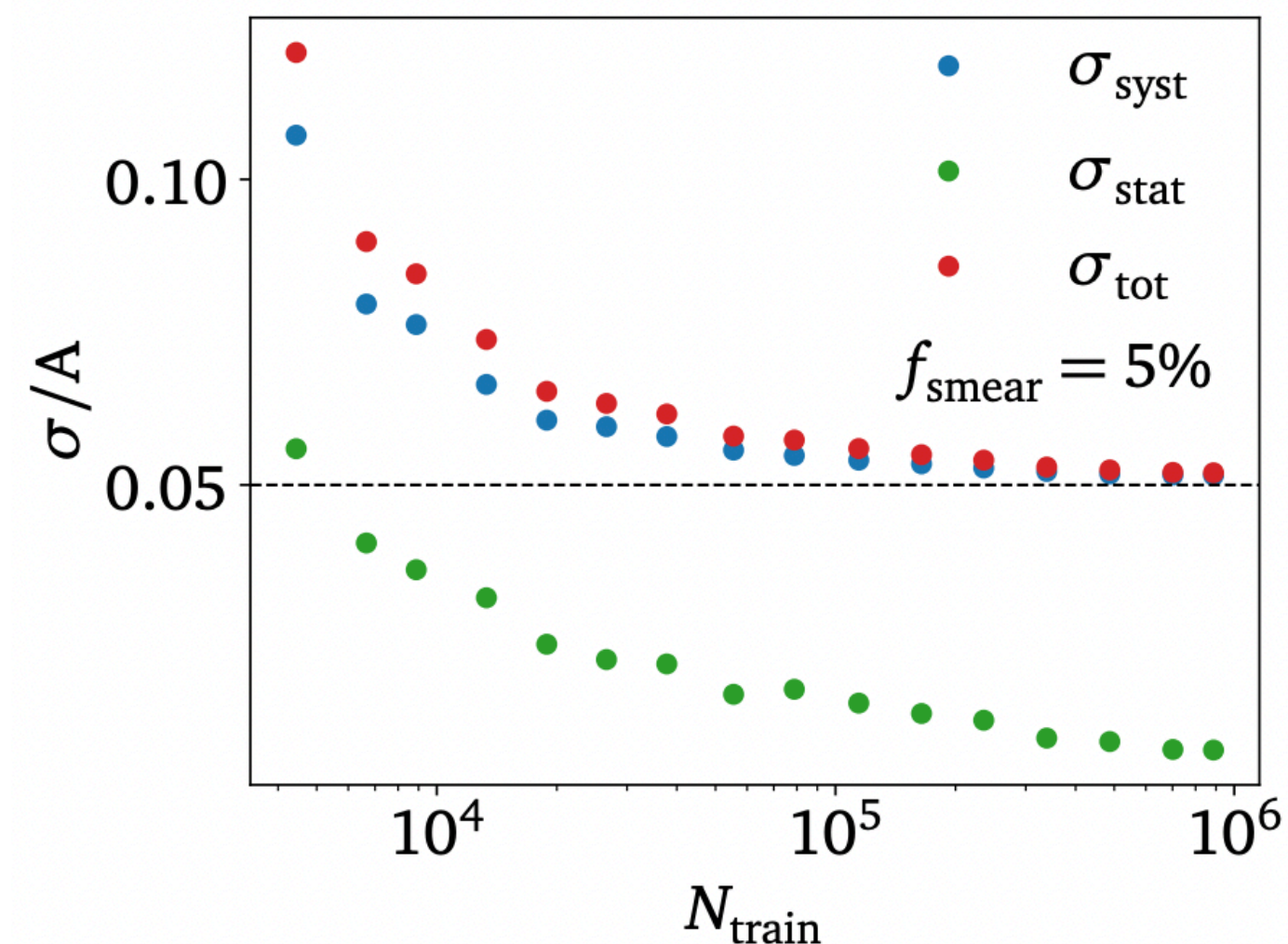
→ Plateaus for large training data

$$\sigma_{\text{stat}}^2 = \int d\theta \, q(\theta) [\bar{A}(x, \theta) - \langle A \rangle(x, \theta)]^2$$

→ Vanishes for identical weights
[= in the limit of infinite data]

Uncertainties from limited training data

[2412.12069](#) H. Bahl, N. Elmer, L. Favaro, et al.



$$\langle A \rangle(x) = \int d\theta \, q(\theta) A(x, \theta)$$

$$\begin{aligned} \sigma_{\text{syst}}^2 &= \int d\theta \, q(\theta) [\bar{A}^2(x, \theta) - \bar{A}(x, \theta)^2] \\ &= \int d\theta \, q(\theta) \sigma_{\text{model}}(\theta)^2 \end{aligned}$$

→ Captures noise, etc.

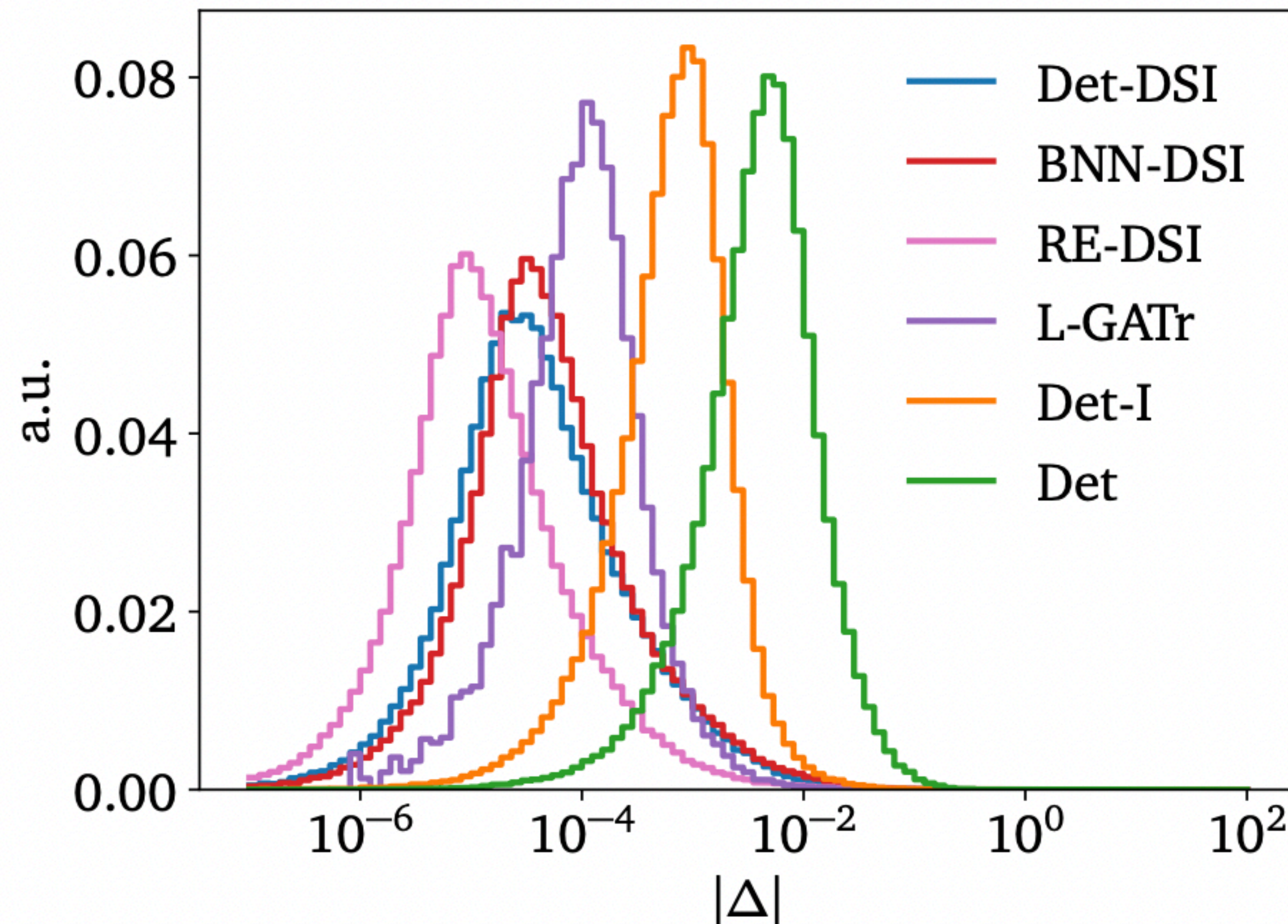
→ Plateaus for large training data

$$\sigma_{\text{stat}}^2 = \int d\theta \, q(\theta) [\bar{A}(x, \theta) - \langle A \rangle(x, \theta)]^2$$

→ Vanishes for identical weights
[= in the limit of infinite data]

What remains? Network size & symmetries (again)

[2412.12069](#) H. Bahl, N. Elmer, L. Favaro, et al.



