Unbinned measurements in global SMEFT fits from machine learning

DIS2022

XXIX International Workshop on Deep-Inelastic Scattering and Related Subjects

Santiago de Compostela, 2-6 May 2022

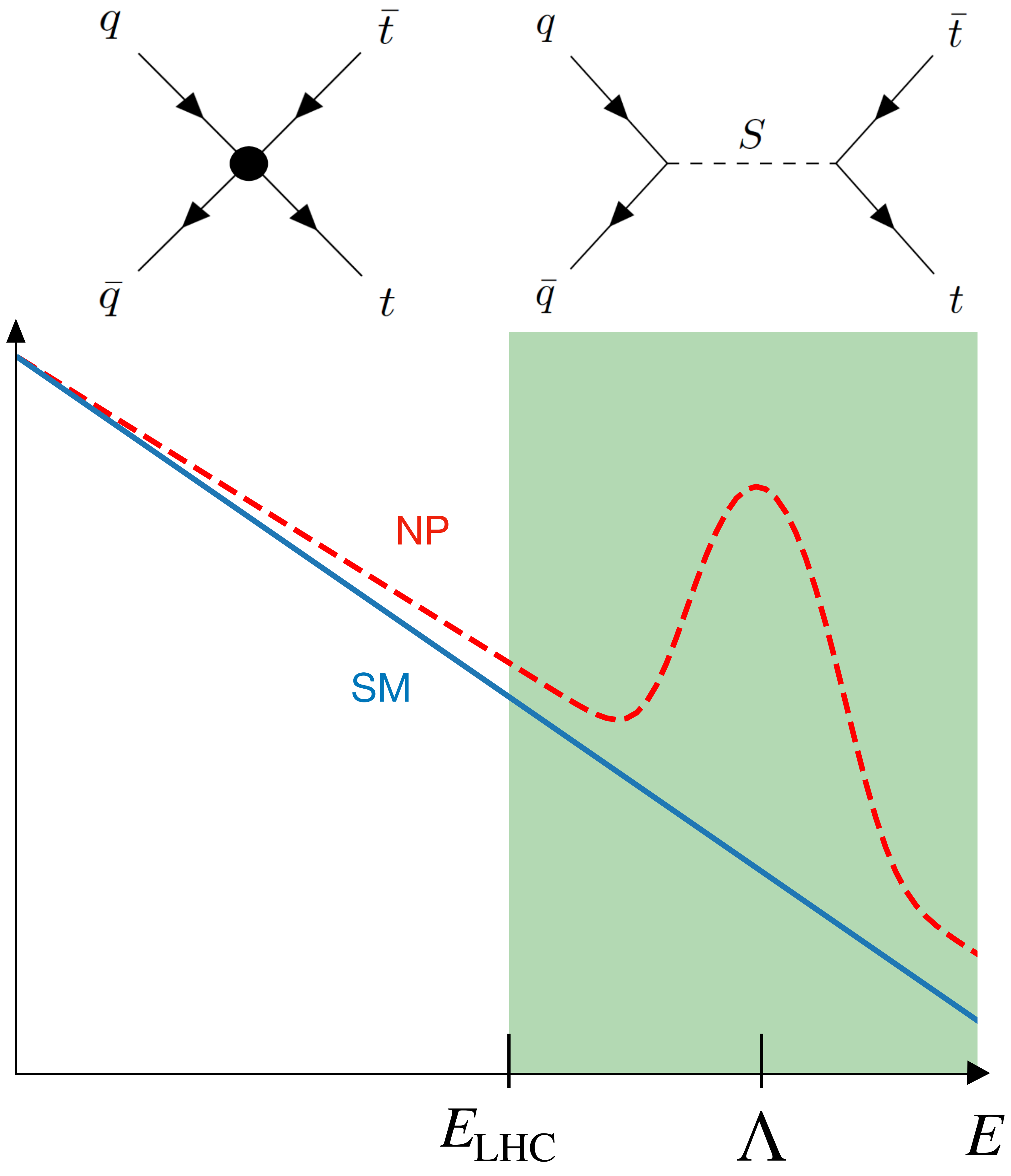
Jaco ter Hoeve

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

The Standard Model Effective Field Theory

- Systematic parameterisation of the **theory space** close to the Standard Model
- Study the **fingerprints** of NP at low energies through higher dimensional operators
- Assumes the **SM field content** and **gauge symmetries**

$$\mathscr{L}_{\mathrm{SMEFT}} = \mathscr{L}_{\mathrm{SM}} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i^{N_{d8}} \frac{b_i}{\Lambda^4}\mathcal{O}_i^{(8)} + \dots$$



