European Research Council

Established by the European Commission

PBSP

MARIA UBIALI

UNIVERSITY OF CAMBRIDGE

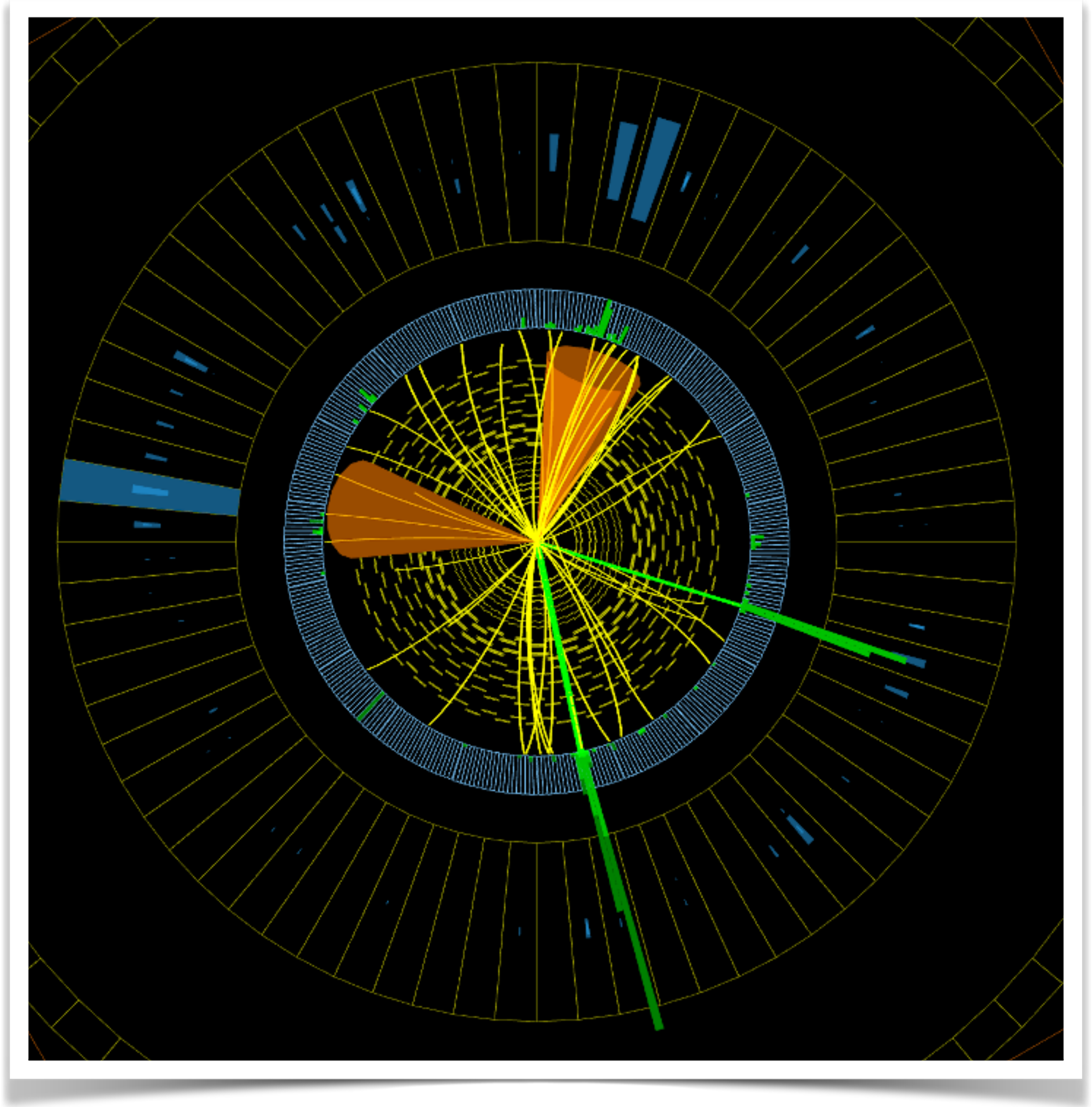
MODERN MACHINE LEARNING AND PARTICLE PHYSICS PHENOMENOLOGY AT THE LHC

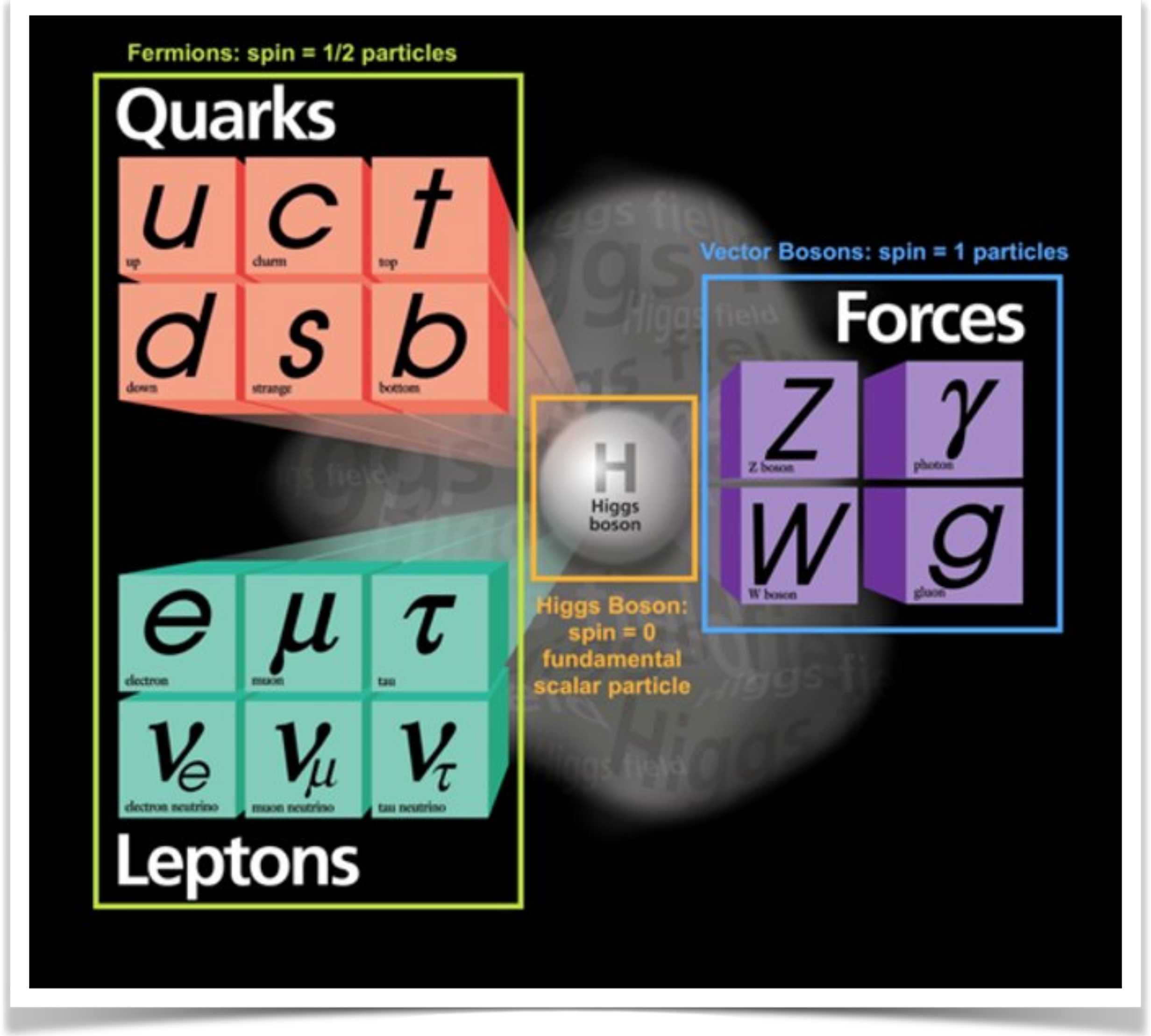
EUCAIFCON 2025, CAGLIARI, ITALY

17TH JUNE 2025

PARTICLE PHYSICS PHENOMENOLOGY AT THE LHC

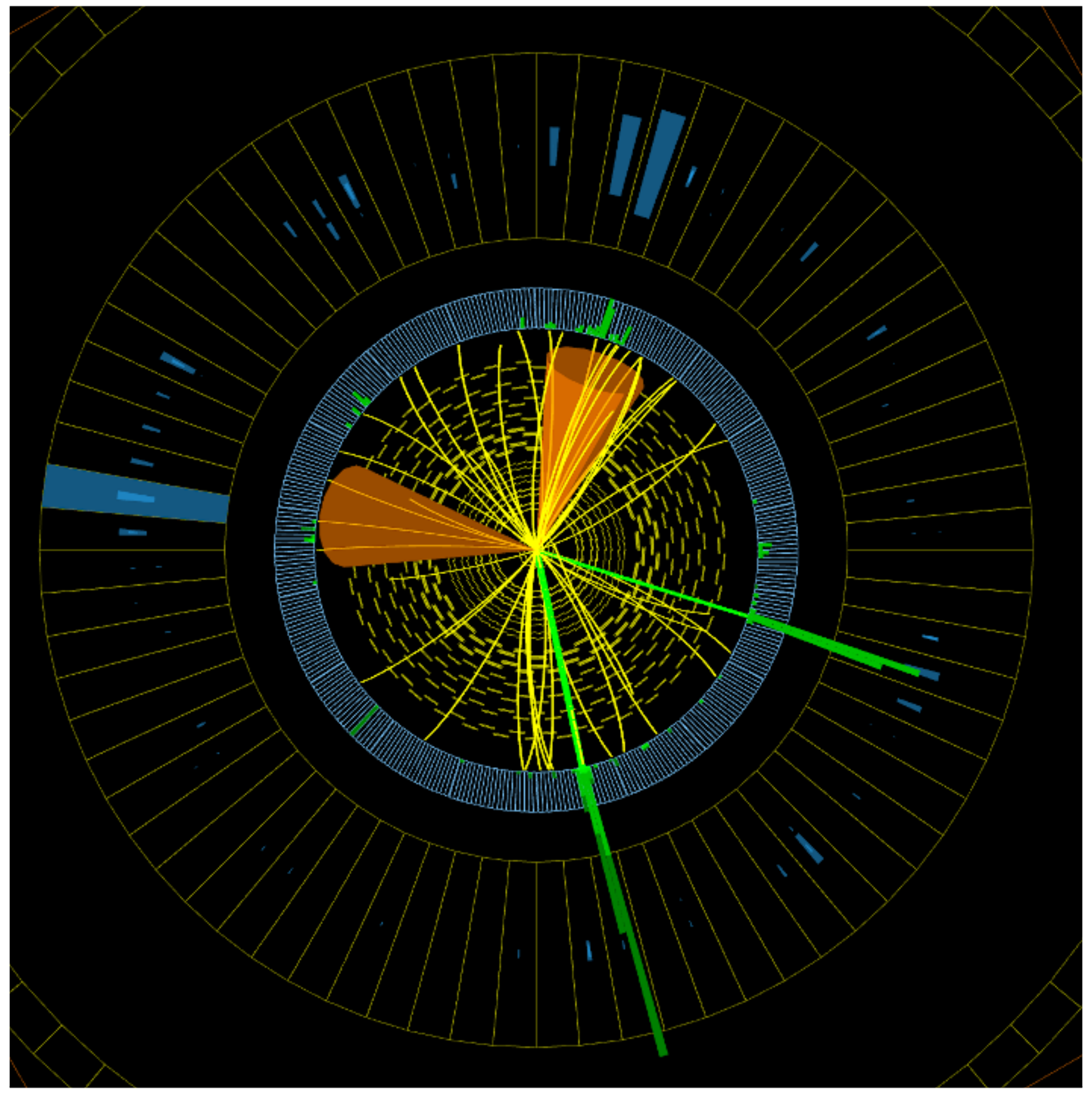
Experimental observations ⟷ HEP - PH ⟷ Theoretical model ⟷ HEP - TH ⟷ Ultimate Theory including gravity

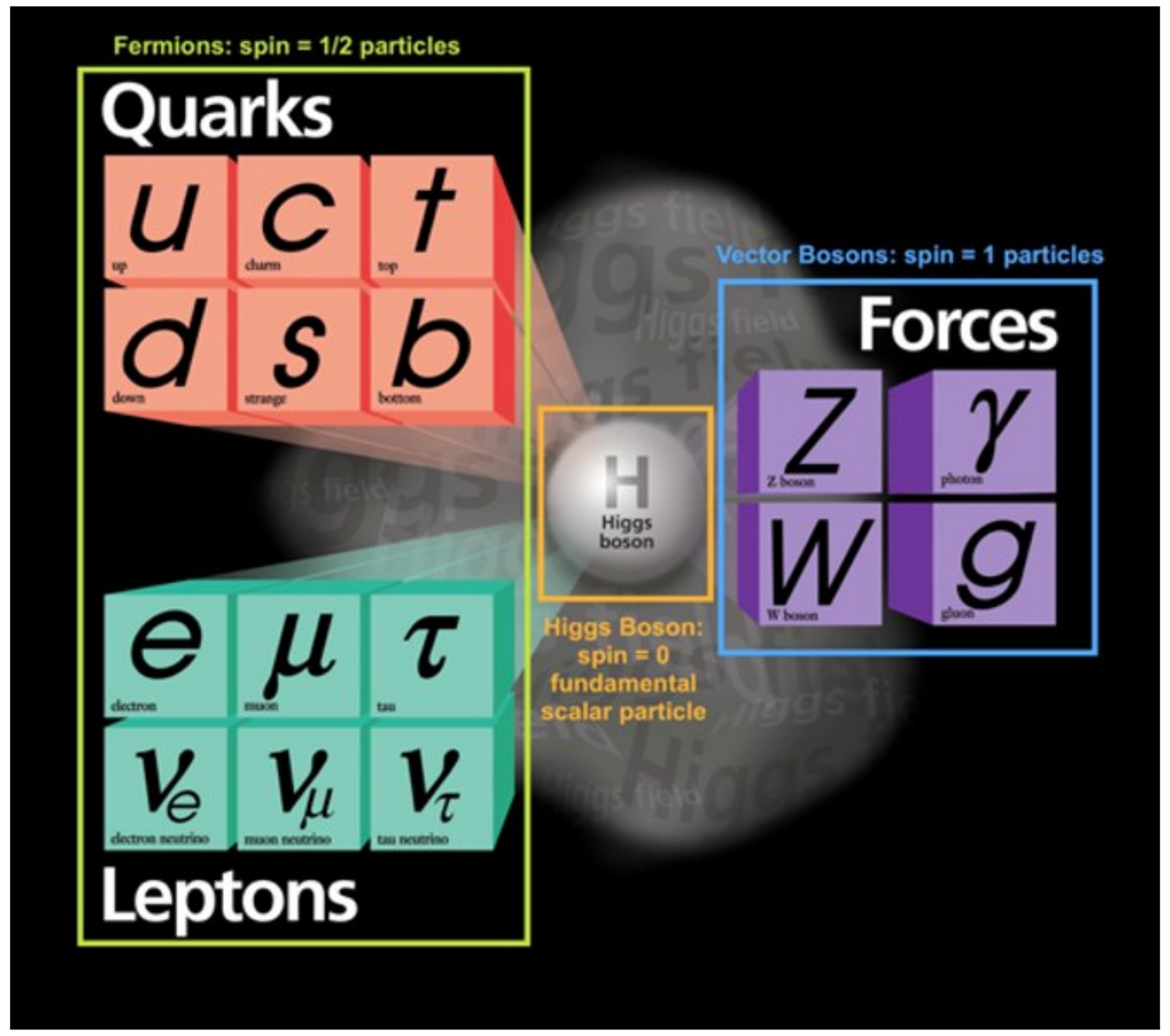




PARTICLE PHYSICS PHENOMENOLOGY AT THE LHC

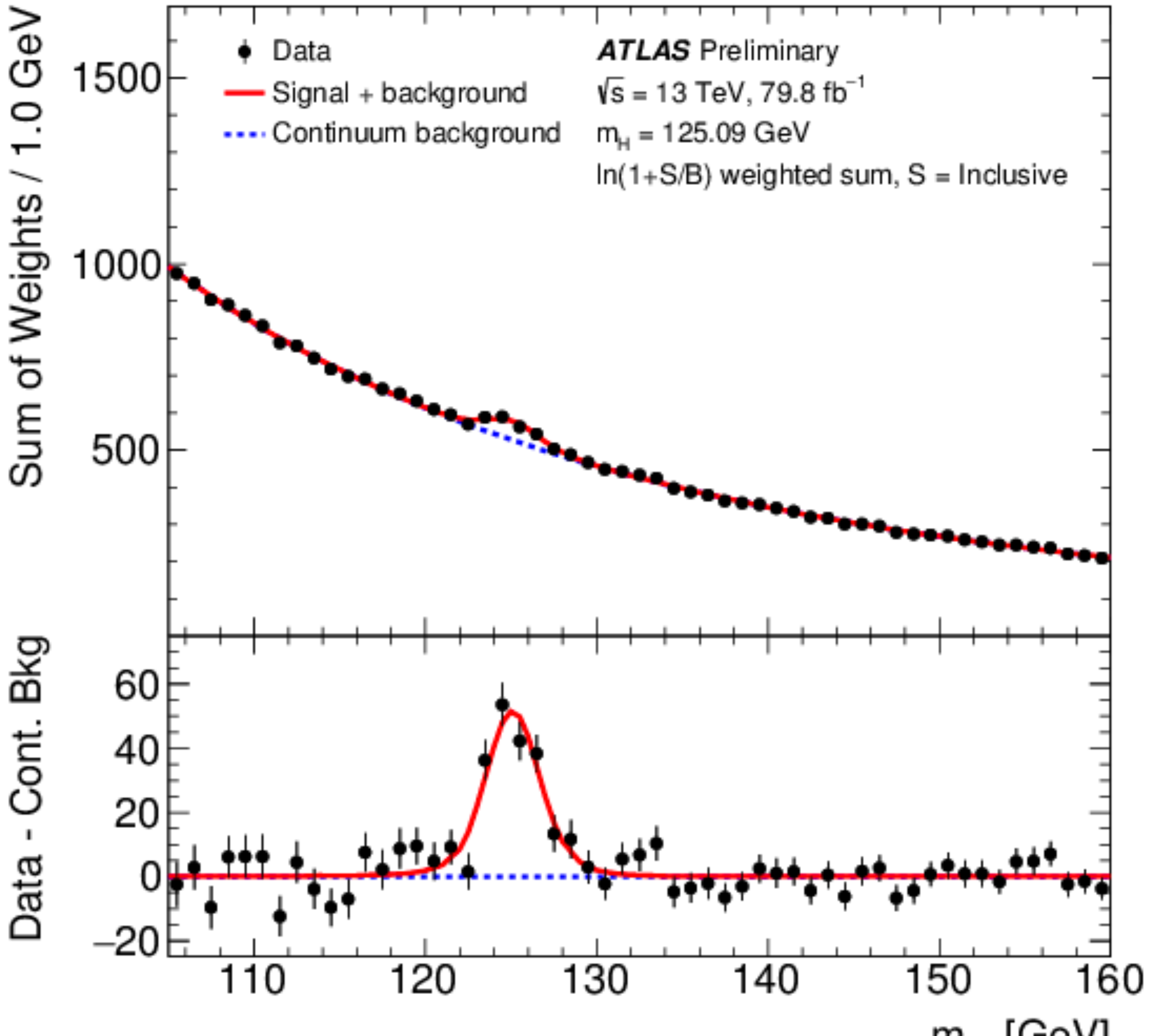
Experimental observations ⟷ HEP - PH ⟷ Theoretical model ⟷ HEP - TH ⟷ Ultimate Theory including gravity





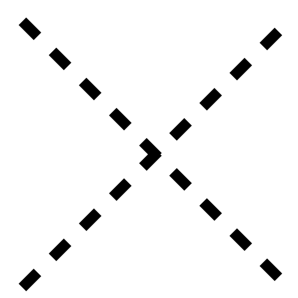
Most examples taken from the research
I am most familiar with, apologies
for not citing your favourite paper

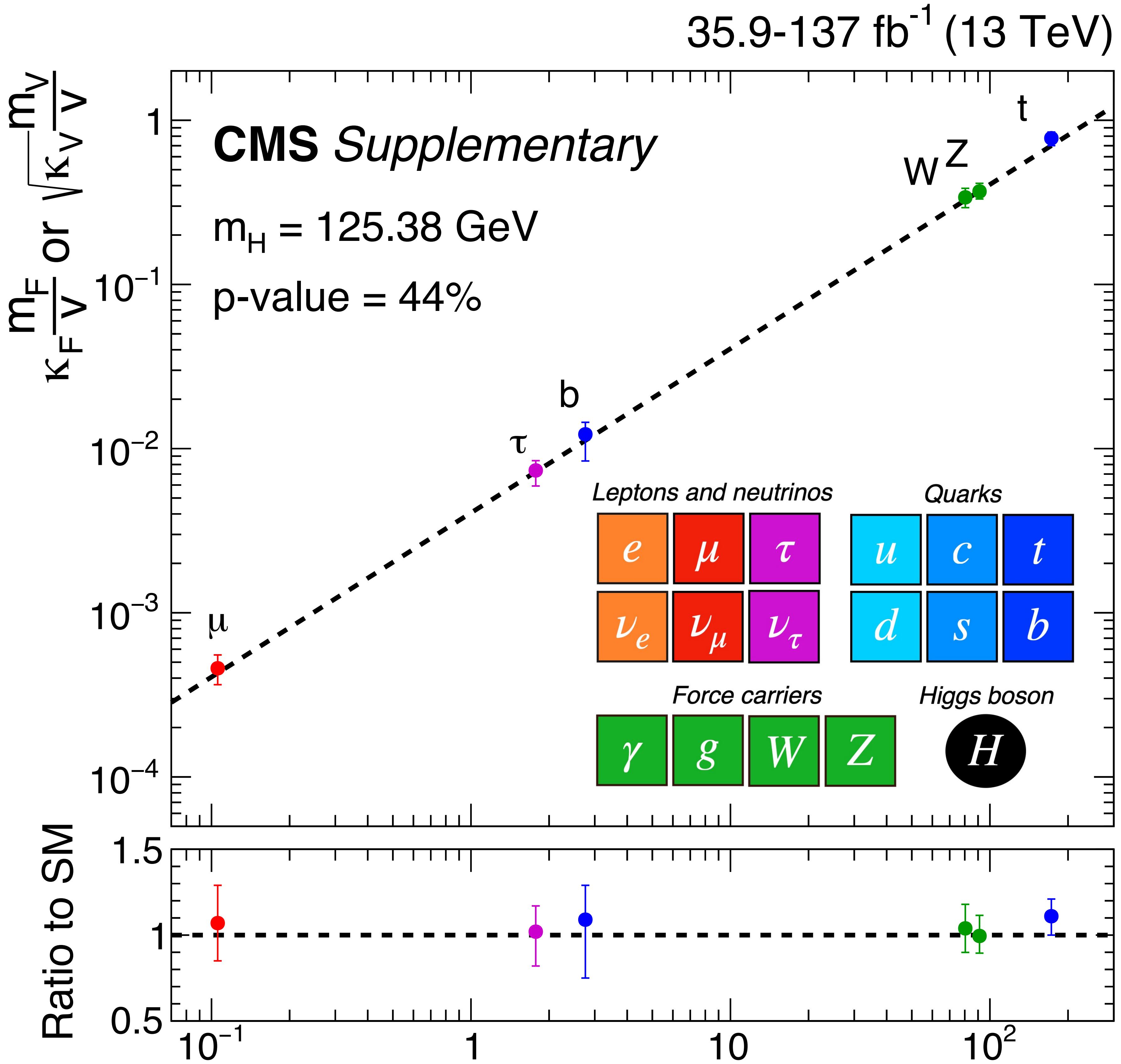
EXPLORING THE HIGH ENERGY FRONTIER



ATLAS 1901.5527

10-50% error &
H self-coupling
unknown

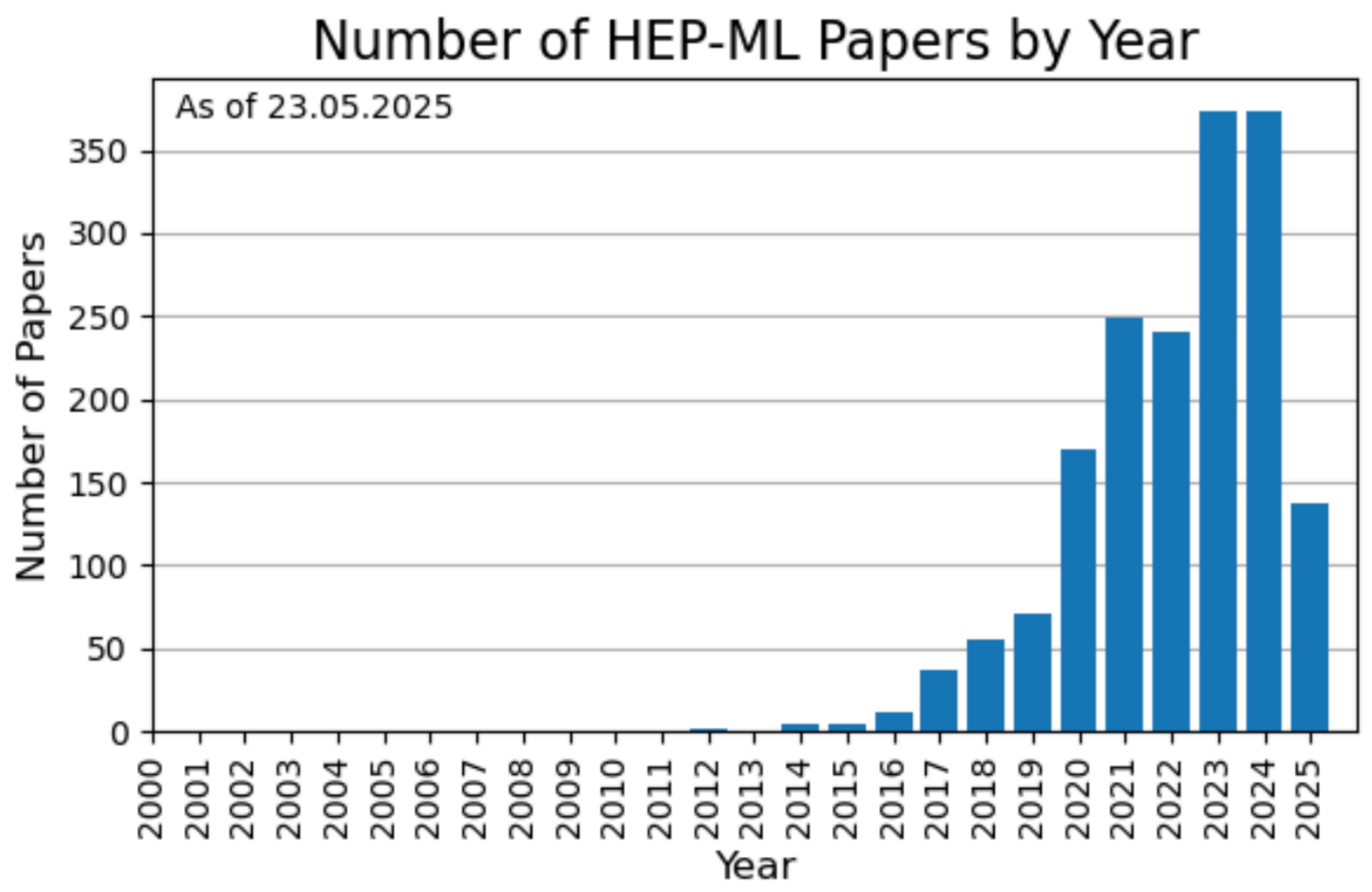




CMS-HIG-19-006

- The discovery of the Higgs boson at the LHC in 2012 opened a new era of exploration of the high energy frontier at particle colliders with new Yukawa force still largely unknown and many puzzles that the SM cannot explain (dark matter, neutrino mass, matter-antimatter asymmetry, gravity…)
- LHC rapidly evolving to a precision machine, capable of measuring small deviations over SM predictions
- Extremely **precise** and **highly correlated data** and **complex theoretical predictions**

EXPLORING THE HIGH ENERGY FRONTIER



A Living Review of Machine Learning for Particle Physics

https://iml-wg.github.io/HEPML-LivingReview/

Modern machine learning is driving a paradigm shift in all aspects of theoretical predictions and phenomenology analyses.

- The discovery of the Higgs boson at the LHC in 2012 opened a new era of exploration of the high energy frontier at particle colliders with new Yukawa force still largely unknown and many puzzles that the SM cannot explain (dark matter, neutrino mass, matter-antimatter asymmetry, gravity…)
- LHC rapidly evolving to a precision machine, capable of measuring small deviations over SM predictions
- Extremely **precise** and **highly correlated data** and **complex theoretical predictions**

OUTLINE

Intro & Part I & Part II

- Theory predictions at the LHC
- Application of ML to parton level predictions
- Beyond parton level predictions

MACHINE LEARNING

PARTICLE PHYSICS PHENOMENOLOGY AND THEORY

- Uncertainty quantification
- Symmetries
- Interpretability

Part III

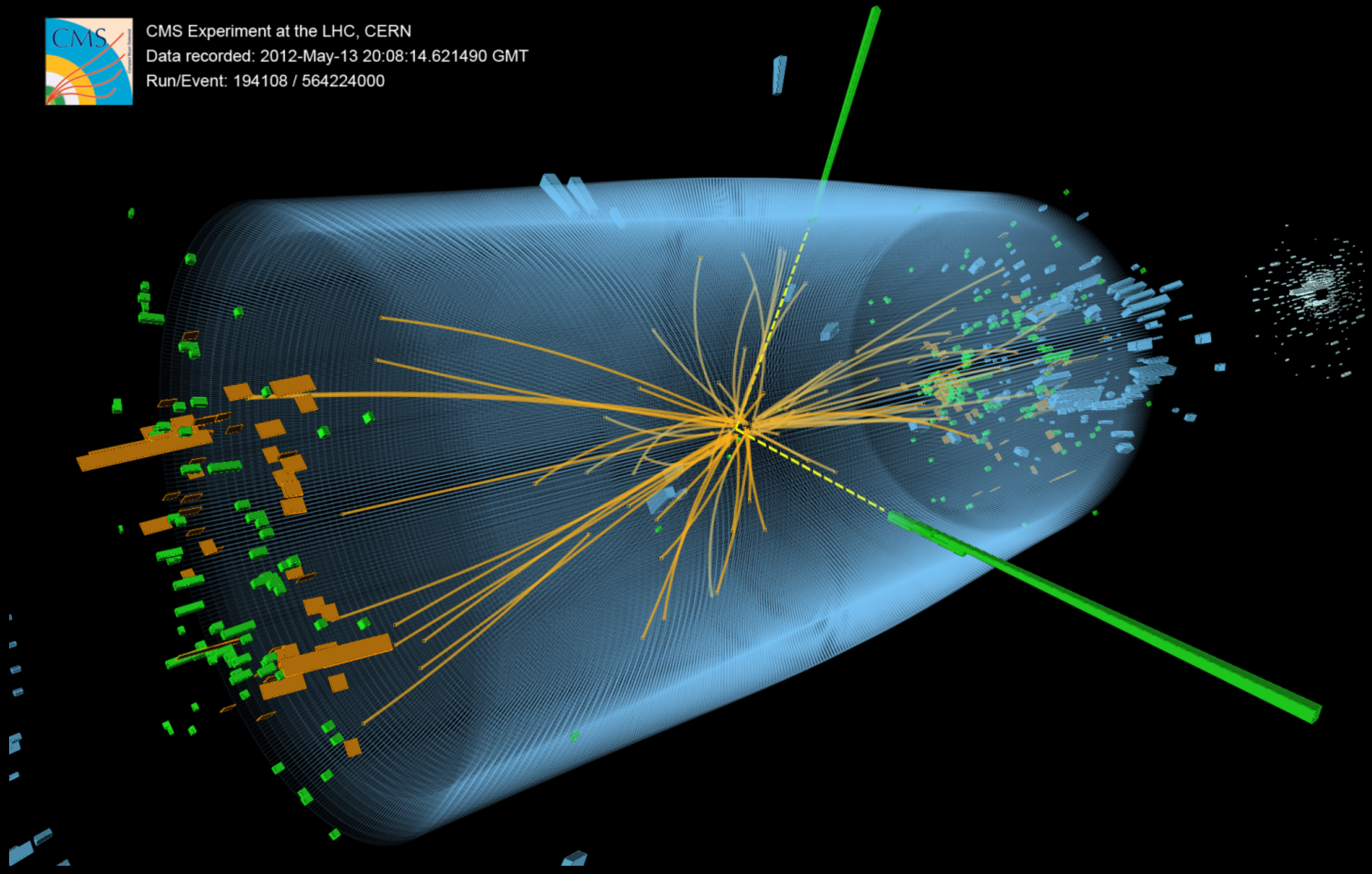
Intro

- Theory predictions at the LHC

MACHINE LEARNING

PARTICLE PHYSICS PHENOMENOLOGY AND THEORY

EXPERIMENTAL DATA AT THE LHC



- Huge volume of data
- High-level and low-level triggers
- Increased luminosity goes with large pile-up

- Must process the events to give theorists useful, robust and least possible biassed information
- Must make data reusable for reinterpretation

Jennifer Ngadiuba's summary talk

THEORY PREDICTIONS AT THE LHC

- All we want is the probability of a given theory **T** given some data **D**
- In Bayesian terms

$$P(T|D) \approx P(\vec{\alpha}|D) = \frac{P(D|\vec{\alpha})\, P(\vec{\alpha})}{P(D)}$$

Likelihood

Prior

Evidence

THEORY PREDICTIONS AT THE LHC

- All we want is the probability of a given theory **T** given some data **D**
- In Bayesian terms

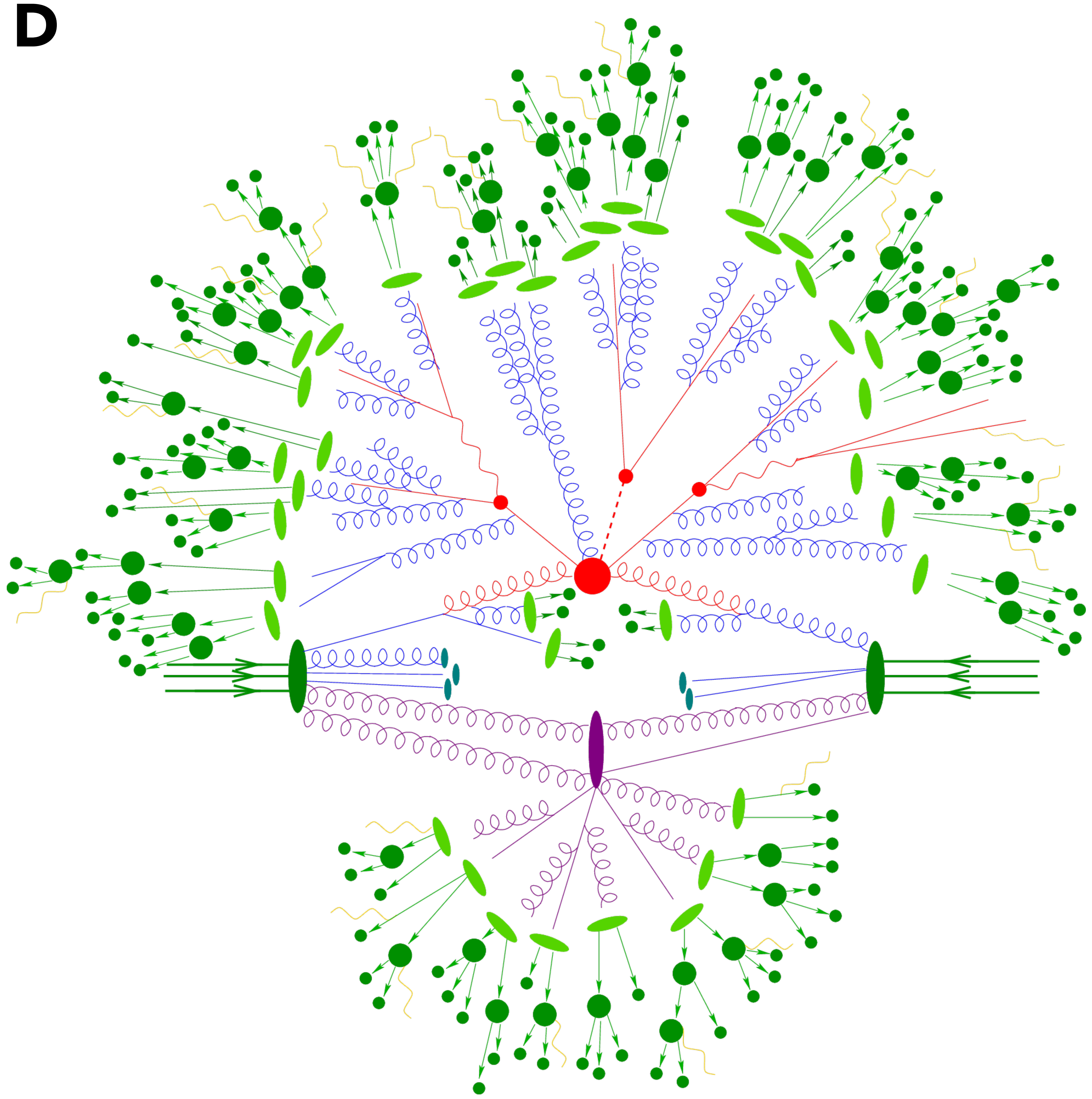
$$P(T|D) \approx P(\vec{\alpha}|D) = \frac{P(D|\vec{\alpha})\, P(\vec{\alpha})}{P(D)}$$