Optimisation of network architecture

No noise

+

Large dataset

+

Include invariants

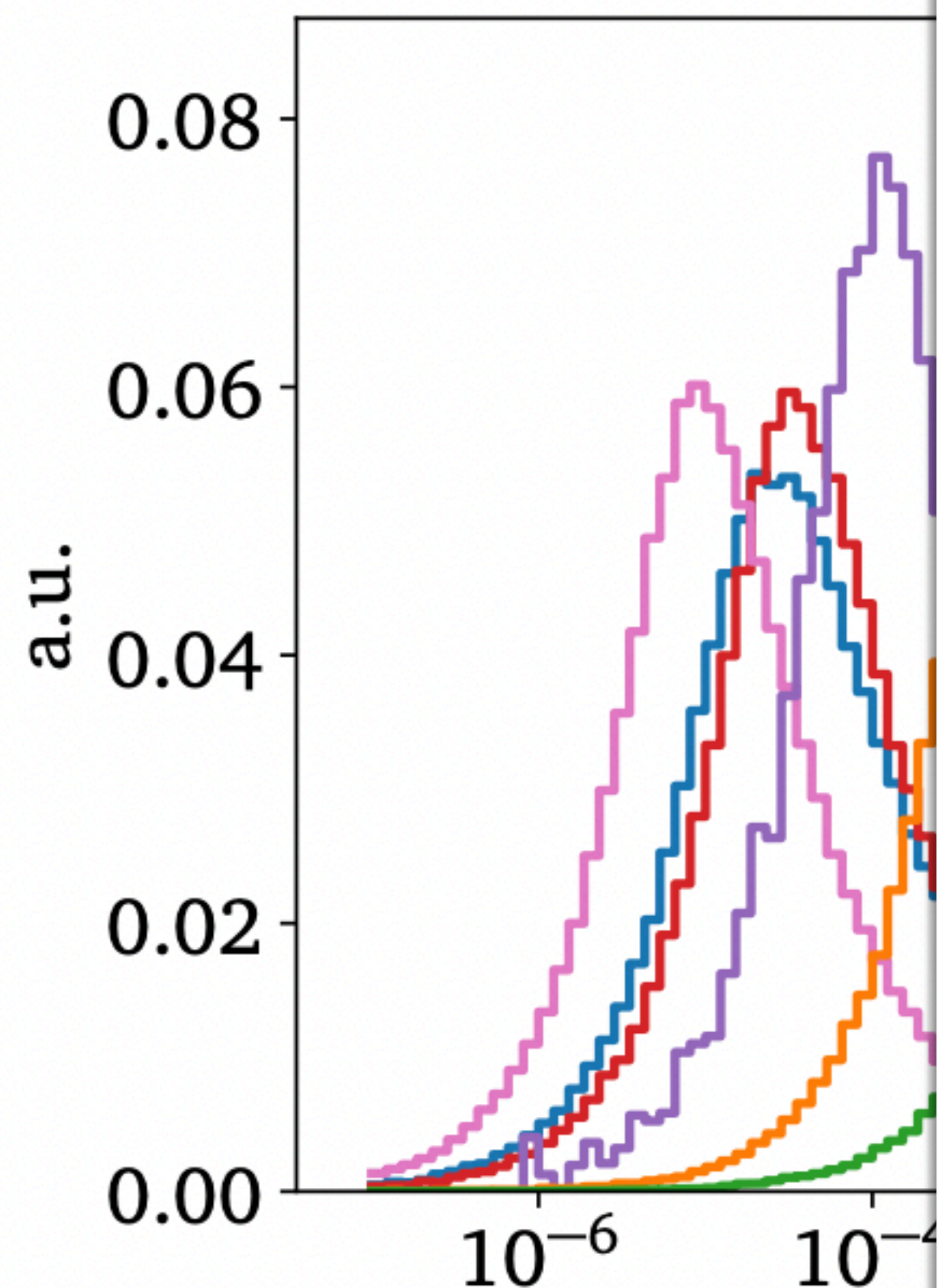
+

Deep sets architecture

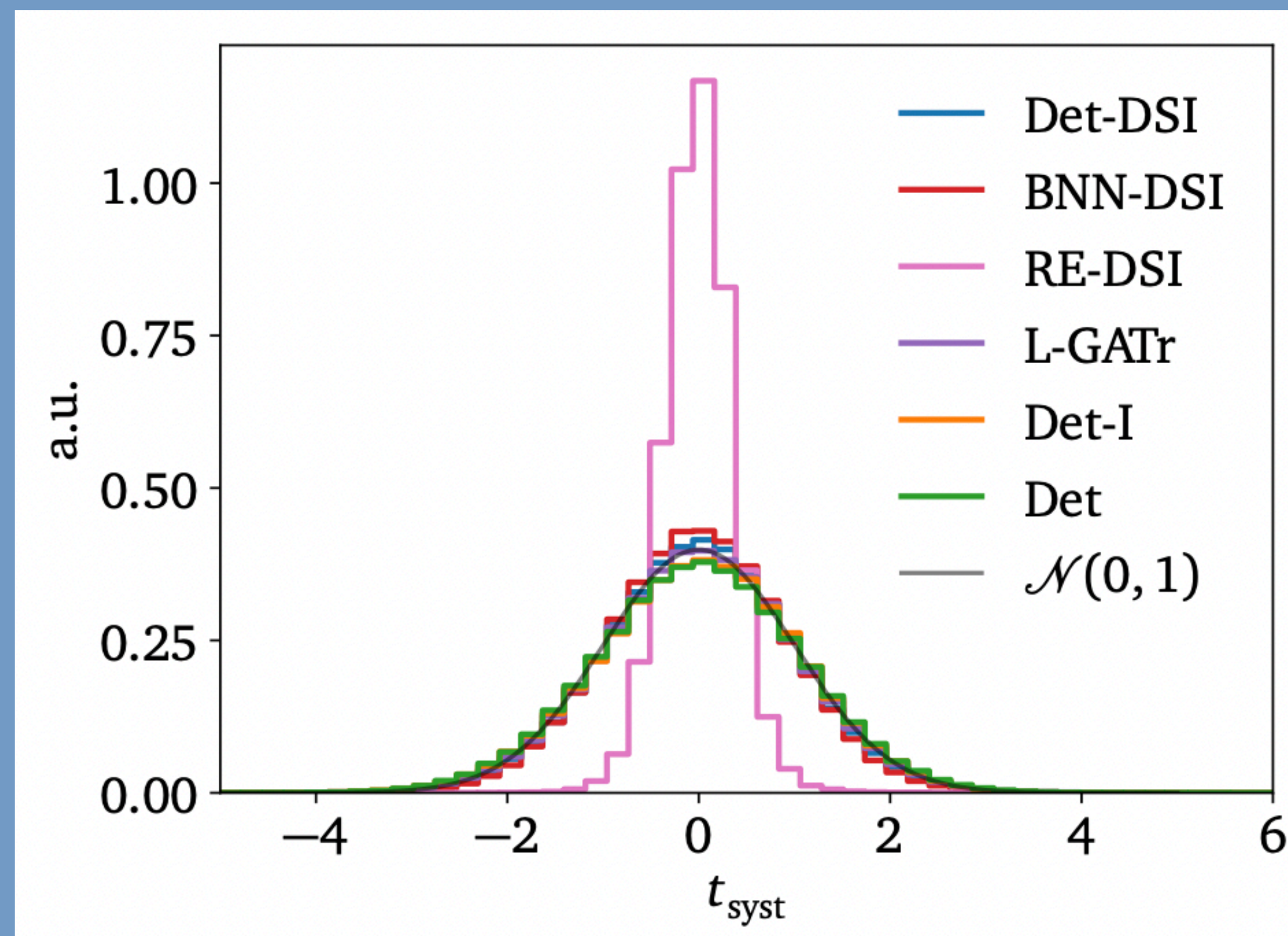
Reaching Precision of 10^{-5} !

What remains? Network size & symmetries (again)

[2412.12069](#) H. Bahl,



Calibrated!



network architecture

noise

dataset

invariants

architecture

precision of 10^{-5} !

1. Precision amplitudes

or

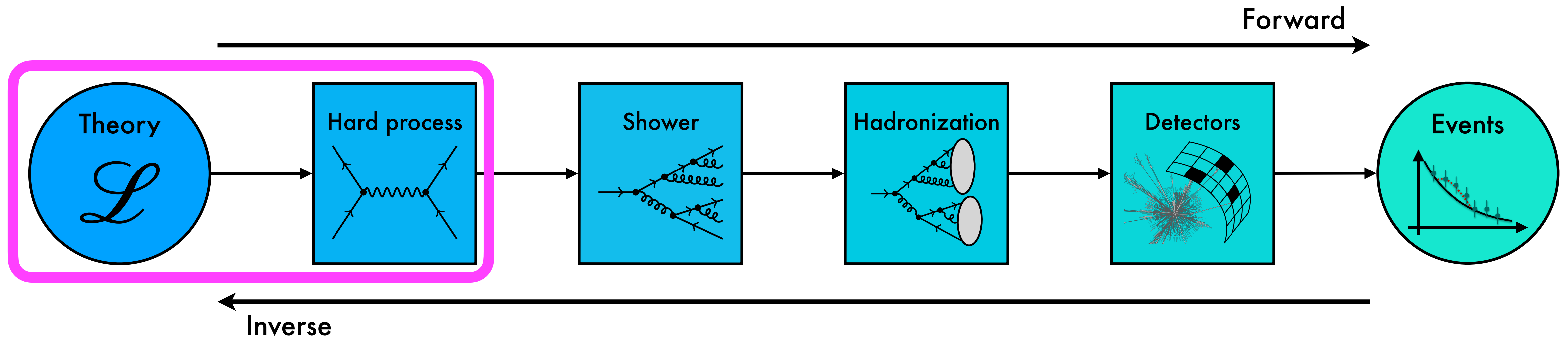
Quantifying uncertainties in regression problems

2. Event generation

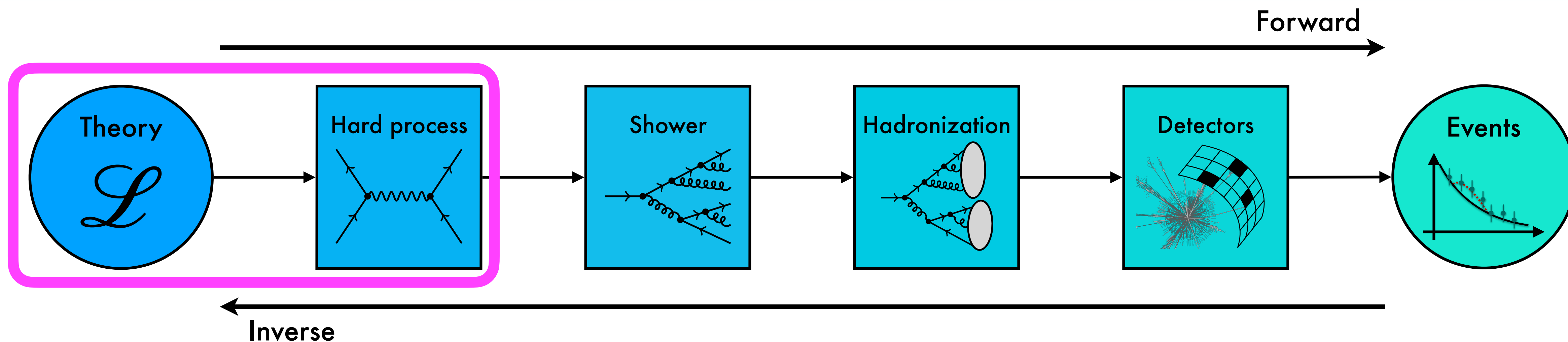
or

Controlling generative networks

Monte Carlo event generation



Monte Carlo event generation



1. Generate

phase space points

→ set of four-momenta p_i

2. Calculate event weight

$$w_{\text{event}} = f(x_1, Q^2) f(x_2, Q^2) \times |\mathcal{M}(x_1, x_2, p_1, \dots, p_n)|^2 \times J(p_i(r))$$

PDF

Matrix element

Phase space

3. Unweighting

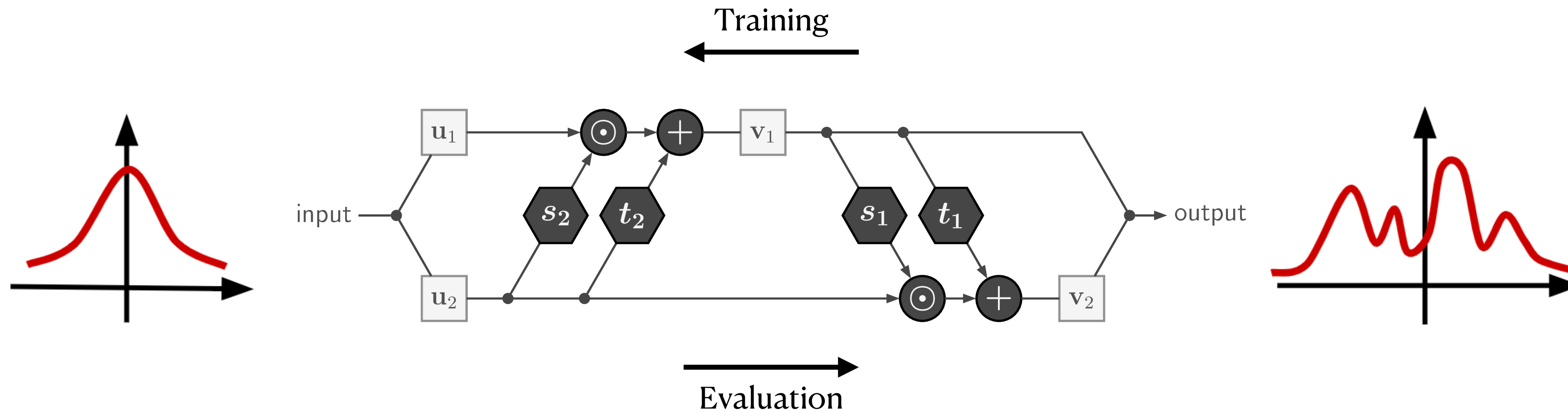
keep events with

$$\frac{w_i}{w_{\text{max}}} > r \in [0,1]$$

Normalizing flows

Invertible networks for complex transformations

- + Bijective mapping
- + Tractable Jacobian $\rightarrow p_x(x) = p_z(z) \cdot J_{NN}$
- + INN $\rightarrow$ flow with fast evaluation in both direction



Training on density $t(x)$
 $\rightarrow$ Minimize difference

$$\begin{aligned}\mathcal{L} &= \log p_x(x)/t(x) \\ &= \log p_z(z(x)) J_{NN} / t(x)\end{aligned}$$

Requires evaluation of $t(x)$

Training on samples x
 $\rightarrow$ Maximize the log-likelihood

$$\begin{aligned}\mathcal{L} &= \log p(\theta | x) \\ &= \log p(z | \theta) + \log J_{NN} + p(\theta) + \text{const}\end{aligned}$$

Normalizing flows

Invertible networks for complex transformations

- + Bijective mapping
- + Tractable Jacobian $\rightarrow p_x(x) = p_z(z) \cdot J_{NN}$
- + Fast evaluation in both direction

Optimized training strategy in MadNIS
[2212.06172, 2311.01548, 2408.01486]
implemented in MadGraph!

Evaluation

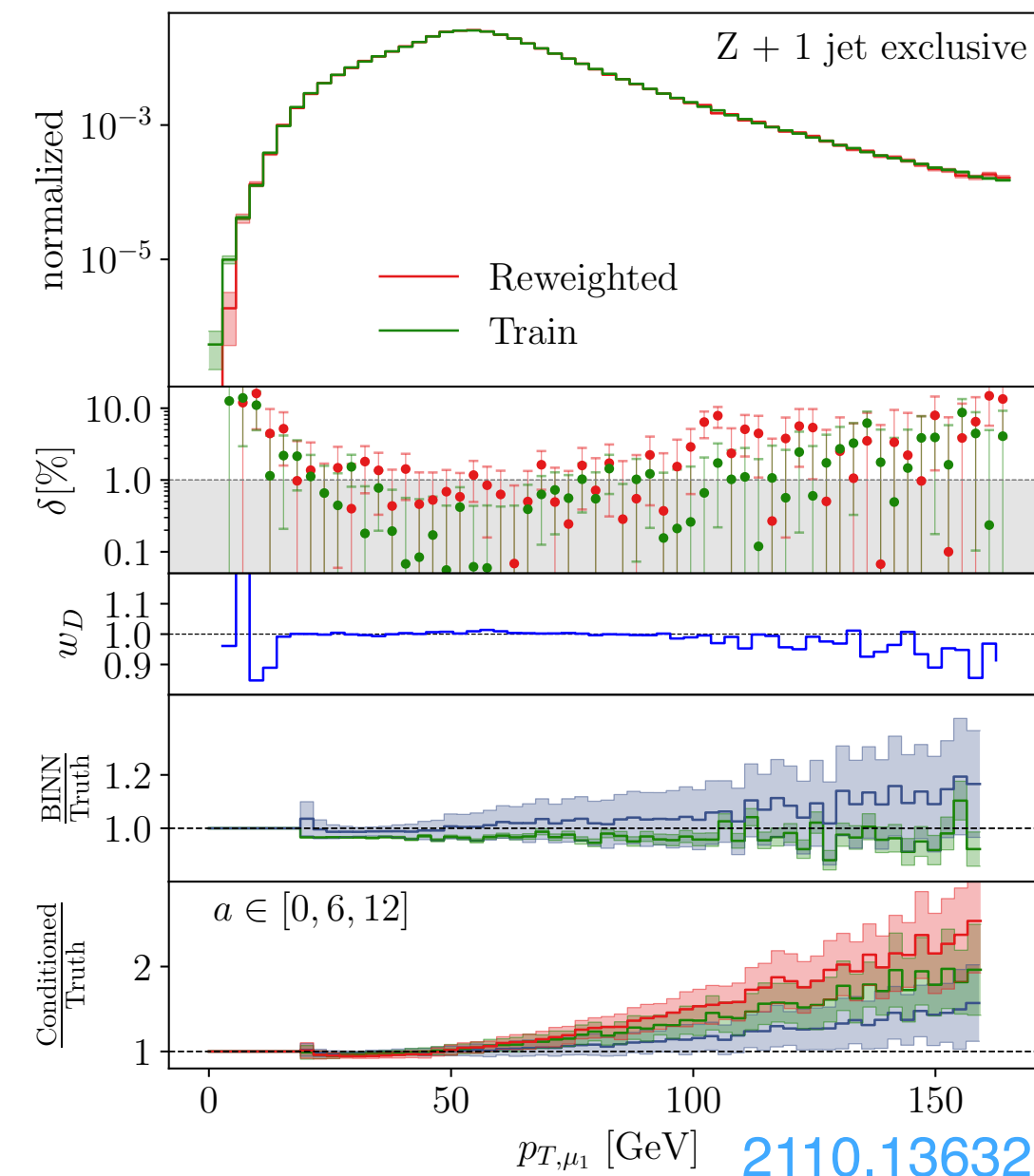
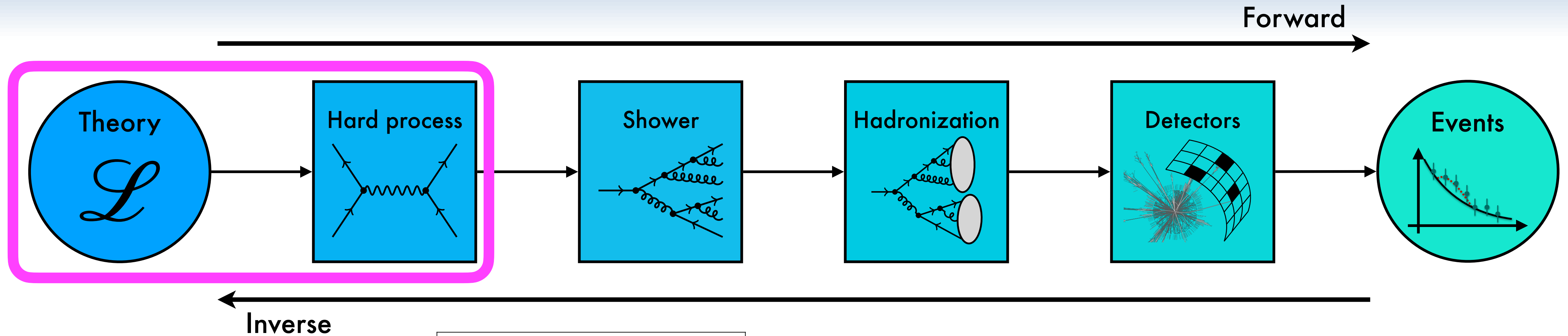
Training on density $t(x)$
 $\rightarrow$ Minimize difference

$$\begin{aligned}\mathcal{L} &= \log p_x(x)/t(x) \\ &= \log p_z(z(x)) J_{NN} / t(x)\end{aligned}$$