The Standard Model Effective Field Theory

- Systematic parameterisation of the **theory space** close to the Standard Model

- Study the **fingerprints** of NP at low energies through higher dimensional operators

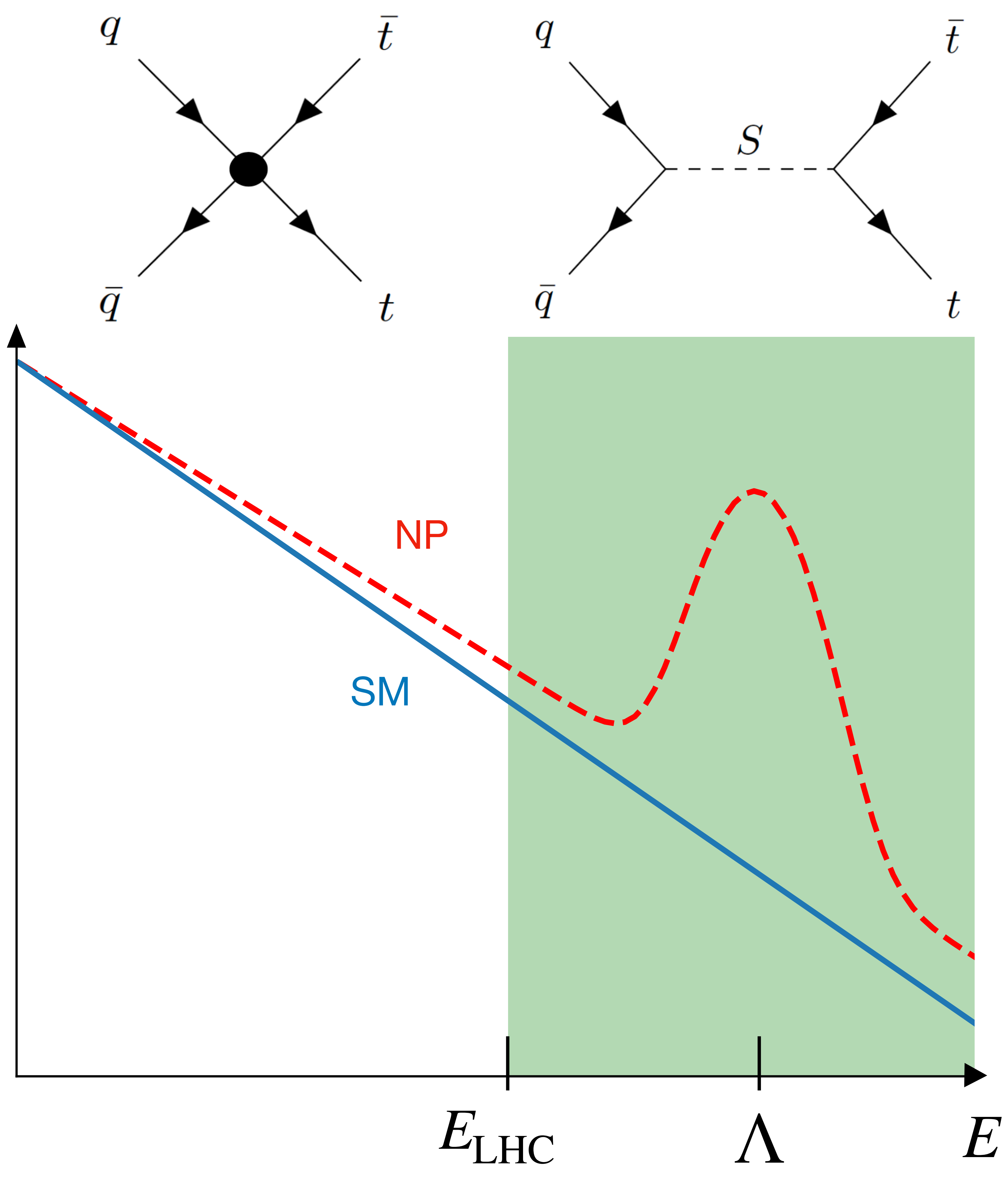
- Assumes the **SM field content** and **gauge symmetries**

$$\mathscr{L}_{\mathrm{SMEFT}} = \mathscr{L}_{\mathrm{SM}} + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2} \mathscr{O}_i^{(6)} + \sum_{i}^{N_{d8}} \frac{b_i}{\Lambda^4} \mathscr{O}_i^{(8)} + \ldots$$

The Standard Model Effective Field Theory

- Systematic parameterisation of the **theory space** close to the Standard Model
- Study the **fingerprints** of NP at low energies through higher dimensional operators
- Assumes the **SM field content** and **gauge symmetries**