Likelihood

Prior

Evidence

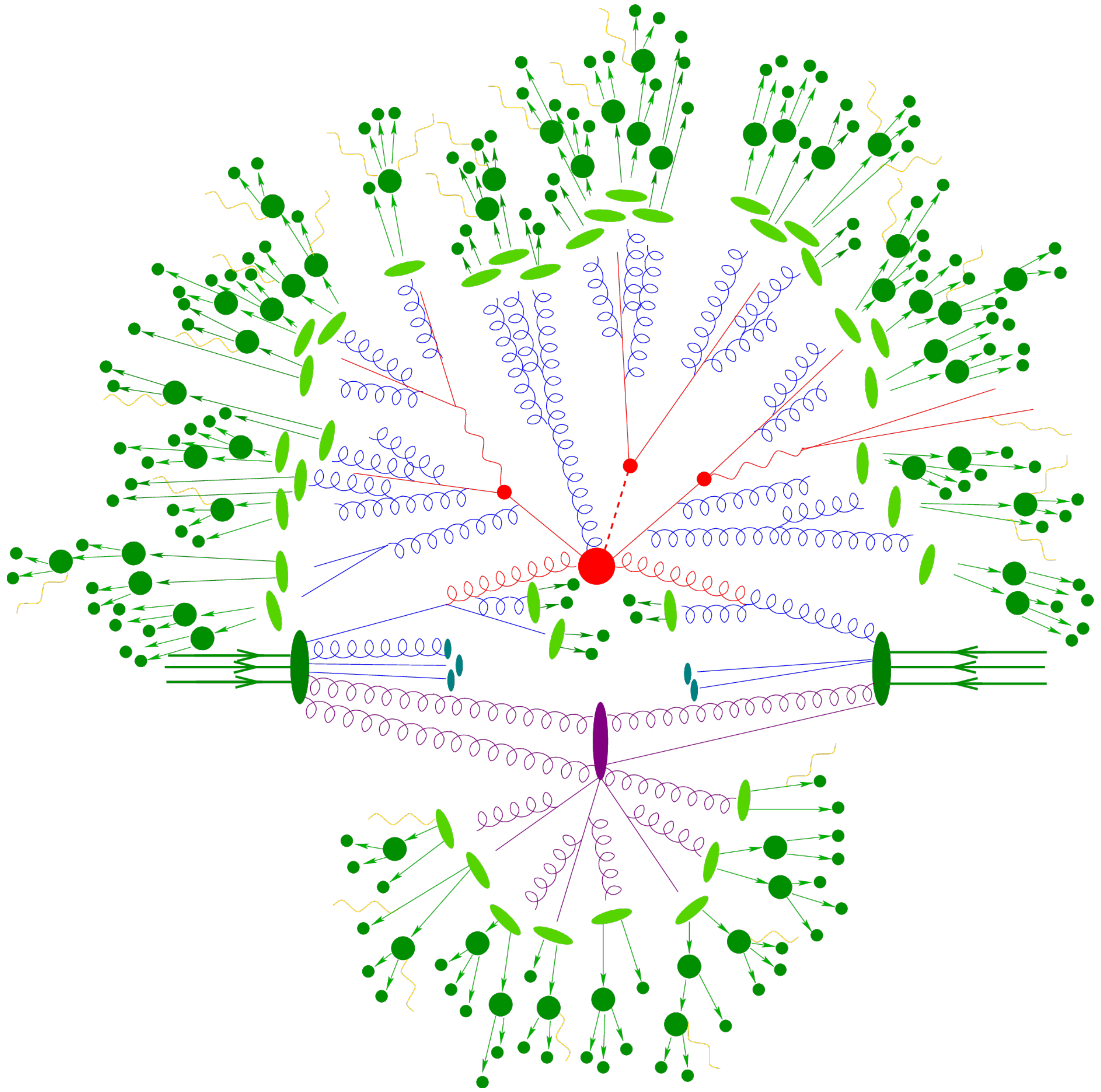
- Consider theoretical predictions at the LHC
- The likelihood itself at the LHC is extremely complicated



Divide et impera

THEORY PREDICTIONS AT THE LHC

Quantum theory
Lagrangian
$\mathcal{L}$ (SM, SM + EFT, BSM)



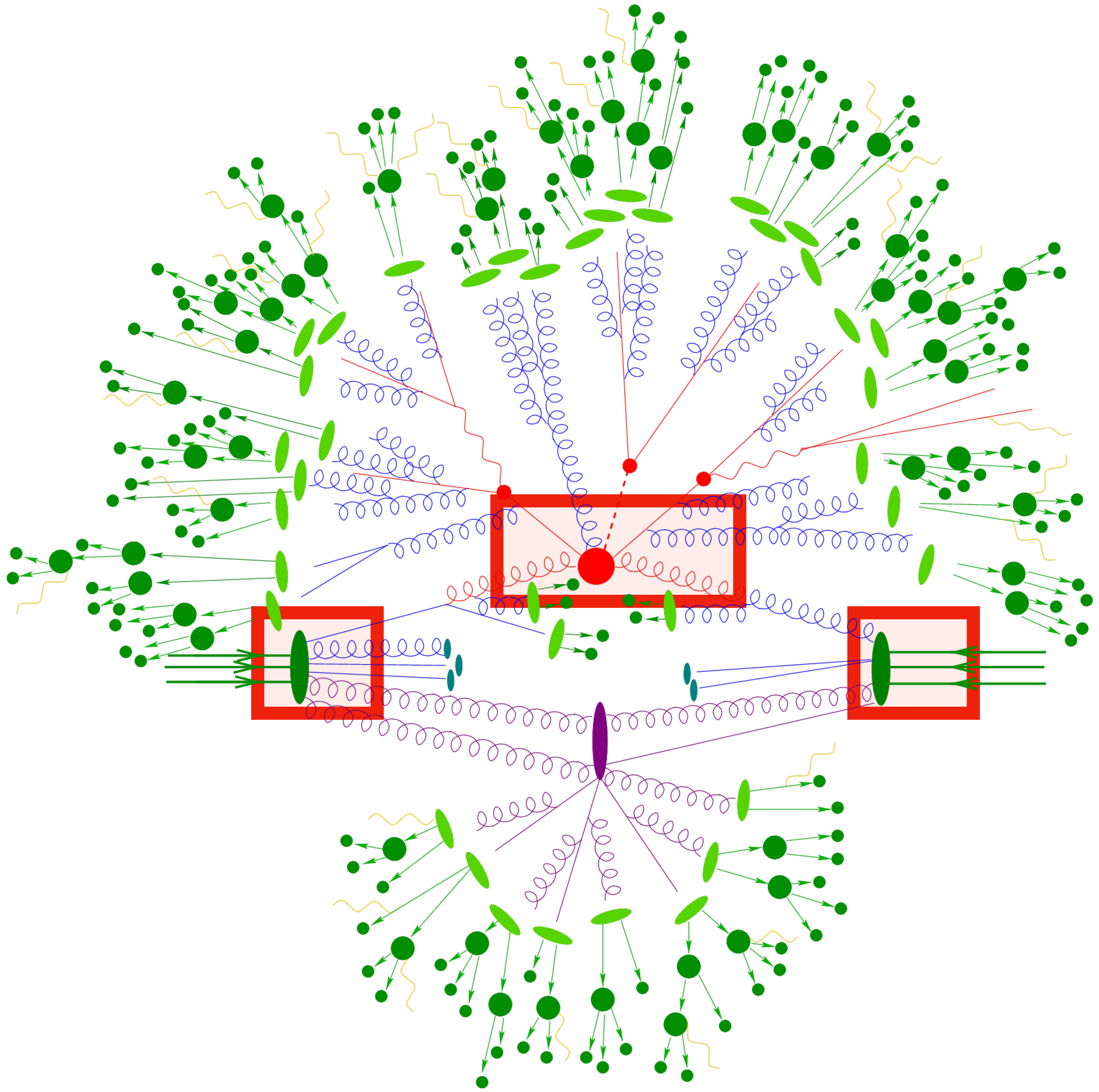
THEORY PREDICTIONS AT THE LHC

Quantum theory
Lagrangian
$\mathcal{L}$ (SM, SM + EFT, BSM)

$\Downarrow$

Scattering amplitudes | Phase space integration | PDFs (proton structure)

Parton level



THEORY PREDICTIONS AT THE LHC

Quantum theory
Lagrangian
$\mathcal{L}$ (SM, SM + EFT, BSM)

Scattering amplitudes

Phase space integration

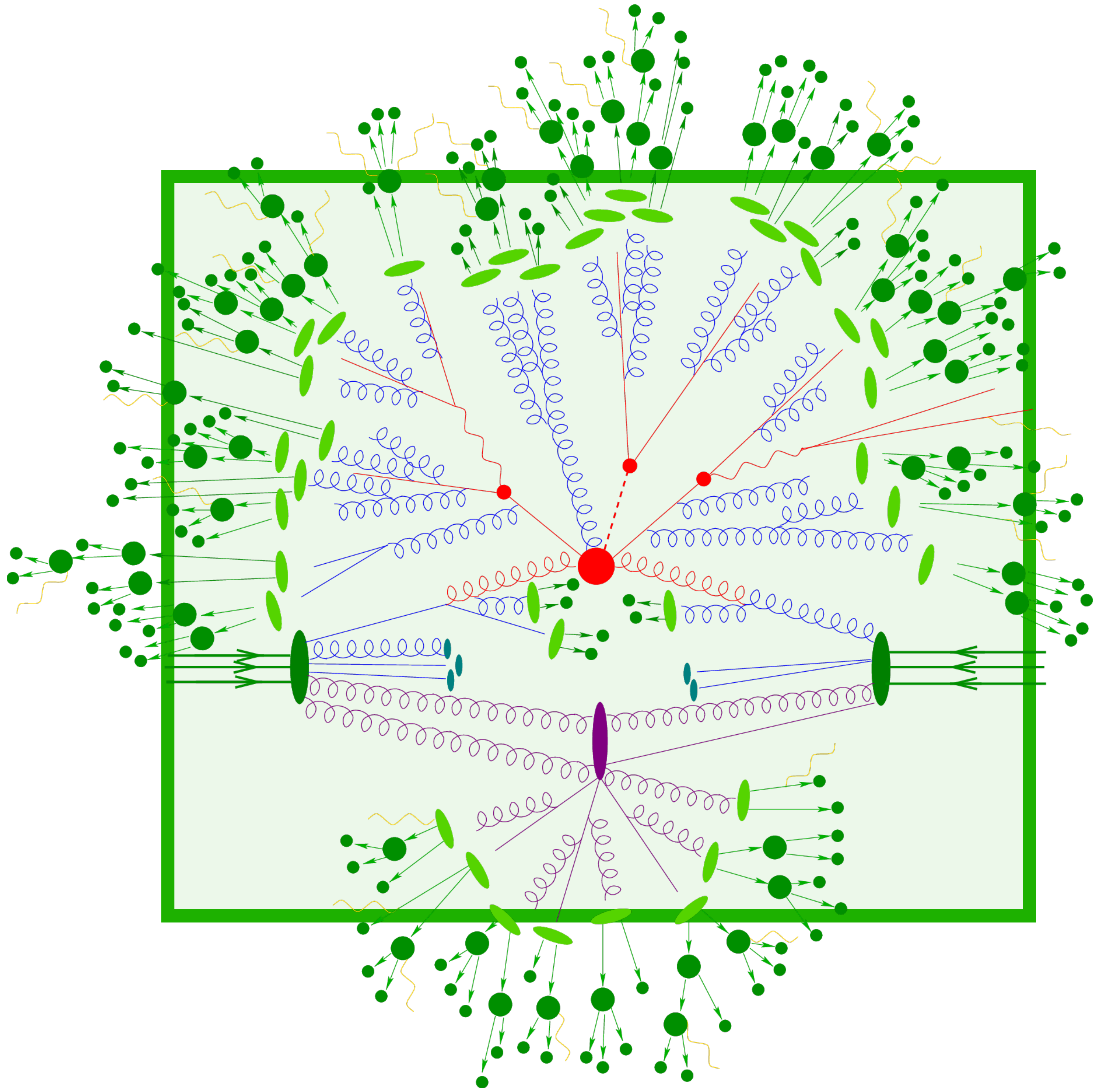
PDFs (proton structure)

Parton level

QCD and QED radiation (Parton Showers)

Fragmentation, Hadronisation, Jets

Particle level



THEORY PREDICTIONS AT THE LHC

Quantum theory
Lagrangian
$\mathcal{L}$ (SM, SM + EFT, BSM)

↓

Scattering amplitudes | Phase space integration | PDFs (proton structure)

Parton level

↓

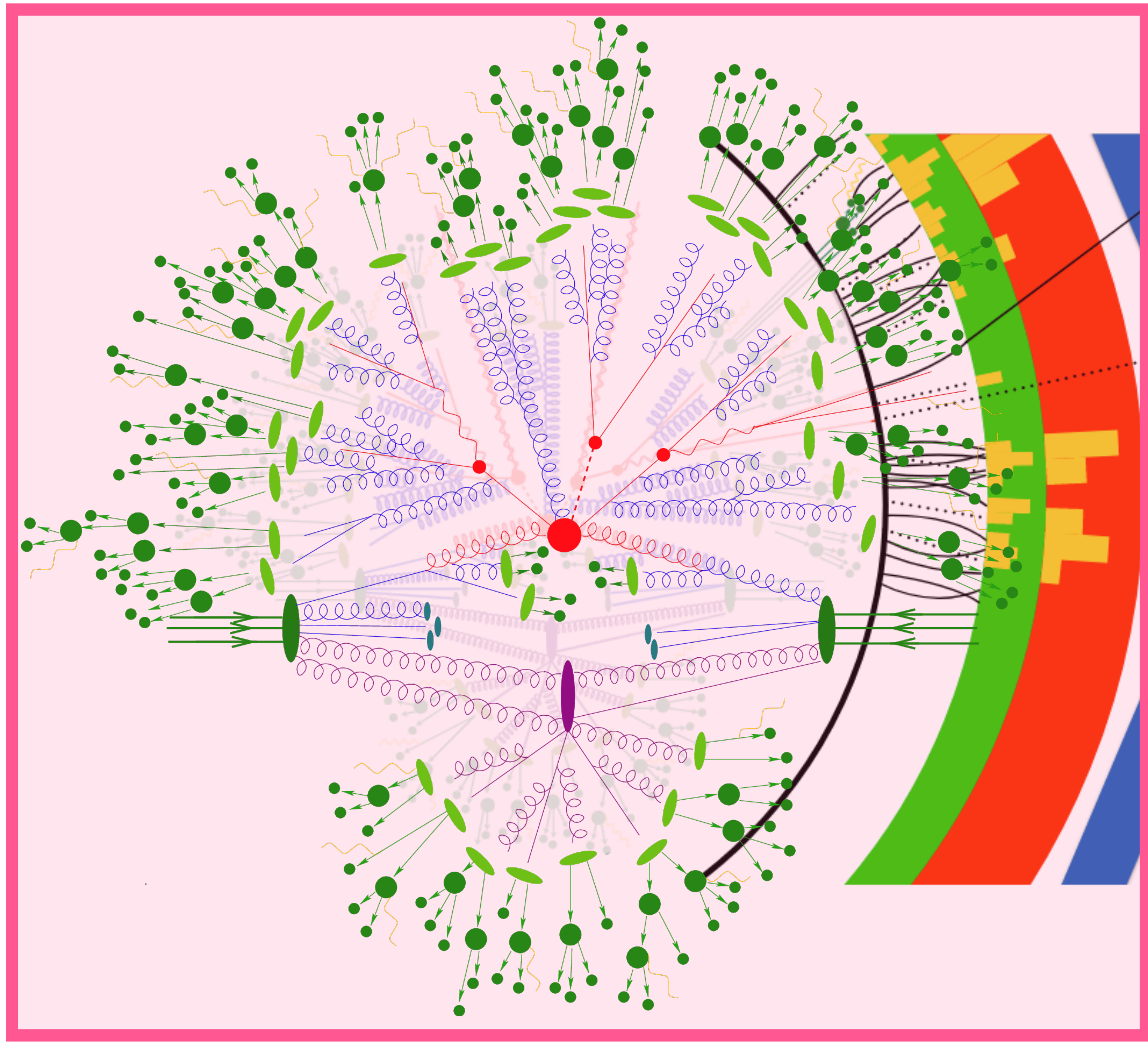
QCD and QED radiation (Parton Showers) | Fragmentation, Hadronisation, Jets

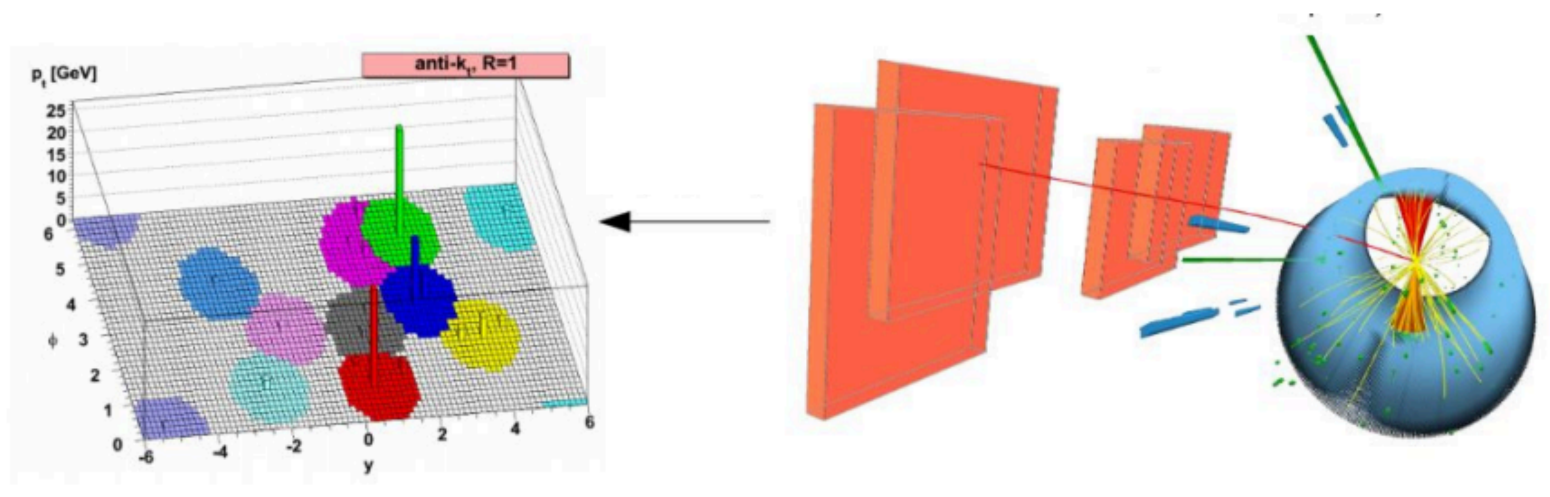
Particle level

↓

Detector simulation | Event reconstruction and selection

Detector level





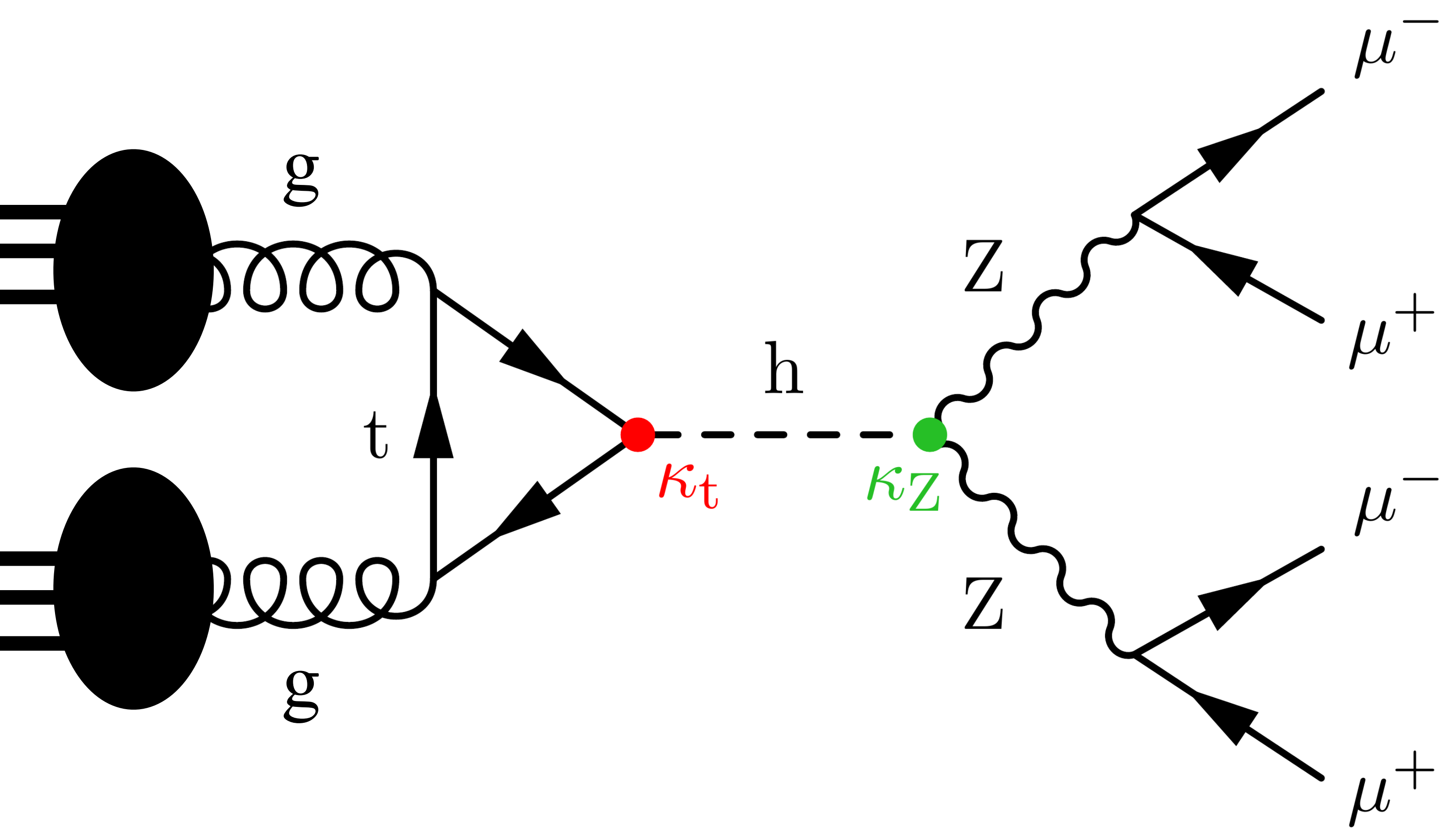
Part I

MACHINE LEARNING

- Application of ML to parton level predictions
  (1) matrix elements
  (2) phase space
  (3) parton distribution functions
  (4) parameter determination

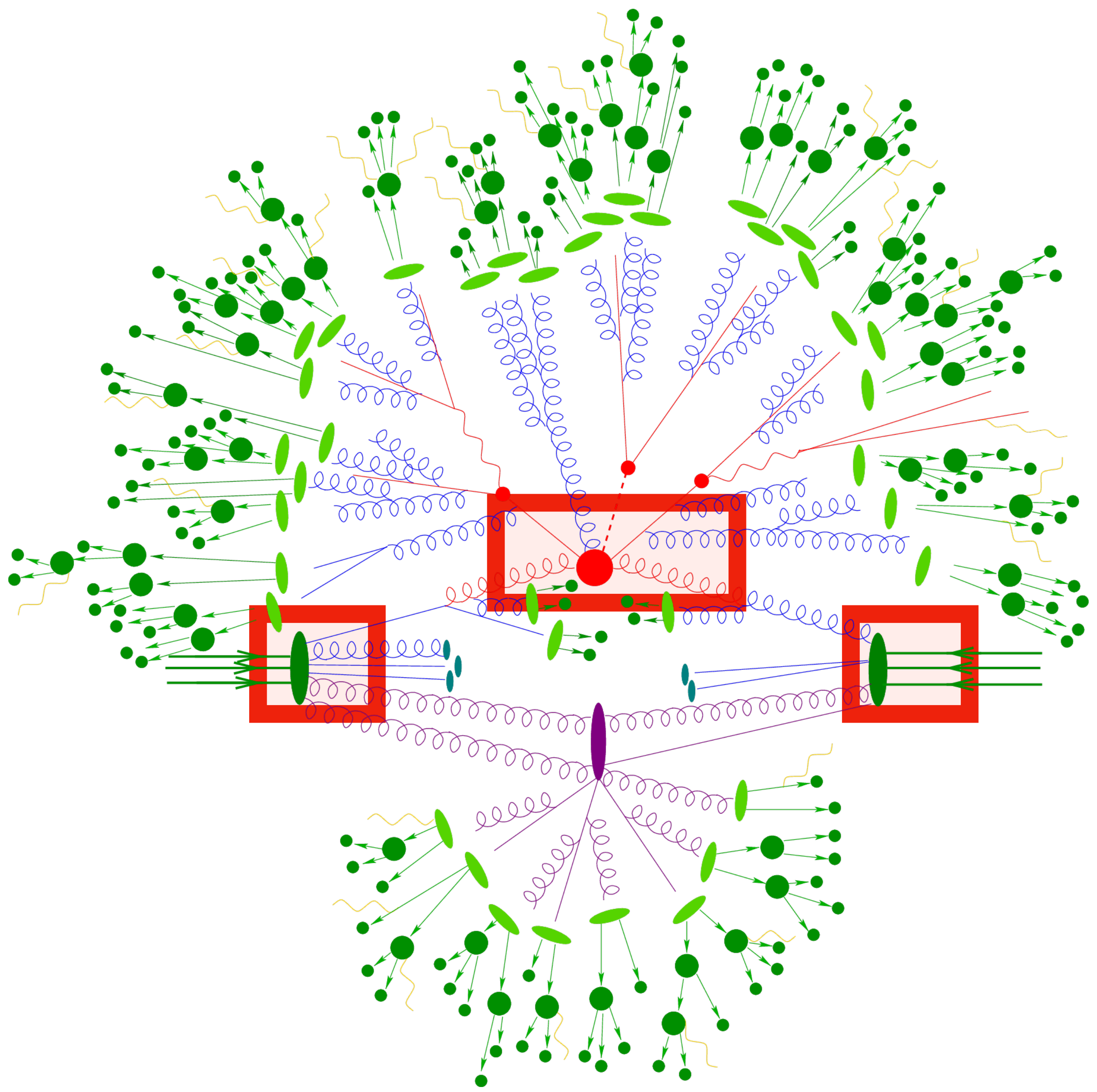
PARTICLE PHYSICS PHENOMENOLOGY AND THEORY

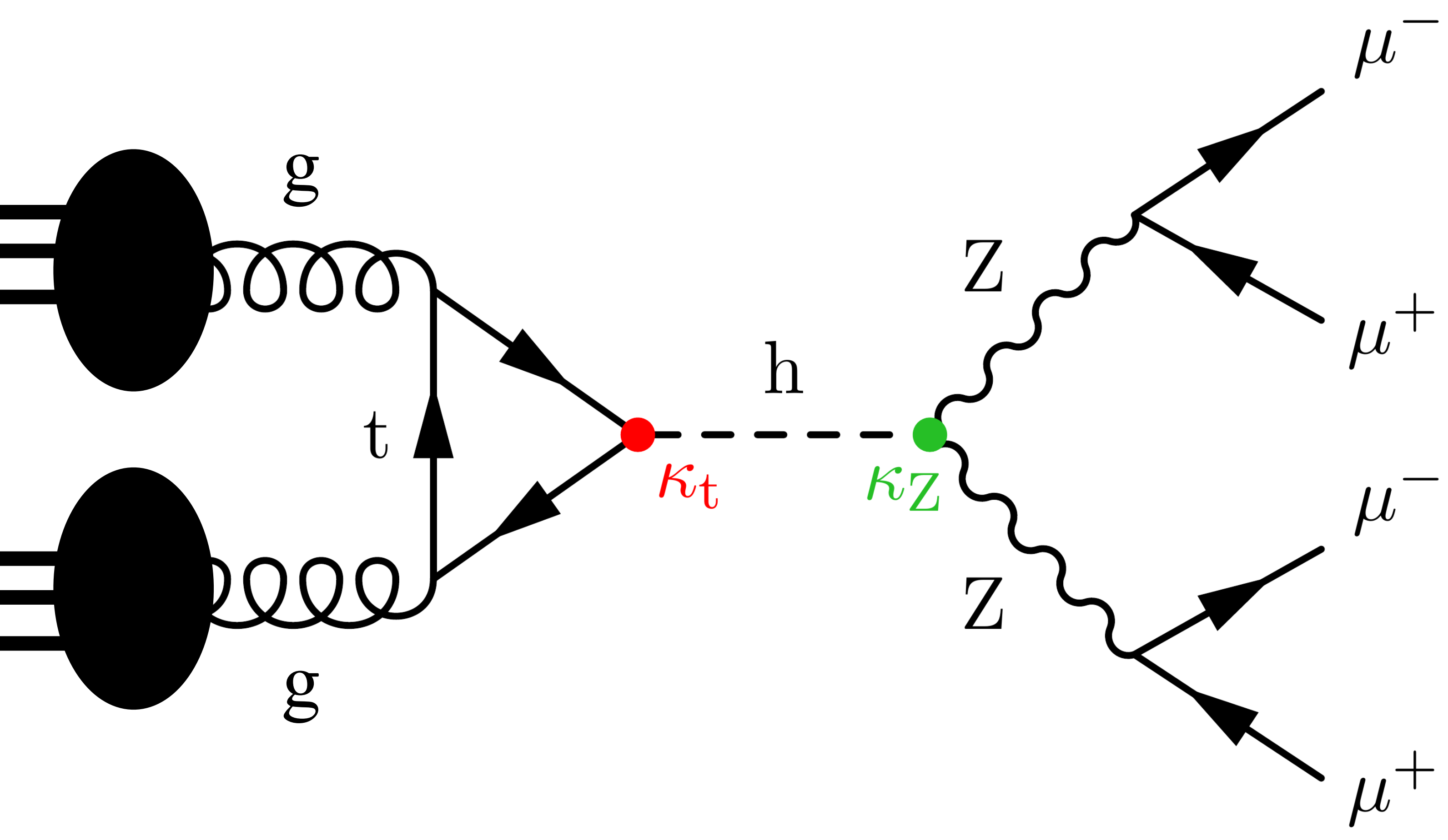
PARTON LEVEL PREDICTIONS



(4) SM & BSM parameters

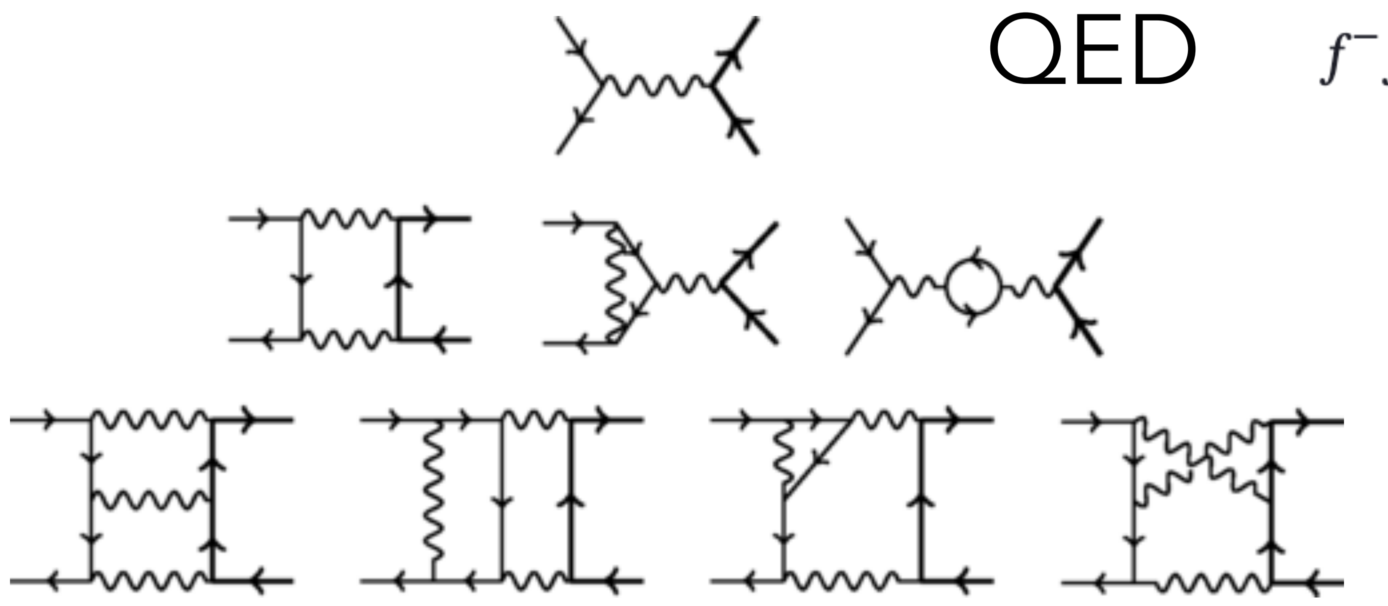
$$d\sigma \approx \underbrace{\left|\mathcal{A}_{ij\to 1\ldots n}(x)\right|^2}_{\text{(1) Scattering amplitude}} \times \underbrace{d\Phi_n}_{\text{(2) n-particles phase space}} \times \underbrace{f_{p/i} \times f_{p/j}}_{\text{(3) Parton Distribution Functions}}$$





$$d\sigma \approx \underbrace{\left|\mathcal{A}_{ij\to 1\ldots n}(x)\right|}^2 \times d\Phi_n \times f_{p/i} \times f_{p/j}$$

