Optimally sensitive observables for global EFT fits

HIGGS 2021 online conference
October 19th

Jaco ter Hoeve

Work done in collaboration with R. Gomez Ambrosio, M. Madigan, J. Rojo, V. Sanz

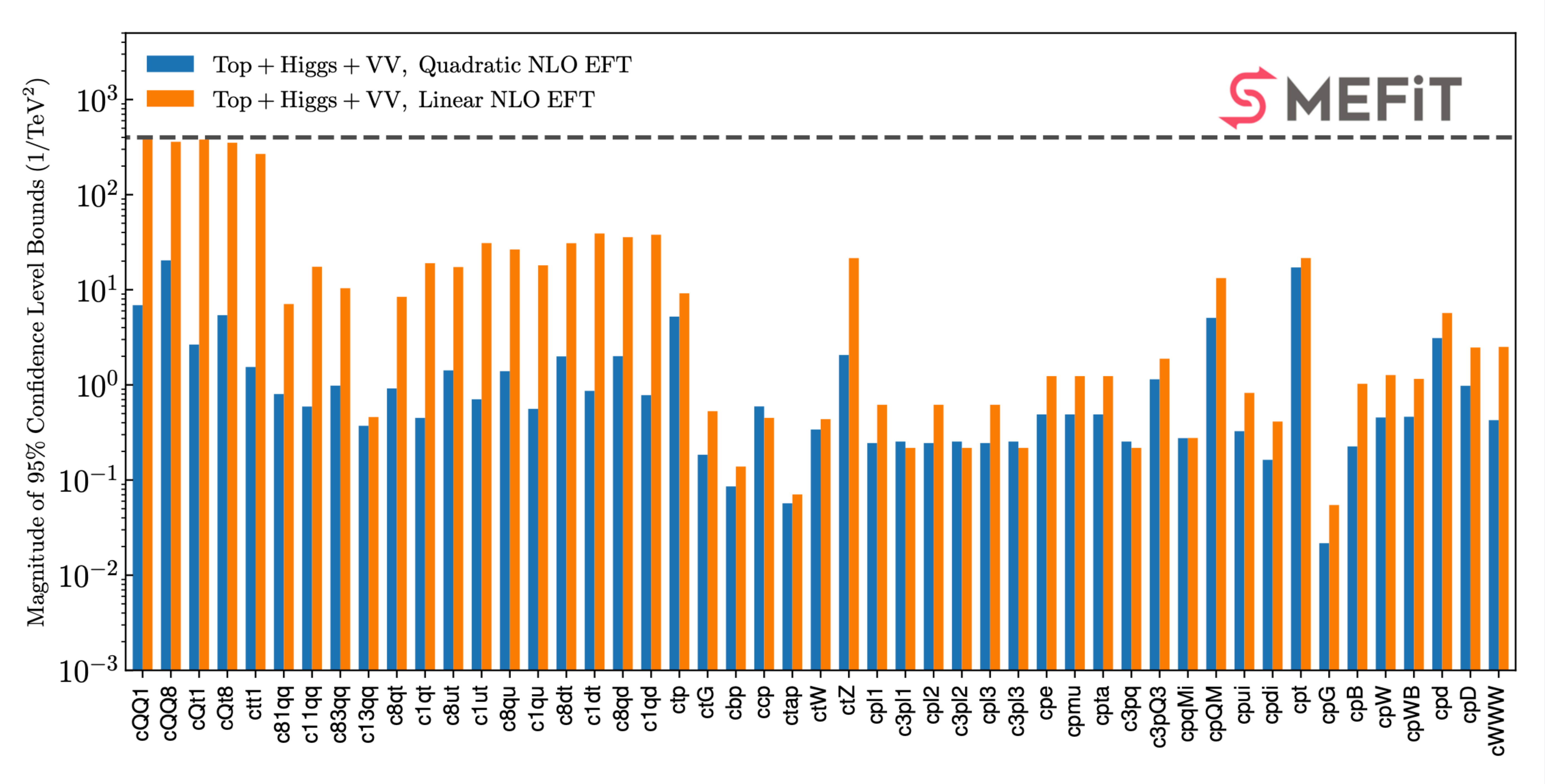
Nikhef

VU VRIJE UNIVERSITEIT AMSTERDAM

NWO

Introduction

- Status of the global EFT program: **Top + Higgs + diboson** data
- Based on traditional unfolded cross section distributions
- Can one construct observables **specifically designed to constrain EFT operators**?



See also the next talk by Ken Mimasu (Fitmaker)!

J.J. Ethier et al. [2105.00006]

Introduction

Key question: given a collider process, how can one define **optimal observables** with the highest sensitivity to EFT coefficients?

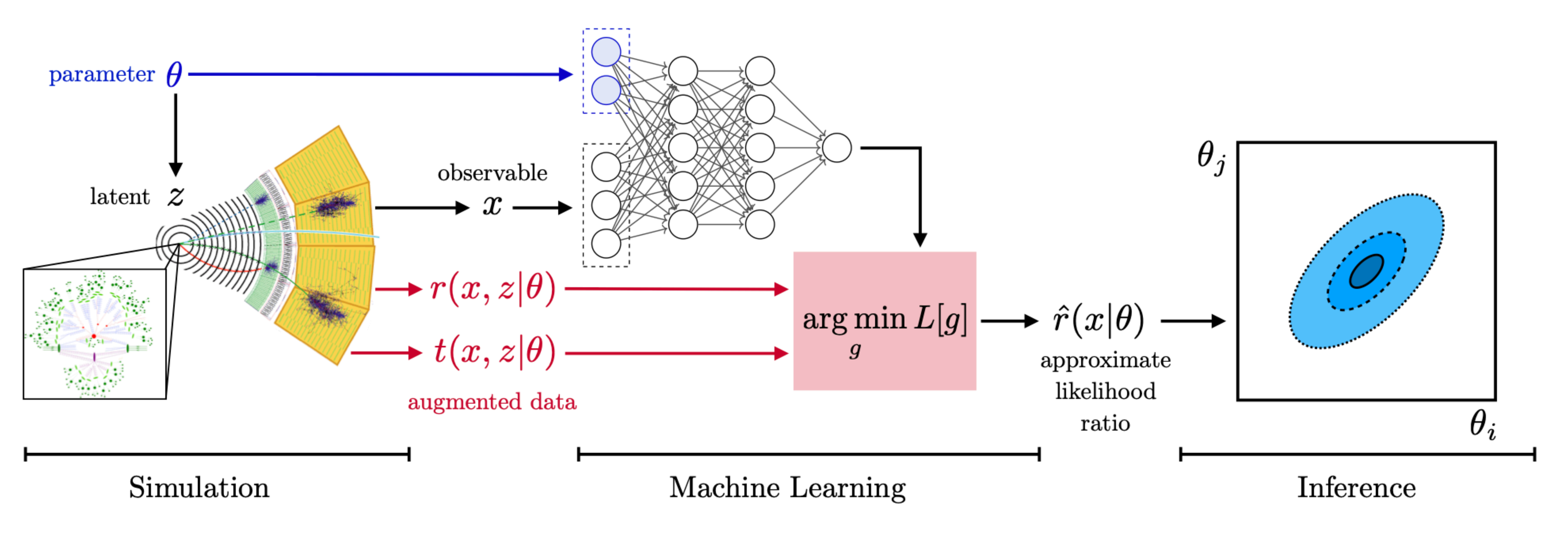
- We lose information in the process of binning
- To what degree can binned analyses achieve statistically optimal bounds?
- Even for bins, the precise choice of binning is not clear

