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

Anja Butter
LPNHE (Paris), ITP (Heidelberg)

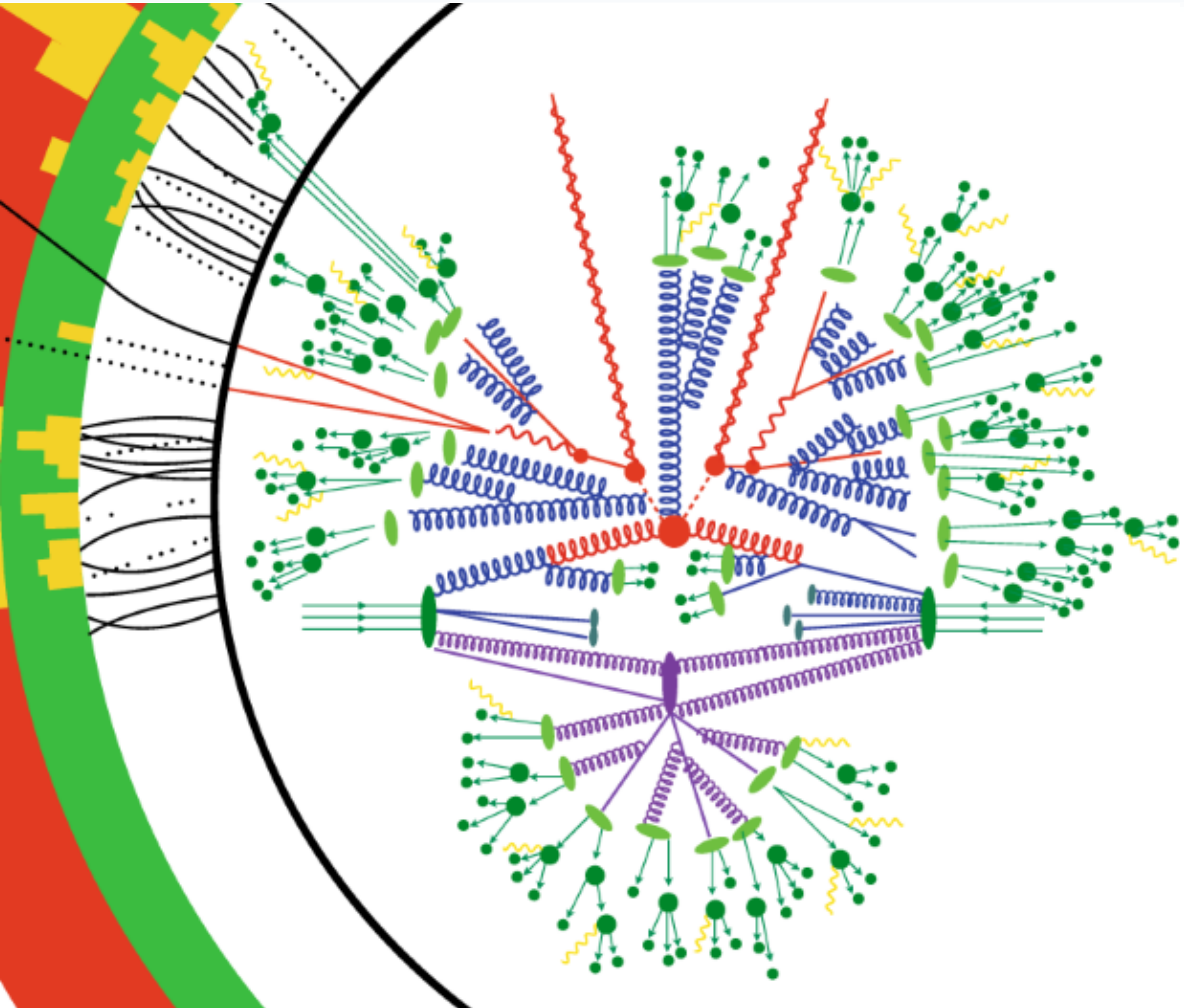
Why use machine learning in particle physics?

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^xLO process, off-shell,...

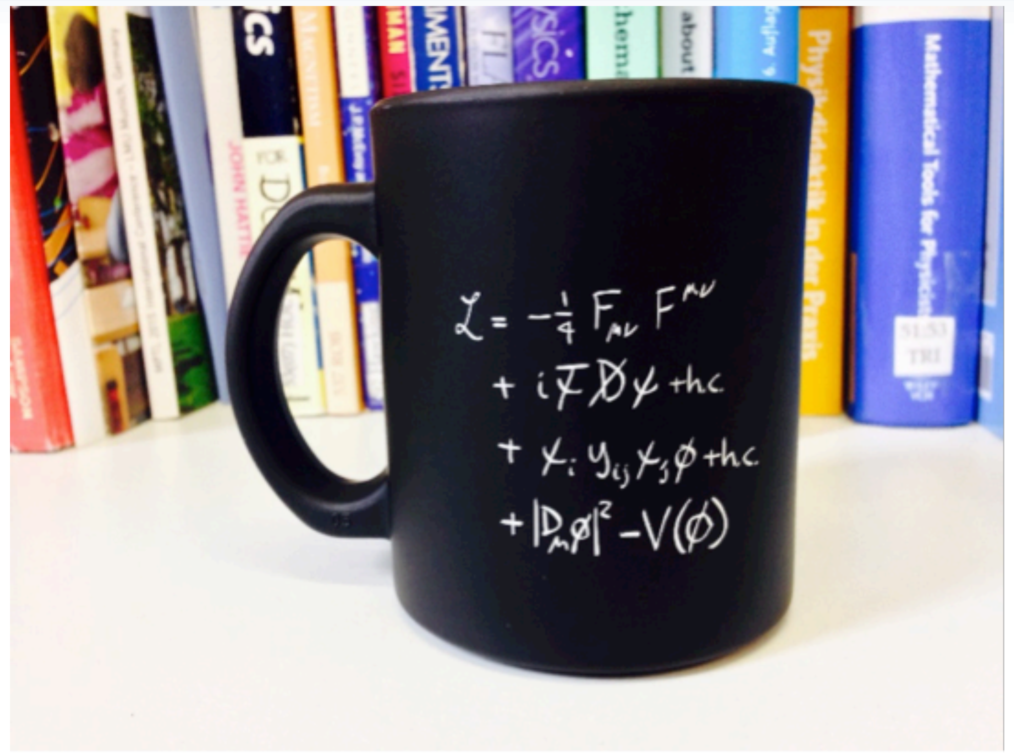
State of the art runtime for individual diagrams

	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.

[2305.19768] G. Heinrich, et al.

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 throug
- New ideas (anomaly
- New dynamic from fast moving developments in ML

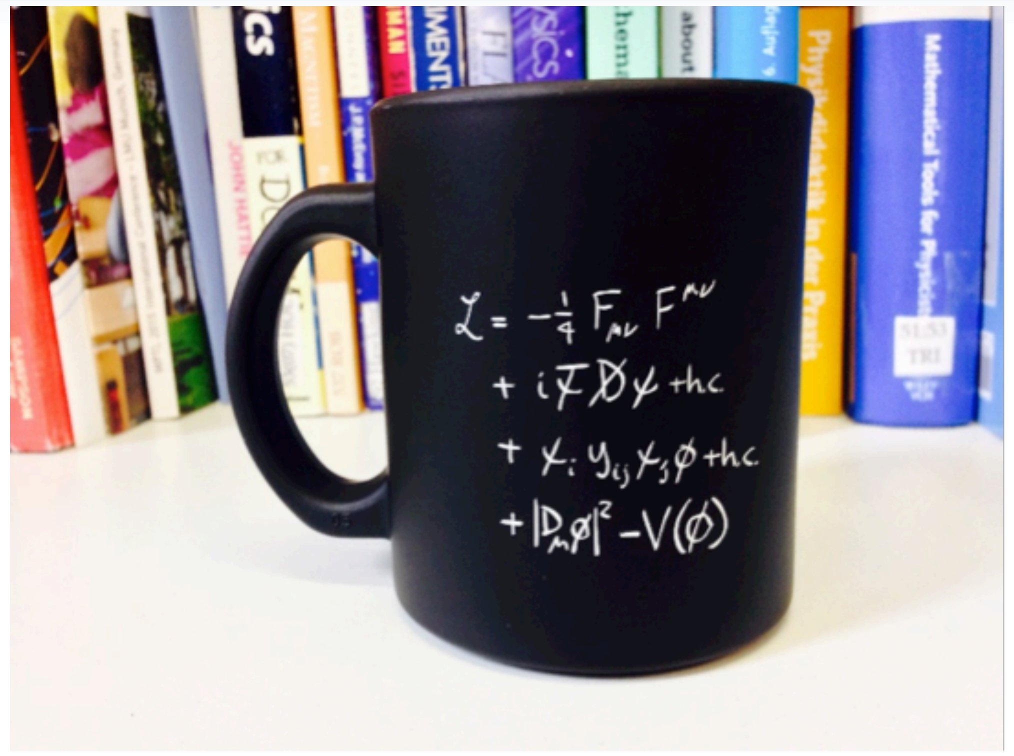
Develop & adapt ML techniques

→ right balance between ML & theory

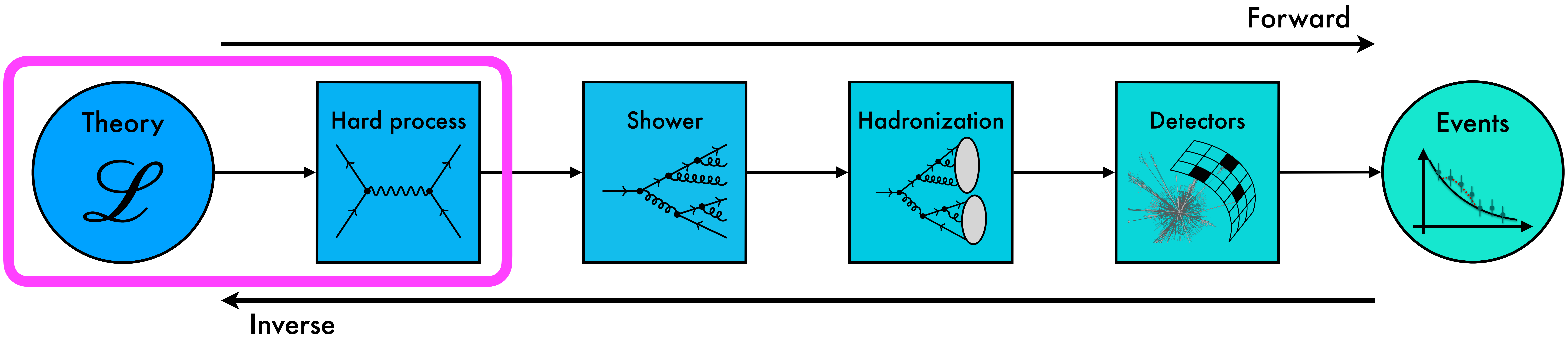
- Fast precision calculations & simulations -

- Optimized data analysis -

- Calibrated uncertainties -

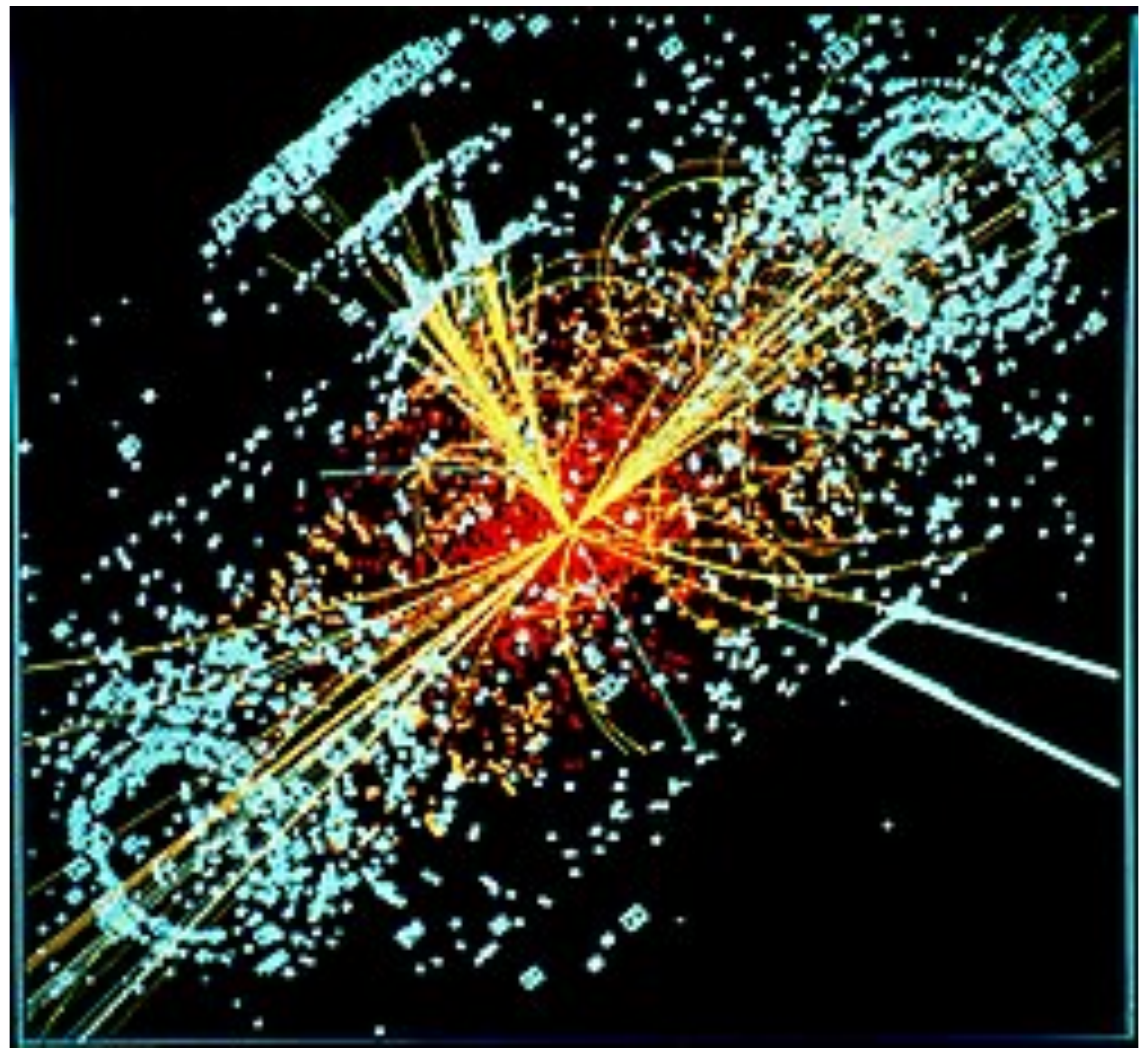


Monte Carlo event generation

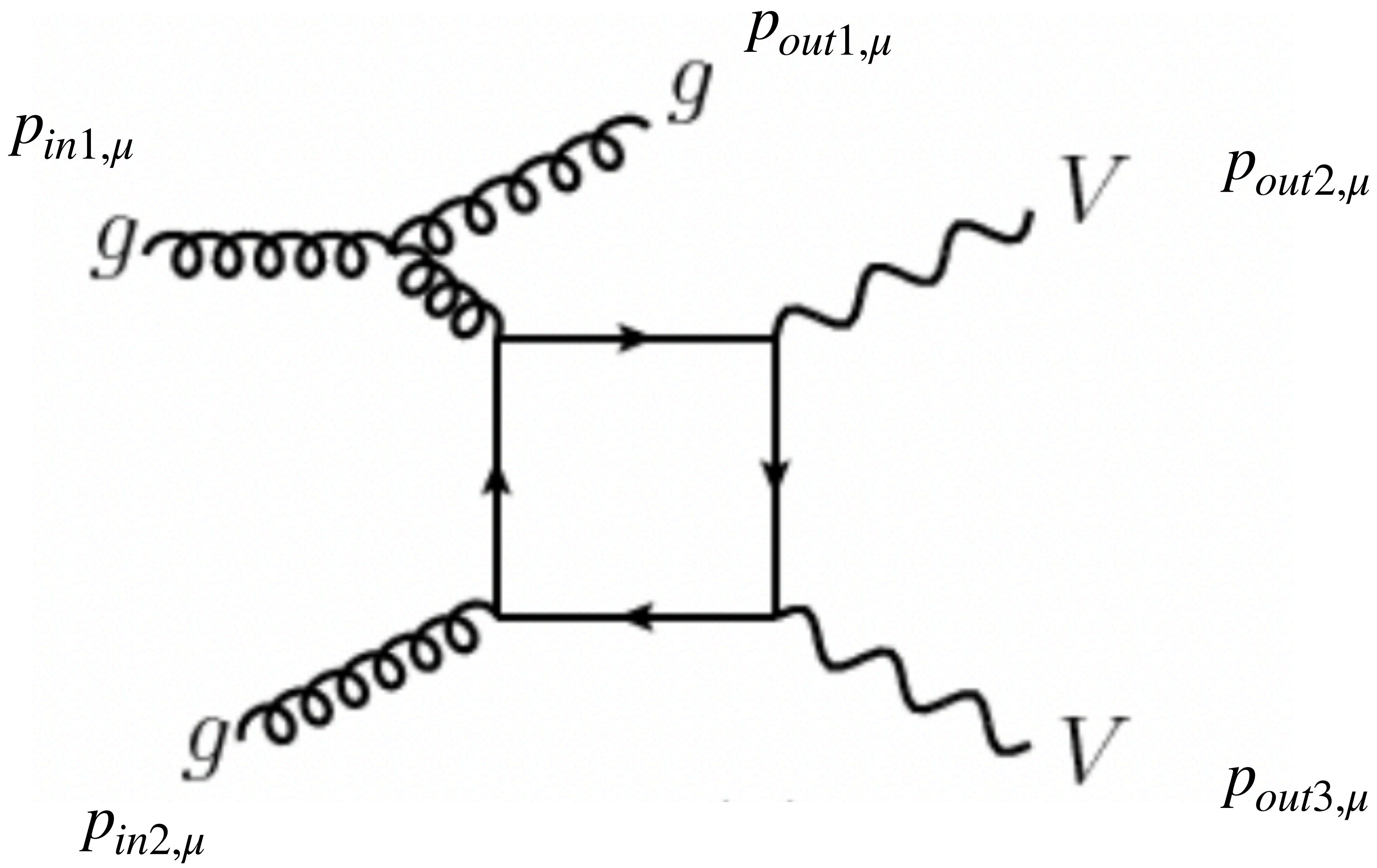


1. Precision amplitudes
or
Quantifying uncertainties in regression problems

Learning Amplitudes - A Regression Problem



Experimentalist's view



Theorist's view

- Amplitudes are @ the heart of LHC physics - probability of particle interactions
- Compute $A(\{p_{i,\mu}\} \,|\, i = 1,\ldots,N_{particles})$ 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 $\mathscr{L}$	Likelihood $\mathscr{L}$
Learnable function	Neural network f_θ	Physics knowledge $f_{physics}$
Data	Sampling $\frac{1}{N}\sum_{i-1}^{N} \mathscr{L}(f_\theta(x_i)\,;x_i)$	Phase space integration $\int \mathrm{d}x\ p(x)\ldots$
Solving the problem	Gradient descent $\theta^i \to \theta^{i+1} = \theta^i - \nabla_{\theta_i}\mathscr{L}$	Functional derivative $\frac{\delta \mathscr{L}}{\delta f} \stackrel{!}{=} 0$

Inspired by J. Thaler

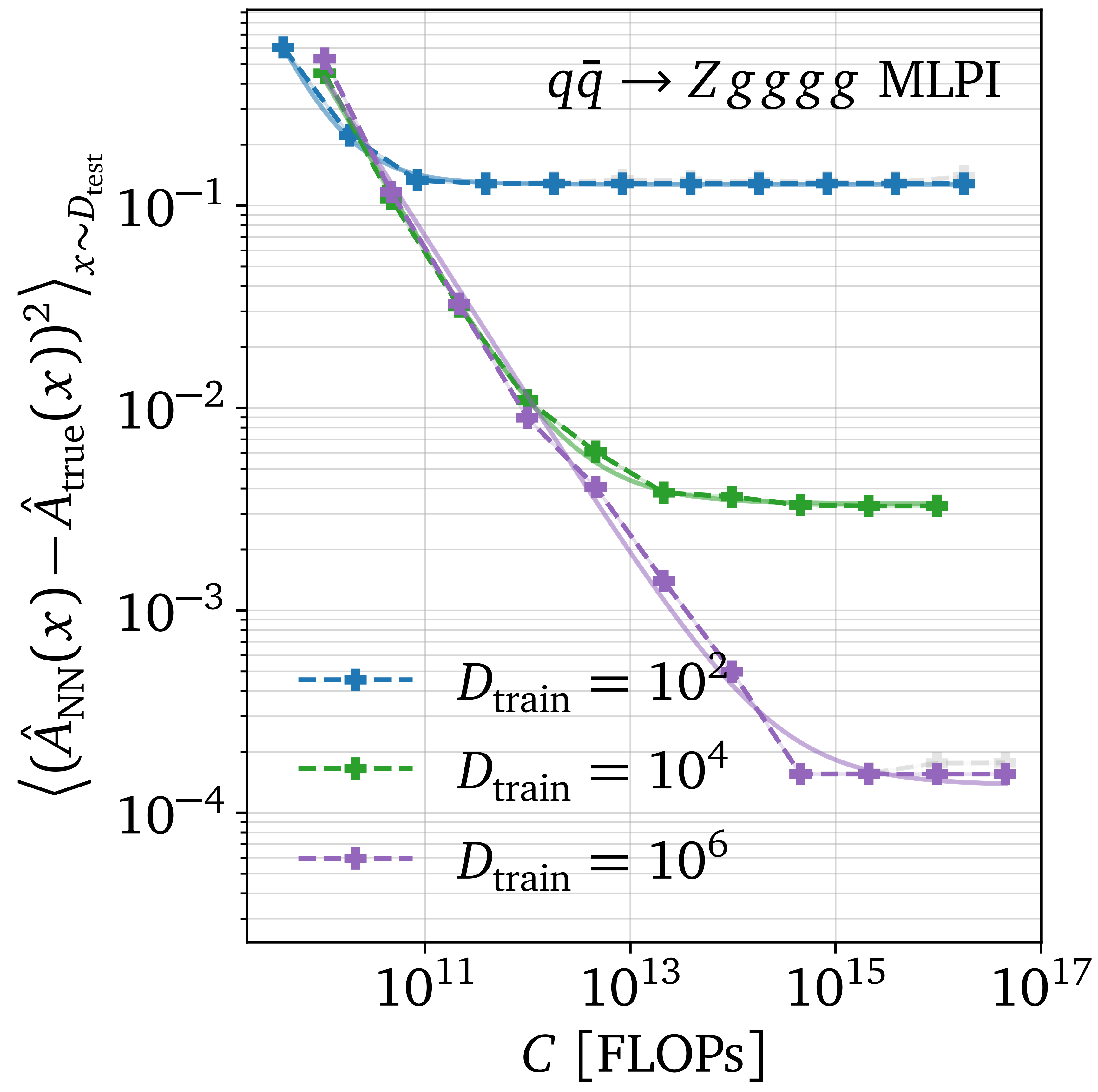
How to achieve high accuracy amplitudes?

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

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

- 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 $q\bar{q} \to Zgggg$

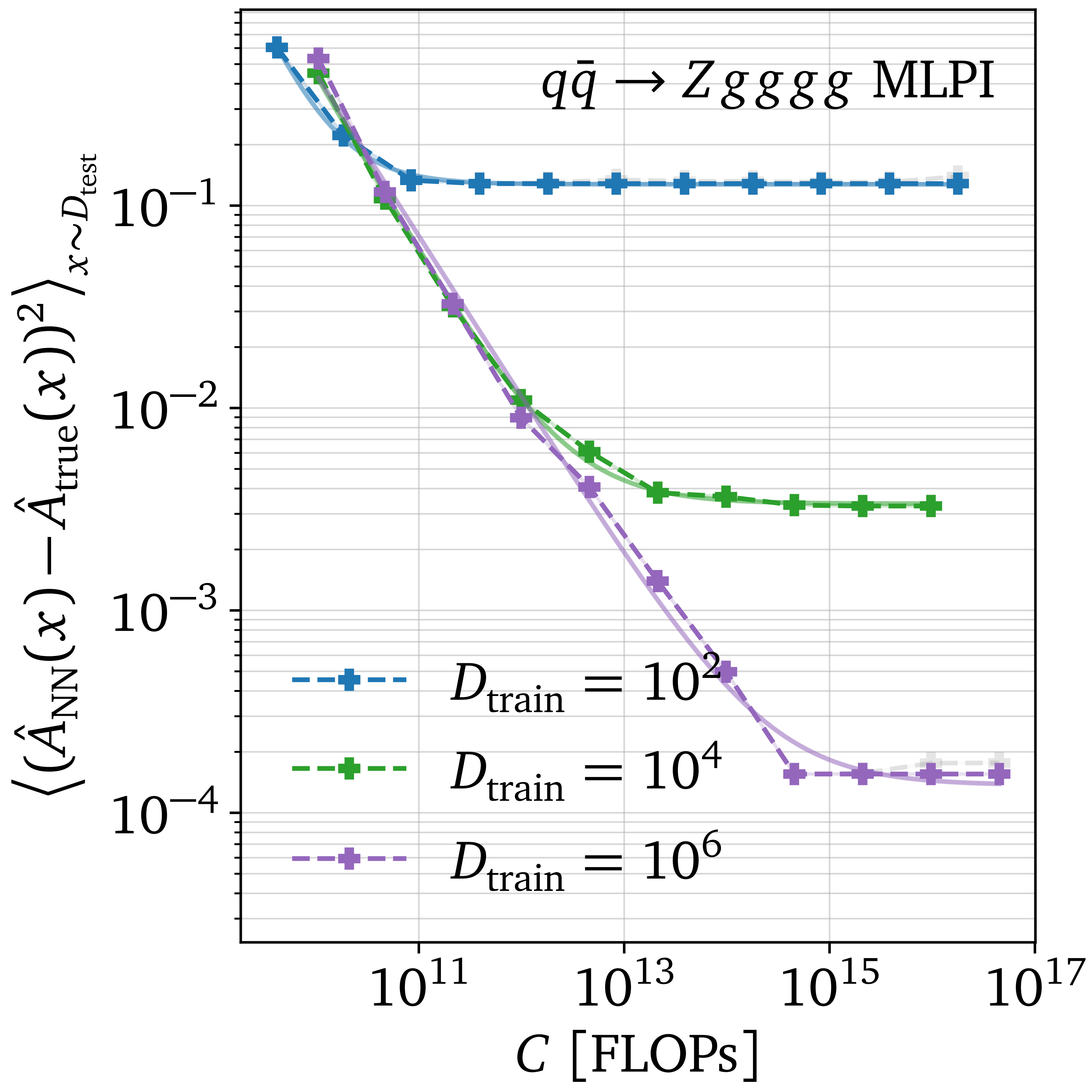
- Reminder:
  - Amplitudes are known exactly
  - Numerical noise negligible

- Setup:
  - Standard NN (MLP)
  - Use invariants $(s, t, u$
  - Evaluate MSE over te