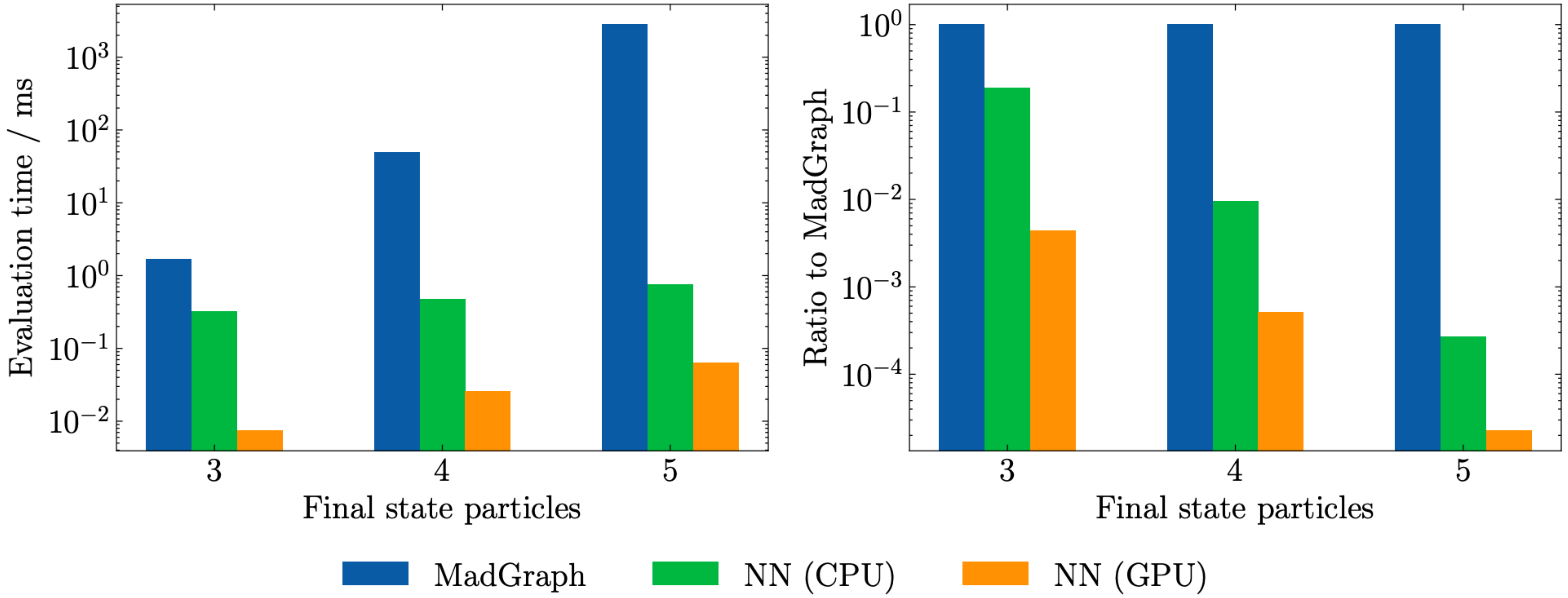
(1) ML & AMPLITUDES

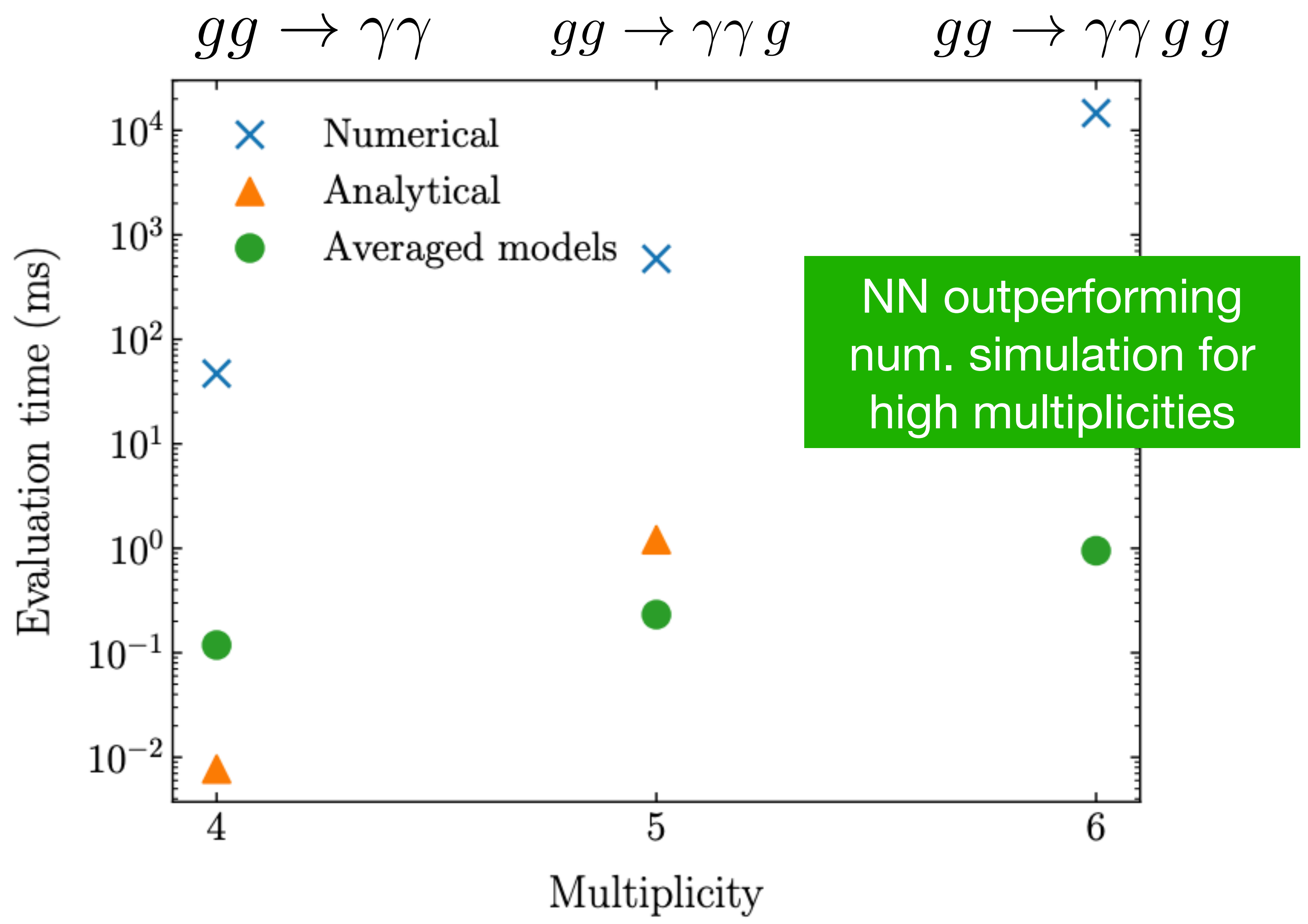


QED $f^- f^+ \to F^- F^+$

$$|\mathcal{A}_{ij\to 1\ldots n}(x)|^2_{\mathrm{true}} \approx |\mathcal{A}_{ij\to 1\ldots n}(x)|^2_{\mathrm{NN}}$$

- **Regression problem =** Exploit NN flexibility to speed up computation of amplitudes: train ML regressors (NNs, or ensemble of NNs) with pre-computed “true” (slow) amplitudes for phase space points x, and use them to predict same amplitudes accurately and fast [arXiv:1912.11055, 2002.07516, 2006.16273, 2008.10949, 2105.04898, 2106.09474, 2109.11964, 2206.08901,2302.04005,2301.13562, 2306.07226 ...]





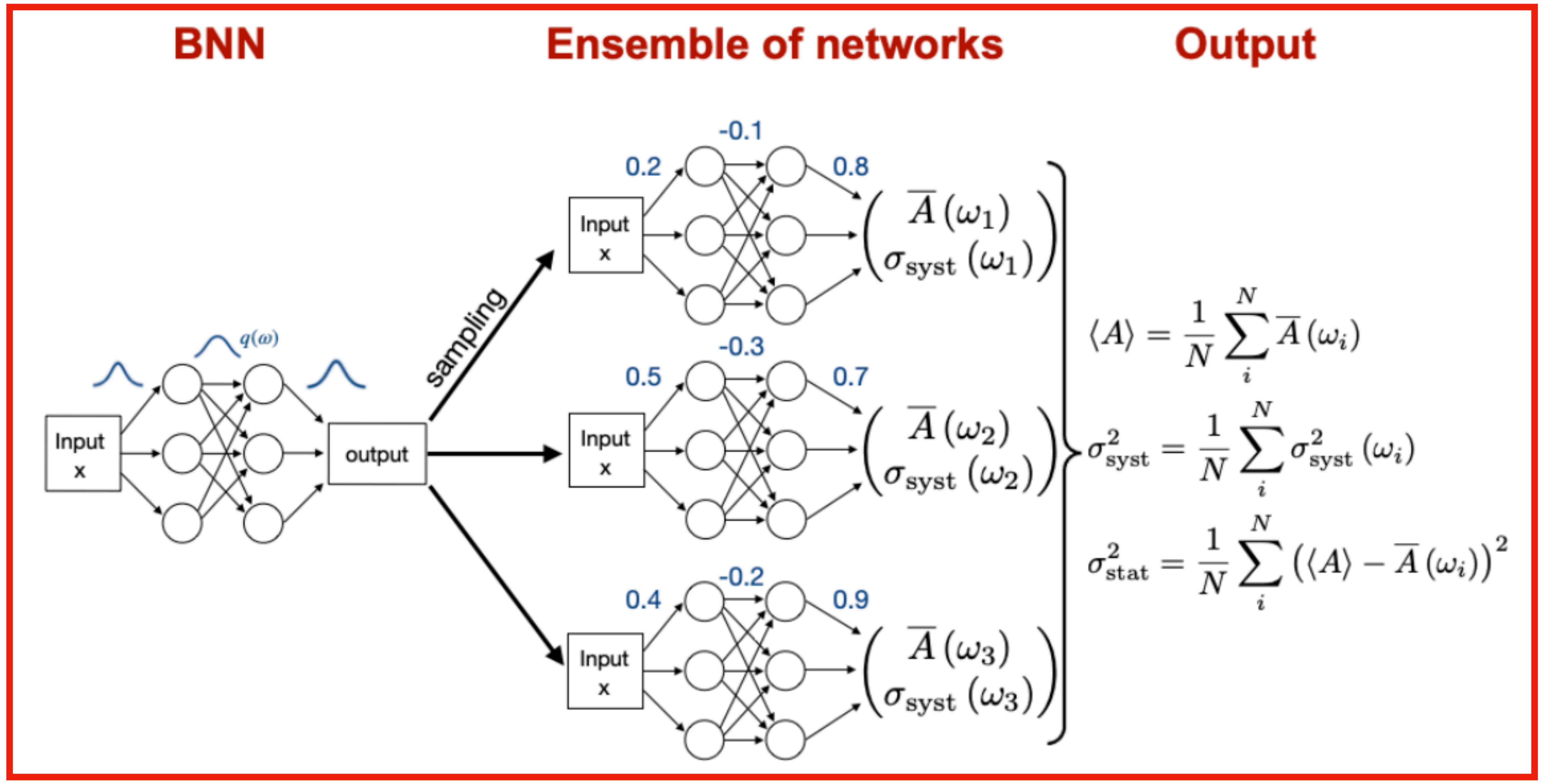
Maitre, Troung arXiv: 2302.04005 $e^+e^- \to Z/\gamma^* \to q\bar{q} + n_g g$

Aylett-Bullock et al, arXiv: 2106.09474

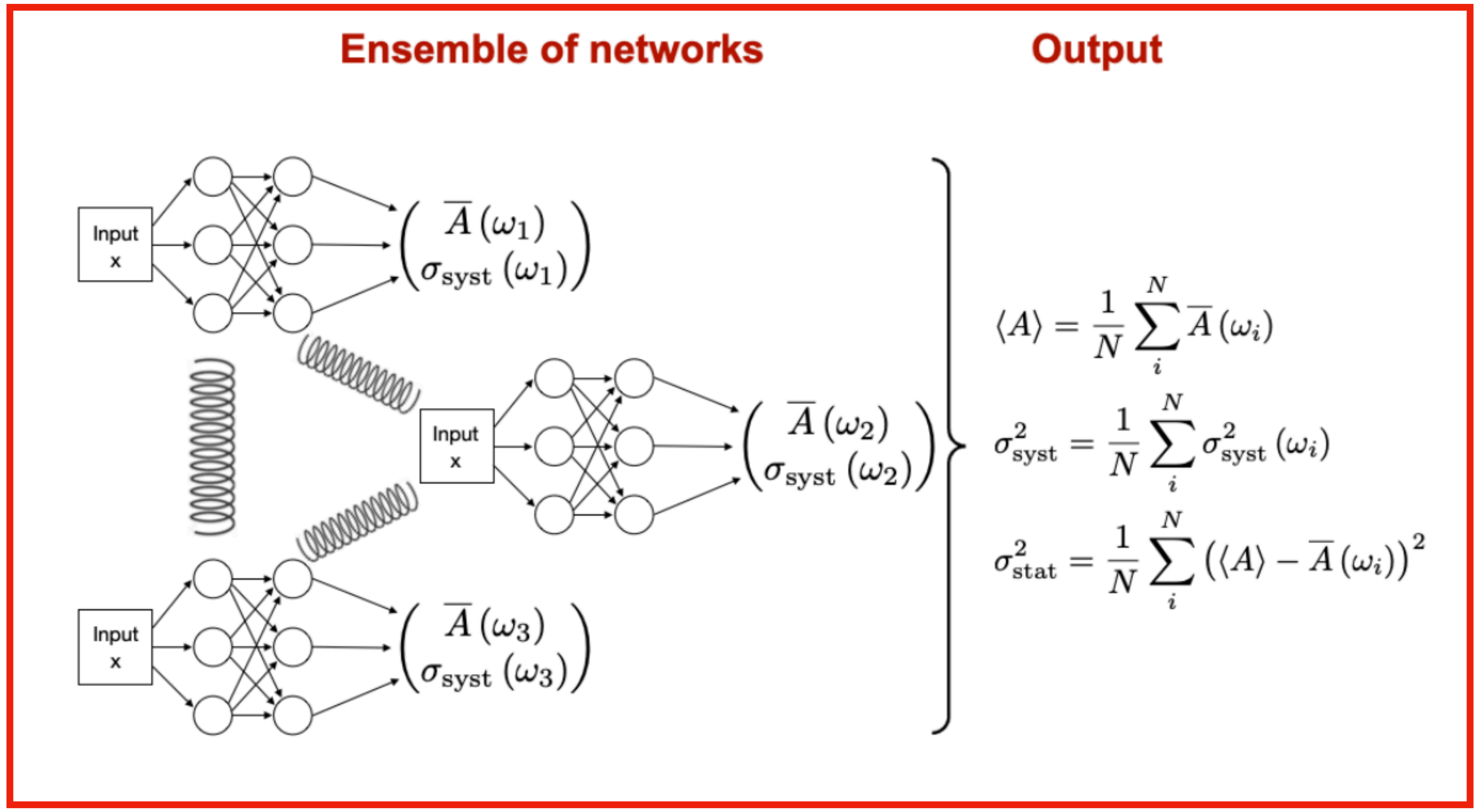
(1) ML & AMPLITUDES

- To generate higher-order scattering amplitudes, ML regressors (ML surrogates) must be **precise** enough to reflect the underlying theory precision.
- Reliable uncertainty estimate is key to use ML surrogates. How to quantify **uncertainty?**
- Heteroscedastic losses (including unknown uncertainty in the loss and learning it in deterministic NN), Bayesian NN (**BNNs**) and Repulsive Ensembles (**REs**) can be used to track systematic and statistical uncertainties and their performance benchmarked [arXiv:2206.14831, 2412.12069 …]

Bayesian NNs



Repulsive ensembles

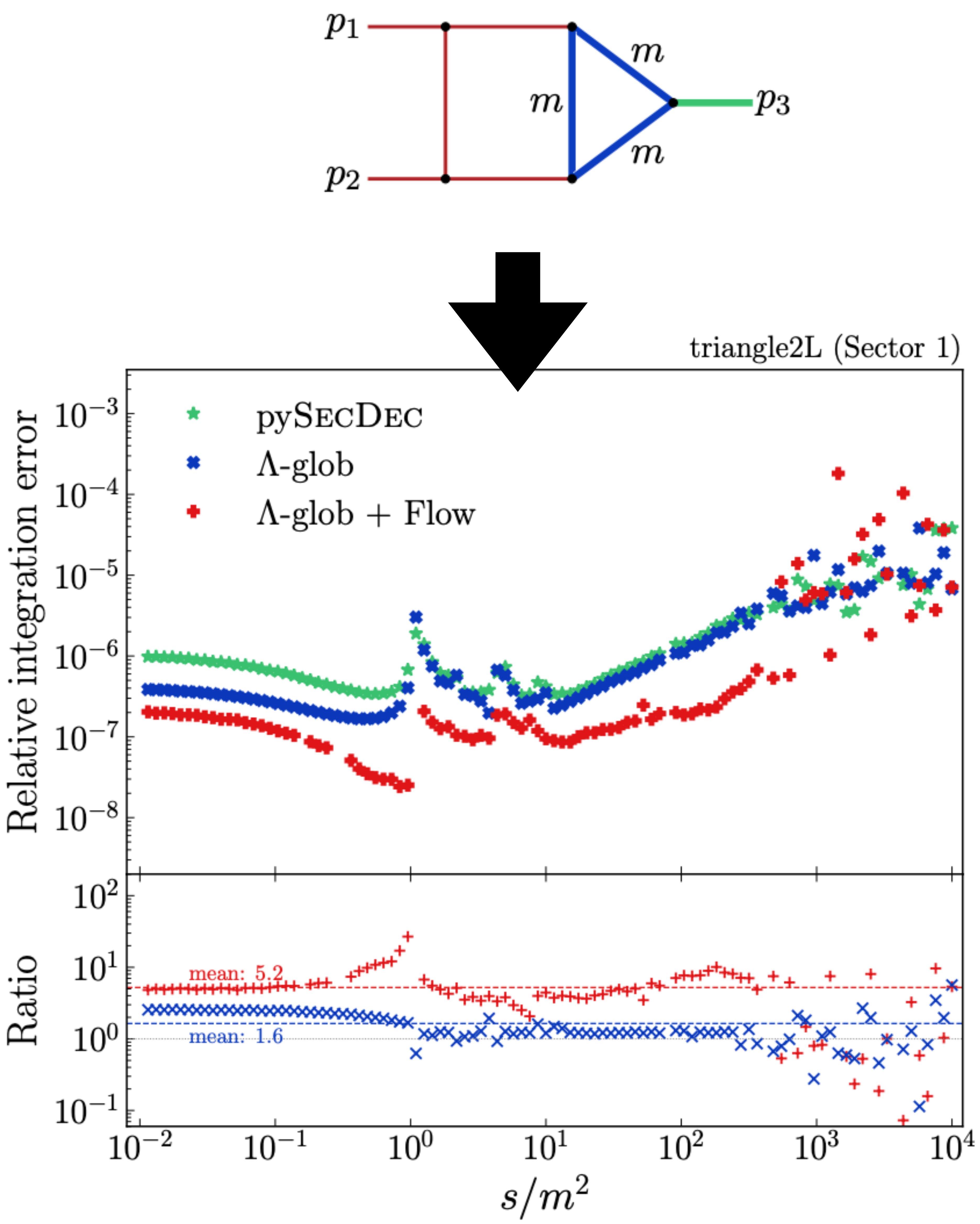


Elmer et al, arXiv: 2412.12069

(1) ML & AMPLITUDES

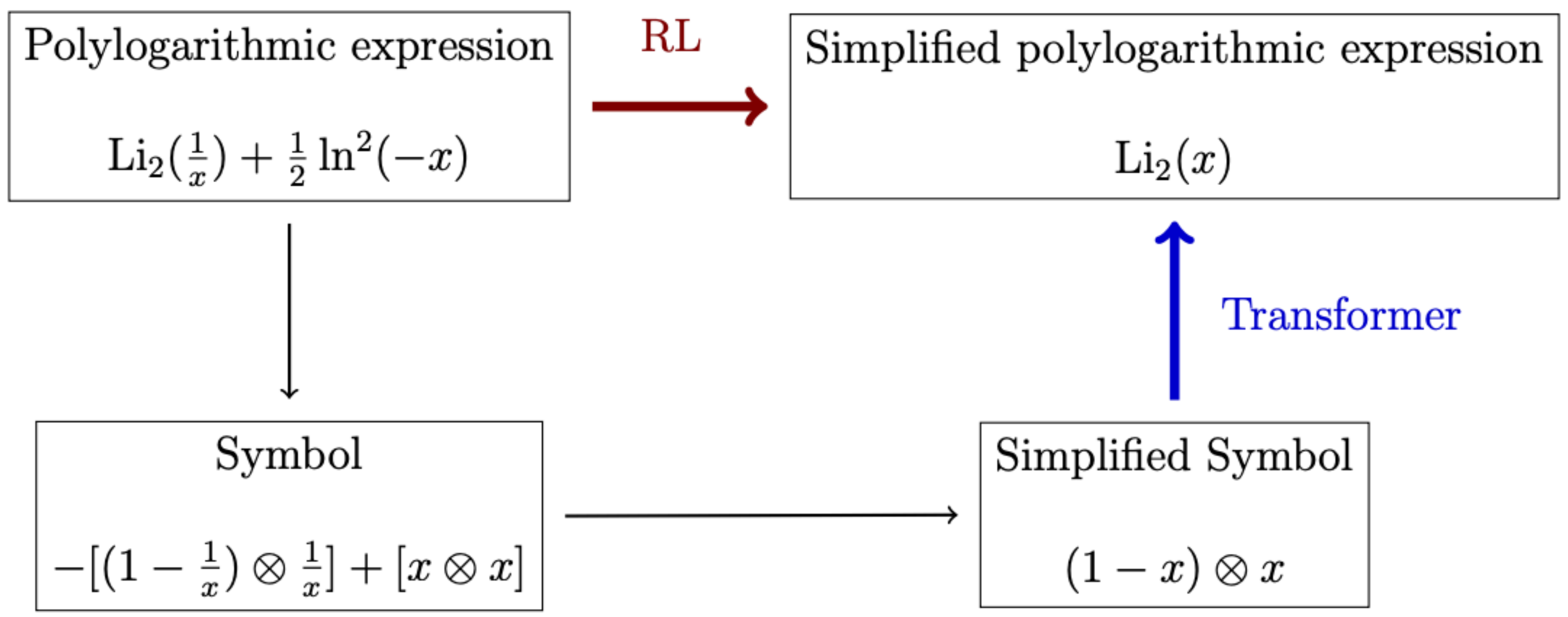
$$|\mathcal{A}_{ij\to 1\ldots n}(x)|^2_{\text{true}} \approx |\mathcal{A}_{ij\to 1\ldots n}(x)|^2_{\text{NN}}$$

- **Multi-loop integrals =** Amplitudes beyond leading order contain loop integrals, and ML can optimise the computation of integrals.
- Integrable singularities on real axis → contour deformation into complex plane.
- NN-assisted algorithms (based on normalising flows) offer great potential to amplify precision of standard contour deformation algorithms. [Winterhalder et al 2112.09145]
- See also [Calisto et al 2312.02067] and [Maitre et al 2211.02834] for alternative approach to compute Feynman integrals by solving numerically the differential equations satisfied by the Feynman integrals

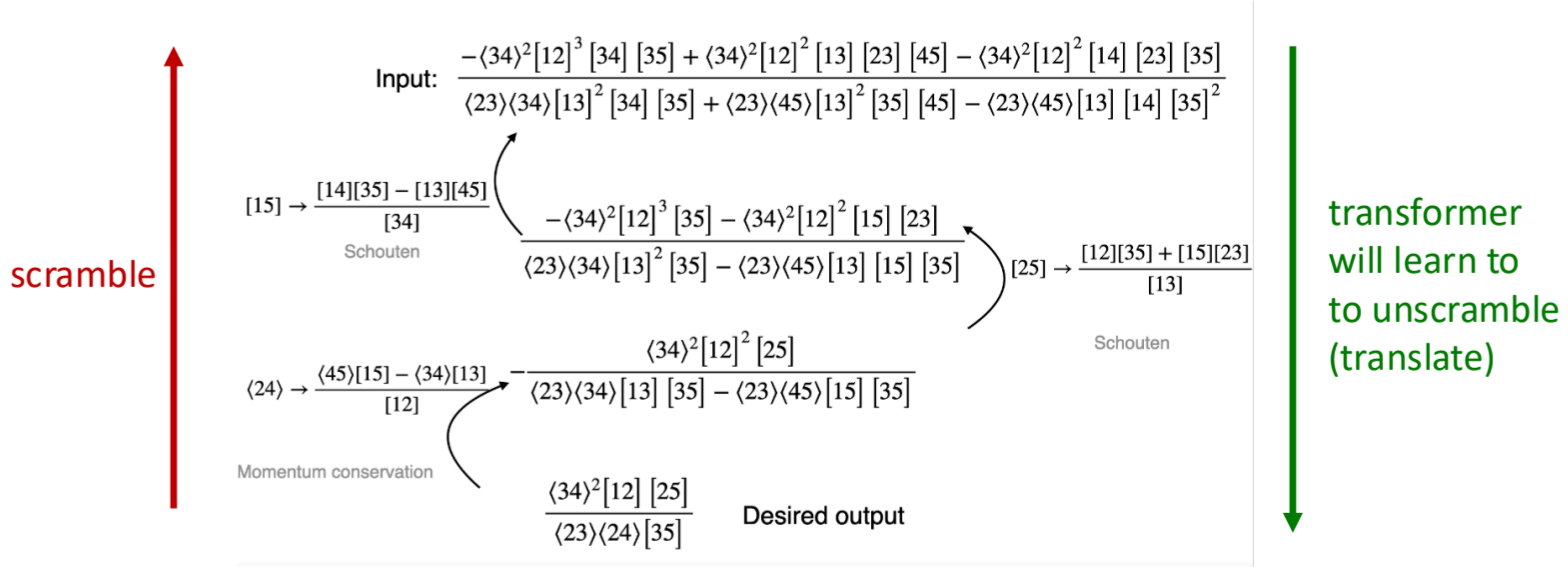


(1) ML & AMPLITUDES

- Integration of scattering amplitudes gives rise to generalised polylogarithms.
- No classical algorithms to **simplify expressions**. Machine learning huge potential [Dersy, Schwartz, Zhang 2206.04115]
  - ➡ Reinforcement Learning: apply known identities like moves in a game and learn them by training.
  - ➡ Transformer Networks: learn to guess the answer by translating from complicated to simple
- Similar techniques applied to simplify scattering amplitudes expressed in terms of spinor-helicity using transformers and contrastive learning [Cheung, Dersy, Schwartz 2408.04720]

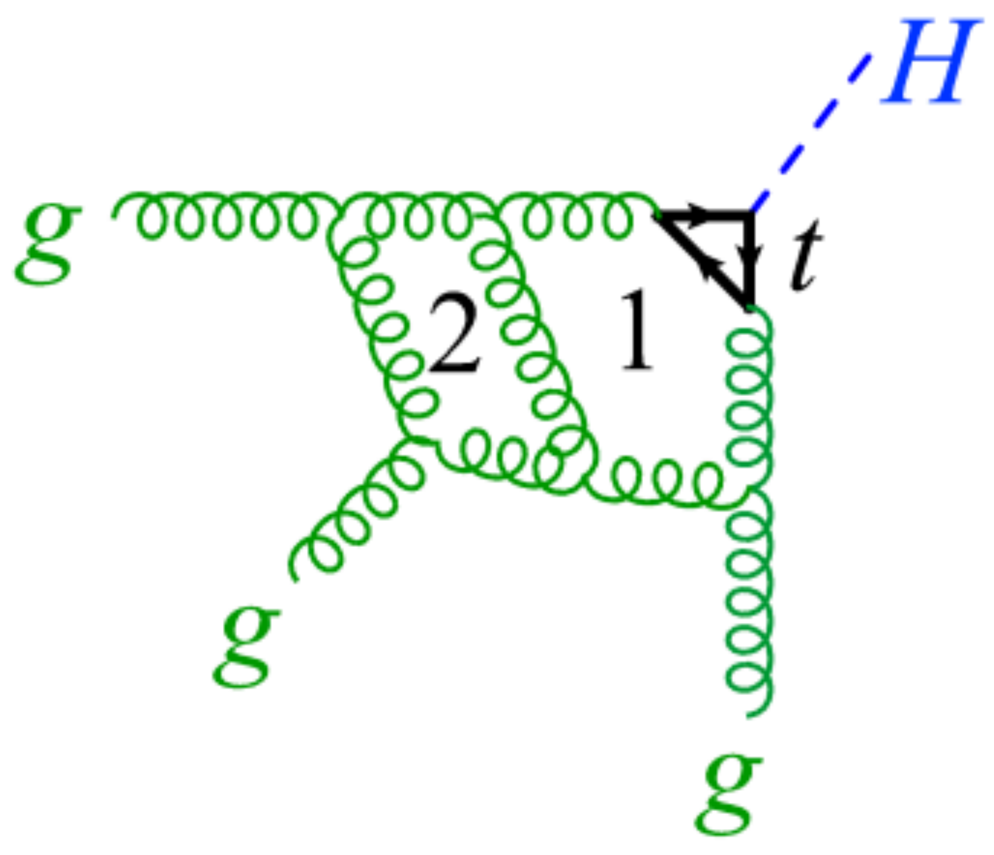


Transformer trained on data generated by scrambling

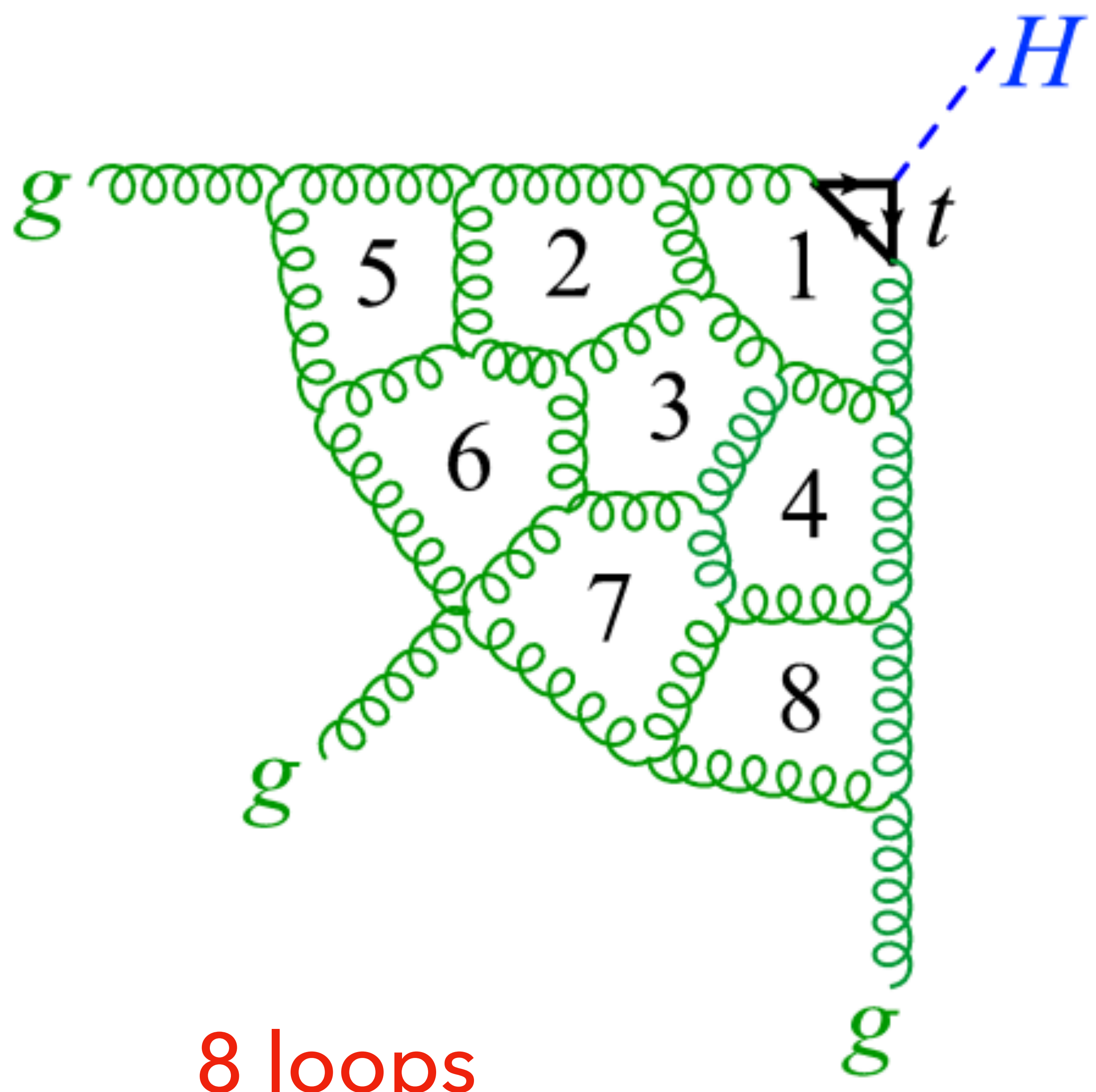


(1) ML & AMPLITUDES

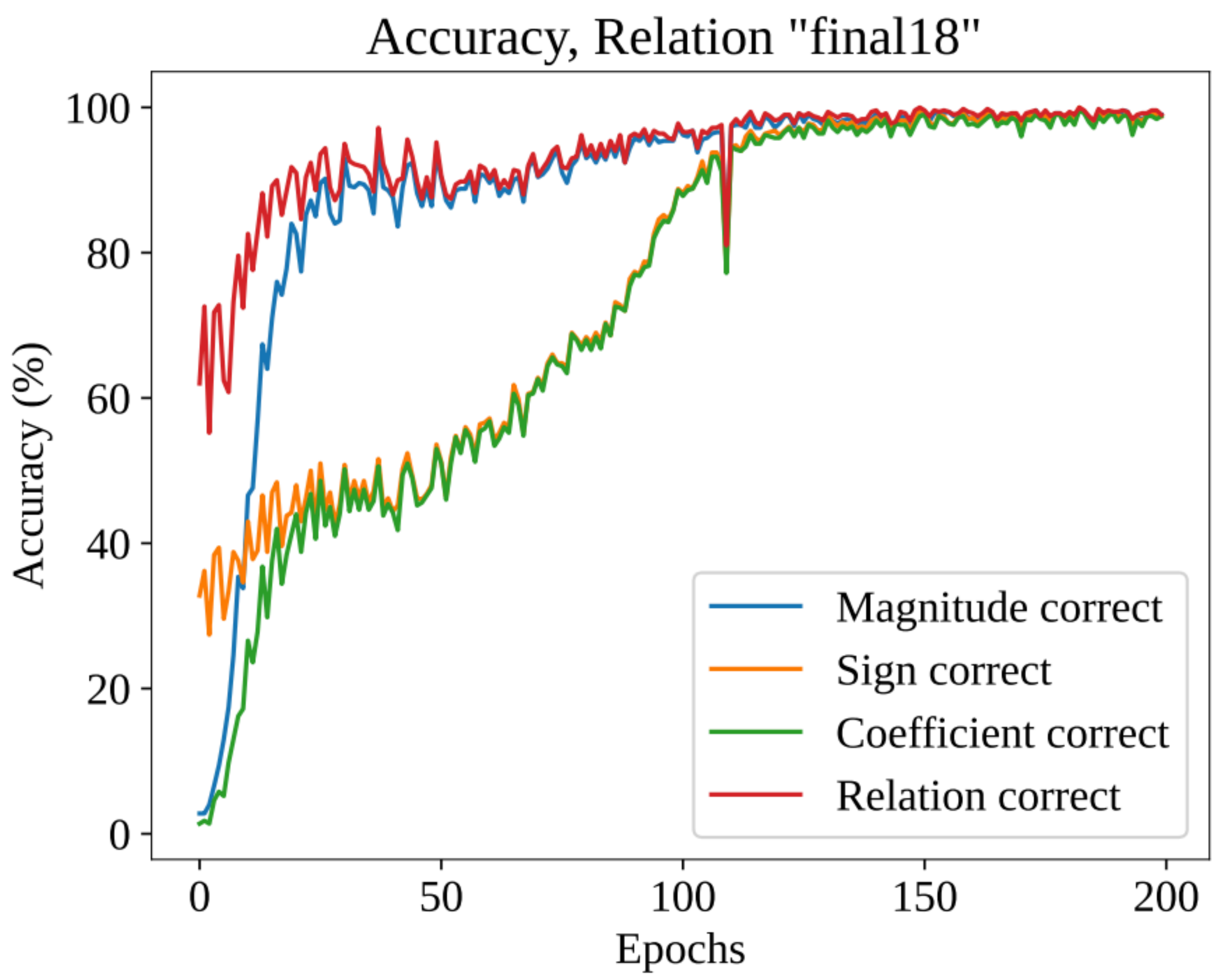
- Use a NN transformer model to **predict coefficients** of elements in the symbols for scattering amplitudes in N = 4 planar Super Yang-Mills (SYM) theory [Cai, Merz et al, arXiv: 2405.06107]
- Two-step learning: coefficient magnitude by grouping, then the true signs.
- Model can predict the coefficients at loop L using only a small subset of related coefficients at loop (L − 1).
- Extendable to other problems where multi-loop data encoded by symbols



