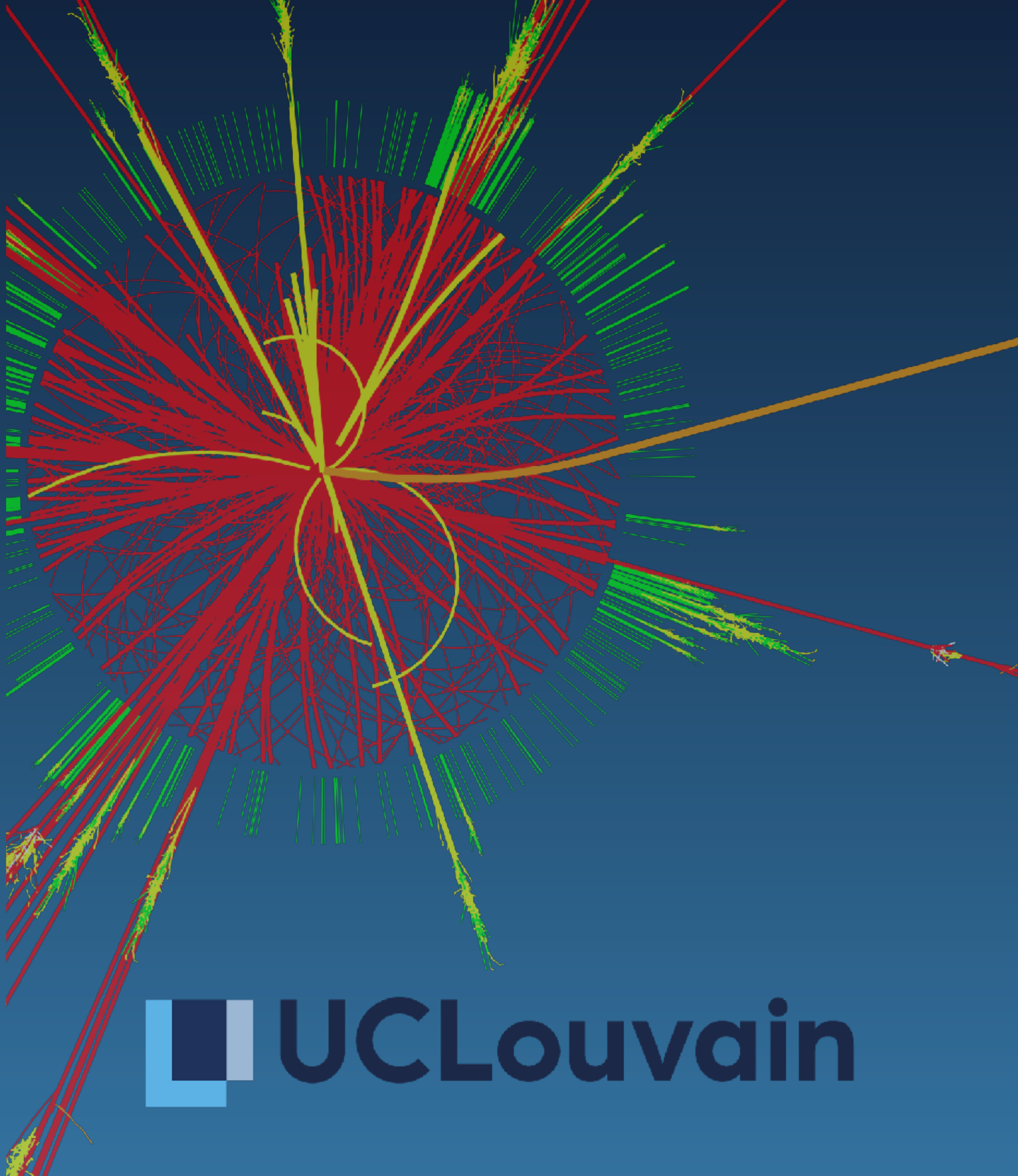


Modern Machine Learning for LHC Physics

Sapientia ex machina?

LHCP2023 — Belgrade

Ramon Winterhalder — UC Louvain

Why should LHCP care about ML?

Why should LHCP care about ML?

> “LHCP should care about machine learning because it can improve data analysis, simulation and modeling, lead to new discoveries, and foster cross-disciplinary collaboration

ChatGPT

Why ML in HEP?

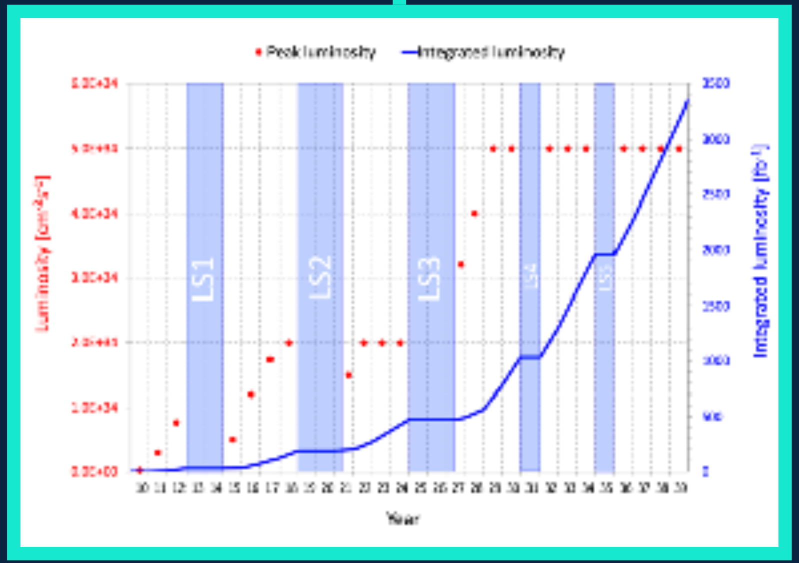
Data volume

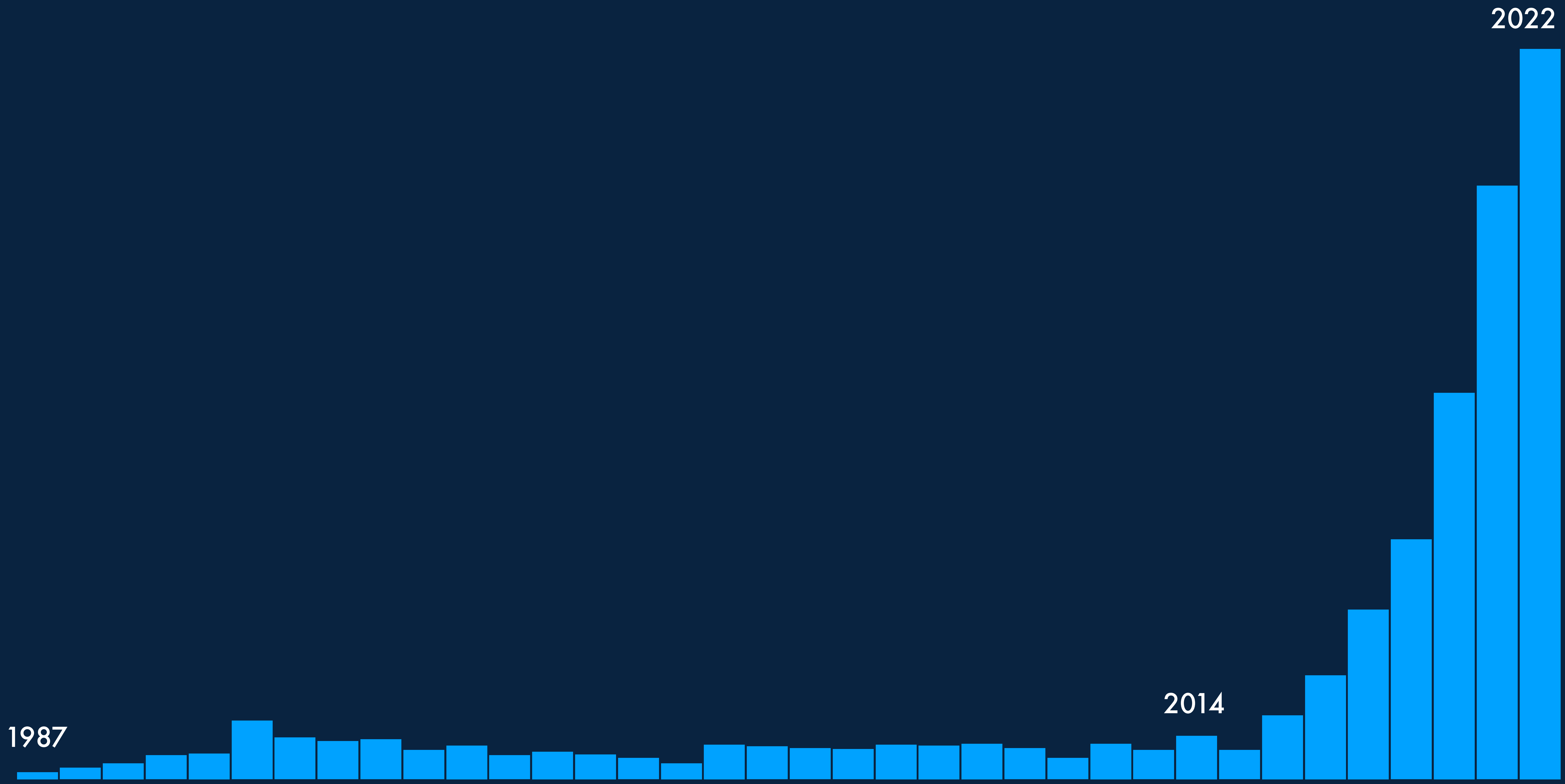
Large amounts of data

1. labeled (Simulation)
2. unlabeled (Detector)

ML wants lots of data

1





Why ML in HEP?

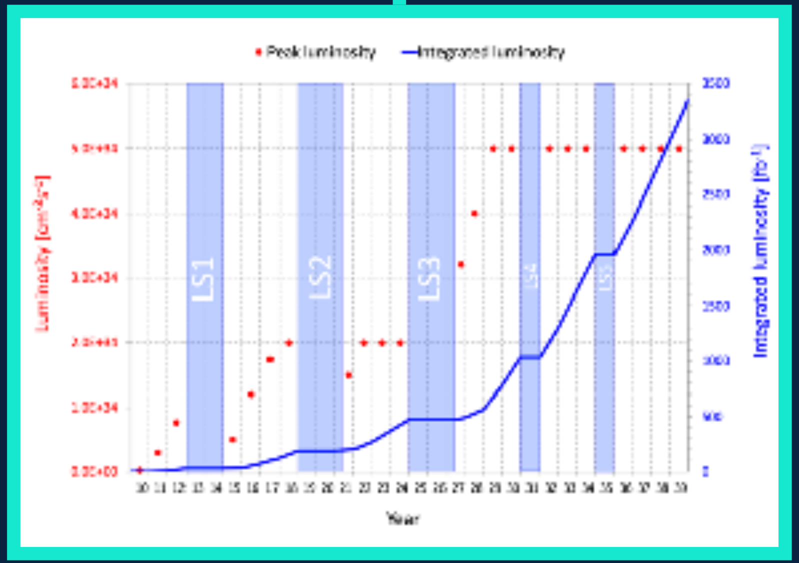
Data volume

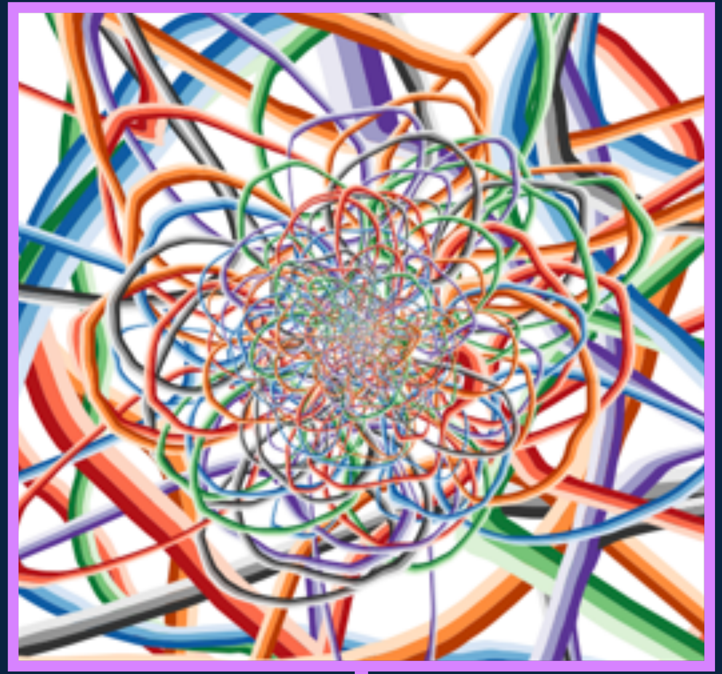
Large amounts of data

1. labeled (Simulation)
2. unlabeled (Detector)

ML wants lots of data

1



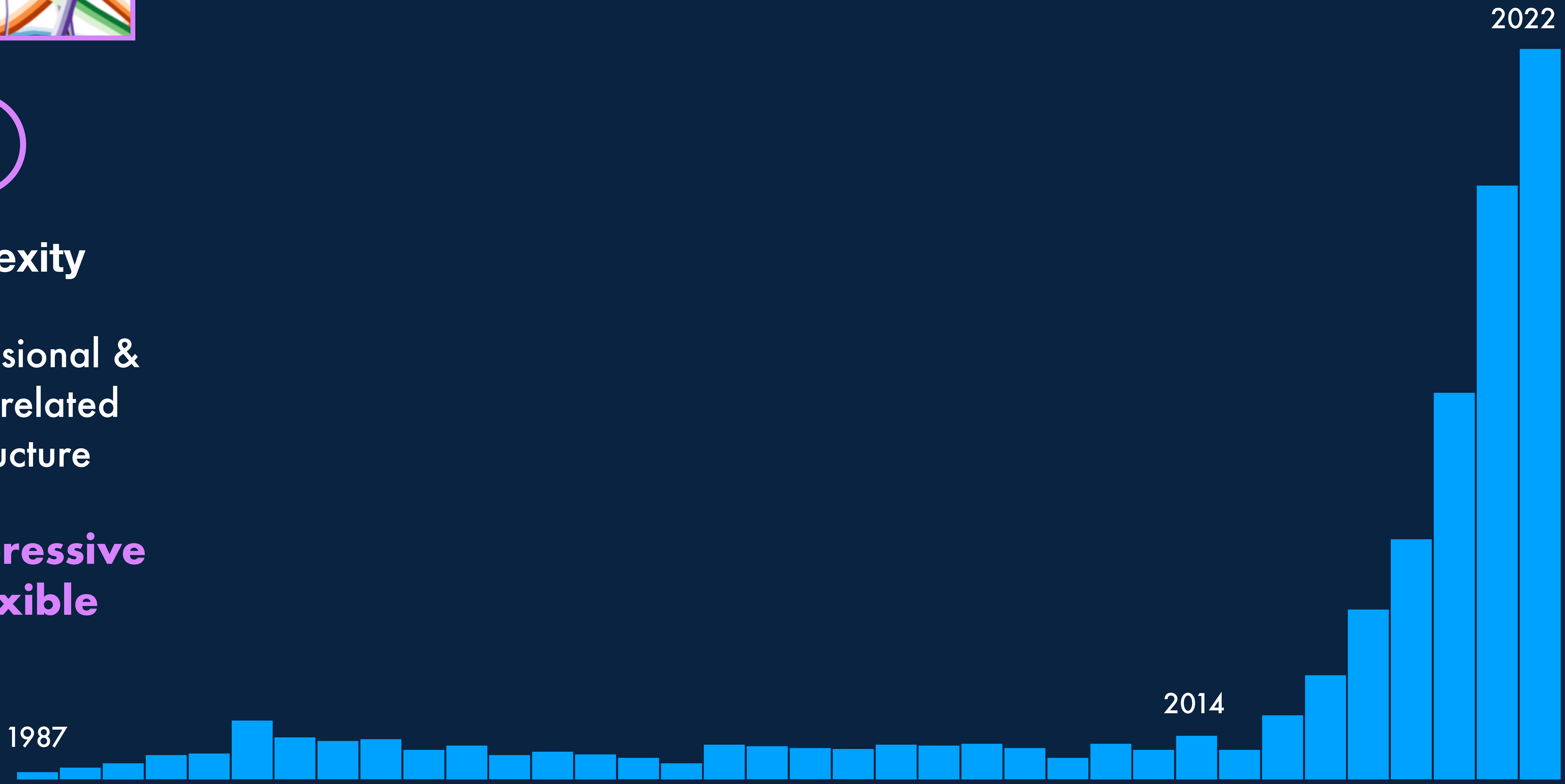


2

Complexity

High-dimensional & highly correlated data structure

ML is expressive and flexible



Why ML in HEP?