Power law scaling

$$L(X, Y, Z) = (X_c/X)^{\alpha_X} + K(Y, Z)$$

for
Model size
Dataset size
Compute

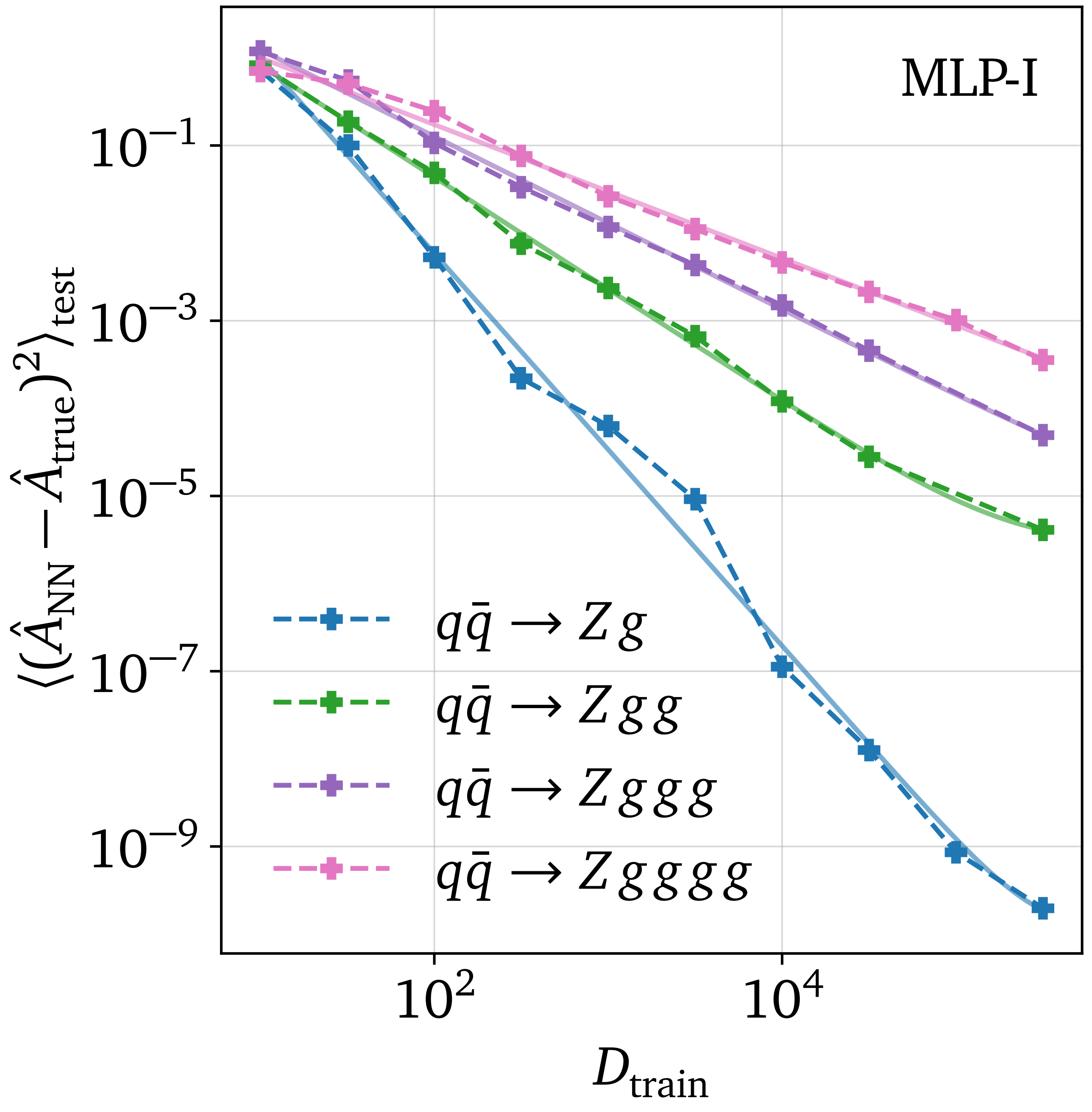


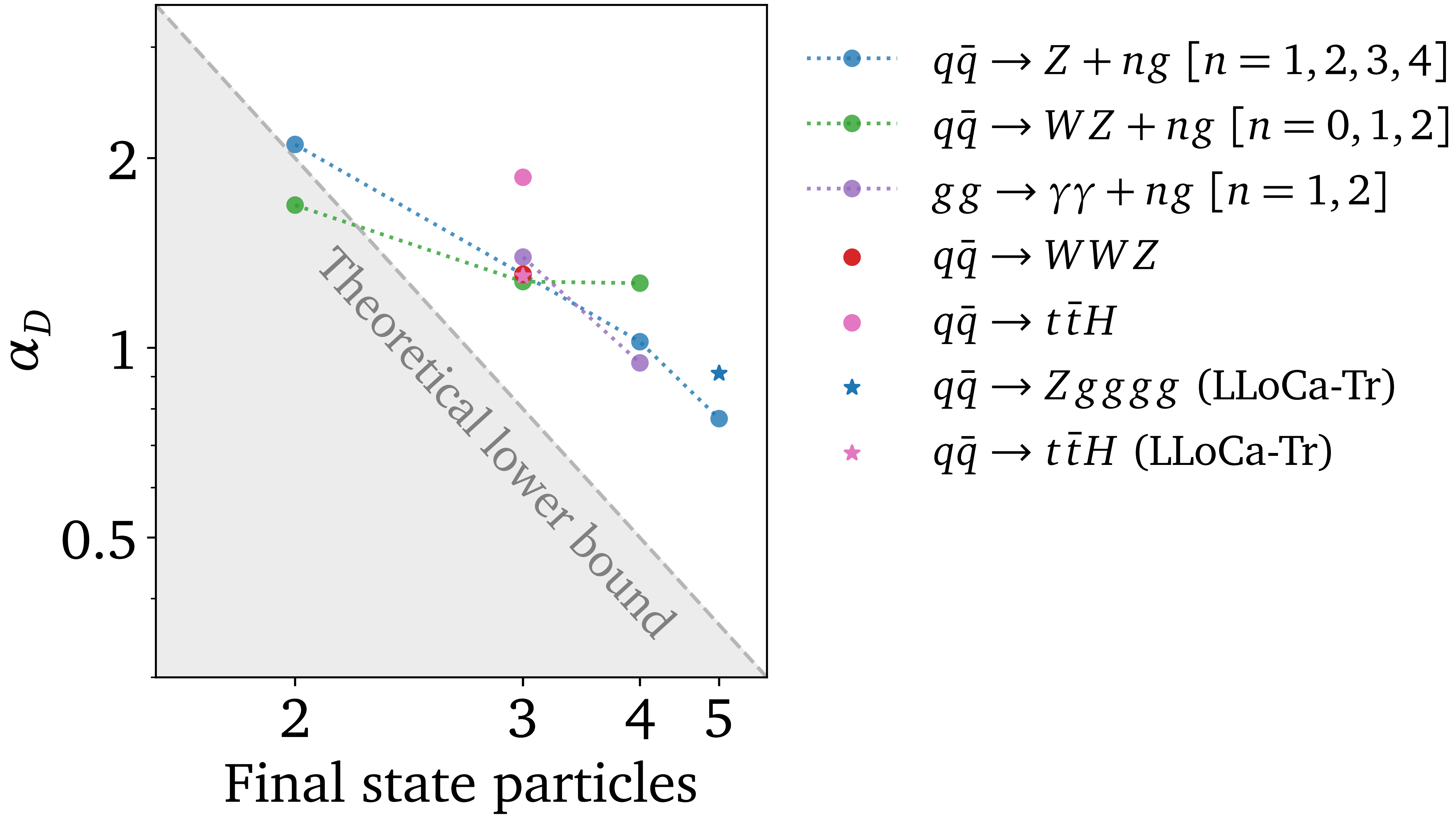
Can we predict the scaling?

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

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

- Scaling depends on degrees of freedom





Including symmetries

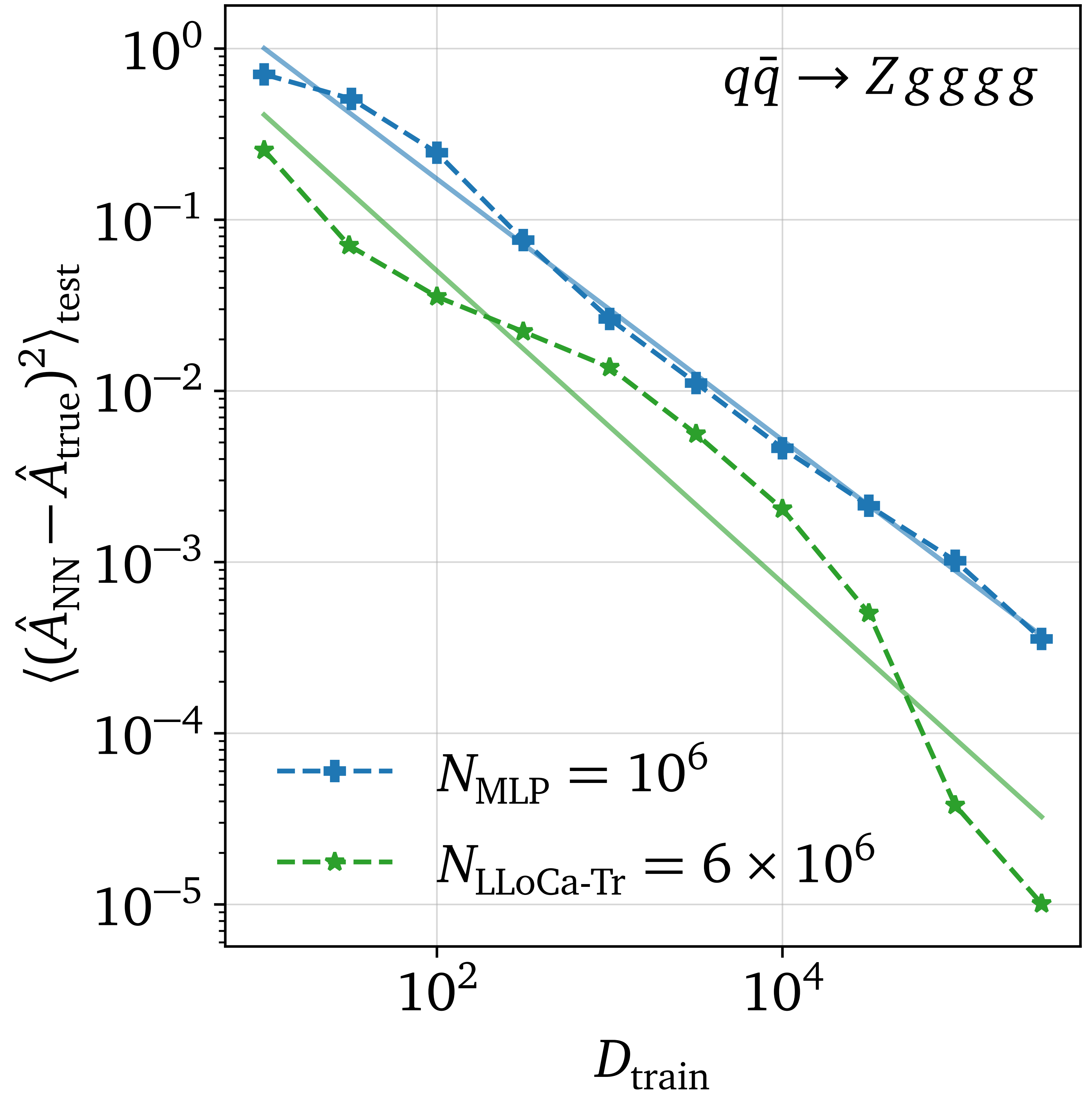
[2601.13308] H. Bahl, V. Bréso, AB, J. Ramirez

LLoCa network

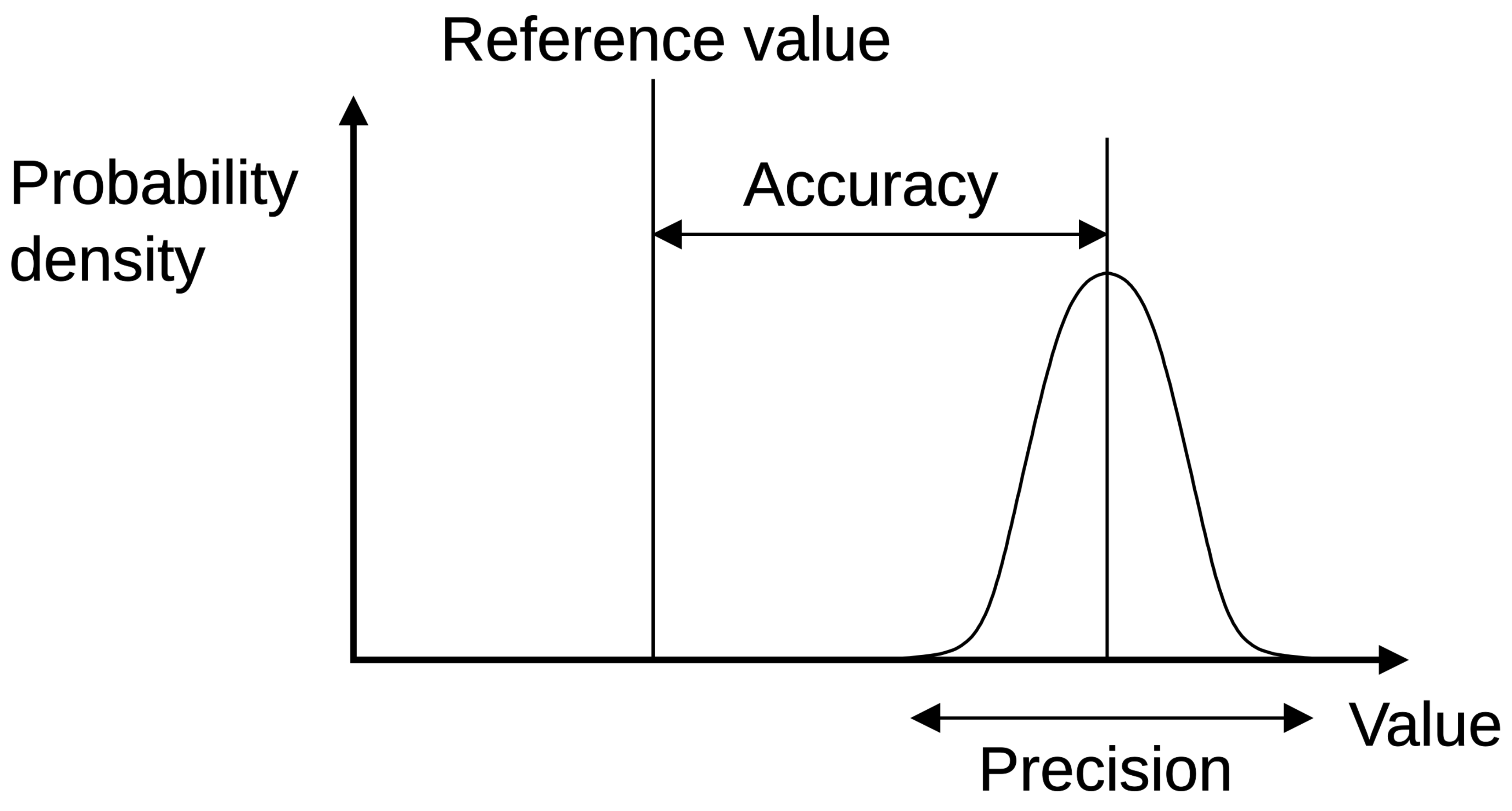
Lorentz Local Canonicalization [2505.20280]

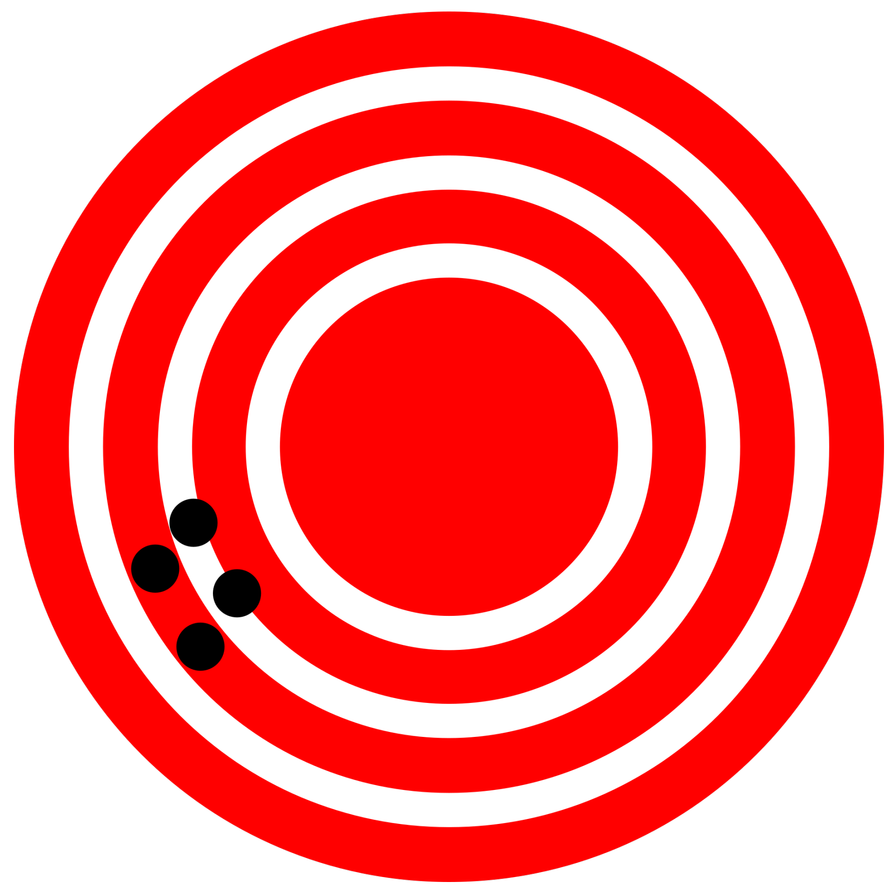
J. Spinner, L. Favaro, P. Lippmann, S. Pitz, G. Gerhartz, T. Plehn, F. A. Hamprecht

- 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

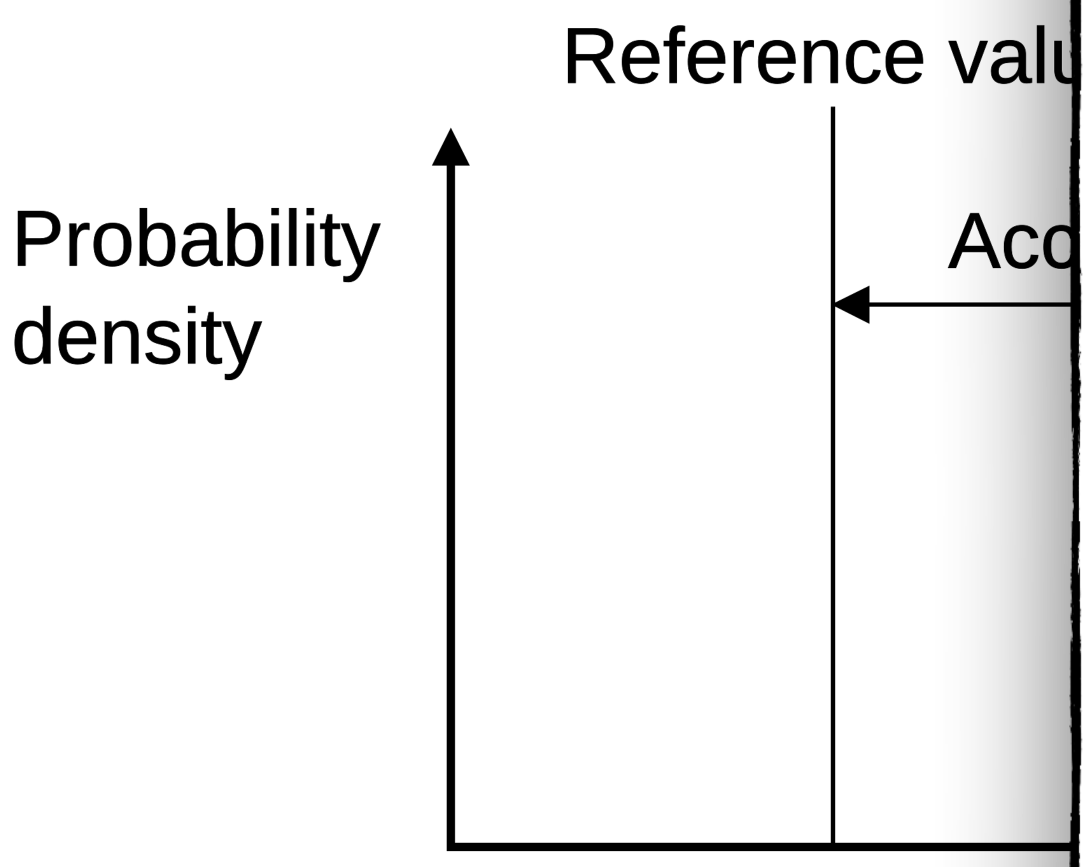




High precision with low accuracy

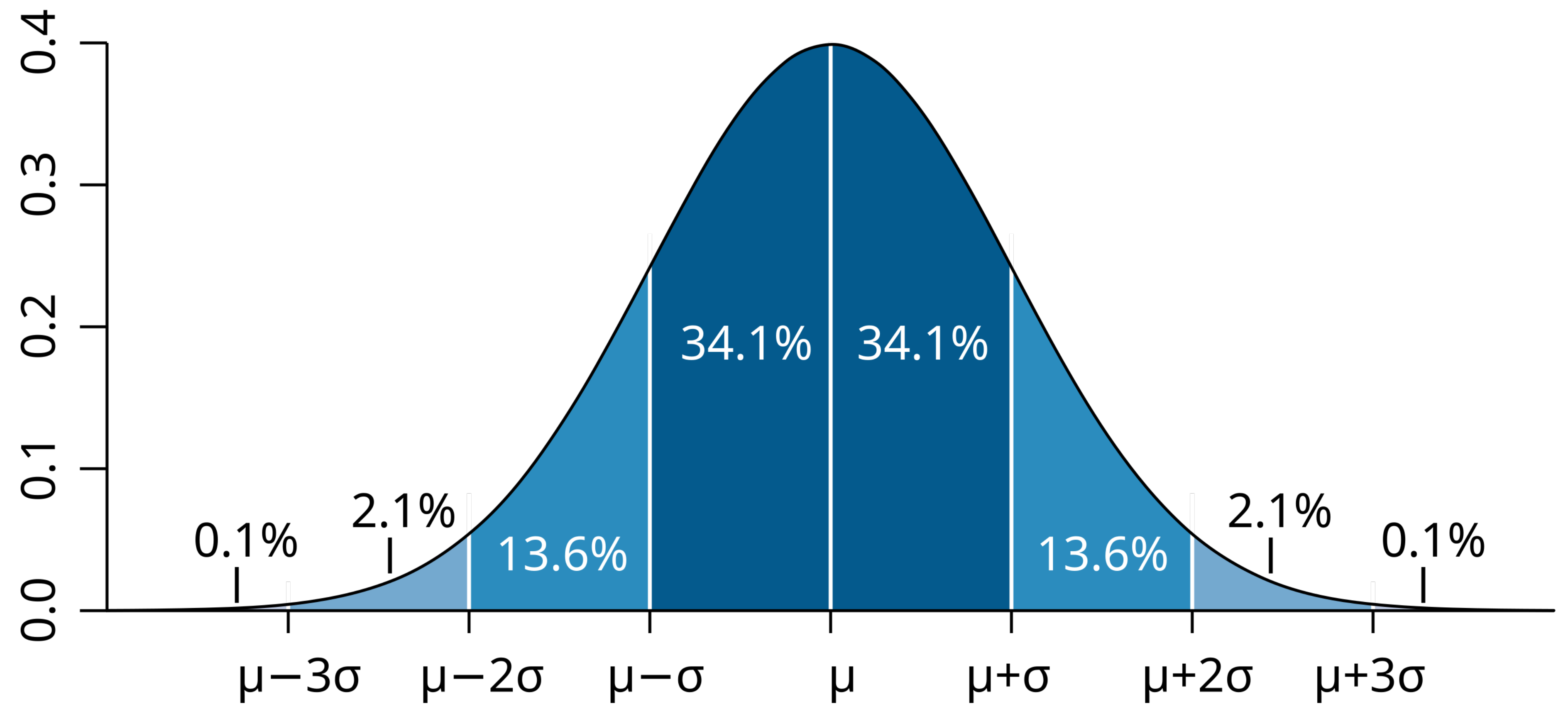
Best of all worlds: high accuracy & high precision

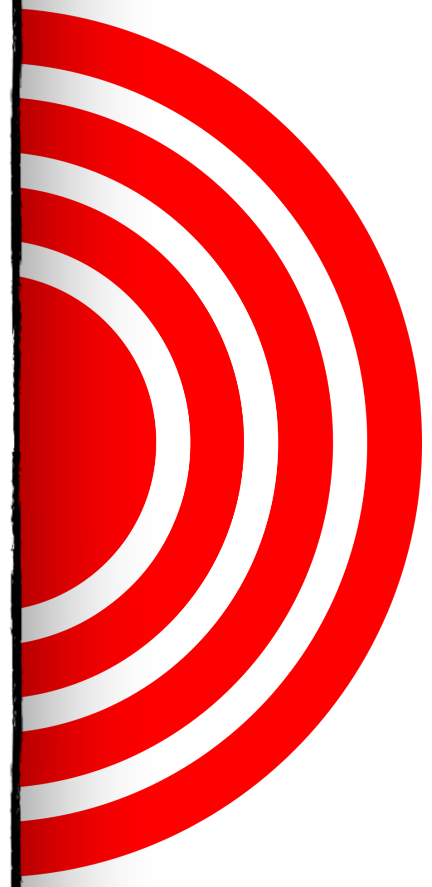
Accuracy vs precision



More important:

Calibrated uncertainties





vith low accuracy

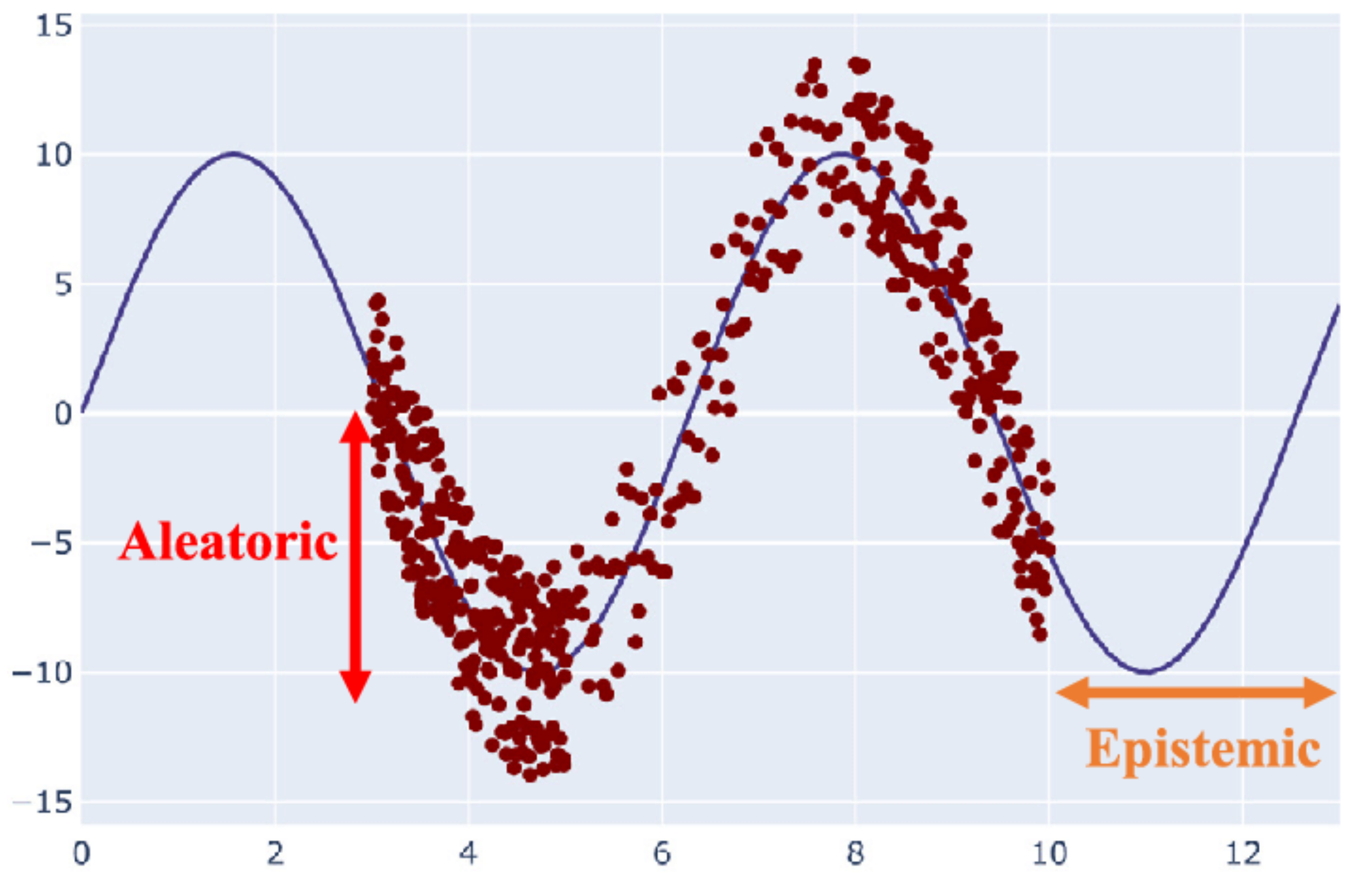
Best of all worlds: high accuracy & high precision

Who is talking about uncertainties ?