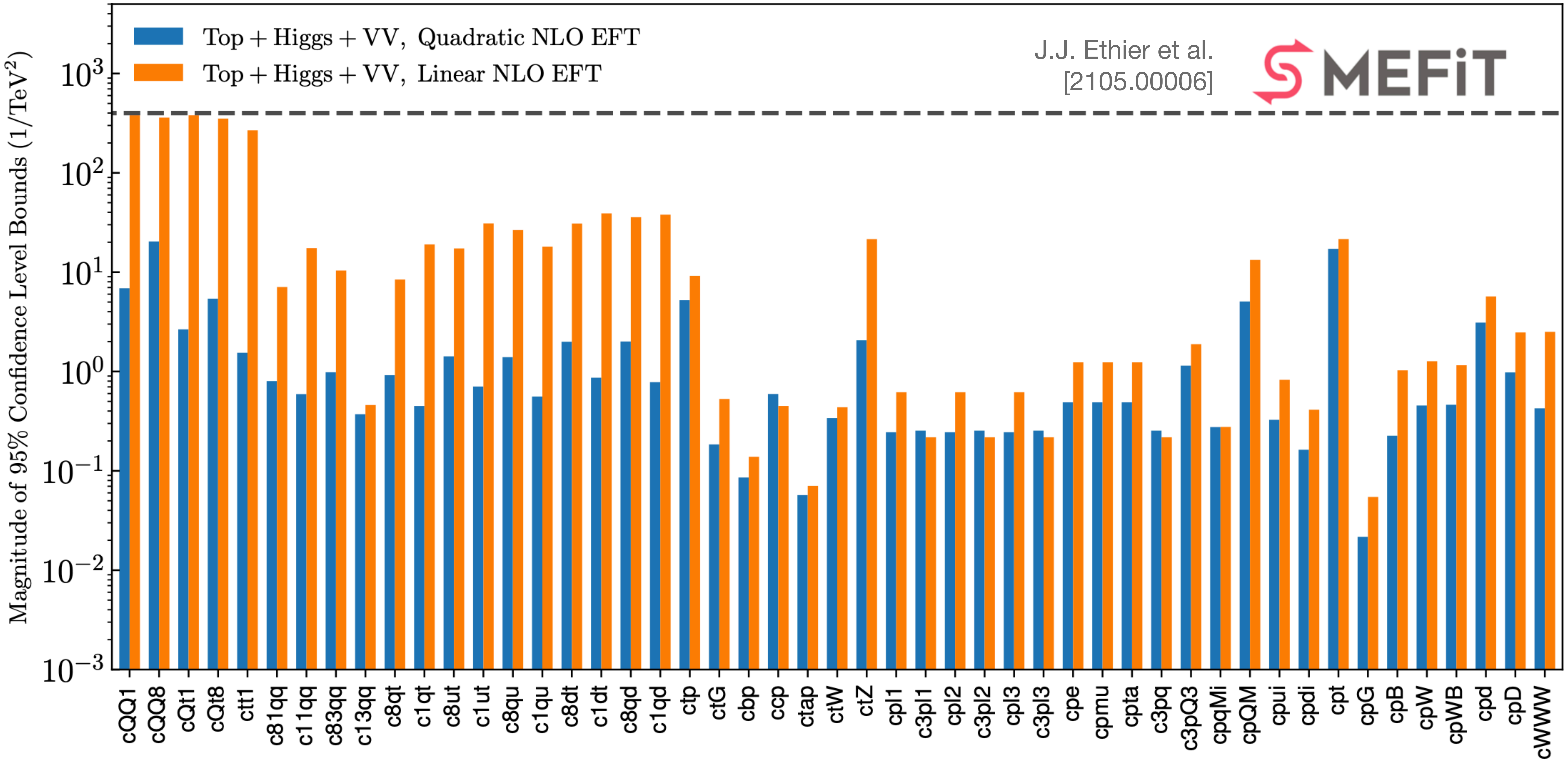
$$\sigma = \sigma_{\mathrm{SM}} + \underbrace{\sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2}\kappa_i}_{\text{Interference}} + \underbrace{\sum_{i,j}^{N_{d6}} \frac{c_i c_j}{\Lambda^4}\tilde{\kappa}_{ij}}_{\text{Quadratic corrections}} + \ldots$$



So what then are global SMEFT fits?

- A global SMEFT analysis needs to explore a **huge** parameter space (2499 at dim 6)
- Studies the intricate interplay between the **top, Higgs** and **diboson** sectors
- Requires state-of-the-art theory calculations in both the SM (NLO QCD + NNLO K-factors) and EFT

So what then are global SMEFT fits?