Training on samples x
 $\rightarrow$ Maximize the log-likelihood

$$\begin{aligned}\mathcal{L} &= \log p(\theta | x) \\ &= \log p(z | \theta) + \log J_{NN} + p(\theta)\end{aligned}$$

Monte Carlo event generation



- Basis: INN, diffusion networks
 - Phase space symmetries in architecture

- **Control** via classifier D

$$\frac{p_{\text{truth}}(x)}{p_{\text{INN}}(x)} = \frac{D(x)}{1 - D(x)}$$

← **ACCURACY**

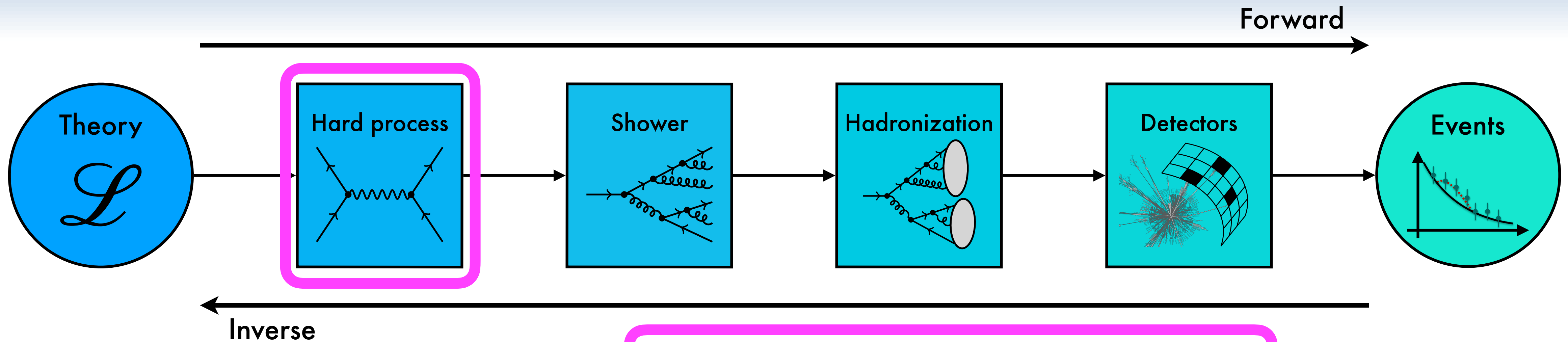
- **Accuracy** via reweighting

➡ **Uncertainty estimation** via Bayesian NN

← **PRECISION**

➡ **Uncertainty propagation** via conditioning

Monte Carlo event generation

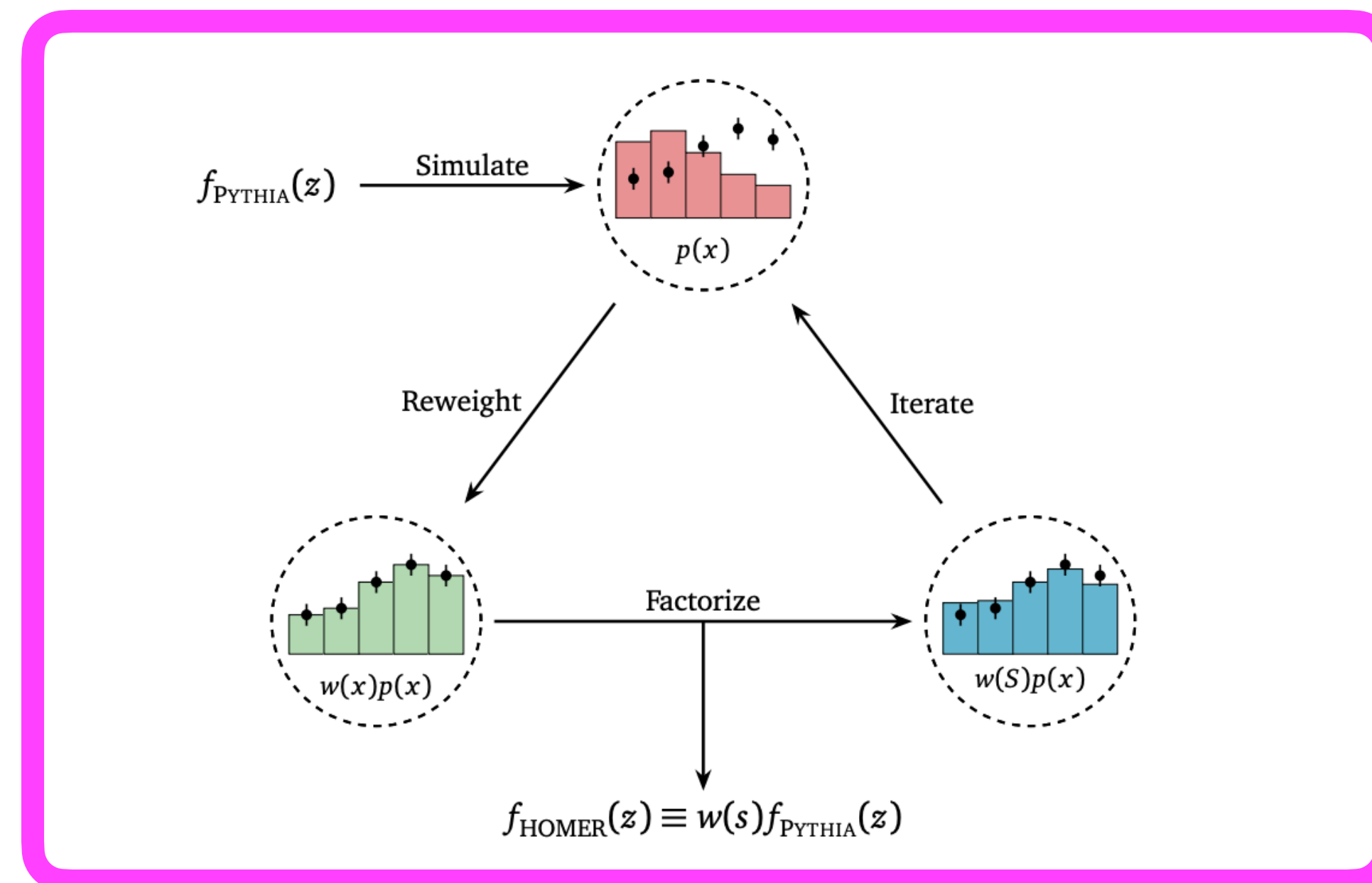


Optimal fragmentation function:

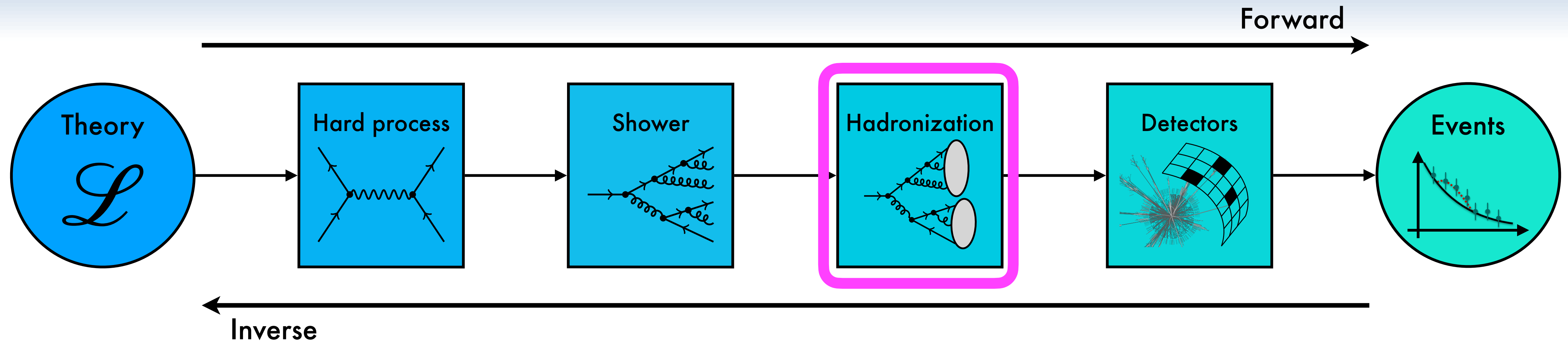
Reweighting PYTHIA to match data

[2509.03592] AB, A. Ore, S. Palacios

Schweitzer, et al.



Monte Carlo event generation



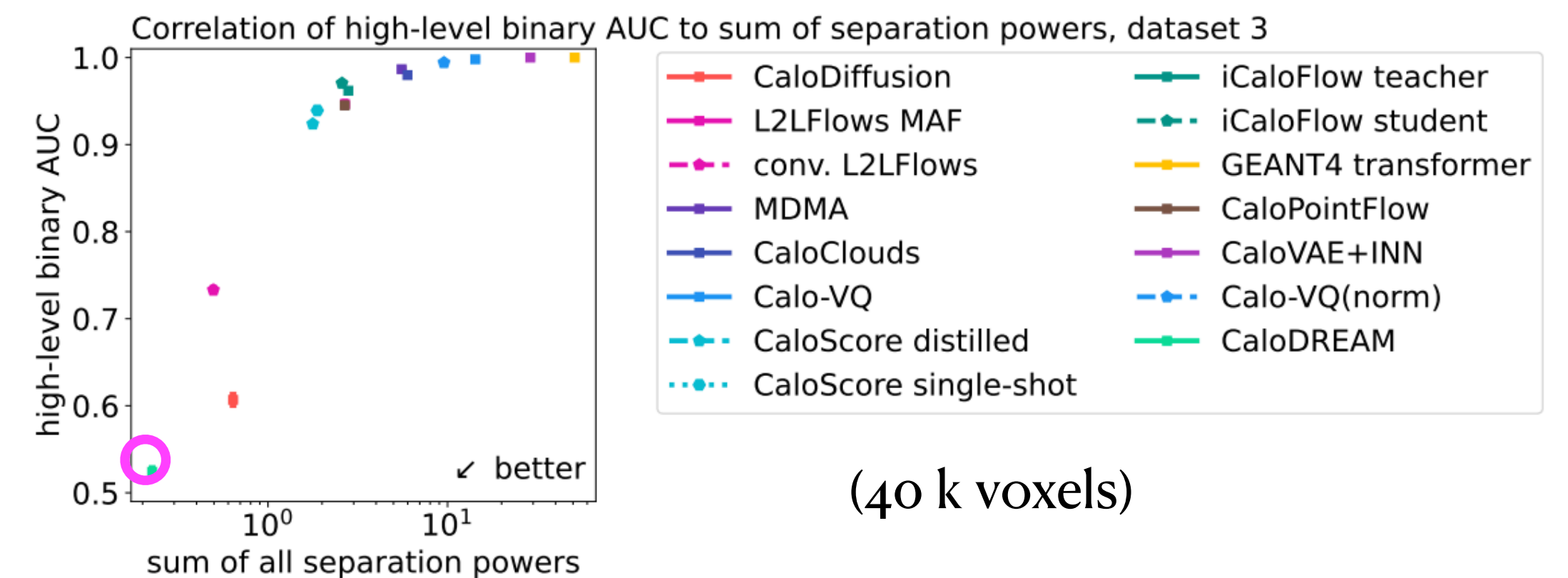
Efficient detector simulations:

- Transfer concept from phase space generation
- Same networks (diffusion + transformers)
- Challenge: high dimensional

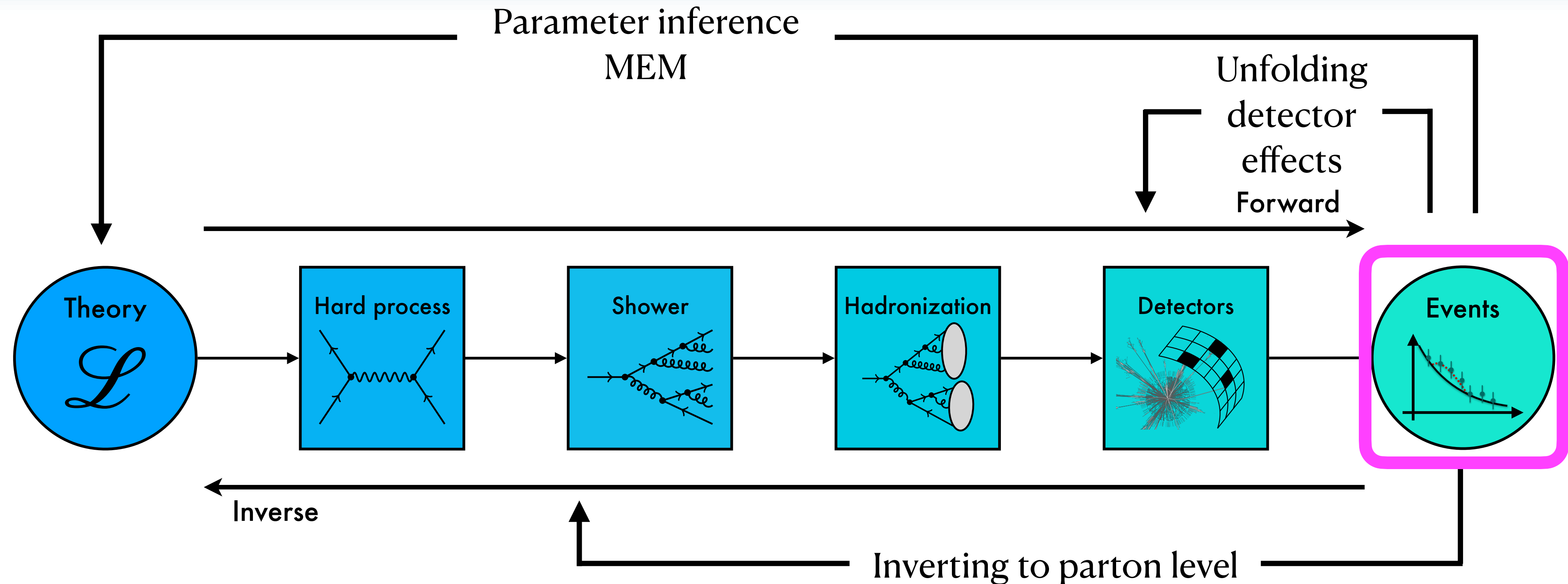
CaloChallenge 2022 [2410.21611]