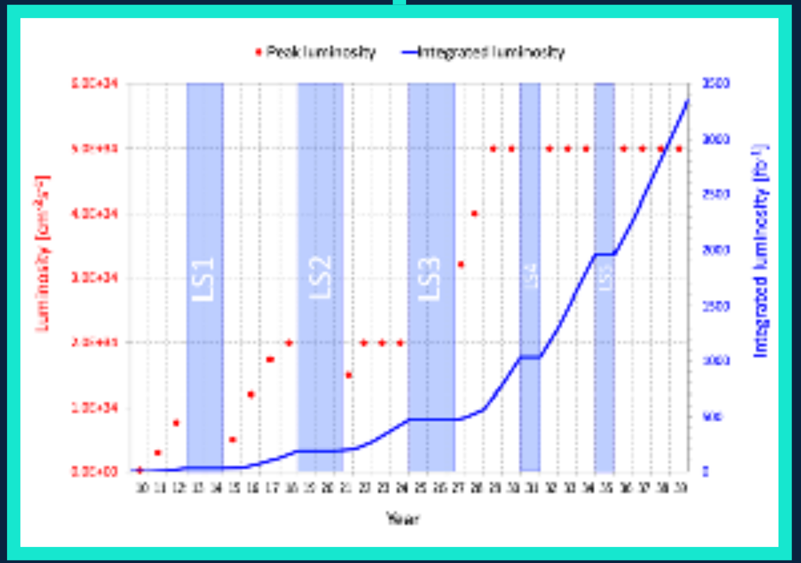
Data volume

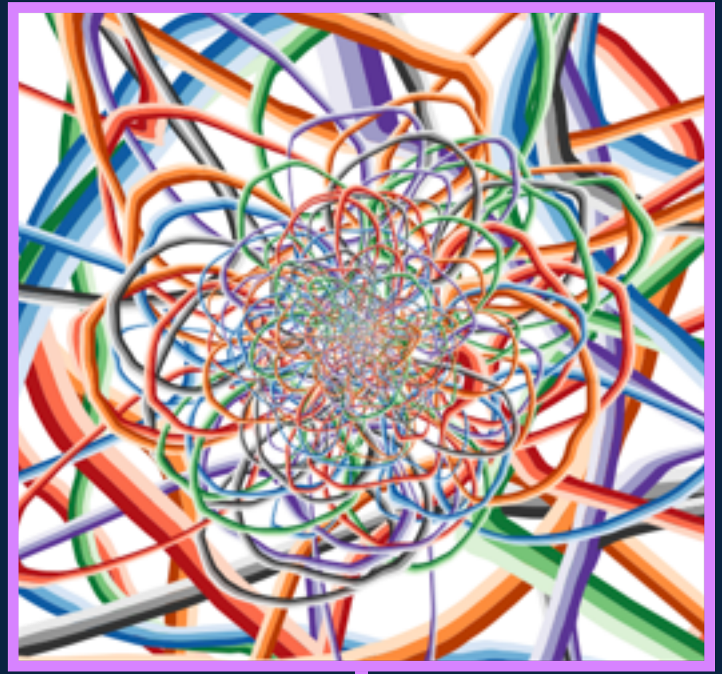
Large amounts of data

1. labeled (Simulation)
2. unlabeled (Detector)

ML wants lots of data

1





2

Complexity

High-dimensional & highly correlated data structure

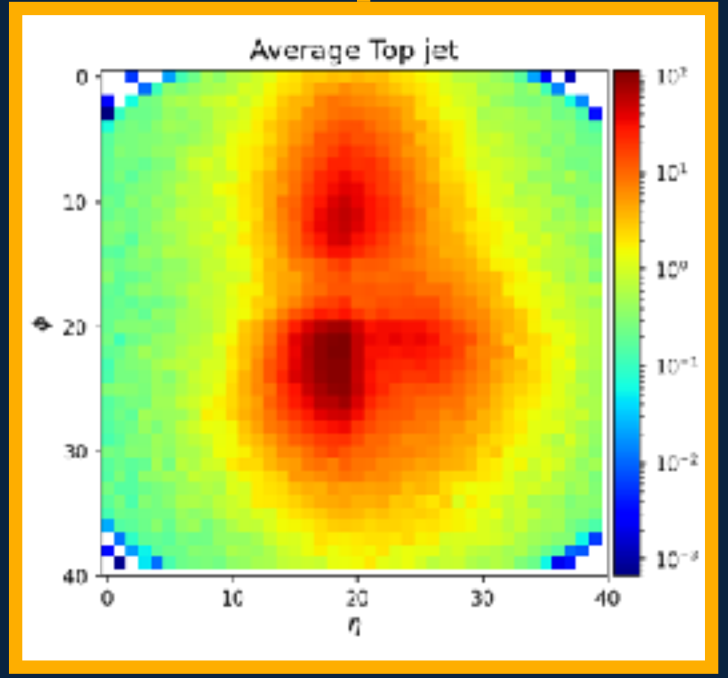
ML is expressive and flexible

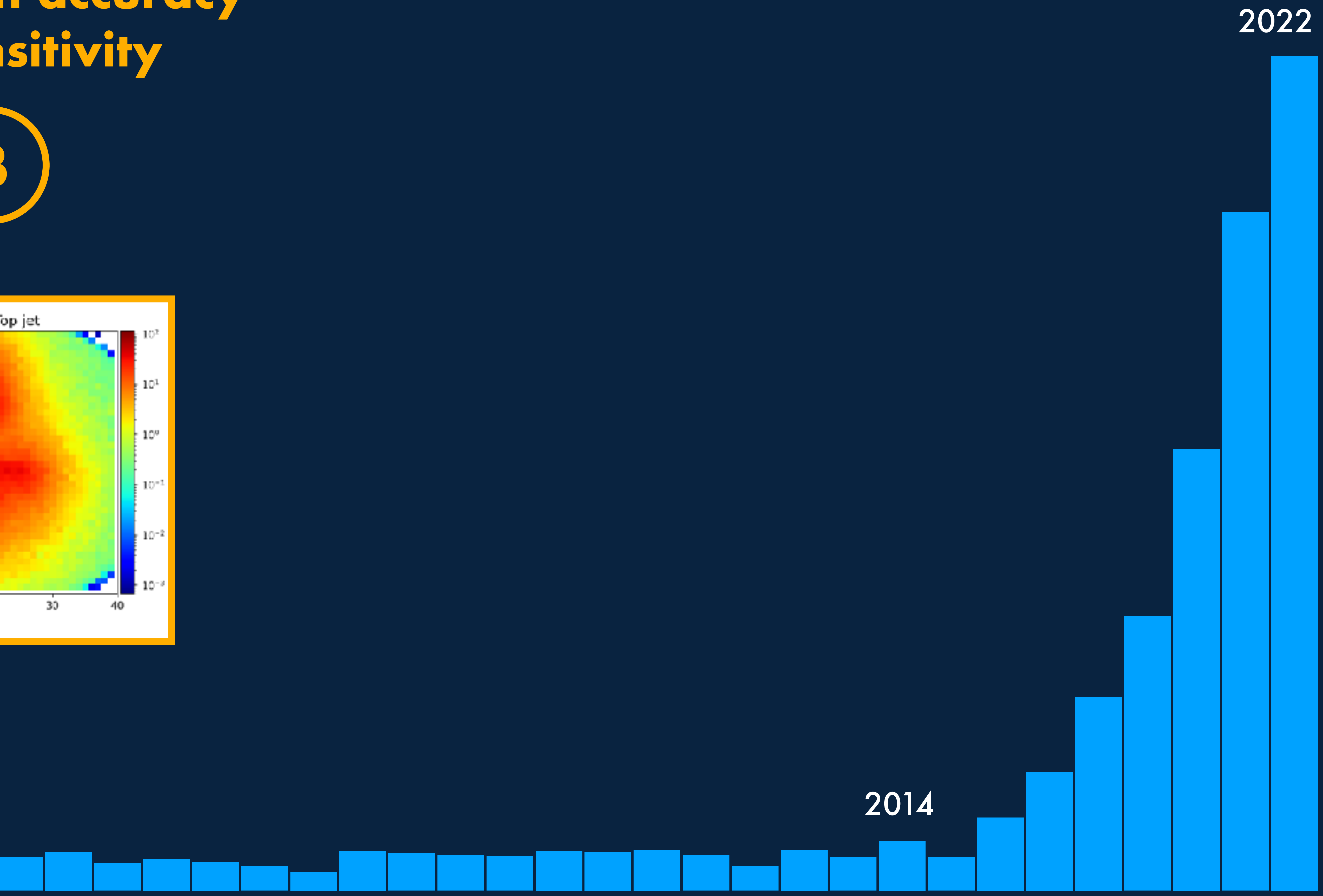
Signal detection

Rare and elusive signals among large backgrounds

ML has high accuracy and sensitivity

3





Why ML in HEP?

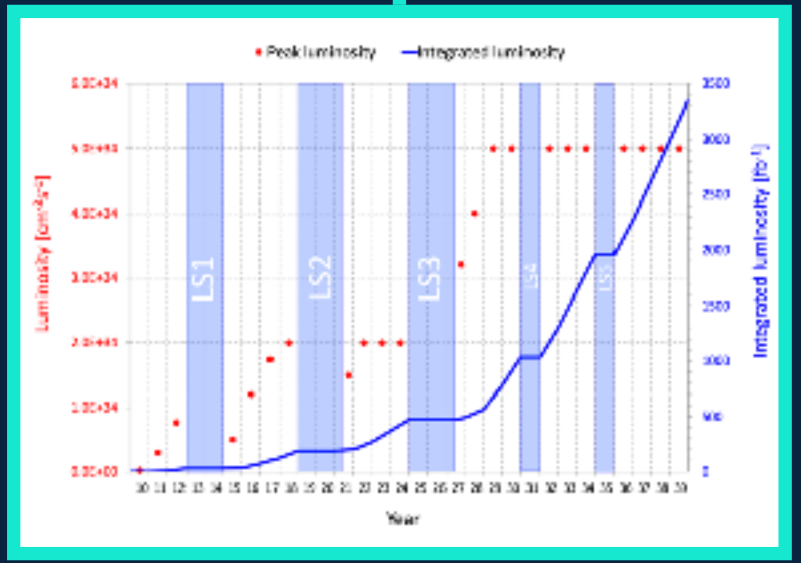
Data volume

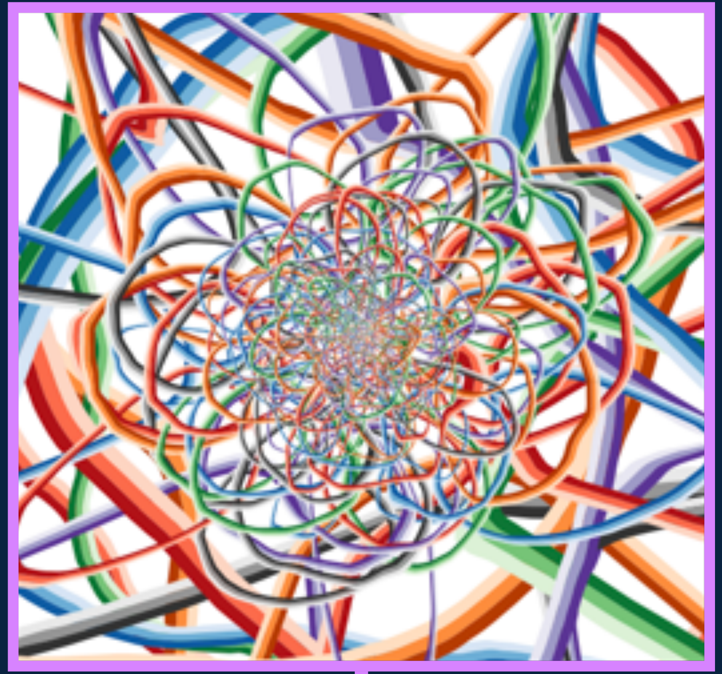
Large amounts of data

1. labeled (Simulation)
2. unlabeled (Detector)

ML wants lots of data

1





2

Complexity

High-dimensional & highly correlated data structure

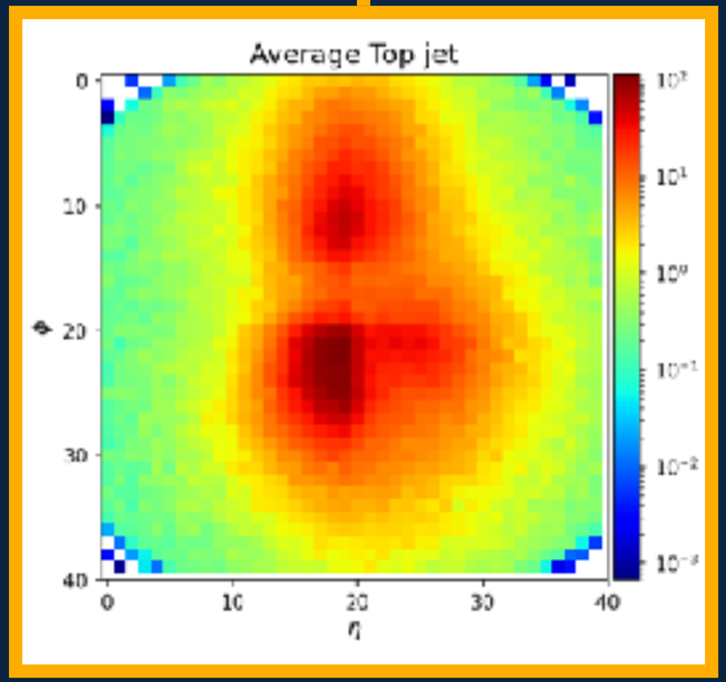
ML is expressive and flexible

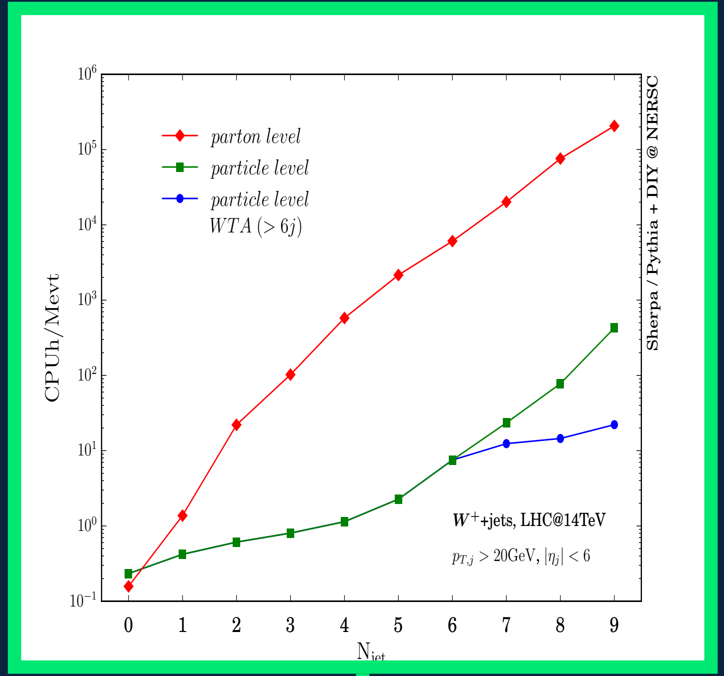
Signal detection

Rare and elusive signals among large backgrounds

ML has high accuracy and sensitivity

3



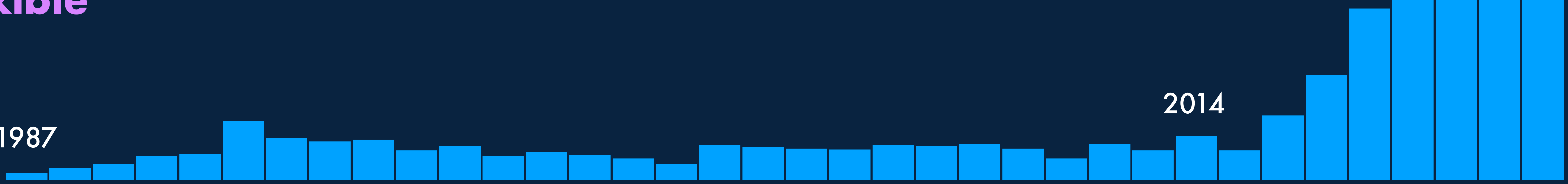


4

Computing Budget

Simulation & analysis is computationally expensive

ML is fast



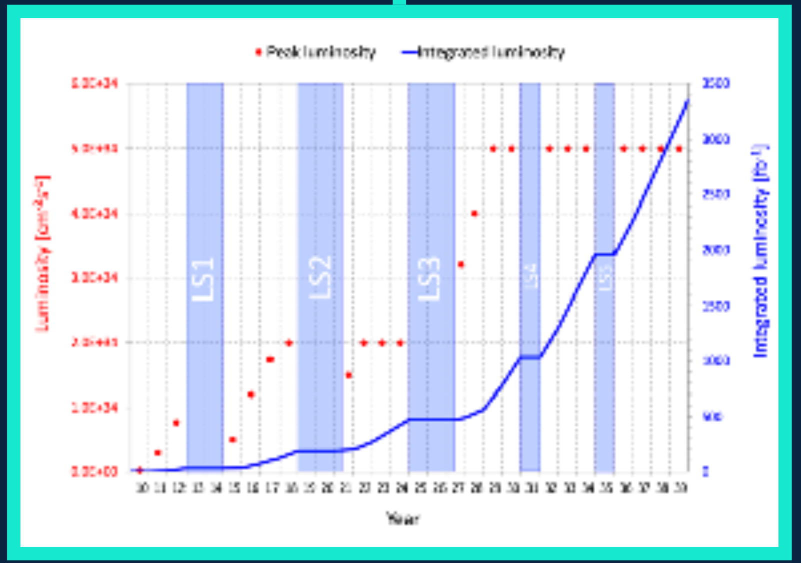
Why ML in HEP?