Physicist	Why ?	How ?
Statistical	Limited data	Poisson
Systematic	Jet energy resolution	Gaussian
Theorie	Higher order calculation	Flat?

ML researcher	Meaning
Aleatoric /Stochastic	Intrinsic noise, irreducible with more data
Epistemic/Systematic	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:

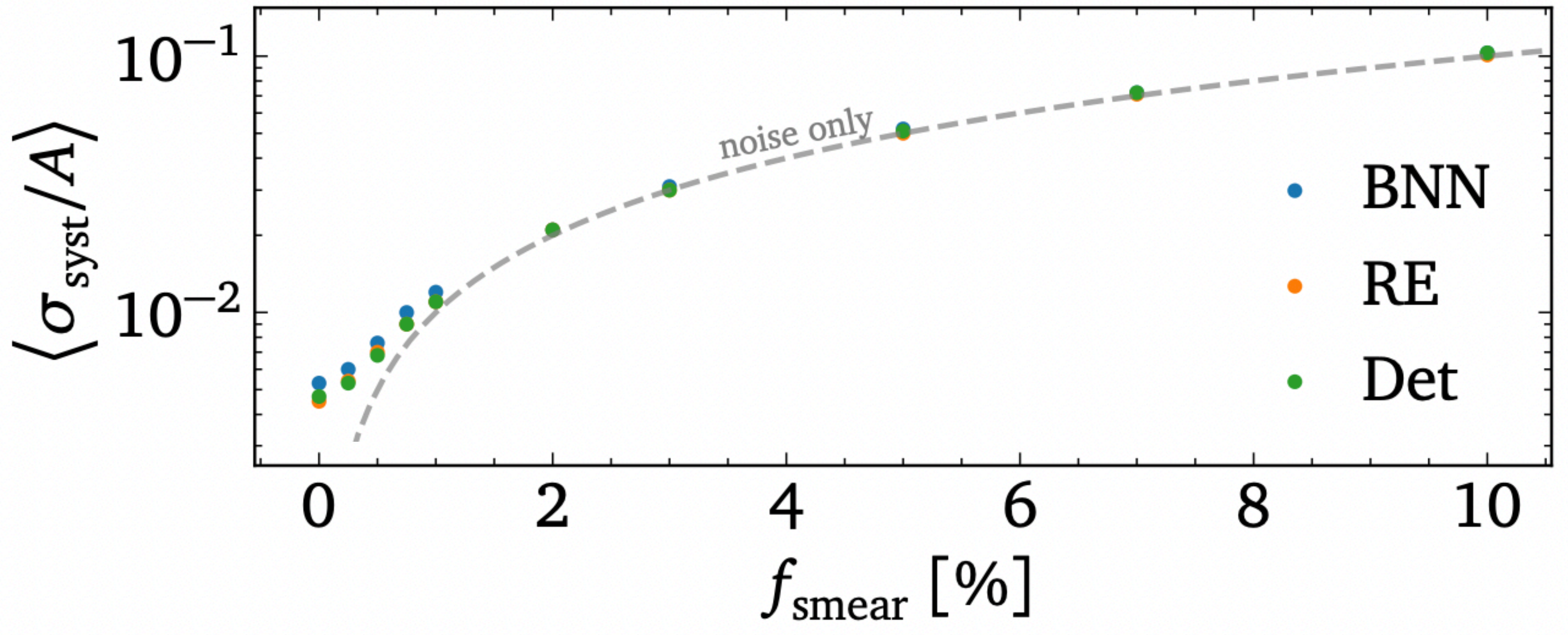
$$p(\theta|x_{\text{train}}) = p(x_{\text{train}}|\theta) \cdot p(\theta) \cdot \text{const}$$

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

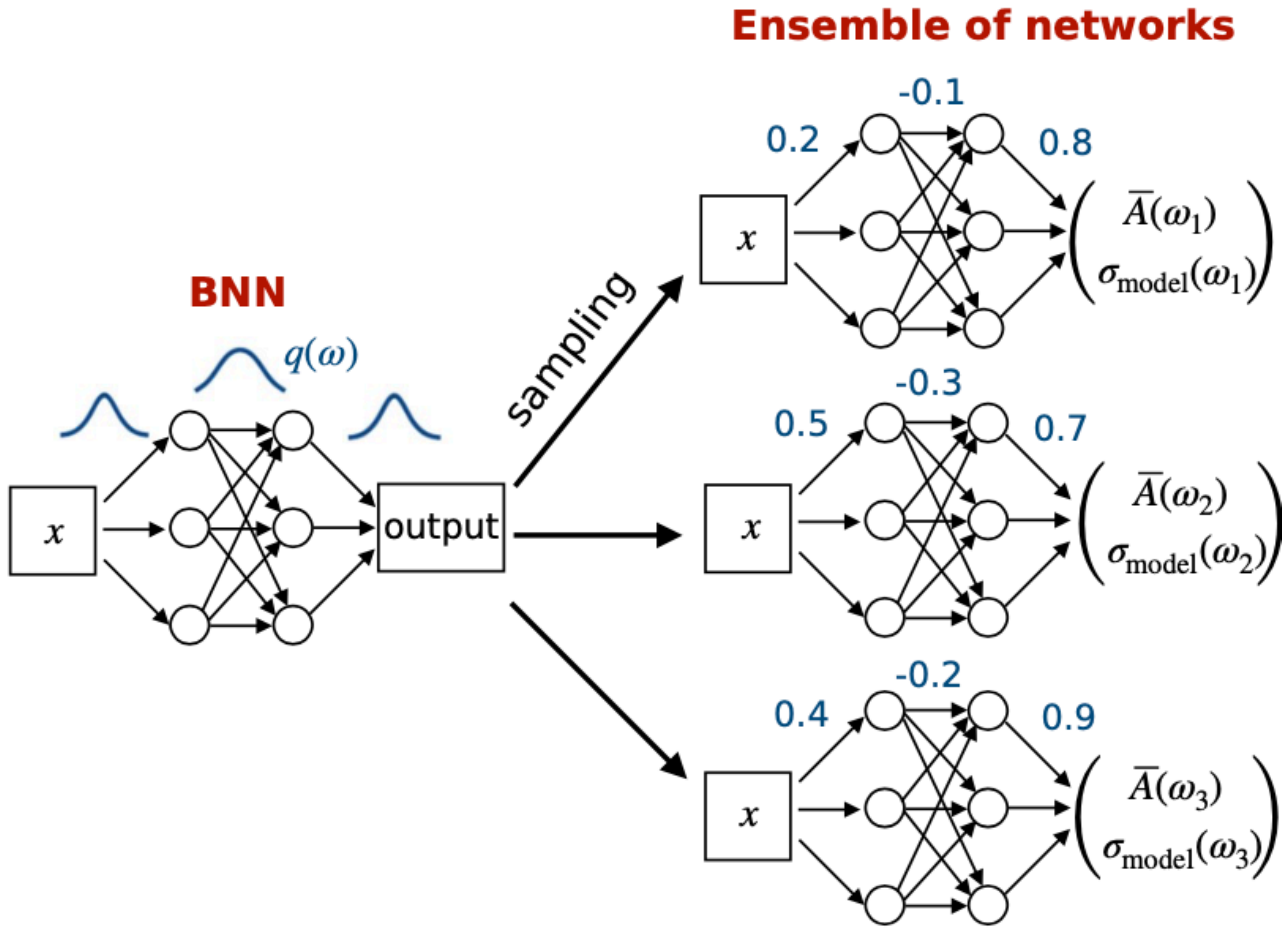


2412.12069 H. Bahl, N. Elmer, L. Favaro, et al.

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

Separating uncertainties with Bayesian networks

$$\boxed{p(A) = \int \mathrm{d}\omega \; p(A\,|\,\omega) p(\omega\,|\,T) \approx \int \mathrm{d}\omega \; p(A\,|\,\omega) q(\omega)}$$



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

$$\sigma^2_{syst} = \int d\theta \; q(\theta) \; [\overline{A^2}(x, \theta) - \bar{A}(x, \theta)^2]$$

$$= \int d\theta \; q(\theta) \; \sigma_{model}(\theta)^2$$

→ Captures noise, etc.

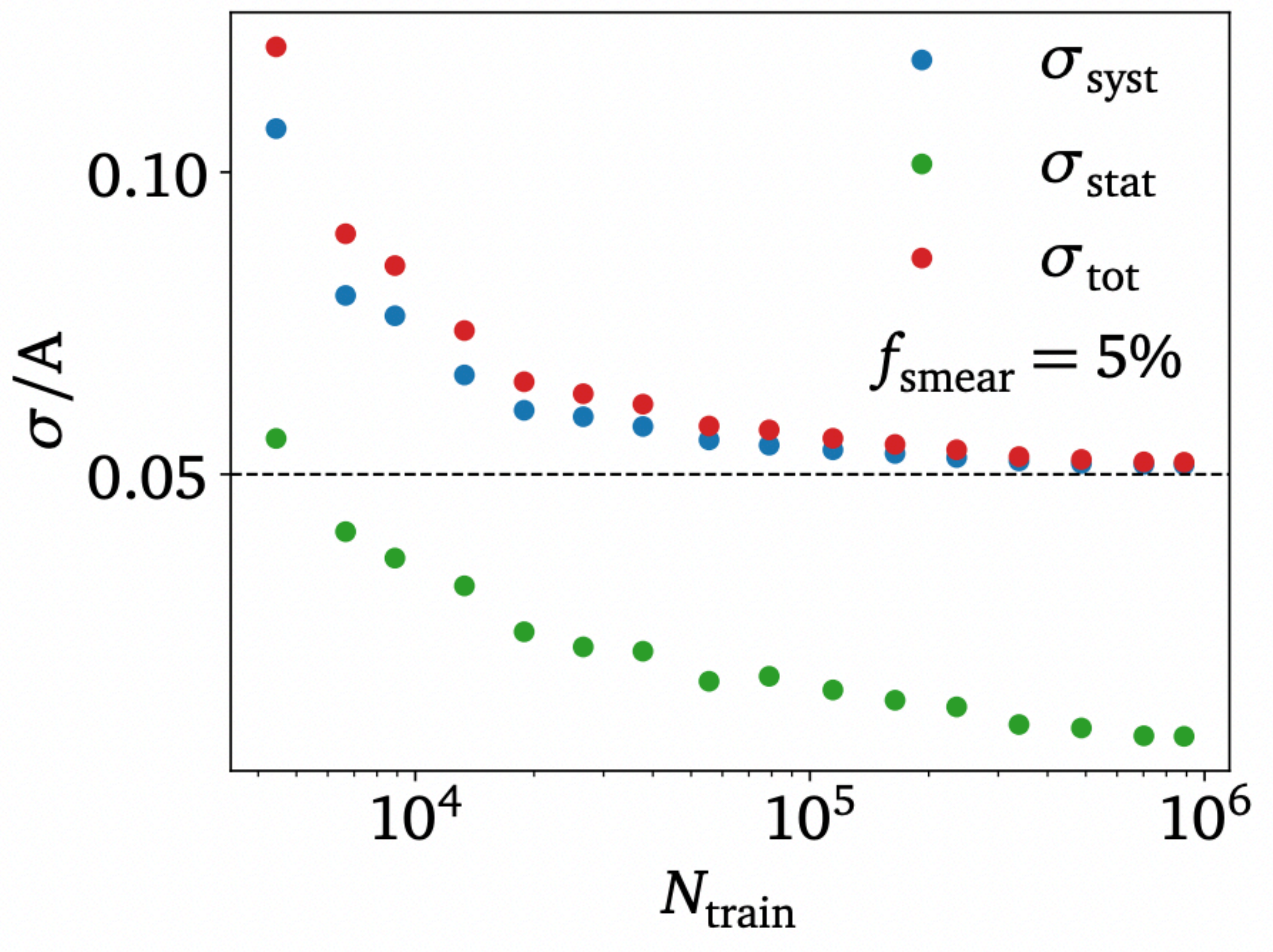
→ Plateaus for large training data

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

→ 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.



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

$$\sigma^2_{syst} = \int d\theta\, q(\theta)\, [\overline{A^2}(x,\theta) - \bar{A}(x,\theta)^2]$$

$$= \int d\theta\, q(\theta)\, \sigma_{model}(\theta)^2$$

→ Captures noise, etc.

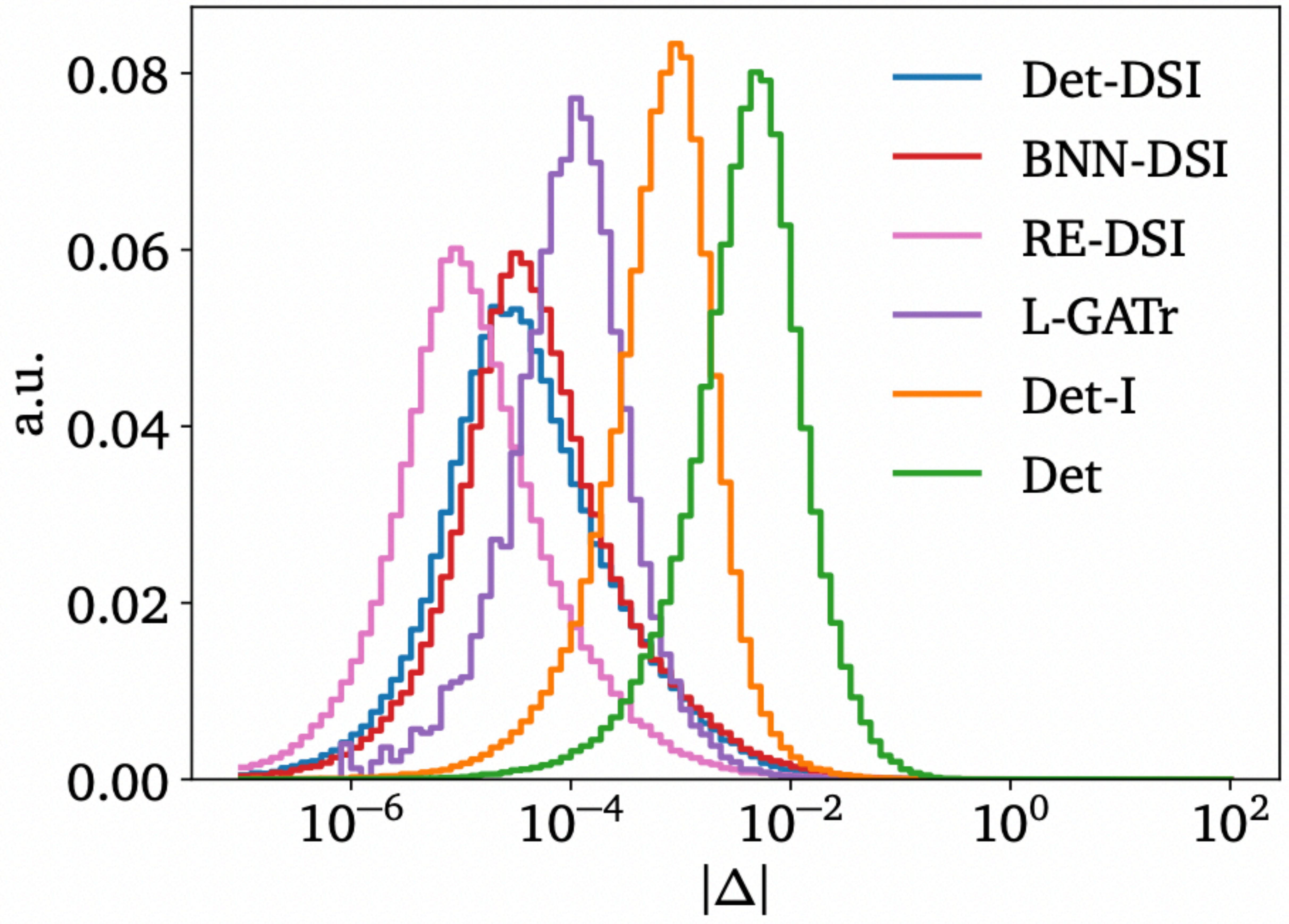
→ Plateaus for large training data

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

→ 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.



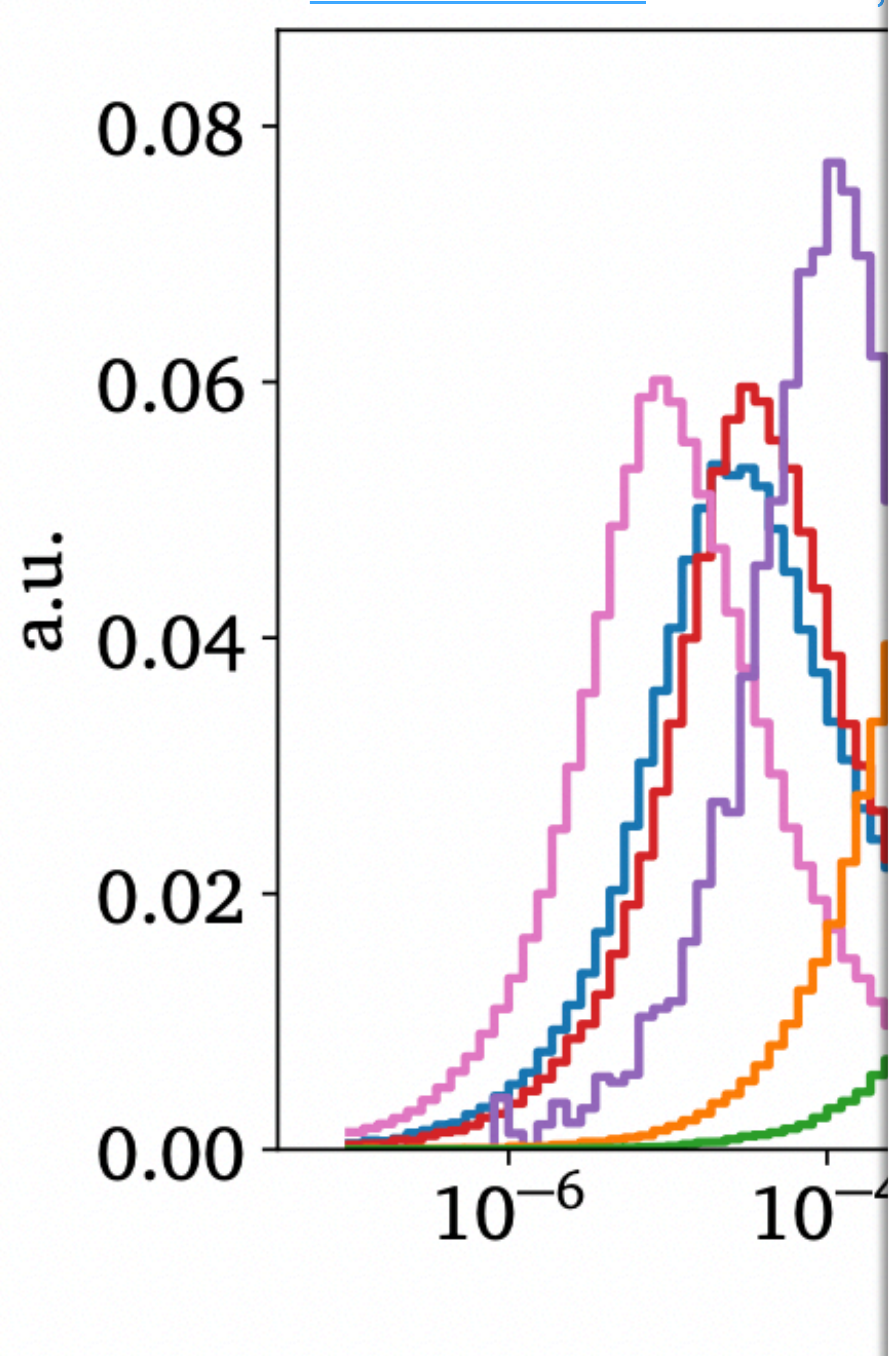
Optimisation of network architecture

No noise
\+ Large dataset
\+ Include invariants
\+ Deep sets architecture

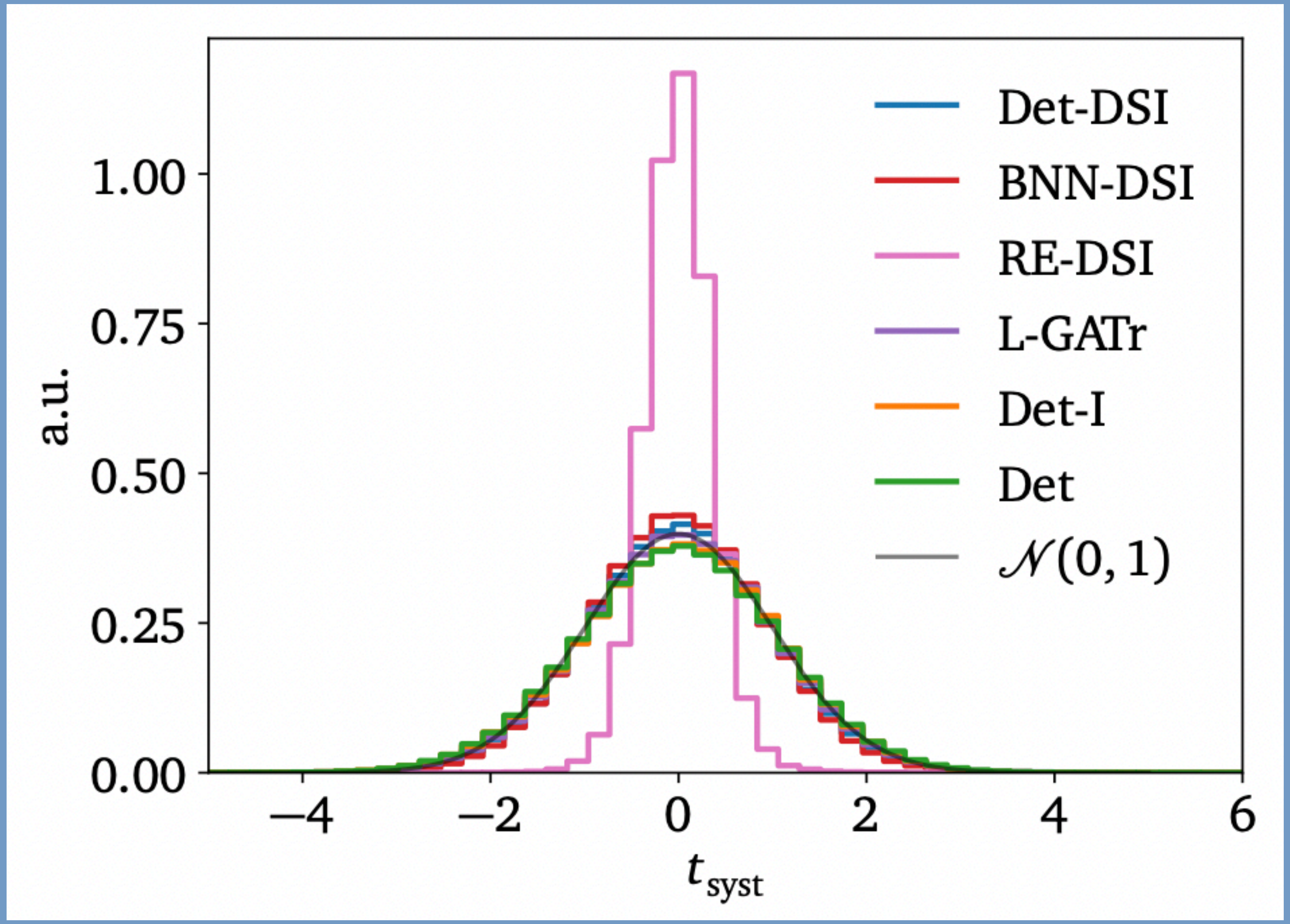
Reaching Precision of 10^{-5} !

What remains? Network size & symmetries (again)

2412.12069 H. Bahl,



Calibrated!



network architecture

noise

ge dataset

e invariants

ts architecture

ecision of 10^{-5} !