J.J. Ethier et al. [2105.00006]

So what then are global SMEFT fits?

- A global SMEFT analysis needs to explore a **huge** parameter space (2499 at dim 6)
- Studies the intricate interplay between the **top, Higgs** and **diboson** sectors
- Requires state-of-the-art theory calculations in both the SM (NLO QCD + NNLO K-factors) and EFT

So what then are global SMEFT fits?

- A global SMEFT analysis needs to explore a **huge** parameter space (2499 at dim 6)
- Studies the intricate interplay between the **top, Higgs** and **diboson** sectors
- Requires state-of-the-art theory calculations in both the SM (NLO QCD + NNLO K-factors) and EFT
- Based on unfolded cross sections **not tailored** for EFT studies!

Can we extend global SMEFT fits with ML based observables?

Statistically optimal observables from ML

Which kind of measurement is **most sensitive** to SMEFT operators?

- Difficult question to answer in general, since SMEFT measurements can be

Statistically optimal observables from ML

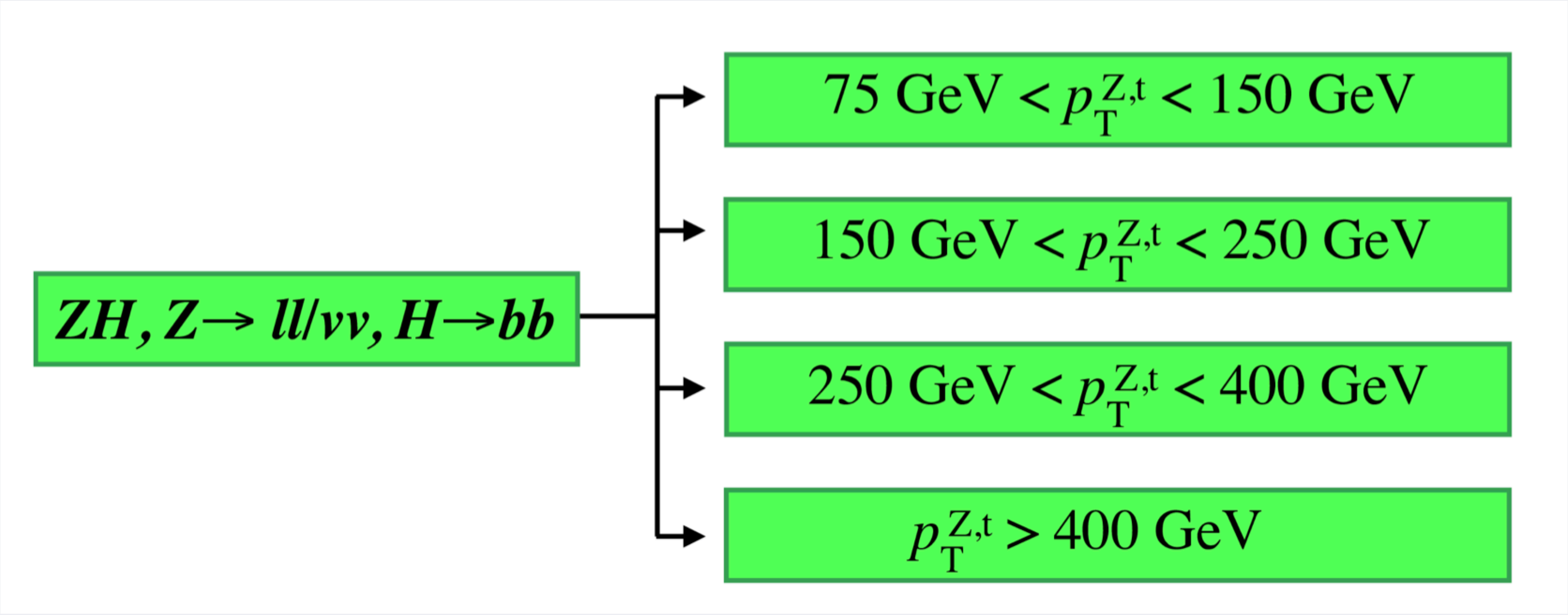
Which kind of measurement is **most sensitive** to SMEFT operators?

- Difficult question to answer in general, since SMEFT measurements can be
  - Inclusive or (1, 2, 3) - differential (in which variables?)

Statistically optimal observables from ML

Which kind of measurement is **most sensitive** to SMEFT operators?