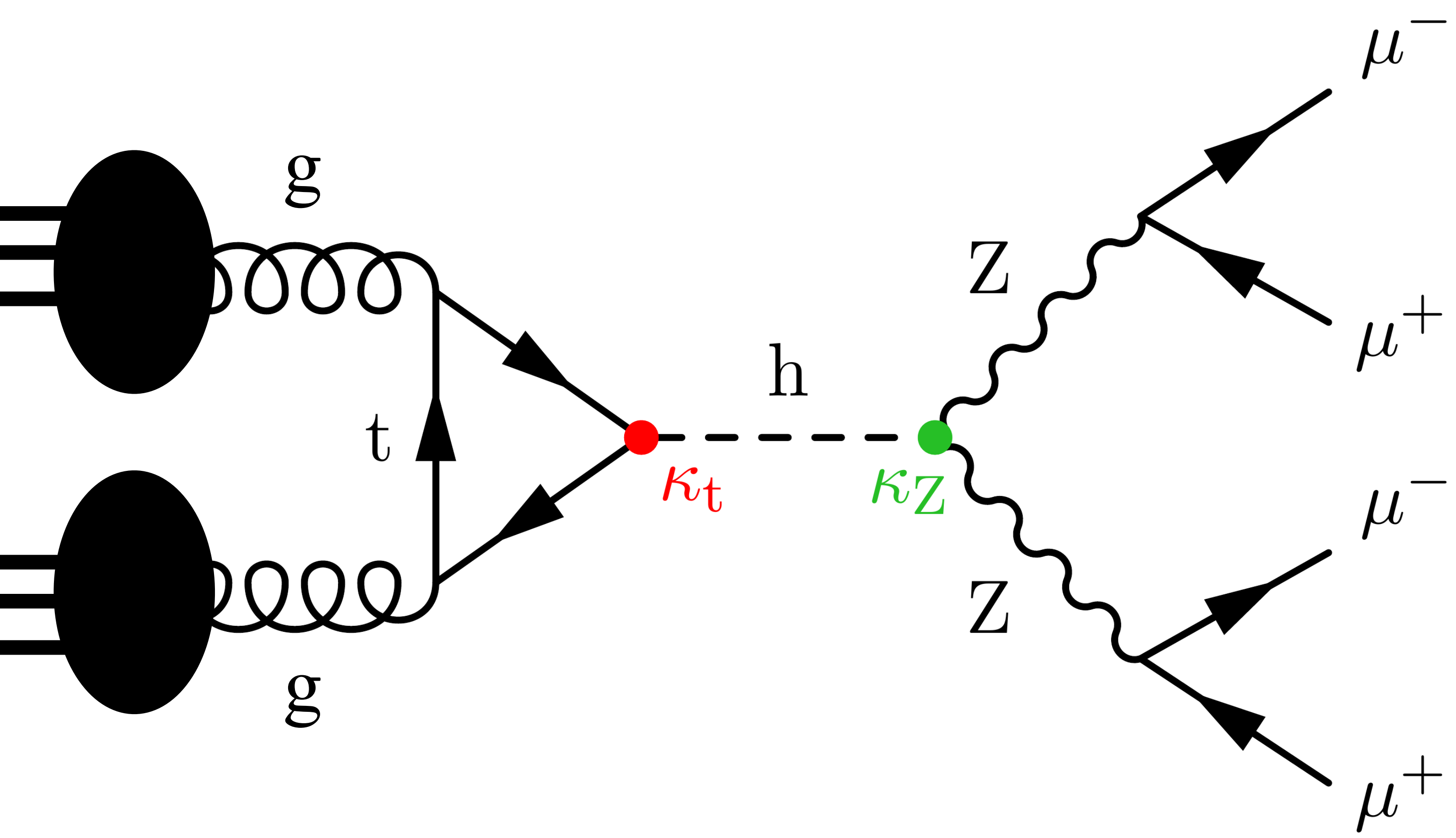
2 loops



8 loops



Cai, Merz et al, arXiv: 2405.06107

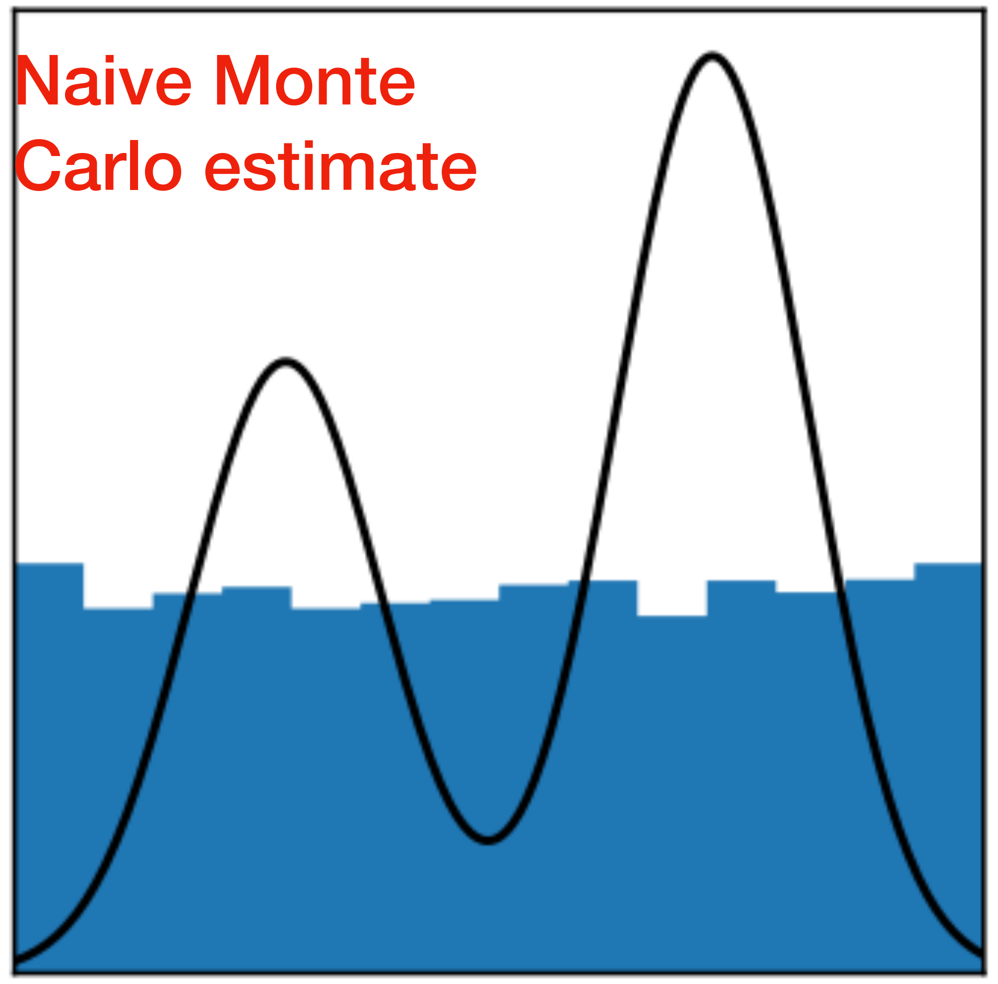


$$d\sigma \approx |\mathcal{A}_{ij\to 1\ldots n}(x)|^2 \times \underbrace{d\Phi_n}_{} \times f_{p/i} \times f_{p/j}$$

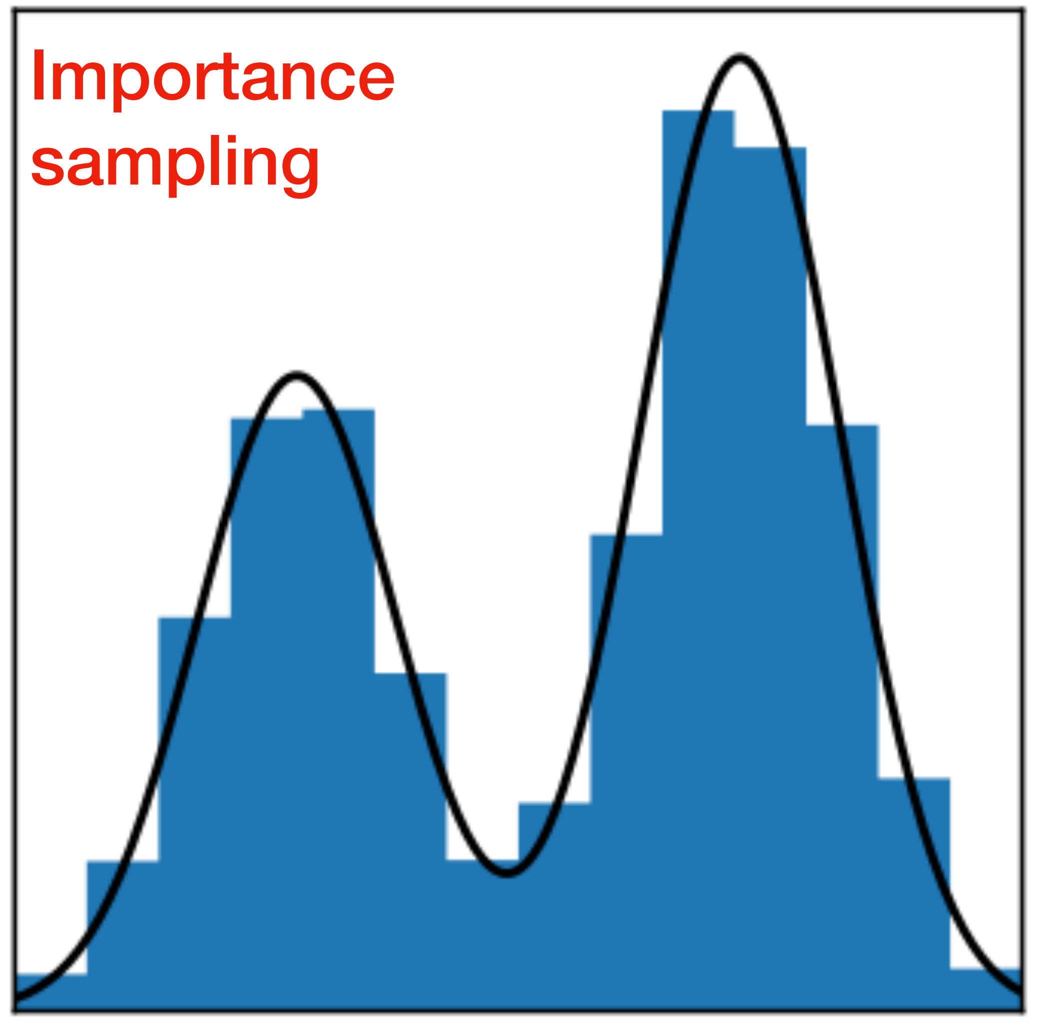
(2) ML & PHASE SPACE

$$\hat{\sigma} \approx \int_{\Phi} |\mathcal{A}_{ij\to 1\ldots n}|^2 d\Phi_n \to I = \int_{\Omega} f(u')\, du'$$

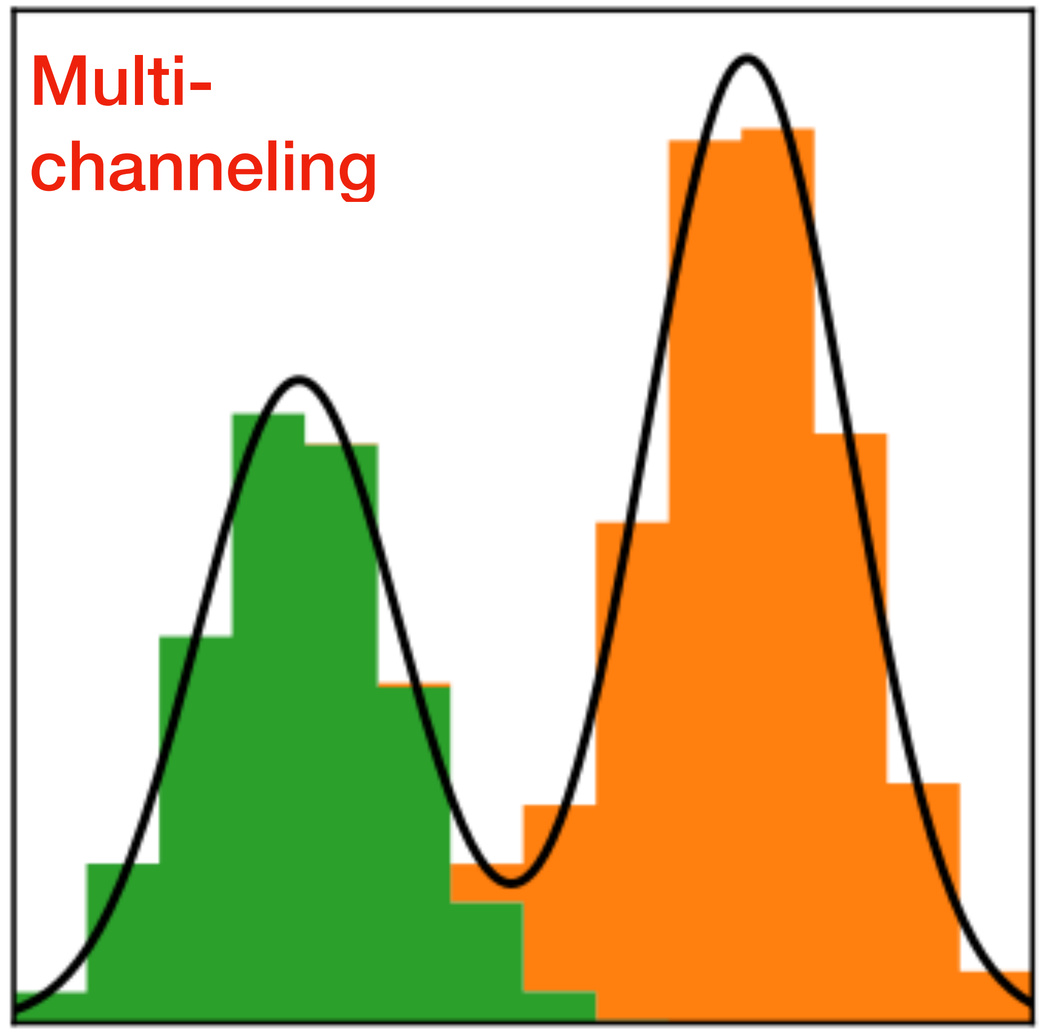
- Key task improve precision and efficiency of integration of squared amplitudes over phase space
- Importance sampling and multi-channeling efficiency depend on suitability of variable transformation and on the judicious choice of channels
- Standard VEGAS approach [G. P Lepage 1978] adaptive importance sampling algorithms that fits bins with equal probability and varying width: cheap to implement but no correlations $\Longrightarrow$ struggles with multimodal functions if peaks not aligned with coordinate axes.



$$I \approx E_N = \frac{1}{N}\sum_{i=1}^{N} f(u_i) = \langle f \rangle_x$$



$$I \approx E_N = \frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)} = \left\langle \frac{f}{g} \right\rangle_{x\sim g(x)}$$



$$I \approx E_N = \frac{1}{N}\sum_{j=1}^{N_c}\sum_{i=1}^{N} \alpha_j \frac{f(x_i)}{g_j(x_i)} = \sum_{j=1}^{N_c} \left\langle \alpha_j \frac{f}{g_j} \right\rangle_{x\sim g_j(x)}$$

Image credit @ T. Heimel, Pheno 2023

(2) ML & PHASE SPACE

- Bijective **Normalising Flows** (NFs) chain of invertible, learnable transformations with exact likelihood from change of variables that allow to redistribute input of random variables z_1 to the mapping functions z_n and better adapt to integrands

[Müller et al,1808.03856, Bothmann et al 2001.05478, Gao et al. 2001.05486, 2001.10028, Chen et al 2009.07819, Pina-Otey et al 2005.12719...]

$$\log\, p(z_n|c) = \log\, p(z_1) + \log\left(\det \frac{\partial z_1(z_n;c)}{\partial z_n}\right)$$

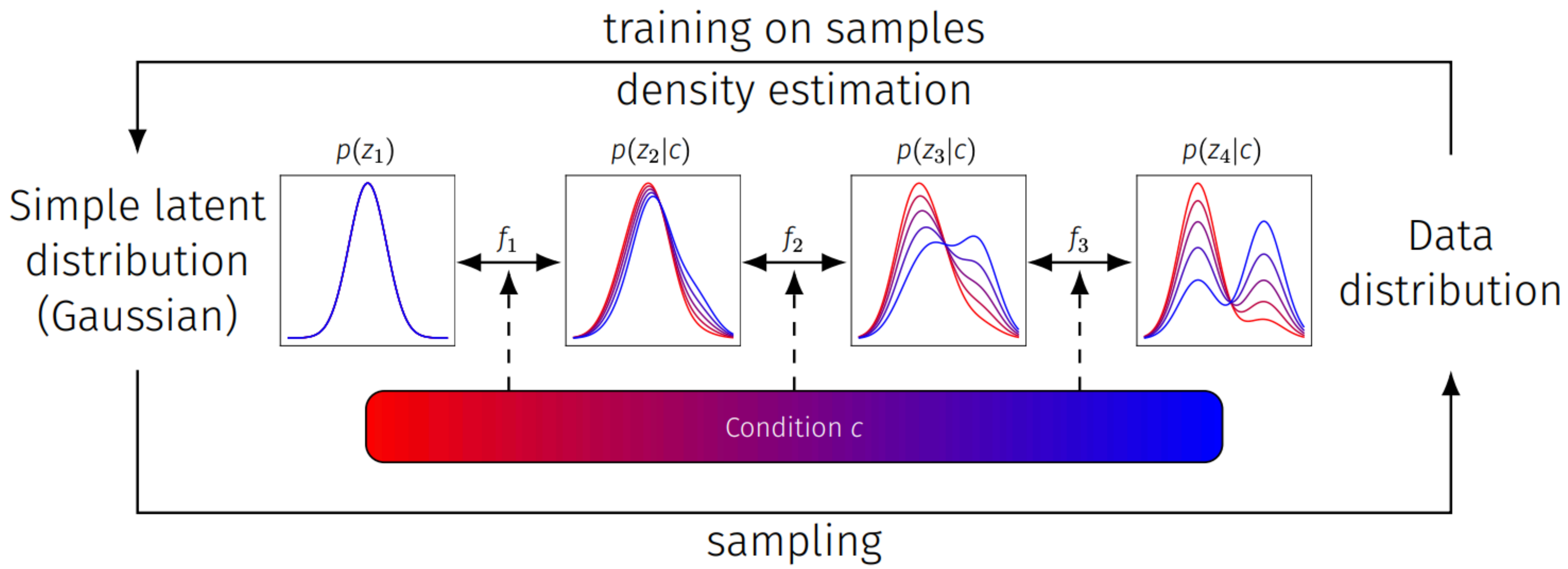


Image credit @ T. Heimel, Pheno 2023

(2) ML & PHASE SPACE

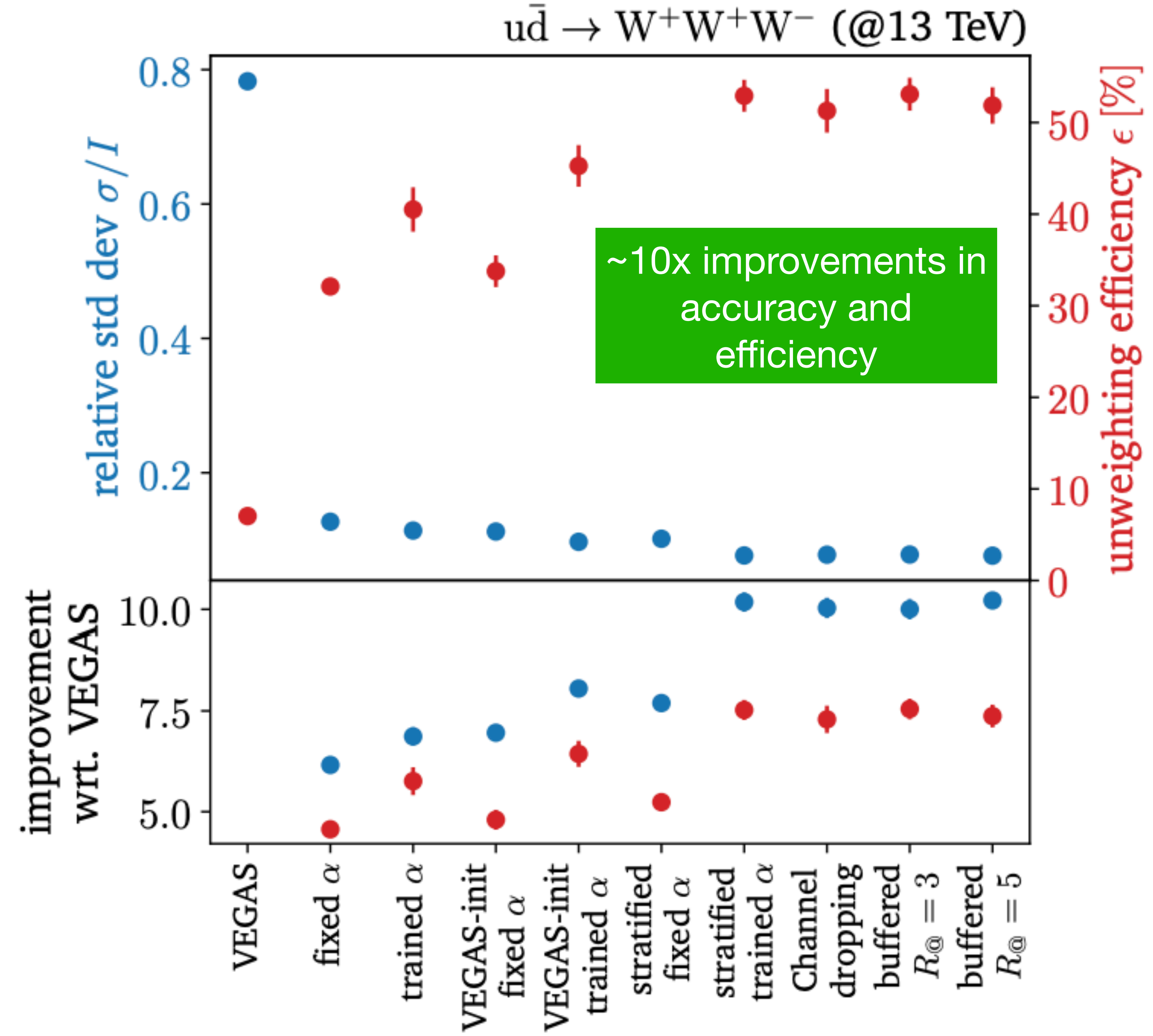
- **MadNIS** framework combines standard automated event generator MadGraph with ML tools (Channel-Weight NN to encode local multichannel weights + Invertible NN = a NFs equally fast in both directions) to make computation faster and more accurate [Heimel, Winterhalter et al 2212.06172, 2311.01548, 2408.01486]

$$I \approx E_N = \sum_j \left\langle \alpha_j(x) \frac{f(x)}{g_j(x)} \right\rangle_{x\sim g_j(x)}$$

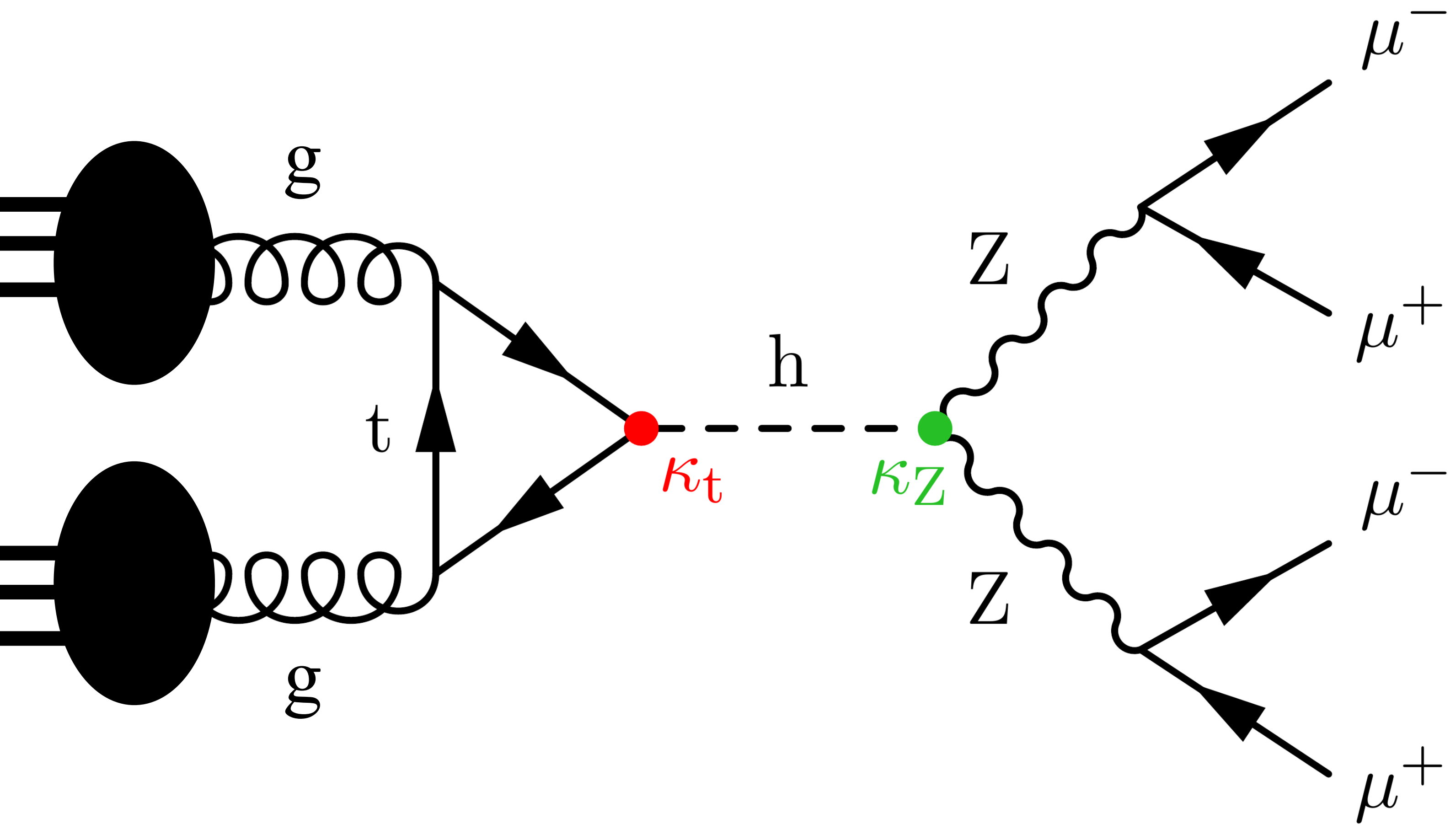
Channel weight from standard MadGraph x function parametrise by NN

Analytic mapping by MadGraph refined with NFs

$$\boxed{x \in \mathbb{R}^D \xleftrightarrow{\text{analytic}} y \in [0,1]^D \xleftrightarrow{\text{INN}} z \in [0,1]^D}$$



Heimel et al, 2311.01548

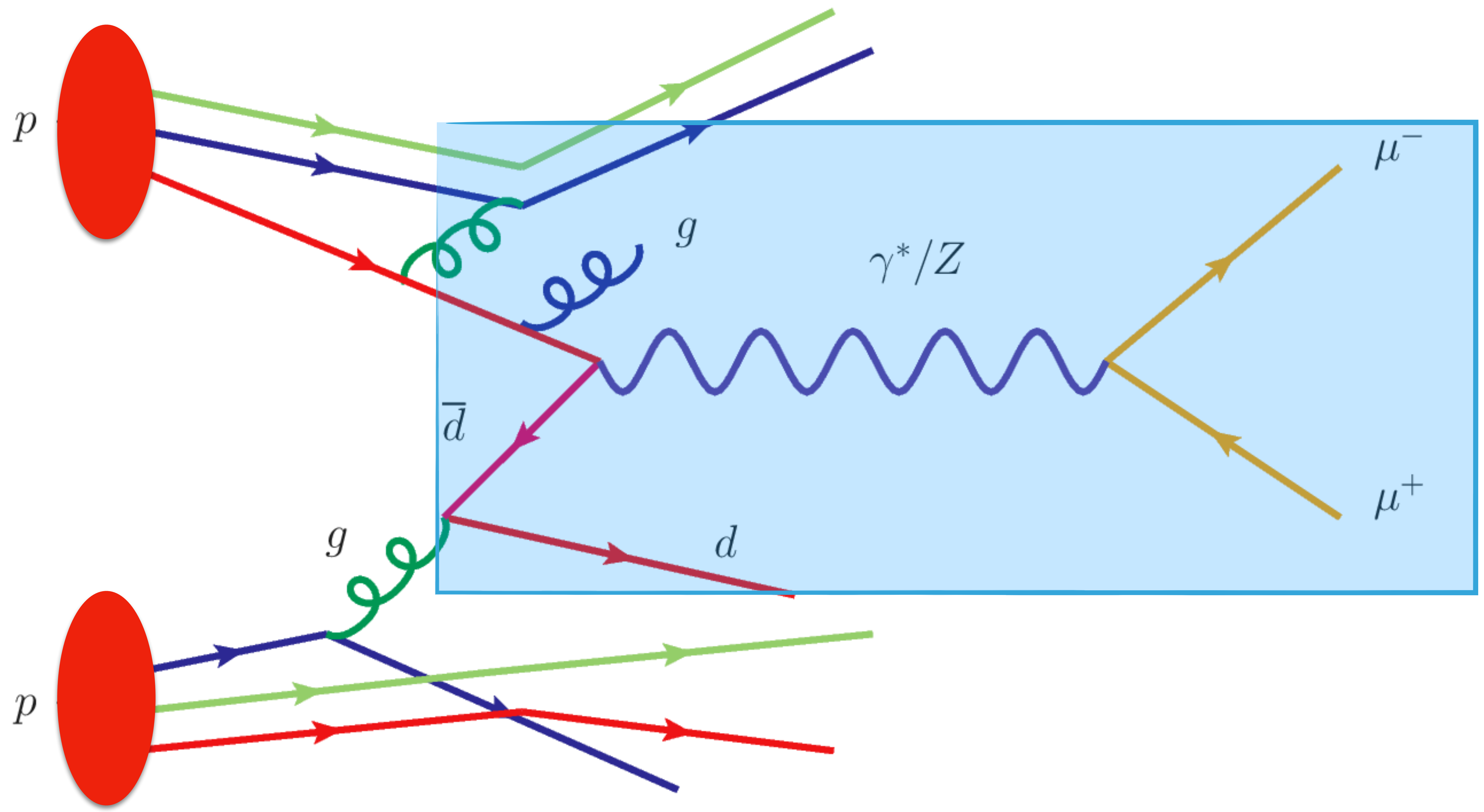


$$d\sigma \approx \left|\mathcal{A}_{ij\to 1\ldots n}(x)\right|^2 \times d\Phi_n \times f_{p/i} \times f_{p/j}$$

(3) ML & PARTON DISTRIBUTION FUNCTIONS

- What is the probability that for a given event a parton **i** is carrying a fraction x_1 of one proton's momentum and parton **j** is carrying a fraction x_2 of the other proton's momentum?
- Parton Distribution Functions (PDFs) carry this information and their **uncertainty** is crucial input at the LHC

Parton Distribution Functions



Amplitudes + Phase space integration

- PDFs are universal
- QCD predicts dependence of PDFs on energy μ

⇒ can extract PDFs from data

$$d\sigma^{pp\to ab} = \sum_{i,j} \boxed{f_i \otimes f_j} \otimes \boxed{d\hat{\sigma}^{ij\to ab}} + \ldots$$

$$f_i(x, \mu)$$

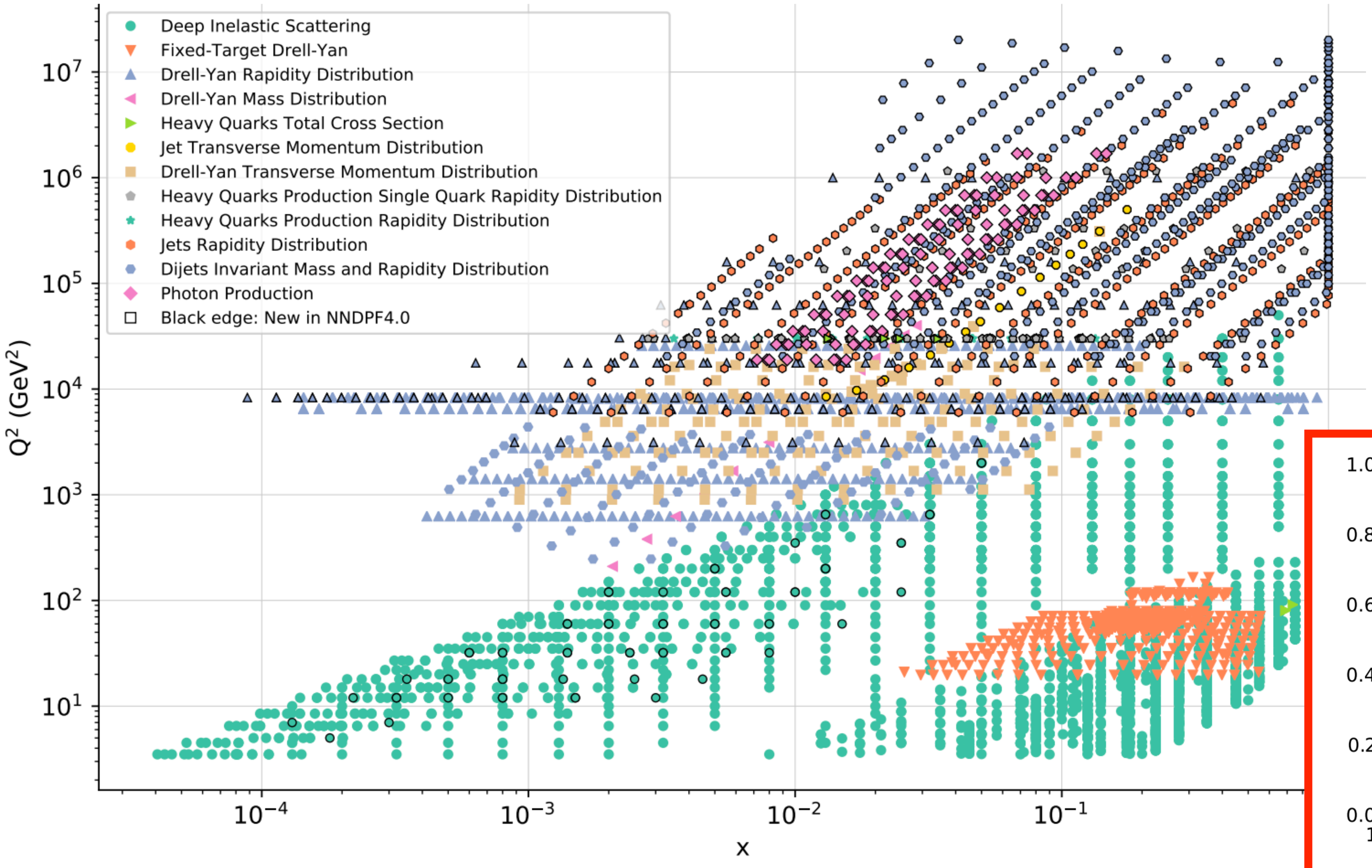
Theory

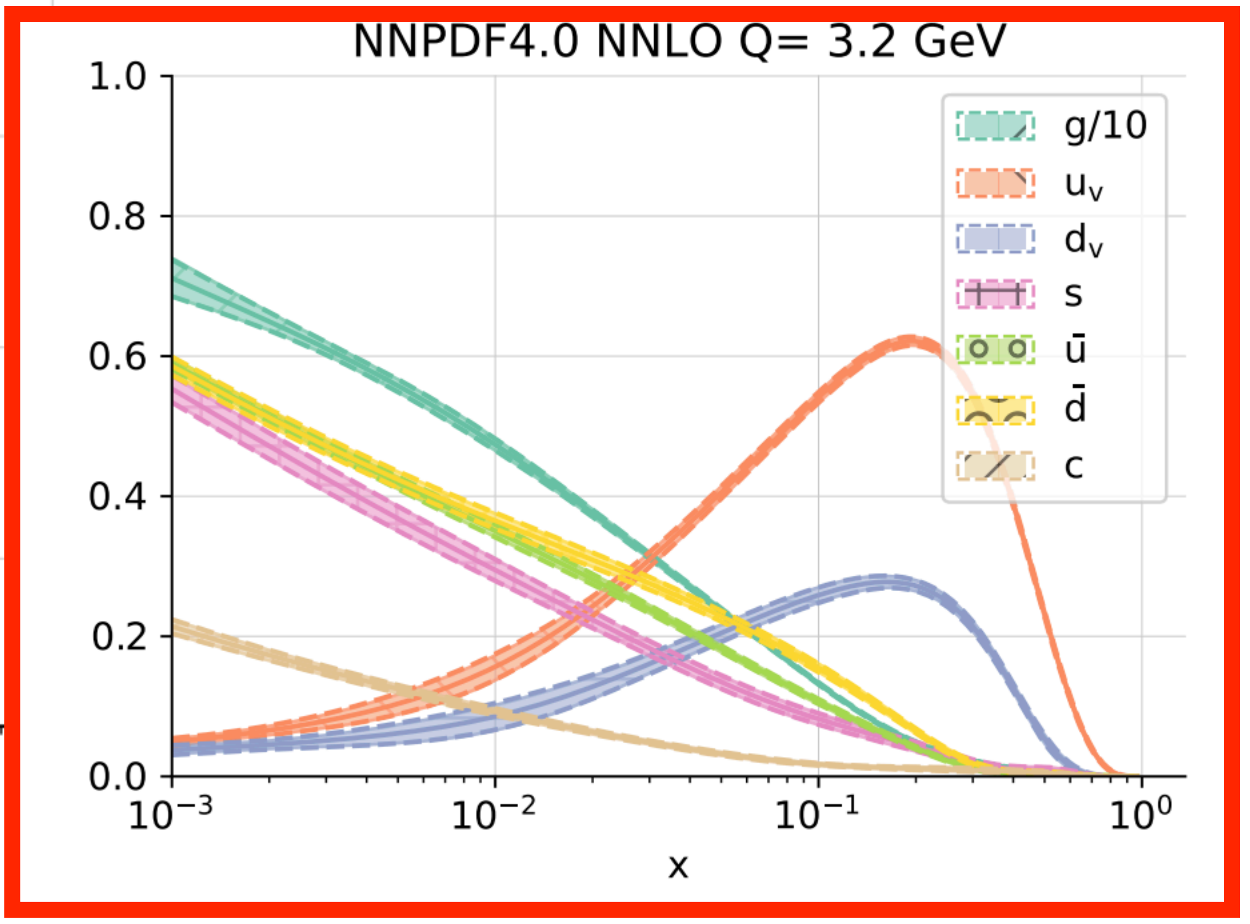
Data

(3) ML & PARTON DISTRIBUTION FUNCTIONS

$$f_i(x, \mu)$$

- Given finite set of discrete N_{dat} experimental data points **D** want to determine some functions **{f}** at a given energy scale $\boldsymbol{\mu}_0$ that theoretical predictions depend on and **estimate their uncertainty**
- Want to find a infinite-dimensional probability density from a finite number of information **(regression & infinite-dimensional inverse problem)**.