1. Precision amplitudes

or

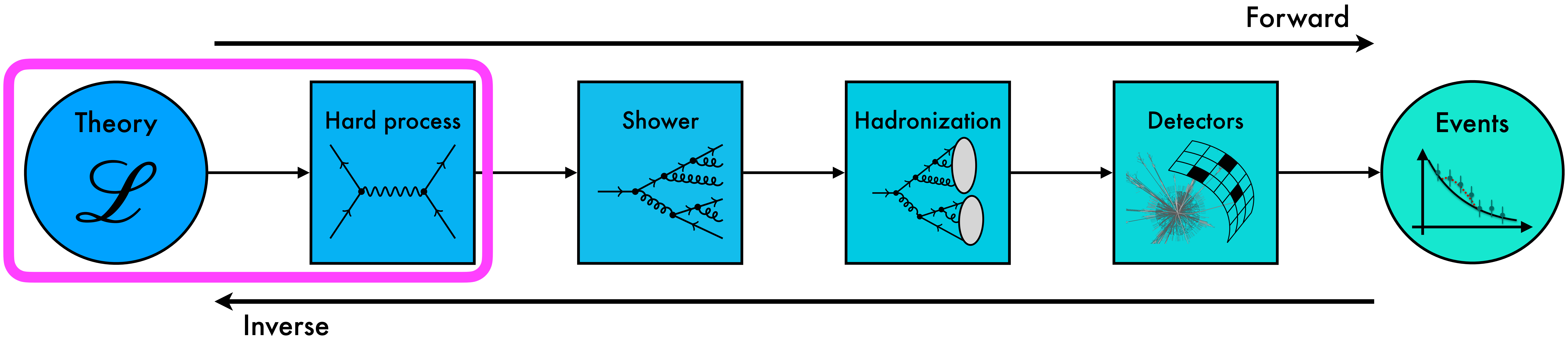
Quantifying uncertainties in regression problems

2. Event generation

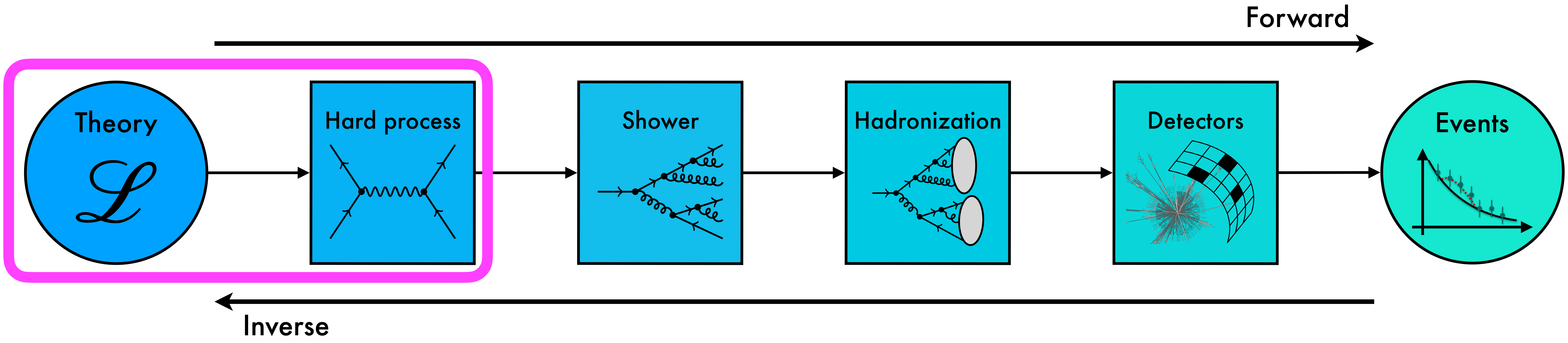
or

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

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

PDF

Matrix element

Phase space

3. Unweighting

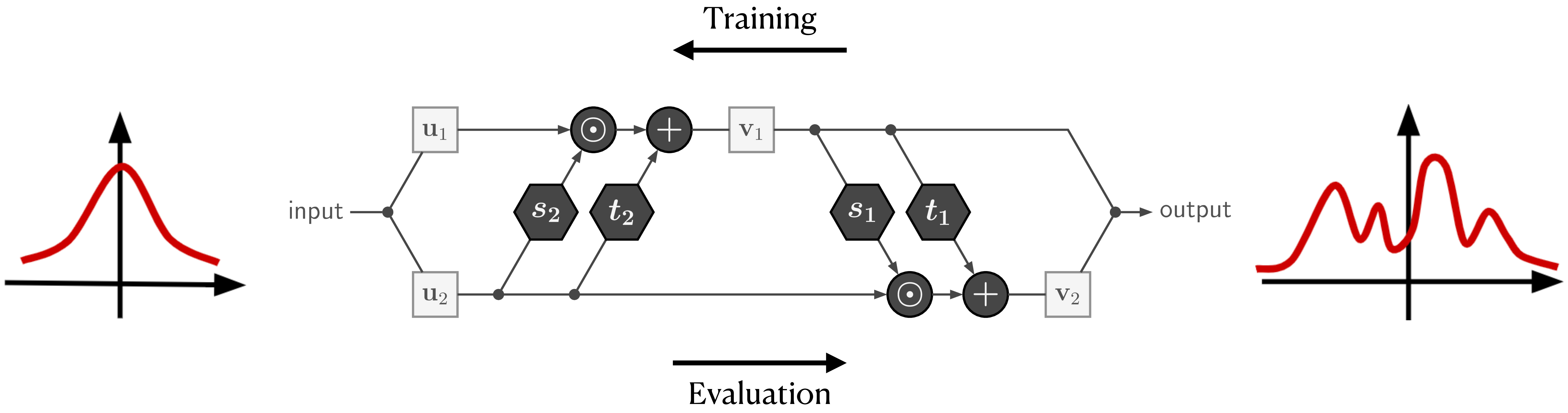
keep events with

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

Normalizing flows

Invertible networks for complex transformations

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



Training on density $t(x)$
→ Minimize difference

$$\mathscr{L} = \log p_x(x)/t(x)$$
$$= \log p_z(z(x))\, J_{NN} \,/\, t(x)$$

Requires evaluation of $t(x)$

Training on samples x
→ Maximize the log-likelihood

$$\mathscr{L} = \log p(\theta\,|\,x)$$
$$= \log p(z\,|\,\theta) + \log J_{NN} + p(\theta) + \text{const}$$

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

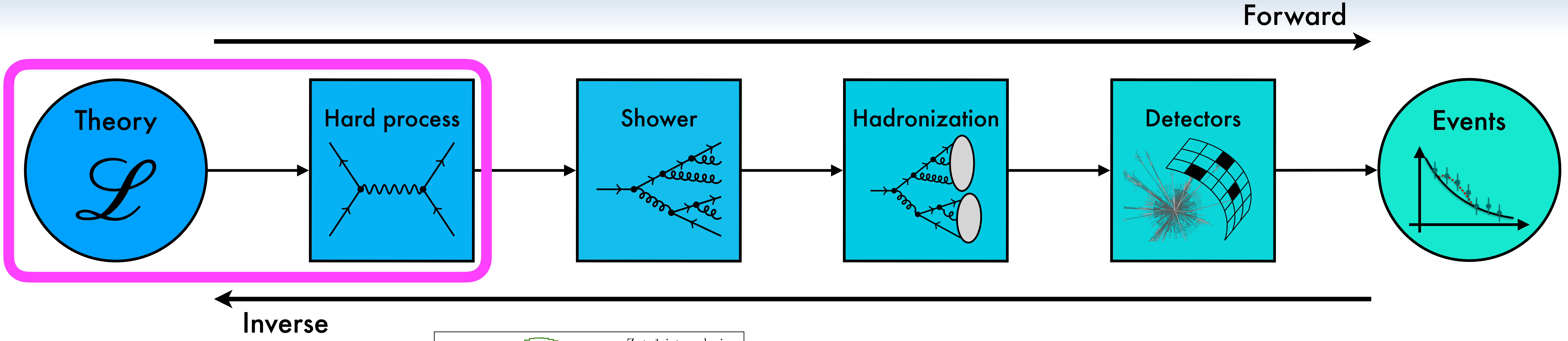
$$\begin{aligned}\mathscr{L} &= \log p_x(x)/t(x) \\ &= \log p_z(z(x))\, J_{NN} \,/\, t(x)\end{aligned}$$

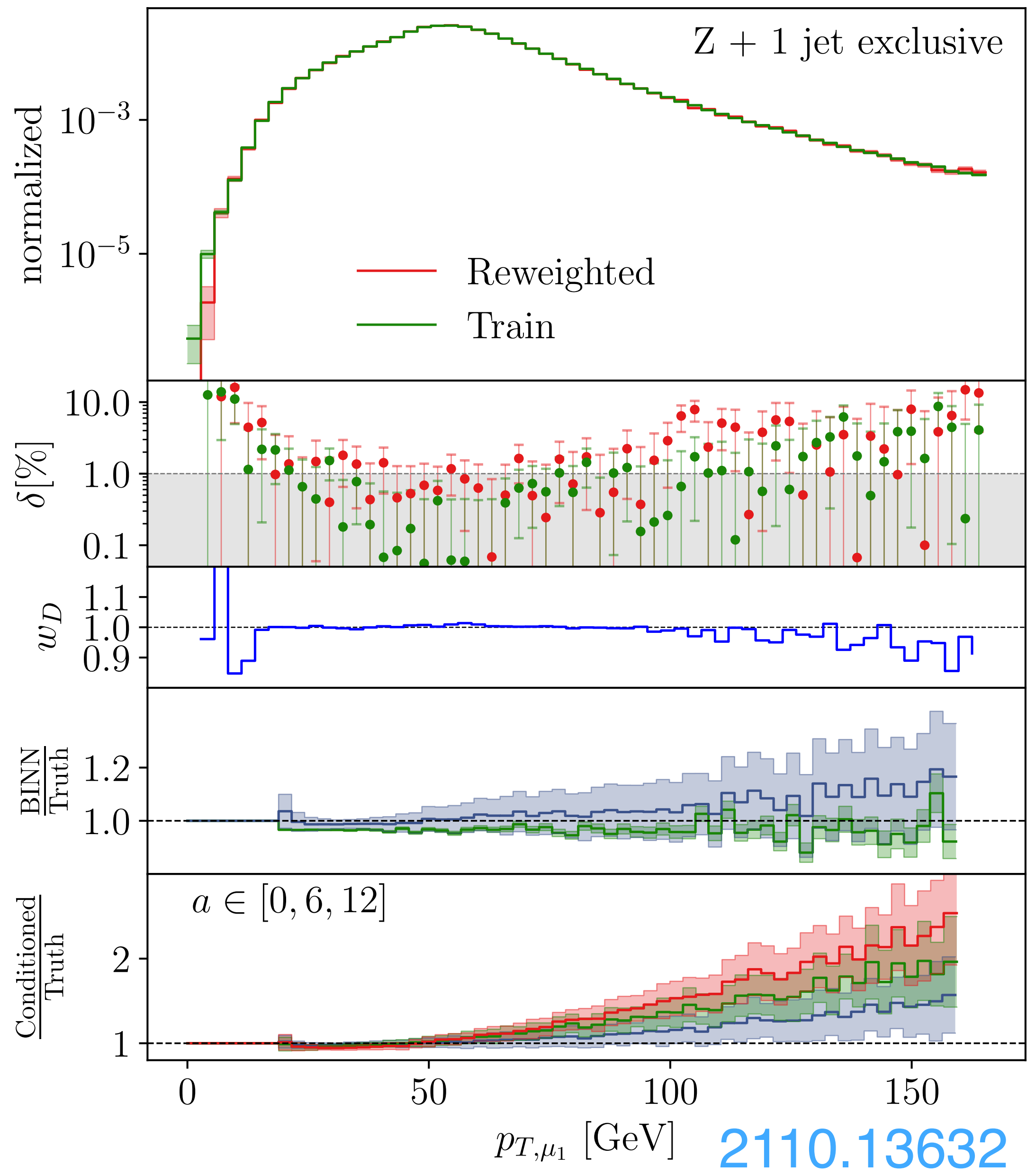
Training on samples x

$\rightarrow$ Maximize the log-likelihood

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

Monte Carlo event generation





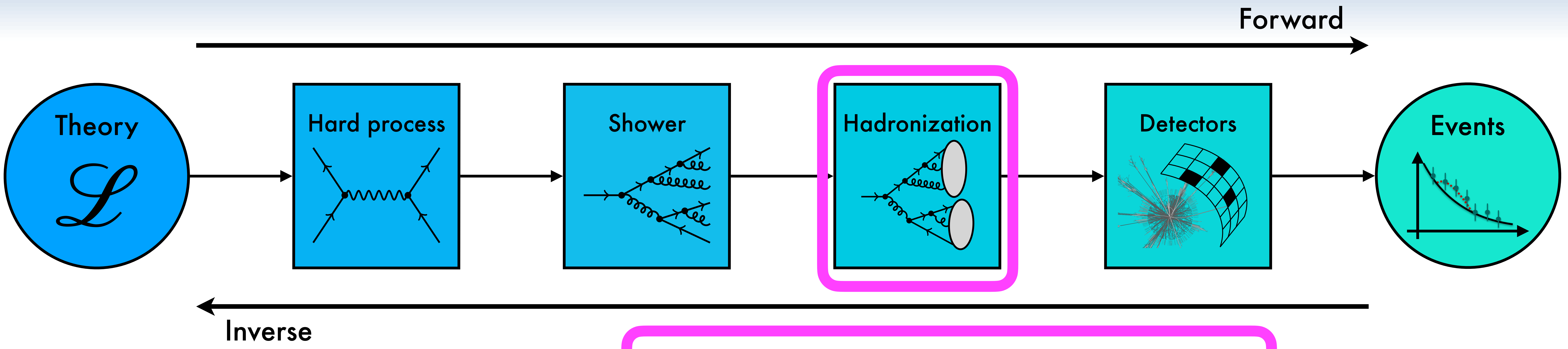
2110.13632

- 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

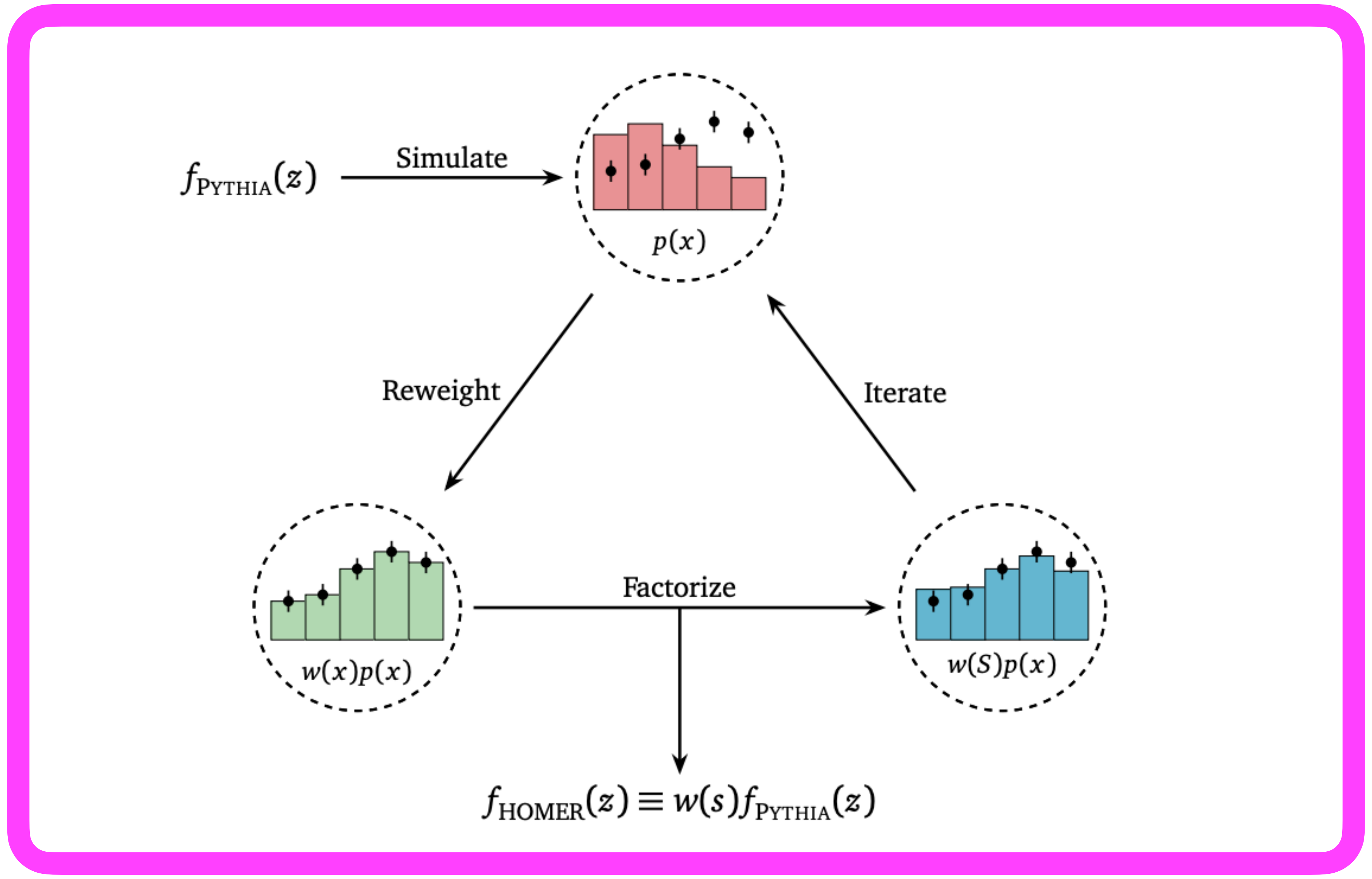
Monte Carlo event generation



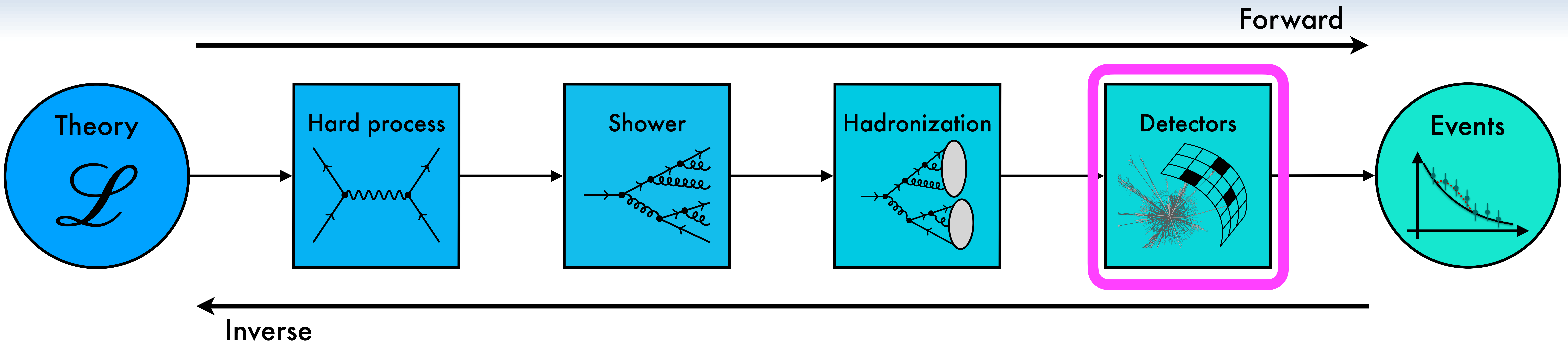
Optimal fragmentation function:

Reweighing PYTHIA to match data

[2509.03592] AB, A. Ore, S. Palacios Schweitzer, et al.



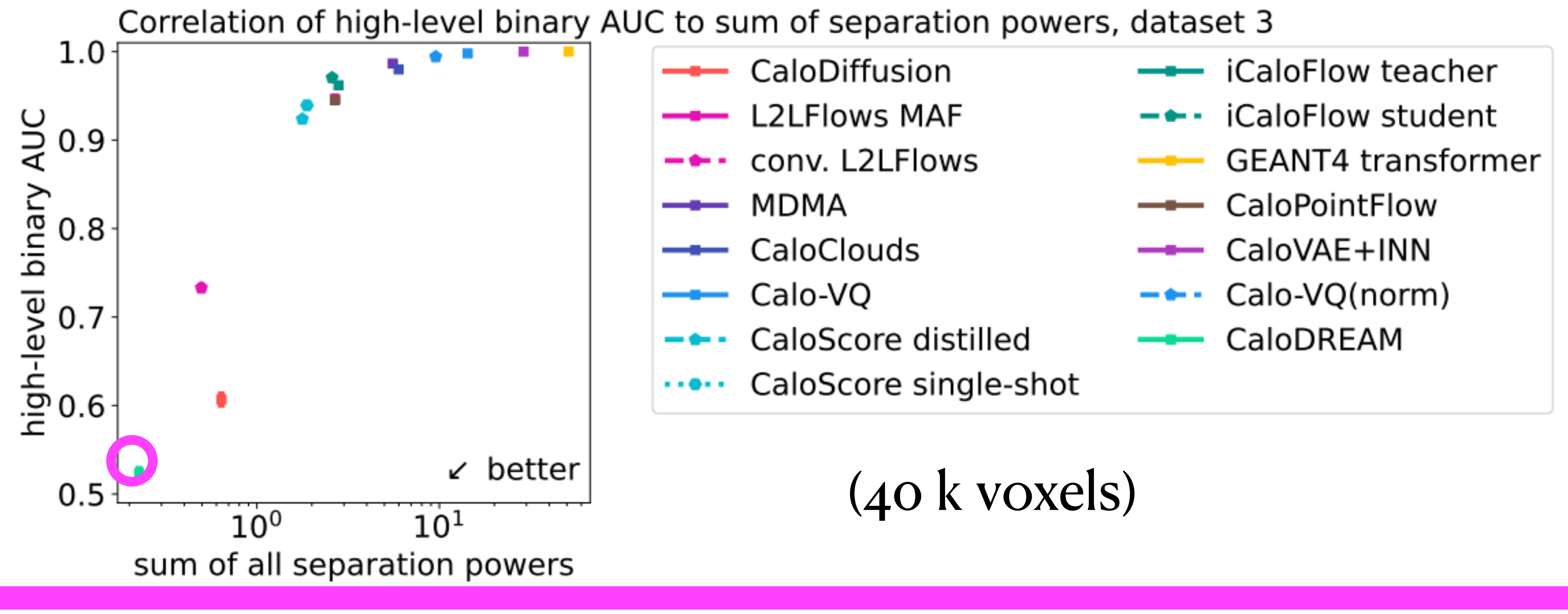
Monte Carlo event generation



Efficient detector simulations:

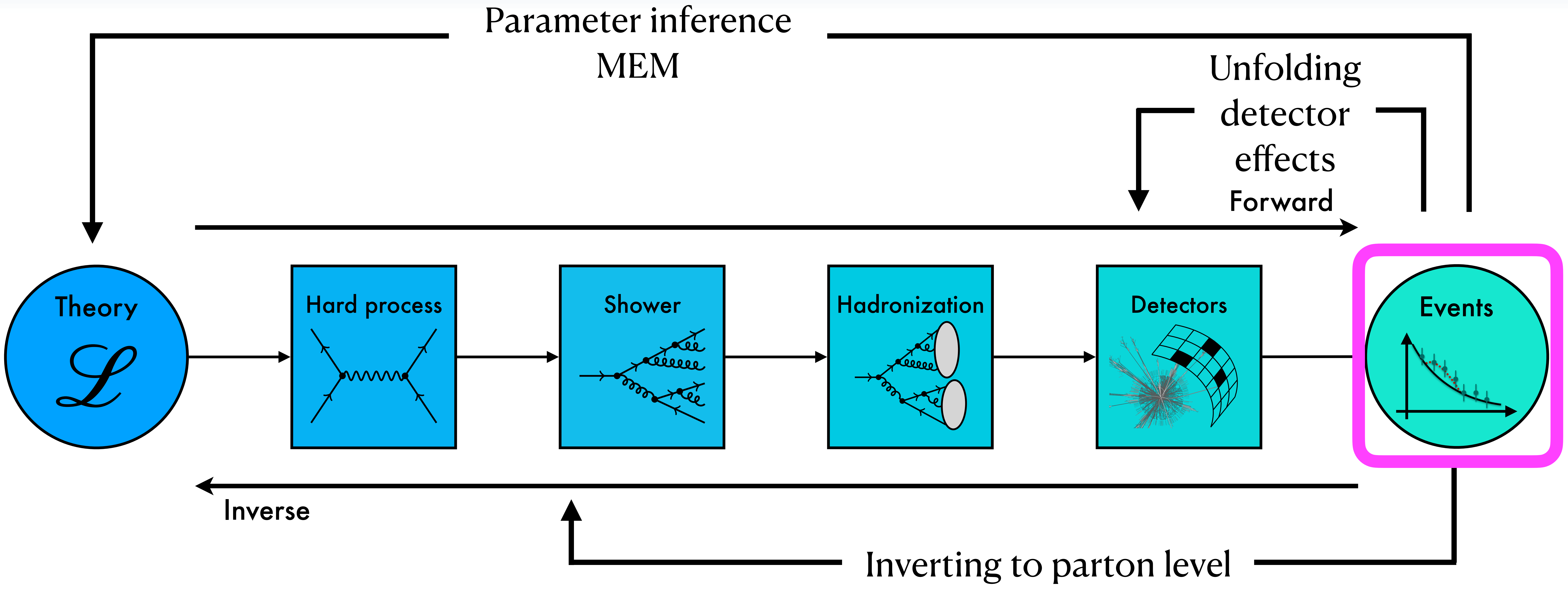
- Transfer concept from phase space generation
- Same networks (diffusion + transformers)
- Challenge: high dimensional

CaloChallenge 2022 [2410.21611]
CaloDREAM [2405.09629]
L. Favaro, A. Ore, S. Palacios Schweitzer, T. Plehn



(40 k voxels)

Inverting the simulation chain



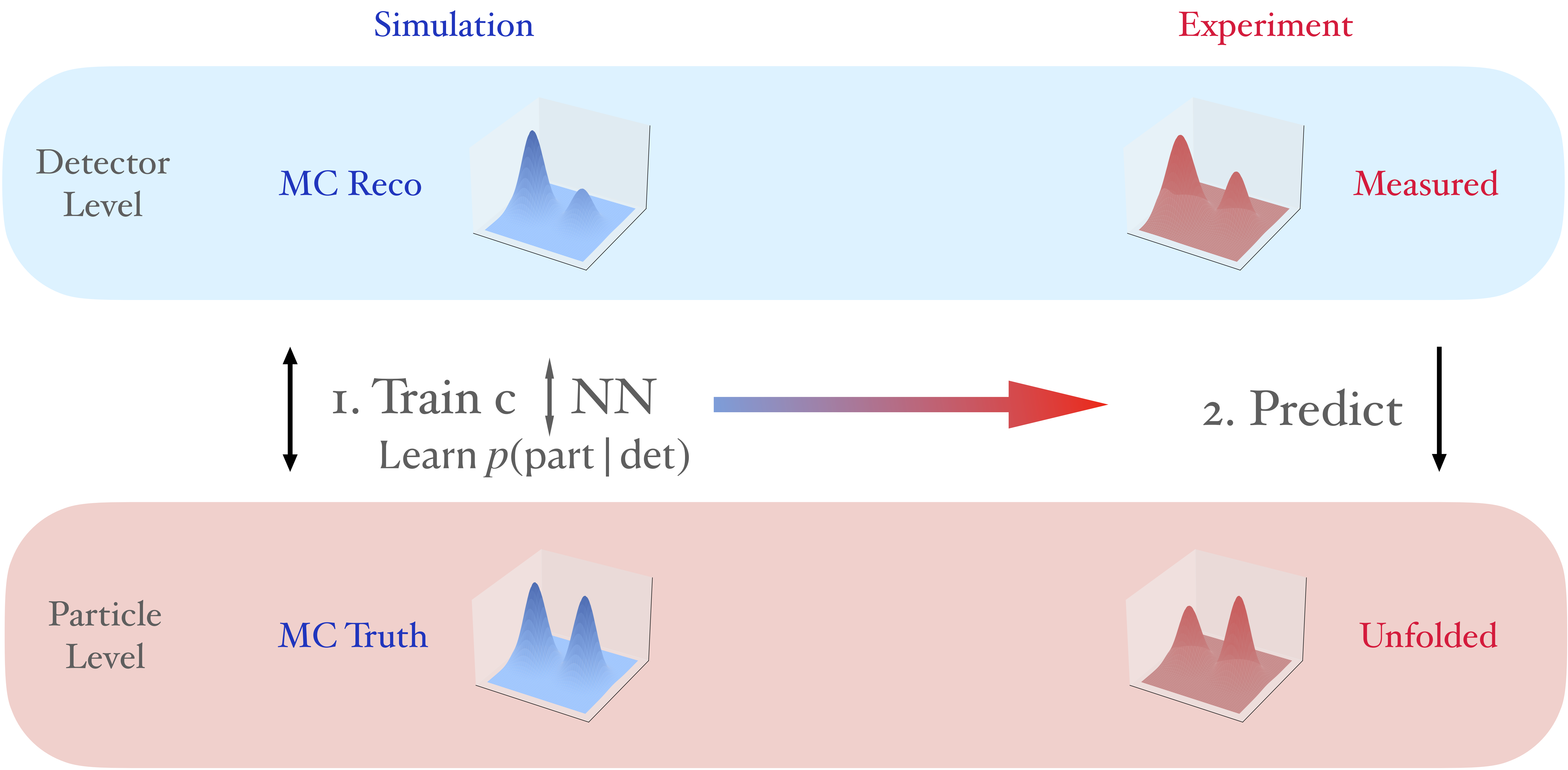
Requirements
- ☐ High - dimensional
- ☐ Bin - independent
- ☐ Statistically well defined

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

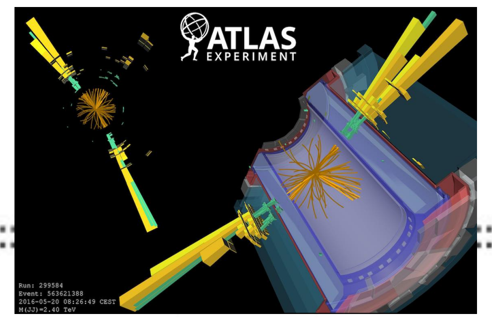
Generative unfolding methods



Generative unfolding

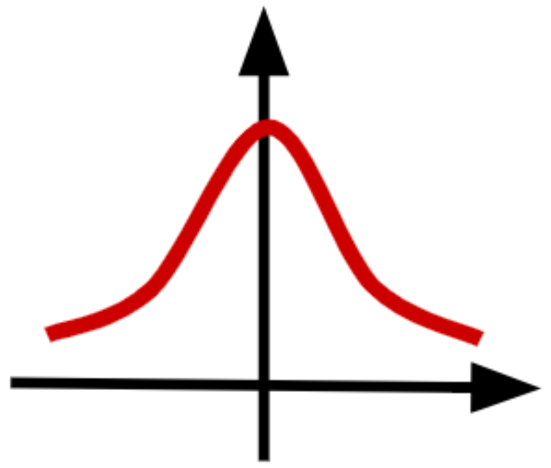
High-dimensional. Bin independent. Robust.