CaloDREAM [2405.09629]

L. Favaro, A. Ore, S. Palacios Schweitzer, T. Plehn



Inverting the simulation chain



Requirements

- ☐ High - dimensional
- ☐ Bin - independent
- ☐ Statistically well defined

1. Precision amplitudes

or

Quantifying uncertainties in regression problems

2. Event generation

or

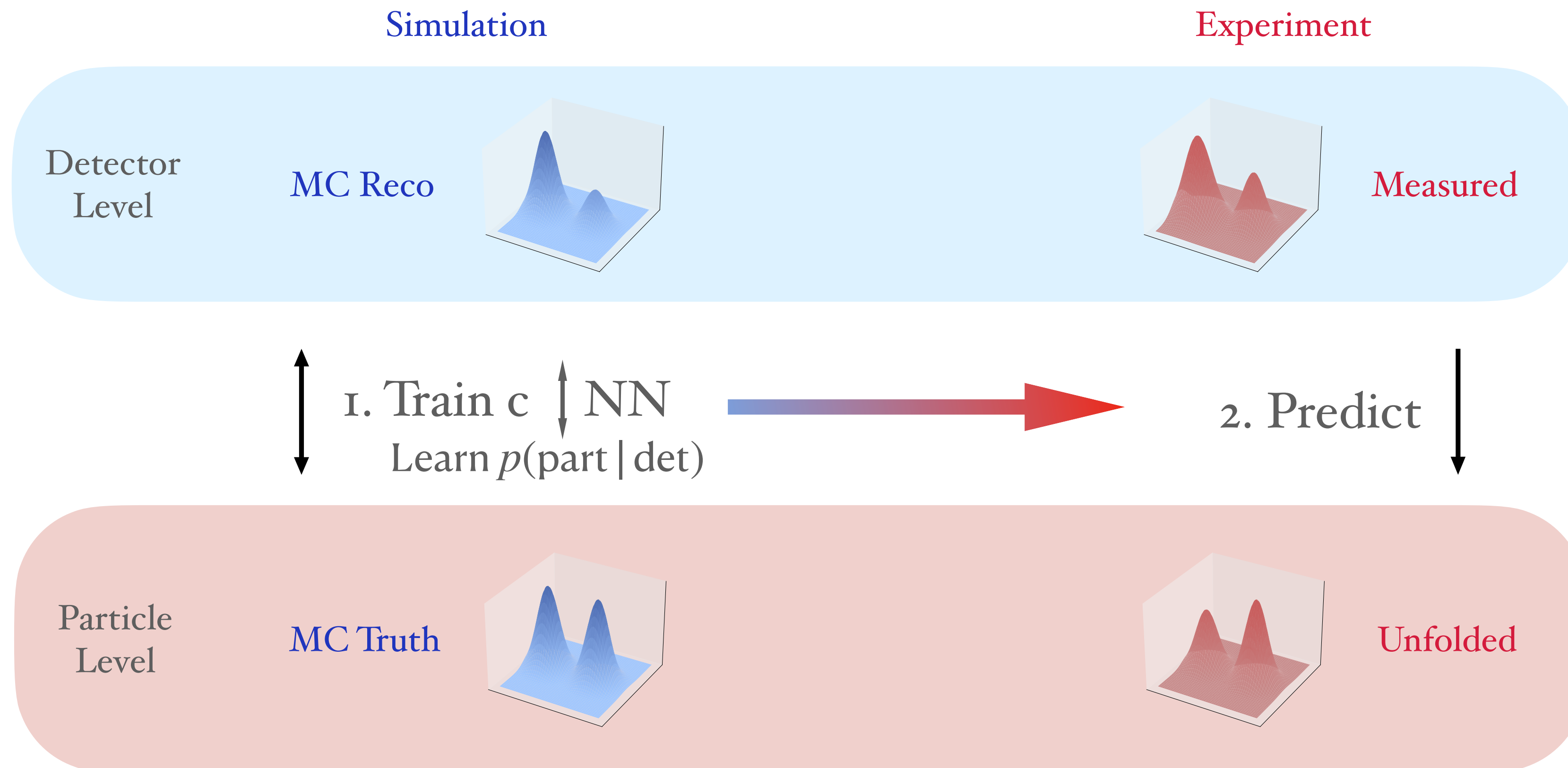
Controlling generative networks

3. Unfolding new physics

or

Overcoming bias in inverse problems

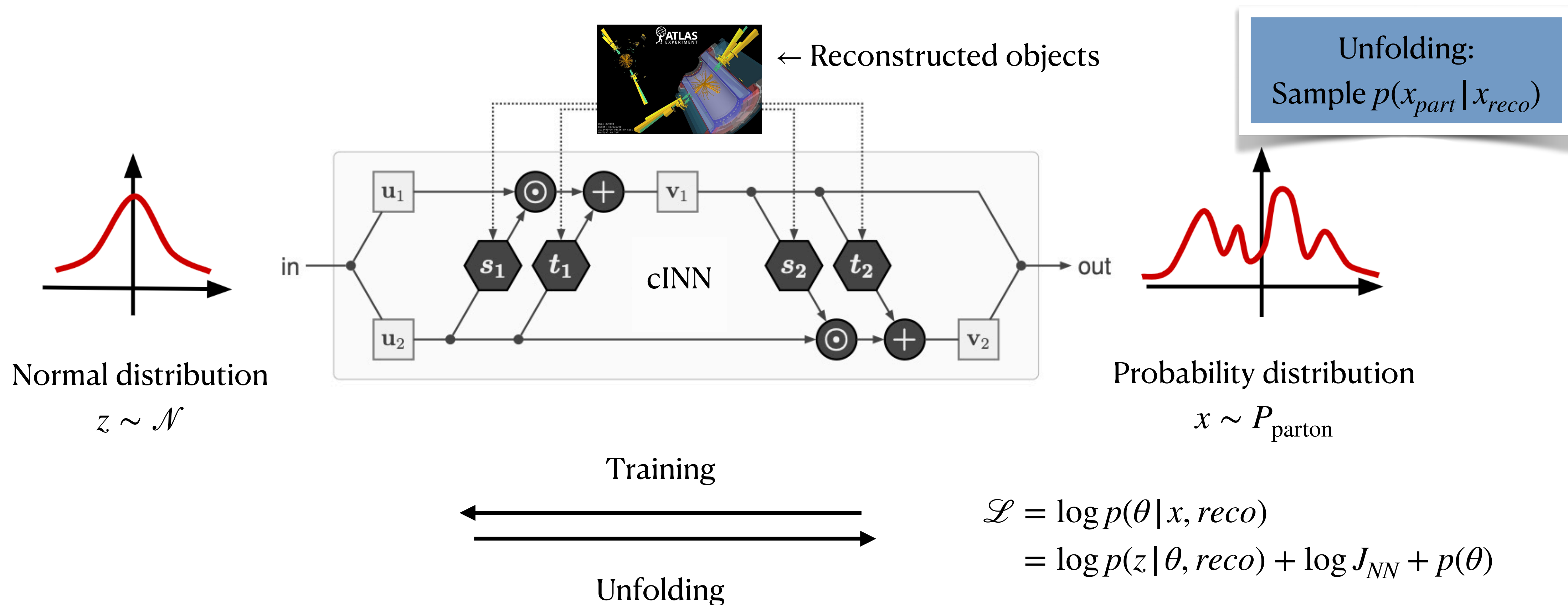
Generative unfolding methods



Generative unfolding

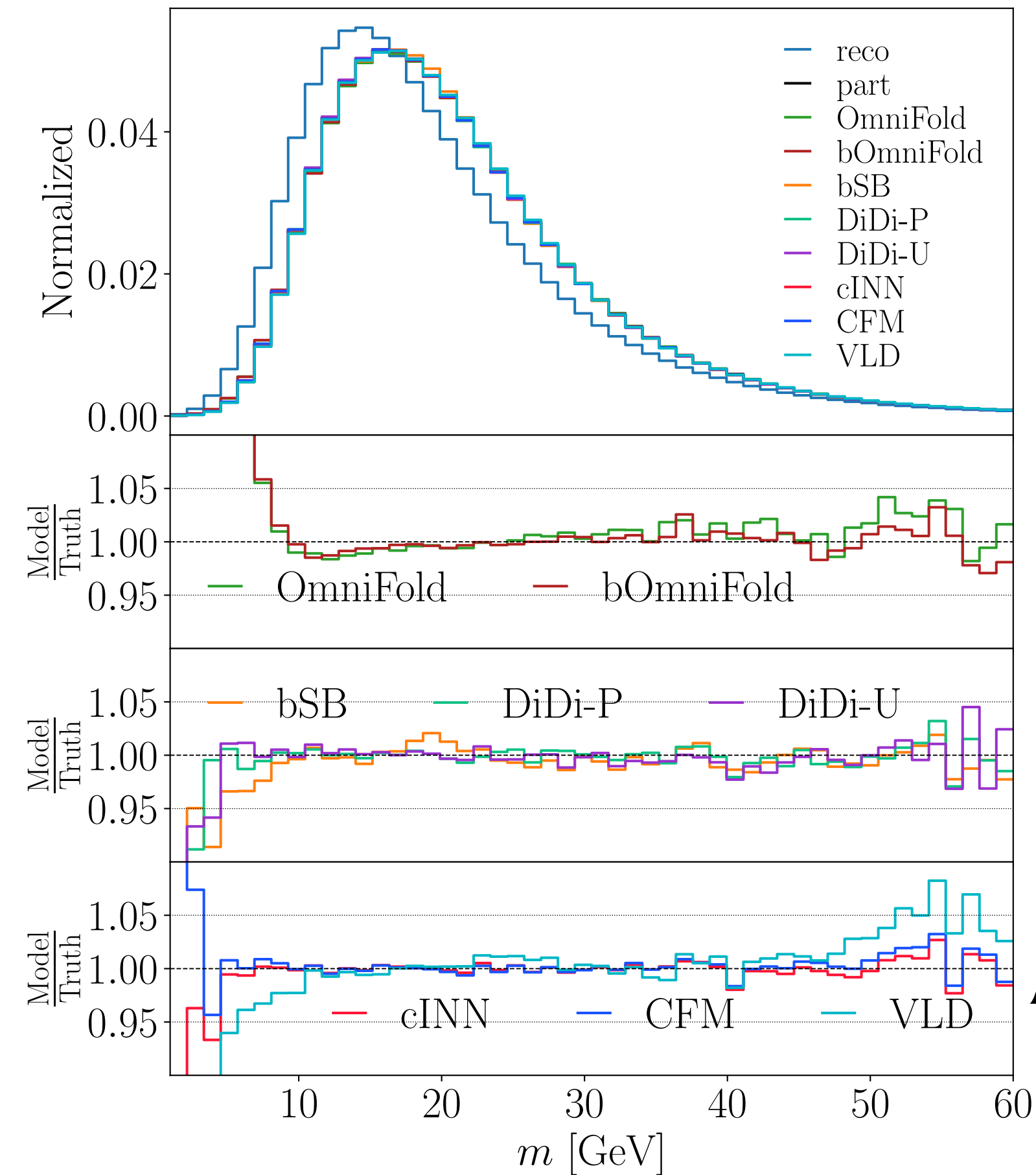
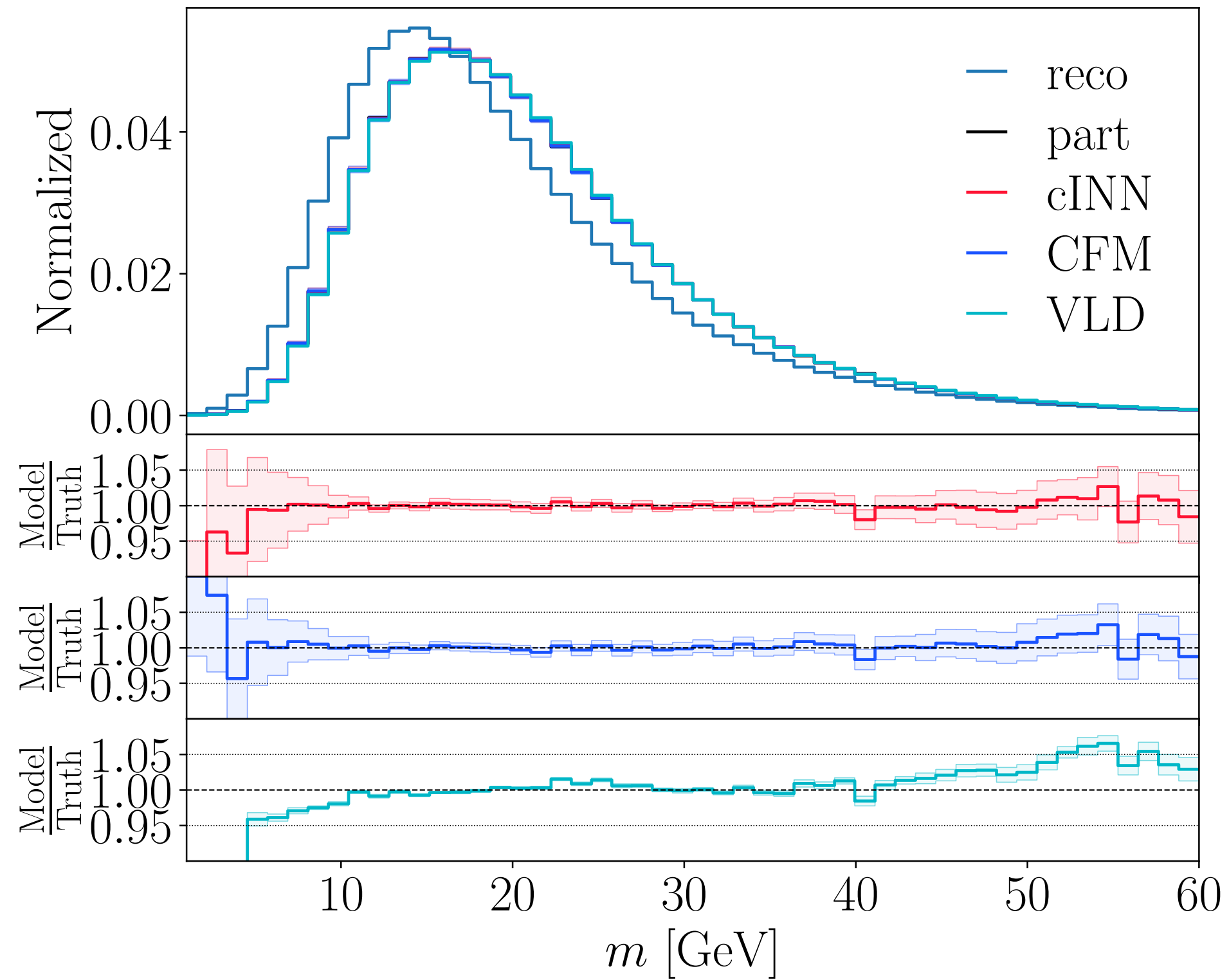
High-dimensional. Bin independent. Robust.