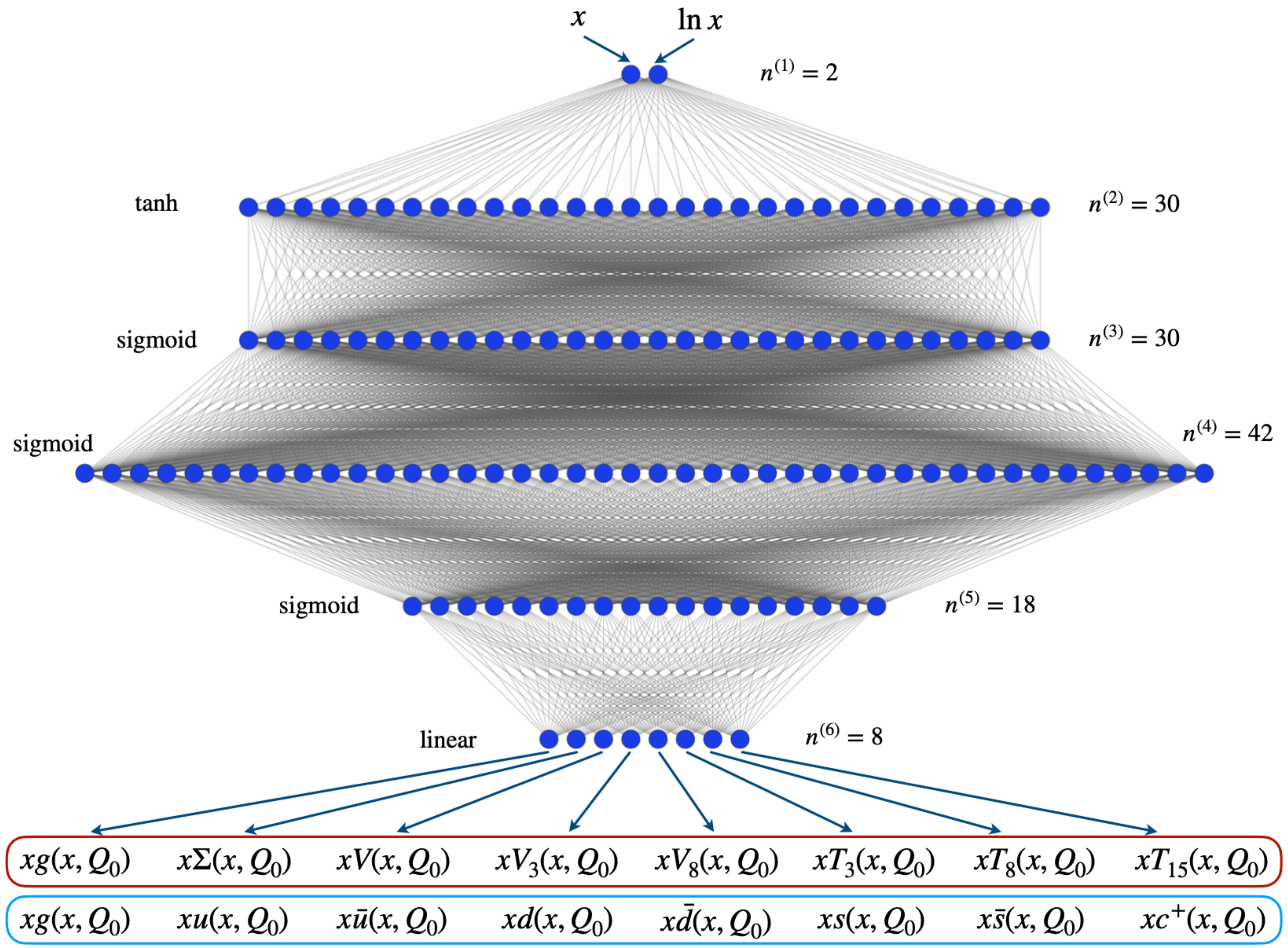




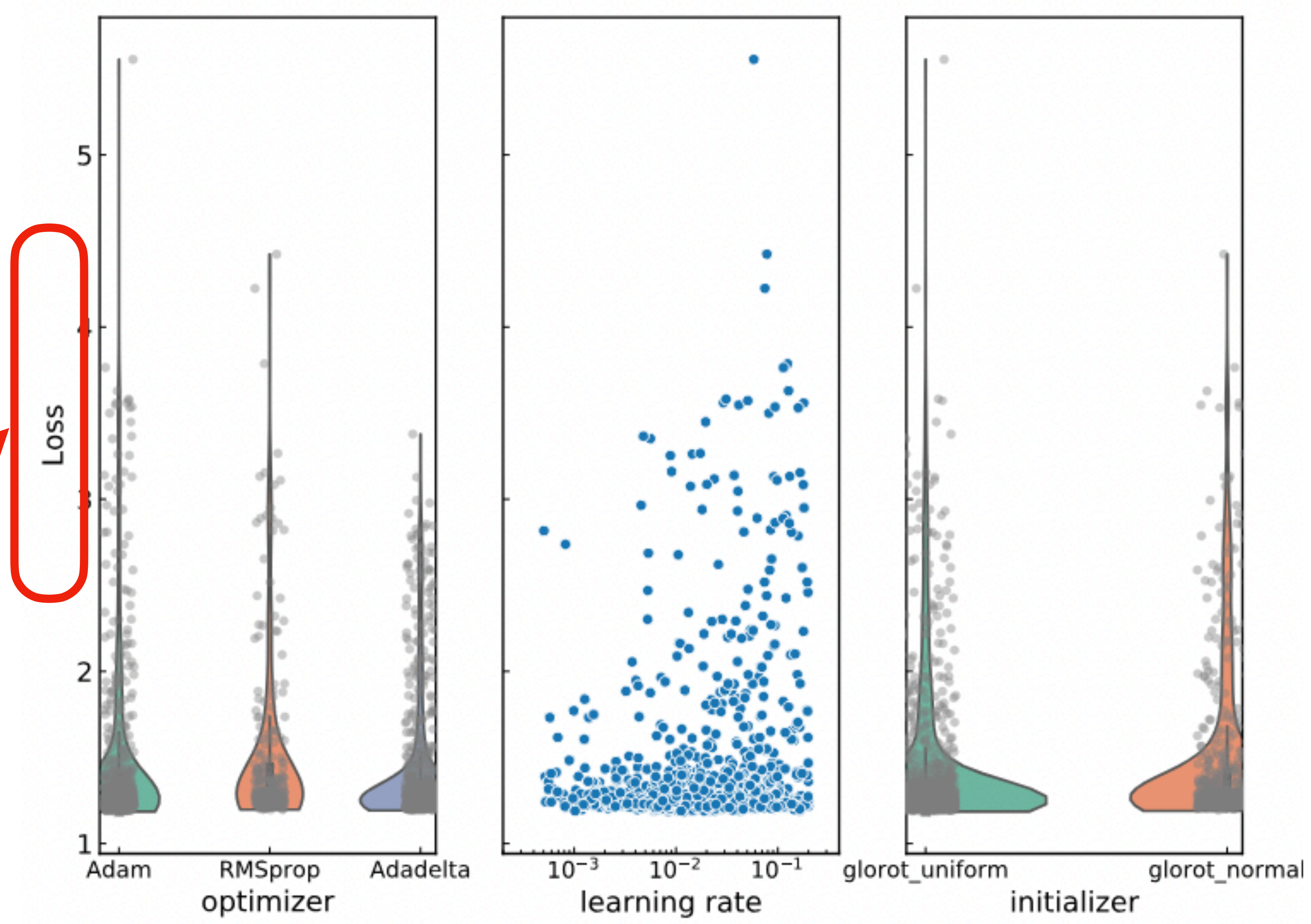
- Overall $n_f = 8$: up, anti-up, down, anti-down, strange, anti-strange, charm, gluon
- O(5000) **parton level** data from electron-proton, electron-nucleon, proton-antiproton and proton-proton experiments.

(3) ML & PARTON DISTRIBUTION FUNCTIONS

- Traditional approach: project infinite dimensional functional space into N_{par}-dimensional parameter space.
- **Problem:** parametrisation might induce bias and uncertainties are inflated by a "tolerance" factor
- With Deep NN can choose parametrisation so large that in principle can fit any conceivable function f
  [Forte, Latorre 2002] [Ball, Del Debbio, Forte, Guffanti, Piccione, Rojo, MU, 2008]
- NNPDF4.0: Single DNN with hyper-parameter optimised via K-fold procedure
  [Ball et al arXiv:2109.02653 + public code publication arXiv:2109.02671] [Cruz-Martinez et al 2410.16248]



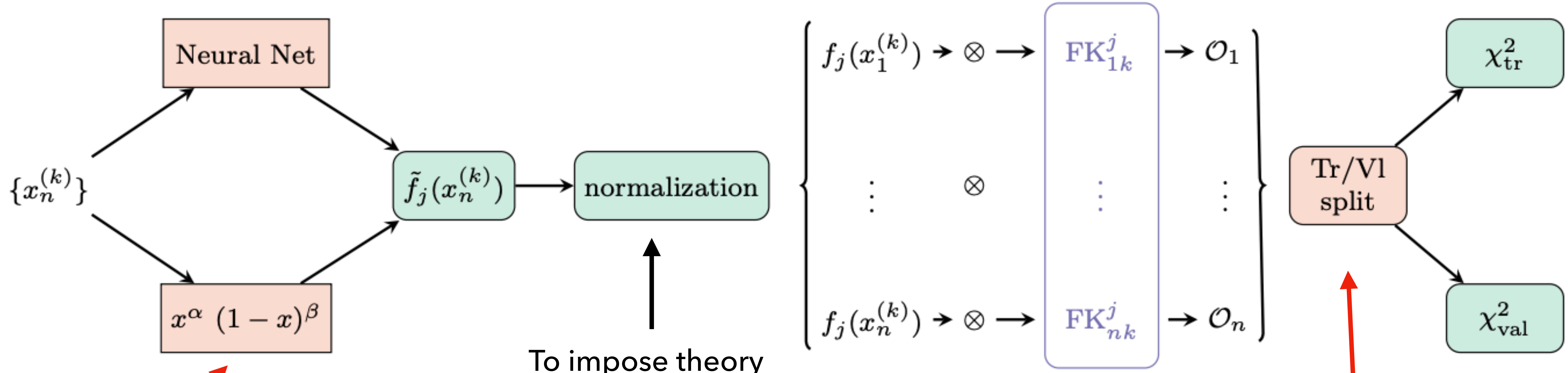
K-fold loss



List of hyper-parameters

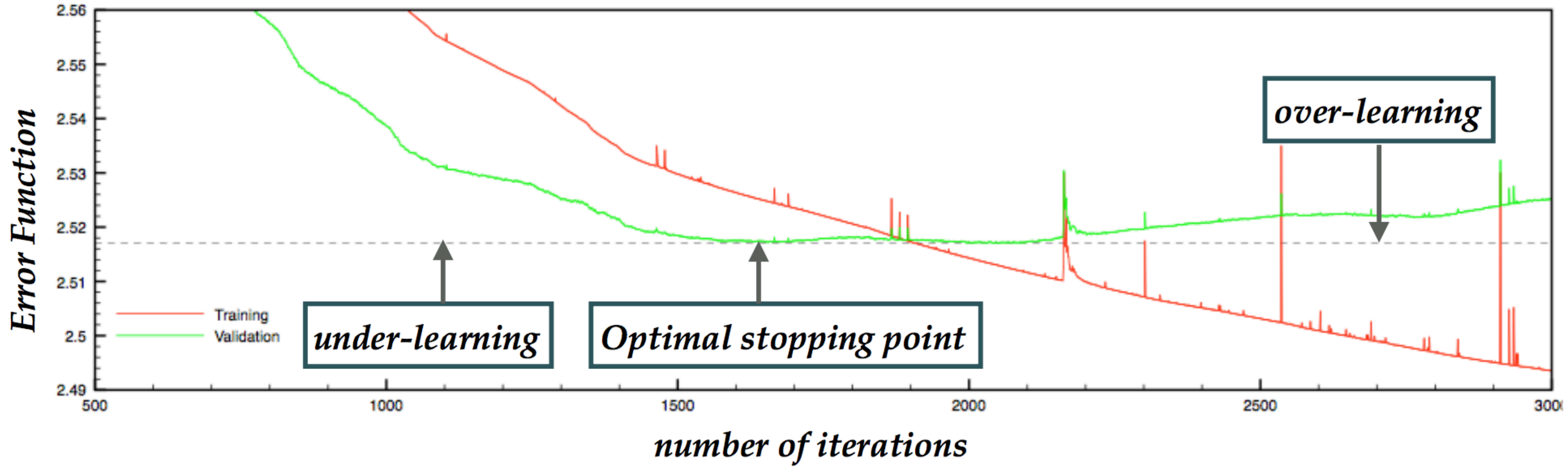
NEURAL NETWORK	FIT OPTIONS
NUMBER OF LAYERS (*)	OPTIMIZER (*)
SIZE OF EACH LAYER	INITIAL LEARNING RATE (*)
DROPOUT	MAXIMUM NUMBER OF EPOCHS (*)
ACTIVATION FUNCTIONS (*)	STOPPING PATIENCE (*)
INITIALIZATION FUNCTIONS (*)	POSITIVITY MULTIPLIER (*)

(3) ML & PARTON DISTRIBUTION FUNCTIONS



To impose theory constraints

α, β pre-processing exponents



(3) ML & PARTON DISTRIBUTION FUNCTIONS

- How to estimate uncertainties?
- Use Monte Carlo - or bootstrap - error propagation by importance sampling [Giele, Kosover 1993] [Forte et al, 2006]
- Throw random pseudo data points about the experimental data $\mathbf{D_0}$, according to a multivariate normal distributions centred on the experimental data $\mathbf{D_0}$ with experimental covariance matrix $\boldsymbol{\Sigma}_{\text{exp}}$.
- For each pseudo data compute the **optimal point on the theory surface** based on training-validation splitting and minimisation stopping procedure and obtain associated parameter values.
- Repeating gives an approximation to the **parameter distribution** by importance sampling.

$$D_0 = t + \eta$$

$$\eta \sim \mathcal{N}(0, \Sigma_{\text{exp}})$$

Vector of central experimental values

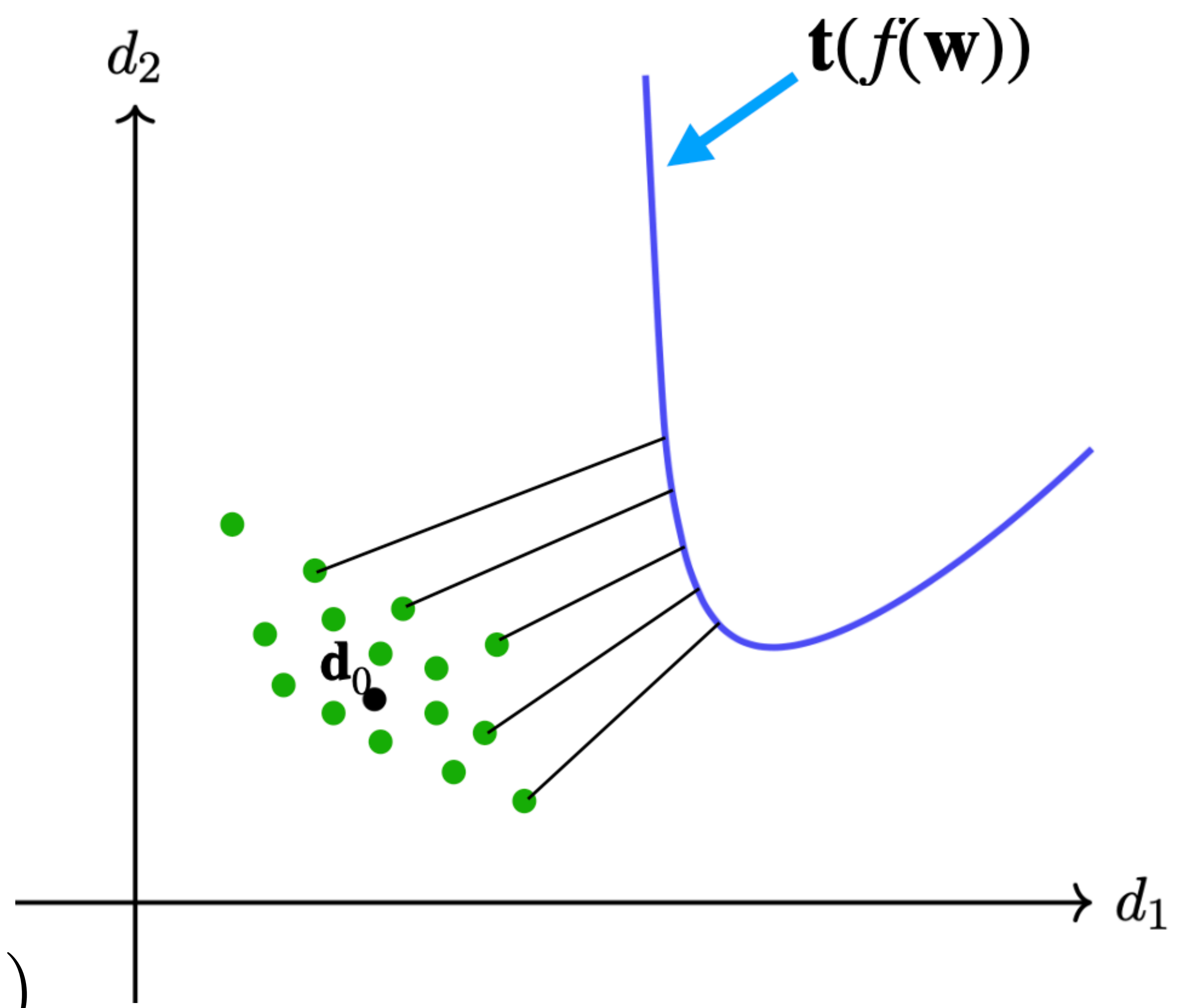
Vector of "true", unknown experimental values

Observational noise

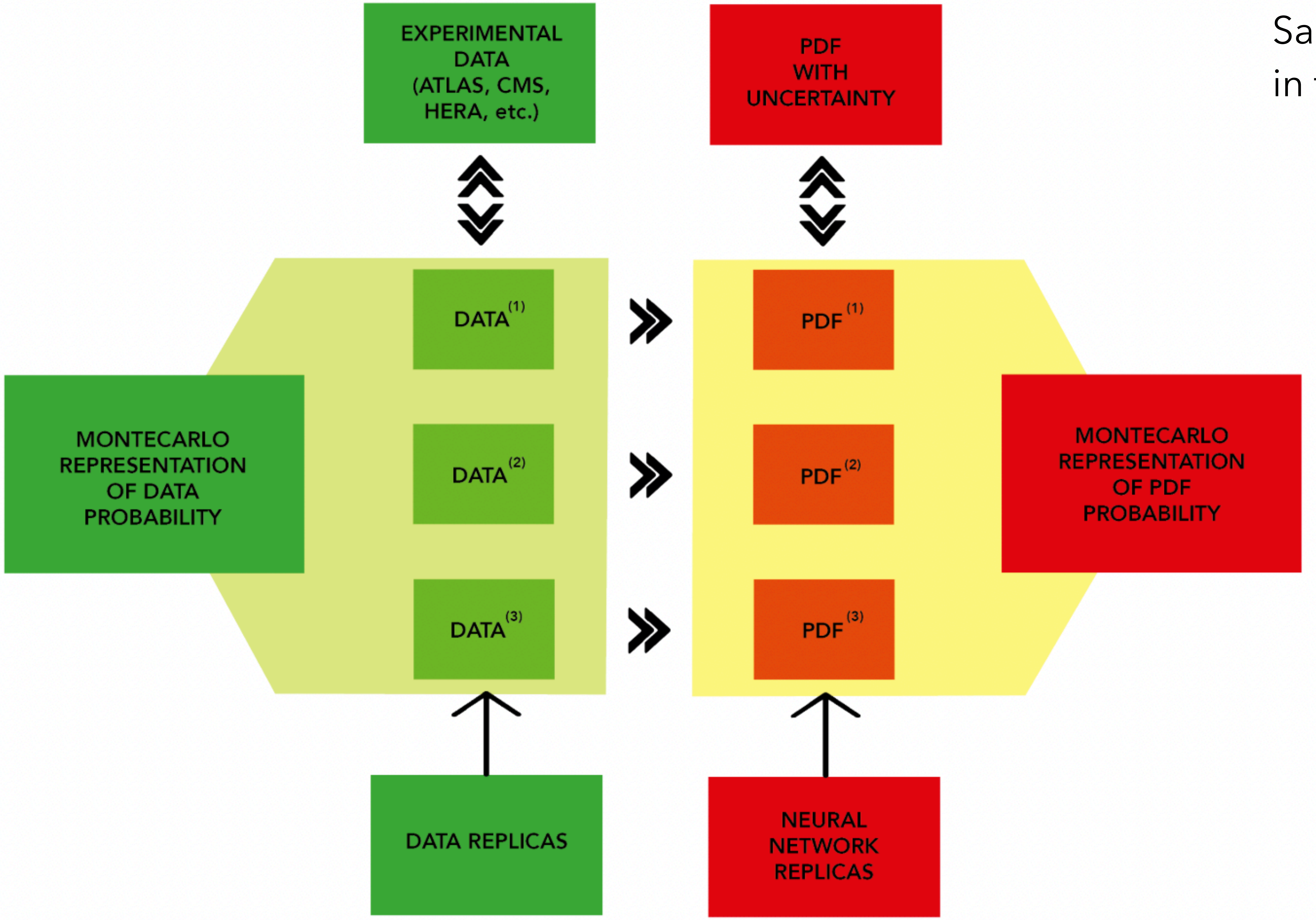
$$\mu^{(k)} = D_0 + \epsilon^{(k)} = t + \eta + \epsilon^{(k)}$$

Pseudo-data replicas, k = 1, ..., N_{rep}

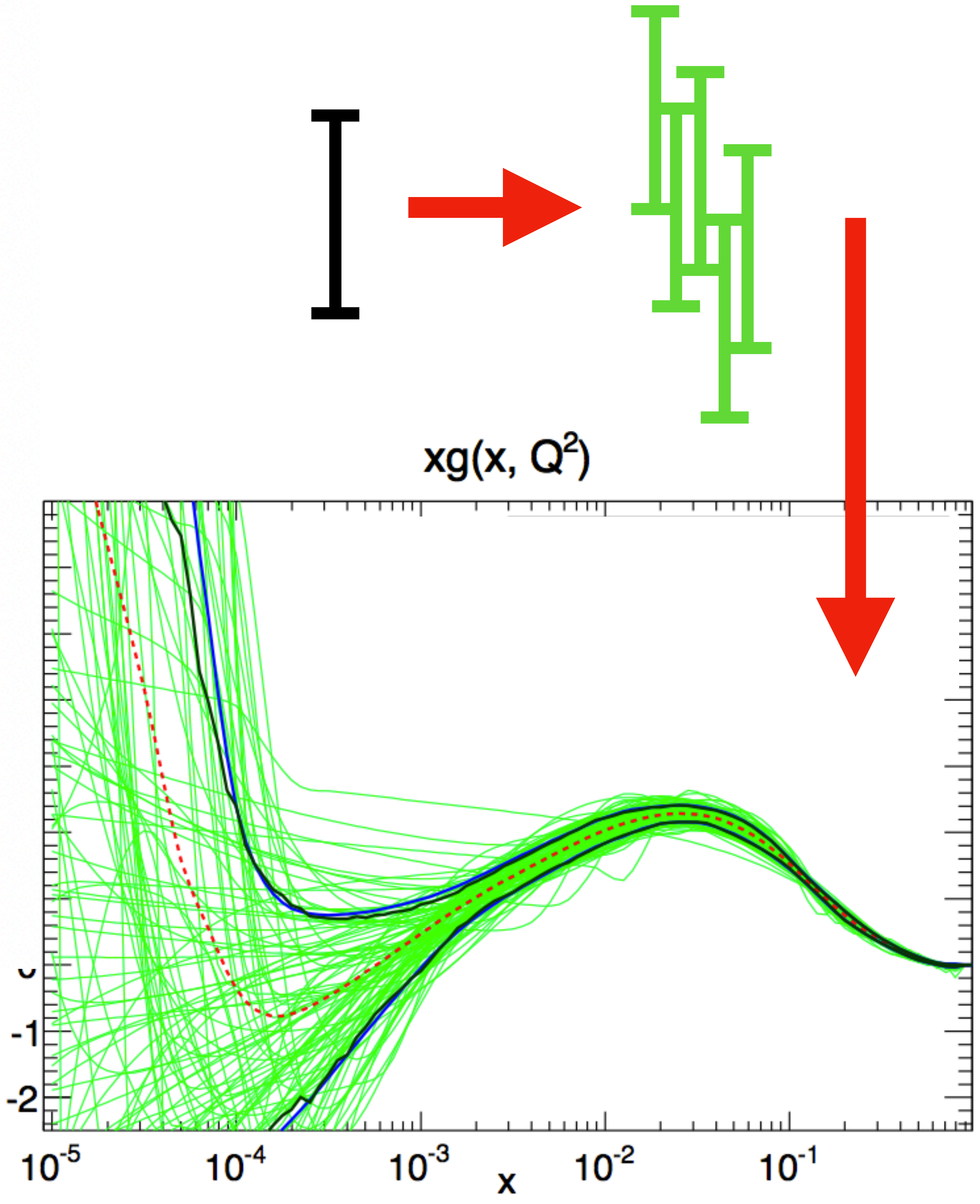
$$\epsilon^{(k)} \sim \mathcal{N}(0, \Sigma_{\text{exp}})$$



(3) ML & PARTON DISTRIBUTION FUNCTIONS

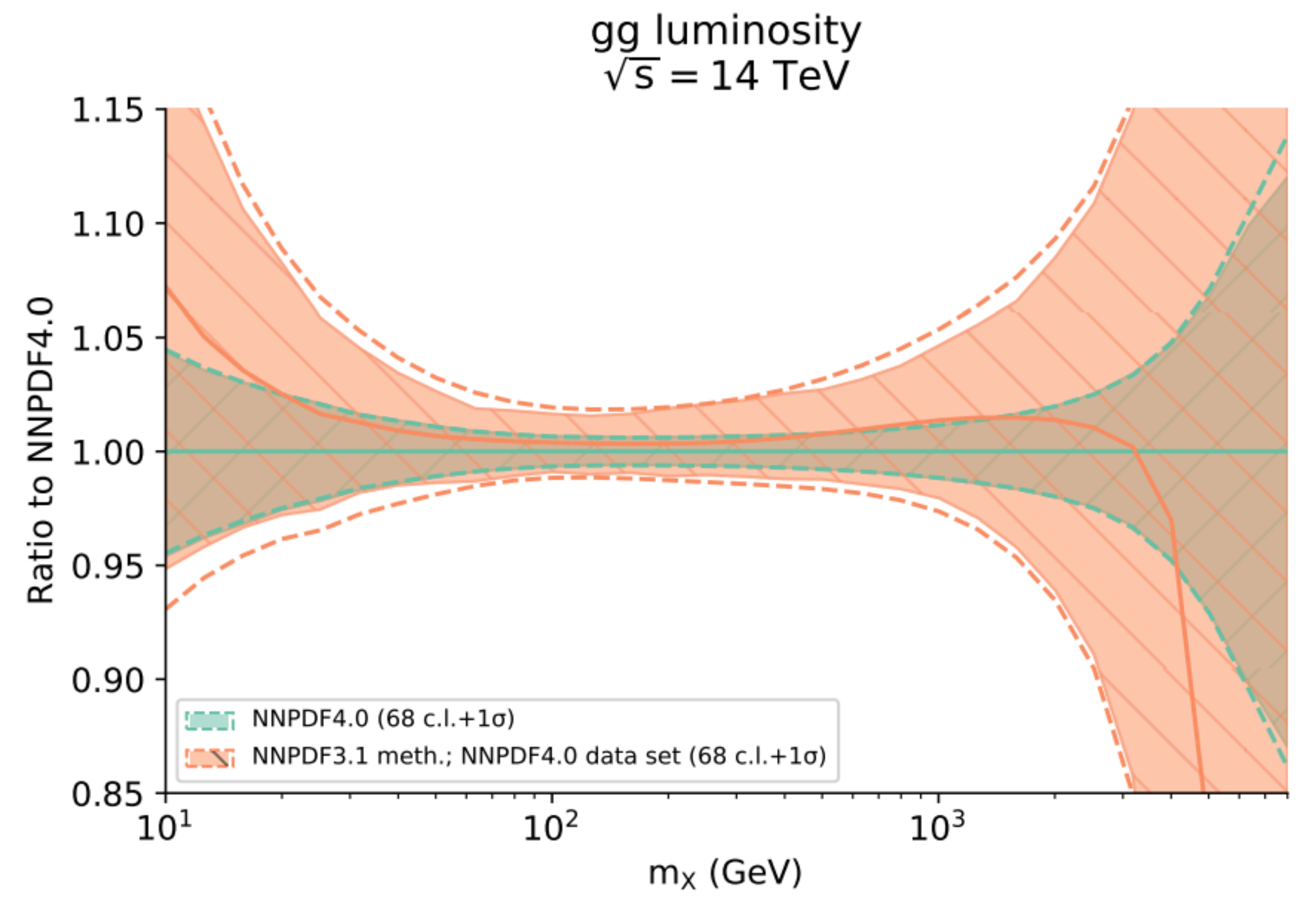


Sampling in the data space projected in the PDF space.



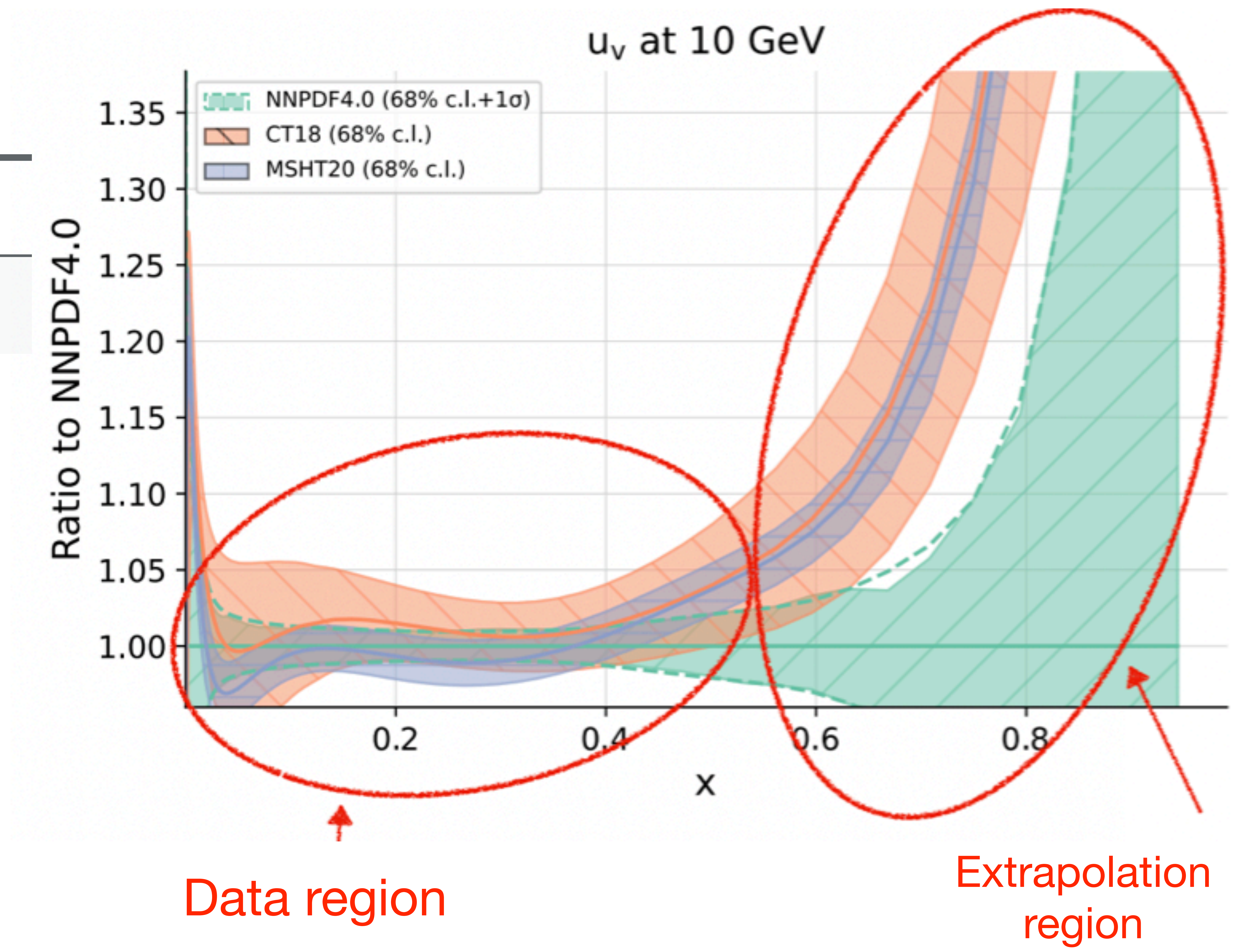
(3) ML & PARTON DISTRIBUTION FUNCTIONS

- Using a ML model instead of a fixed parametrization allows to include more data (hence more information) in the fit.
- Fit yields smaller uncertainties in the data region with same input dataset and better fit-quality.