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

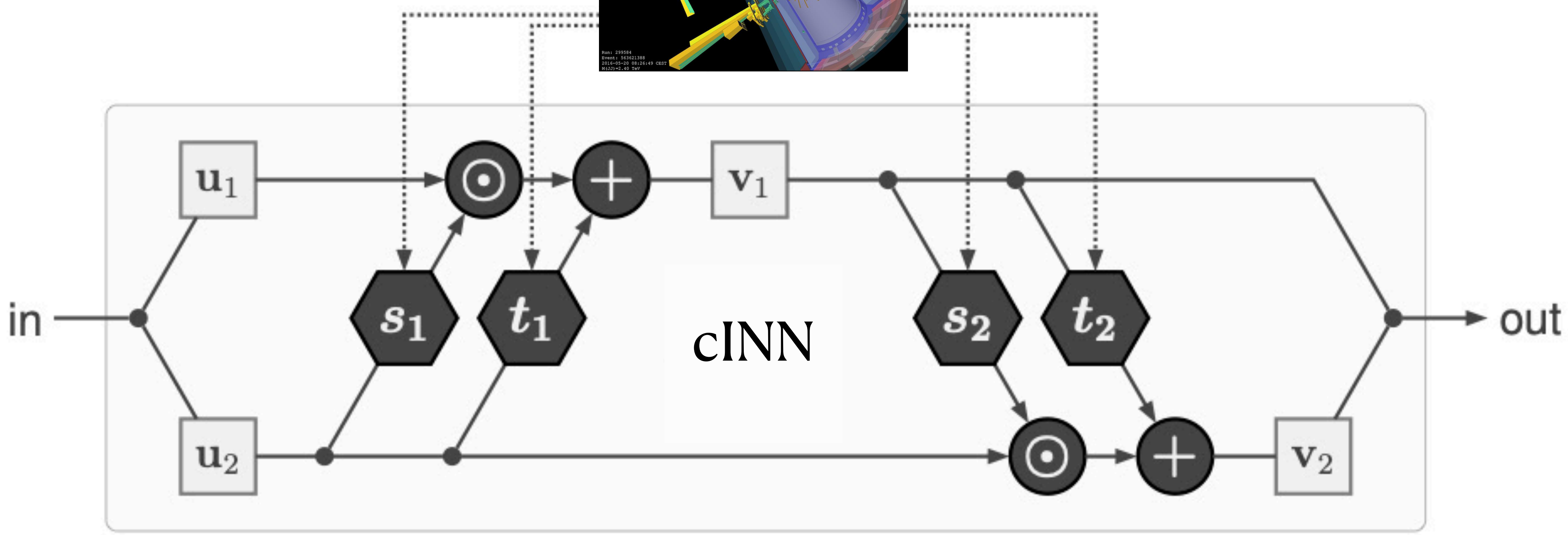


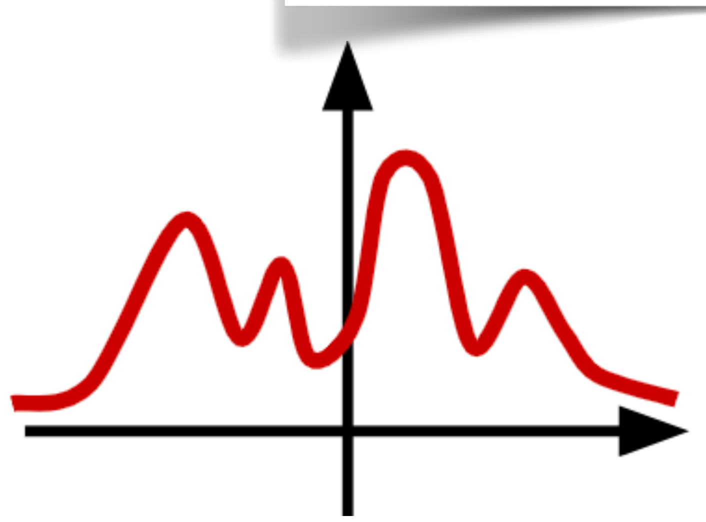
← Reconstructed objects

Unfolding:
Sample $p(x_{part} \,|\, x_{reco})$



Normal distribution
$z \sim \mathcal{N}$





Probability distribution
$x \sim P_{\text{parton}}$

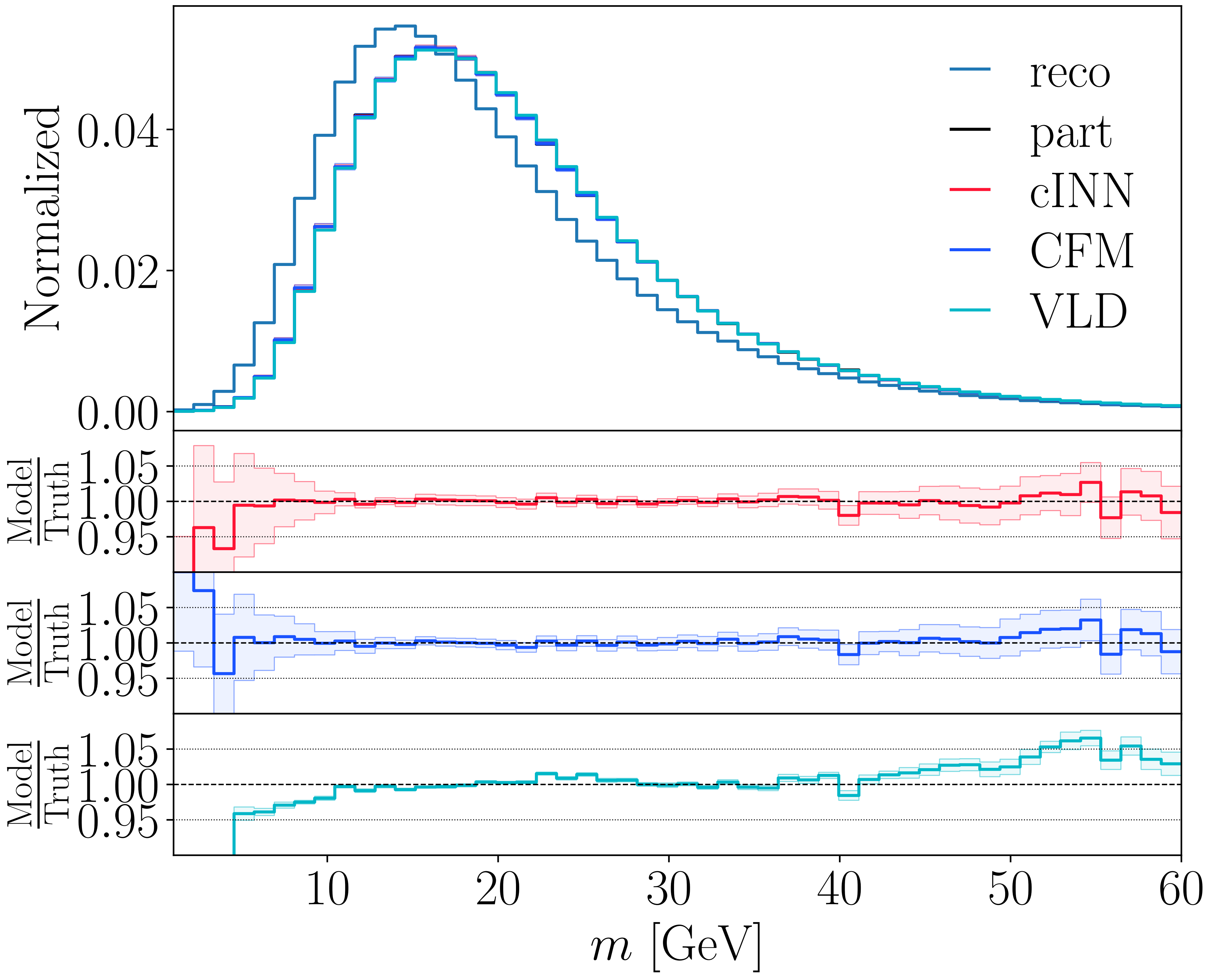
Training

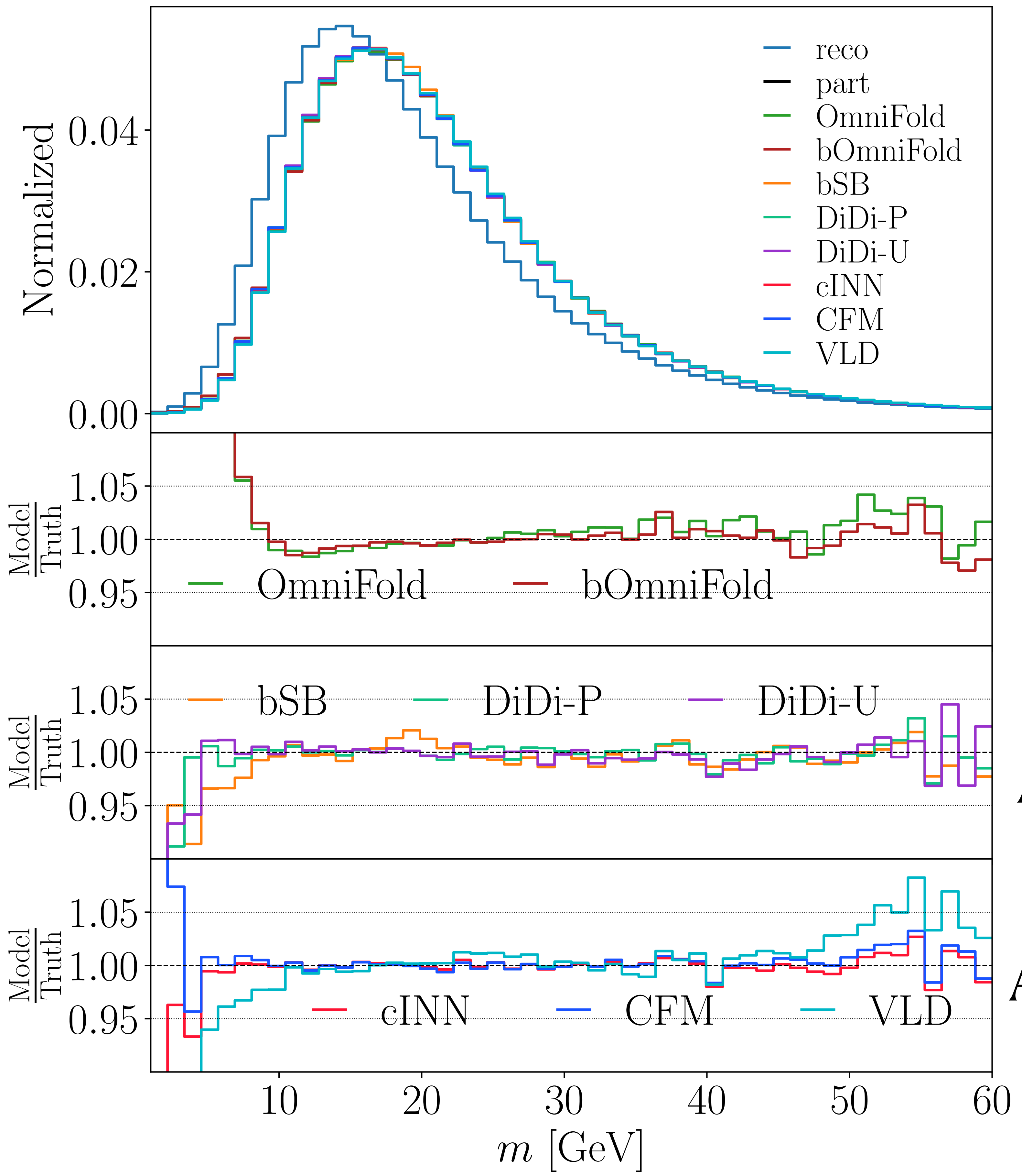
Unfolding

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

Unfolding Z+jets events

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





[1911.09107]
A. Andreassen, et al.

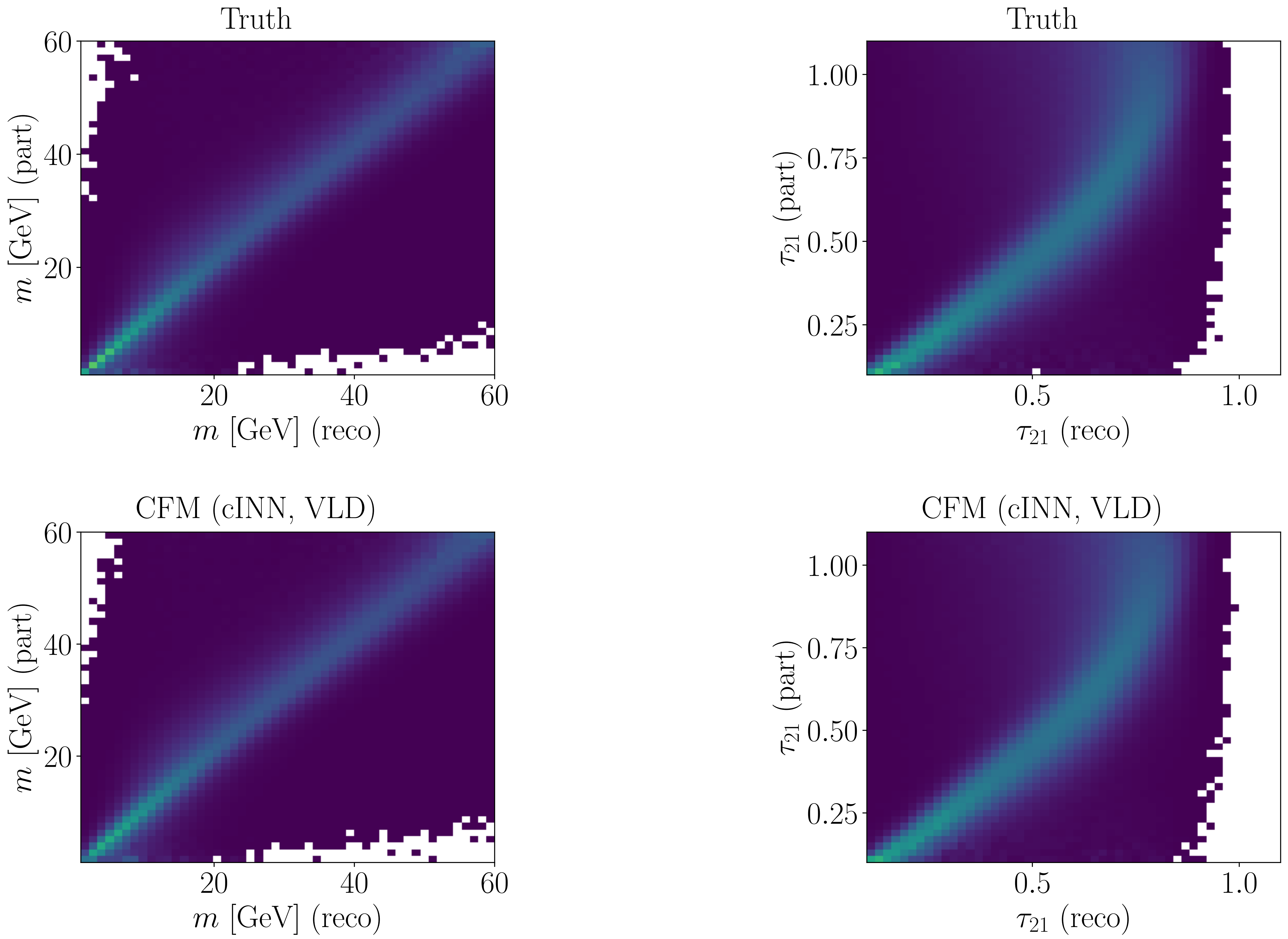
[2411.02495]
AB, S. Diefenbacher, et al.

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

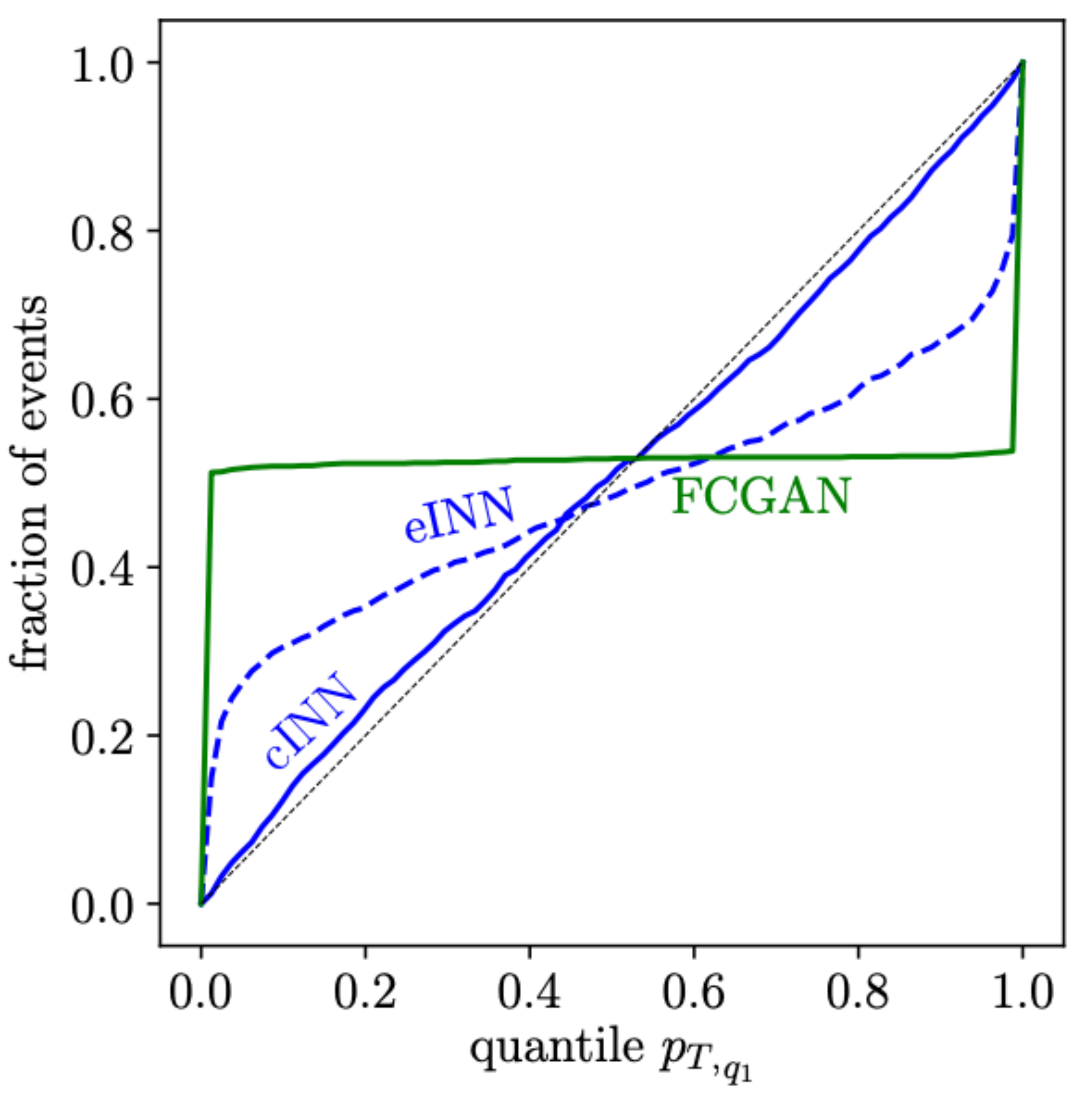


[2404.18807] N. Hütsch, et al.

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]

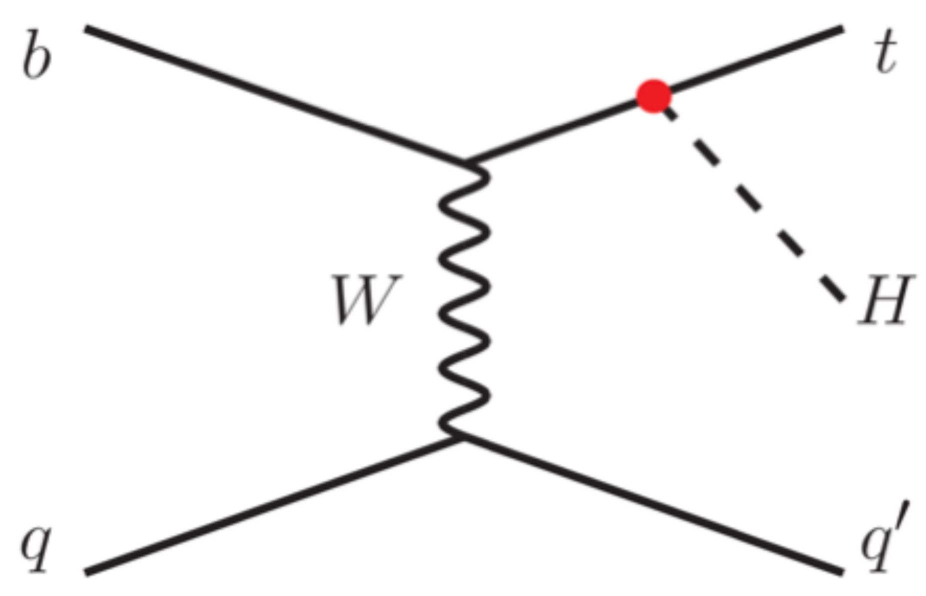
Matrix element method is based on untractable likelihood

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

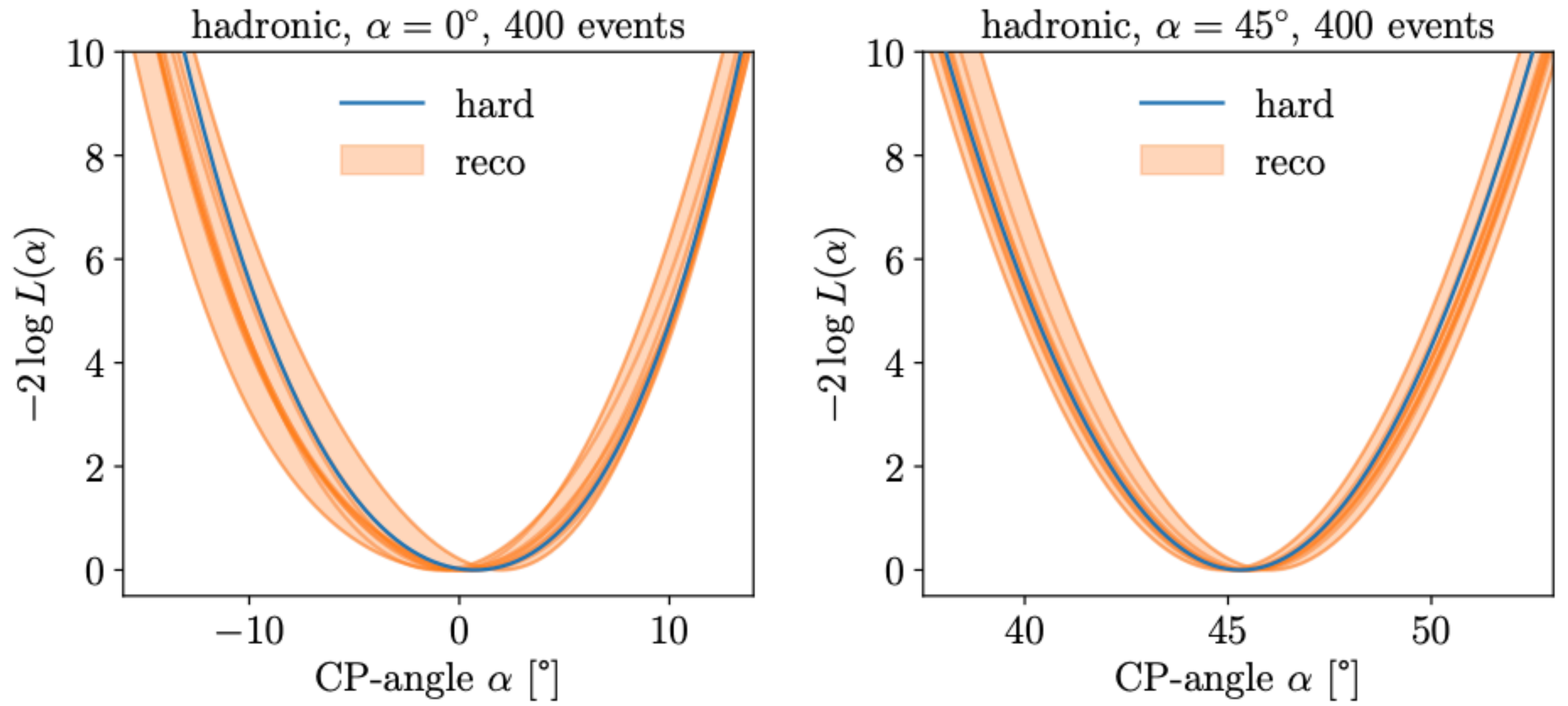
Problem: integration over full phase space of the hard scattering

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

Fitting data with uncertainties

$f(x) = y \pm \sigma$

for

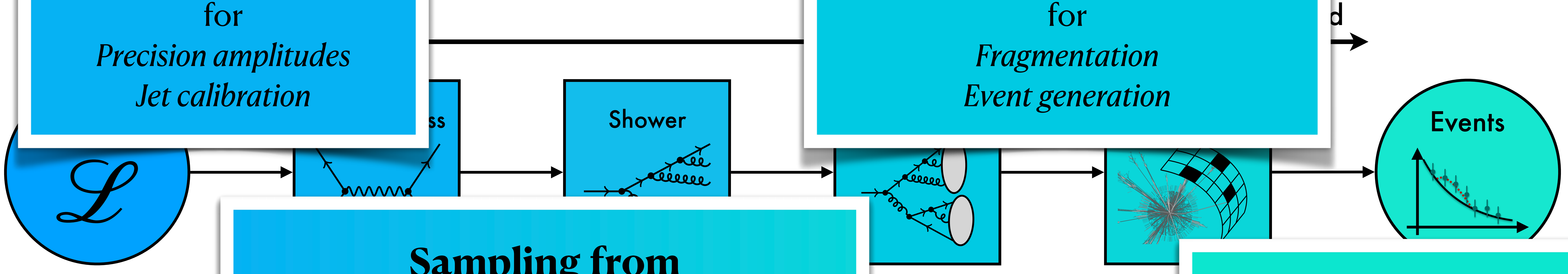
Precision amplitudes
Jet calibration

Reweighting simulations to match data

$x \sim p(x)$

for

Fragmentation
Event generation



Sampling from accurately learned distributions

$x \sim p(x)$

for

Phase space generation
Event generation
Detector simulation

Unfolding unbinned measurements

$p(x \mid y)$

for

Unfolding to particle/parton level
Matrix element method

What's next?

What's next?

MadAgents

Tilman Plehn[1,2], Daniel Schiller[1], and Nikita Schmal[1]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, Germany

February 11, 2026

Abstract

We uncover an effective and communicative set of agents working with MadGraph. Agentic installation, learning-by-doing training, and user support provide easy access to state-of-the-art simulations and accelerate LHC research. We show in detail how MadAgents interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MadAgents automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

https://github.com/heidelberg-hepml/MadAgents

What’s next?