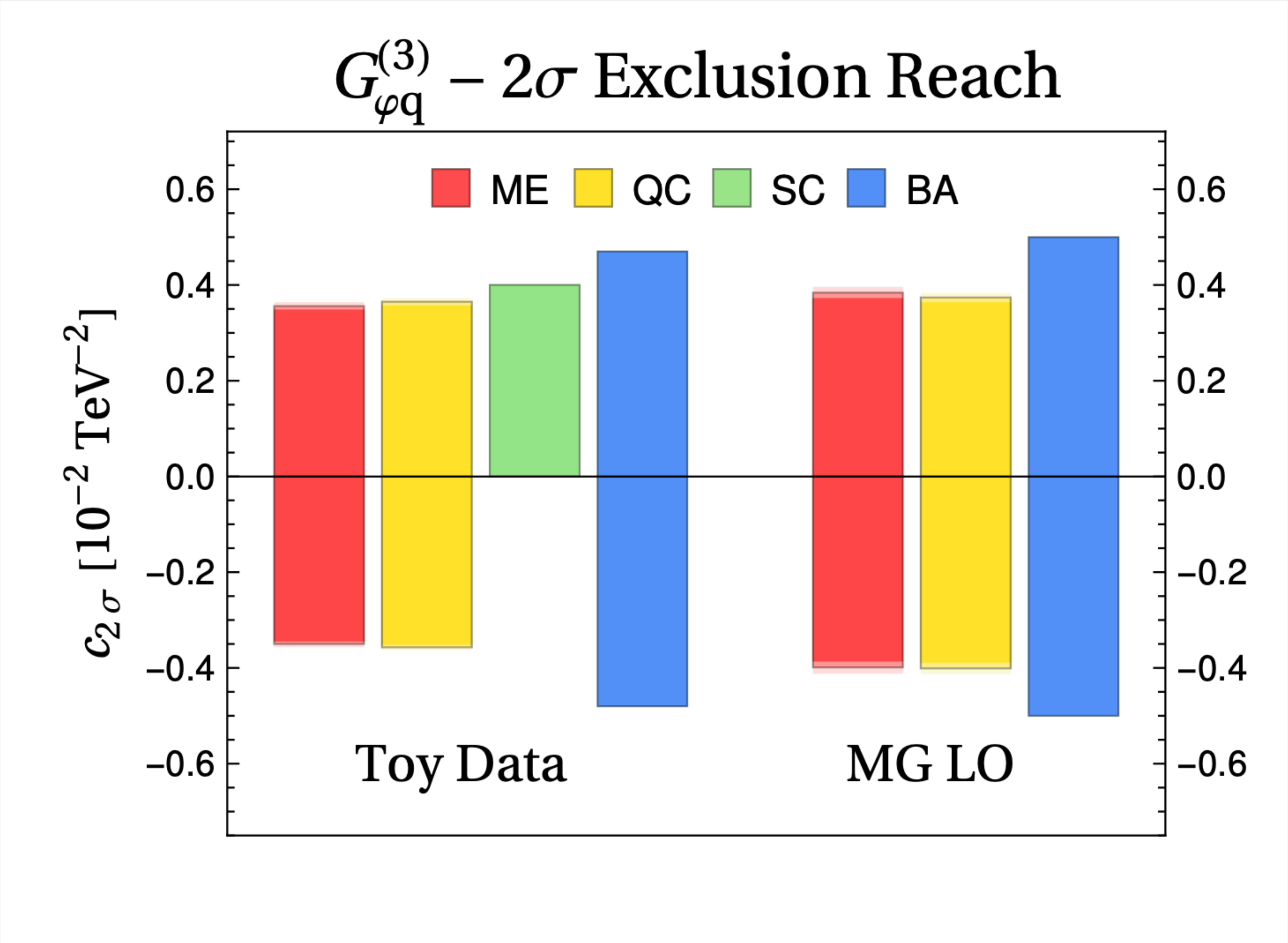
Goal: develop statistically optimal observables and integrate them into global EFT fits

Related work

- The **likelihood** (ratio) as central object
- Matrix Element Method (MEM): transfer functions
- Parameterise the likelihood ratio with **Neural Networks**
- Current studies are limited to a **small number of EFT coefficients**

S. Chen, A. Glioti, G. Panico, A. Wulzer [2007.10356]





J. Brehmer, K. Cranmer [2010.06439]

Finding optimal observables

Neyman-Pearson: the most powerful statistical test at fixed size (significance level) between two *simple* hypotheses H_0 and H_1 is the (log) likelihood ratio:

$$t_c(D) \equiv \log \frac{\mathscr{L}(H_1 \mid D)}{\mathscr{L}(H_0 \mid D)}$$

- Any other test statistic has less power, i.e. gives suboptimal bounds
- No longer applies in case of systematics: profile likelihood ratio (WIP)

Finding optimal observables

Key idea: train a NN classifier to learn the *extended* likelihood ratio

SM hypothesis

Number of events

$$t_c \equiv \log \frac{\mathscr{L}(H_1 | D)}{\mathscr{L}(H_0 | D)} = \nu^{\text{eft}} - \nu^{\text{sm}} - \sum_{i=1}^{n} \log \frac{d\sigma(x_i, c)}{d\sigma(x_i, 0)}$$

Extended likelihood ratio

EFT hypothesis (null)

Expected number of events under the SM

Cross section ratio

The events x_i can be invariant masses, rapidities, scattering angles, p_T, ...

Binary classifier

- Train a **classifier** by minimising the cross entropy (or the quadratic loss) loss functional

$$L[f(x)] = -\int dx \frac{d\sigma_0}{dx} \log(1-f) - \int dx \frac{d\sigma_1}{dx} \log f$$

which gives

The choice of loss functional is not unique!