1 Data volume

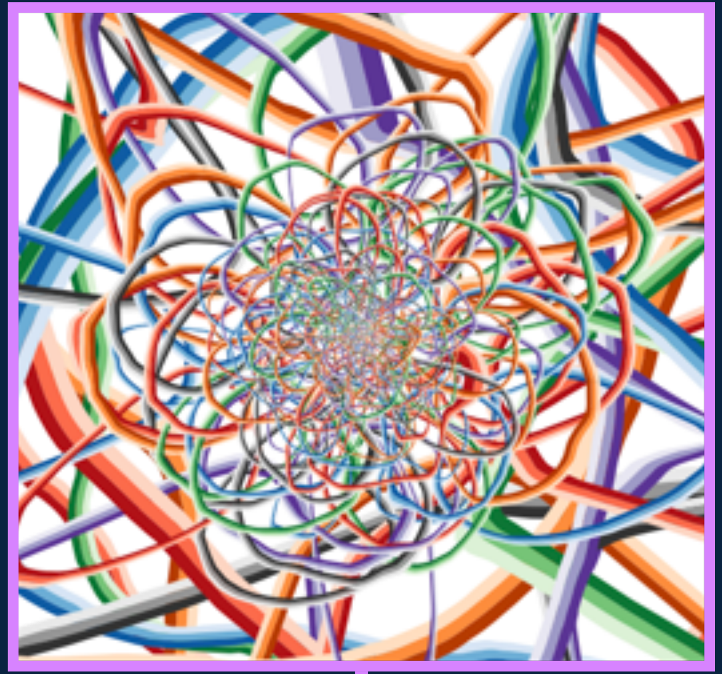
Large amounts of data

1. labeled (Simulation)
2. unlabeled (Detector)

ML wants lots of data



2 Complexity



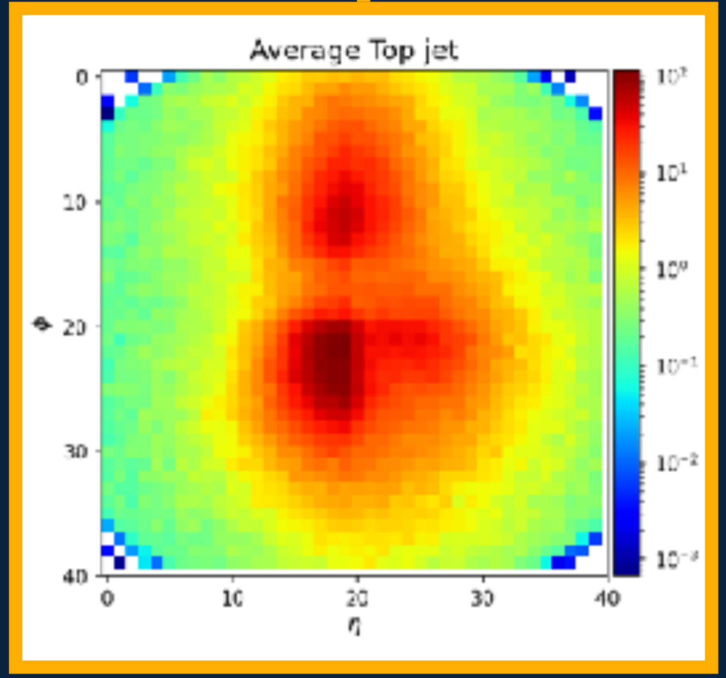
High-dimensional & highly correlated data structure

ML is expressive and flexible

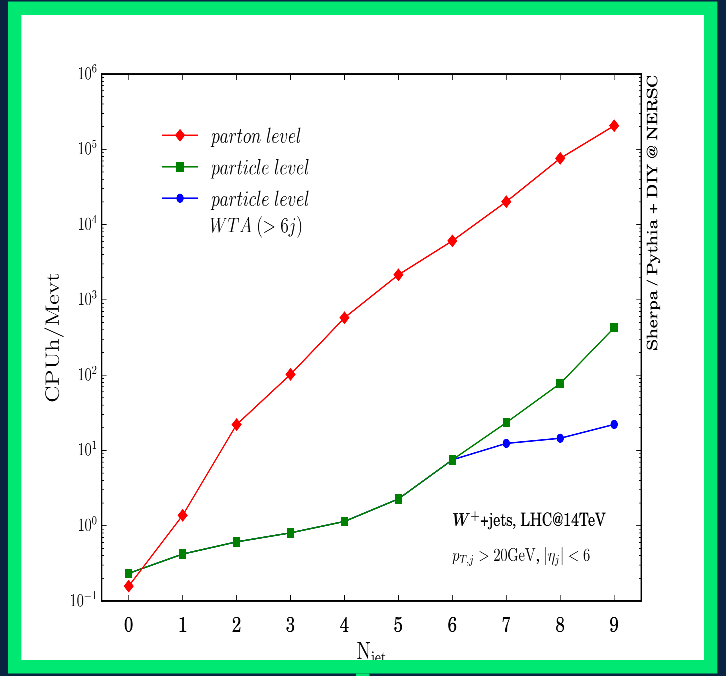
3 Signal detection

Rare and elusive signals among large backgrounds

ML has high accuracy and sensitivity



4 Computing Budget



Simulation & analysis is computationally expensive

ML is fast

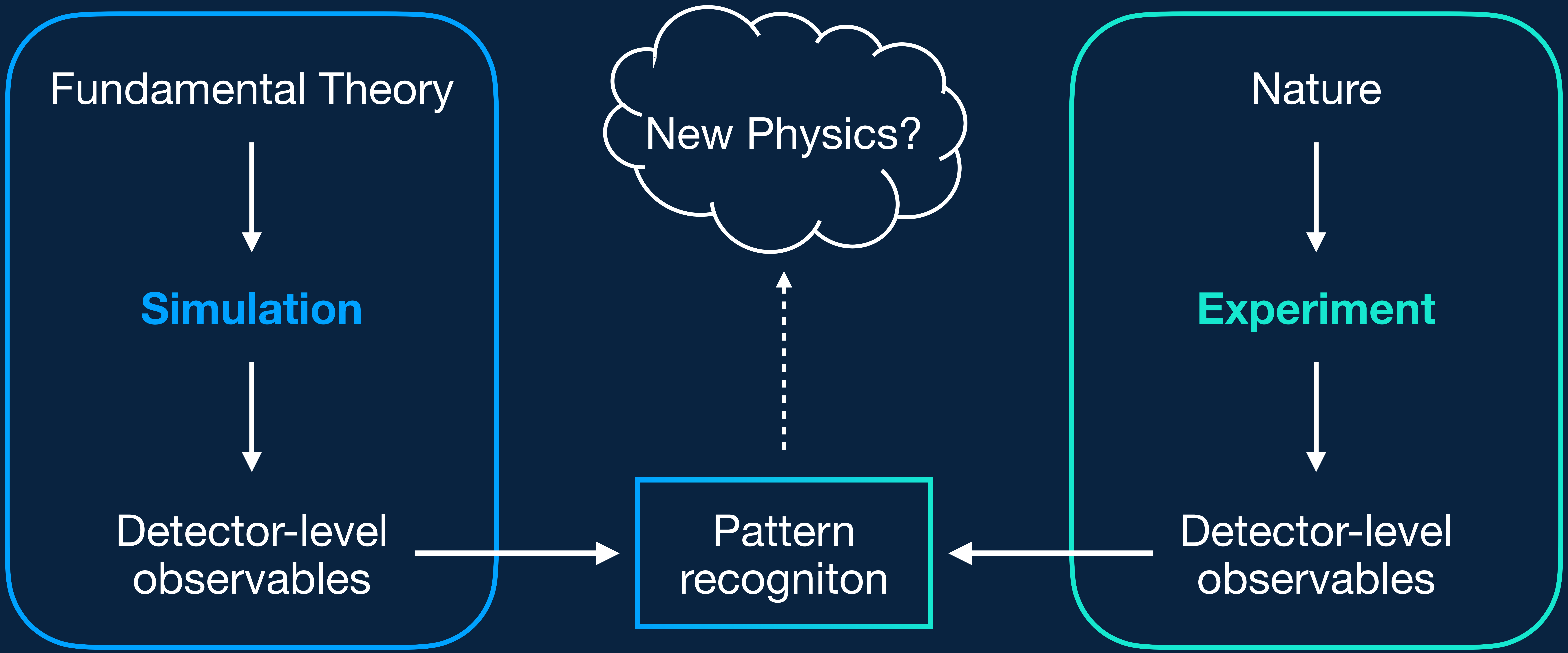
5 Increasing interest

> 150 paper/year

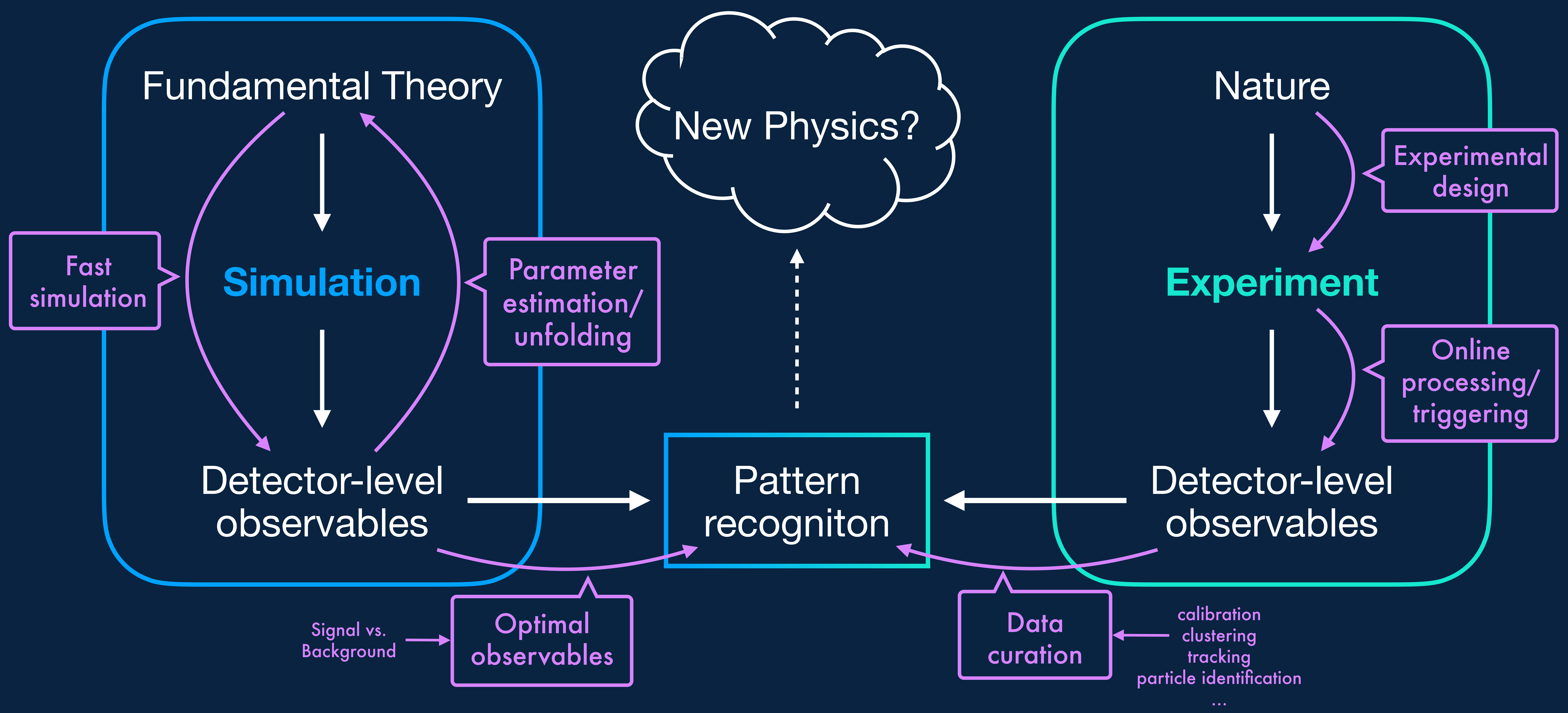
Future of HEP?

ML is fun!

LHC analysis (oversimplified)



LHC analysis + ML

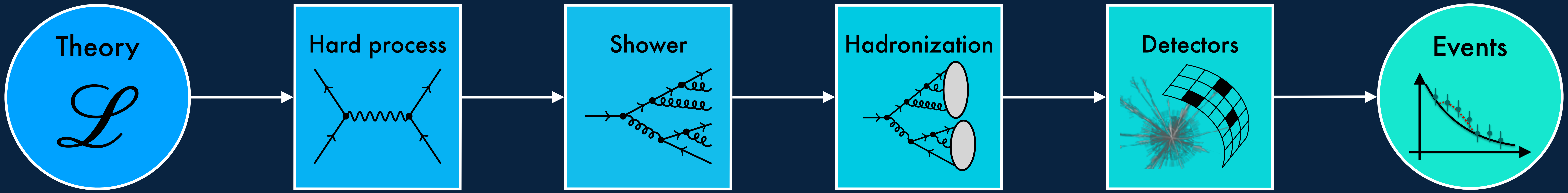



Machine Learning
Supervised Learning
Classification
Point Clouds [2102.05073]
PELICAN [2211.00454]
Energy Flow Networks [1810.05165]
ParticleNet [1902.08570]
Bayesian Tagger [1904.10004]
Regression
Bayesian Regression [2003.11099]
NNPDF [2109.02653]
Matrix Elements [2206.14831]
Symbolic regression [2109.10414]
Reinforcement Learning
Flavor structure [2304.14176]
Simplifying Polylogs [2206.04115]
String vacua & landscape [1903.11616, 2111.11466]
Model Building [2103.04759]
Analytic continuation [2112.13011]
Unsupervised Learning
Clustering
Jet Clustering [2008.06064]
3D Pixel Clustering [2007.03083]
Anomaly Detection
AnomalyCLR [2301.04660]
LaCATHODE [2210.14924]
Normalized AE [2206.14225]
Generative Models
NF
Precision Generation [2110.13632]
MEM [2210.00019]
MADNIS [2212.06172]
CaloFlow I-IV [2106.05285,…]
ELSA [2305.07696]
VAE
OTUS [2101.08944]
Jet Simulation [2203.00520]
Diffusion Models
PC-JeDi [2303.05376]
FPCD [2304.01266]
DDPM & CFM [2305.10475]
GAN
EPiC-GAN [2301.08128]
How to GAN [1907.03764]
CaloGAN [1712.10321]
Transformer
JetGPT [2305.10475]

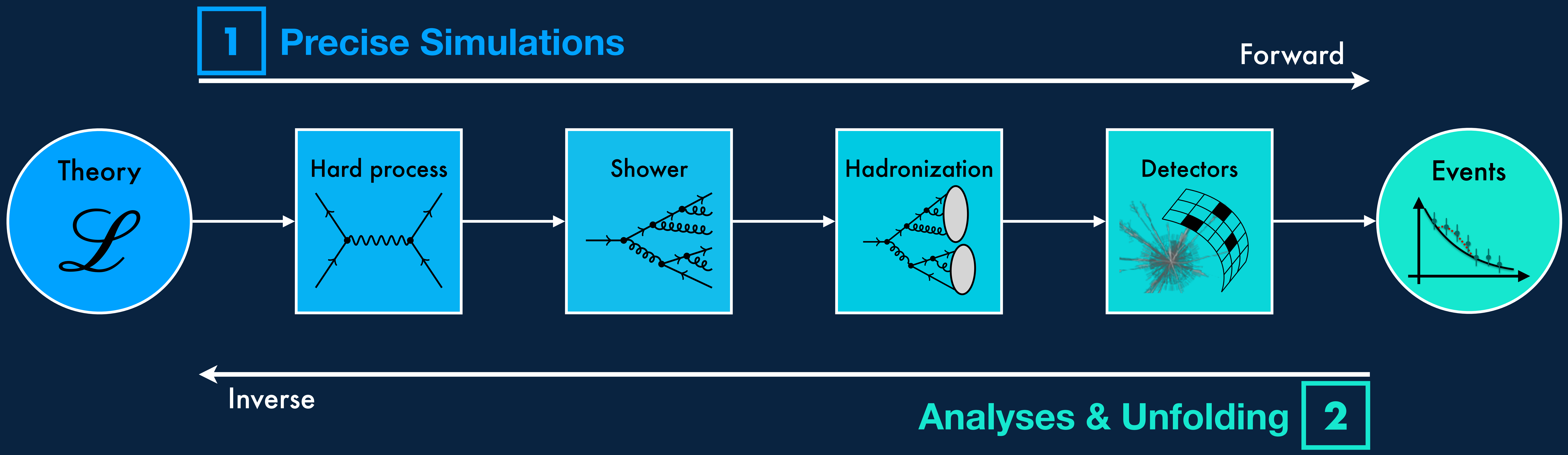
Generative Models
VAE
OTUS [2101.08944]
Jet Simulation [2203.00520]
NF
Precision Generation [2110.13632]
MEM [2210.00019]
MADNIS [2212.06172]
CaloFlow I-IV [2106.05285,…]
ELSA [2305.07696
Diffusion Models
PC-JeDi [2303.05376]
FPCD [2304.01266]
DDPM & CFM [2305.10475]
GAN
EPiC-GAN [2301.08128]
How to GAN [1907.03764]
CaloGAN [1712.10321]
Transformer
JetGPT [2305.10475]
Unsupervised Learning
Clustering
Jet Clustering [2008.06064]
3D Pixel Clustering [2007.03083]
Anomaly Detection
AnomalyCLR [2301.04660]
LaCATHODE [2210.14924]
Normalized AE [2206.14225]
Reinforcement Learning
Flavor structure [2304.14176]
Simplifying Polylogs [2206.04115]
String vacua & landscape [1903.11616, 2111.11466]
Model Building [2103.04759]
Analytic continuation [2112.13011]

Precision Generation
[2110.13632]
MEM
[2210.00019]
MADNIS
[2212.06172]
NF
CaloFlow I-IV
[2106.05285,...]
OTUS
[2101.08944]
ELSA
[2305.07696]
Jet Simulation
[2203.00520]
VAE
PC-JeDi
[2303.05376]
FPCD
[2304.01266]
Generative Models
Diffusion Models
EPiC-GAN
[2301.08128]
GAN
How to GAN
[1907.03764]
DDPM & CFM
[2305.10475]
CaloGAN
[1712.10321]
Transformer
JetGPT
[2305.10475]
?
Theory
$\mathscr{L}$
Hard process
Shower
Hadronization
Detectors
Events