- Difficult question to answer in general, since SMEFT measurements can be
  - Inclusive or (1, 2, 3) - differential (in which variables?)
  - Binned (choice of binning?) or unbinned


$ZH, Z\to ll/\nu\nu, H\to bb$
75 GeV $< p_T^{Z,t} <$ 150 GeV
150 GeV $< p_T^{Z,t} <$ 250 GeV
250 GeV $< p_T^{Z,t} <$ 400 GeV
$p_T^{Z,t} >$ 400 GeV

ATLAS-CONF-2021-051

Statistically optimal observables from ML

Which kind of measurement is **most sensitive** to SMEFT operators?

- Difficult question to answer in general, since SMEFT measurements can be
  - Inclusive or (1, 2, 3) - differential (in which variables?)
  - Binned (choice of binning?) or unbinned

Statistically optimal observables from ML

Which kind of measurement is **most sensitive** to SMEFT operators?

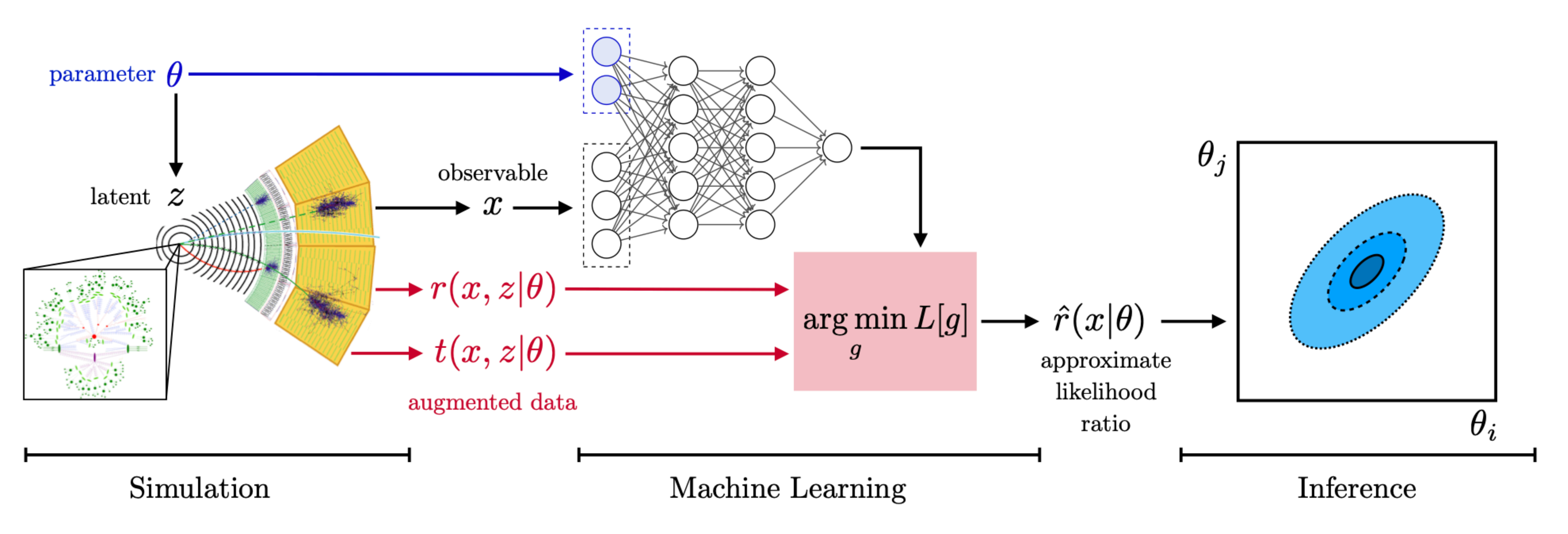
- Difficult question to answer in general, since SMEFT measurements can be
  - Inclusive or (1, 2, 3) - differential (in which variables?)
  - Binned (choice of binning?) or unbinned

Goal: deploy unbinned measurements to determine the optimal sensitivity using ML techniques