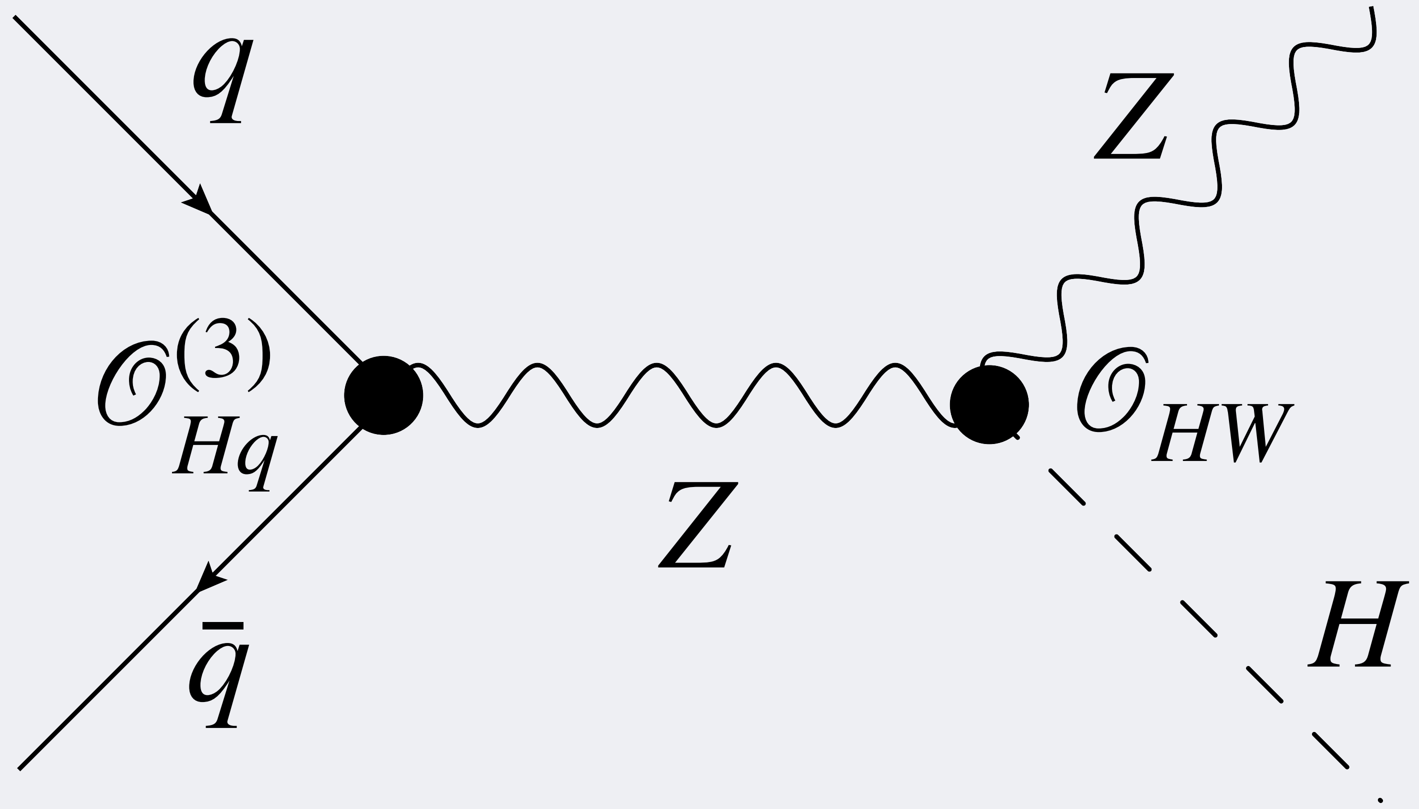
$$\frac{\delta L}{\delta f(x')} = \frac{d\sigma_0}{1-f} - \frac{d\sigma_1}{f} = 0 \implies \hat{f} = \frac{1}{1 + d\sigma_0/d\sigma_1}$$

- This is a **one-to-one estimator** of the likelihood ratio!

VH production

- Need access to underlying truth to assess the NN accuracy
- Efficient **pipeline** using **FeynRules, SMEFTsim, FeynArts / FormCalc** to obtain analytical predictions
- LO parton level, but the method is applicable to any final state
- Study $\mathcal{O}_{HW}, \mathcal{O}_{HWB}, \mathcal{O}_{HB}, \mathcal{O}_{HD}$ and $\mathcal{O}_{Hq}^{(3)}$ up to $\mathcal{O}\left(\Lambda^{-4}\right)$ differential in the rapidity and invariant mass m_{VH}

$$\frac{d\sigma}{dm_{VH}dY} = \frac{2m_{VH}}{s}\left[\sum_f f_f\left(x_1, Q\right) f_{\bar{f}}\left(x_2, Q\right) \hat{\sigma}_{q\bar{q}\to VH}\right]$$



VH production: FormCalc

```
dim6QuadSimp = FullSimplify[dim6Quad]
```

$$\frac{1}{216\, cth^6\, MZ2\, (MZ2 - S)^2\, S^2\, sth^6\, T^2\, U^2\, \Lambda^4}$$

$$\pi v^2 \Big(-48\, Alfa2\, cHq3\, cth^2\, \pi\, S\, sth^2\, T^2\, U^2 \Big(3\, cHq3\, (MZ2 - S)\, S\, \big(MZ2^2 + MZ2\, (S - T - U) + T\, U\big) + cHW\, cth^2 \big(-3\, S + 2\, (MZ2 + S)\, sth^2\big) \big(MZ2^3 + MH2\, (3\, MZ2 - 2\, S - T)\, T + MH2^2\, (-MZ2 + T) + MZ2^2\, (6\, S - 3\, U) -$$

$$MH2\, (MZ2 + T)\, U + S\, T\, (S + T + U) + MZ2 \big(S^2 - (T - 2\, U)\, (T + U) - S\, (5\, T + U)\big)\big) + sth \big(cHWB\, cth \big(-MZ2 - 2\, S + 2\, (MZ2 + S)\, sth^2\big) + cHB\, sth \big(2\, cth^2\, (-MZ2 + S) + S \big(-3 + 4\, sth^2\big)\big)\big)$$

$$\big(MZ2^3 + MH2\, (3\, MZ2 - 2\, S - T)\, T + MH2^2\, (-MZ2 + T) + MZ2^2\, (6\, S - 3\, U) - MH2\, (MZ2 + T)\, U + S\, T\, (S + T + U) + MZ2 \big(S^2 - (T - 2\, U)\, (T + U) - S\, (5\, T + U)\big)\big)\Big)\, v^2 +$$

$$Alfa \Big(24\, cHq3\, cth^4\, (MZ2 - S)\, S\, sth^4 \big(cHW\, cth^2 \big(-3\, S + 4\, MZ2\, sth^2\big) + sth \big(-cHWB\, cth \big(S + MZ2 \big(2 - 4\, sth^2\big)\big) + cHB\, sth \big(4\, cth^2\, (-MZ2 + S) + S \big(-3 + 4\, sth^2\big)\big)\big)\big)\, T^2\, U^2 \big(MZ2^3 + MH2\, (3\, MZ2 - 2\, S - T)\, T +$$

$$MH2^2\, (-MZ2 + T) + MZ2^2\, (6\, S - 3\, U) - MH2\, (MZ2 + T)\, U + S\, T\, (S + T + U) + MZ2 \big(S^2 - (T - 2\, U)\, (T + U) - S\, (5\, T + U)\big)\big) + 8\, cth^4\, sth^4 \big(cHW^2\, cth^4 \big(9\, S^2 - 24\, MZ2\, S\, sth^2 + 32\, MZ2^2\, sth^4\big) +$$

$$2\, cHW\, cth^2\, sth \big(cHWB\, cth \big(3\, S\, (2\, MZ2 + S) - 8\, MZ2\, (2\, MZ2 + S)\, sth^2 + 32\, MZ2^2\, sth^4\big) + cHB\, sth \big(4\, cth^2\, (MZ2 - S) \big(3\, S - 8\, MZ2\, sth^2\big) + S \big(9\, S - 12\, (MZ2 + S)\, sth^2 + 32\, MZ2\, sth^4\big)\big)\big) +$$

$$sth^2 \big(cHWB^2\, cth^2 \big(5\, S^2 + 8\, MZ2^2 \big(1 - 2\, sth^2\big)^2 + 4\, MZ2\, S \big(-1 + 2\, sth^2\big)\big) + cHB^2\, sth^2 \big(32\, cth^4\, (MZ2 - S)^2 - 8\, cth^2\, (MZ2 - S)\, S \big(-3 + 8\, sth^2\big) + S^2 \big(9 - 24\, sth^2 + 32\, sth^4\big)\big) +$$