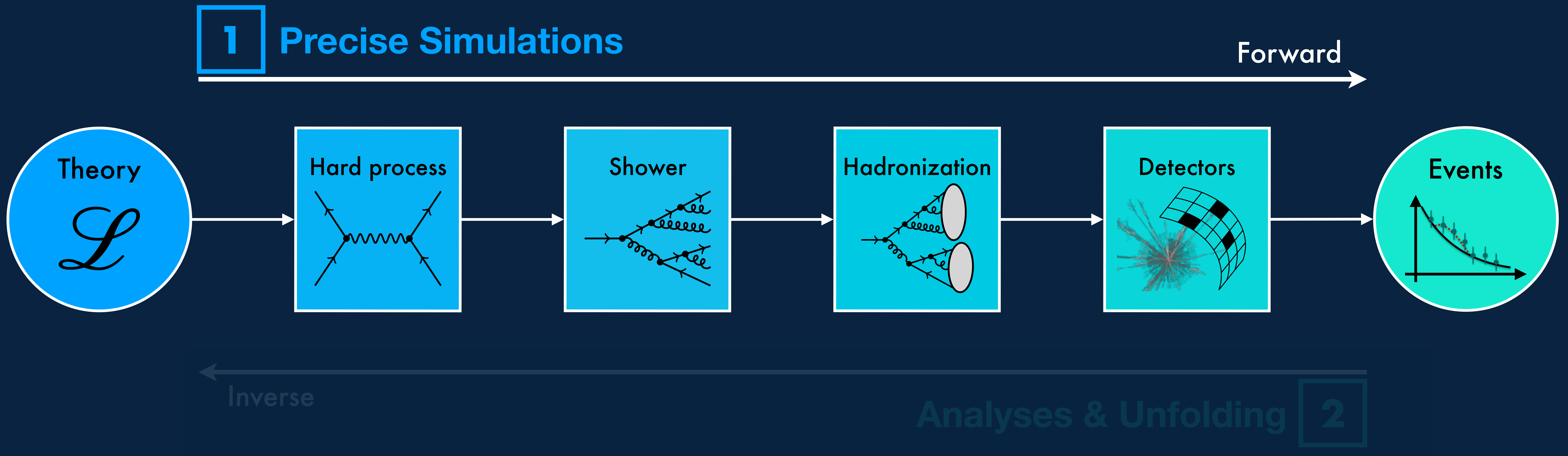
ML aided simulation chain



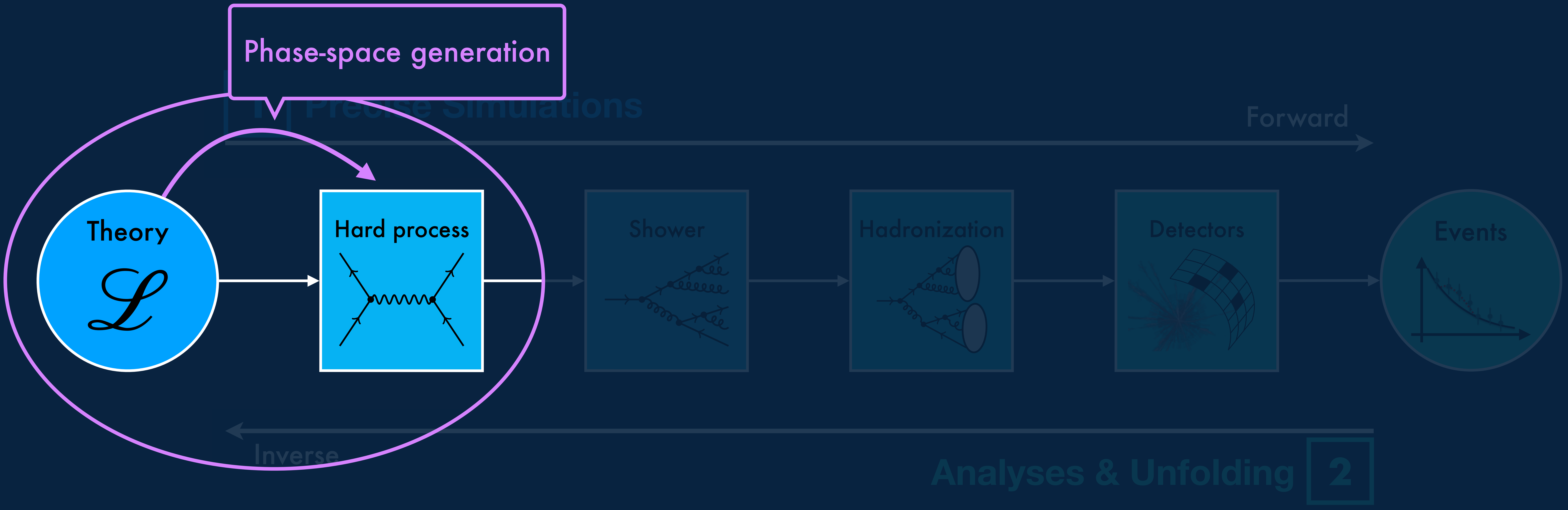
ML aided simulation chain



ML improved simulations

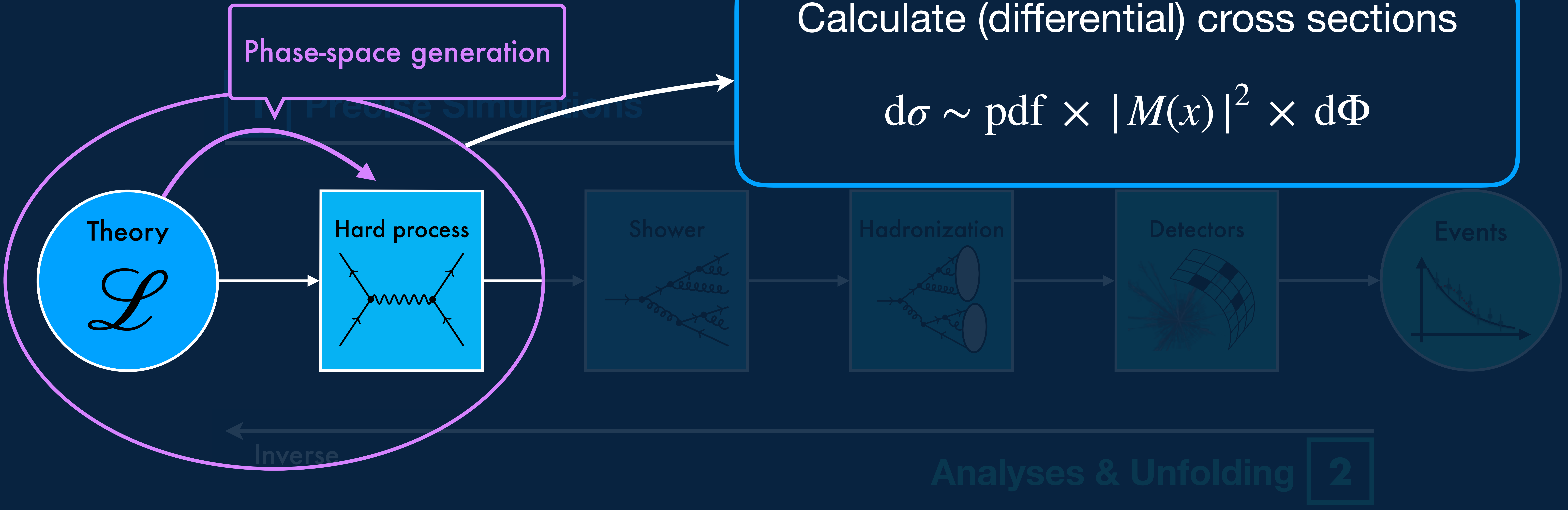


ML improved simulations



BDT [1707.00028, …], NN [1810.11509, 2009.07819, …]
NF [2001.05486, 2001.05478, 2001.10028, 2005.12719, 2112.09145, 2212.06172, …]

ML improved simulations

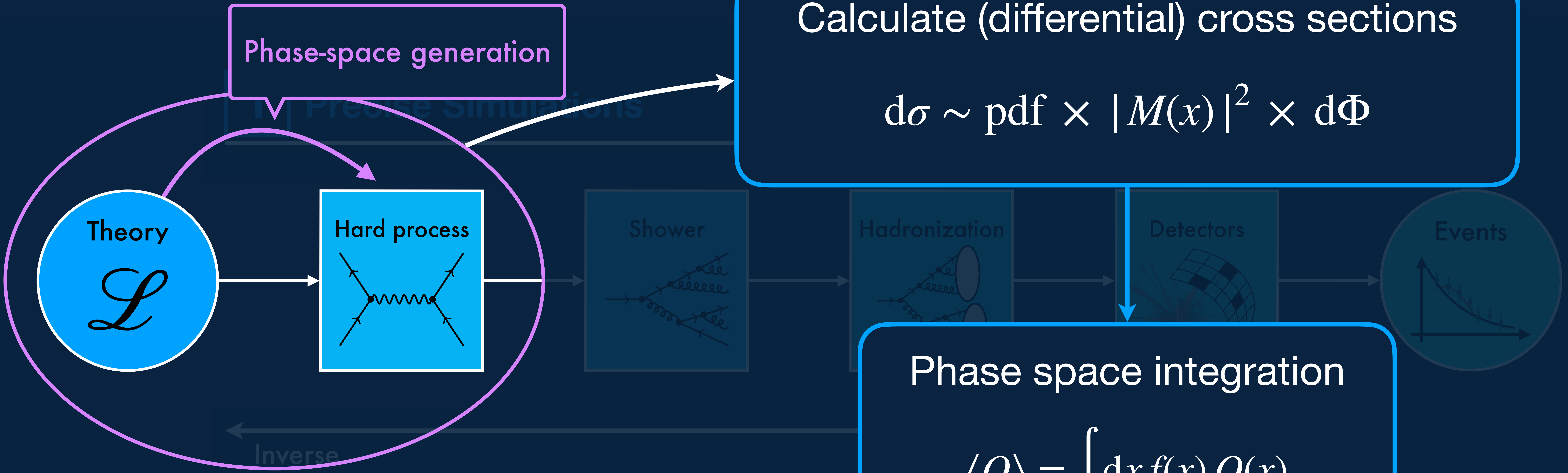


Calculate (differential) cross sections

$$d\sigma \sim \text{pdf} \times |M(x)|^2 \times d\Phi$$

BDT [1707.00028, …], NN [1810.11509, 2009.07819, …]
NF [2001.05486, 2001.05478, 2001.10028, 2005.12719, 2112.09145, 2212.06172, …]

ML improved simulations



Calculate (differential) cross sections

$$d\sigma \sim \text{pdf} \times |M(x)|^2 \times d\Phi$$

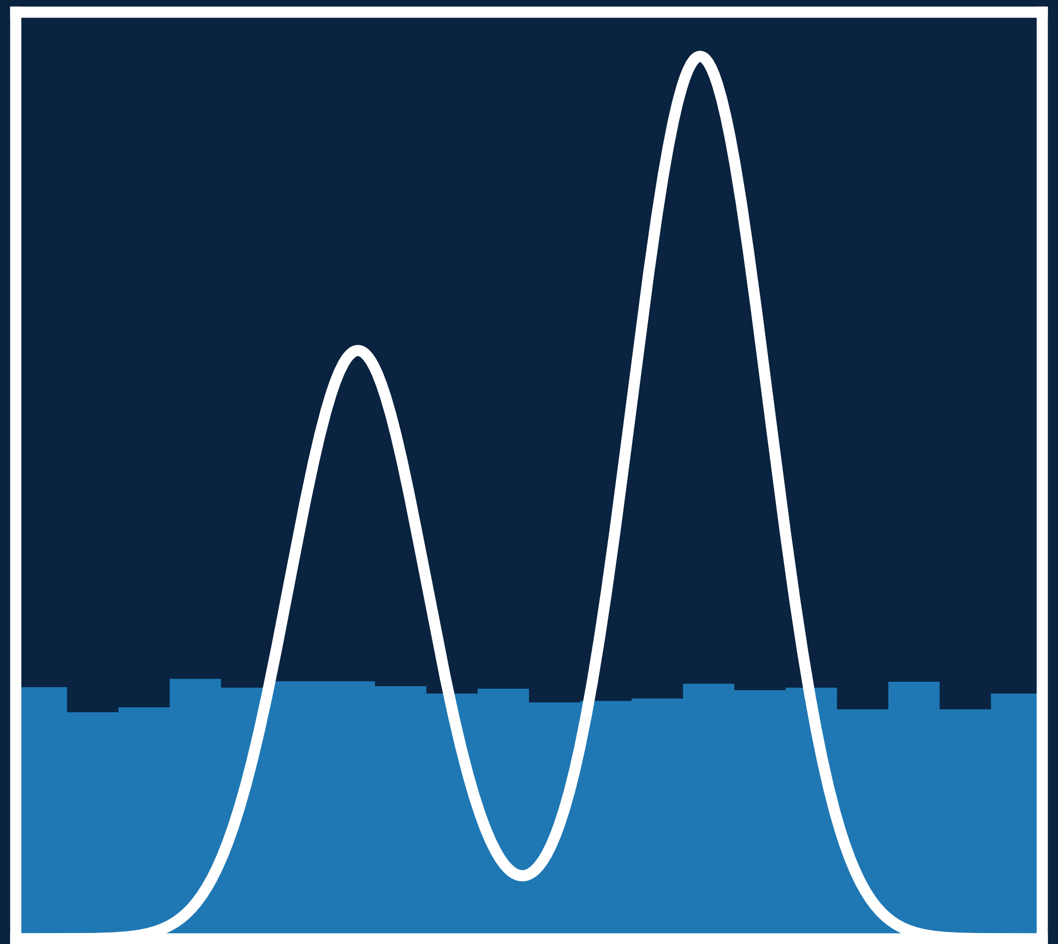
Phase space integration

$$\langle O \rangle = \int dx\, f(x)\, O(x)$$

BDT [1707.00028, …], NN [1810.11509, 2009.07819, …]
NF [2001.05486, 2001.05478, 2001.10028, 2005.12719, 2112.09145, 2212.06172, …]