MadAgents

Tilman Plehn[1,2], Daniel Schiller[1], and … hmal[1]

1 Institut für Theoretische Physik, U… Germany
2 Interdisciplinary Center for Scienti… Heidelberg, Germany

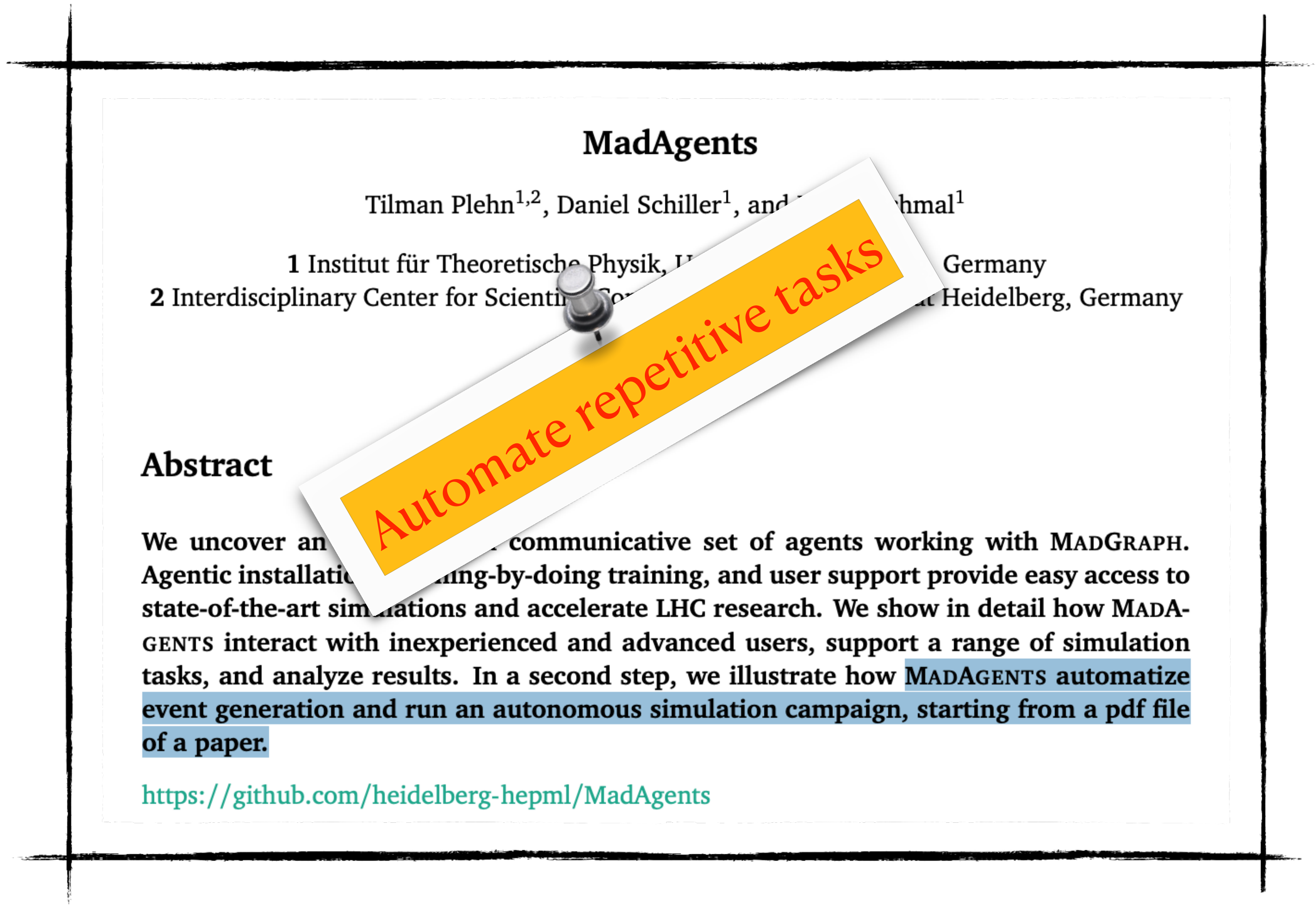
Automate repetitive tasks

Abstract

We uncover an … communicative set of agents working with MadGraph. Agentic installati… ing-by-doing training, and user support provide easy access to state-of-the-art sim…ations and accelerate LHC research. We show in detail how MadAgents interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MadAgents automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

https://github.com/heidelberg-hepml/MadAgents

What's next?



MadAgents

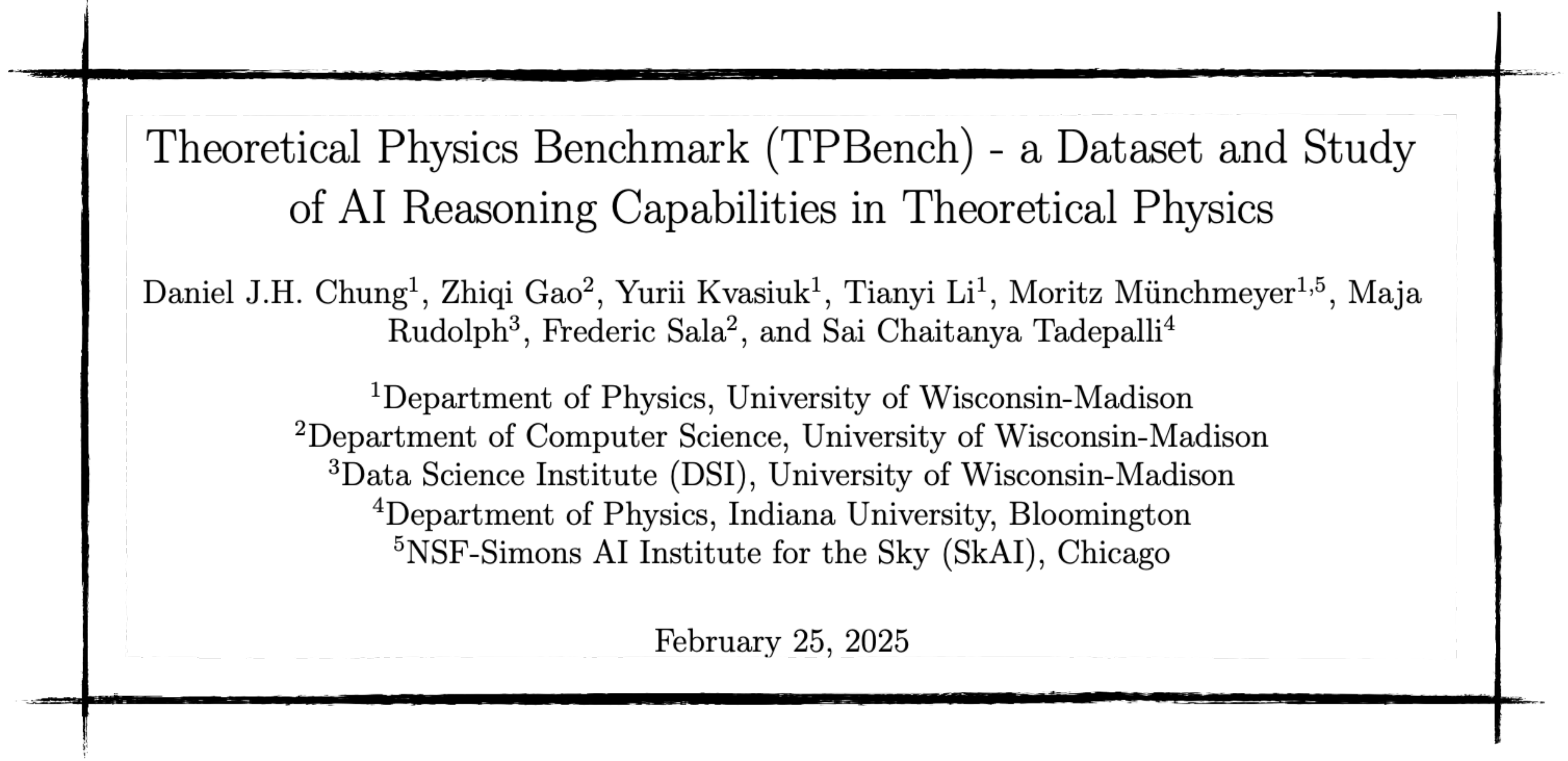
Tilman Plehn[1,2], Daniel Schiller[1], and … hmal[1]

1 Institut für Theoretische Physik, … Germany
2 Interdisciplinary Center for Scienti… Heidelberg, Germany

Abstract

We uncover an … communicative set of agents working with MadGraph. Agentic installatio… ning-by-doing training, and user support provide easy access to state-of-the-art sim…ations and accelerate LHC research. We show in detail how MadAgents interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MadAgents automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

https://github.com/heidelberg-hepml/MadAgents

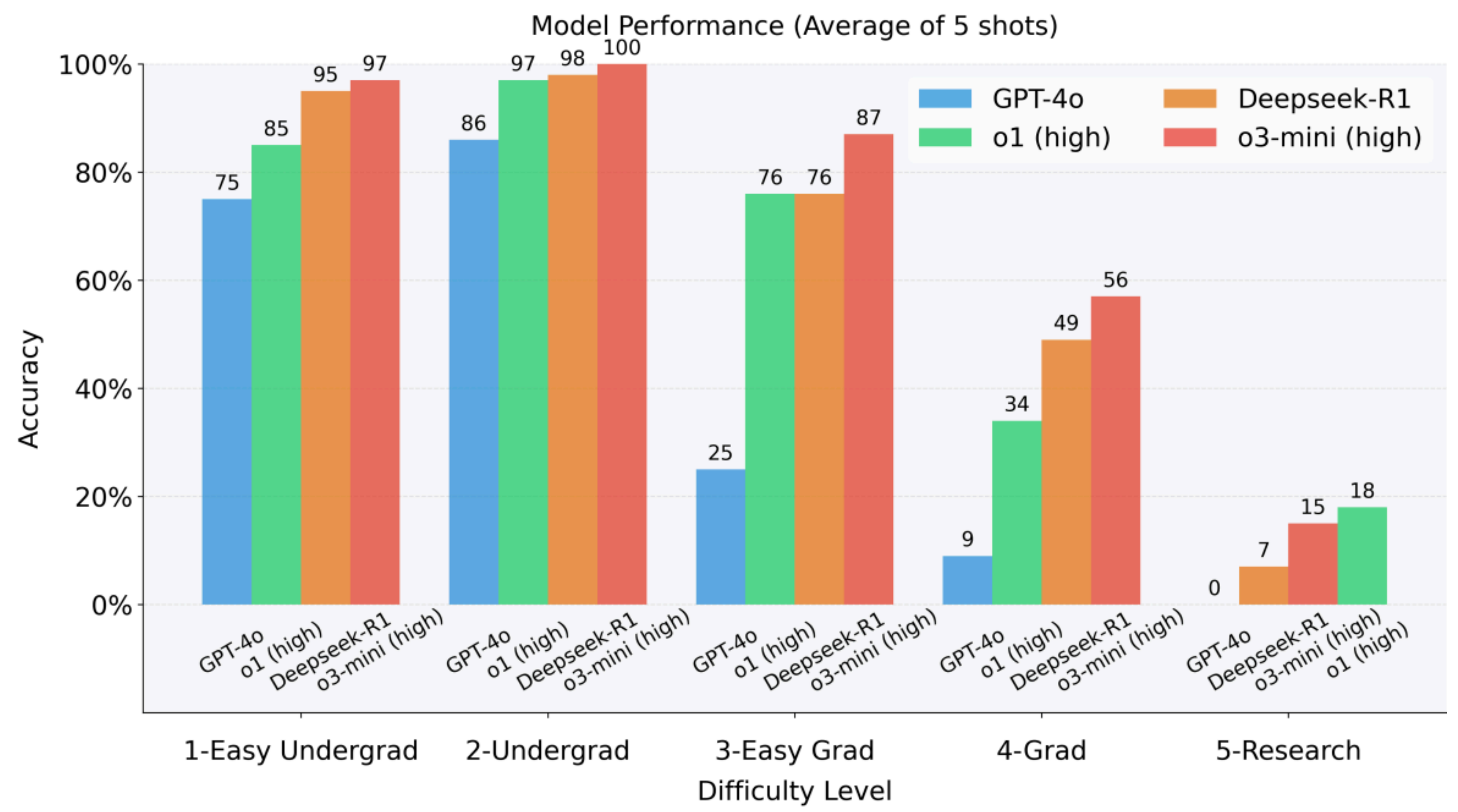


Theoretical Physics Benchmark (TPBench) - a Dataset and Study of AI Reasoning Capabilities in Theoretical Physics

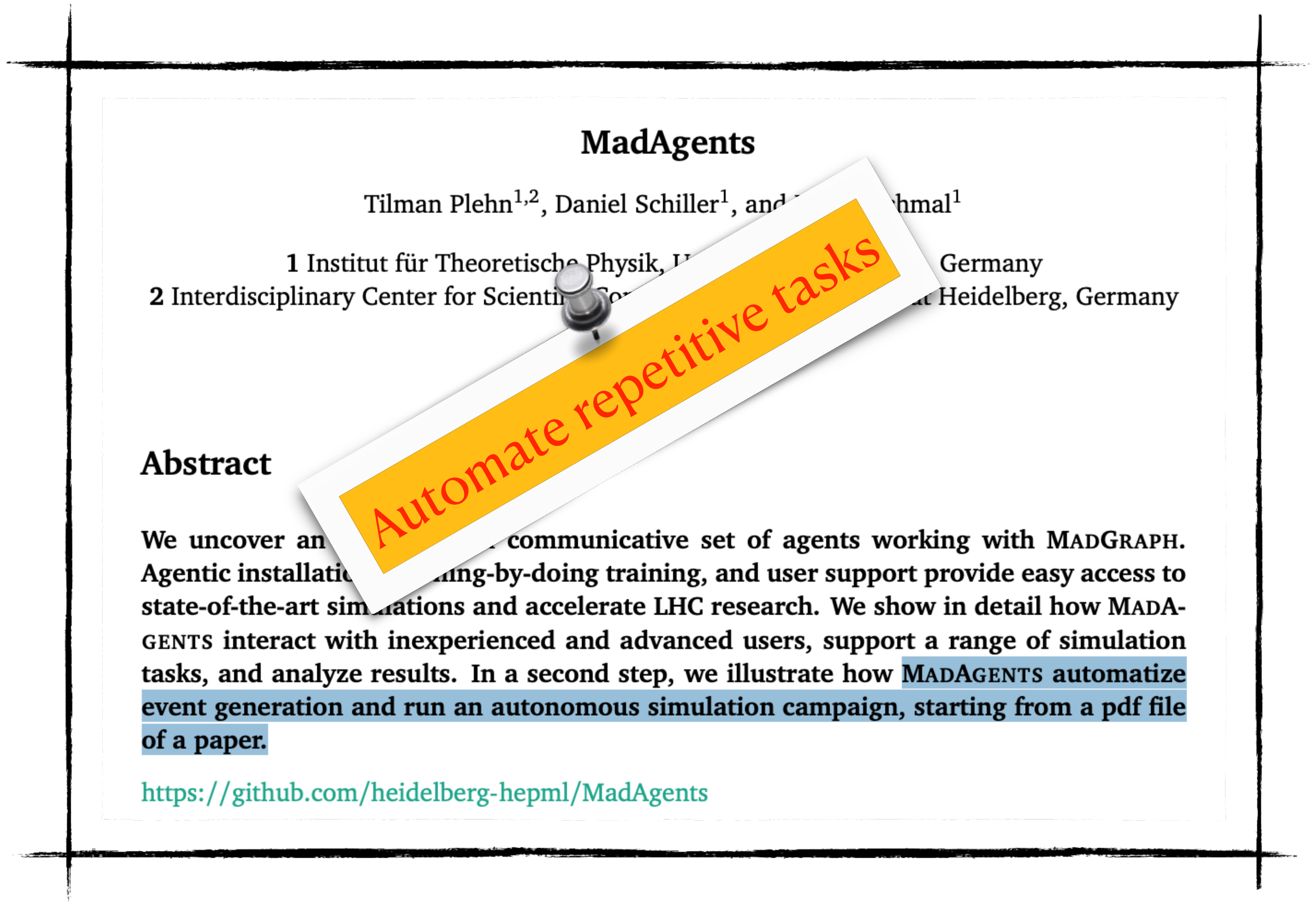
Daniel J.H. Chung[1], Zhiqi Gao[2], Yurii Kvasiuk[1], Tianyi Li[1], Moritz Münchmeyer[1,5], Maja Rudolph[3], Frederic Sala[2], and Sai Chaitanya Tadepalli[4]

[1]Department of Physics, University of Wisconsin-Madison
[2]Department of Computer Science, University of Wisconsin-Madison
[3]Data Science Institute (DSI), University of Wisconsin-Madison
[4]Department of Physics, Indiana University, Bloomington
[5]NSF-Simons AI Institute for the Sky (SkAI), Chicago

February 25, 2025



What's next?



MadAgents

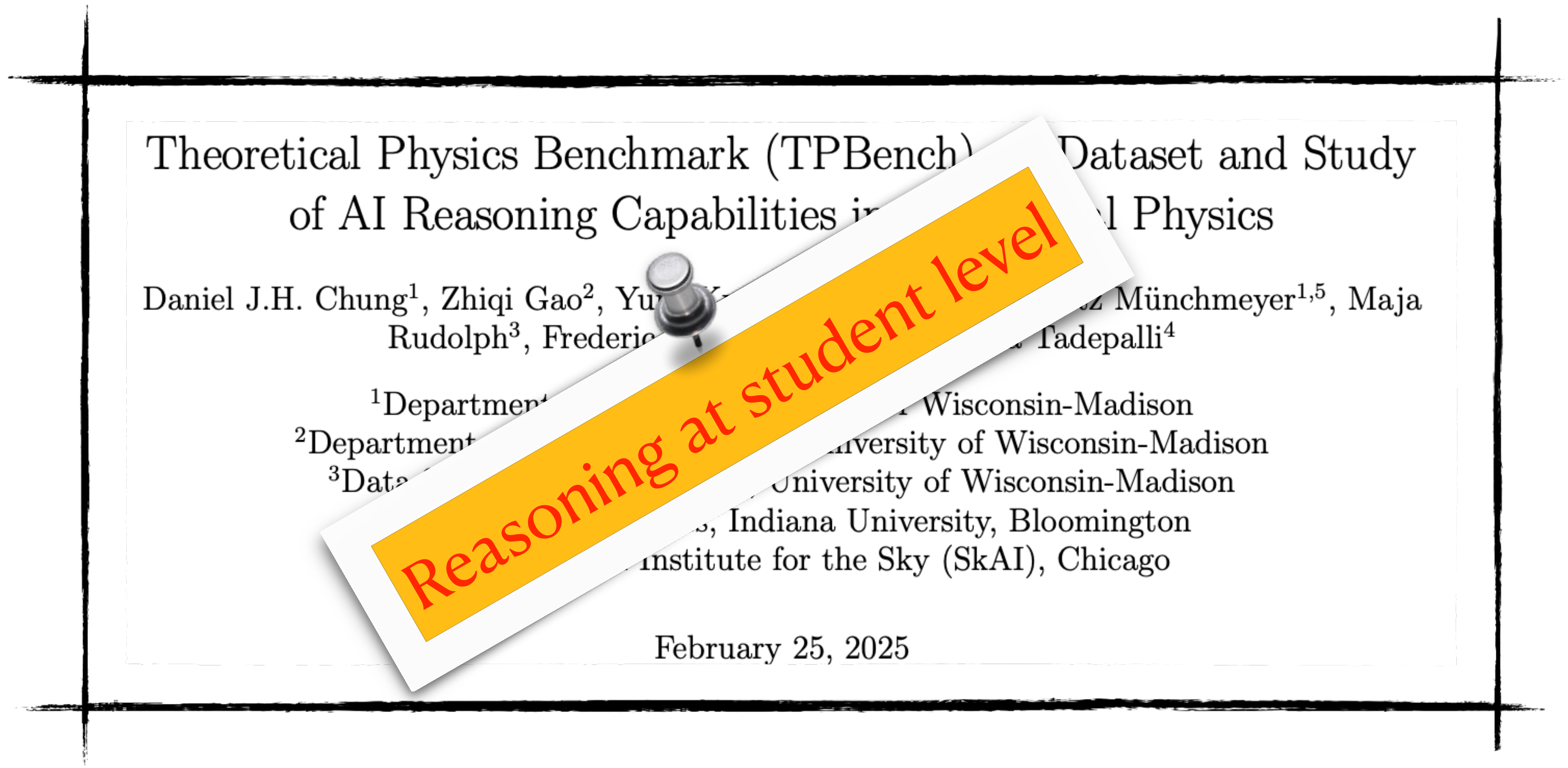
Tilman Plehn[1,2], Daniel Schiller[1], and ...hmal[1]

1 Institut für Theoretische Physik, ... Germany
2 Interdisciplinary Center for Scienti... Heidelberg, Germany

Abstract

We uncover an ... communicative set of agents working with MADGRAPH. Agentic installati... ...ing-by-doing training, and user support provide easy access to state-of-the-art sim...ations and accelerate LHC research. We show in detail how MADAGENTS interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MADAGENTS automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

https://github.com/heidelberg-hepml/MadAgents

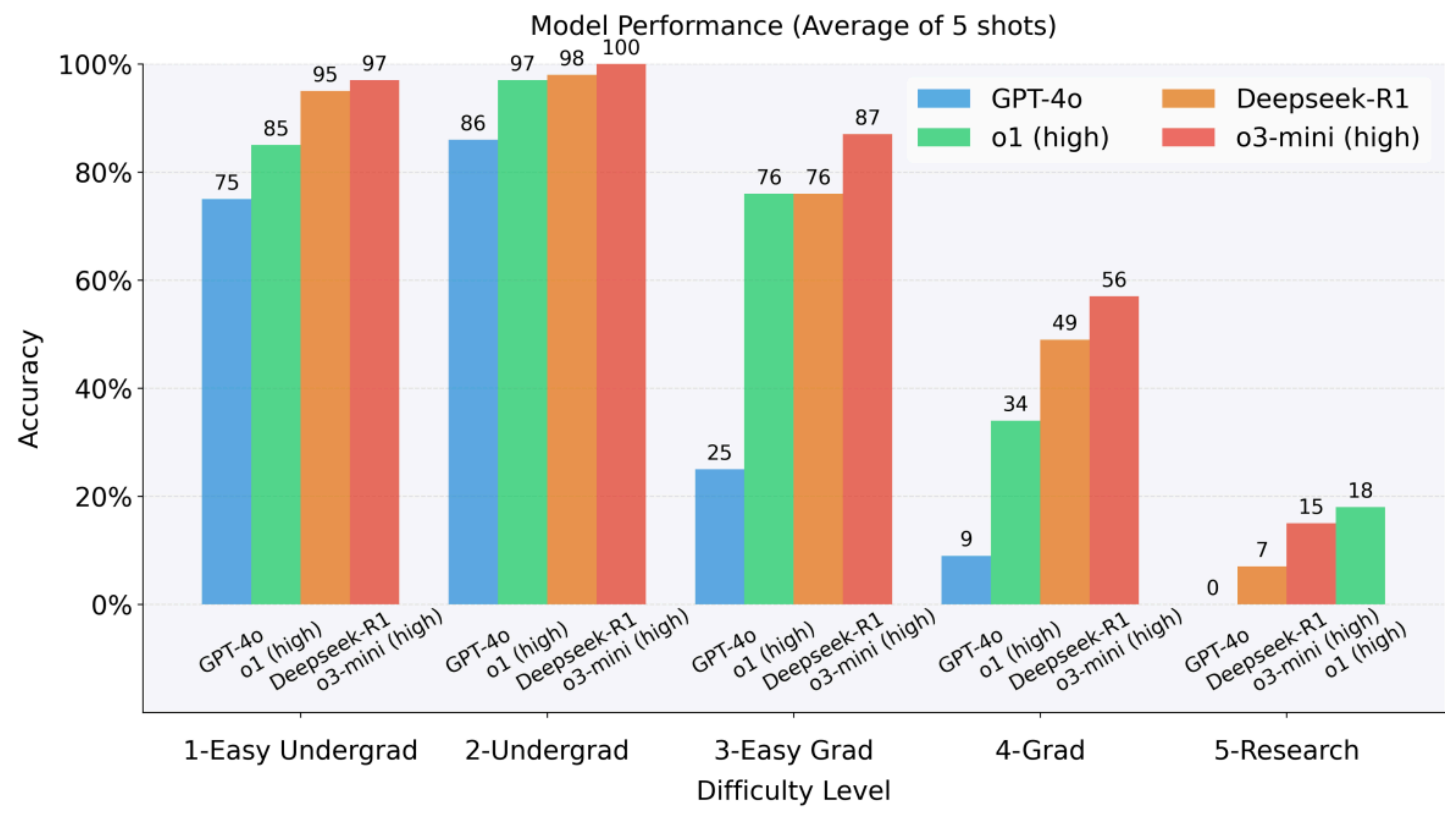


Theoretical Physics Benchmark (TPBench) ... Dataset and Study of AI Reasoning Capabilities i... ...l Physics

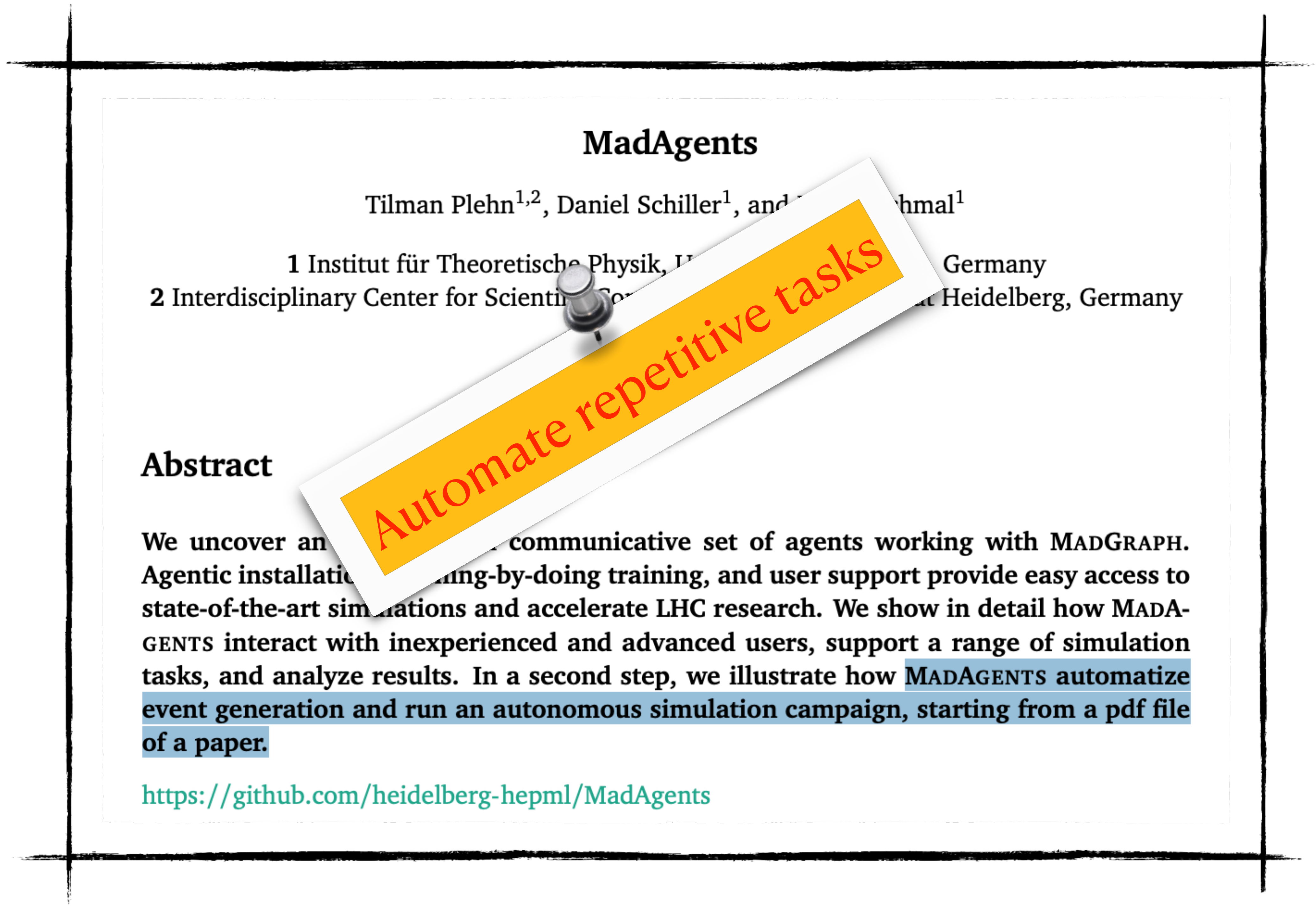
Daniel J.H. Chung[1], Zhiqi Gao[2], Yu... ...z Münchmeyer[1,5], Maja Rudolph[3], Frederi... ...Tadepalli[4]

[1]Departmen... ...Wisconsin-Madison
[2]Departmen... ...iversity of Wisconsin-Madison
[3]Data... ...University of Wisconsin-Madison
...s, Indiana University, Bloomington
...Institute for the Sky (SkAI), Chicago

February 25, 2025



What's next?