DATA \ METHODOLOGY	NNPDF3.1	NNPDF4.0
NNPDF3.1 (4093)	1.19	1.12
NNPDF4.0 (4491)	1.25	1.17

- DNN model exhibits smaller uncertainty in data region compared to traditional PDF sets but larger uncertainties in extrapolation region.
- General Adversarial Network (GANs) used to provide compressed PDF set with less replicas and as accurate as the full ensemble [Carrazza et al 1504.06469, 2104.04535]



(3) ML & PARTON DISTRIBUTION FUNCTIONS

- NNPDF4.0 public code for testing and reproducibility https://github.com/NNPDF & https://docs.nnpdf.science/
- DNN methodology scrutinised with statistical closure test to assess faithfulness of uncertainties in the data region, response of DNN to inconsistent datasets and generalisation tests.
  [Barontini et al, 2503.17447, Chiefa et al 2501.10359]

Generate several "run of the universe"

$$D_0 = t + \eta$$

$$t = G(w^0)$$

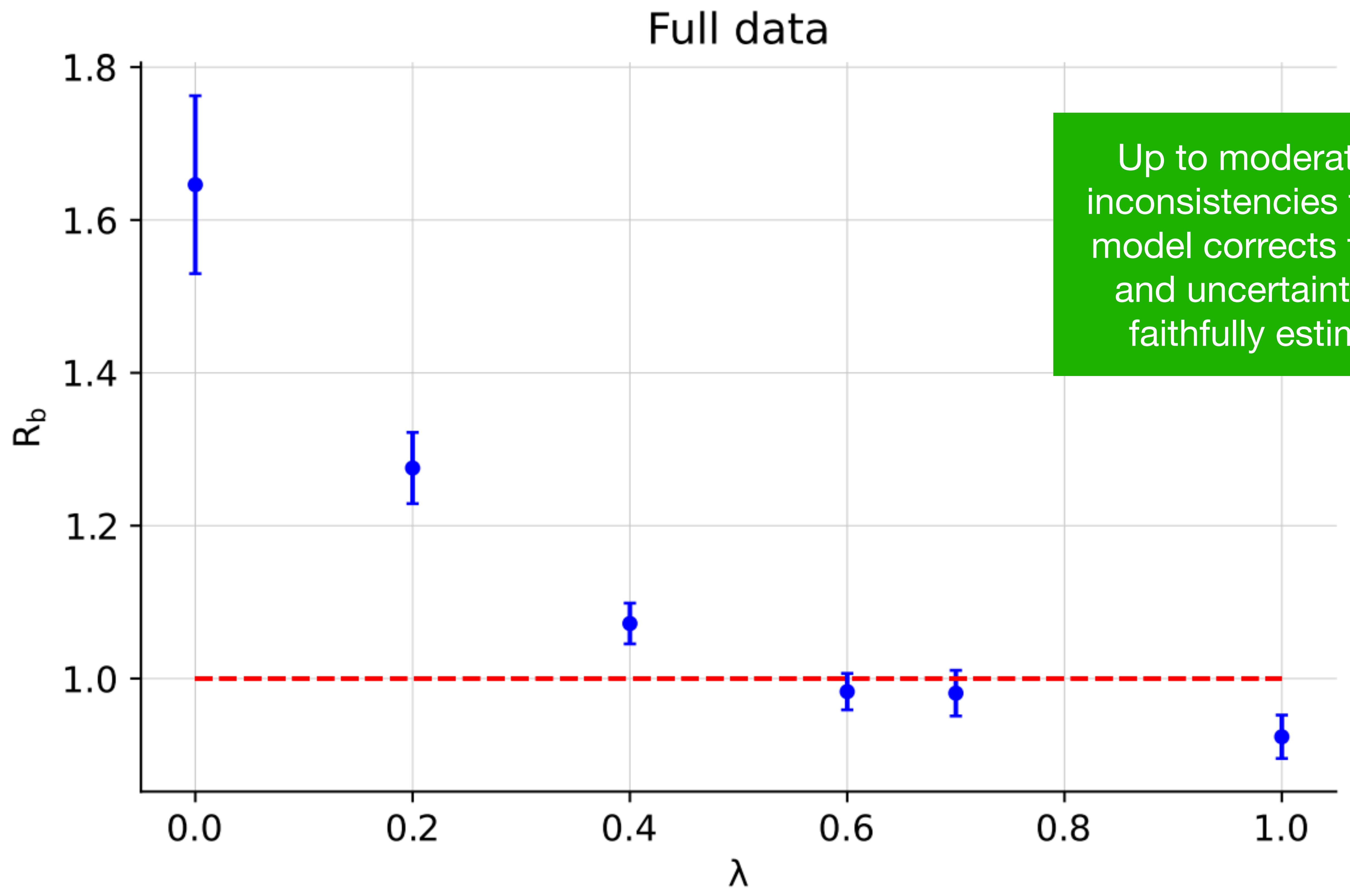
Experimental noise

Pretend know the "truth"

$$\hat{\sigma}_{\rm NNLO} \otimes (f \otimes f)_{\text{“true” PDF}}$$

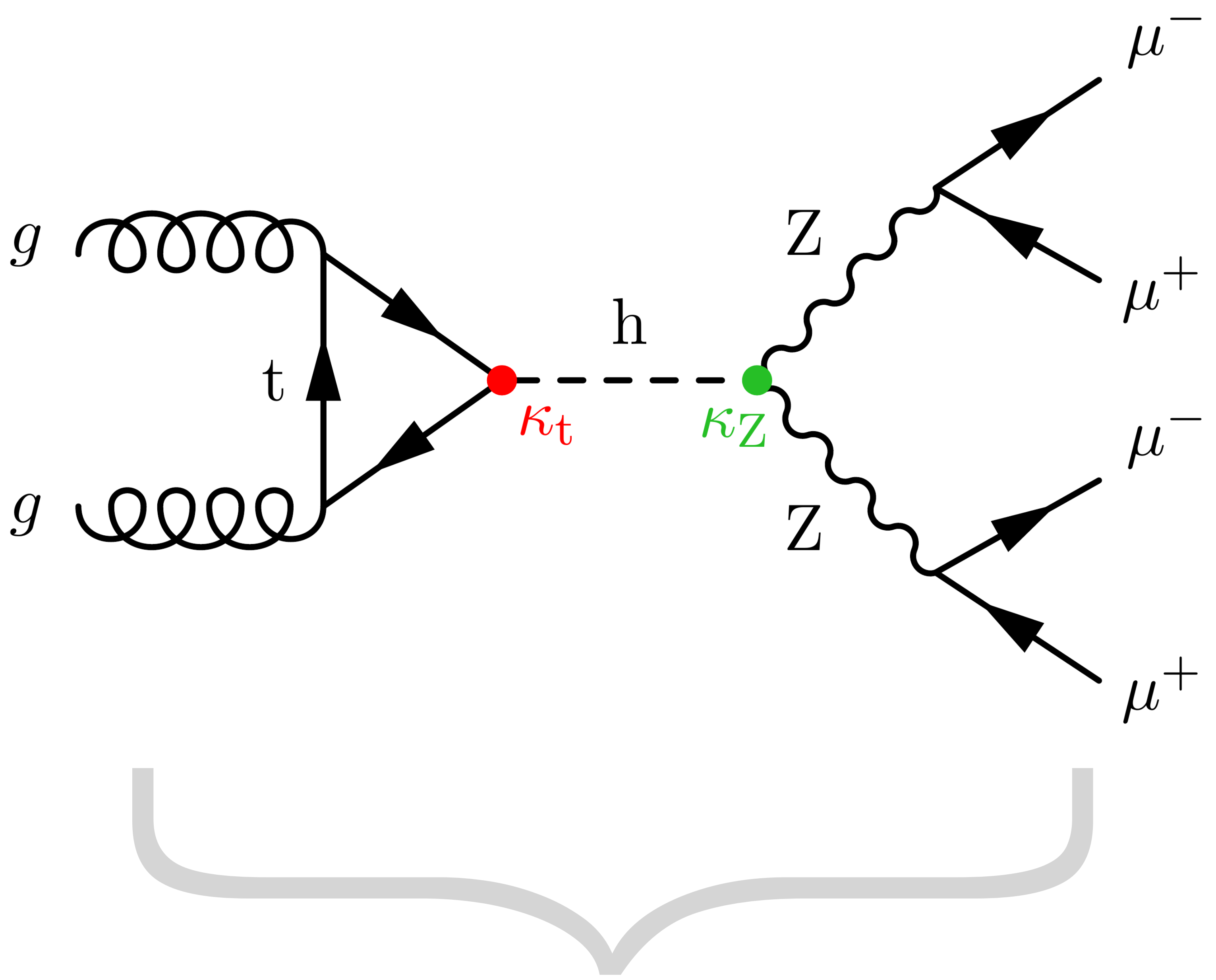
Propagate uncertainties via MC sampling

Measure normalised bias R_b = mean square deviation of predictions from the "truth" in units of predicted standard deviation



Up to moderate data inconsistencies the DNN model corrects for them and uncertainties still faithfully estimated!

Barontini et al, 2503.17447



$$d\sigma \approx \left|\mathcal{A}_{ij\to 1\ldots n}(x)\right|^2 \times d\Phi_n \times f_{p/i} \times f_{p/j}$$

(4) ML & PARAMETER DETERMINATION

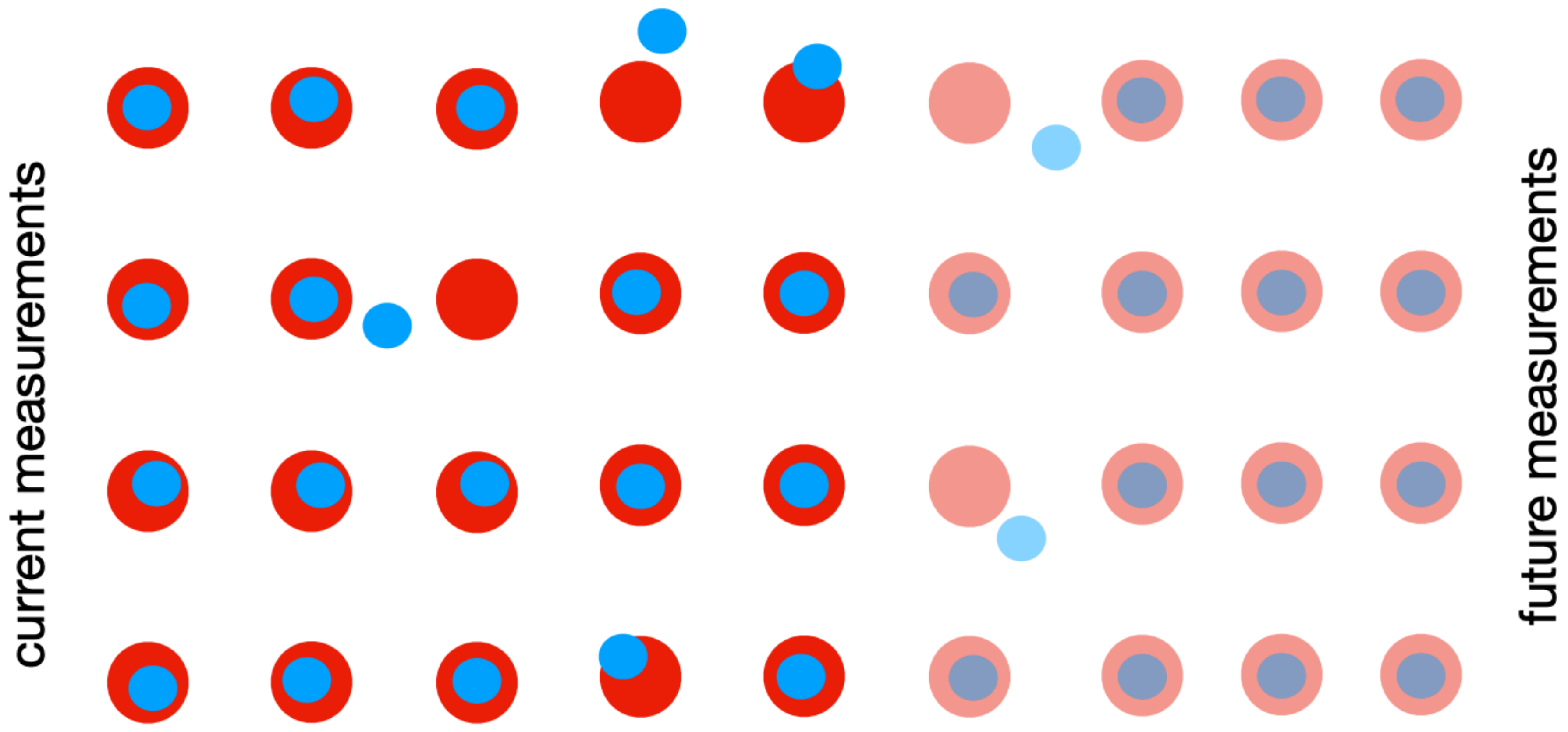
$$\Delta\mathrm{Obs}_n = \mathrm{Obs}_n^{\mathrm{EXP}} - \mathrm{Obs}_n^{\mathrm{SM}} = \frac{1}{\Lambda^2}\sum_i \boxed{a_{n,i}^{(6)}(\mu)}\, c_i^{(6)}(\mu) + \mathcal{O}\left(\frac{1}{\Lambda^4}\right)$$

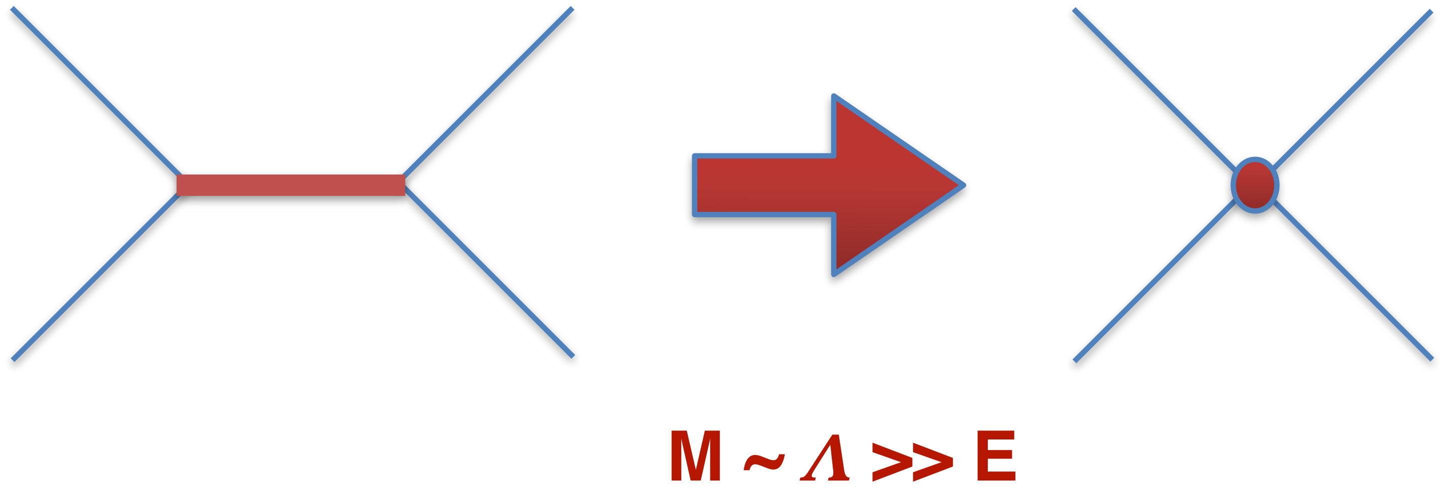
Experimental measurements

SM predictions

Wilson Coefficients parametrising the deviations due to heavy BSM physics

Image credit: F. Maltoni





$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{d=5}^{N}\sum_{i=1}^{n_d} \frac{c_{i,d}}{\Lambda^{d-4}}\mathcal{O}_i^{(d)}$$

(4) ML & PARAMETER DETERMINATION

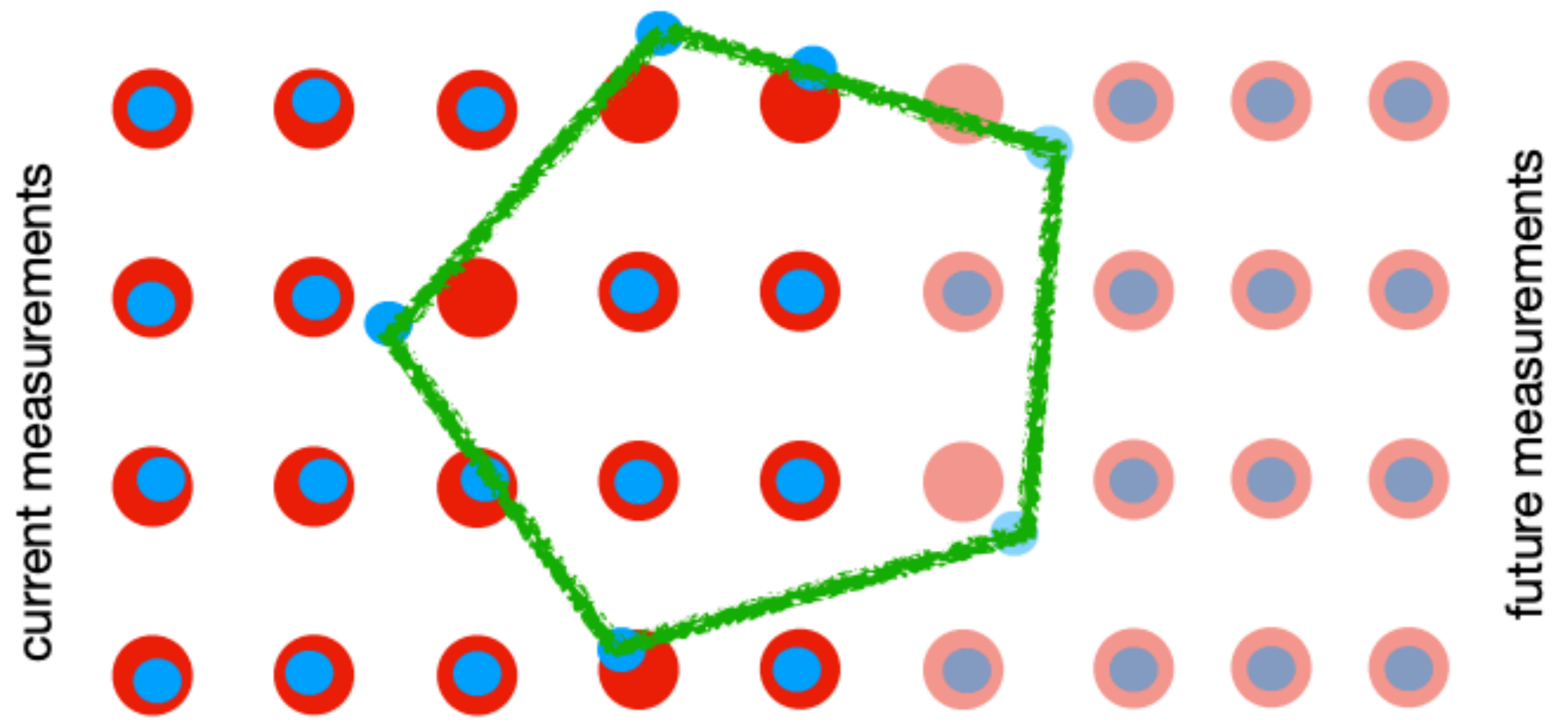
$$\Delta \text{Obs}_n = \text{Obs}_n^{\text{EXP}} - \text{Obs}_n^{\text{SM}} = \frac{1}{\Lambda^2}\sum_i \boxed{a^{(6)}_{n,i}(\mu)}\, c^{(6)}_i(\mu) + \mathcal{O}\left(\frac{1}{\Lambda^4}\right)$$

Experimental measurements

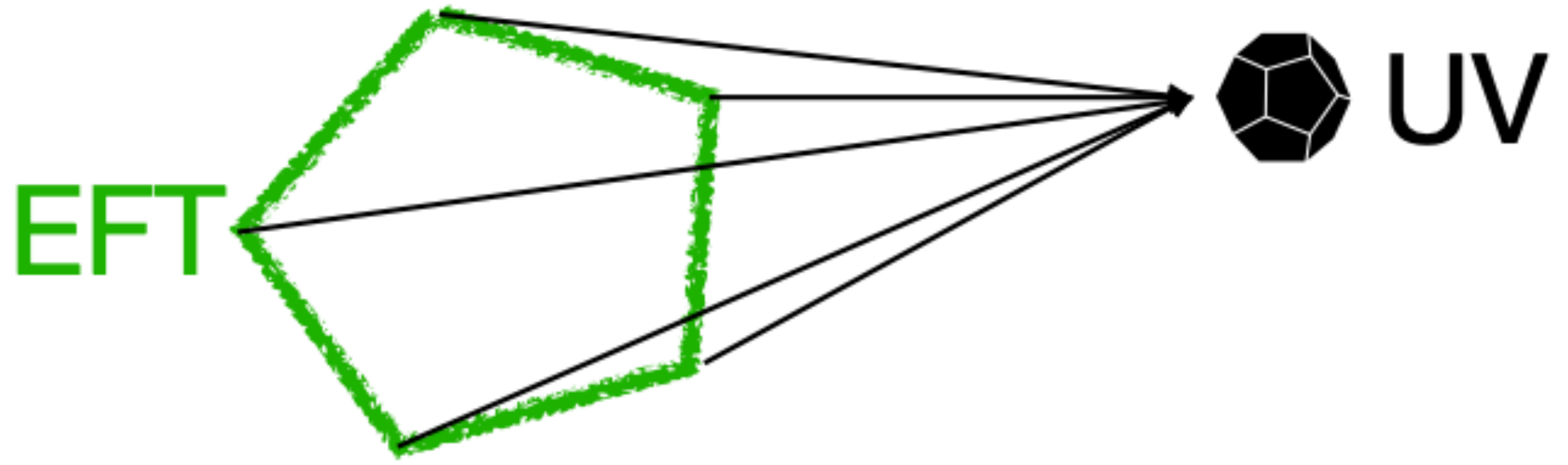
SM predictions

Wilson Coefficients parametrising the deviations due to heavy BSM physics

Image credit: F. Maltoni

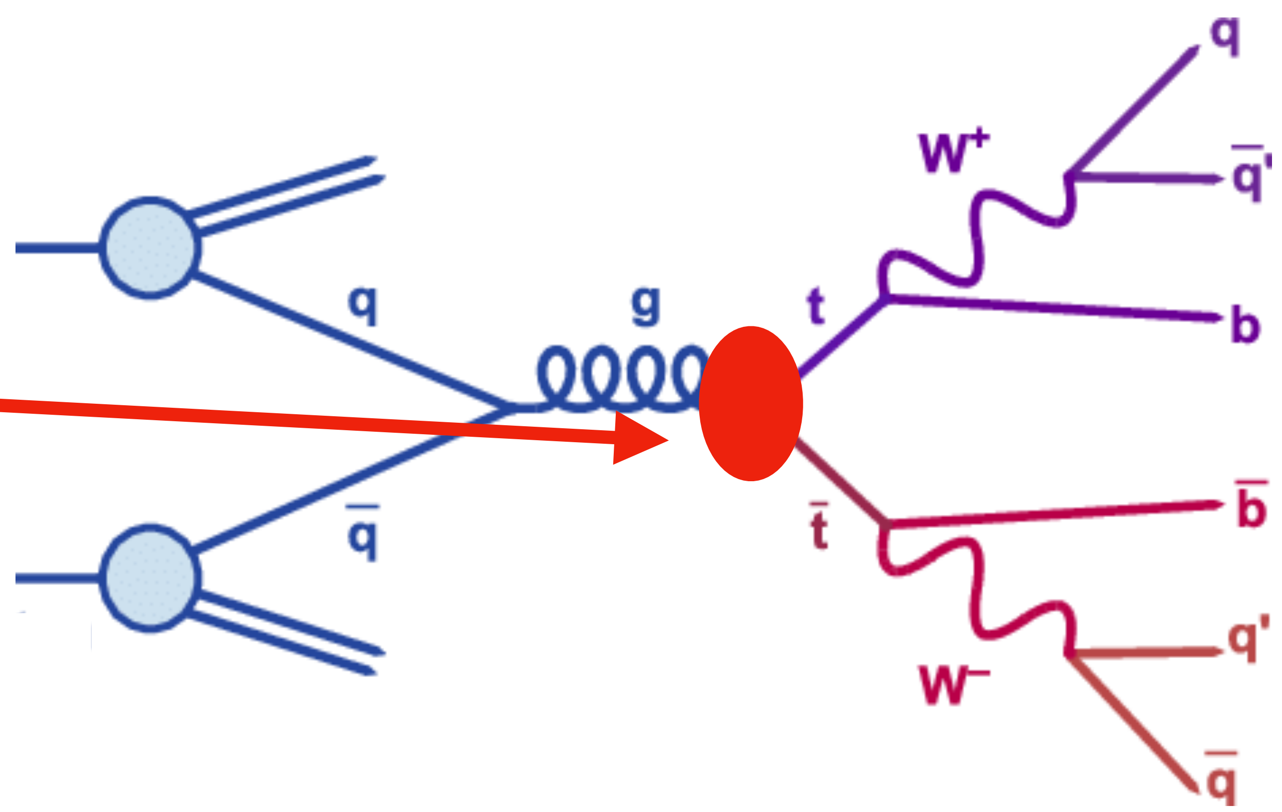


As more precise data and SM theoretical predictions become available, we can identify patterns of small differences induced by new physics and from there deduce what is the new model that causes a given pattern!

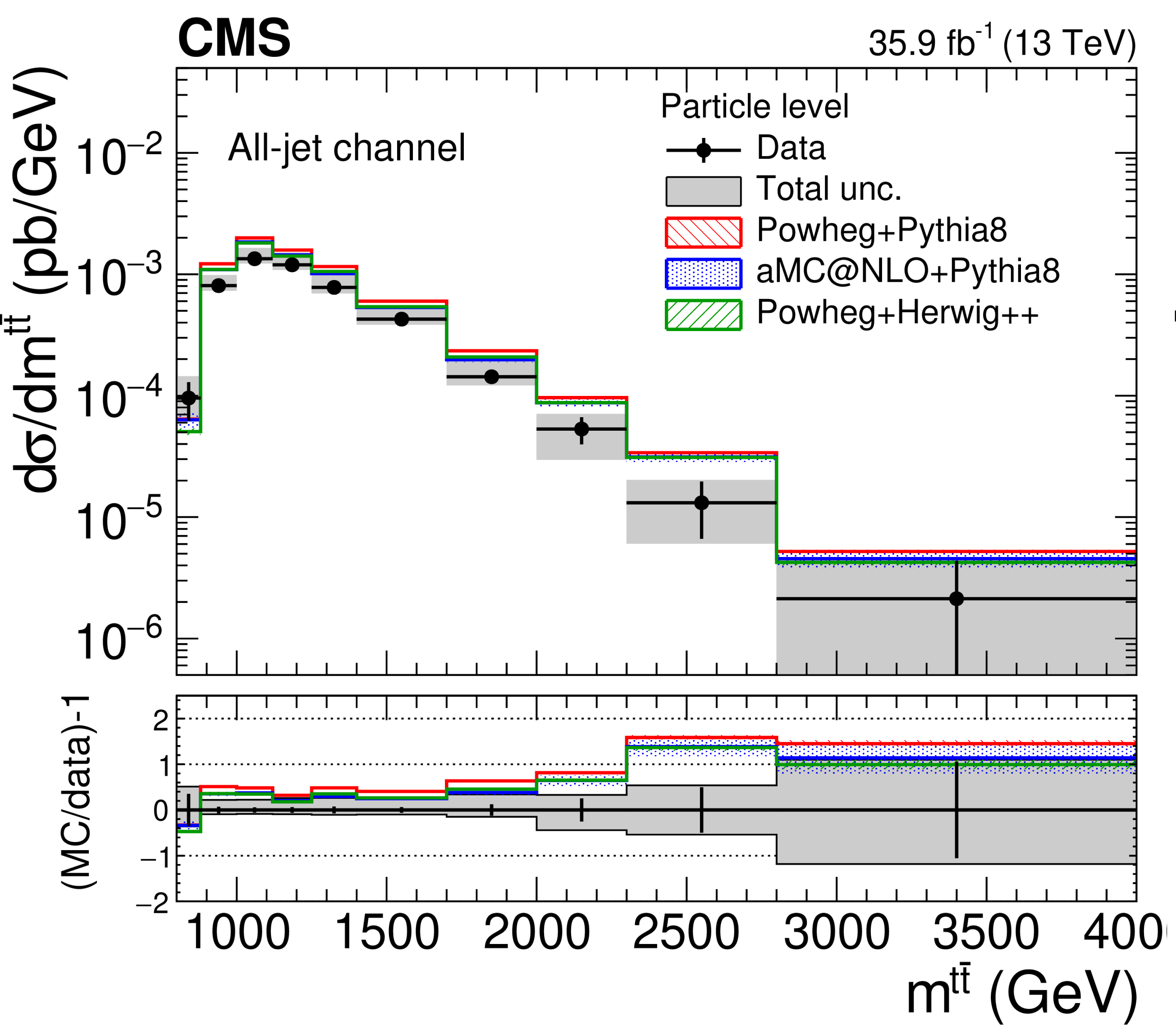


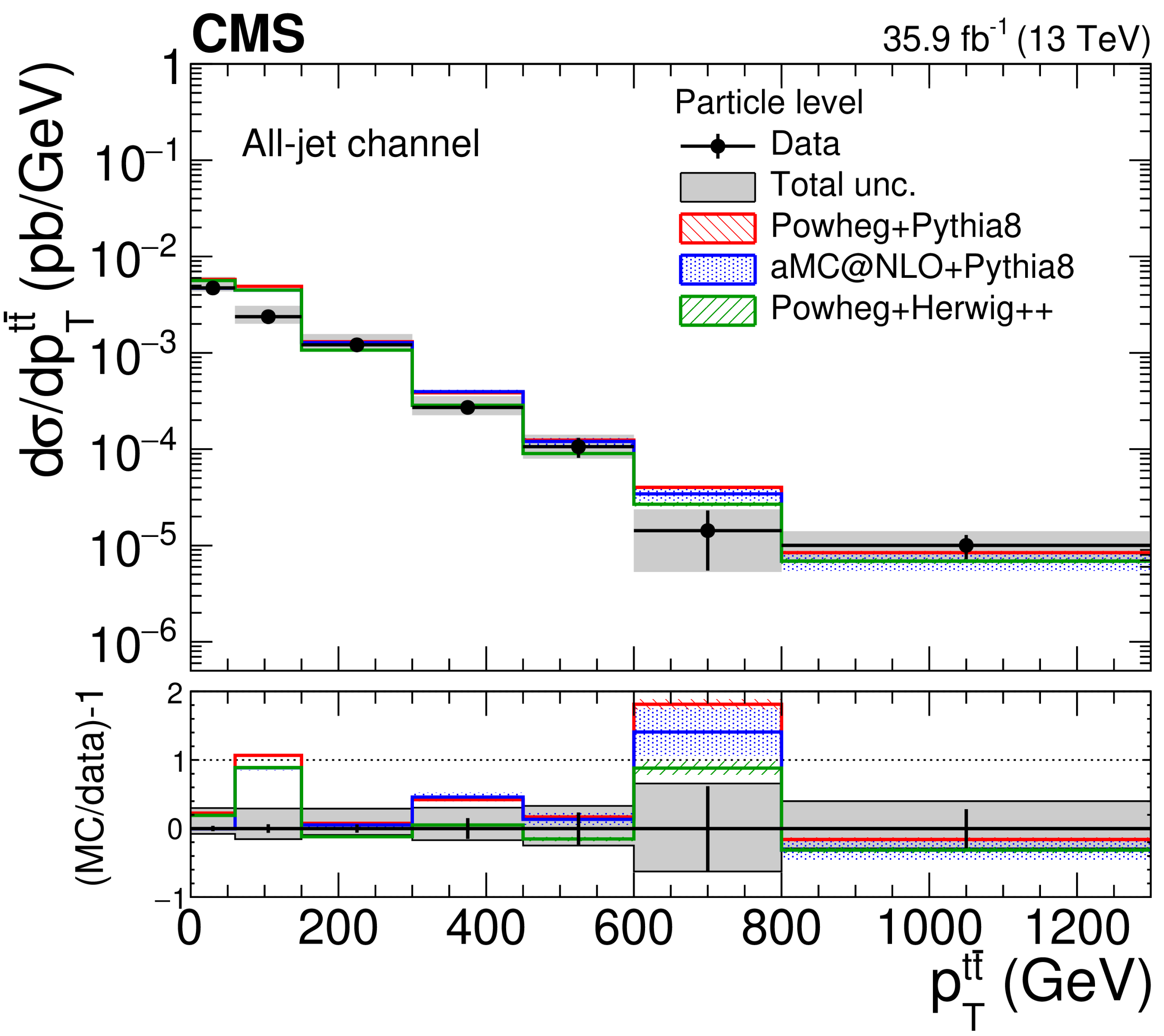
(4) ML & PARAMETER DETERMINATION

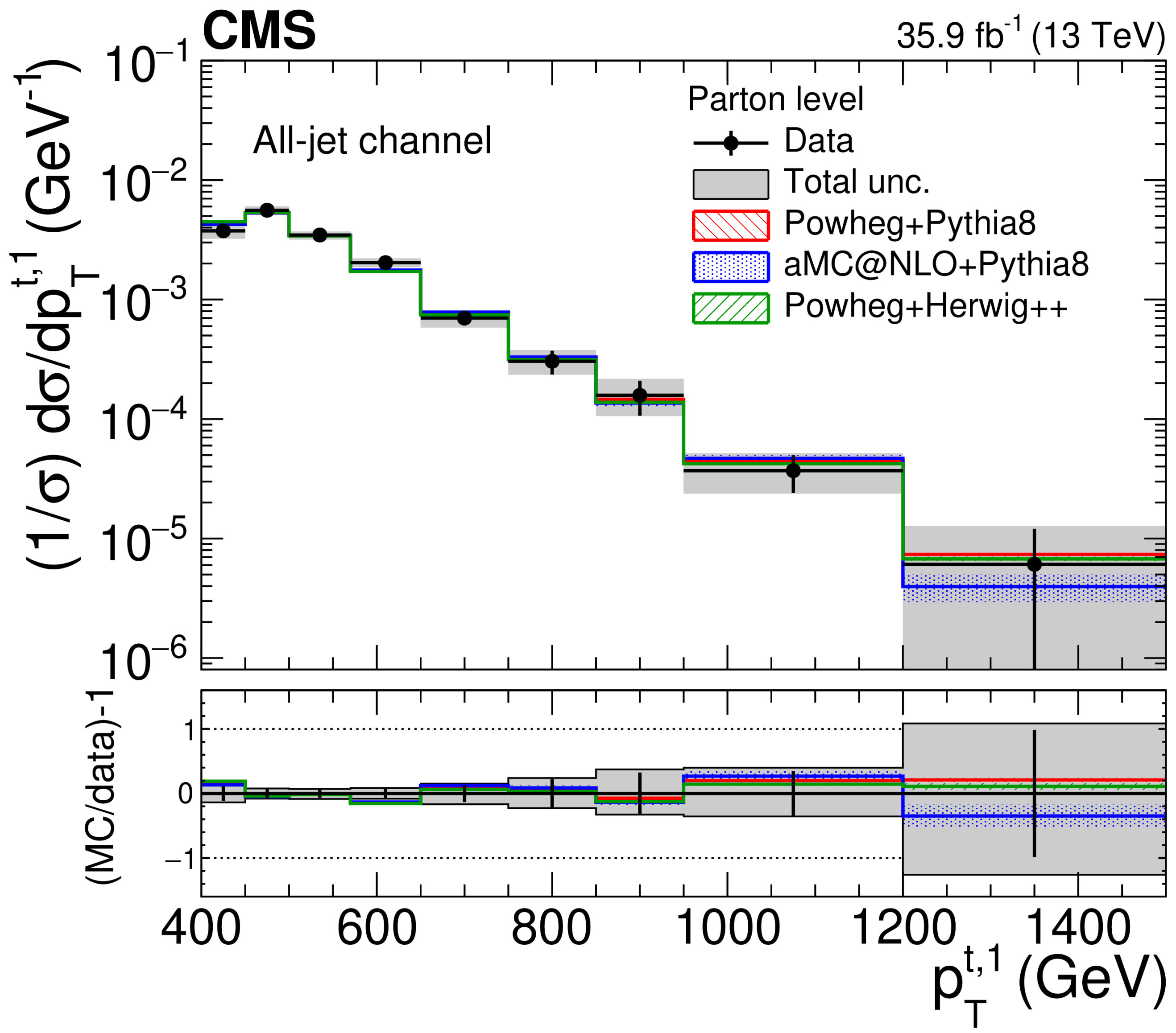
- Is there any sign of BSM physics (non-zero Wilson coefficient) in the gluon-top coupling?
- Which measurement is the most sensitive to these parameters?
- Inclusive, single to multi-differential, which variables, which binning?