Given a reconstructed event:
What is the probability distribution at particle level?



Unfolding Z+jets events

Observables $m, \tau_{21}, w, N, \log \rho, z_g$



[1911.09107]

A. Andreassen, et al.

[2411.02495]

AB, S. Diefenbacher, et al.

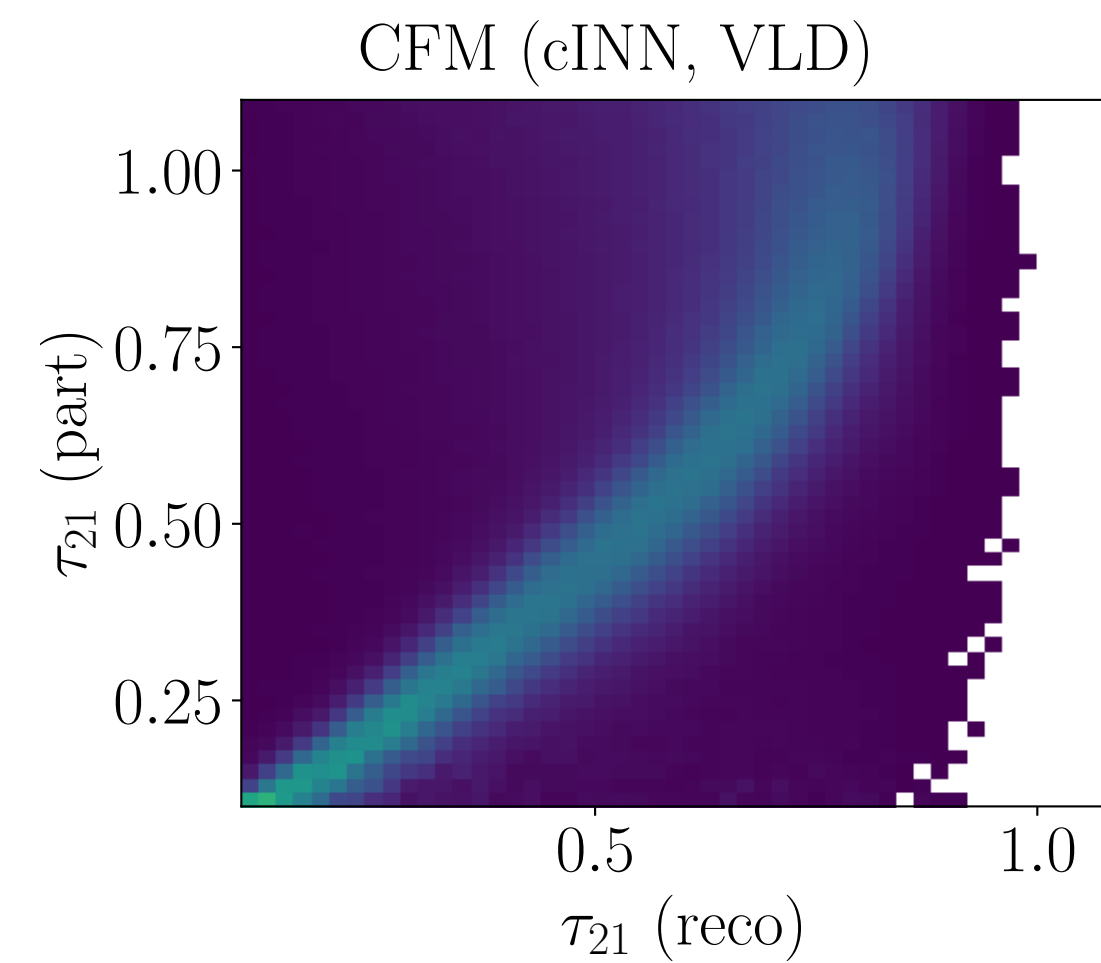
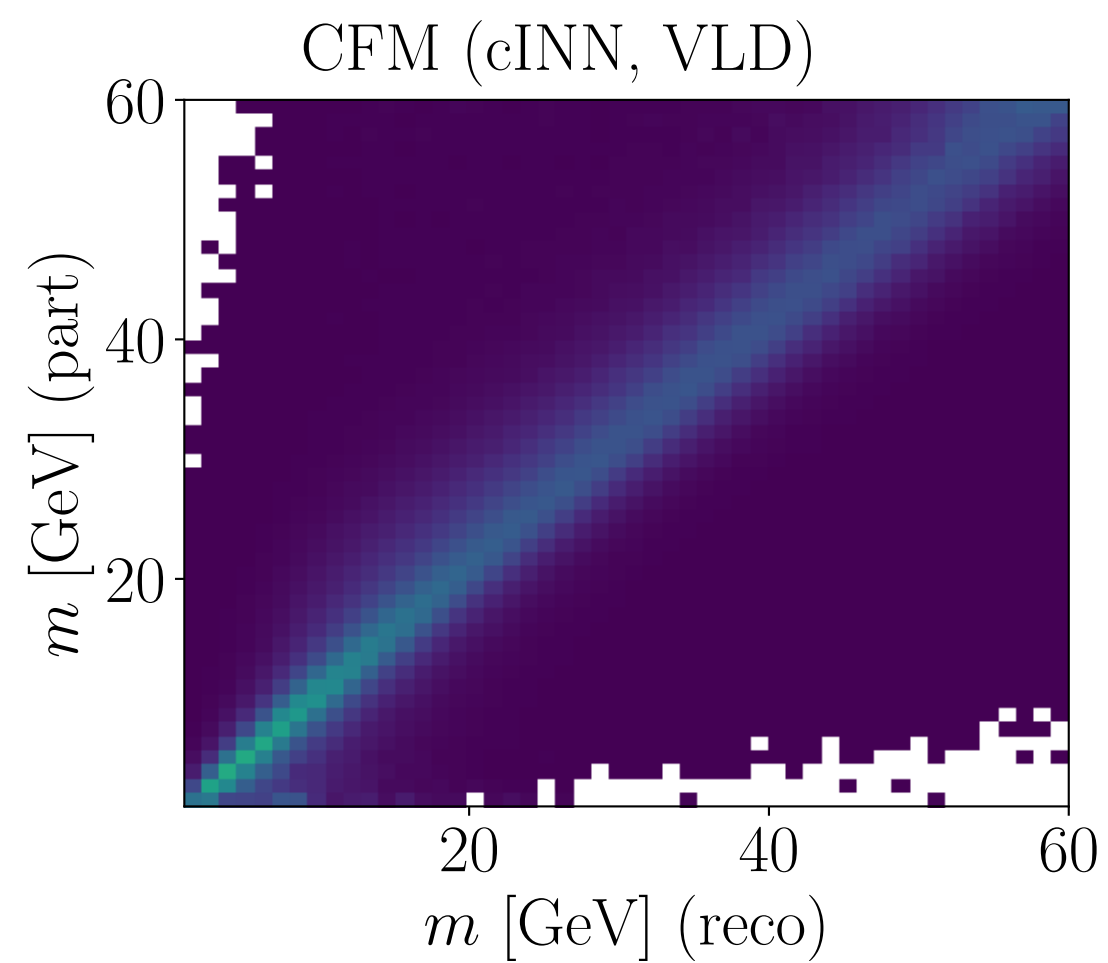
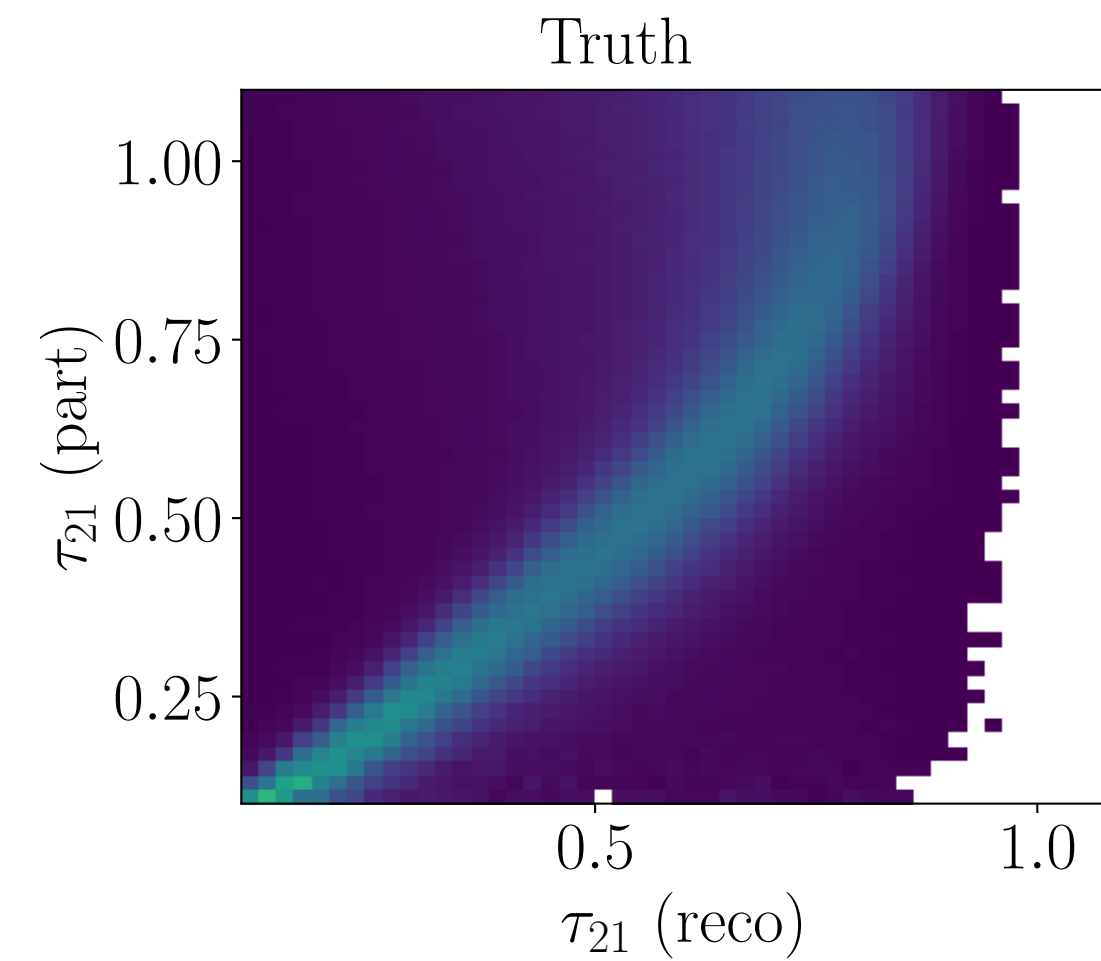
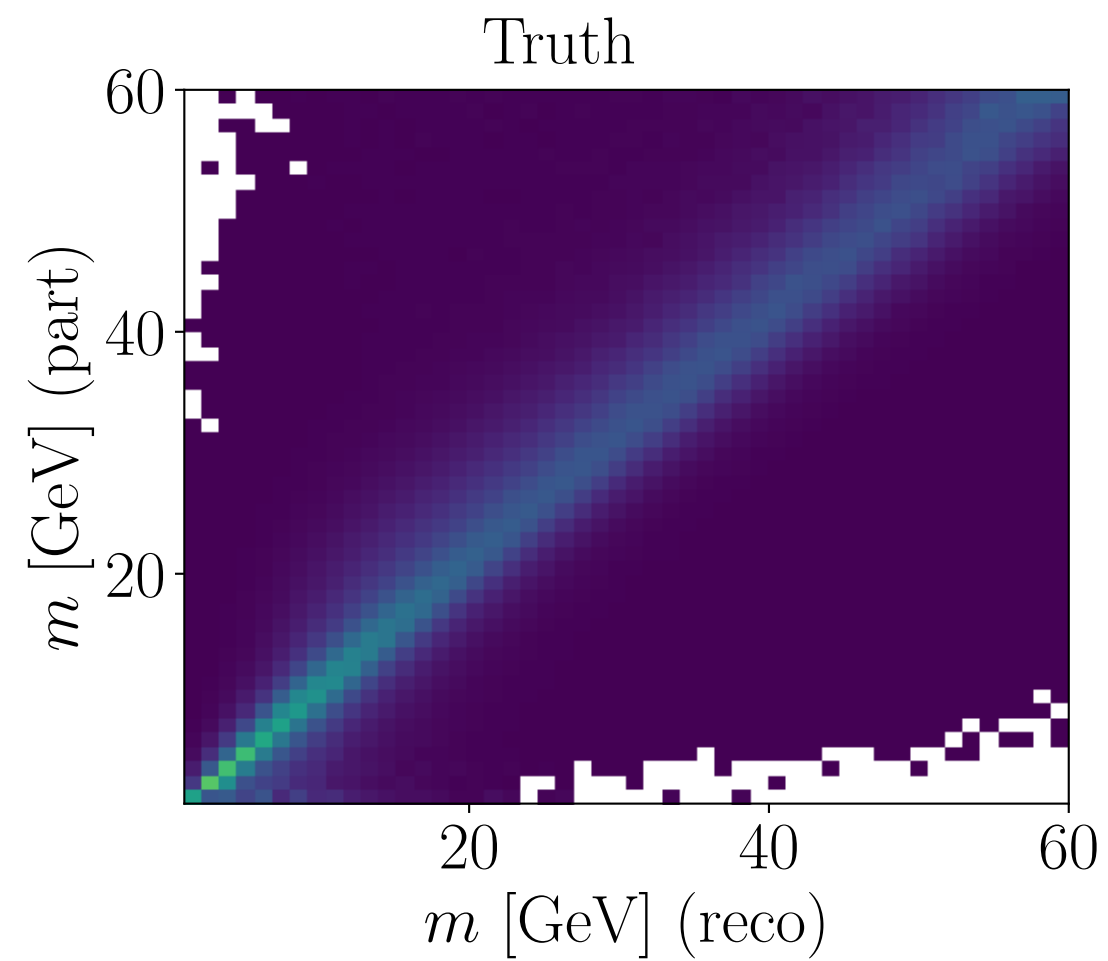
VLD: [2305.10399]

A. Shmakov, K. Greif, et al.

The Landscape of Unfolding with ML [2404.18807] N. Hütsch, et al.