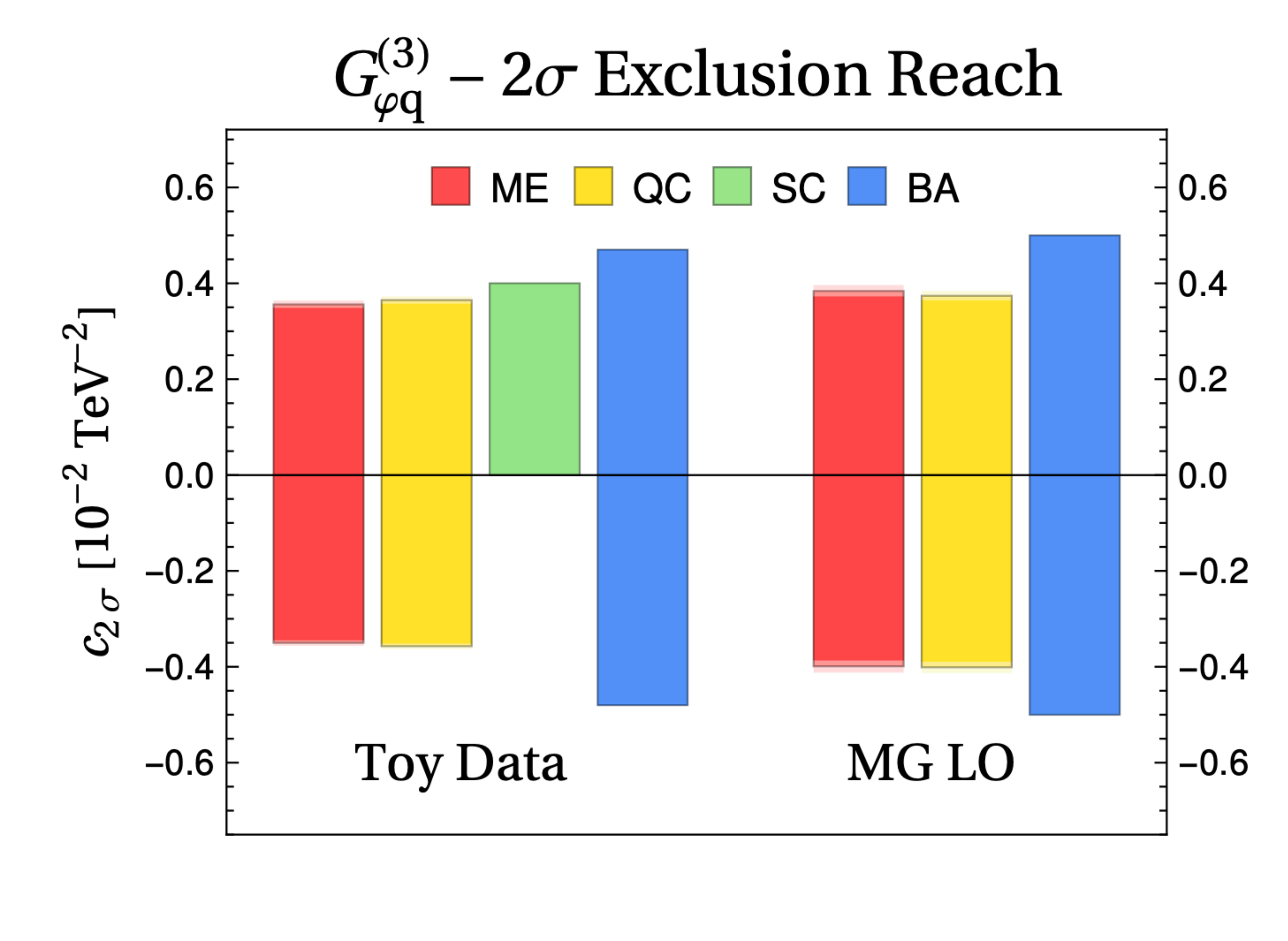
Useful diagnosis tool!

Related work

- The **likelihood ratio** as central object
- Parameterise the likelihood ratio with **Neural Networks**
- Current studies are limited to only a **small number of EFT coefficients**



J. Brehmer et al. [2010.06439]



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

Statistically optimal observables from ML

Carry out EFT analysis with **different** variants of the **same** measurement

$$\log \mathscr{L}_{\text{binned}}(c) = -\frac{1}{2}\sum_{i=1}^{n_{\text{bins}}} \frac{(n_i - \nu_i)^2}{\nu_i}$$

Binned Gaussian Likelihood

$$\log \mathscr{L}_{\text{binned}}(c) = \sum_{i=1}^{n_{\text{bins}}} n_i \log \nu_i(c) - \nu_i(c)$$

Binned Poissonian Likelihood

Unbinned extended likelihood

$$\log \mathscr{L}_{\text{unbinned}}(\nu, c) = -\,\nu(c) + \sum_{i=1}^{N} \log \nu(c) g(x_i, c) \qquad g(x_i, c) \equiv \frac{1}{\sigma(X, c)}\frac{d\sigma(x, c)}{dx}$$

Statistically optimal observables from ML

Carry out EFT analysis with **different** variants of the **same** measurement

$$\log \mathscr{L}_{\text{binned}}(c) = -\frac{1}{2}\sum_{i=1}^{n_{\text{bins}}} \frac{(n_i - \nu_i)^2}{\nu_i}$$