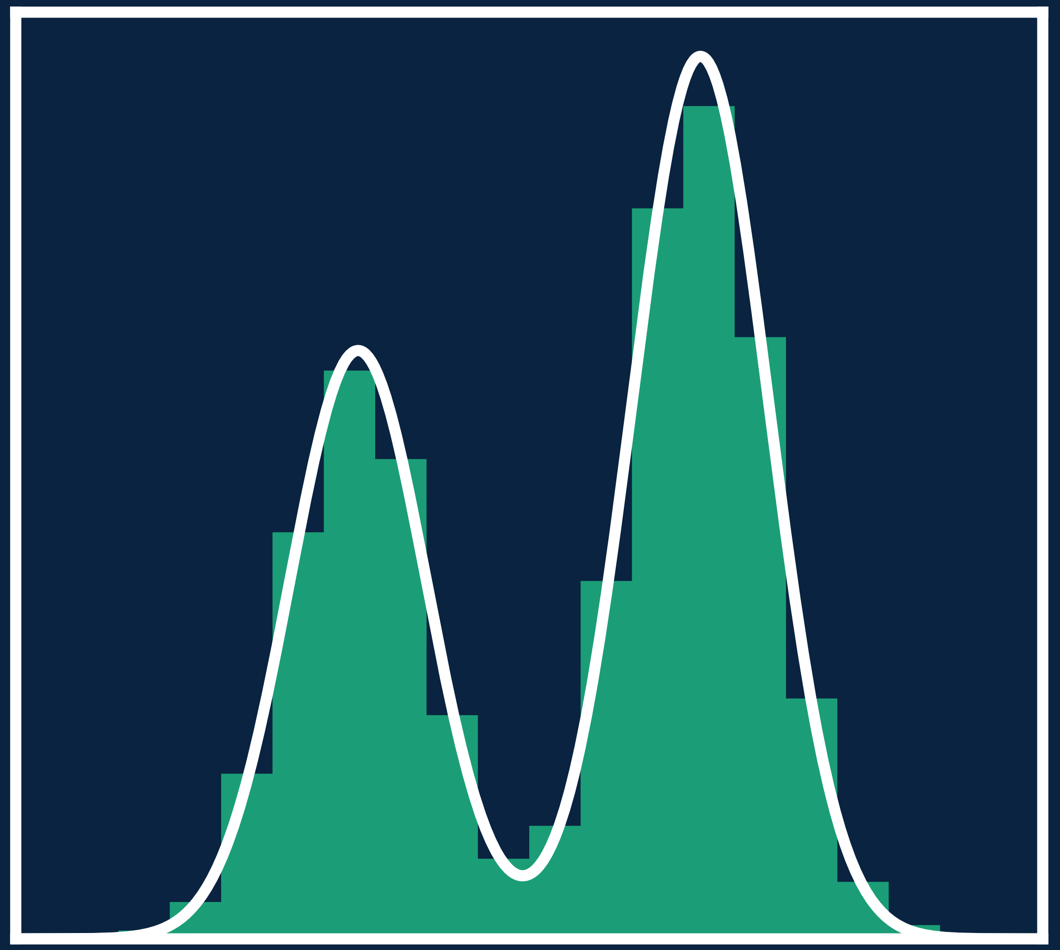
Monte Carlo integration

$$I = \int \mathrm{d}x\, f(x)$$



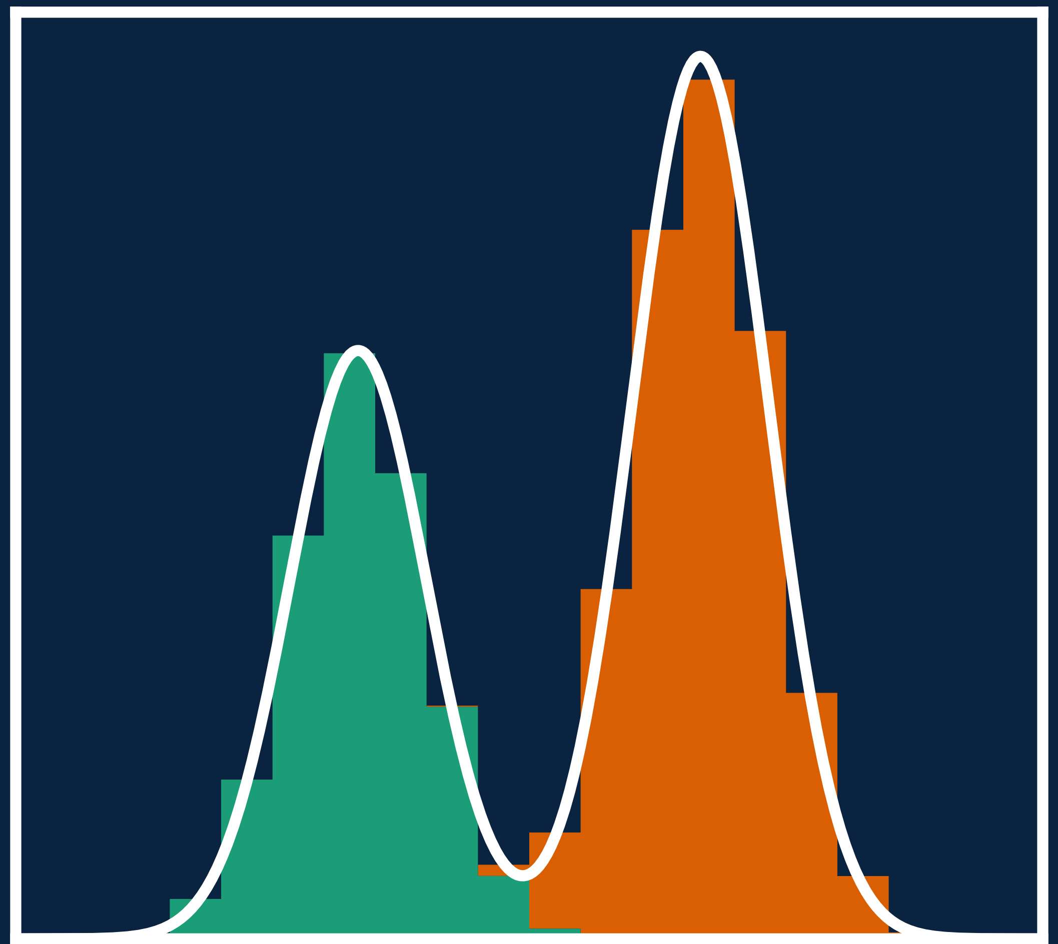
Flat sampling:
inefficient

$$I = \langle f(x) \rangle_{x \sim \mathrm{unif}}$$



Importance sampling:
find g close to f

$$I = \left\langle \frac{f(x)}{p(x)} \right\rangle_{x \sim p(x)}$$



Multi-channel:
one map for each channel

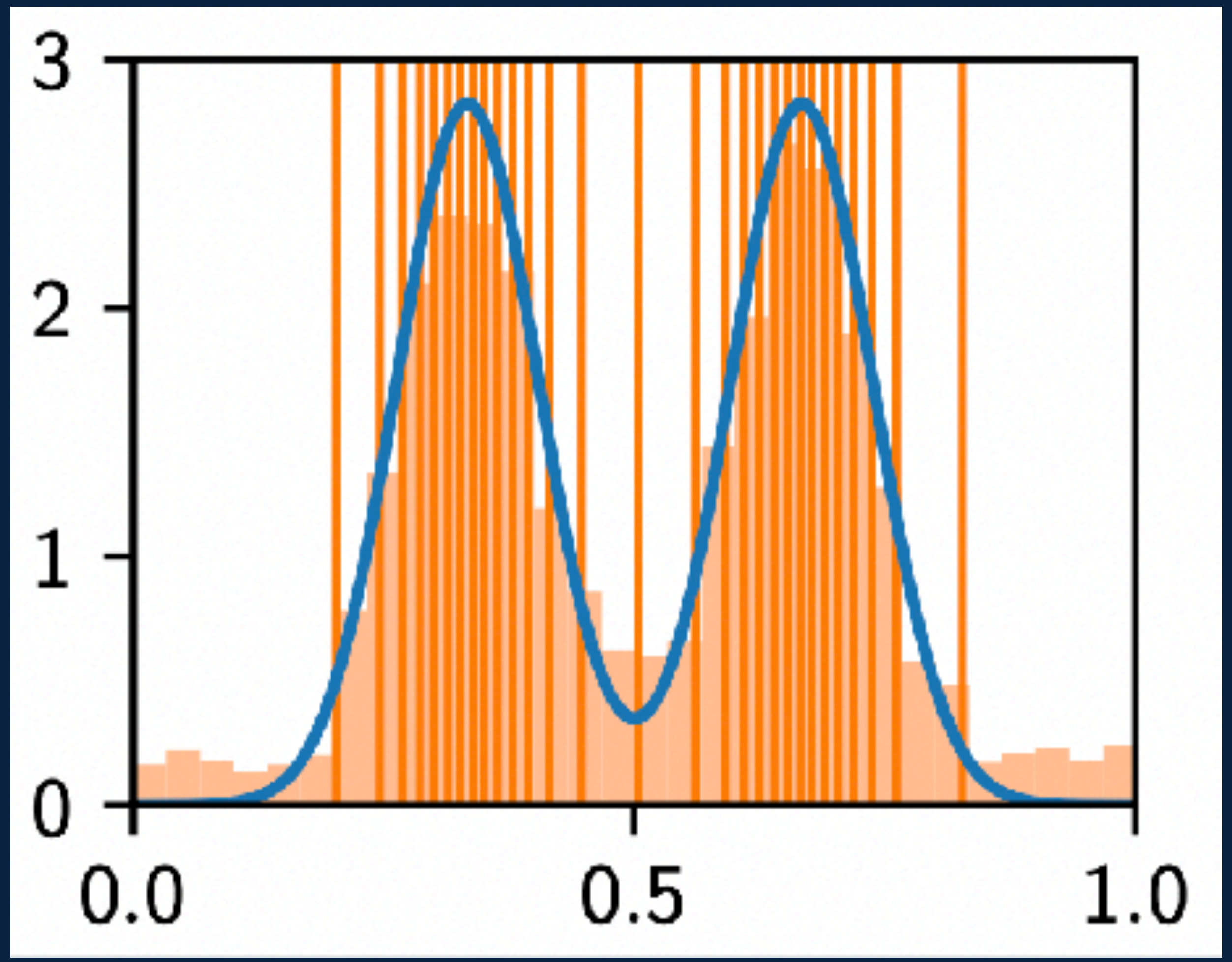
$$I = \sum_i \left\langle \alpha_i(x) \frac{f(x)}{p_i(x)} \right\rangle_{x \sim p_i(x)}$$

Importance sampling — VEGAS

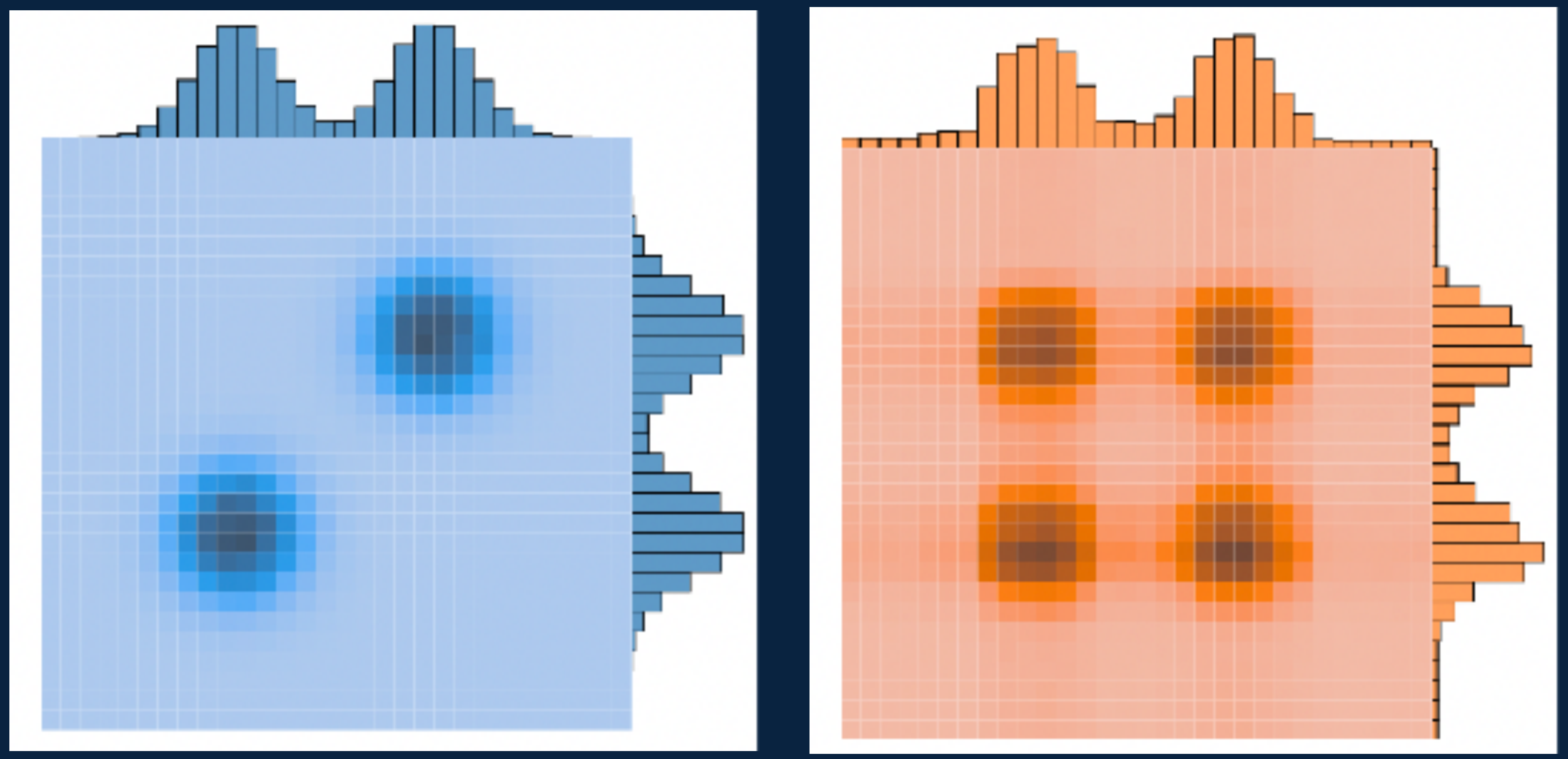
Factorize probability

$$p(x) = p(x_1)\cdots p(x_n)$$

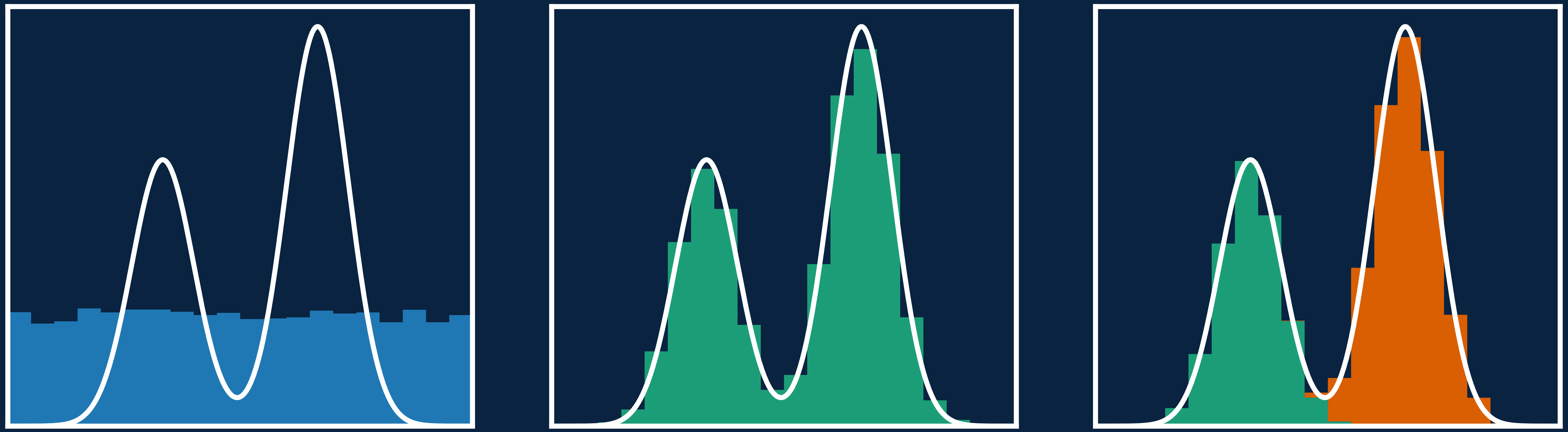
Fit bins with equal probability
and varying width



- ⊕ Computationally cheap
- ⊖ High-dim and rich peaking functions → **slow convergence**
- ⊖ Peaks not aligned with grid axes → **phantom peaks**



MadNIS — Neural importance sampling

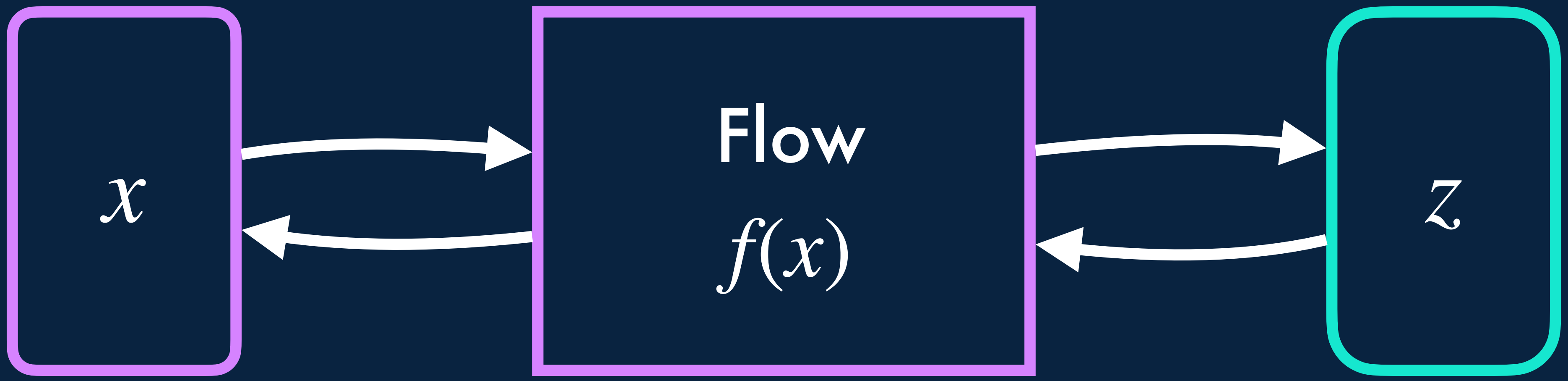


Flat sampling | Importance Sampling | Multi-channel

$$I = \sum_i \left\langle \alpha_i(x) \frac{f(x)}{p_i(x)} \right\rangle_{x\sim p_i(x)}$$

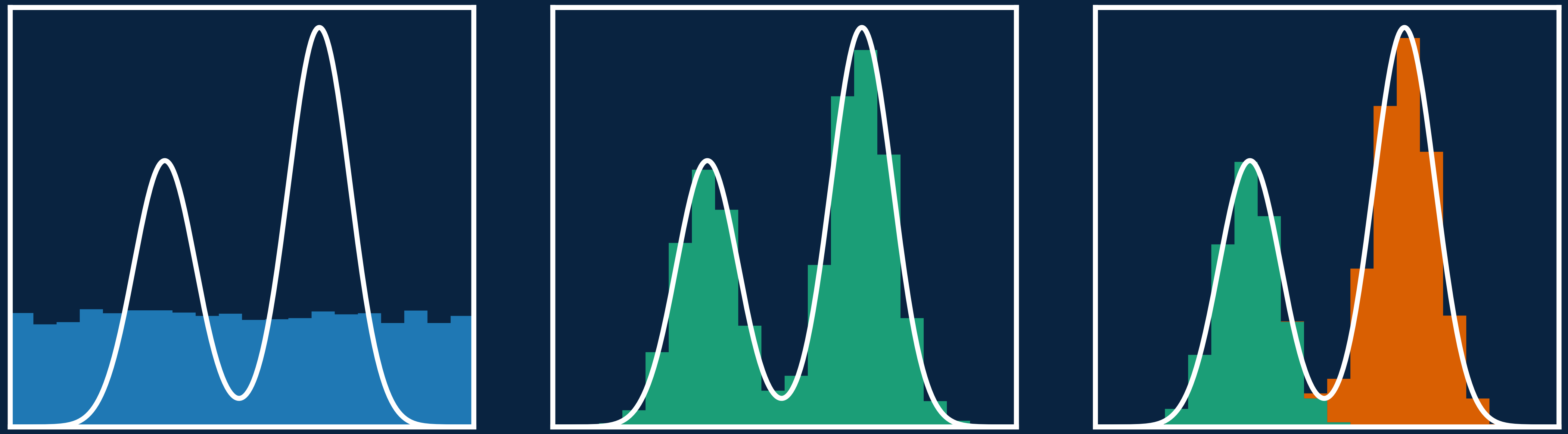
Parametrize with **NN**

Parametrize with **NF**



Sampling probability: $p(x) = \left| \frac{\partial f(x)}{\partial x} \right|$

MadNIS — Neural importance sampling



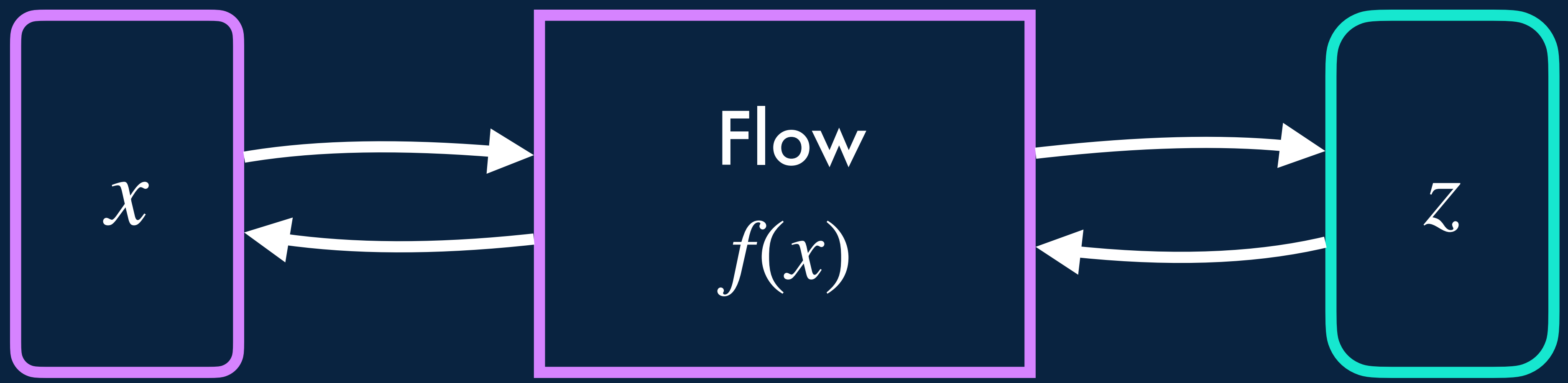
Flat sampling | Importance Sampling | Multi-channel

$$I = \sum_i \left\langle \alpha_i(x) \frac{f(x)}{p_i(x)} \right\rangle_{x \sim p_i(x)}$$

Parametrize with **NN**

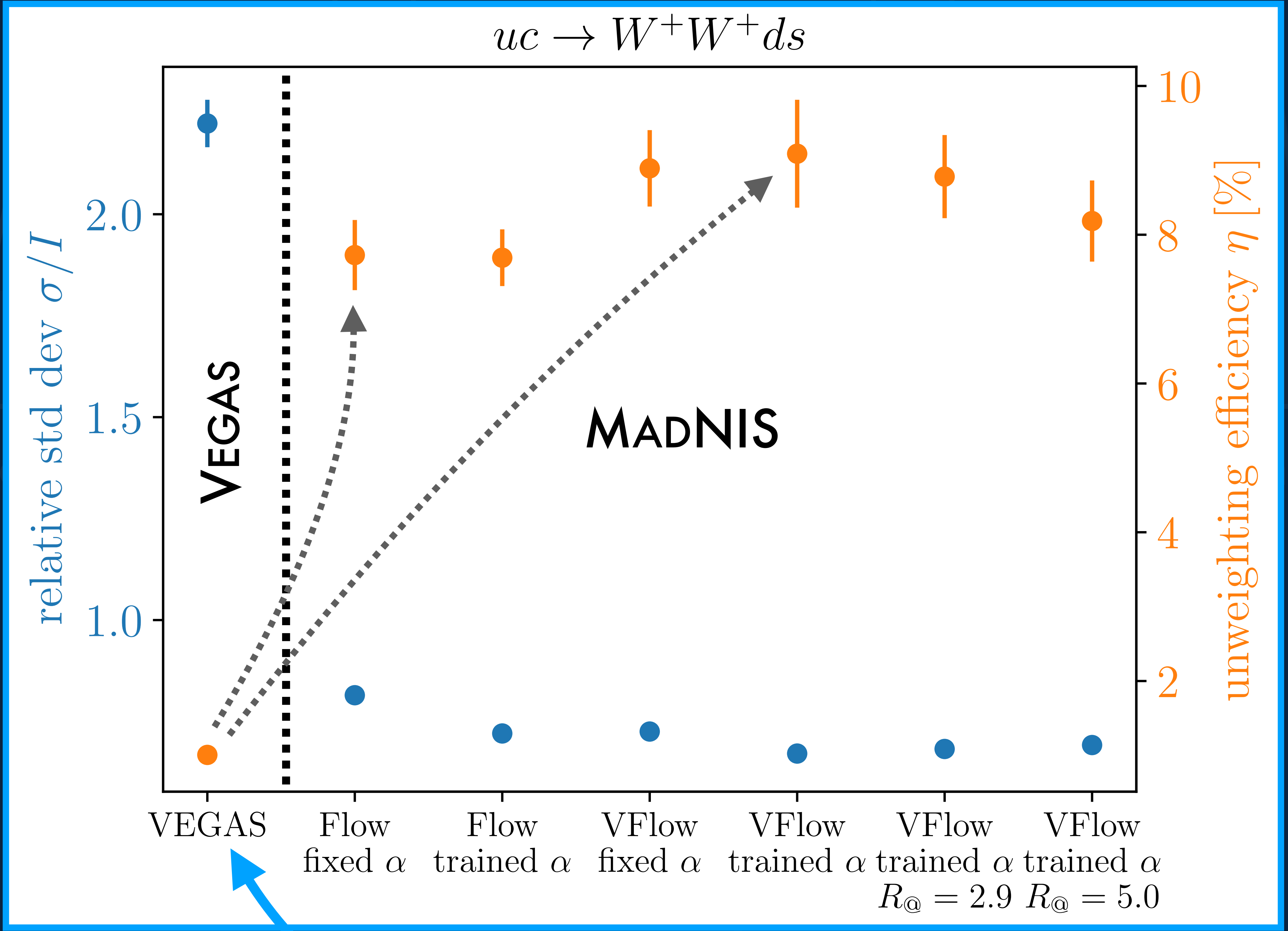
Parametrize with **NF**

⊕ unbinned & no grids
→ **no "phantom peaks"**
⊕ invertible & tractable Jacobians
→ **fast training and eval**