$$2\, cHB\, cHWB\, cth\, sth \big(S \big(6\, MZ2 + 3\, S + 4\, (-7\, MZ2 + S)\, sth^2 + 32\, MZ2\, sth^4\big) + 4\, cth^2\, (-MZ2 + S) \big(S + MZ2 \big(-4 + 8\, sth^2\big)\big)\big)\big)\big)\, T^2\, U^2$$

$$\big(MZ2^3\, (4\, S + T - 2\, U) - (MH2 - S)^2\, T\, (MH2 - S - T - U) + MZ2^2 \big(4\, S^2 + MH2\, T - (T - 2\, U)\, (T + U) - S\, (3\, T + 2\, U)\big) - MZ2 \big(MH2^2\, (T - 2\, U) + S\, (3\, S\, T + 2\, (T - U)\, U) + 2\, MH2\, (S\, (-2\, T + U) + U\, (T + U))\big)\big)\Big) +$$

$$9\, cHq3^2\, S^2 \Big(8\, Alfa2\, \pi^2\, T^2\, U^2 \big(MZ2^2 + MZ2\, (S - T - U) + T\, U\big)\, v^4 - cth^4\, (MZ2 - S)^2\, sth^4$$

$$\big(-8\, T^2\, U^2 \big(MZ2^2 + MZ2\, (S - T - U) + T\, U\big) - \big(2\, S\, T^2\, U^2 - 3\, MH2\, MZ2^2 \big(T^2 + U^2\big) - MZ2^3 \big(T^2 + U^2\big) + 2\, MZ2\, T\, U \big(T^2 + U^2\big) + MZ2^2\, (S + T + U) \big(T^2 + U^2\big)\big)\, v^2\, yu^2\big)\Big)\Big)$$

```
dim6LinSimp = FullSimplify[dim6Lin]
```

$$\frac{1}{108\, cth^6\, MZ2\, (MZ2 - S)^2\, S\, sth^6\, T^2\, U^2\, \Lambda^2}$$

$$\pi v^2 \Big(-4\, Alfa2\, \pi\, T^2\, U^2 \Big(cHW\, cth^4\, sth^2 \big(-9\, S + 12\, (MZ2 + S)\, sth^2 - 32\, MZ2\, sth^4\big) \big(MZ2^3 + MH2\, (3\, MZ2 - 2\, S - T)\, T + MH2^2\, (-MZ2 + T) + MZ2^2\, (6\, S - 3\, U) - MH2\, (MZ2 + T)\, U + S\, T\, (S + T + U) +$$

$$MZ2 \big(S^2 - (T - 2\, U)\, (T + U) - S\, (5\, T + U)\big)\big) - cth^2\, sth^3 \big(cHWB\, cth \big(6\, MZ2 + 3\, S + 4\, (-7\, MZ2 + S)\, sth^2 + 32\, MZ2\, sth^4\big) + cHB\, sth \big(-4\, cth^2\, (MZ2 - S) \big(-3 + 8\, sth^2\big) + S \big(9 - 24\, sth^2 + 32\, sth^4\big)\big)\big)$$

$$\big(MZ2^3 + MH2\, (3\, MZ2 - 2\, S - T)\, T + MH2^2\, (-MZ2 + T) + MZ2^2\, (6\, S - 3\, U) - MH2\, (MZ2 + T)\, U + S\, T\, (S + T + U) + MZ2 \big(S^2 - (T - 2\, U)\, (T + U) - S\, (5\, T + U)\big)\big) +$$

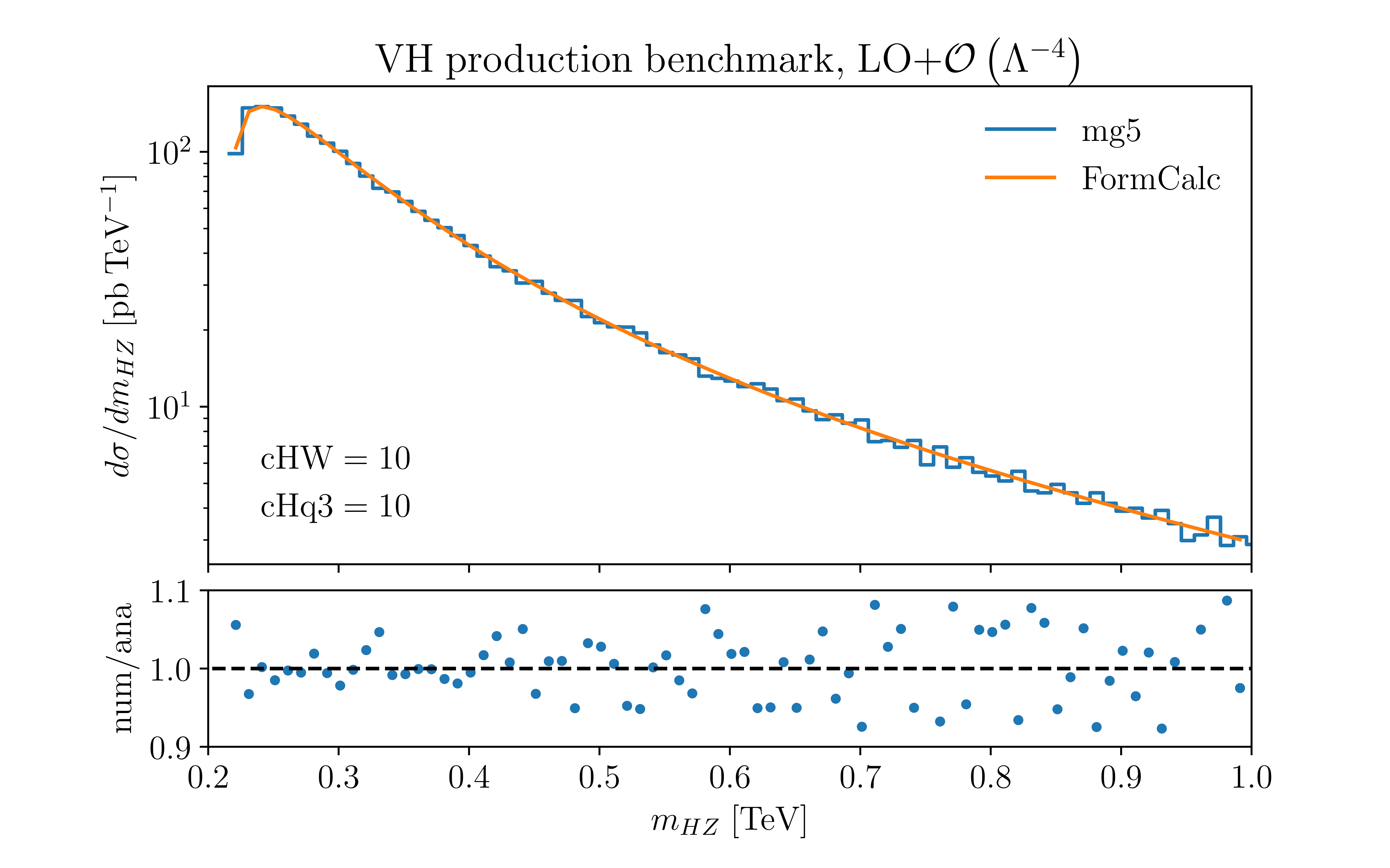
$$6\, cHq3\, S \big(-3 + 4\, sth^2\big) \big(MZ2^2 + MZ2\, (S - T - U) + T\, U\big) \big(cth^2\, (-MZ2 + S)\, sth^2 + Alfa\, \pi\, v^2\big)\Big) +$$