Binned Gaussian Likelihood

$$\log \mathscr{L}_{\text{binned}}(c) = \sum_{i=1}^{n_{\text{bins}}} n_i \log \nu_i(c) - \nu_i(c)$$

Binned Poissonian Likelihood

Unbinned extended likelihood

$$\log \mathscr{L}_{\text{unbinned}}(\nu, c) = -\nu(c) + \sum_{i=1}^{N} \log \nu(c) g(x_i, c) \qquad g(x_i, c) \equiv \frac{1}{\sigma(X, c)} \frac{d\sigma(x, c)}{dx}$$

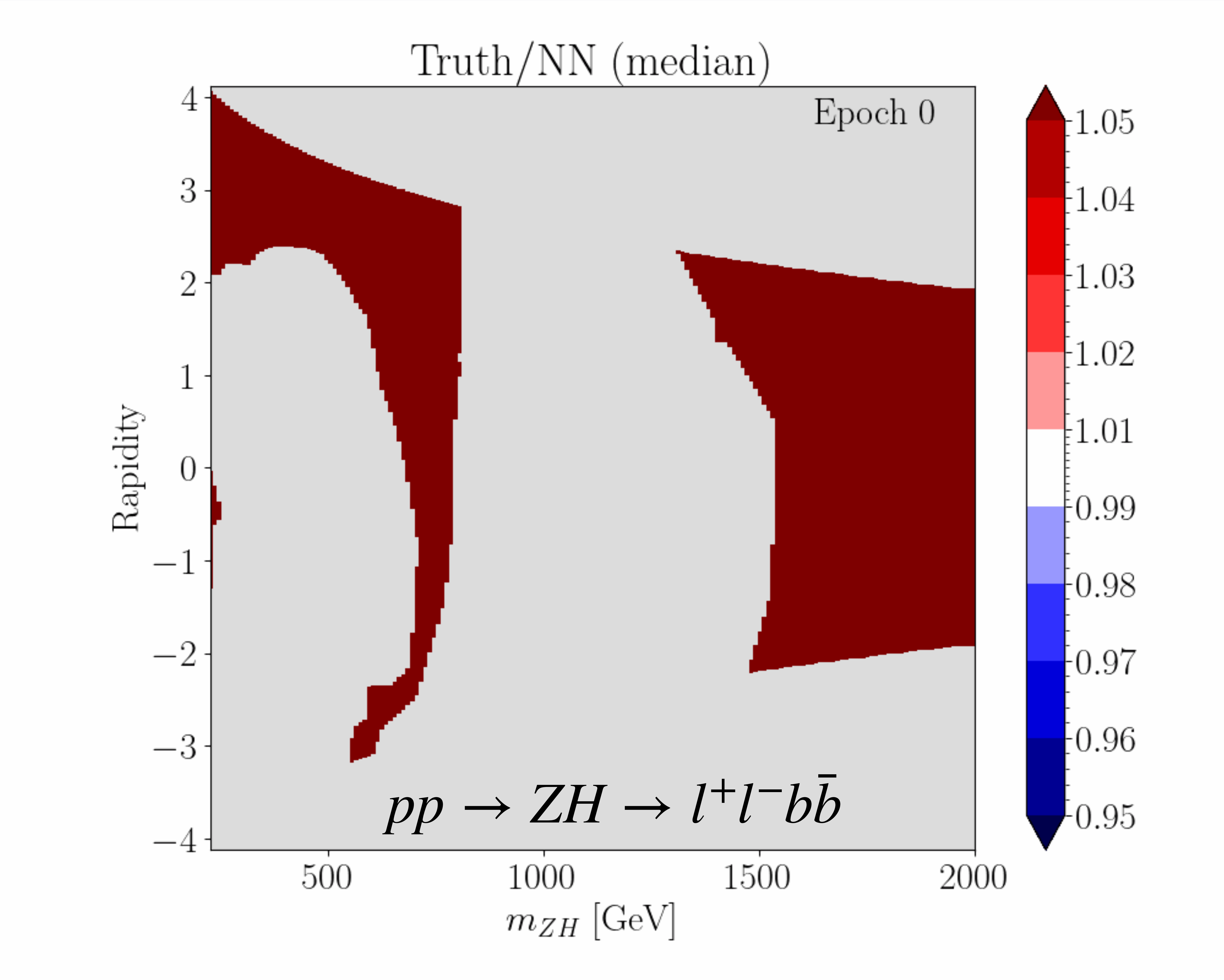
We consider processes that are dominated by statistical uncertainties, adding systematics is WIP