Correlations

Z+jets: reco vs particle level, jet mass & subjettiness ratio

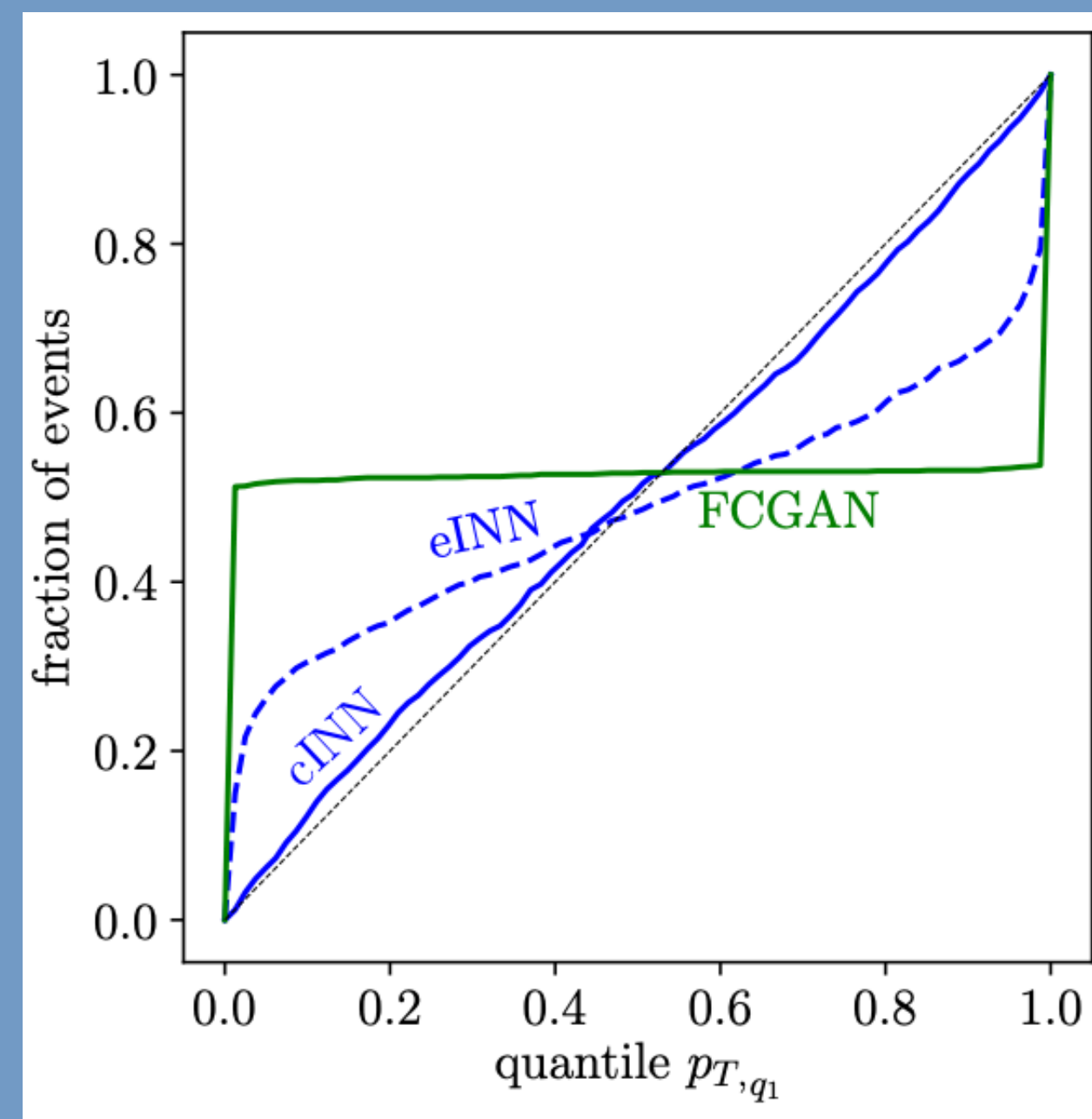


[2404.18807] N. Hütsch, et al.

Correlations

Calibrated!

[2006.06685] M. Bellagente, AB, G. Kasieczka, et al



[2404.18807] N. Hütsch, et al.

Beyond unfolding: Enabling the MEM

[2210.00019, 2310.07752]

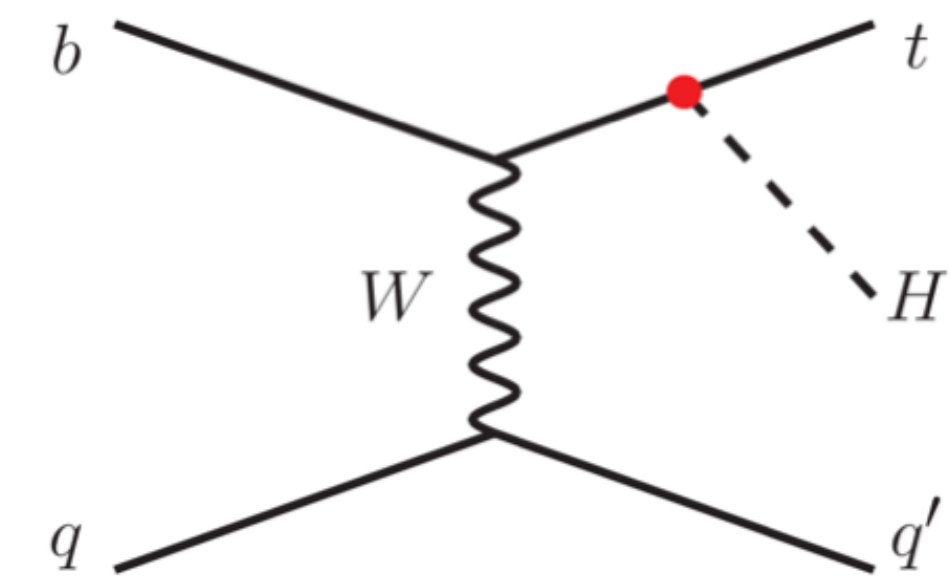
Matrix element method is based on untractable likelihood

$$p(x_{\text{reco}}|\alpha) = \int dx_{\text{hard}} \underbrace{p(x_{\text{hard}}|\alpha)}_{\text{diff. CS}} \underbrace{p(x_{\text{reco}}|x_{\text{hard}}, \alpha)}_{\text{estimate with network}}$$

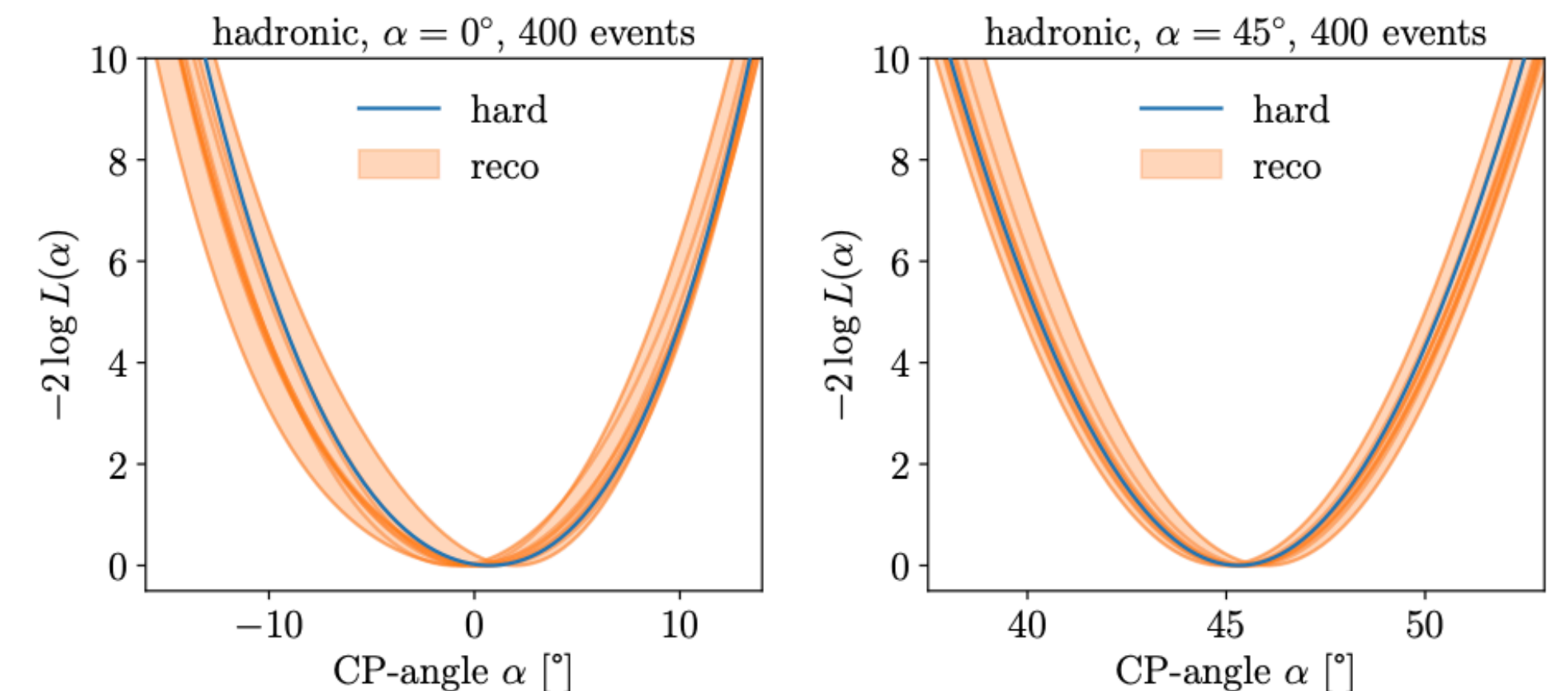
Problem: integration over full phase space of the hard scattering

Solution: Use unfolding cINN to sample x_{hard}

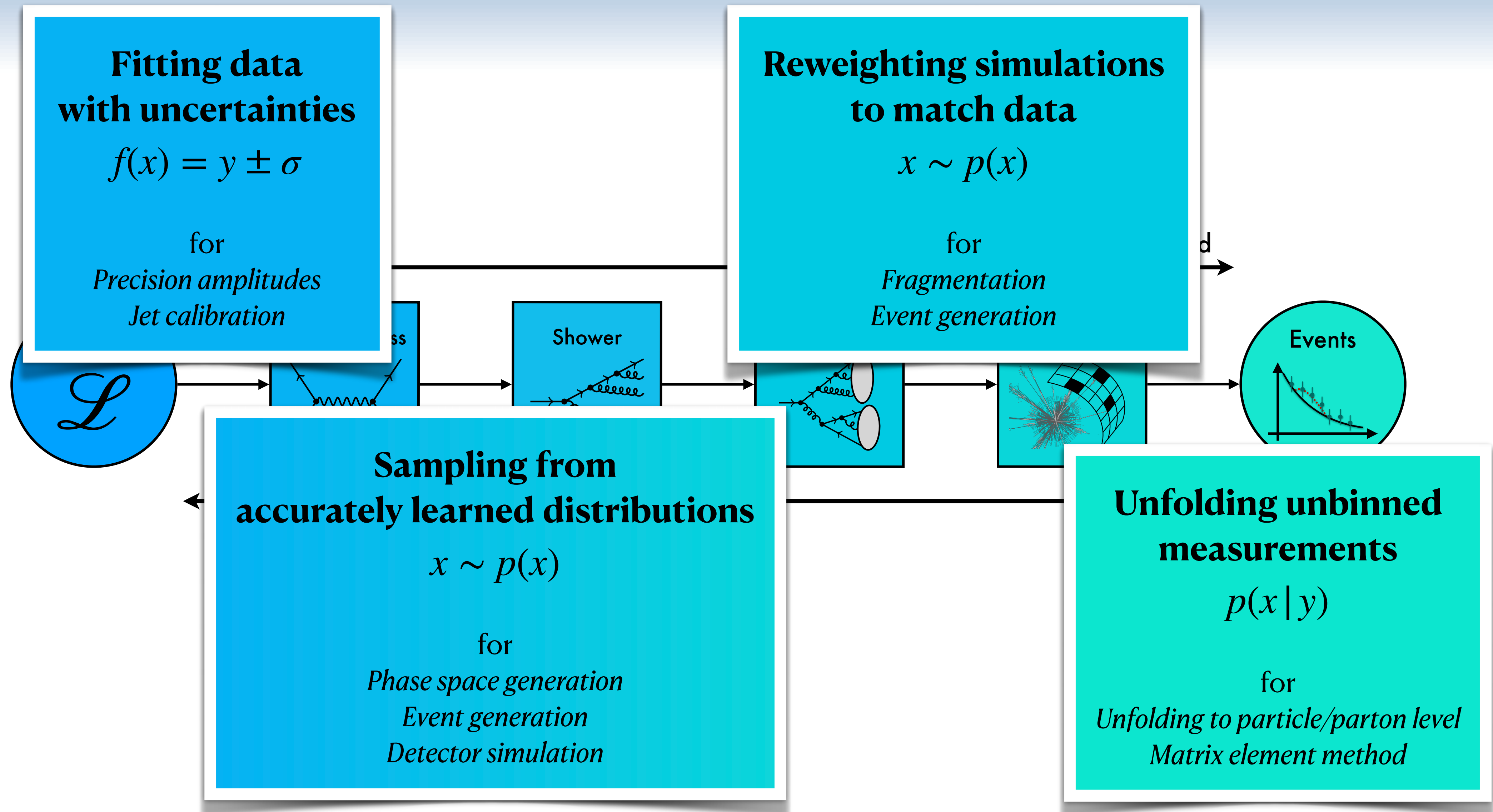
$$p(x_{\text{reco}}|\alpha) = \left\langle \frac{1}{q(x_{\text{hard}})} p(x_{\text{hard}}|\alpha) p(x_{\text{reco}}|x_{\text{hard}}, \alpha) \right\rangle_{x_{\text{hard}} \sim q(x_{\text{hard}})}$$



Single Higgs production
with anomalous non-CP conserving Higgs coupling



ML enables



What's next?