$$3\, Alfa\, cHq3\, cth^4\, (MZ2 - S)^2\, S\, sth^4 \Big(MZ2^2 \big(-3 + 4\, sth^2\big)\, (MZ2 - S - T)\, T^2 + MZ2\, T \big(8\, (MZ2 - S)\, (MZ2 + 2\, S)\, sth^2 + 3\, MZ2\, T - 4\, (3\, MZ2 + 4\, S)\, sth^2\, T + 2 \big(3 - 4\, sth^2\big)\, T^2\big)\, U +$$

$$\big(MZ2^2\, (MZ2 - S) \big(-3 + 4\, sth^2\big) + MZ2 \big(3\, MZ2 - 4\, (3\, MZ2 + 4\, S)\, sth^2\big)\, T + 2 \big(3\, S - 8\, MZ2\, sth^2\big)\, T^2\big)\, U^2 -$$

$$MZ2 \big(-3 + 4\, sth^2\big)\, (MZ2 + 2\, T)\, U^3 + 3\, MH2\, MZ2^2 \big(\big(-3 + 4\, sth^2\big)\, T^2 + 8\, sth^2\, T\, U + \big(-3 + 4\, sth^2\big)\, U^2\big)\Big)\, yu^2\Big)$$

VH production: benchmark



Training the likelihood ratio

We separate the learning problem by exploiting the structure inherent to the EFT parameter space:

$$r(x, \mathbf{c}) = 1 + c_1\alpha_1(x) + c_2\alpha_2(x) + c_1^2\beta_{11}(x) + c_1c_2\beta_{12}(x) + \beta_{22}c_2^2$$

1. Train the linear coefficient functions in parallel
2. Switch on quadratic corrections and train the quadratic coefficients
3. The cross terms can finally be extracted

New: this allows for efficient scaling (quadratically) and parallel training for n EFT parameters

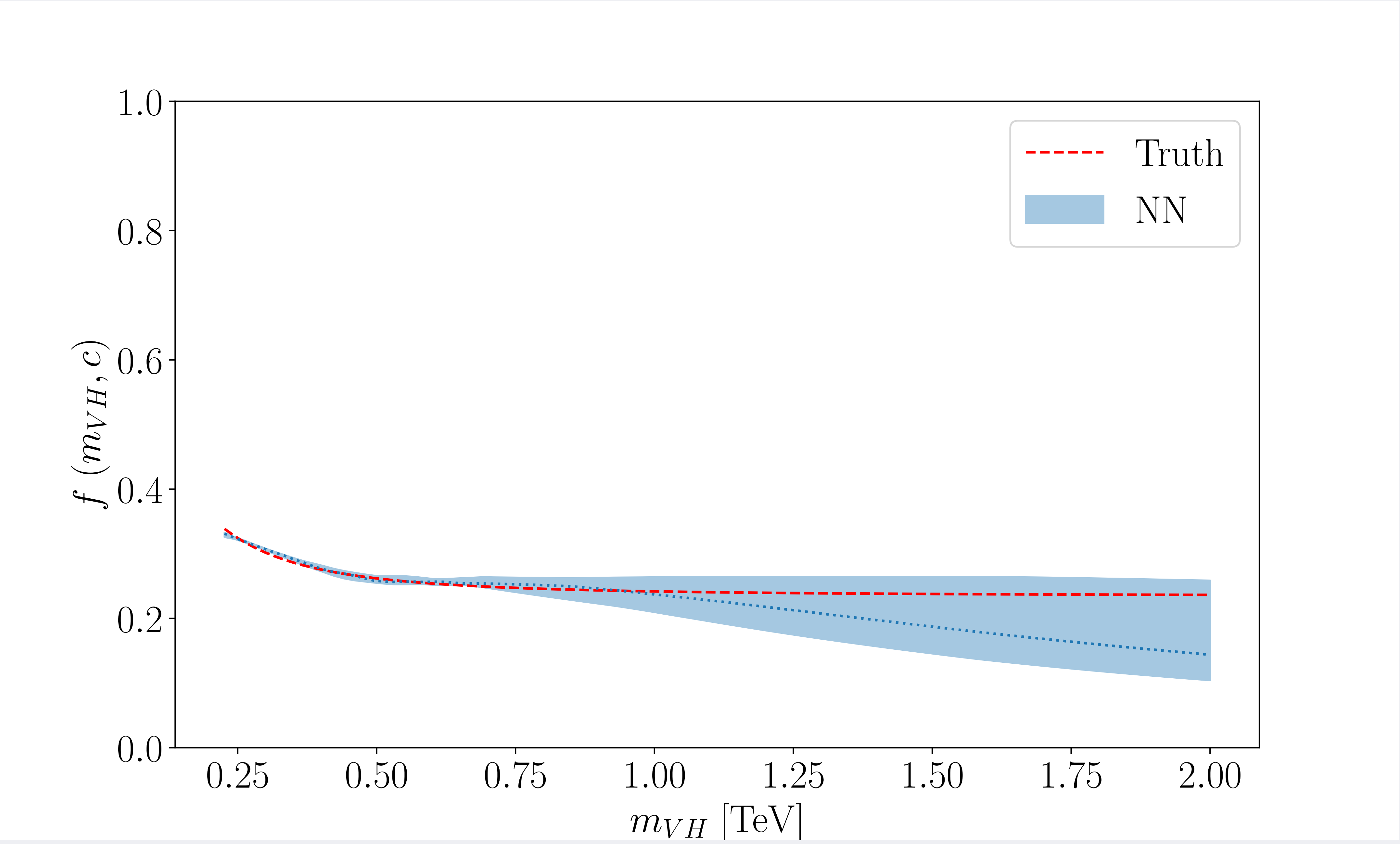
NN model uncertainties

- We systematically assess the **model uncertainties** associated to the NN parameterisation of the likelihood ratio