CMS collaboration arXiv:2008.07860

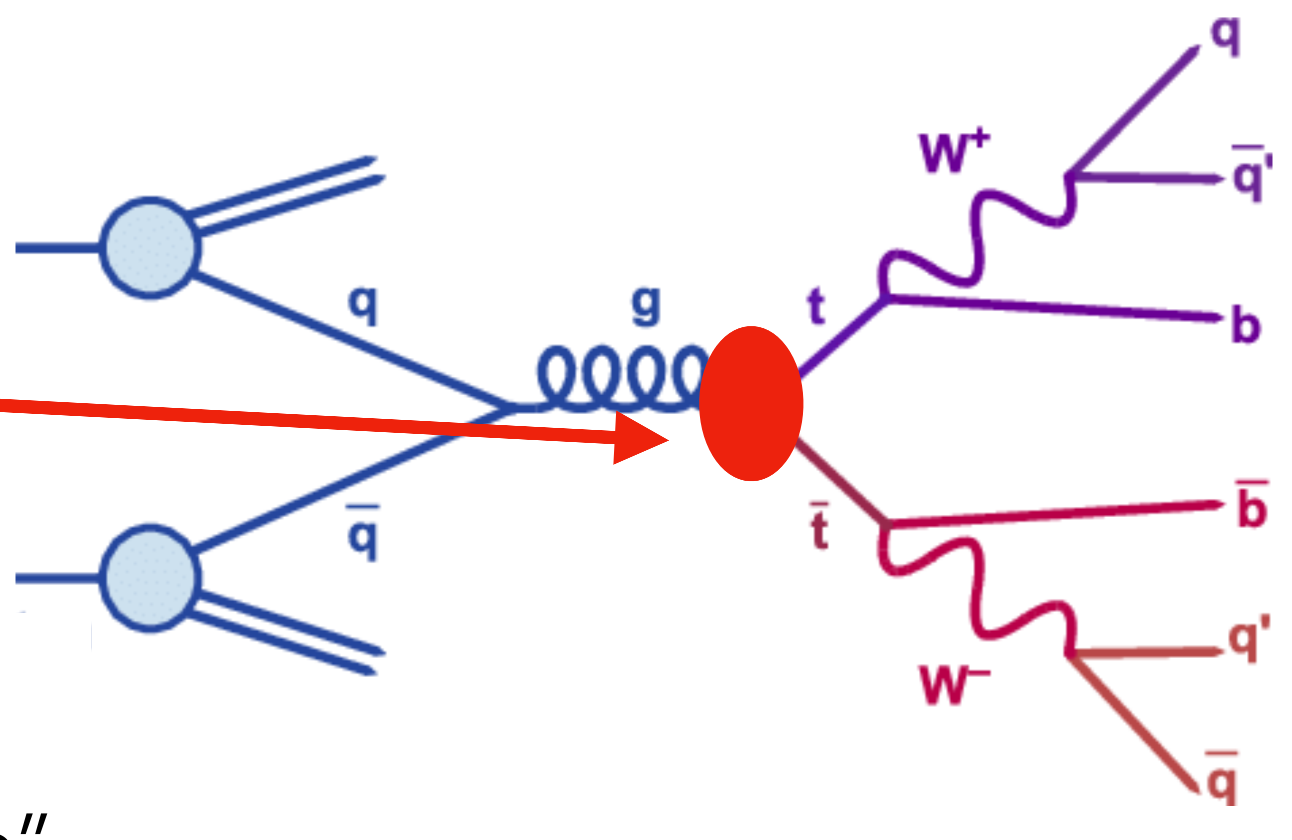






plus many other distributions…

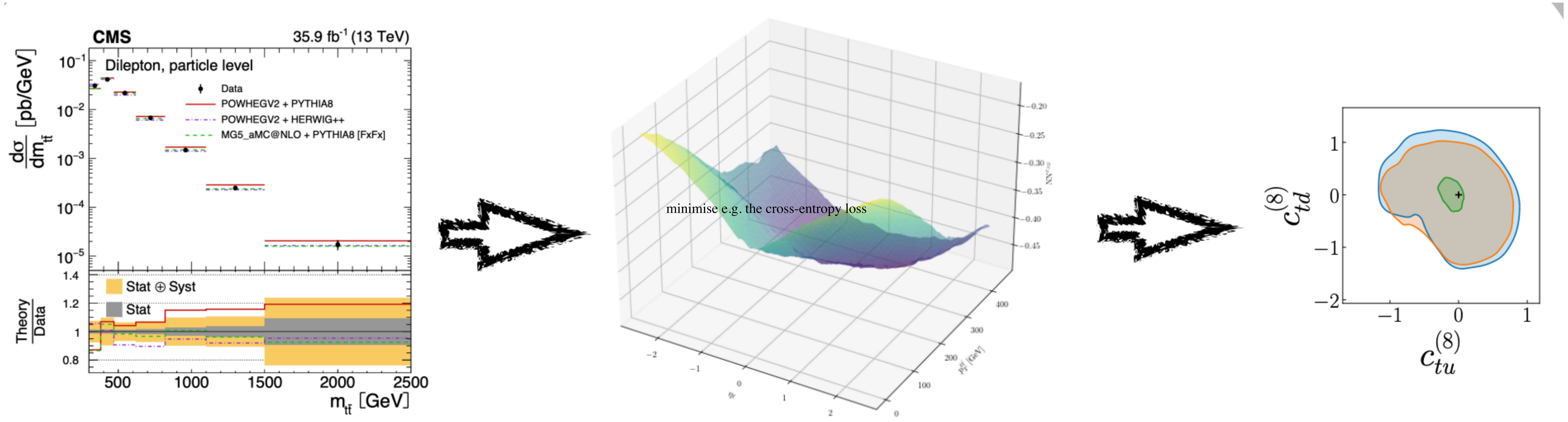
(4) ML & PARAMETER DETERMINATION



- Is there any sign of BSM physics (non-zero Wilson coefficient) in the gluon-top coupling?
- Which measurement is the most sensitive to these parameters?
- Inclusive, single to multi-differential, which variables, which binning?
- **Humans** can only visualise things in one/two variables but **ML tools** can "see" in an arbitrarily large number of dimensions and find optimal variables and binning to "see" new physics

$$\boxed{r(x_i;\mu_1,\mu_0) = \frac{p(x_i|\mu_1)}{p(x_i|\mu_0)}}$$

- Generate likelihood (parton level) for each phase space point μ_1: SMEFTI vs μ_0: SM
- Train NN to approximate ratio and compute signal strength
- Example of Neural Simulation Based Inference [Gomez Ambrosio et al, arXiv:2211.02058]



Binned, univariate

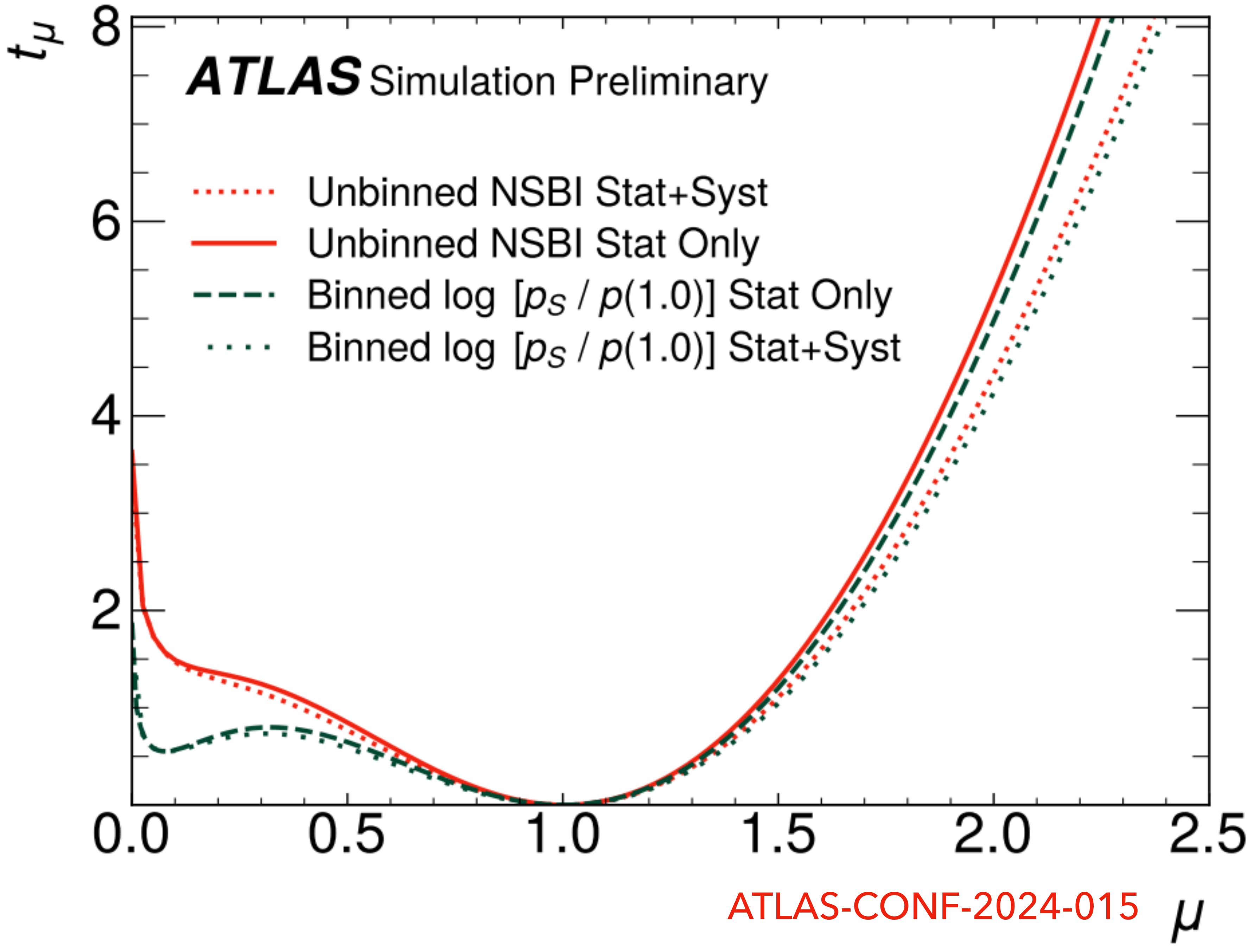
Unbinned, multivariate

Optimal inference

J. ter Hoeve, Cometa General Meeting 2024

(4) ML & PARAMETER DETERMINATION

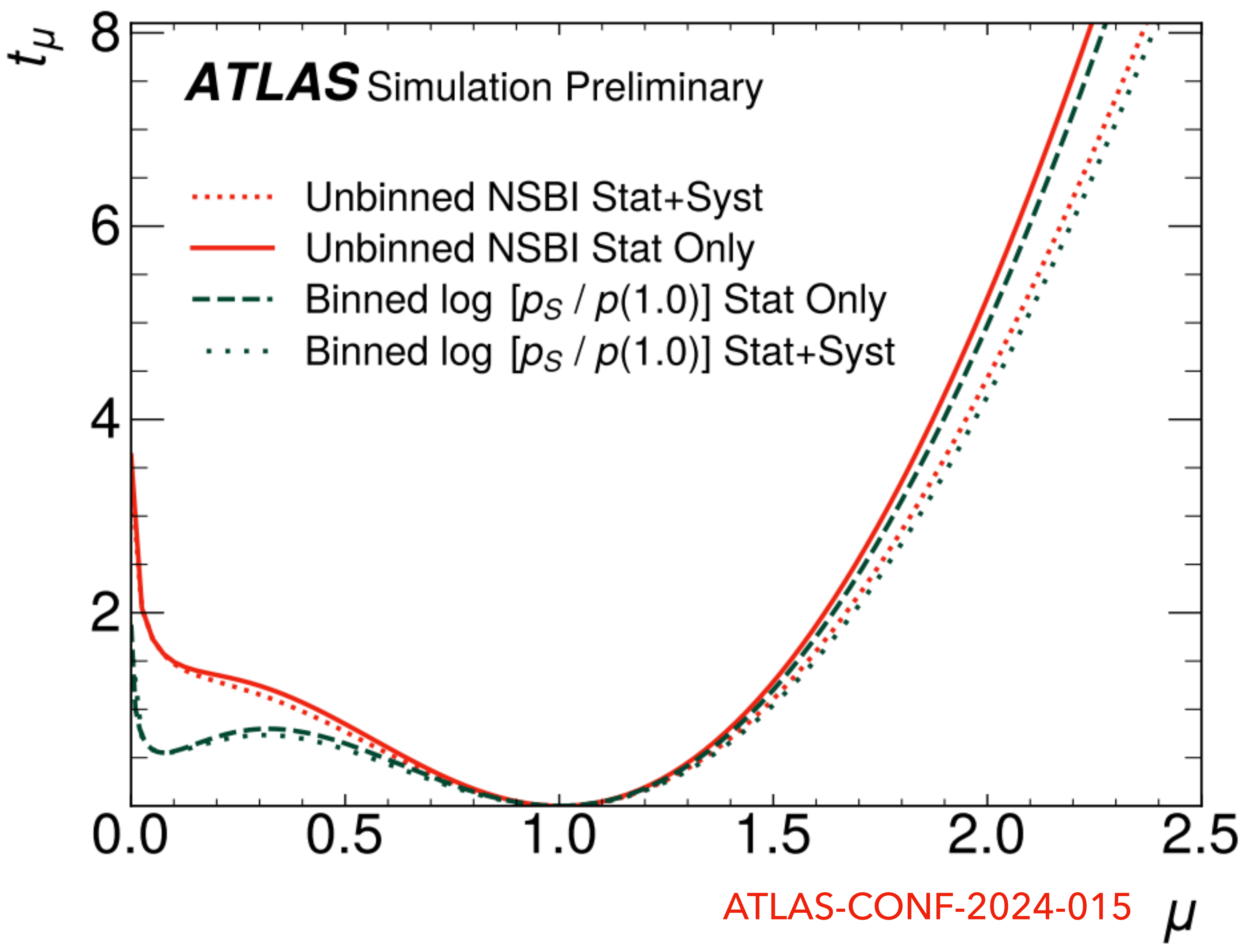
- Similar works
  - ➡ MadMiner series [J.Brehmer, K.Cranmer, G.Louppe … 1907.10621, 1805.00020, …]
  - ➡ Parametrised classifiers for SMEFT [A. Glioti et al 2007.10356]
  - ➡ Learning the EFT likelihood with tree boosting [R. Schöfbeck et al 2205.12976]
  - ➡ Back to the Formula [A. Butter et al 2109.10414]
  - ➡ Boosted likelihood learning with event reweighing [A. Glioti et al 2308.05704]
  - ➡ Designing Observables for Measurements with Deep Learning [O.Long et al 2310.0871
  - ➡ ATLAS analysis for off-shell Higgs to Z boson signal strength [ATLAS-CONF-2024-015]



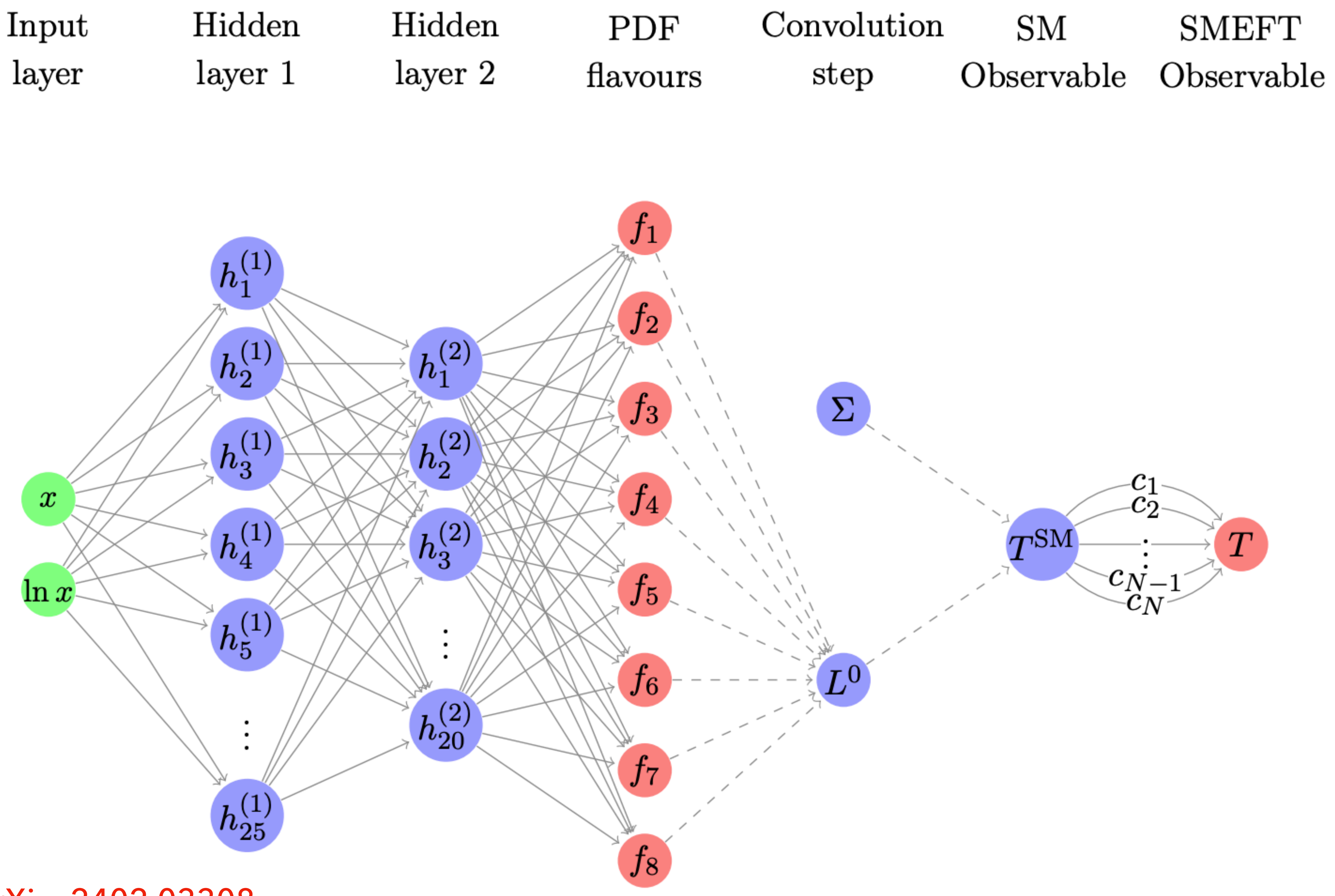
ATLAS-CONF-2024-015

(4) ML & PARAMETER DETERMINATION

- Similar works
  - ➡ MadMiner series [J.Brehmer, K.Cranmer, G.Louppe … 1907.10621, 1805.00020, …]
  - ➡ Parametrised classifiers for SMEFT [A. Glioti et al 2007.10356]
  - ➡ Learning the EFT likelihood with tree boosting [R. Schöfbeck et al 2205.12976]
  - ➡ Back to the Formula [A. Butter et al 2109.10414]
  - ➡ Boosted likelihood learning with event reweighing [A. Glioti et al 2308.05704]
  - ➡ Designing Observables for Measurements with Deep Learning [O.Long et al 2310.0871
  - ➡ ATLAS analysis for off-shell Higgs to Z boson signal strength [ATLAS-CONF-2024-015]



ATLAS-CONF-2024-015



M. Costantini et al arXiv: 2402.03308

- Other directions
  - ➡ Simultaneous fit of SMEFT parameters and PDFs using a deep NN
    https://github.com/HEP-PBSP/SIMUnet
    [Costantini et al, 2402.03308]
  - ➡ Bayesian simultaneous fits of PDFs and parameters with flexible ML framework
    [Colibri, Costantini et al, upcoming]

Part II

- Beyond parton level predictions

MACHINE LEARNING

PARTICLE PHYSICS PHENOMENOLOGY AND THEORY


Quantum theory
Lagrangian
$\mathcal{L}$ (SM, SM + EFT, BSM)
Scattering amplitudes
Phase space integration
PDFs (proton structure)
Parton level
QCD and QED radiation (Parton Showers)
Fragmentation, Hadronisation, Jets
Particle level
Detector simulation
Event reconstruction and selection
Detector level

FROM PARTON TO DETECTORS AND BACK

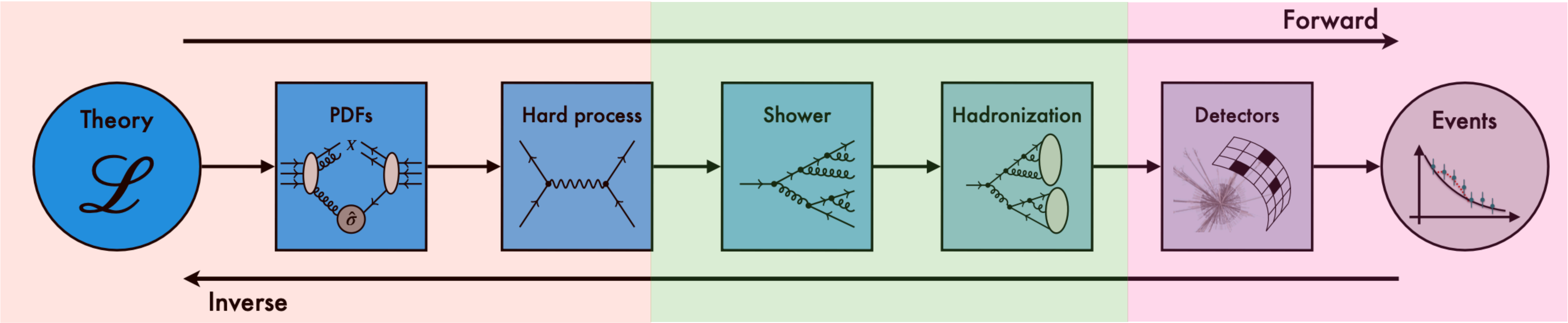


Image Credit
@Plehn et al 2211.01421

- End-to-end ML surrogates for fast HEP event simulations learn multiple steps at once [see Snowmass Report 2203.07460 for a review]

FROM PARTON TO DETECTORS AND BACK

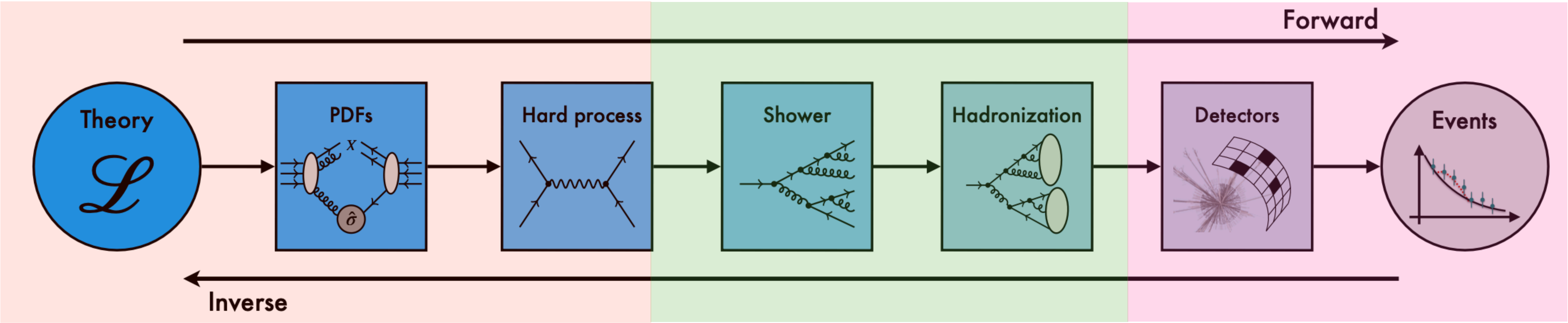


Image Credit
@Plehn et al 2211.01421

- End-to-end ML surrogates for fast HEP event simulations learn multiple steps at once [see Snowmass Report 2203.07460 for a review]
- First attempts based on GANs and VAEs [1901.00875,1901.05282,1903.02433,1907.03764,1912.02748,2001.11103...]
- Improved speed and efficiency with Normalising Flows [2011.13445,2110.13632,2211.13630,2104.04543,...]
- High precision with Diffusion and Transformer models [2303.05376,2305.10475,2307.06836...]
- Key properties: conditional GANs and Transformers allow inversion of simulation chain from detector back to parton level and methods like BNNs and classifiers can be applied for error control [2305.07696...]

FROM PARTON TO DETECTORS AND BACK

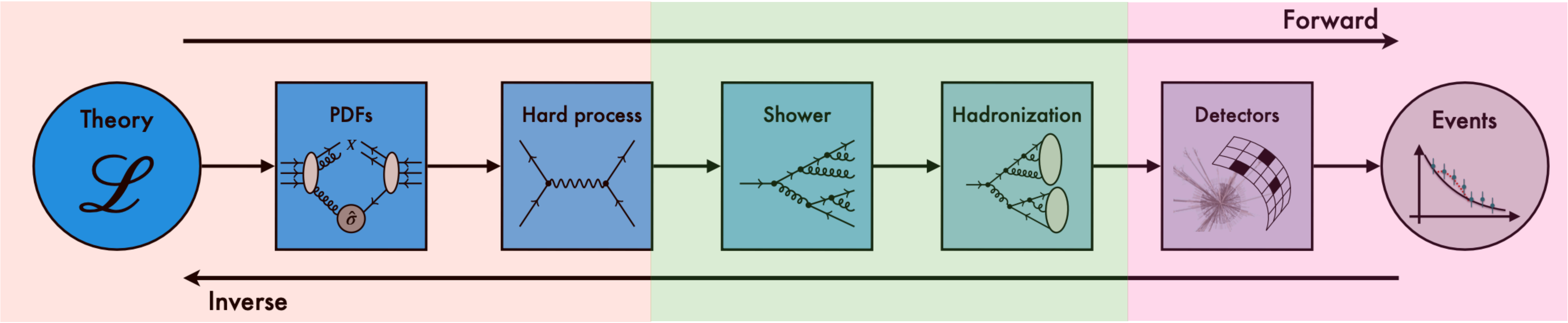
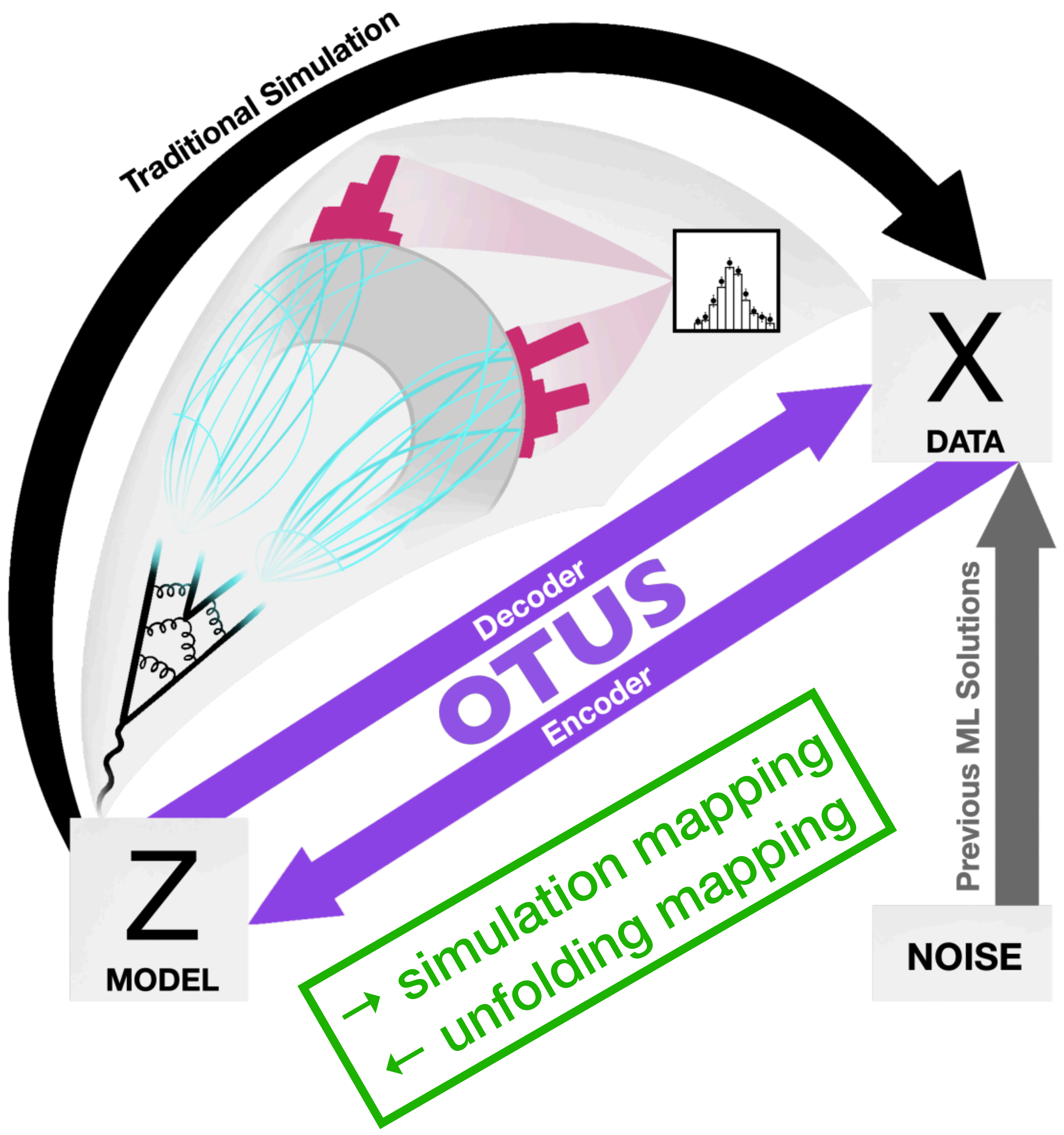


Image Credit
@Plehn et al 2211.01421

- End-to-end ML surrogates for fast HEP event simulations learn multiple steps at once [see Snowmass Report 2203.07460 for a review]
- First attempts based on GANs and VAEs
  [1901.00875,1901.05282,1903.02433,1907.03764,1912.02748,2001.11103…]
- Improved speed and efficiency with Normalising Flows
  [2011.13445,2110.13632,2211.13630,2104.04543,…]
- High precision with Diffusion and Transformer models
  [2303.05376,2305.10475,2307.06836…]
- Key properties: conditional GANs and Transformers allow inversion of simulation chain from detector back to parton level and methods like BNNs and classifiers can be applied for error control [2305.07696…]
- Alternative approach OTUS (Optimal Transport based Unfolding and Simulation) based on probabilistic auto encoders learns mapping from parton level Z to reconstructed objects X without requiring paired event sampling (x,z) [2101.08944…]



Howard et al 2101.08944

FROM PARTON TO DETECTORS AND BACK

Event reconstruction Detector & Particle level Parton level

$$P(D|\vec{\alpha}) = \int dz_D\, dz_S\, dz_P\, p(D|z_D)\, p(z_D|z_S)\, p(z_S|z_P)\, p(z_P|\vec{\alpha})$$

- Traditional analyses from reconstructed events to parton level lose some info as they rely on binned data and "hand-crafted" observables
- Full likelihood intractable but access to it would yield unbinned, multi-varied observables and optimal use of info depending on theory parameter

FROM PARTON TO DETECTORS AND BACK

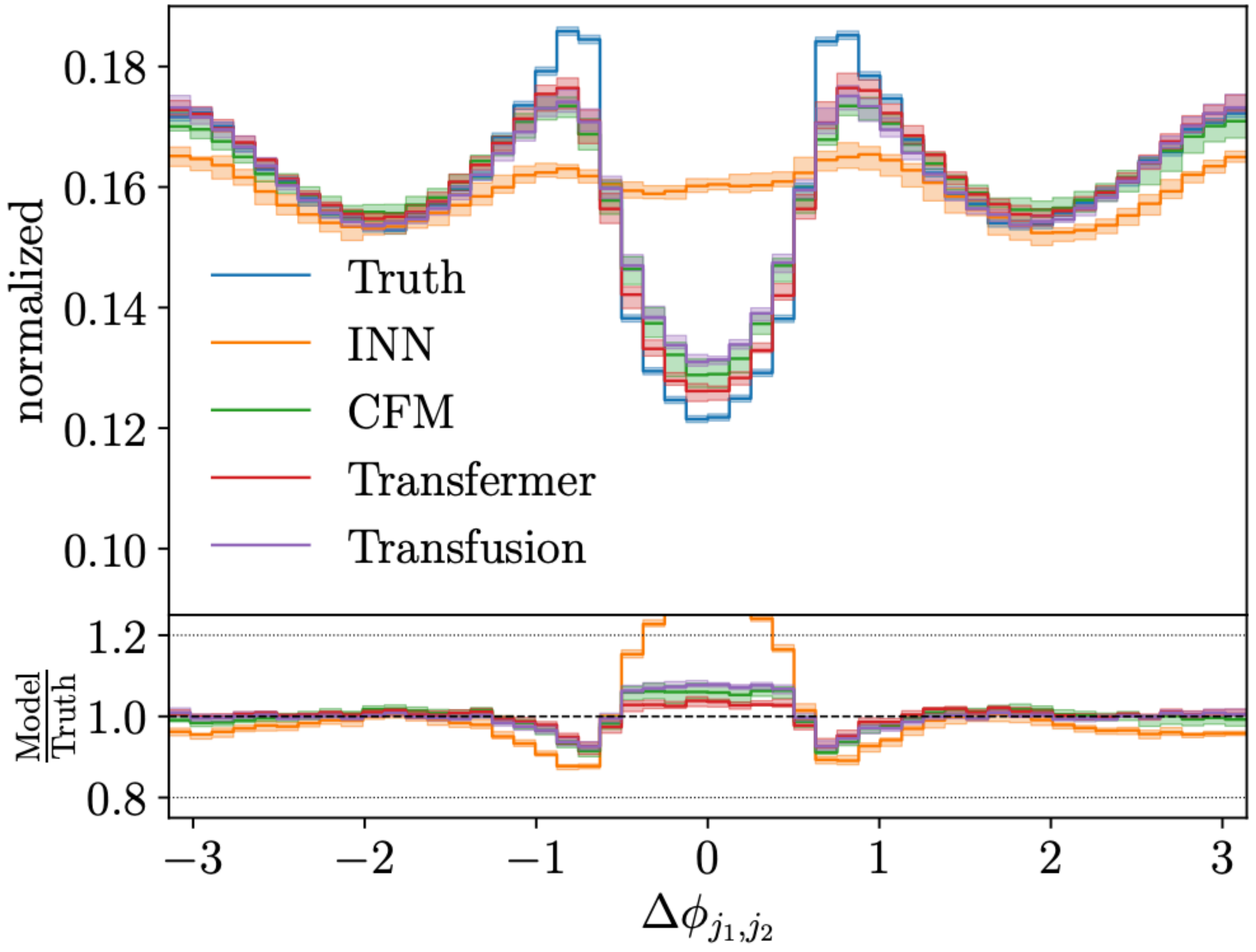
Event reconstruction | Detector & | Particle level | Parton level

$$P(D|\vec{\alpha}) = \int dz_D\, dz_S\, dz_P\, p(D|z_D)\, p(z_D|z_S)\, p(z_S|z_P)\, p(z_P|\vec{\alpha})$$