Tilman Plehn^{1,2}

1 Institut für Theoretische Physik
2 Interdisciplinary Center for Science and Technology

Abstract

We uncover an... con...
Agentic installation...
state-of-the-art simulations and ac...
GENTS interact with inexperienced...
tasks, and analyze results. In a se...
event generation and run an auto...
of a paper.

<https://github.com/heidelberg-hep>

Resummation of the C-Parameter Sudakov Shoulder Using Effective Field Theory

Matthew D. Schwartz^{1,2}

¹Department of Physics, Harvard University, Cambridge, MA 02138, USA
²Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

schwartz@physics.harvard.edu

AI RESEARCH ASSISTANT Claude Opus 4.5 (Anthropic)

February 25, 2025

Author Contribution

M.D.S. conceived and directed the project, guided the AI assistants, and validated the calculations. Claude Opus 4.5, an AI research assistant developed by Anthropic, performed all calculations including the SCET factorization theorem derivation, one-loop soft and jet function calculations, EVENT2 Monte Carlo simulations, numerical analysis, figure generation, and manuscript preparation. The work was conducted using Claude Code, Anthropic's agentic coding tool. M.D.S. is fully responsible for the scientific content and integrity of this paper.

Benchmark (TPBench) and Study of Capabilities in Physics

^{1,2}, Yuhang Yao^{1,2}, Michael Münchmeyer^{1,5}, Maja Federer^{1,2}, and Ananya Adipal⁴

¹University of Wisconsin-Madison
²University of Wisconsin-Madison
³University of Wisconsin-Madison
⁴Indiana University, Bloomington
⁵Institute for the Sky (SkAI), Chicago

February 25, 2025

Model Performance (Average of 5 shots)

Difficulty Level	Kimi-K2-Thinking	GPT-5 (high)	Grok-4	Claude-Opus-4.7 (max)	GPT-5.5 (xhigh)	o3 (high)	Gemini-3 Pro	Deepseek-V4-Pro
1-Easy Undergrad	76	78	84	91	91	91	93	100
2-Undergrad	59	64	70	75	75	75	75	75
3-Easy Grad	17	25	31	41	42	44	44	55

Automated

AI research assistant?

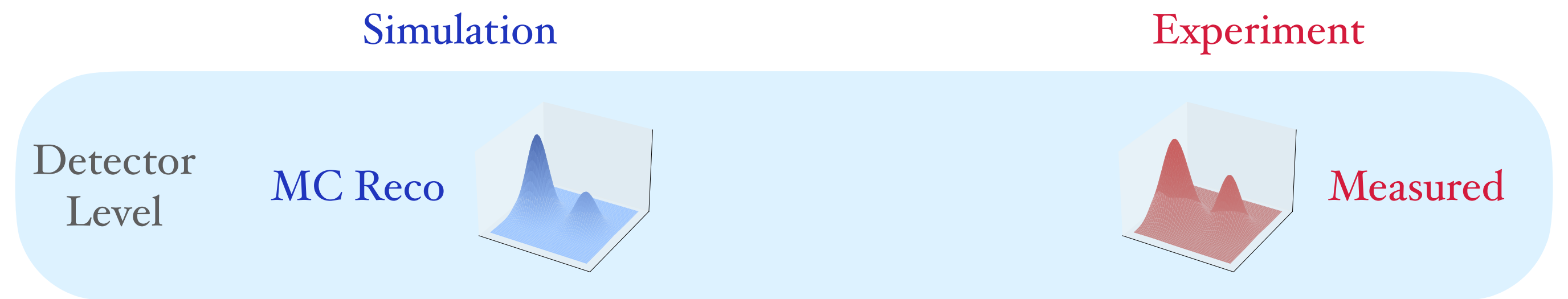
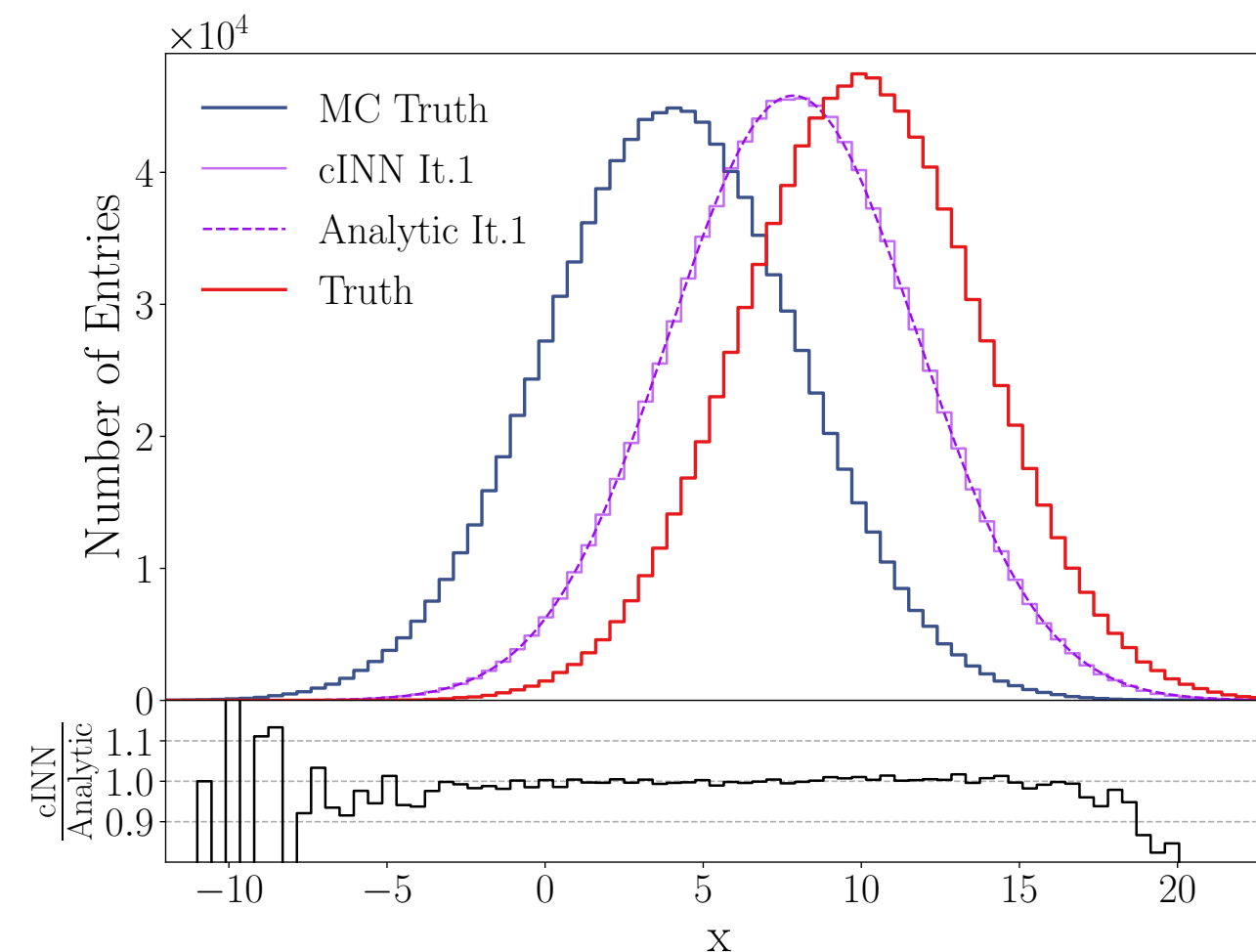
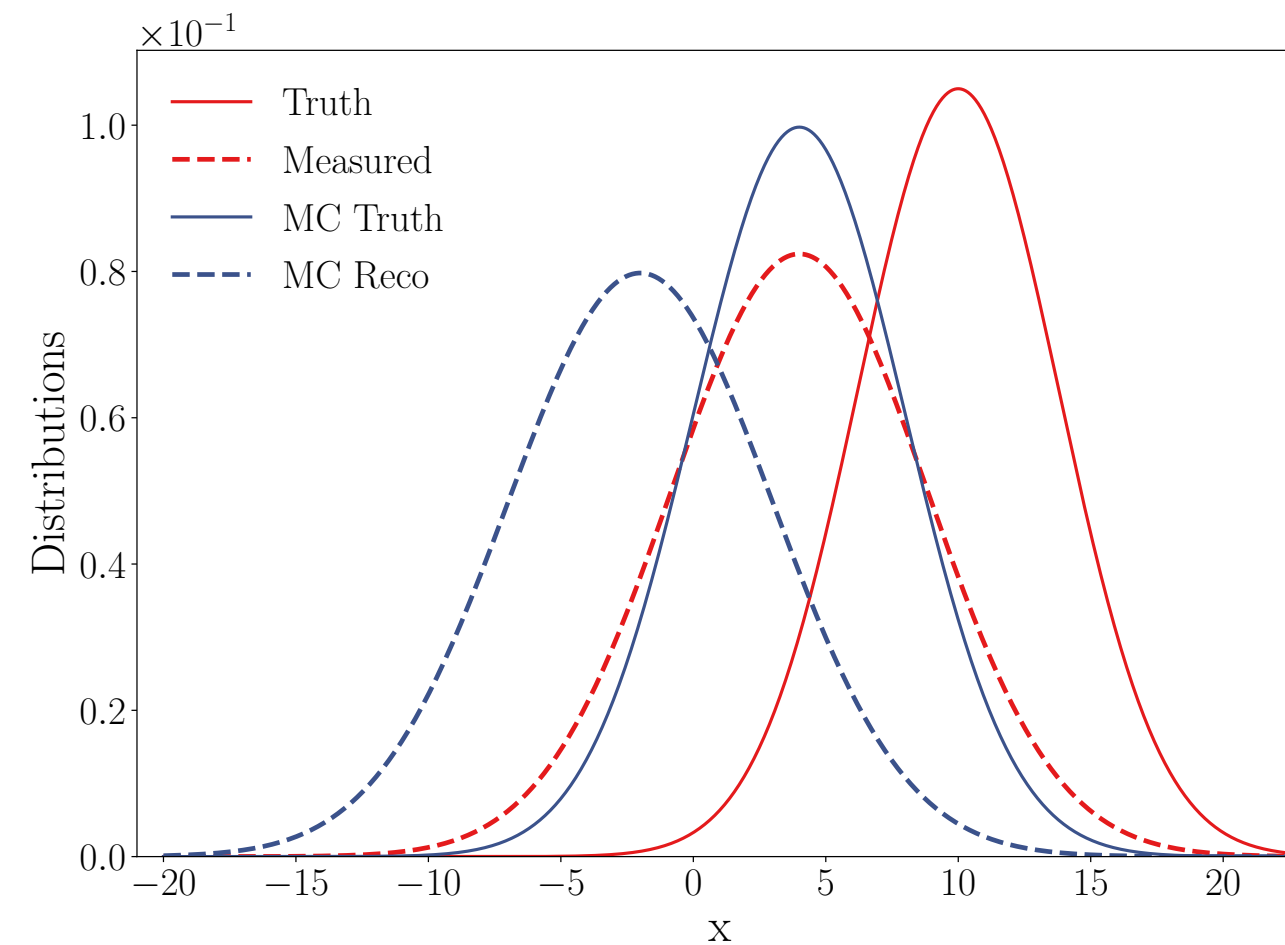
If everything is uncertain,