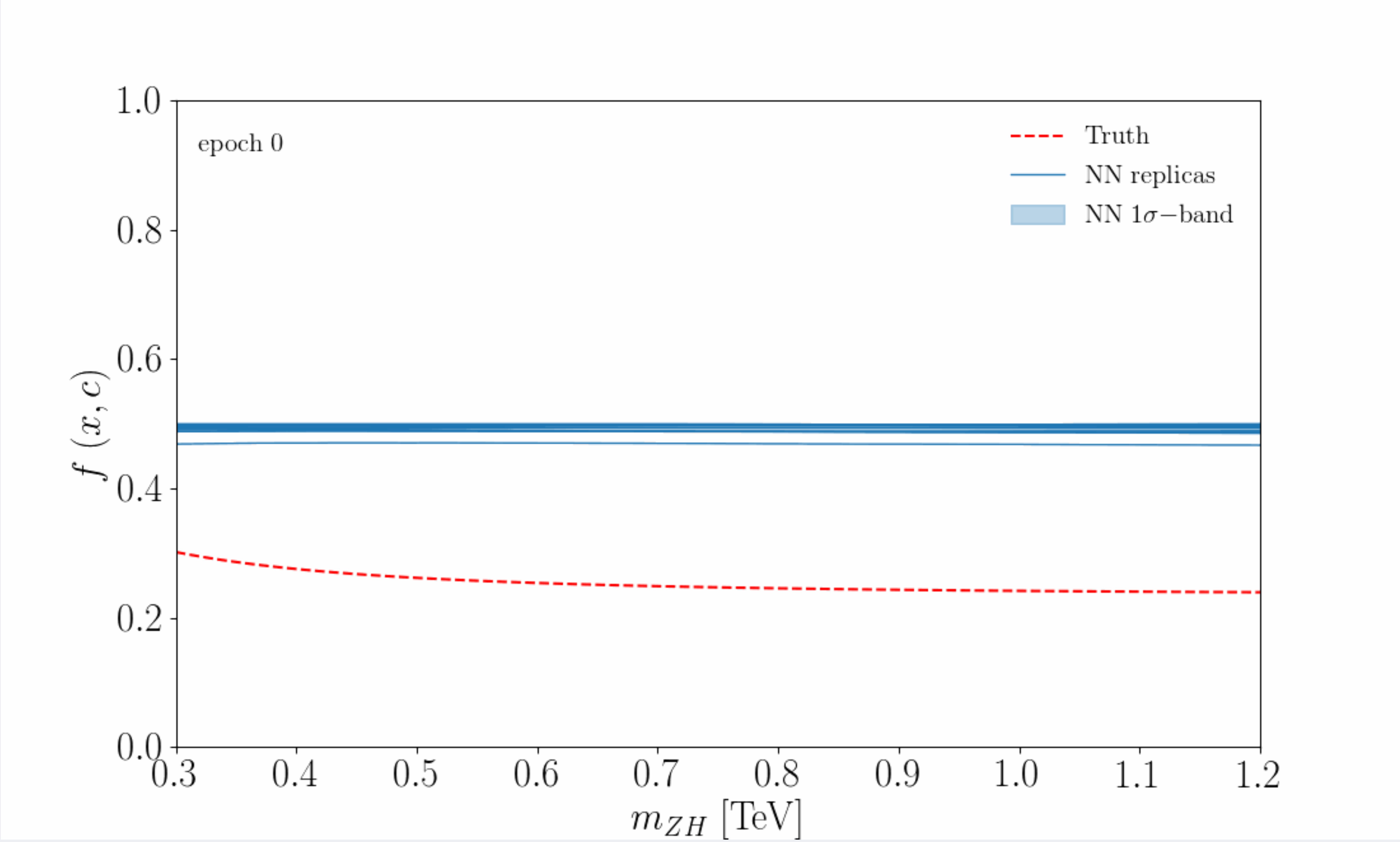
- **Replica**: an independent MC training set to propagate the error to the space of models