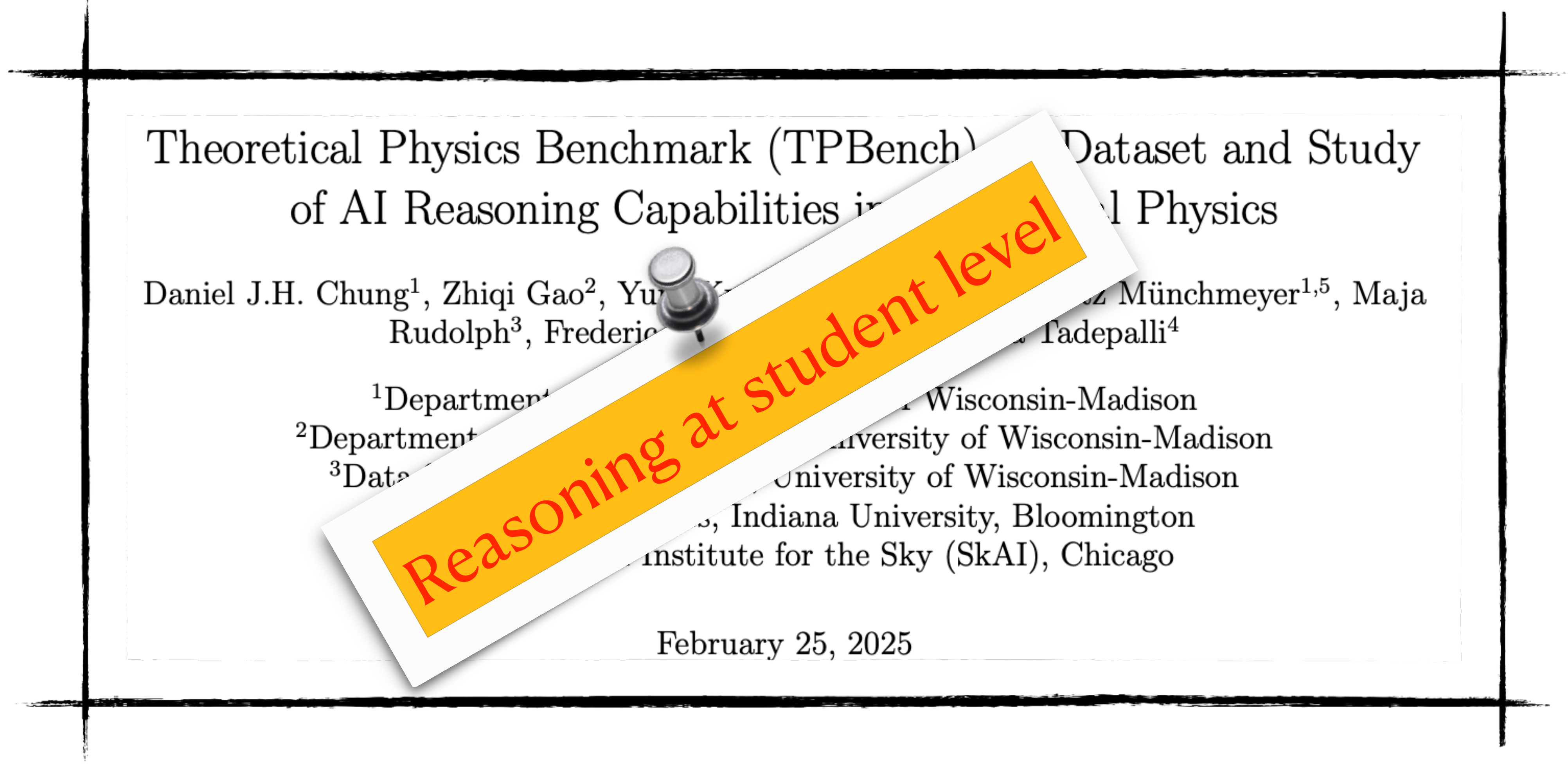

MadAgents

Automate repetitive tasks

Abstract

communicative set of agents working with MadGraph. state-of-the-art sim ations and accelerate LHC research. We show in detail how MadAgents interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MadAgents automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

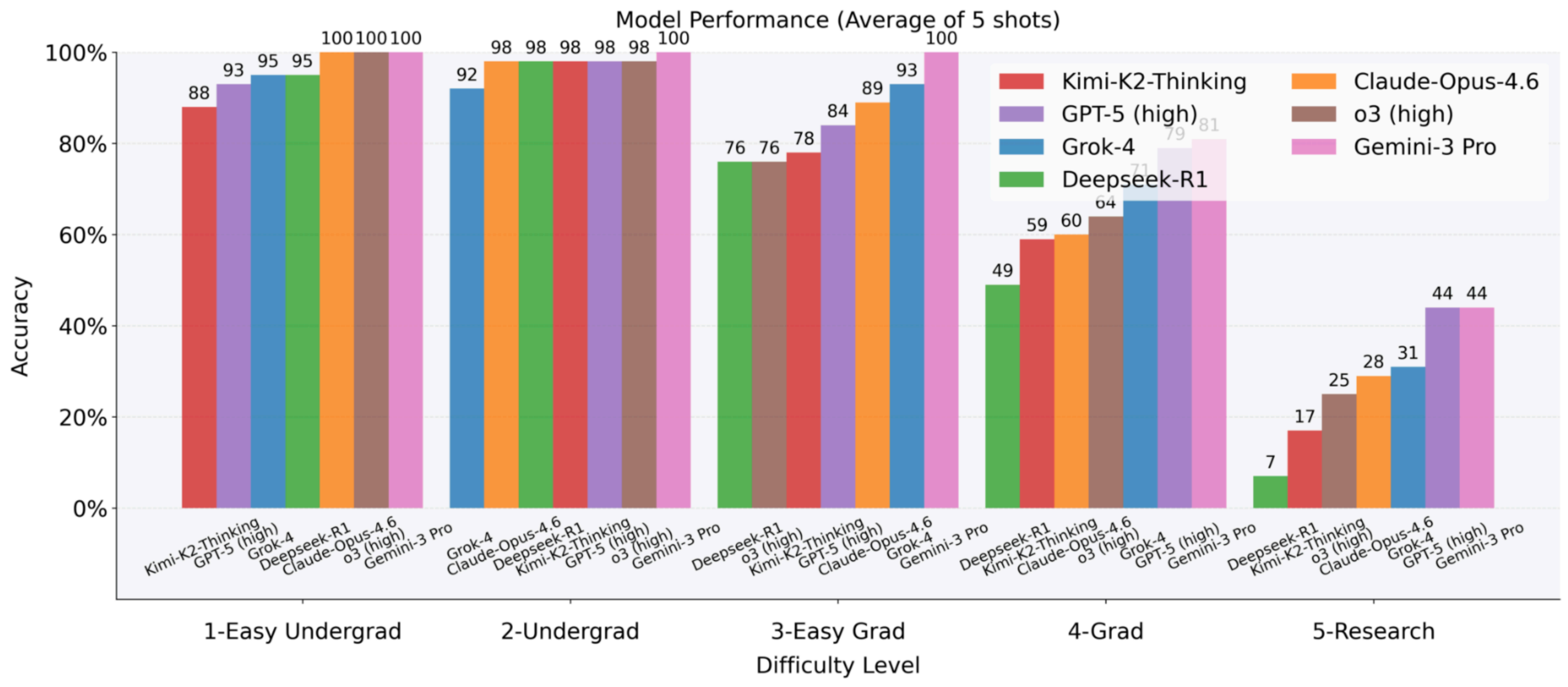
https://github.com/heidelberg-hepml/MadAgents


Theoretical Physics Benchmark (TPBench) Dataset and Study of AI Reasoning Capabilities in Physics

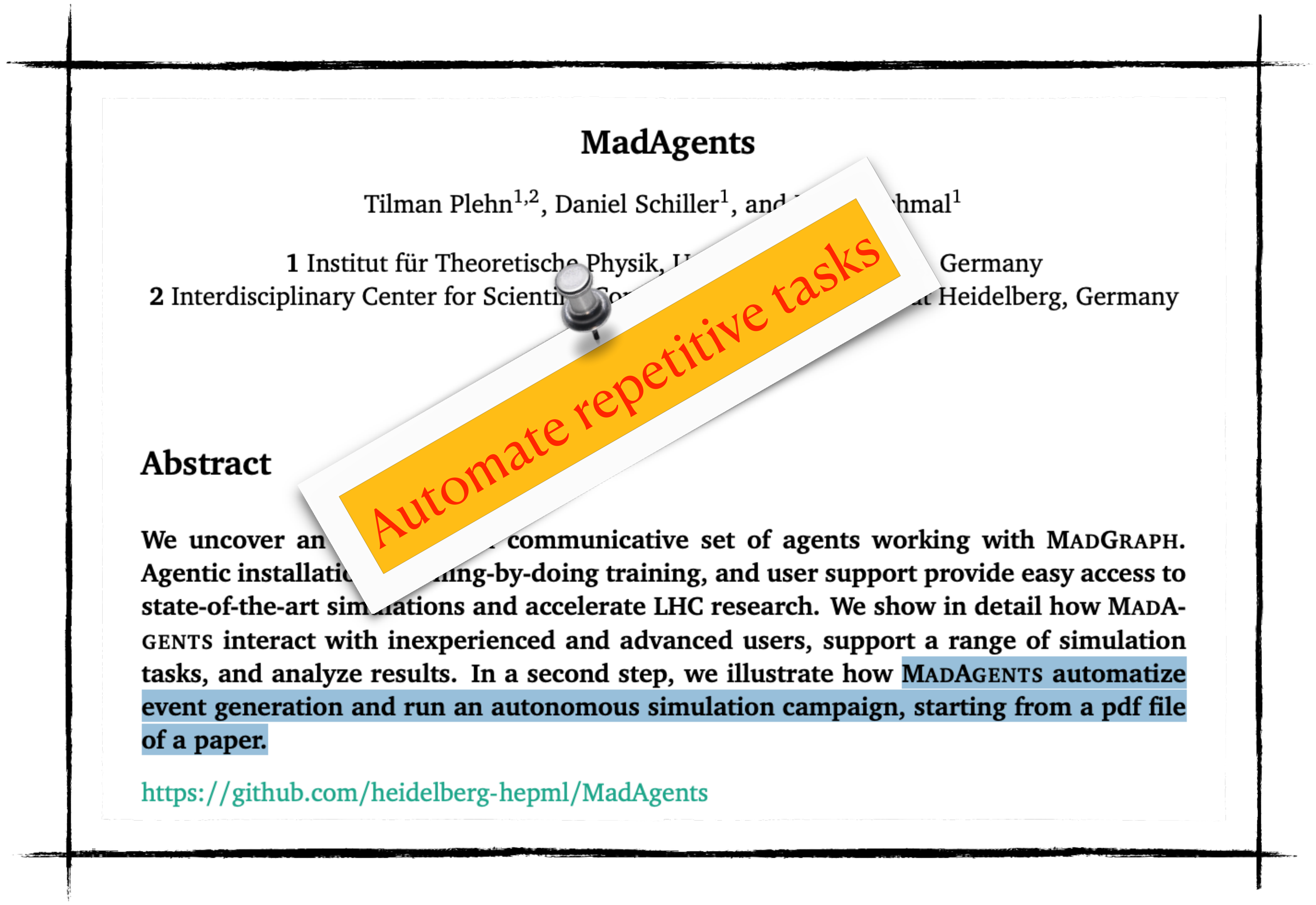
Reasoning at student level

Institute for the Sky (SkAI), Chicago

February 25, 2025



What's next?


MadAgents

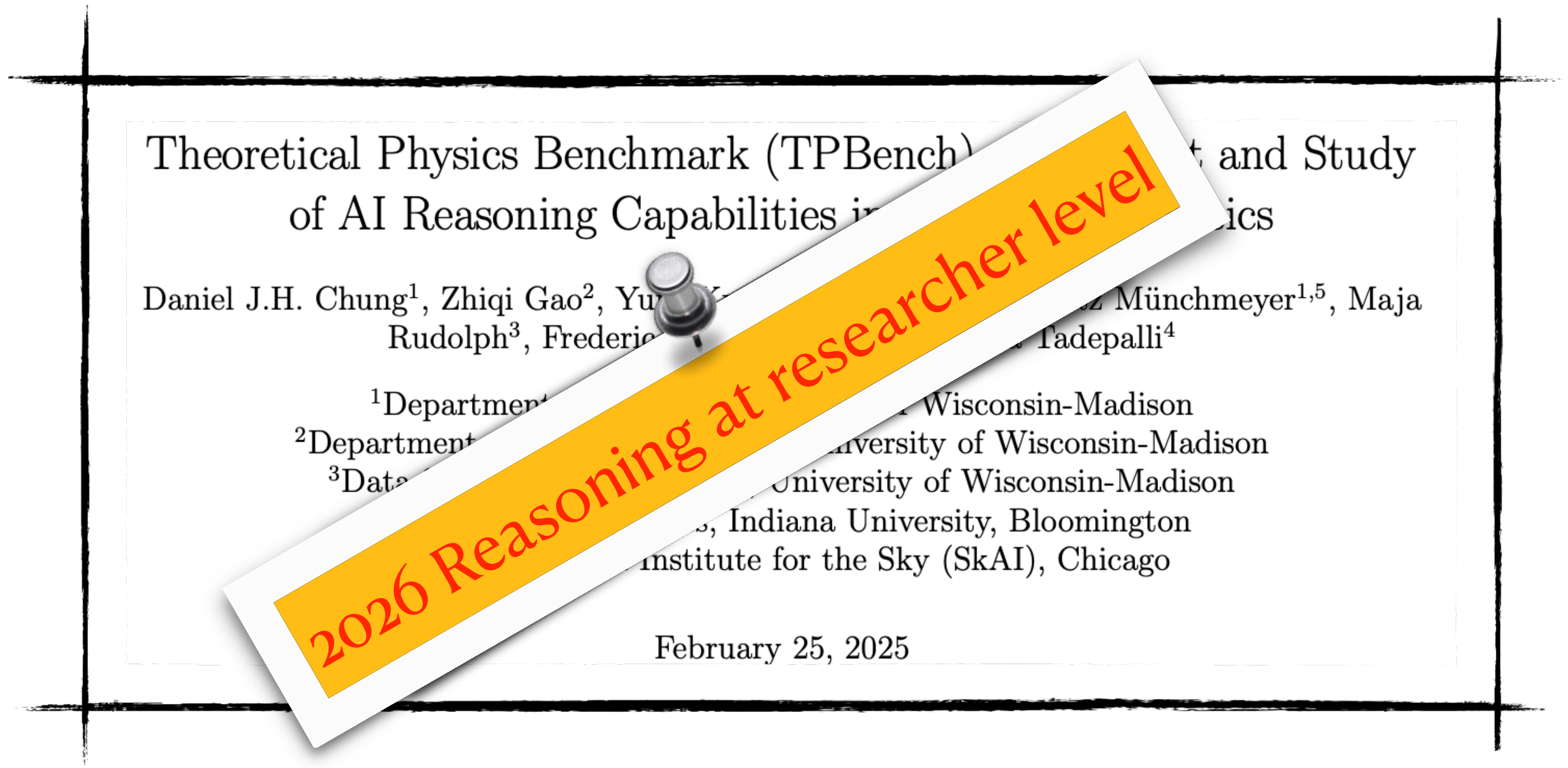
Tilman Plehn[1,2], Daniel Schiller[1], and …hmal[1]

1 Institut für Theoretische Physik, U… Germany
2 Interdisciplinary Center for Scienti… Heidelberg, Germany

Abstract

We uncover an … communicative set of agents working with MadGraph. Agentic installati…ing-by-doing training, and user support provide easy access to state-of-the-art sim…ations and accelerate LHC research. We show in detail how MadAgents interact with inexperienced and advanced users, support a range of simulation tasks, and analyze results. In a second step, we illustrate how MadAgents automatize event generation and run an autonomous simulation campaign, starting from a pdf file of a paper.

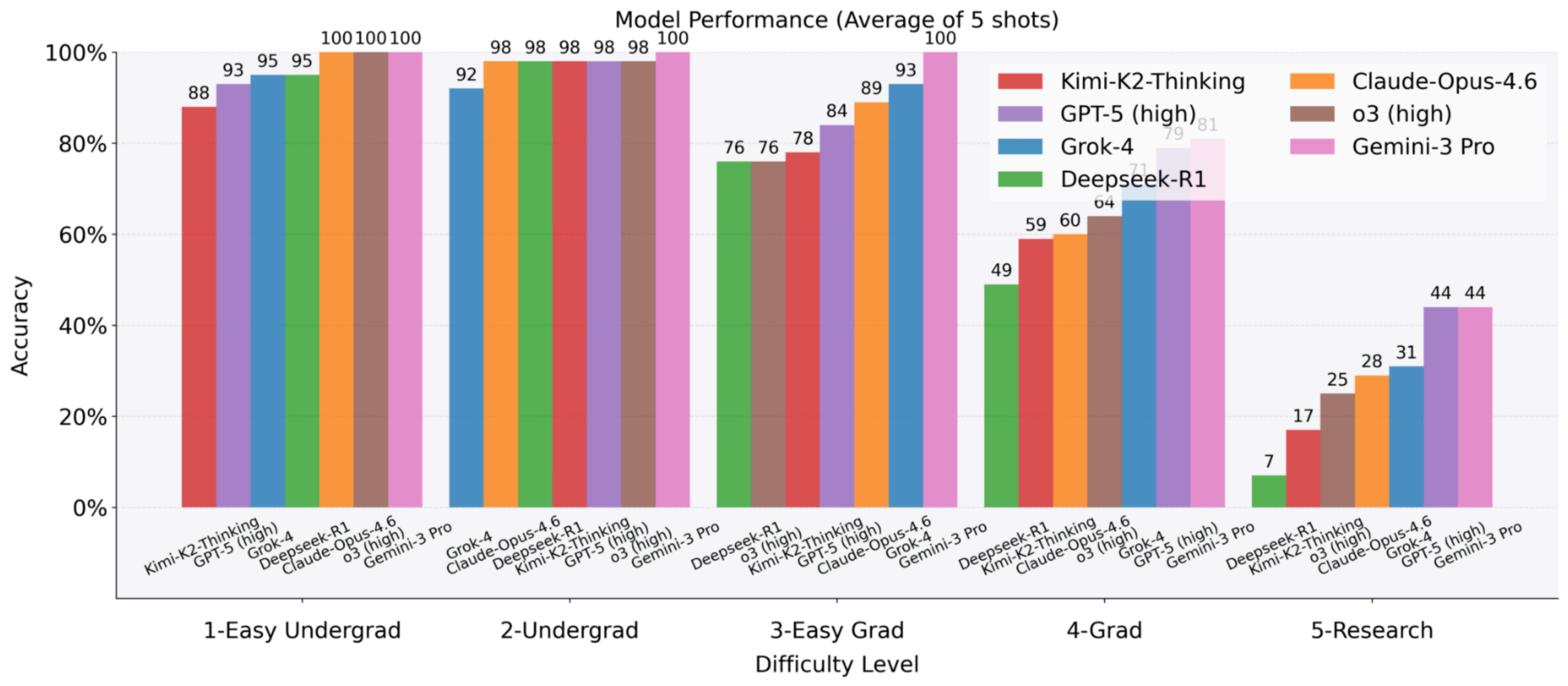
https://github.com/heidelberg-hepml/MadAgents


Theoretical Physics Benchmark (TPBench) … and Study of AI Reasoning Capabilities i… …ics

Daniel J.H. Chung[1], Zhiqi Gao[2], Yu… …z Münchmeyer[1,5], Maja Rudolph[3], Frederic… …Tadepalli[4]

[1]Departmen… …Wisconsin-Madison
[2]Departmen… …iversity of Wisconsin-Madison
[3]Data… …University of Wisconsin-Madison
…s, Indiana University, Bloomington
…Institute for the Sky (SkAI), Chicago

February 25, 2025



What's next?

Tilman Plehn[1,2],

1 Institut für Theoretisc

2 Interdisciplinary Center for Scien

Abstract

Automat

We uncover an … con
Agentic installati… ng-by-d
state-of-the-art sim… ations 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-hepr

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

Matthew D. Schwartz[1,2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*

[2] *Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)*

schwartz@g.harvard.edu

AI RESEARCH ASSISTANT: Claude Opus 4.5 (Anthropic)

January 7, 2026

Author Contributions

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.

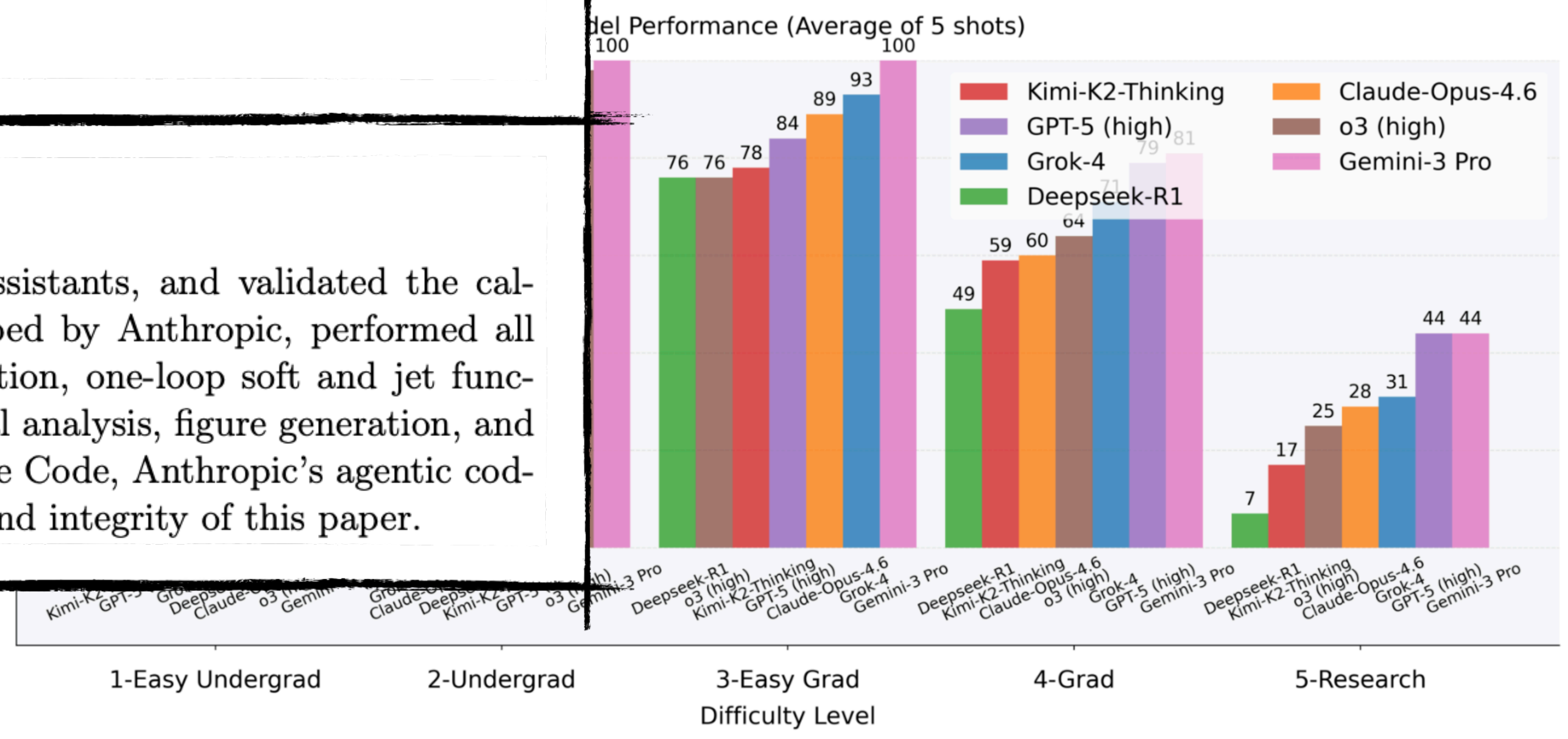
enchmark (TPBench) … and Study
g Capabilities i… ics

…[2], Yu… Münchmeyer[1,5], Maja
ederic… Tadepalli[4]

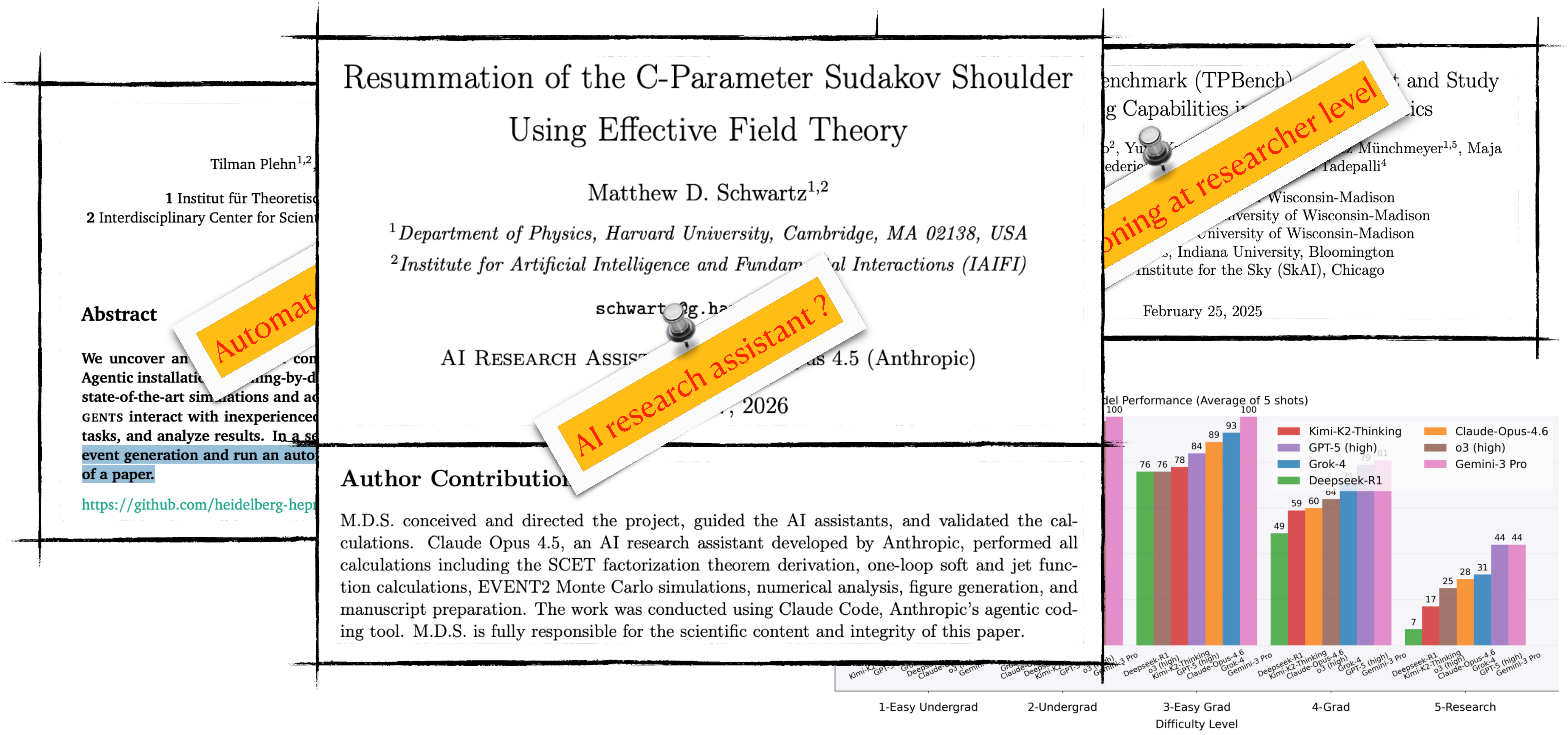
…Wisconsin-Madison
…iversity of Wisconsin-Madison
…University of Wisconsin-Madison
…, Indiana University, Bloomington
Institute for the Sky (SkAI), Chicago

February 25, 2025

…oning at researcher level



What’s next?