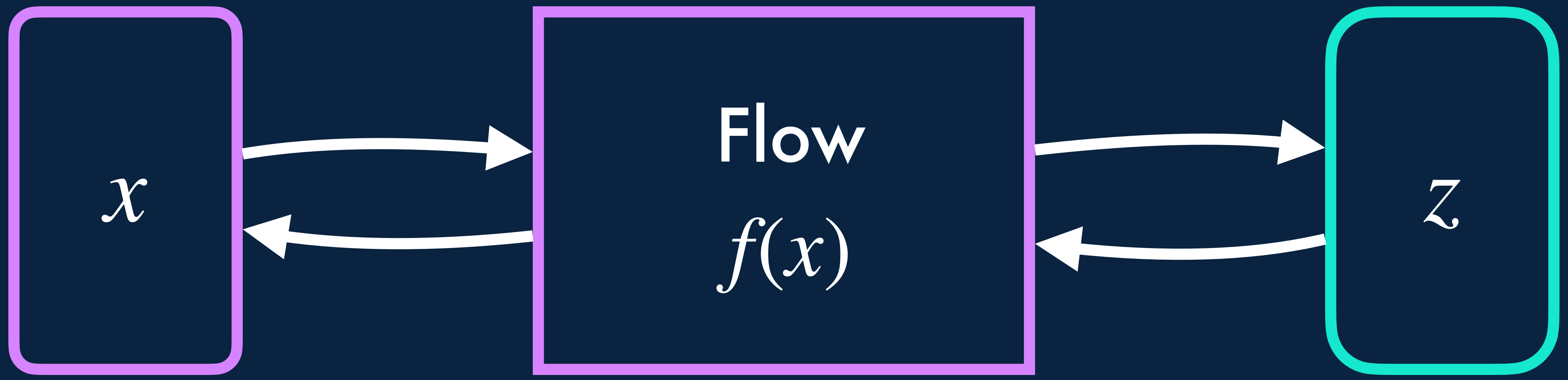
Sampling probability: $p(x) = \left| \frac{\partial f(x)}{\partial x} \right|$

MadNIS — Neural importance sampling



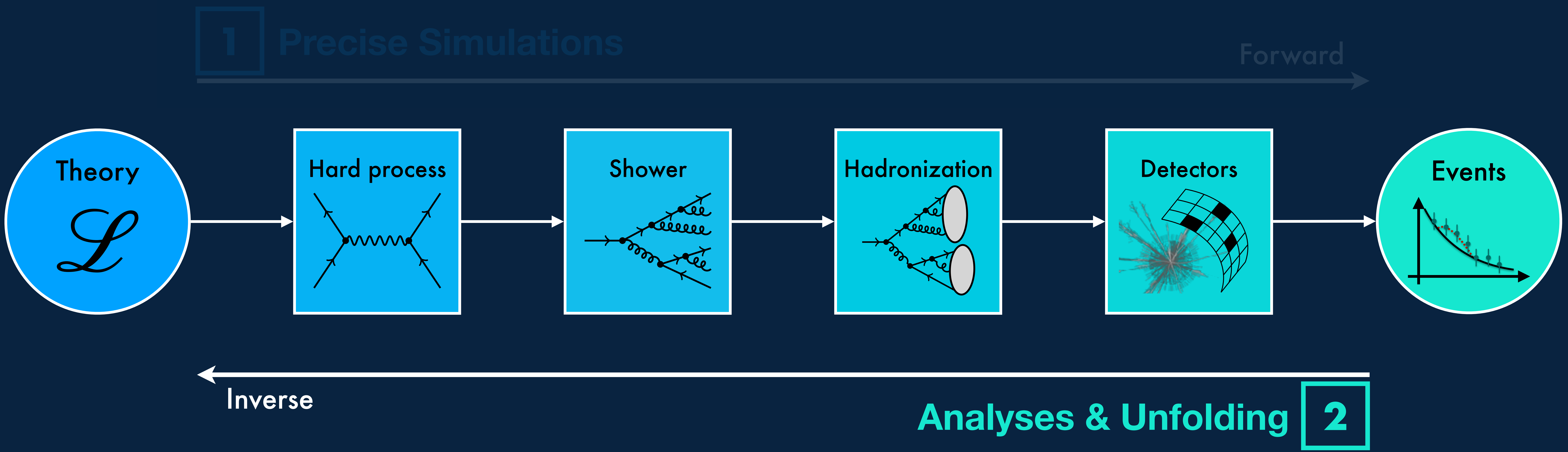
Unweighting efficiency improved up to **factor ~10** compared to Vegas

⊕ unbinned & no grids
→ **no "phantom peaks"**
⊕ invertible & tractable Jacobians
→ **fast training and eval**

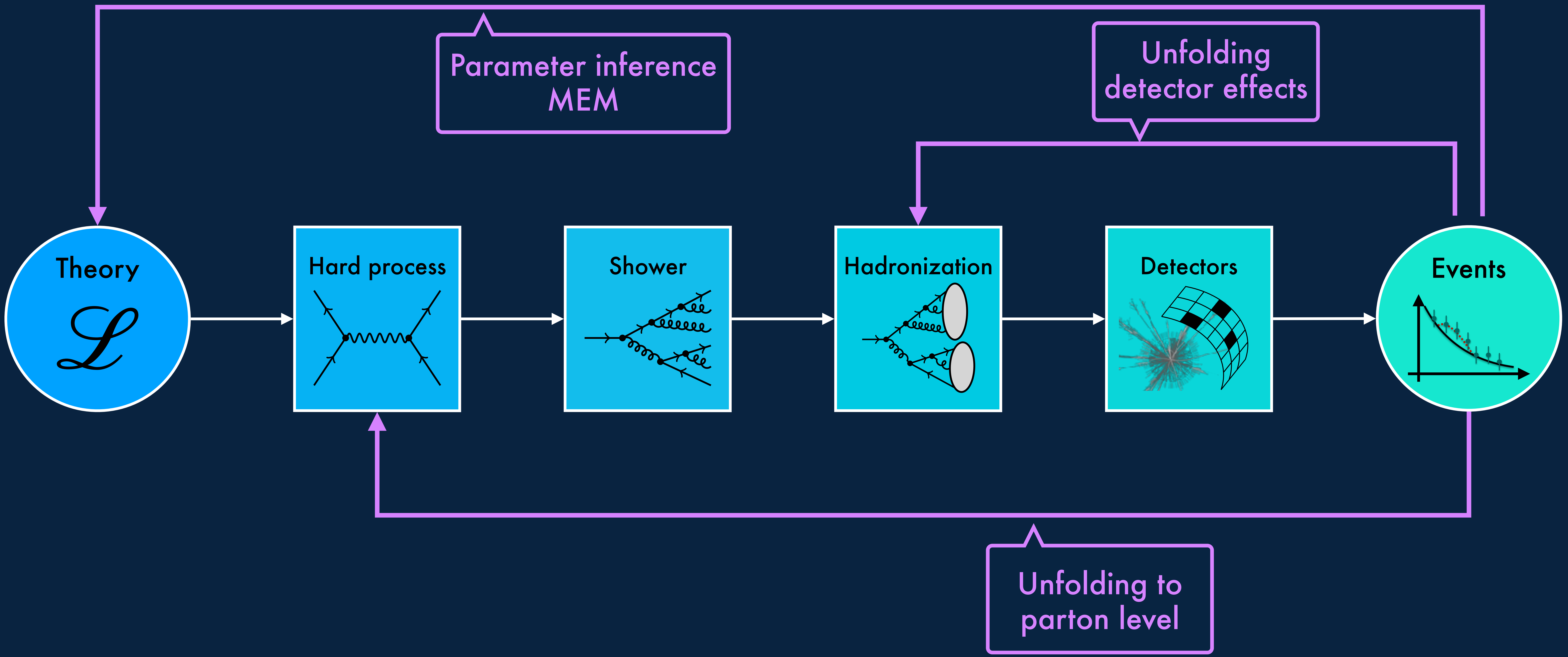


Sampling probability: $p(x) = \left| \frac{\partial f(x)}{\partial x} \right|$

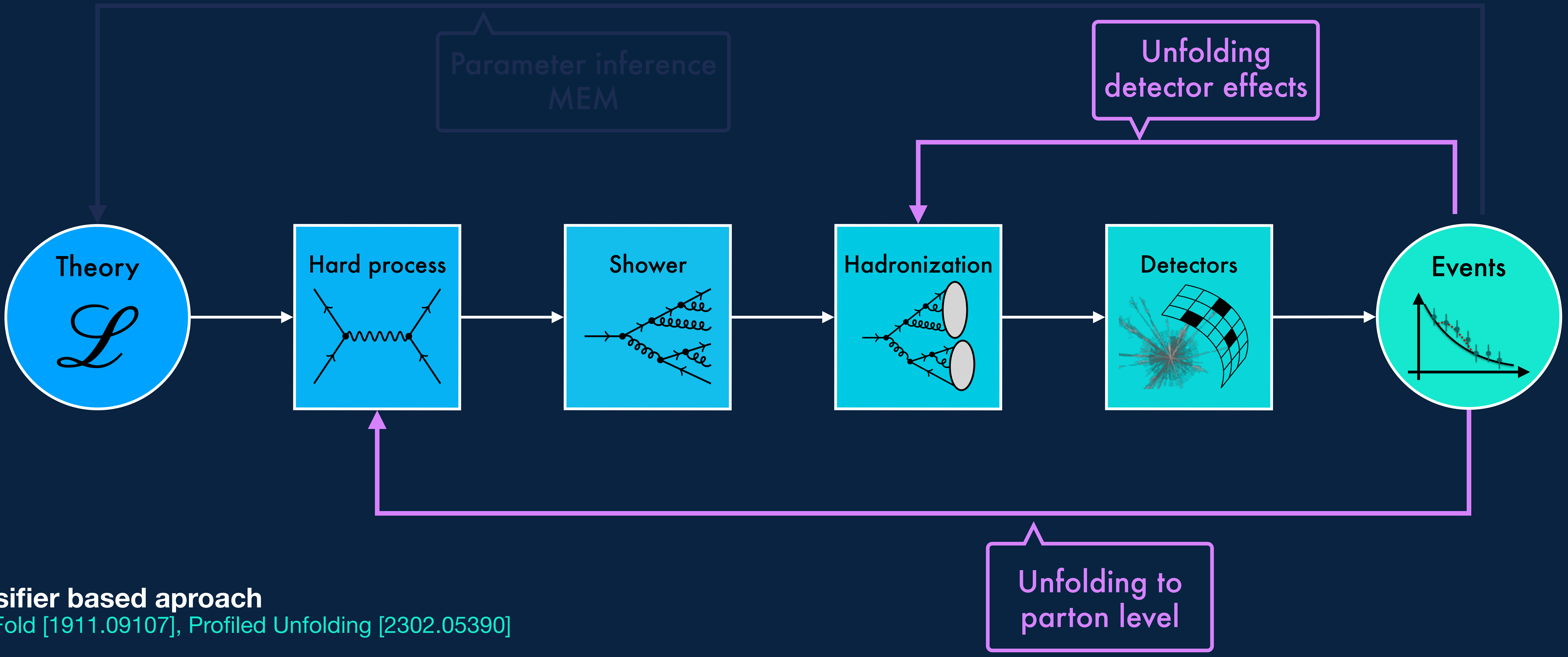
Inverting the simulation chain



Inverting the simulation chain



Unfolding at the LHC

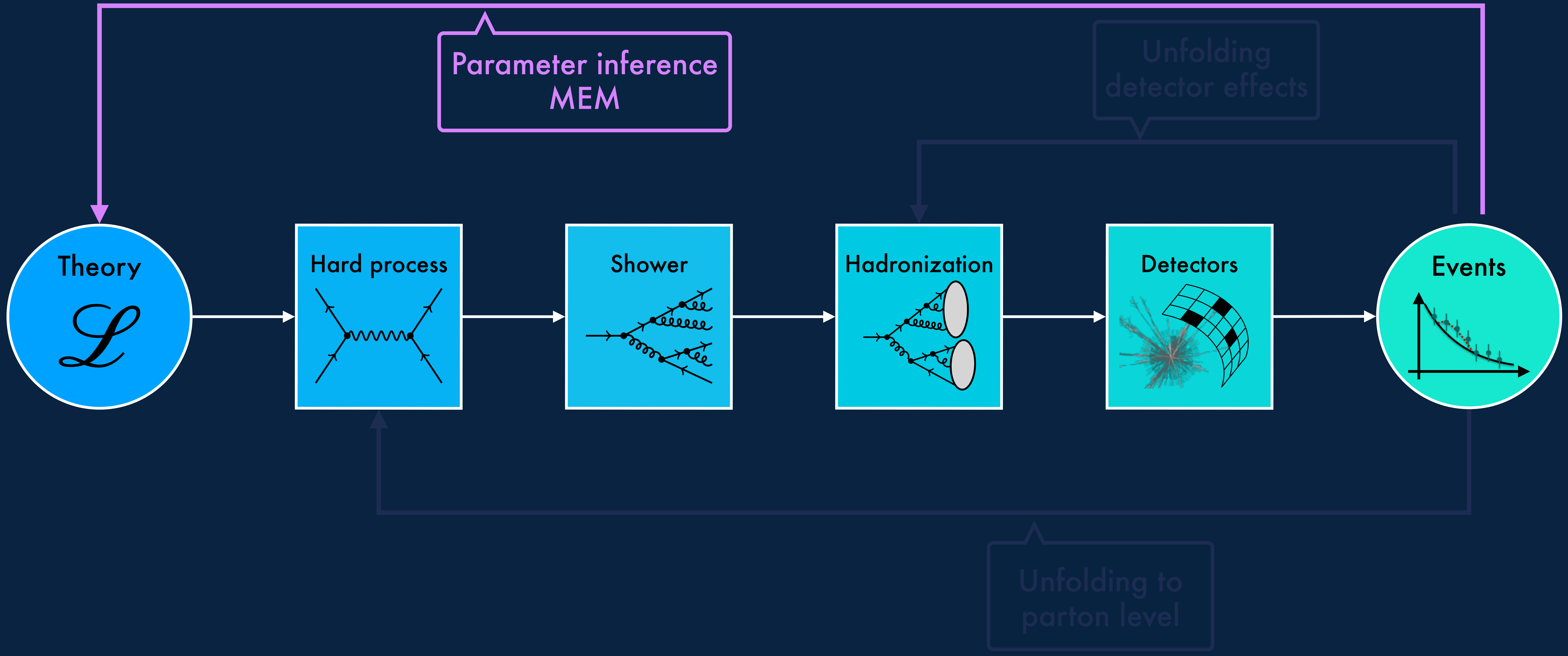


Classifier based aproach
OmniFold [1911.09107], Profiled Unfolding [2302.05390]

Density based approach
FCGAN [1912.00477], cINN [2006.06685],
IcINN [2212.08674], OTUS [2101.08944]

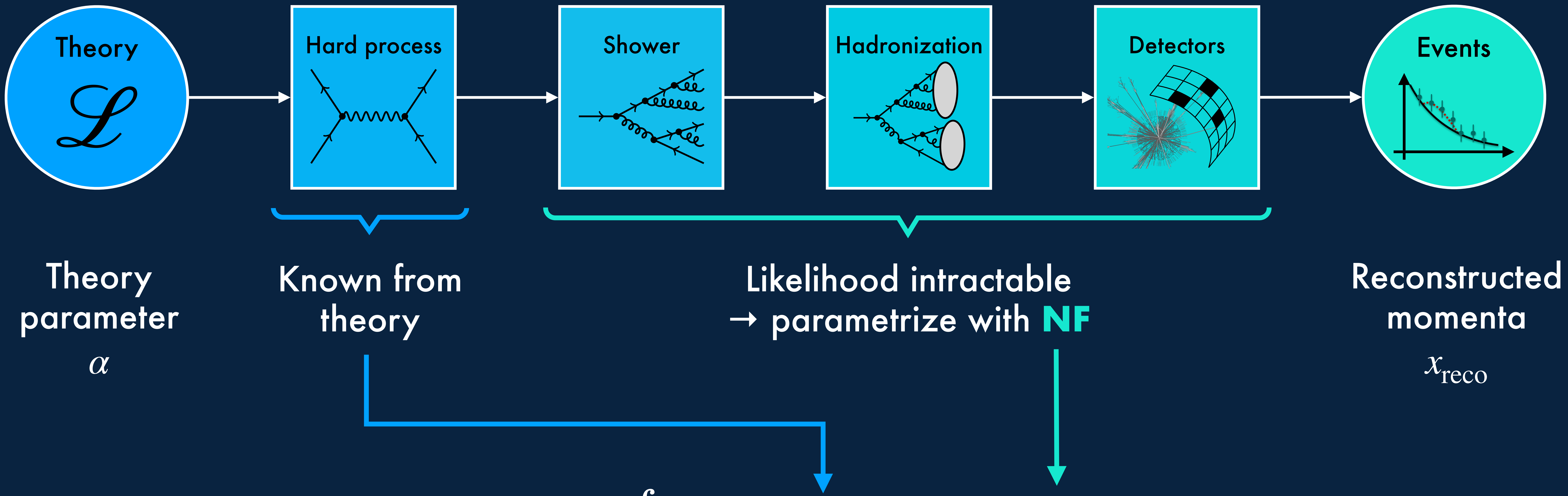
→ **Detailed Comparison:**
Arratia *et al* 2022 *JINST* 17 P01024 [2109.13243]