- Train 30 independent replicas **in parallel**

- Translate to the error on the Wilson coefficients

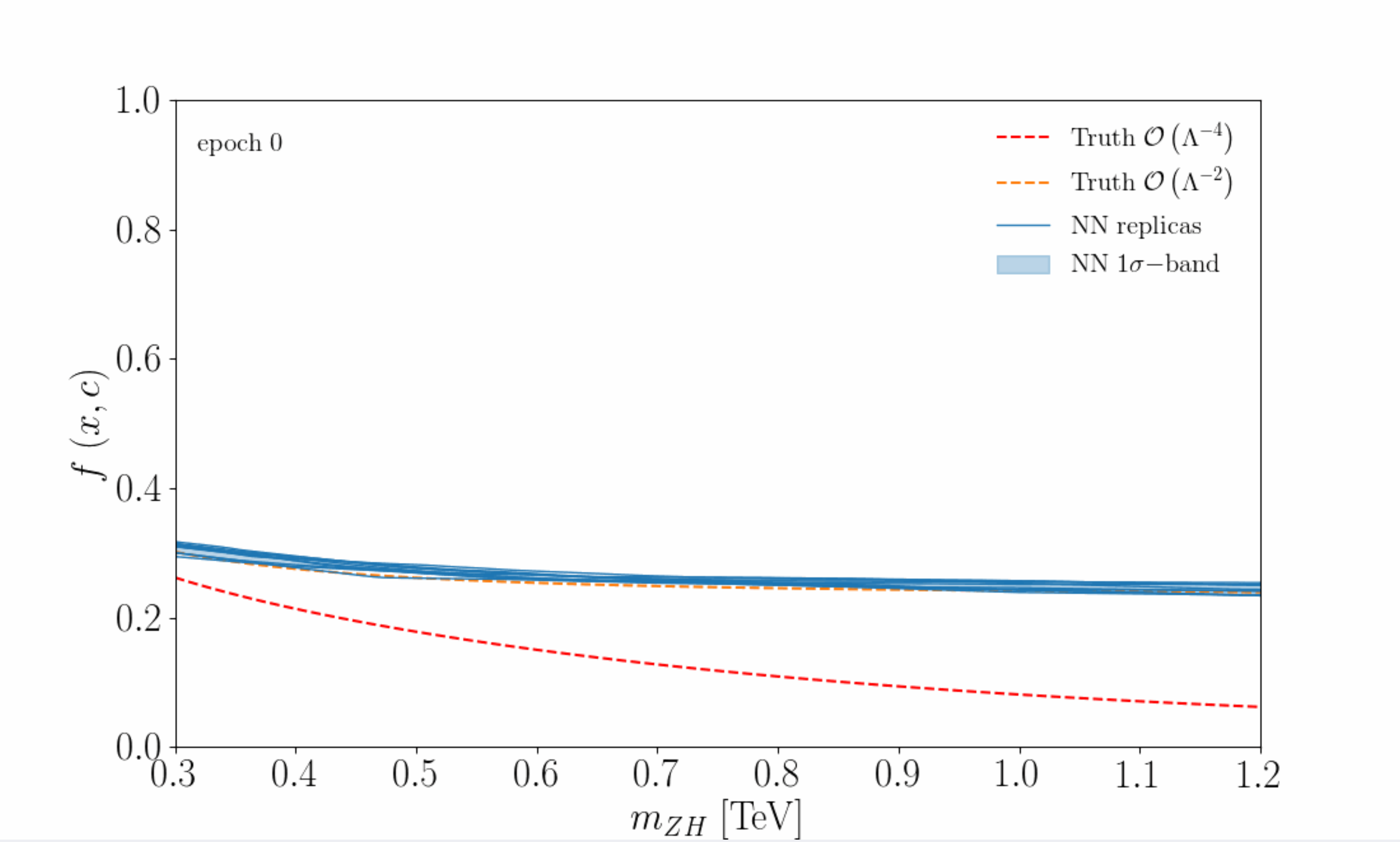


Seeing the training at work



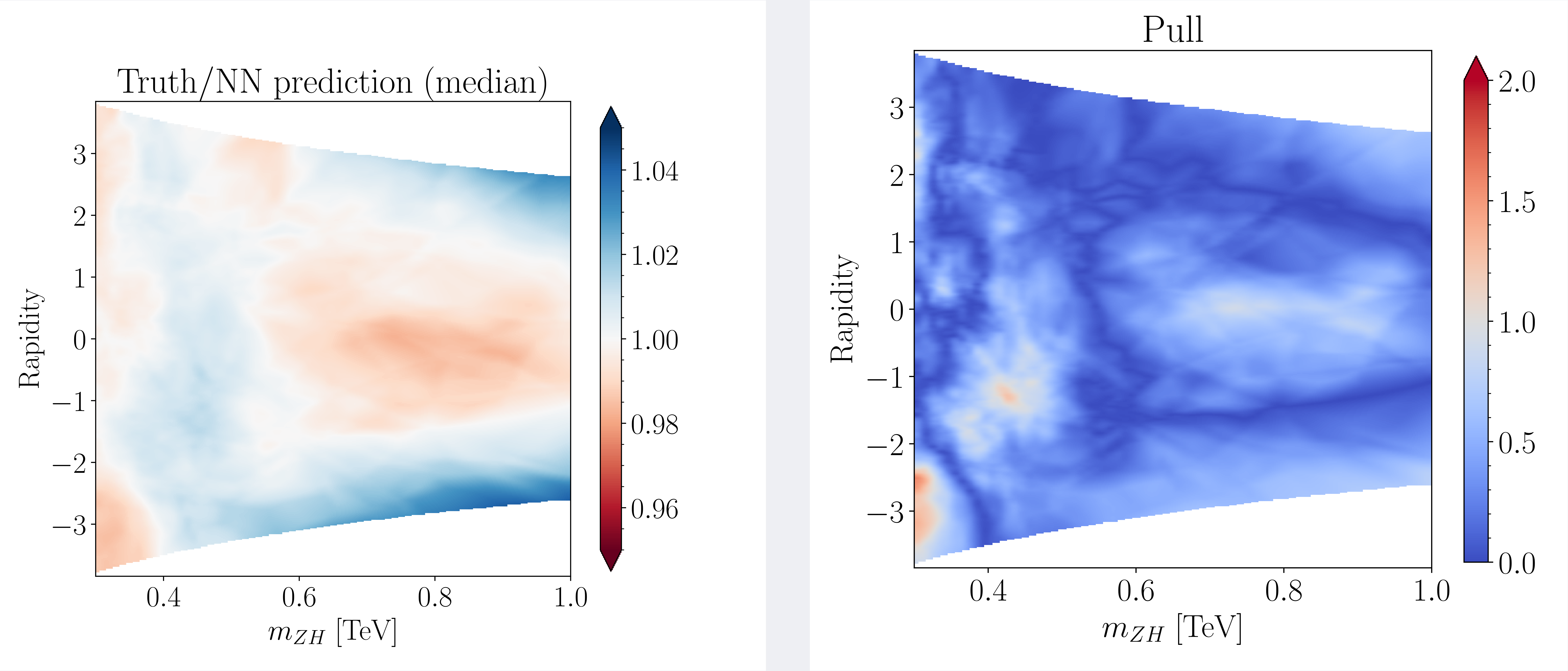
- Trained on 30 replicas
- 100K events in SM and EFT
- Cross validation
- Architecture: {2, 5x30, 1} with ReLU activation functions
- Standardised training data to zero mean and unit variance