Tilman Plehn[1,2],
1 Institut für Theoretisc
2 Interdisciplinary Center for Scient
Abstract
We uncover an
Agentic installati
state-of-the-art sim ations 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-hepr
Automat
Resummation of the C-Parameter Sudakov Shoulder
Using Effective Field Theory
Matthew D. Schwartz[1,2]
[1] Department of Physics, Harvard University, Cambridge, MA 02138, USA
[2] Institute for Artificial Intelligence and Fundam al Interactions (IAIFI)
AI Research Assis s 4.5 (Anthropic)
, 2026
AI research assistant ?
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.
enchmark (TPBench) and Study
g Capabilities i ics
Münchmeyer[1,5], Maja
Tadepalli[4]
Wisconsin-Madison
iversity of Wisconsin-Madison
University of Wisconsin-Madison
, Indiana University, Bloomington
Institute for the Sky (SkAI), Chicago
February 25, 2025
oning at researcher level
del Performance (Average of 5 shots)
Kimi-K2-Thinking
GPT-5 (high)
Grok-4
Deepseek-R1
Claude-Opus-4.6
o3 (high)
Gemini-3 Pro
1-Easy Undergrad
2-Undergrad
3-Easy Grad
4-Grad
5-Research
Difficulty Level

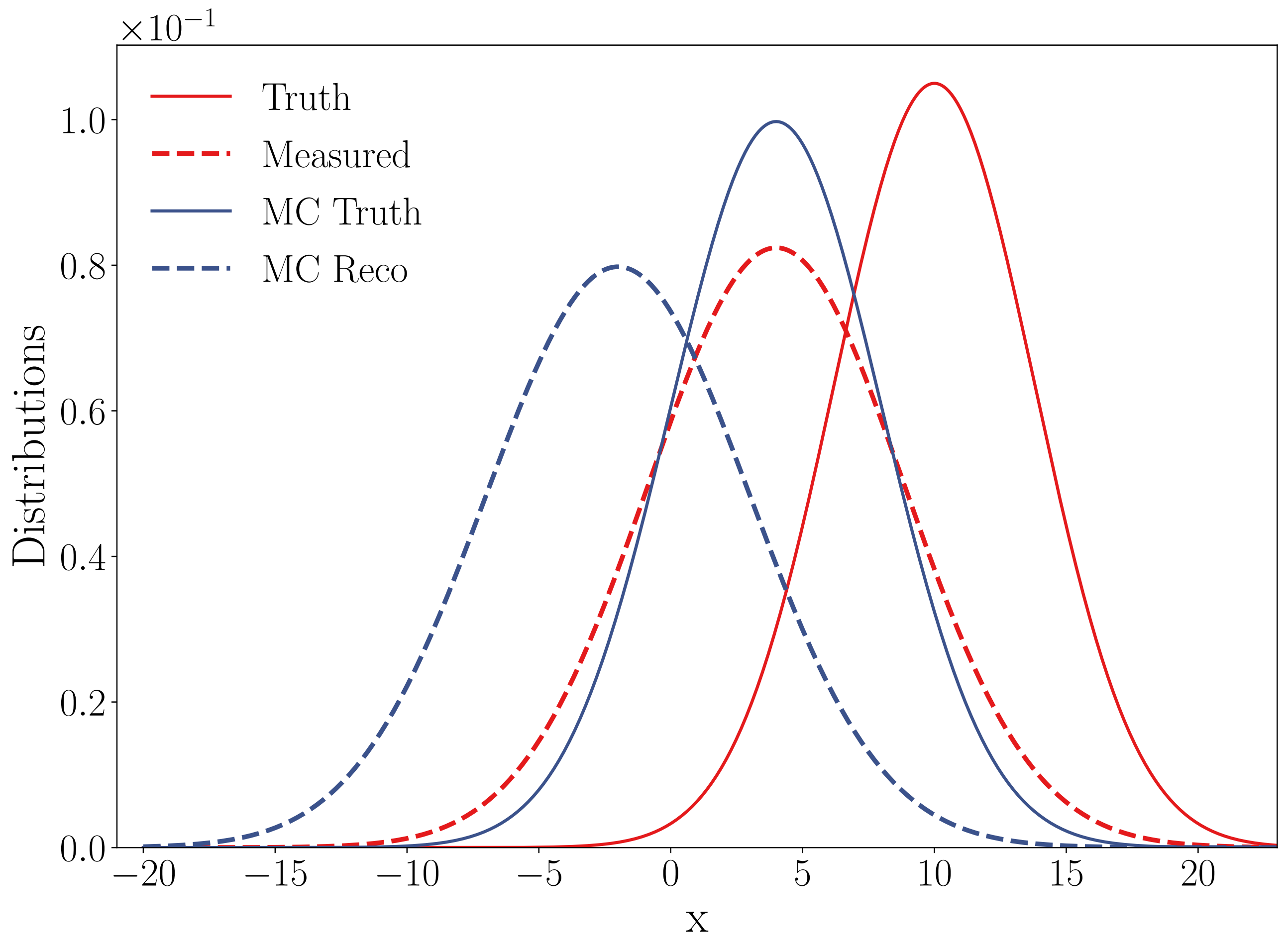
If everything is uncertain,

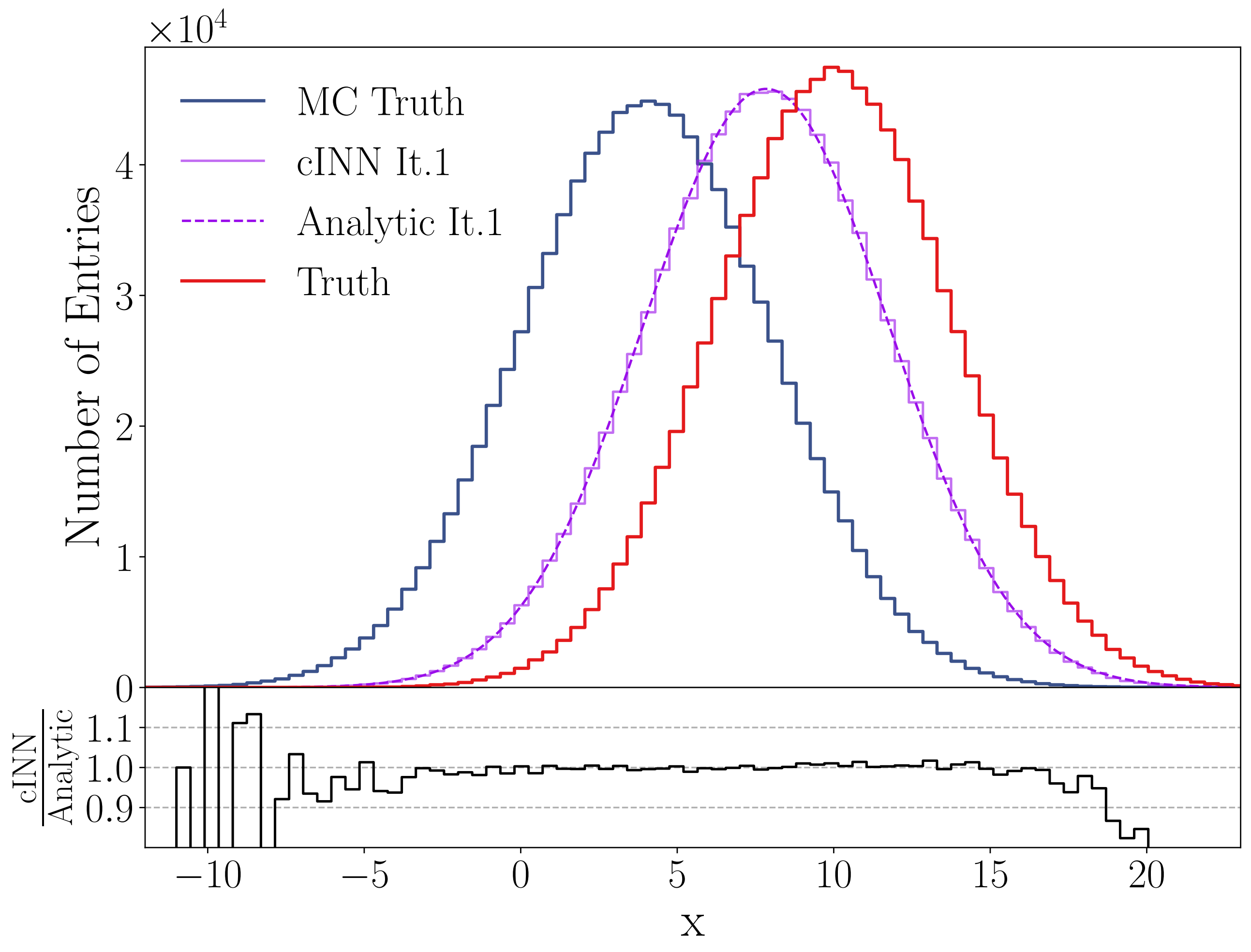
anything is possible

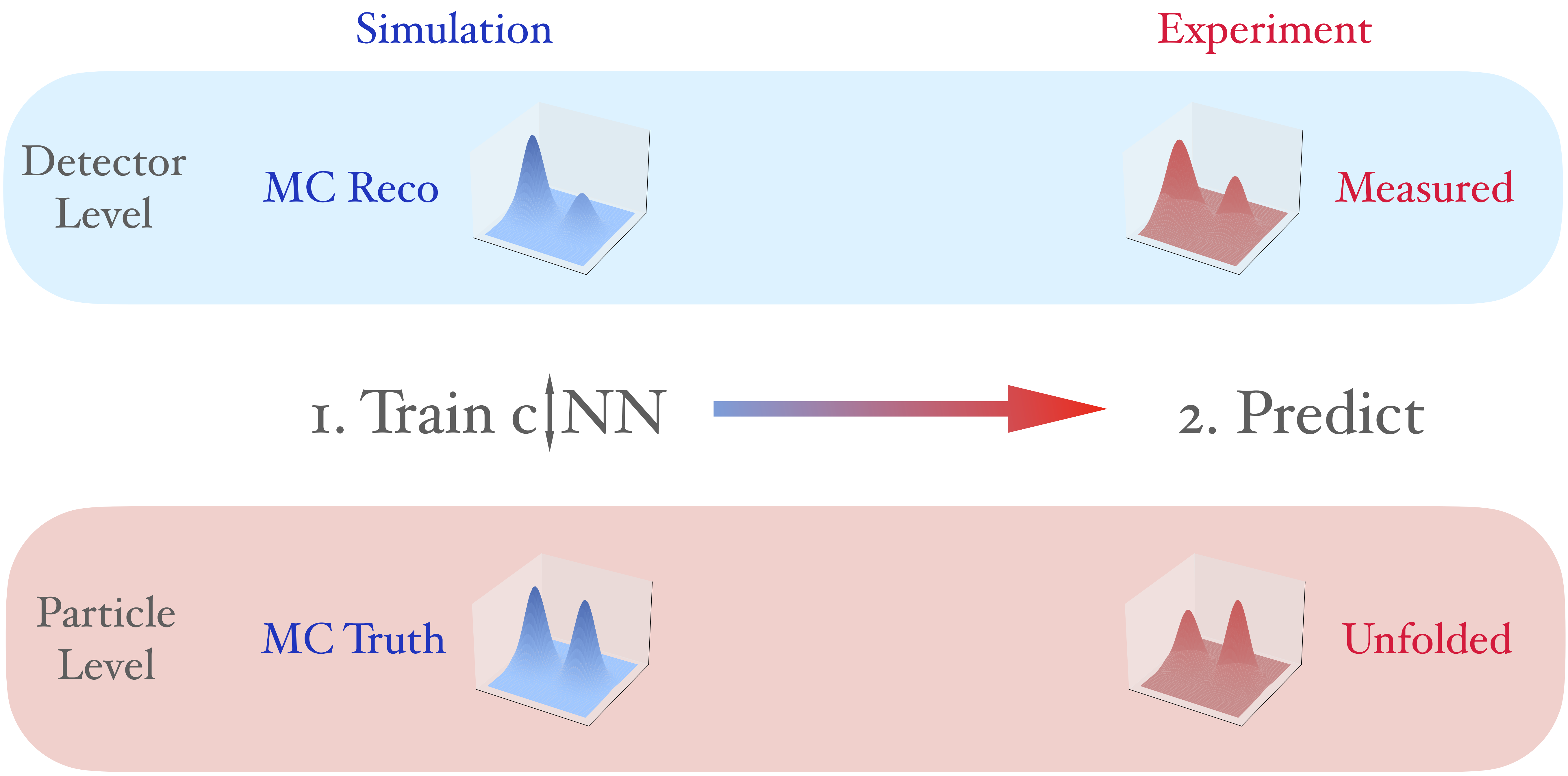
Margaret Drabble

Limitations of direct unfolding with generative NNs

Prior dependence



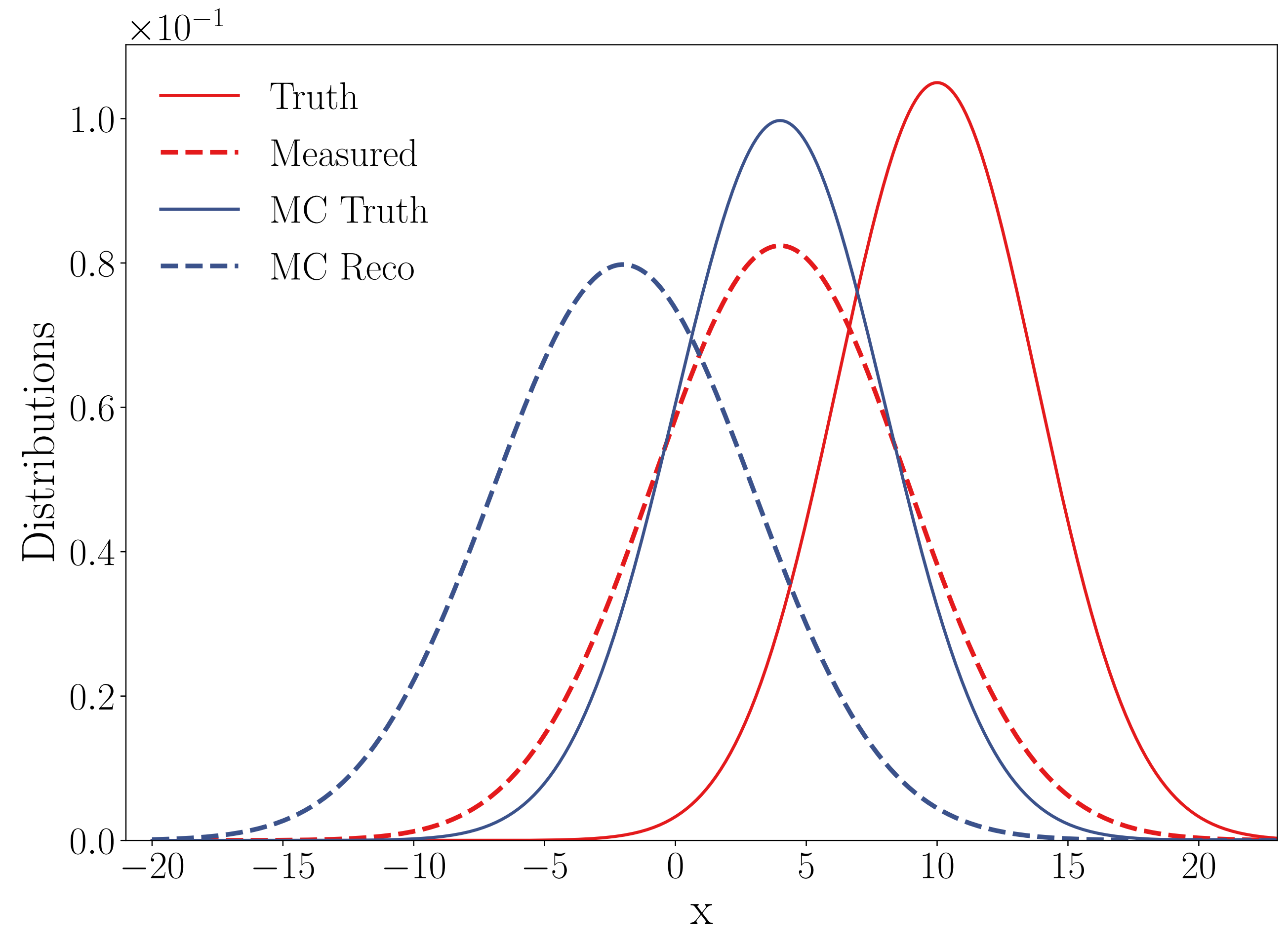


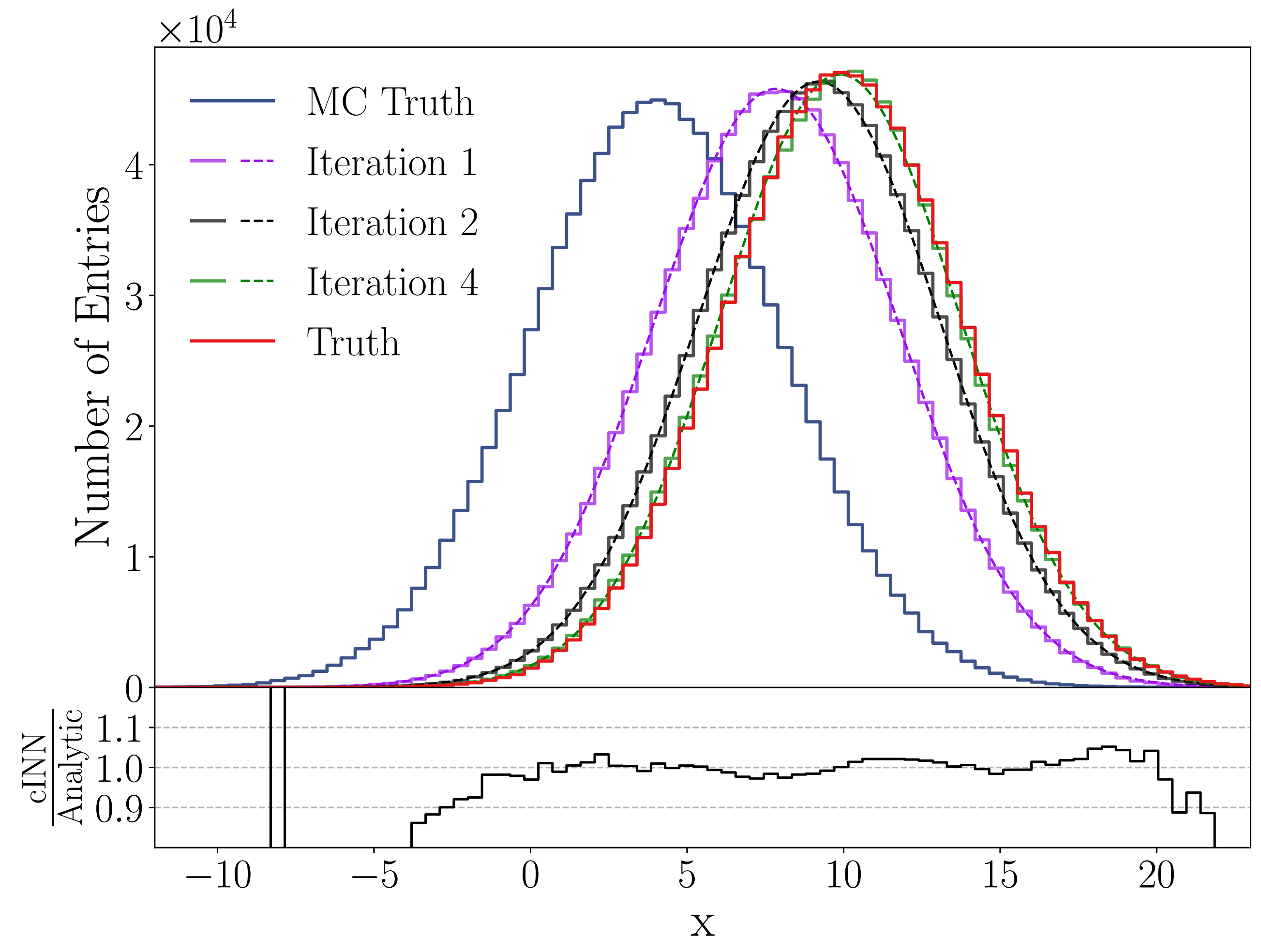


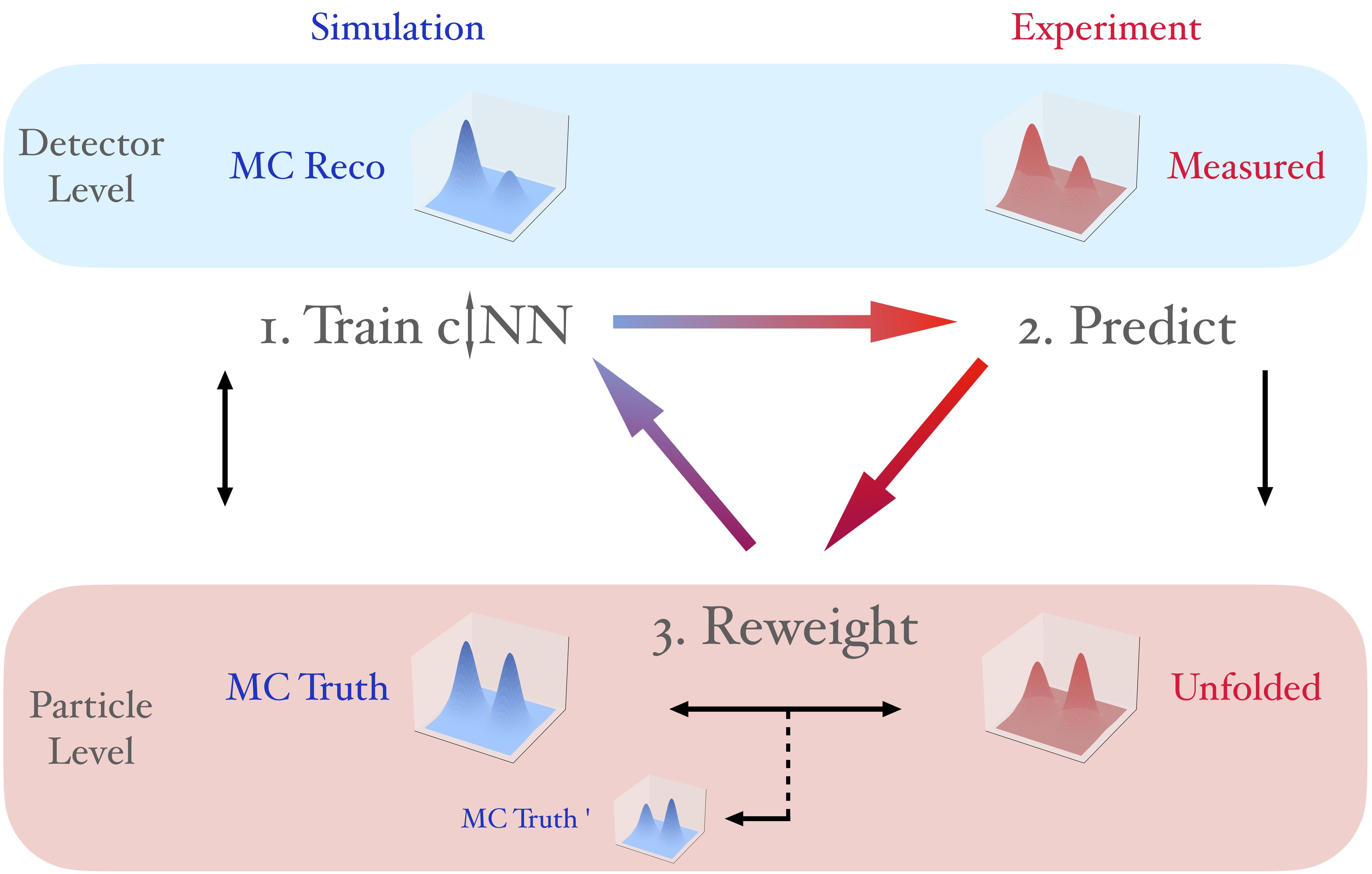
$$p(\text{part}\,|\,\text{det}) = \frac{p(\text{det}\,|\,\text{part}) \cdot p(\text{part})}{p(\text{det})}$$

Conditional iterative unfolding

[2212.08674] M. Backes, et al.







SPINUP

Simulation-Prior Independent Neural Unfolding Procedure

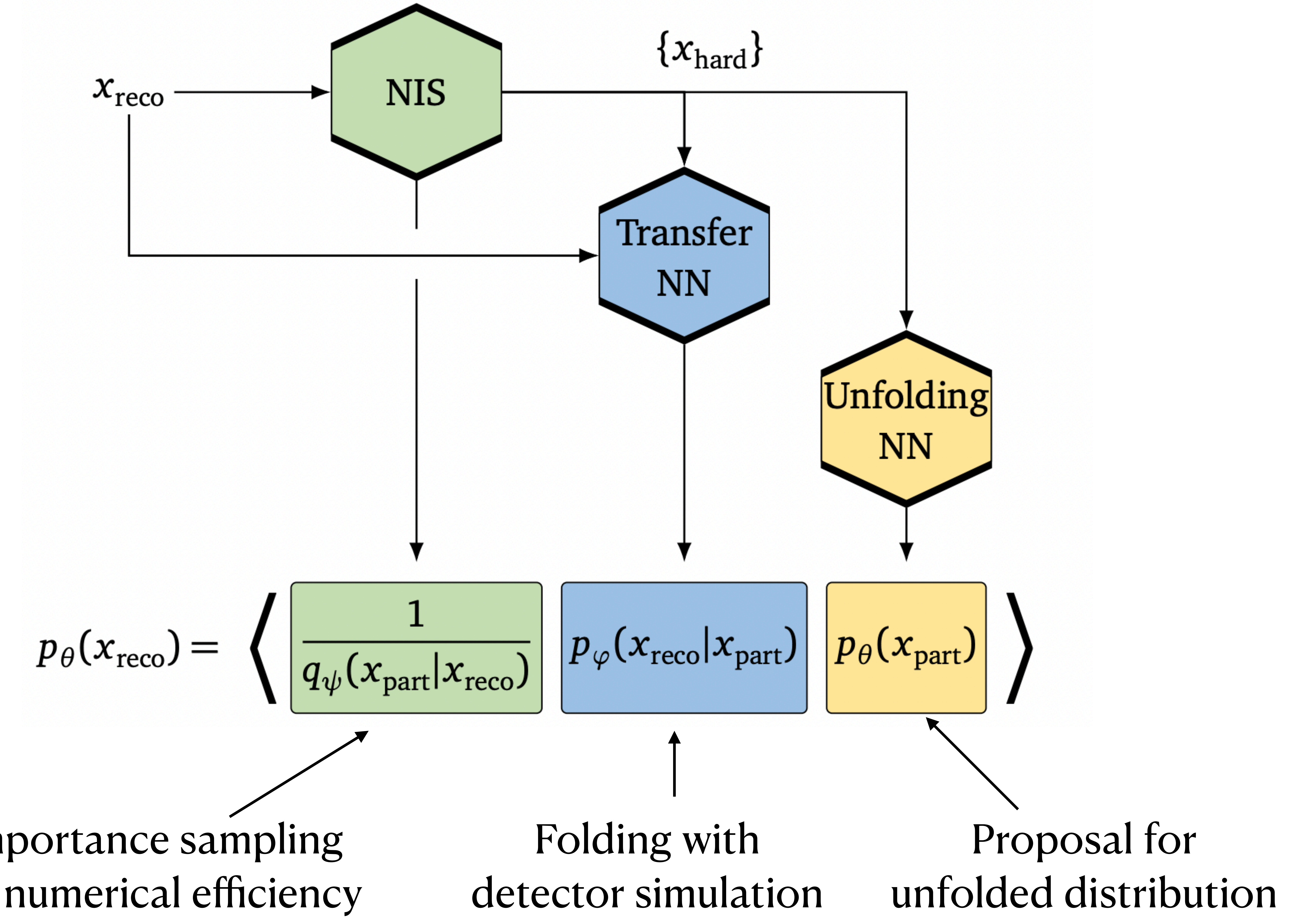
[2507.15084] AB, T Heimel , N. Huetschı , M. Kagan, T. Plehn

Ansatz

$$\underbrace{p_{data}(x_{reco})}_{\boxed{\text{DATA}}} \overset{!}{=} \underbrace{p_\theta(x_{reco})}_{\boxed{\text{“folded” unfolded distribution}}}$$

$$= \int dx_{part}\; p(x_{reco} \,|\, x_{part})\; p_\theta(x_{part})$$

→ Minimize KL divergence



SPINUP

Simulation-Prior Independent Neural Unfolding Procedure

Ensembling for lower bound on uncertainty from information loss

Drastic example:

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