Inverting the simulation chain



Historically → Tevatron
Top mass: D0 (98', 04'), CDF 06', Fiedler et al. [1003.1316]
Single-top: Review [1710.10699]

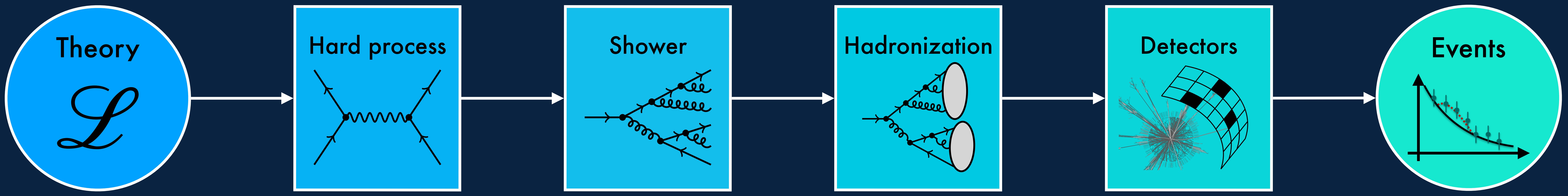
Inference with normalizing flows



MEM master formula:

$$p(x_\text{reco}\,|\,\alpha) = \int \mathrm{d}x_\text{hard}\; p(x_\text{hard}\,|\,\alpha)\, p(x_\text{reco}\,|\,x_\text{hard})$$

Inference with normalizing flows



Theory parameter α

Known from theory

Likelihood intractable
→ parametrize with **NF**

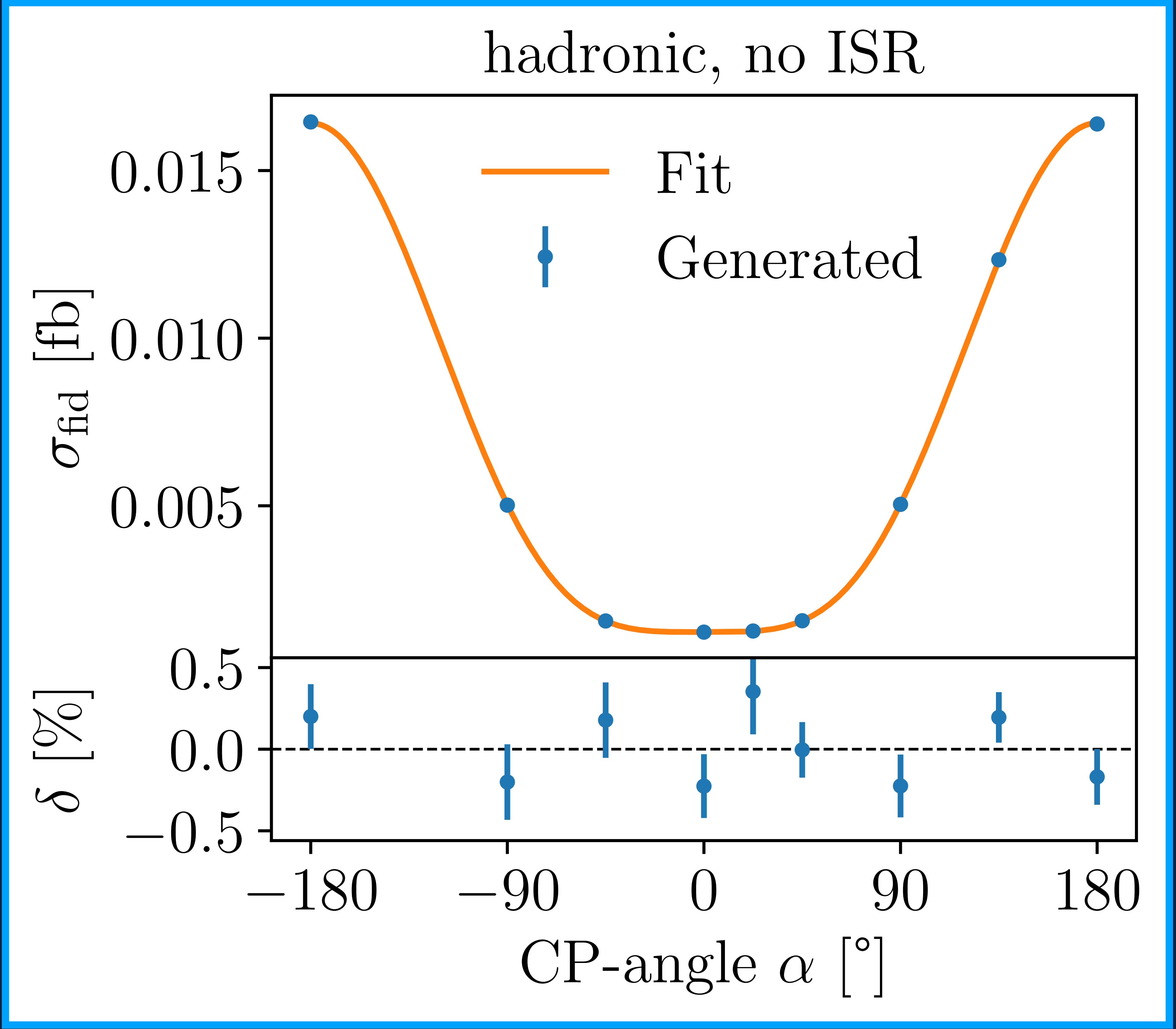
Reconstructed momenta x_{reco}

MEM master formula: $$p(x_{\mathrm{reco}}\,|\,\alpha) = \int \mathrm{d}x_{\mathrm{hard}}\, p(x_{\mathrm{hard}}\,|\,\alpha)\, p(x_{\mathrm{reco}}\,|\,x_{\mathrm{hard}})$$

In practice → perform integral **numerically**

Matrix element method

Heidelberg/Louvain [2210.00019, 23XX.xxxxx]



tHj production:

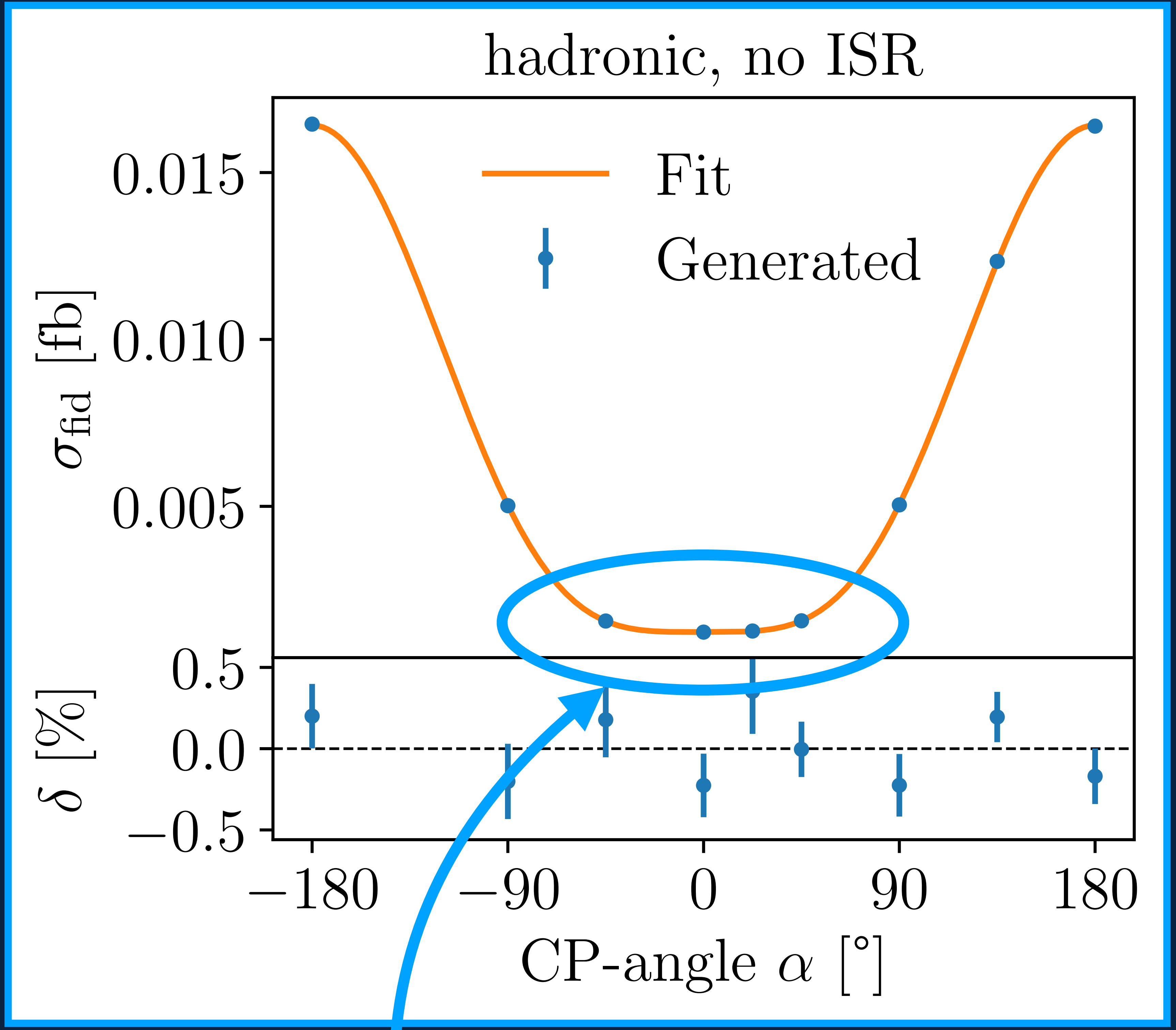
$$\begin{aligned} \mathrm{pp} &\to \mathrm{tH\,j} \\ &\to (\mathrm{bW})\,(\gamma\gamma)\,\mathrm{j} \end{aligned}$$

$$\mathscr{L}_{\mathrm{t\bar{t}H}} = -\frac{y_t}{\sqrt{2}}\left[\cos\alpha\;\bar{\mathrm{t}}\mathrm{t} + \frac{2}{3}\mathrm{i}\sin\alpha\;\bar{\mathrm{t}}\gamma_5\mathrm{t}\right]\mathrm{H}$$

Anomalous coupling
with **CP-angle** α

Matrix element method

Heidelberg/Louvain [2210.00019, 23XX.xxxxx]



tHj production:

$$\mathrm{pp} \to \mathrm{tH\,j}$$
$$\to (\mathrm{bW})\,(\gamma\gamma)\,\mathrm{j}$$

$$\mathscr{L}_{\mathrm{t\bar{t}H}} = -\frac{y_t}{\sqrt{2}}\left[\cos\alpha\;\bar{\mathrm{t}}\mathrm{t} + \frac{2}{3}\mathrm{i}\sin\alpha\;\bar{\mathrm{t}}\gamma_5\mathrm{t}\right]\mathrm{H}$$

Anomalous coupling with **CP-angle** α

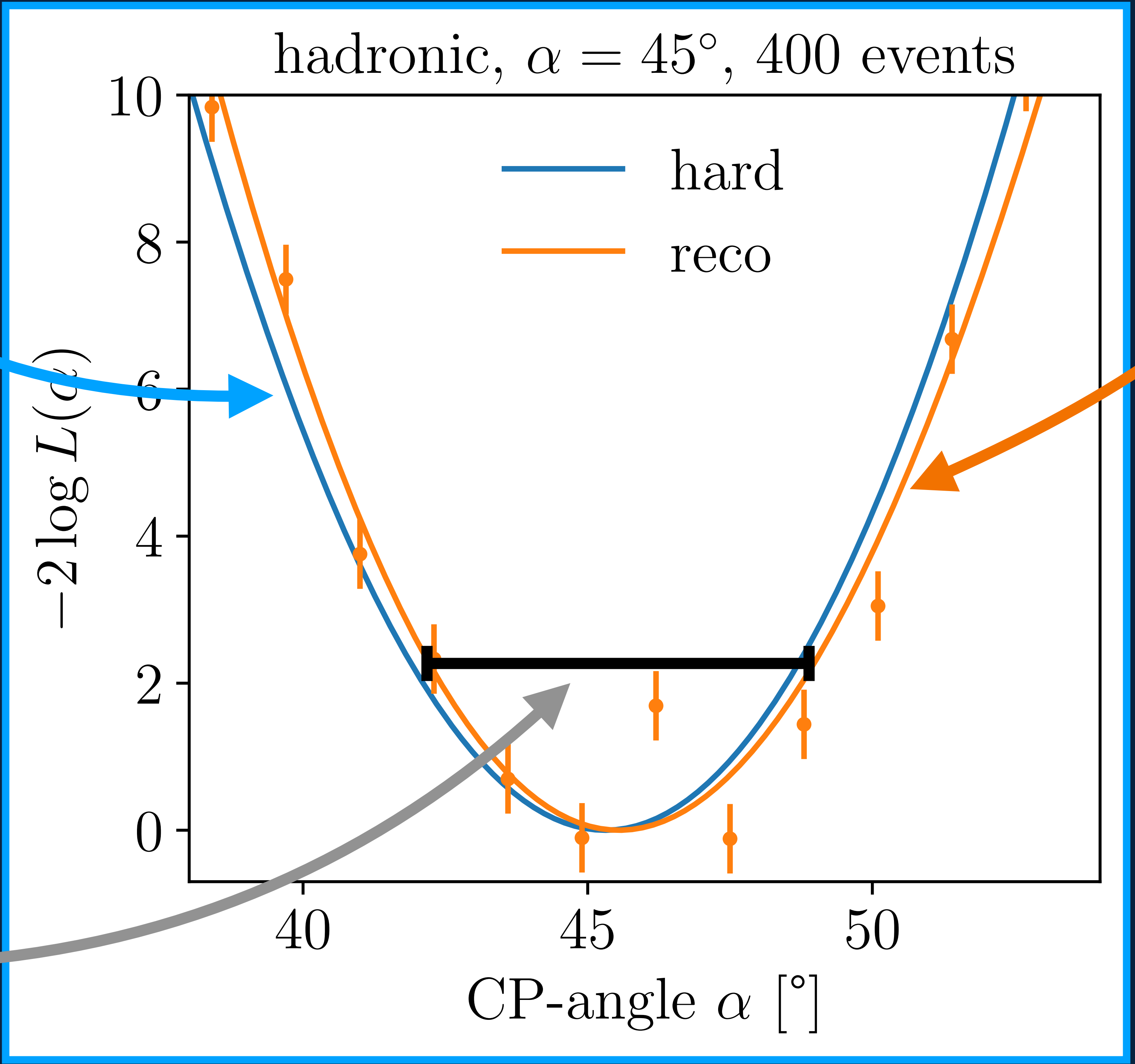
Around SM ($\alpha = 0$):

- ⊖ low total cross section (few events)
- ⊖ low variation of rate
- ⊕ kinematics sensitive → ideal use case for **MEM**

Matrix element method

Hard-scattering truth for comparison

Heidelberg/Louvain [2210.00019, 23XX.xxxxx]



Result from MEM

Statistical uncertainty

tHj production:

$$pp \to tH\,j \to (bW)\,(\gamma\gamma)\,j$$

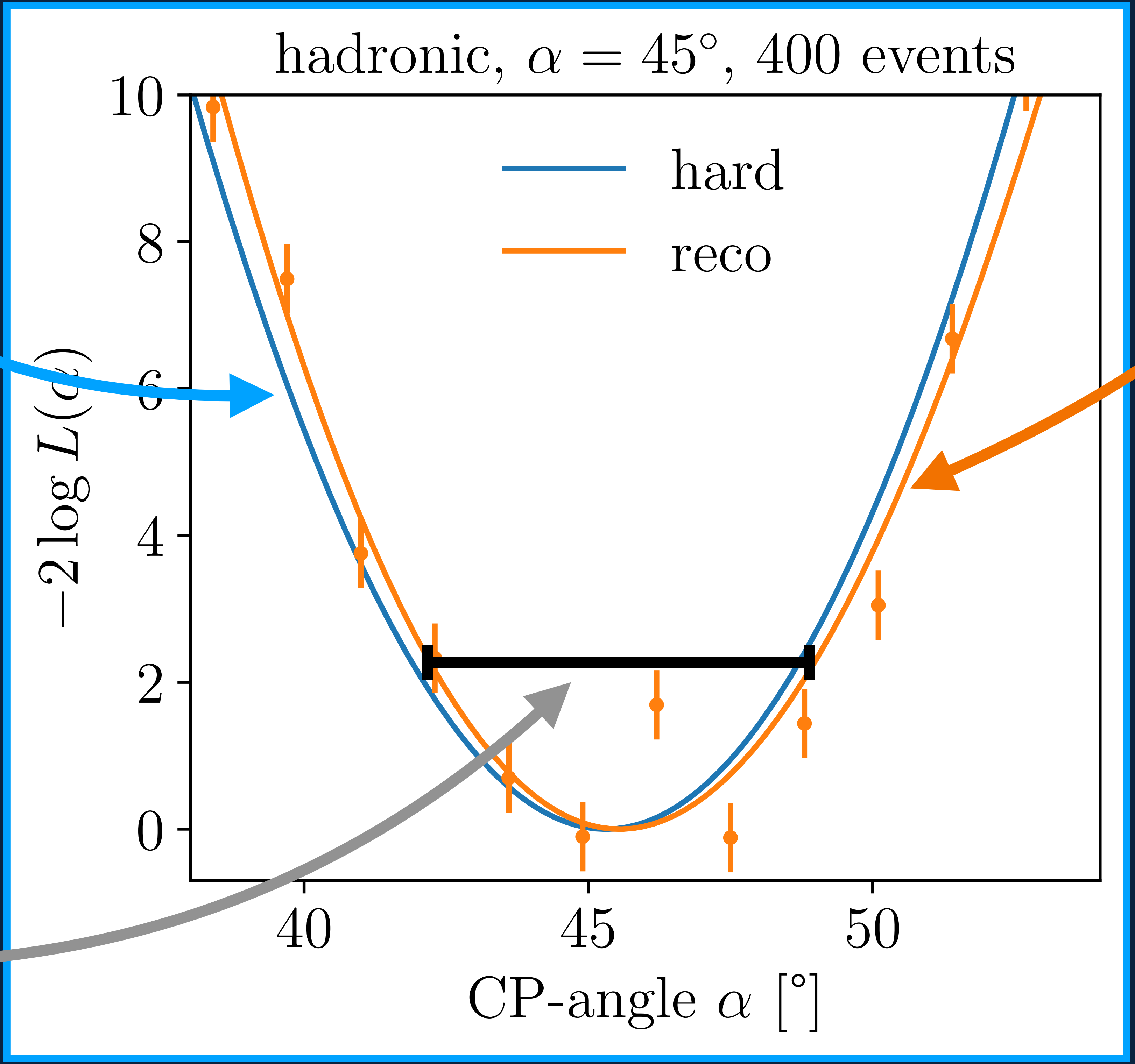
$$\mathscr{L}_{t\bar{t}H} = -\frac{y_t}{\sqrt{2}}\left[\cos\alpha\,\bar{t}t + \frac{2}{3}i\sin\alpha\,\bar{t}\gamma_5 t\right]H$$