Adding quadratic corrections

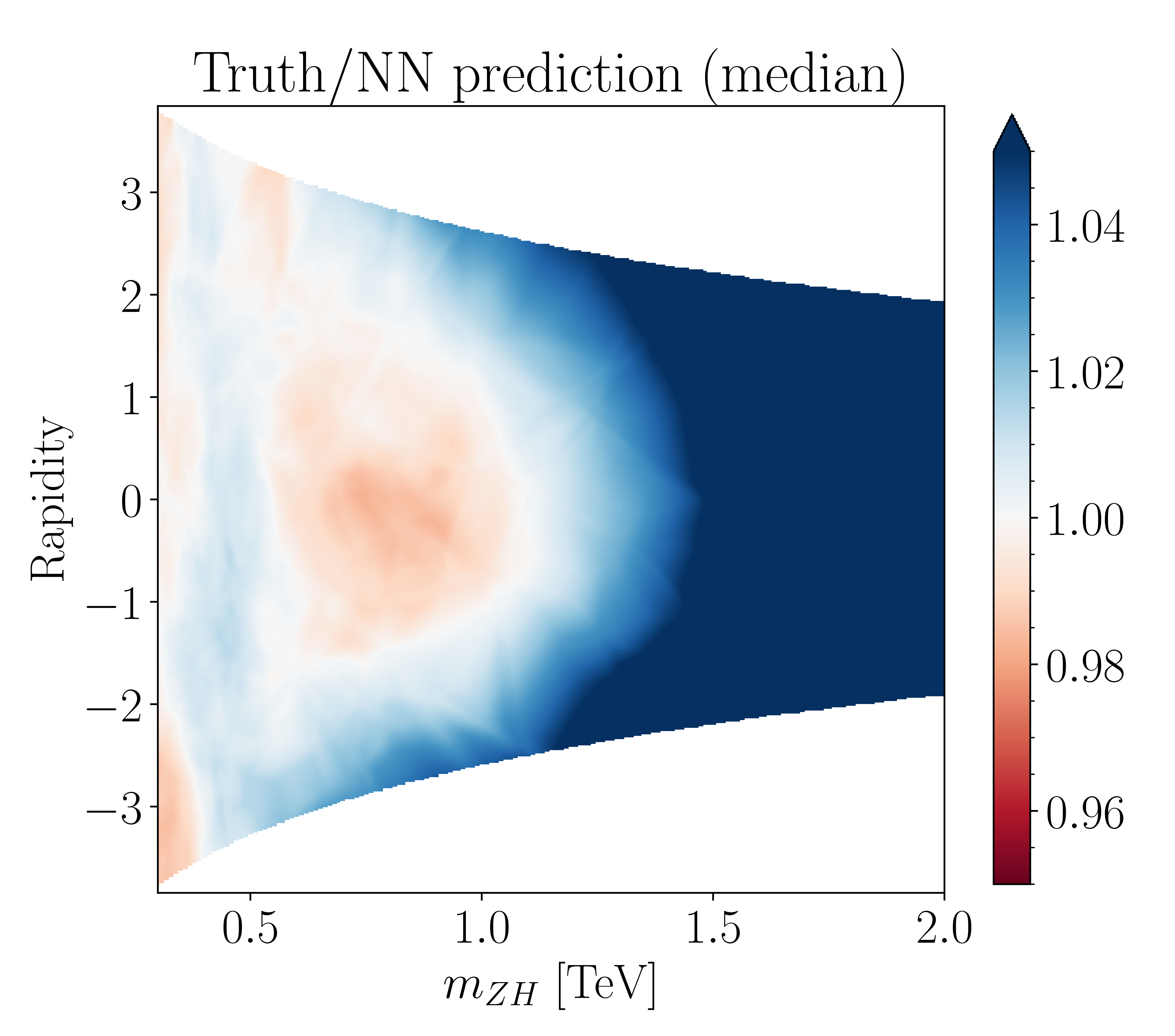


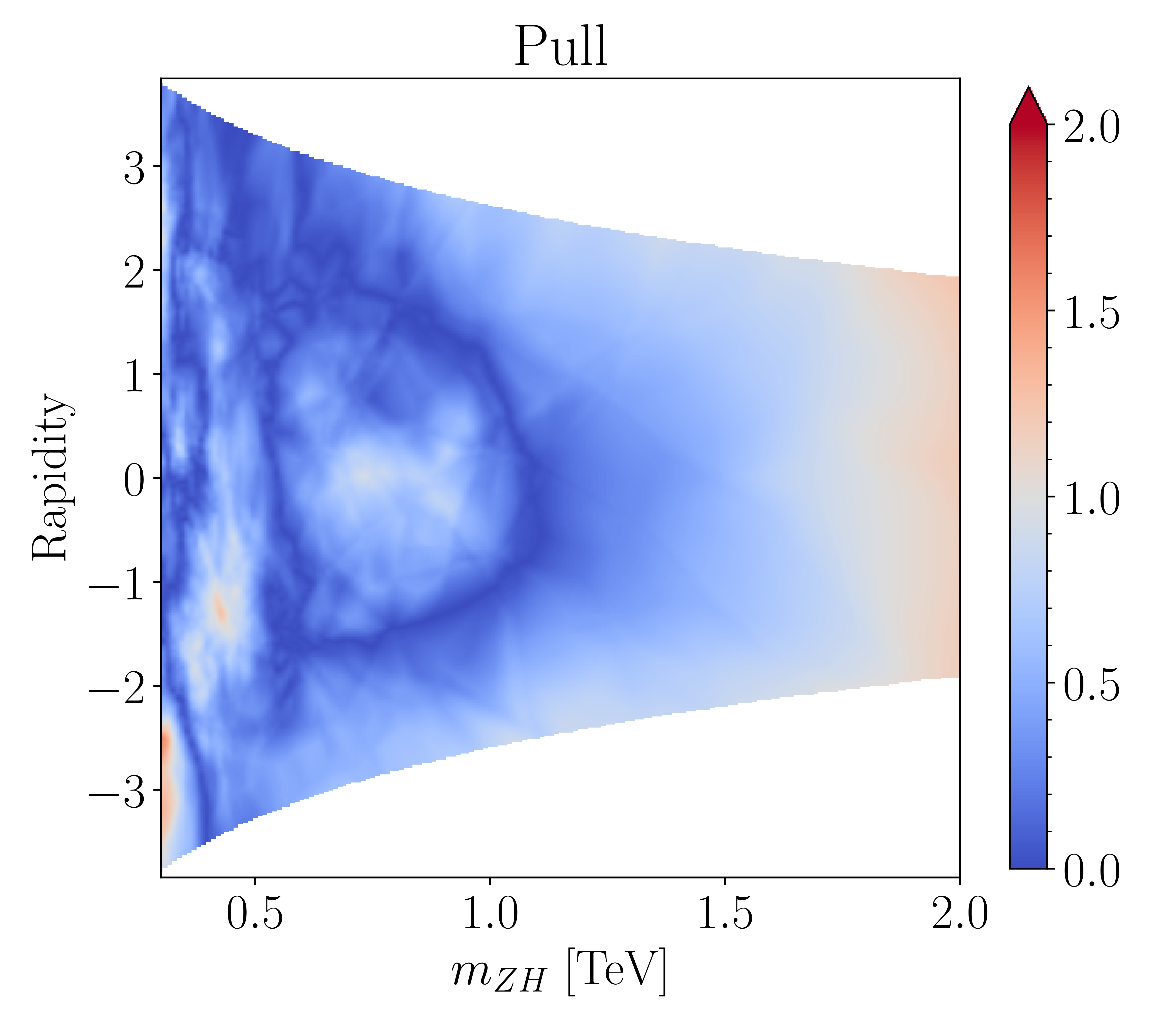
- Trained on 30 replicas
- 100K events in SM and EFT
- Cross validation
- Architecture: {2, 5x30, 1} with ReLU activation functions
- Standardised training data to zero mean and unit variance

Training performances

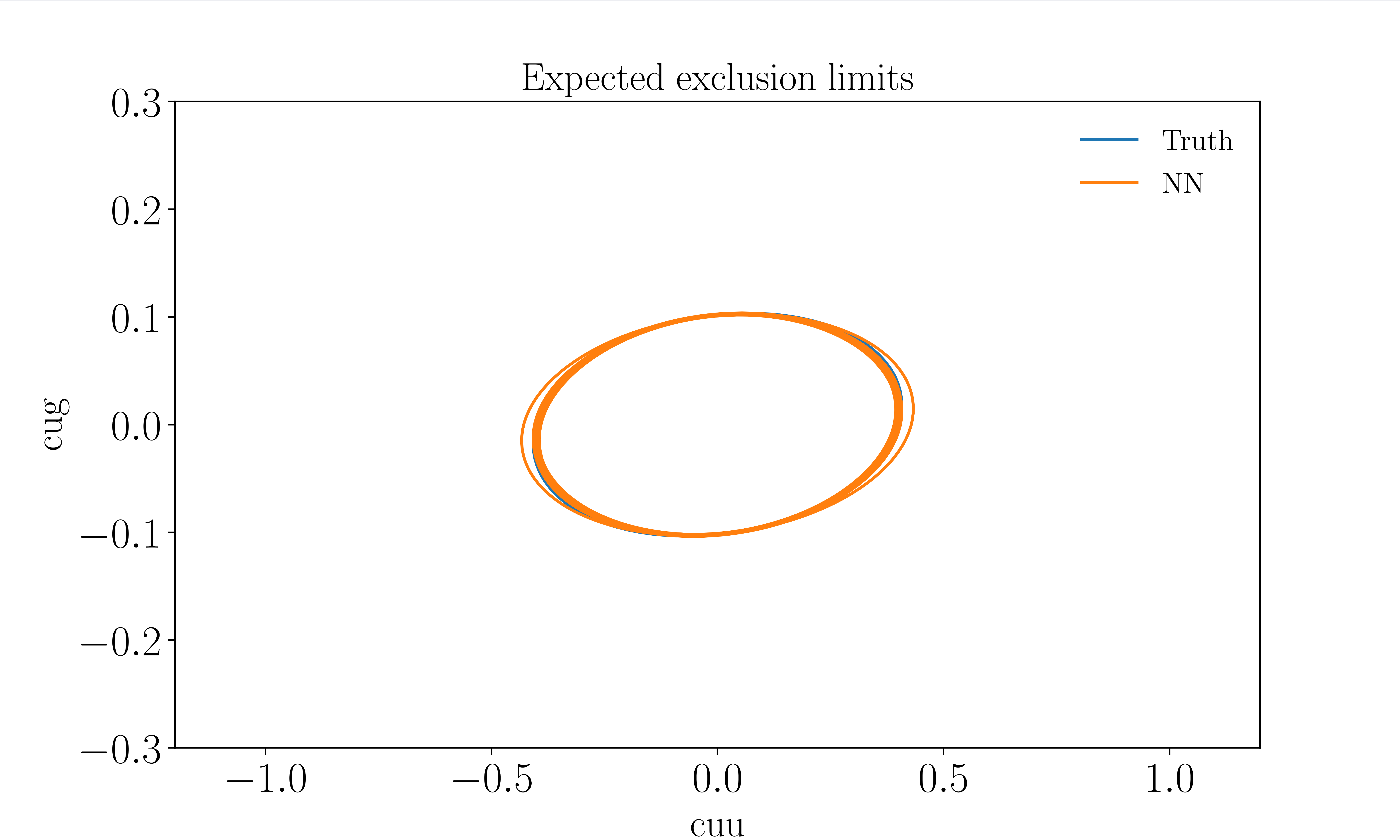


Training performances





Limits: NN versus Truth



Summary

Key question: given a collider process, how can one define **optimal observables** with the highest sensitivity to EFT coefficients?

- Train a surrogate of the **likelihood ratio**
- **Efficient scaling** properties to n EFT parameters necessary for global EFT fits
- Good reconstruction **performances** throughout phase space
- Outlook: Include **systematics** with the profile likelihood ratio

Thank you!