anything is possible

Margaret Drabble

Limitations of direct unfolding with generative NNs

Prior dependence



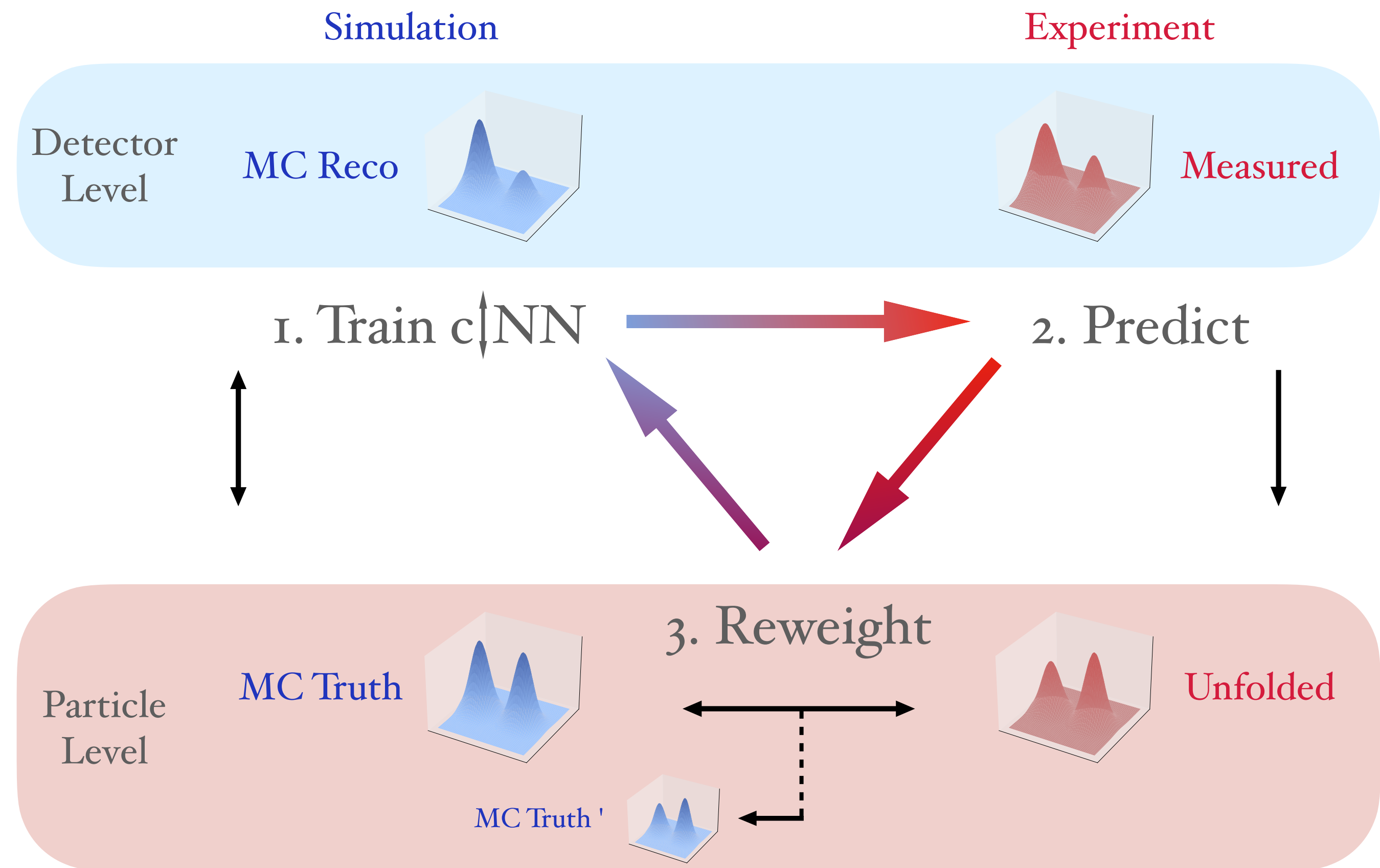
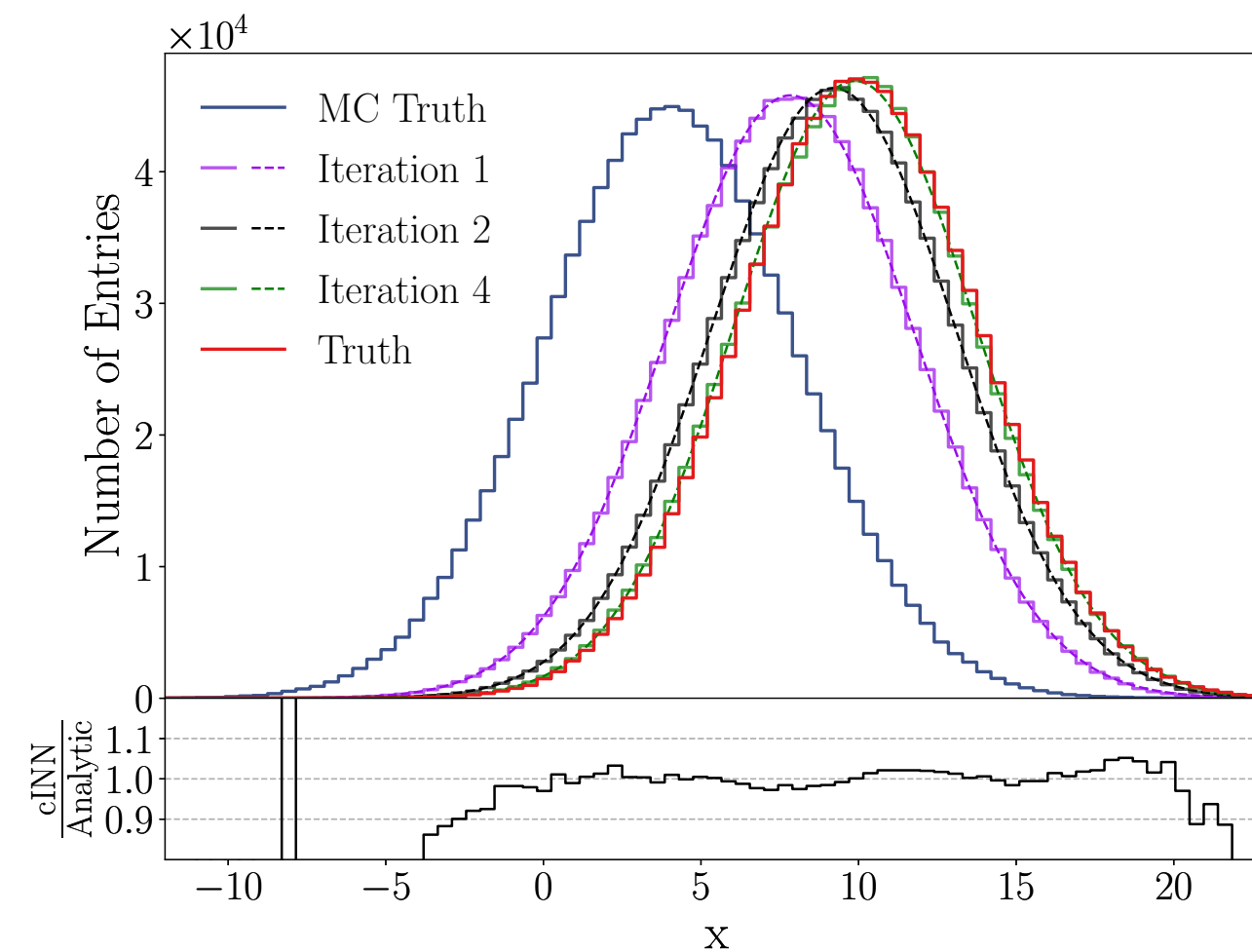
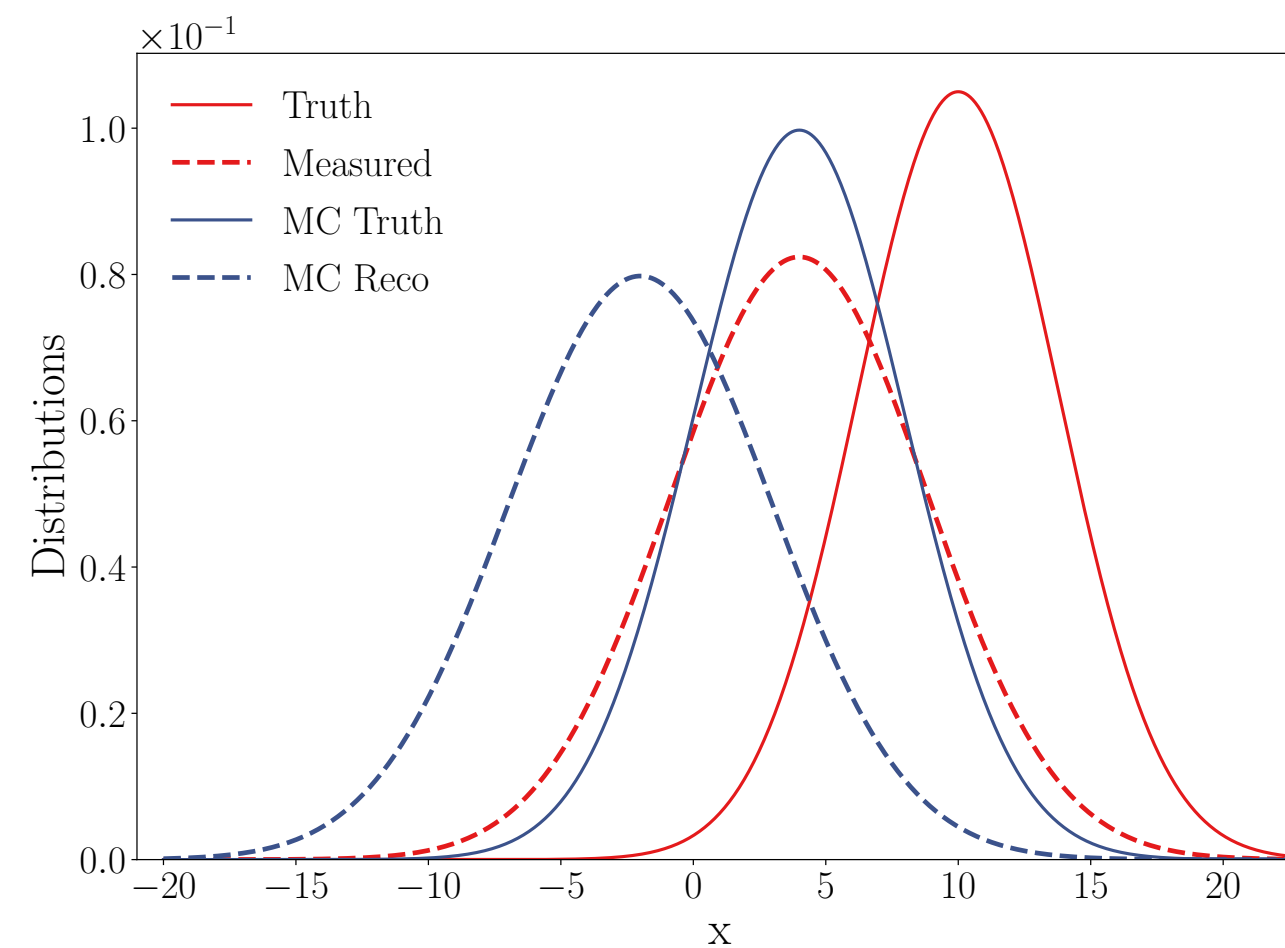
1. Train $c|NN$ $\longrightarrow$ 2. Predict



$$p(\text{part} | \text{det}) = \frac{p(\text{det} | \text{part}) \cdot p(\text{part})}{p(\text{det})}$$

Conditional iterative unfolding

[2212.08674] M. Backes, et al.



SPINUP

Simulation-Prior Independent Neural Unfolding Procedure

[2507.15084] AB, T Heimerl, N. Huetschi, M. Kagan, T. Plehn

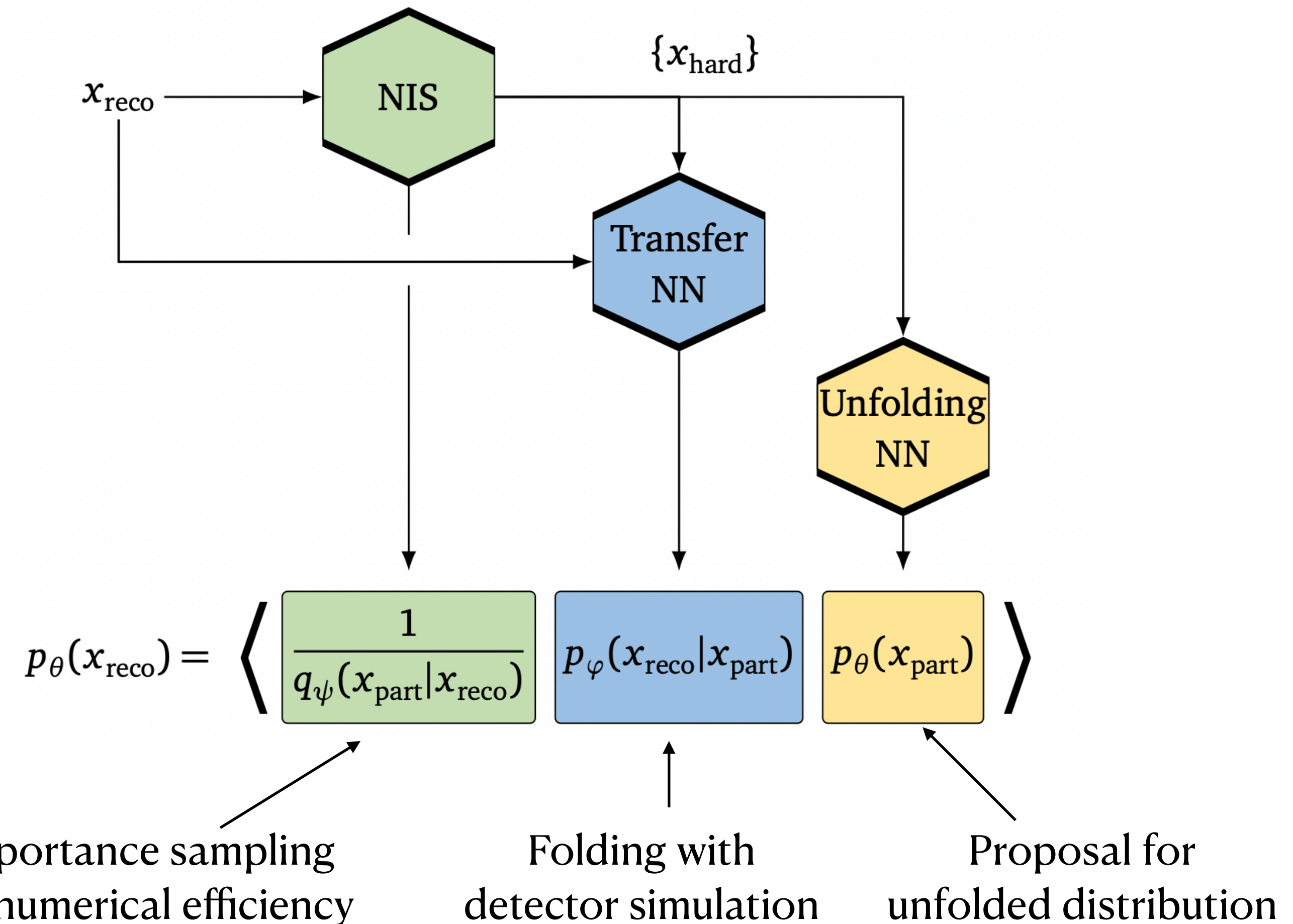
Ansatz

DATA

“folded” unfolded distribution

$$\begin{aligned}
 p_{data}(x_{reco}) &= p_{\theta}(x_{reco}) \\
 &= \int dx_{part} p(x_{reco} | x_{part}) p_{\theta}(x_{part})
 \end{aligned}$$

→ Minimize KL divergence



SPINUP

Simulation-Prior Independent Neural Unfolding Procedure

Ensembling for lower bound on uncertainty from information loss

Drastic example:

Forward mapping $x_{reco} = |x_{part}| + \mathcal{N}(0, 0.2)$