Anomalous coupling with **CP-angle** α

Matrix element method

Heidelberg/Louvain [2210.00019, 23XX.xxxxx]

Hard scattering truth for comparison



Result from MEM

Statistical uncertainty

tHj production:

$$\begin{aligned} pp &\to tHj \\ &\to (bW)\,(\gamma\gamma)\,j \end{aligned}$$

$$\mathscr{L}_{t\bar{t}H} = -\frac{y_t}{\sqrt{2}}\left[\cos\alpha\,\bar{t}t + \frac{2}{3} i \sin\alpha\,\bar{t}\gamma_5 t\right]H$$

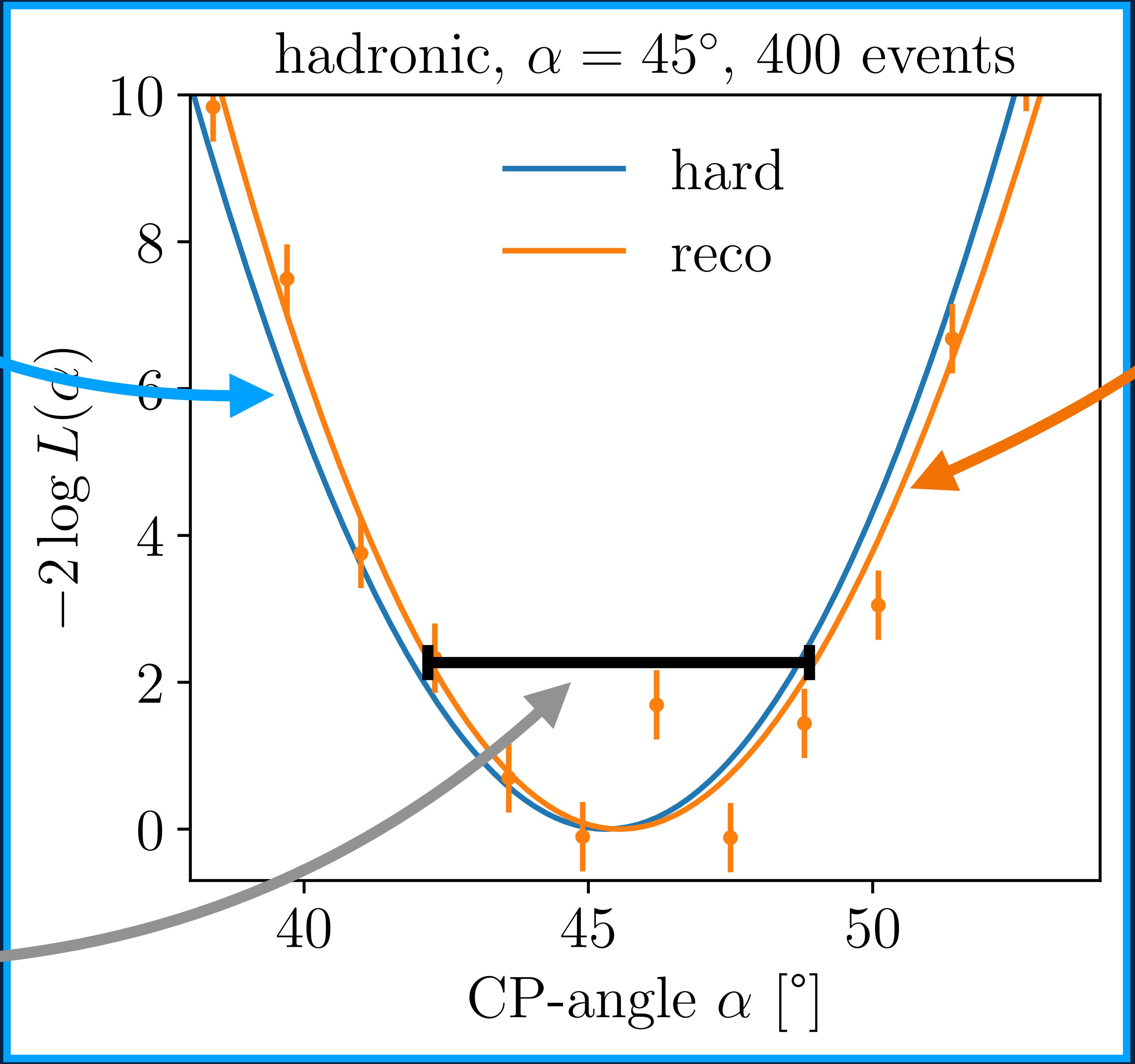
Anomalous coupling with **CP-angle** α

Uncertainties from training of neural network?
→ **Bayesian neural networks**

Matrix element method

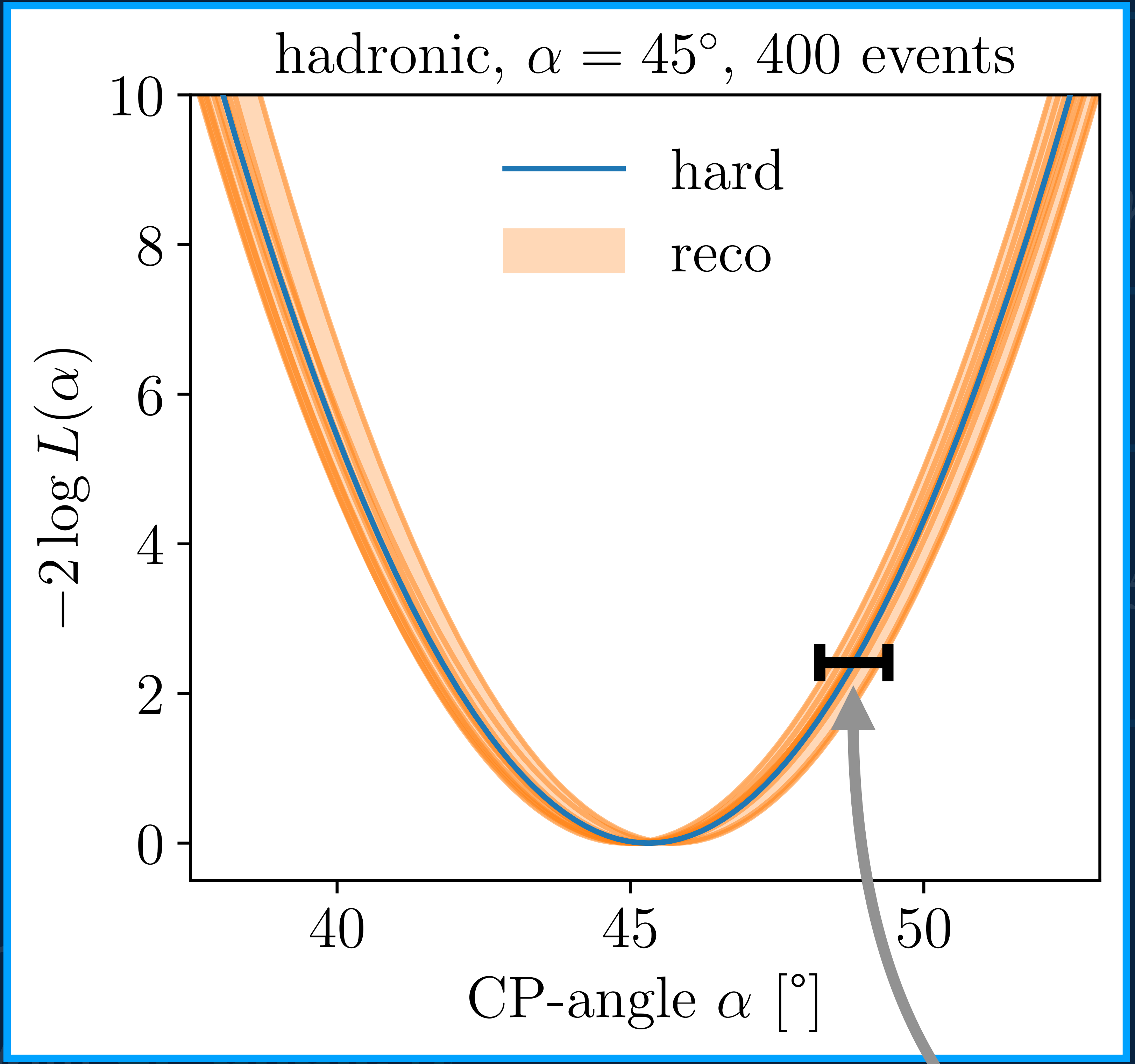
Hard scattering truth for comparison

Heidelberg/Louvain [2210.00019, 23XX.xxxxx]



Result from MEM

Statistical uncertainty



Systematic uncertainty from training

Uncertainties from training of neural network?
→ Bayesian neural networks

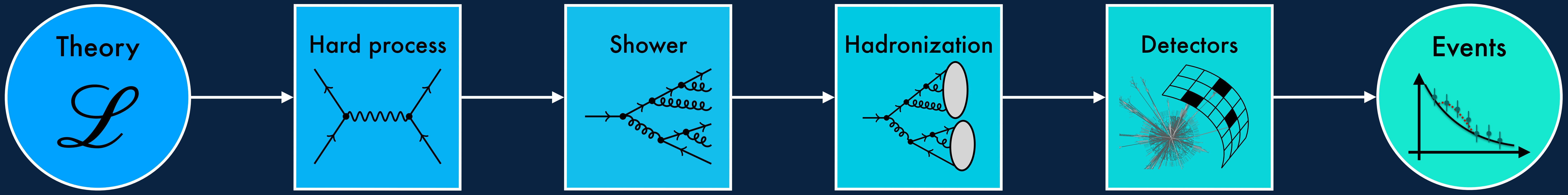
Summary and Outlook

Take-home message

- **Fast** and **precise** predictions with ML-based simulations
- Normalizing flows provide statistically **well-defined likelihoods** for inference
- Account for **uncertainties** with **Bayesian neural networks**

Future exercises

- **Full integration** of ML-based simulations into standard tools → **MadGraph,….**
- Make everything run on the **GPU and differentiable** (MadJax - Heinrich et al. [2203.00057])
- Foster collaboration between **theory** and **experiment**



Summary and Outlook

SciPost | SciPost Phys. 14, 079 (2023)

Machine learning and LHC event generation

Anja Butter[1,2], Tilman Plehn[1], Steffen Schumann[3], Simon Badger[4], Sascha Caron[5,6] Kyle Cranmer[7,8], Francesco Armando Di Bello[9], Etienne Dreyer[10], Stefano Forte[11], Sanmay Ganguly[12], Dorival Gonçalves[13], Eilam Gross[10], Theo Heimel[1], Gudrun Heinrich[14], Lukas Heinrich[15], Alexander Held[16], Stefan Höche[17], Jessica N. Howard[18], Philip Ilten[19], Joshua Isaacson[17], Timo Janßen[3], Stephen Jones[20], Marumi Kado[9,21], Michael Kagan[22], Gregor Kasieczka[23], Felix Kling[24], Sabine Kraml[25], Claudius Krause[26], Frank Krauss[20], Kevin Kröninger[27], Rahool Kumar Barman[13], Michel Luchmann[1], Vitaly Magerya[14], Daniel Maitre[20], Bogdan Malaescu[2], Fabio Maltoni[28,29], Till Martini[30], Olivier Mattelaer[28], Benjamin Nachman[31,32], Sebastian Pitz[1], Juan Rojo[6,33], Matthew Schwartz[34], David Shih[25], Frank Siegert[35], Roy Stegeman[11], Bob Stienen[5], Jesse Thaler[36], Rob Verheyen[37], Daniel Whiteson[18], Ramon Winterhalder[28], and Jure Zupan[19]

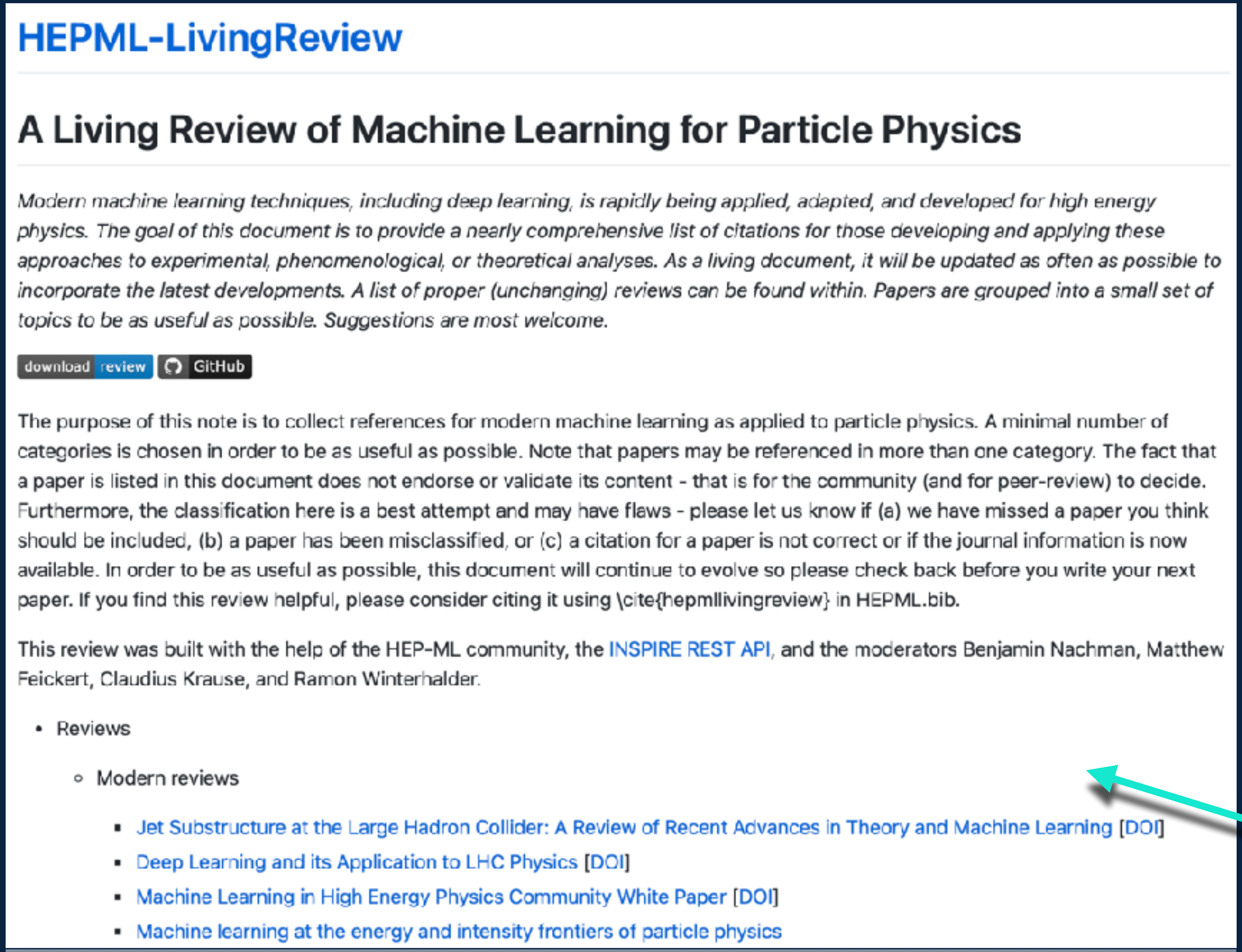
Abstract

First-principle simulations are at the heart of the high-energy physics research program. They link the vast data output of multi-purpose detectors with fundamental theory predictions and interpretation. This review illustrates a wide range of applications of modern machine learning to event generation and simulation-based inference, including conceptional developments driven by the specific requirements of particle physics. New ideas and tools developed at the interface of particle physics and machine learning will improve the speed and precision of forward simulations, handle the complexity of collision data, and enhance inference as an inverse simulation problem.

Future exercises

- **Full integration** of ML-based simulations into standard tools → **MadGraph,….**
- Make everything run on the **GPU and differentiable** (MadJax - Heinrich et al. [2203.00057])
- Foster collaboration between **theory** and **experiment**
- More details in our **Snowmass report**

Summary and Outlook



HEPML-LivingReview

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

download review | GitHub

The purpose of this note is to collect references for modern machine learning as applied to particle physics. A minimal number of categories is chosen in order to be as useful as possible. Note that papers may be referenced in more than one category. The fact that a paper is listed in this document does not endorse or validate its content - that is for the community (and for peer-review) to decide. Furthermore, the classification here is a best attempt and may have flaws - please let us know if (a) we have missed a paper you think should be included, (b) a paper has been misclassified, or (c) a citation for a paper is not correct or if the journal information is now available. In order to be as useful as possible, this document will continue to evolve so please check back before you write your next paper. If you find this review helpful, please consider citing it using \cite{hepmllivingreview} in HEPML.bib.

This review was built with the help of the HEP-ML community, the INSPIRE REST API, and the moderators Benjamin Nachman, Matthew Feickert, Claudius Krause, and Ramon Winterhalder.

- Reviews
  - Modern reviews
    - Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning [DOI]
    - Deep Learning and its Application to LHC Physics [DOI]
    - Machine Learning in High Energy Physics Community White Paper [DOI]
    - Machine learning at the energy and intensity frontiers of particle physics

Future exercises

- **Full integration** of ML-based simulations into standard tools → **MadGraph,….**
- Make everything run on the **GPU and differentiable** (MadJax - Heinrich et al. [2203.00057])
- Foster collaboration between **theory** and **experiment**
- More details in our **Snowmass report**
- Stay tuned for many other **ML4HEP applications**