Efficient Monte Carlo sampling | Classifier Networks | Transfer Function (Normalising Flows + Transformers) | Theory input

- Traditional analyses from reconstructed events to parton level lose some info as they rely on binned data and "hand-crafted" observables
- Full likelihood intractable but access to it would yield unbinned, multi-varied observables and optimal use of info depending on theory parameter
- Matrix Element Method (MEM) builds likelihood using matrix elements from theory and transfer functions & allows to infer fundamental parameters of directly from reconstructed events
  - ➡ **Transfer Function** (from z_P to z_D) for the intractable part of the likelihood
  - ➡ **Classifier** to parametrise acceptance probability
  - ➡ **Generative NN** to integrate

while using theory input for the parton level event [2310.07752,2210.00019…]



Heimel et al 2310.07752

MACHINE LEARNING

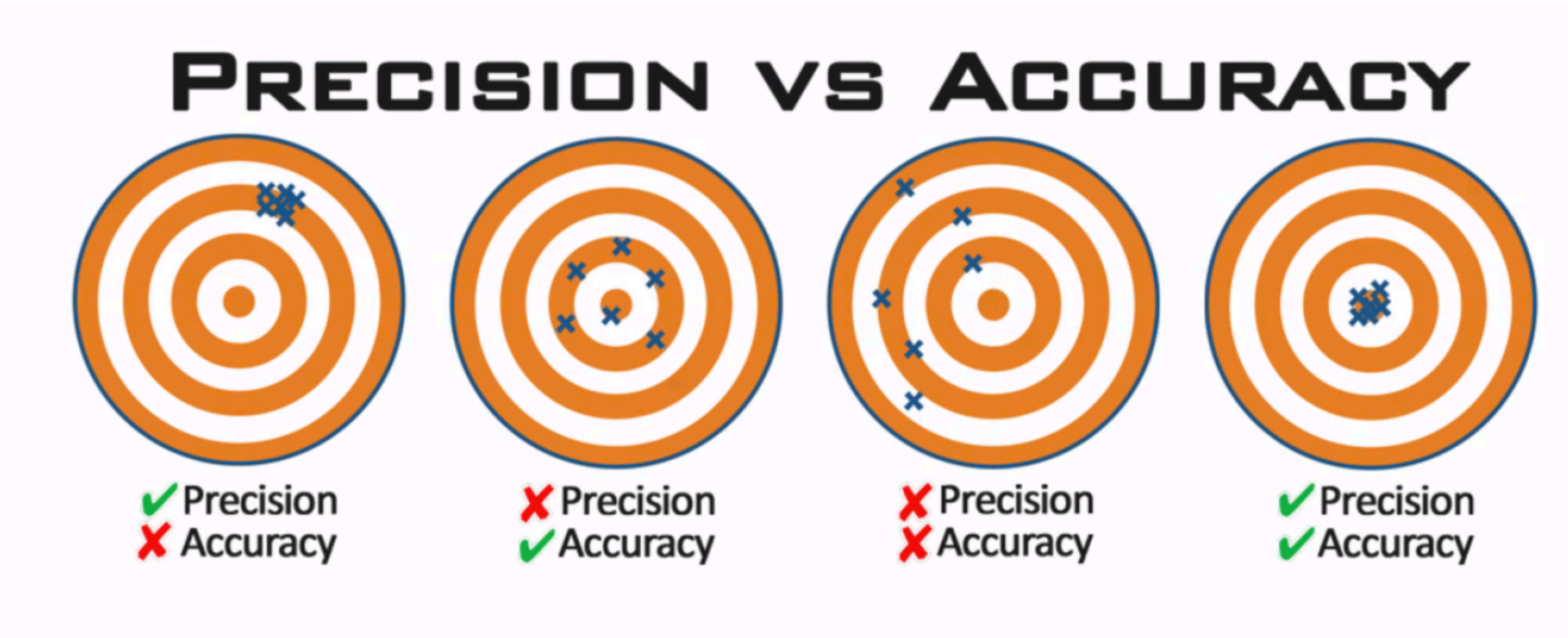
PARTICLE PHYSICS
PHENOMENOLOGY AND THEORY

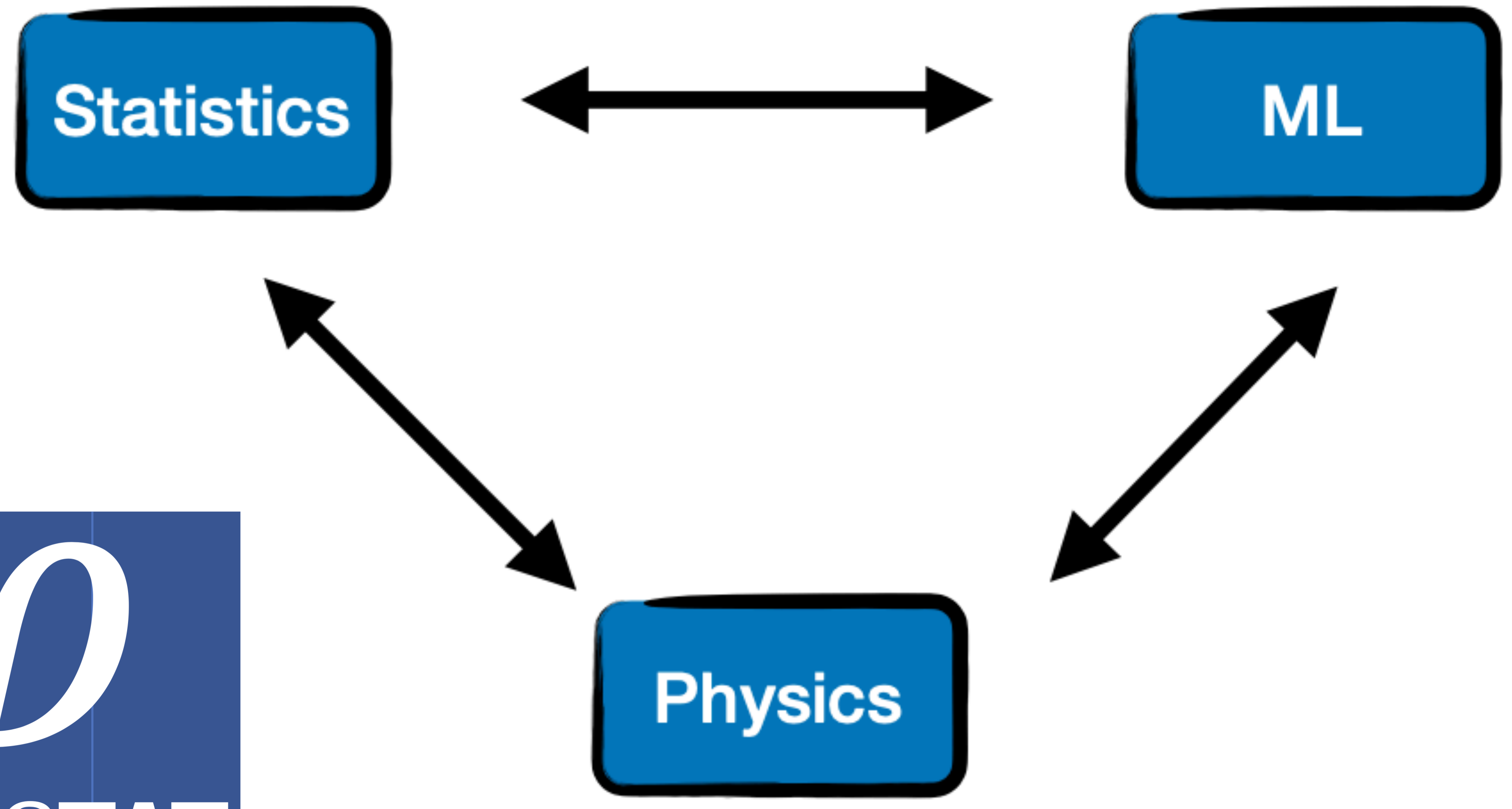
- Uncertainty quantification
- Symmetries
- Interpretability

Part II

(I) UNCERTAINTY QUANTIFICATION

- Modern Machine Learning essential in particle physics phenomenology and theory and will be used more and more in HEP theory chain
- Do these tools provide optimal and resilient results including a **comprehensive uncertainty treatment** that makes them both precise and accurate?





- HEP can be at the forefront of development from deterministic ML to probabilistic ML

Input x → Fixed output y

Input x → Probabilistic output p(y|x)

- Seen several approaches in a number of contexts: bootstrapping, heteroscedastic losses, Bayesian Neural Networks, Repulsive Ensembles, Posterior Sampling
- Key to understand various sources of uncertainties, benchmark uncertainty quantifications, understand dependence on priors, perform statistical tests [VERaIPHY initiative - stay tuned!]

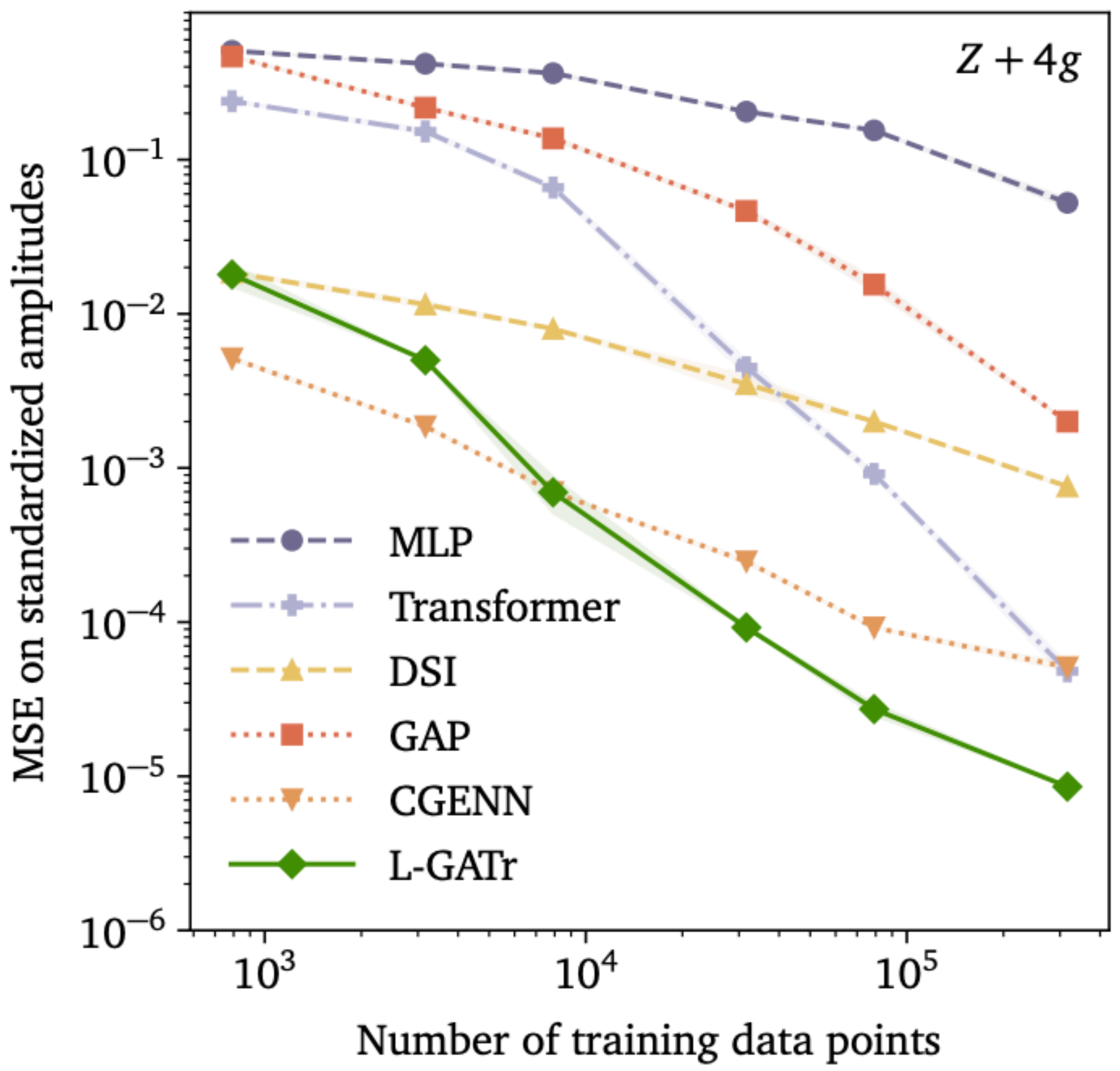
L. Heinrich
PhyStat: Statistics meets Machine Learning 2024

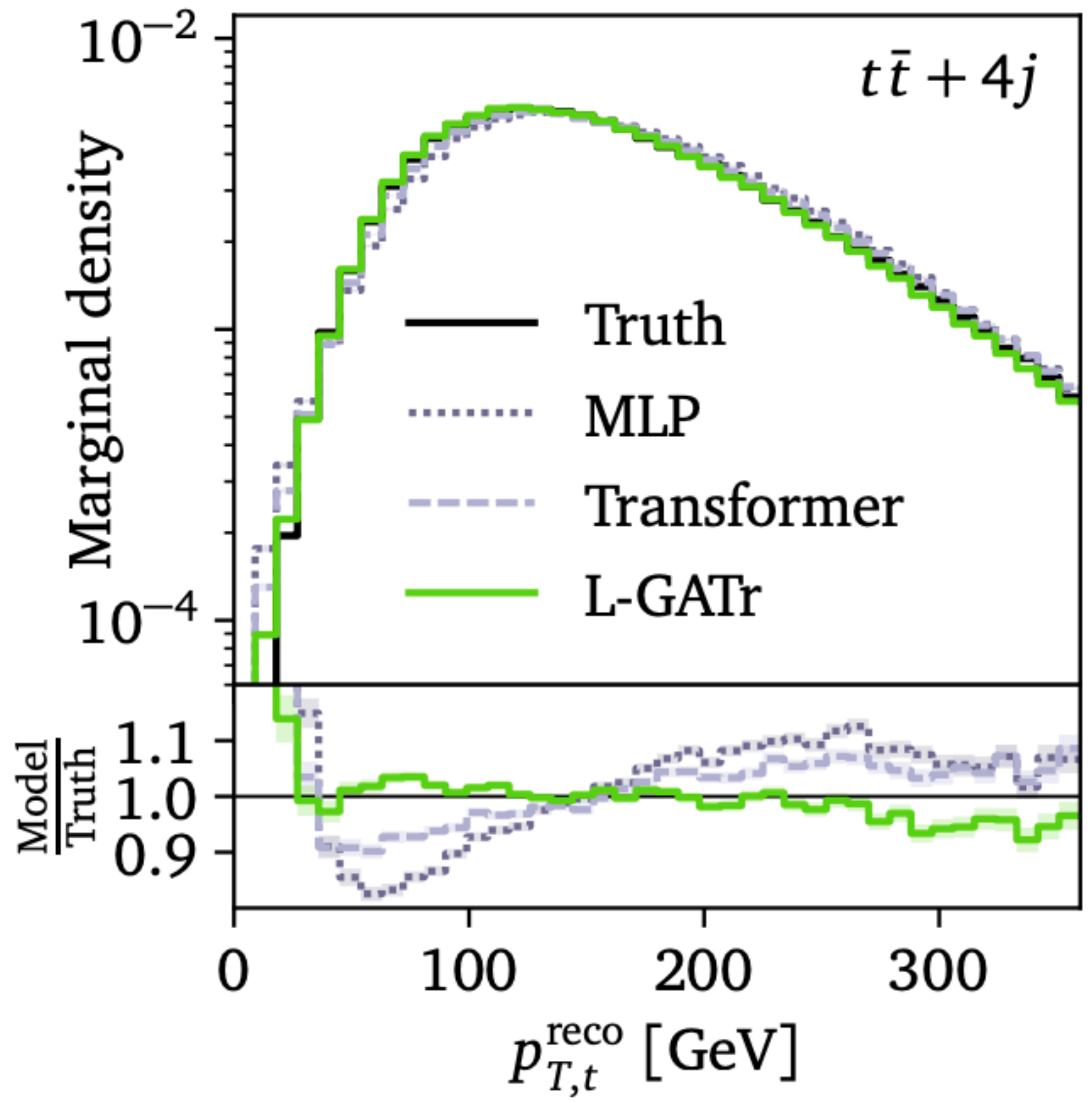
(II) SYMMETRY MEETS MACHINE LEARNING

Symmetries → Machine Learning

- In physics we know that nature follows some symmetries (Lorentz-invariance, local gauge symmetries defining QFT…)
- HEP can provide a smart way to force theory into ML architecture (smart inductive bias)
- For example, Lorentz-equivariant transformers can provide appropriate internal or latent representation of the phase space points x to NN in several contexts (regression, classification, generation…) by encoding Lorentz-equivariance into architecture. [Brehmer et al 2305.18415, De Haan et al 2311.04774, Spinner et al 2405.14806, Brehmer et al 2411.00446]

Amplitude regression
$qq \to Z + 4g$

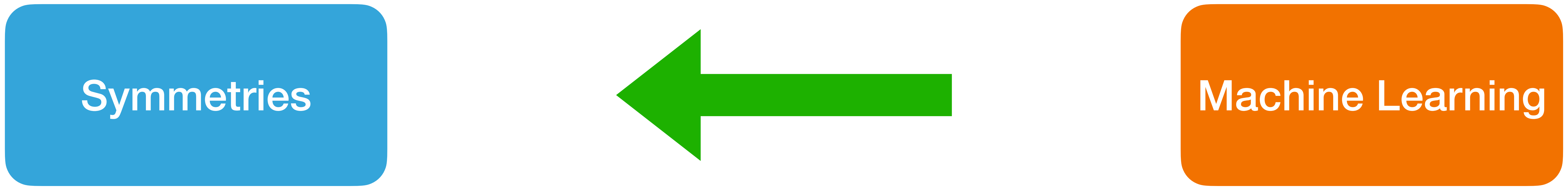




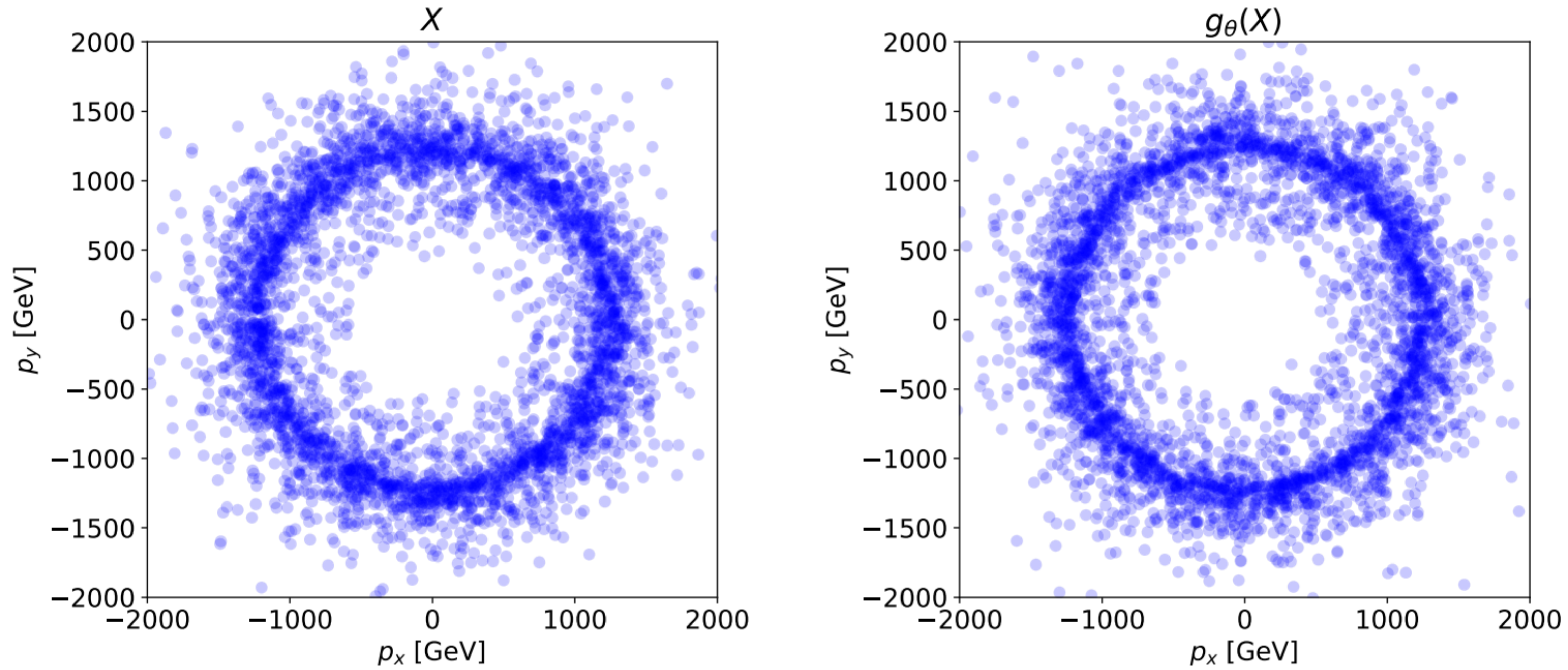
Event generation
$pp \to tT + 4j$

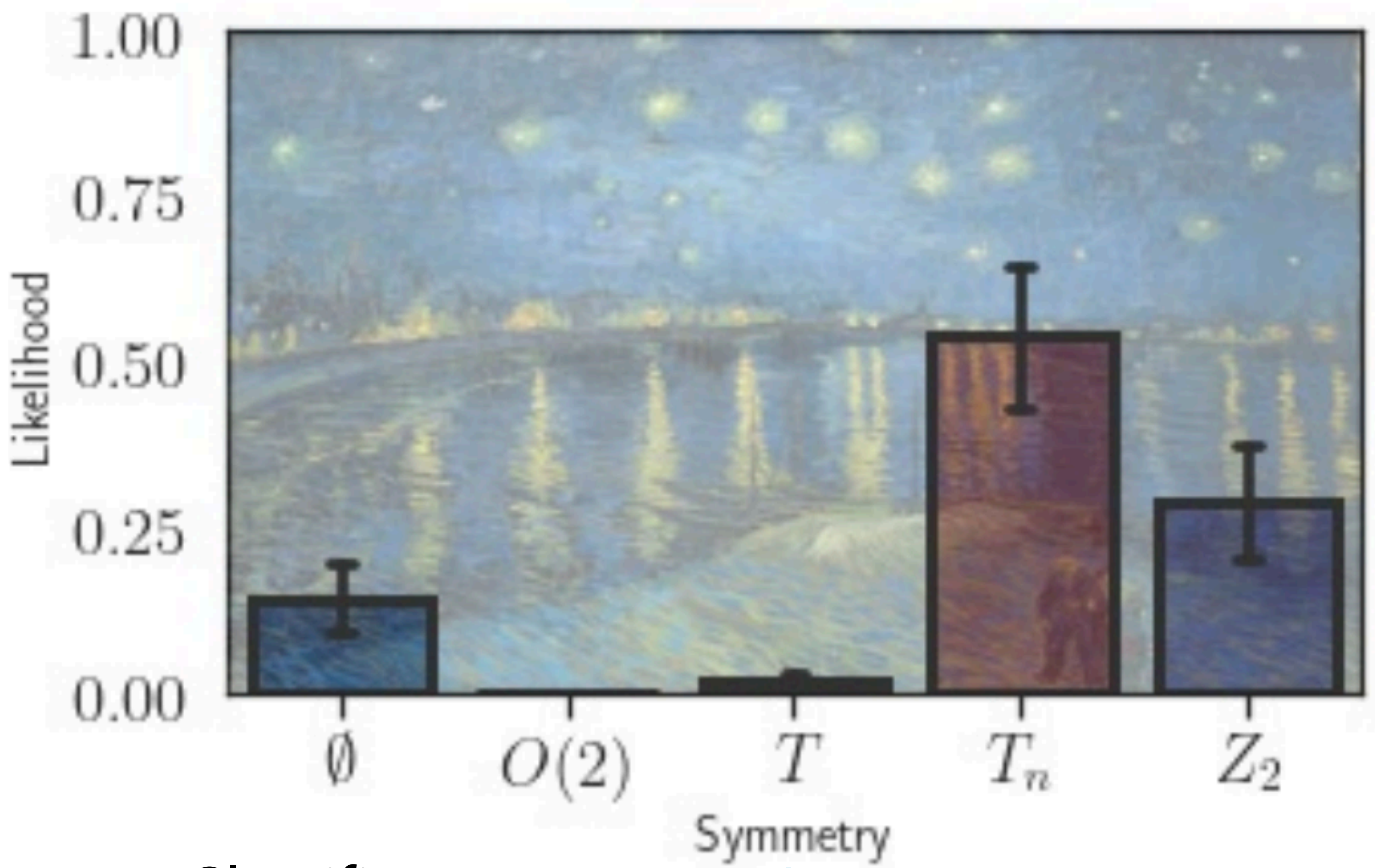
Bremer et al, 2411.00446

(II) SYMMETRY MEETS MACHINE LEARNING



- Discovery of a symmetry signifies the existence of a fundamental principle and manifests itself in the form of physical laws and selection rules
- Detecting symmetries with neural networks is exciting frontier in theoretical physics, seminal efforts with classifiers and Symmetry GAN explored as deep learning approach to discover symmetry [Betzler + Krippendorf 2002.05169, Krippendorf et al 2003.13679, Barenboim, Hirn, Sanz 2103.06115; Desai et al 2112.05722]





Desai et al 2112.05722

Symmetry GAN

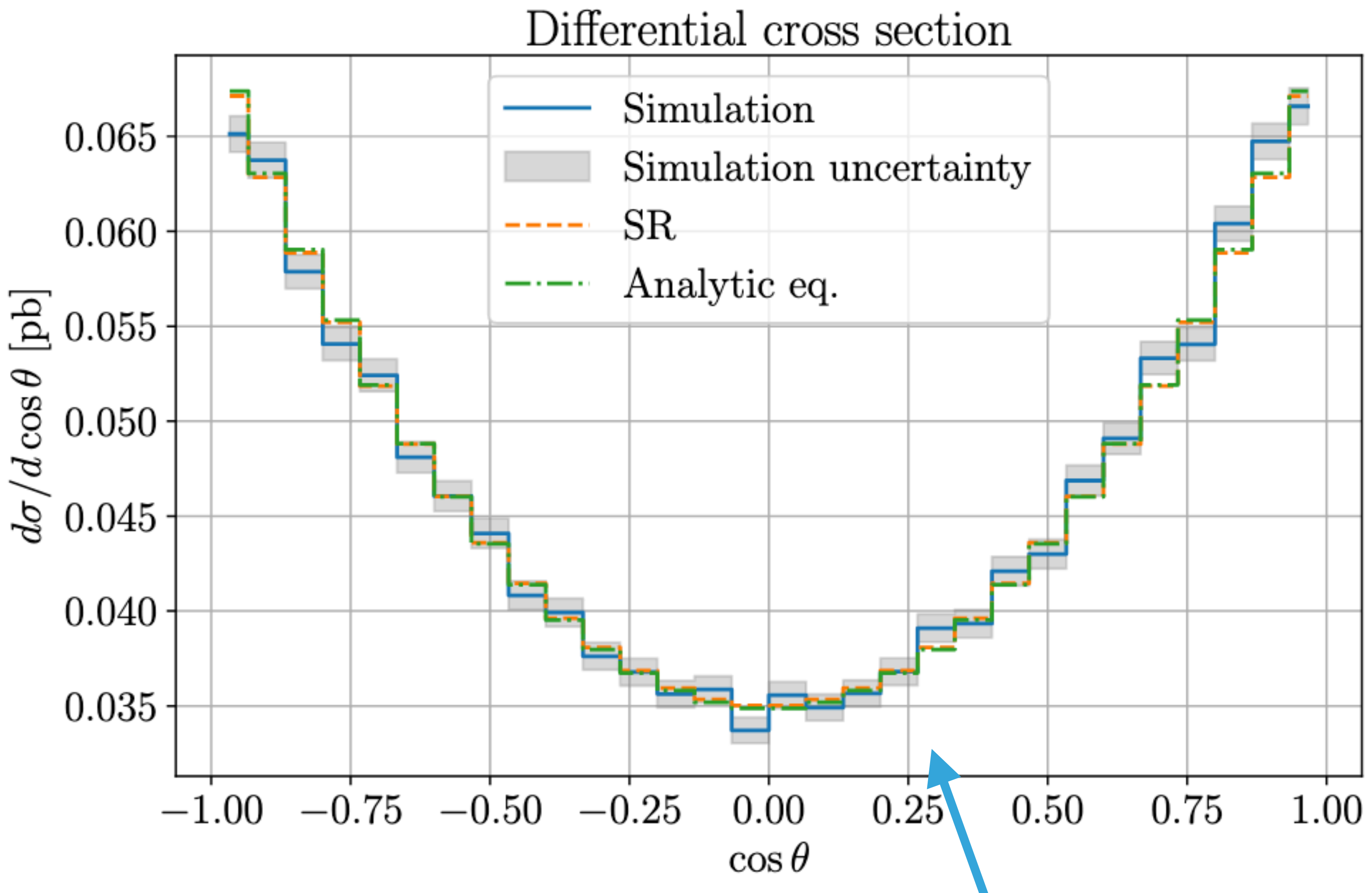
Classifiers

Barenboim, Hirn, Sanz 2103.06115

(III) INTERPRETABILITY

- Physicists want to understand and strive for simplicity and unity (Identify low-rank structures in high-dimensional datasets [J. Thaler @ PhyStat 2024]
- Not all problems require analytical formulae, in some cases numerical simulations and **"black-box" NN** outputs are fine as long as we qualitatively and quantitatively assess sources of statistical and systematic uncertainties
- But formulae help (e. g. better at extrapolation): **symbolic regression** combines benefits of ML and analytic formulas by learning complex functions from low-level or high-dimensional data and expressing them analytically. [Butter Plehn 2109.10414, Lu et al 2210.02184, Tsoi et al 2411.09851, Morales-Alvarado et al 2412.07839, Singireddy et al 2504.13289, Makke et al 2501.07123…]
- See also understanding of DNN using principles of QFT [2402.13321, 2408.00082…] or cosmological dynamics [Krippendorf, Spannowsky 2202.11104…]

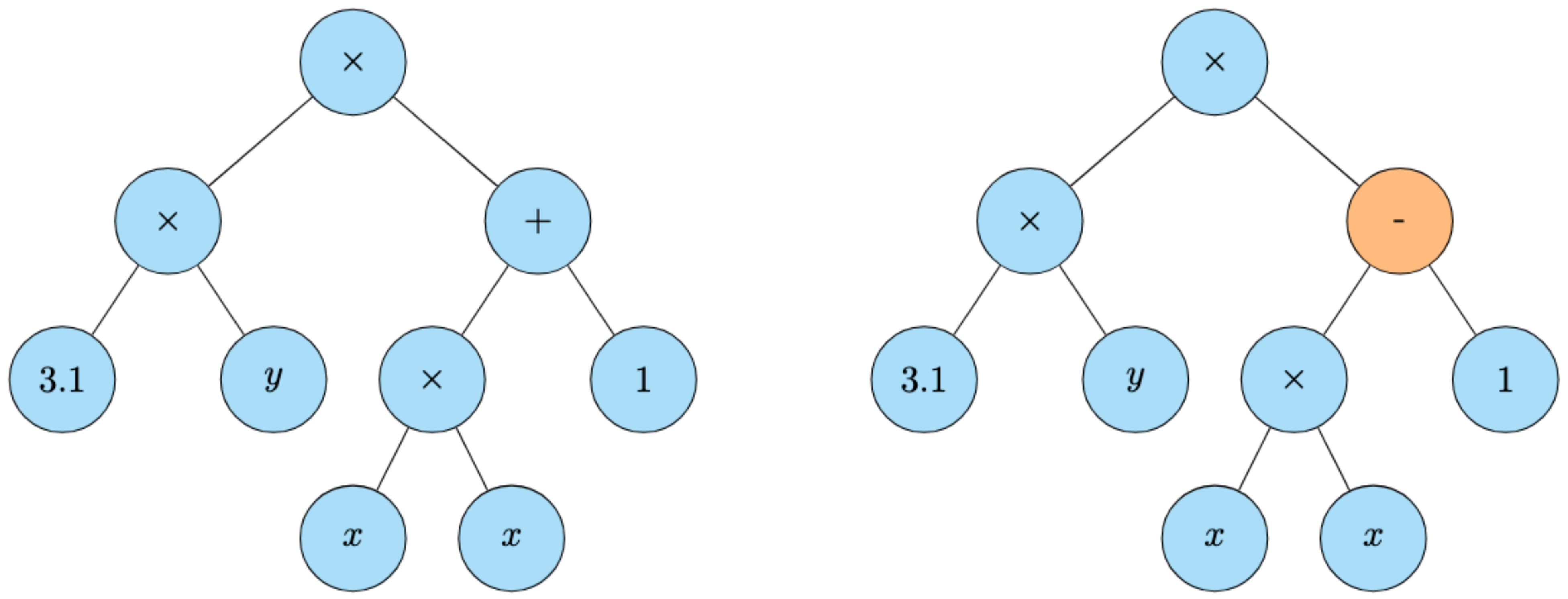
Using PySR [Cranmer 2305.01582] a multi-population evolutionary algorithm that evaluates symbolic expressions, determine angular coefficients in Drell-Yan events at the LHC



$e^+e^- \to \mu^+\mu^-$

$$\boxed{\frac{d\sigma}{d\cos\theta} = \sigma_0(1+\cos^2\theta)}$$

Example of expression tree and one mutation
3.1 * y * (x^2 + 1) → 3.1 * y * (x^2 - 1)



Morales-Alvarado et al, arXiv 2412.07839 + 2506.xxxxx

CONCLUSIONS AND OUTLOOK

- Revolution of LHC physics through modern machine learning (ML) is happening right now!

→ Modern Machine Learning combining regression, classification generation and conditional generation, are beneficial in every step of theory predictions at the LHC.

← Data quantity & of quality more advanced than in most of other contexts (control on systematic uncertainties and correlations) gives labeled and well understood data, HEP perfect playground for ML

Frontiers
- Uncertainty quantification
- ML helps us make the most out of data: how to transition from data science to symbolic problems?
- FunSearch meets Theoretical Physics? Large Physics Model?

No time to talk about
- *Exciting ML applications and development in lattice QCD*
- *ML and string theory*
- *Unfolding and specific parton level tools*
  [see for example 2504.18126, 2502.02670,2403.03245,2410.21611…]

CONCLUSIONS AND OUTLOOK

- Revolution of LHC physics through modern machine learning (ML) is happening right now!

→ Modern Machine Learning combining regression, classification generation and conditional generation, are beneficial in every step of theory predictions at the LHC.

← Data quantity & of quality more advanced than in most of other contexts (control on systematic uncertainties and correlations) gives labeled and well understood data, HEP perfect playground for ML

Frontiers

- Uncertainty quantification
- ML helps us make the most out of data: how to transition from data science to symbolic problems?
- FunSearch meets Theoretical Physics? Large Physics Model?

No time to talk about

- *Exciting ML applications and development in lattice QCD*
- *ML and string theory*
- *Unfolding and specific parton level tools*
  [see for example 2504.18126, 2502.02670,2403.03245,2410.21611…]

"We can only see a short distance ahead, but we can see plenty there that needs to be done."
– Alan Turing