Statistically optimal observables from ML

Dependence of cross-section* on all (independent) kinematic variables and all EFT coefficients

$$g(x_i, c) \equiv \frac{1}{\sigma(X, c)} \frac{d\sigma(x, c)}{dx}$$

is parameterised by a feed-forward NN trained to Monte-Carlo simulations, **benchmarked** with analytical calculations



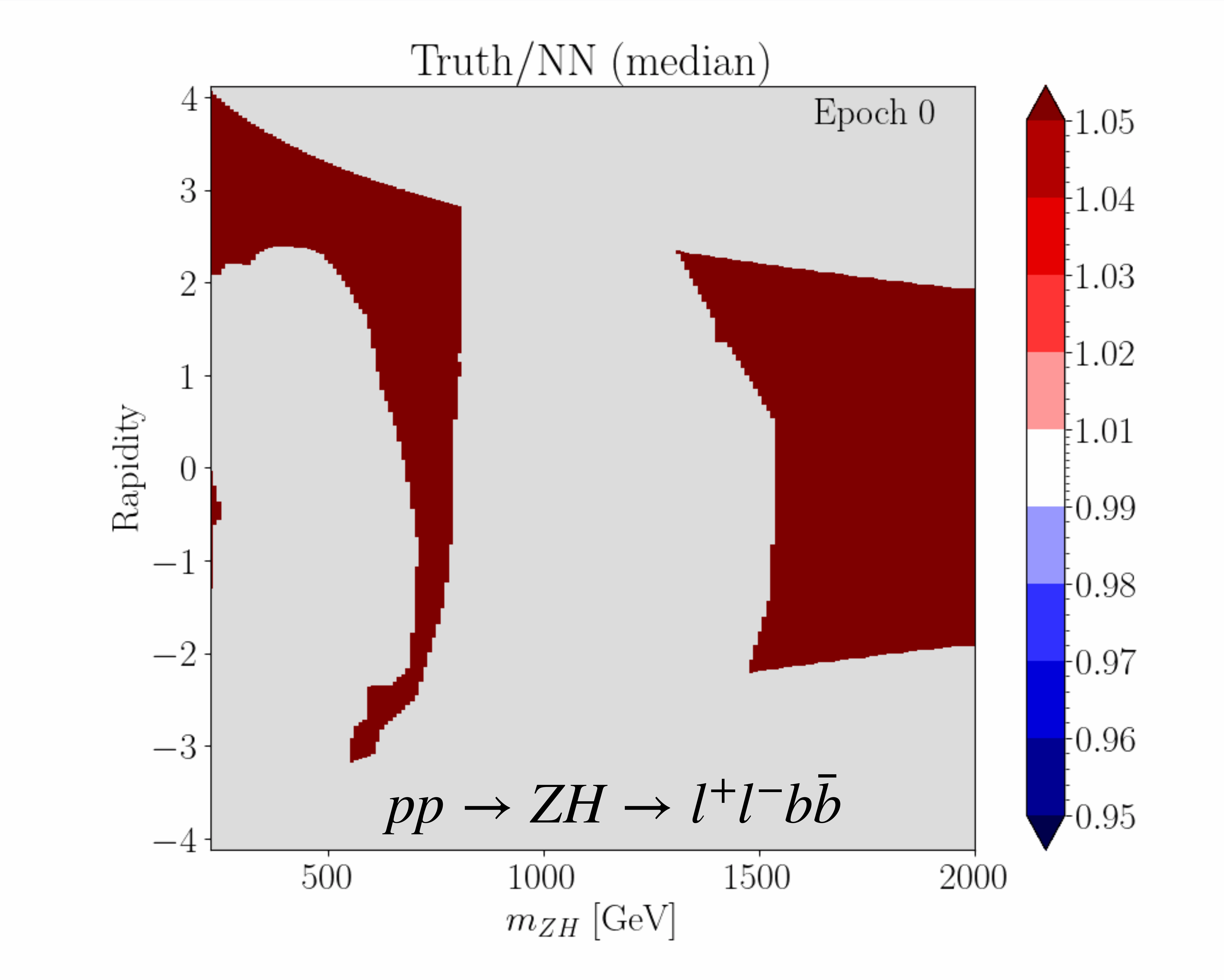
*Actually the likelihood ratio $r(x, c) = d\sigma(x, c)/d\sigma(x,0)$

Statistically optimal observables from ML

Dependence of cross-section* on all (independent) kinematic variables and all EFT coefficients

$$g(x_i, c) \equiv \frac{1}{\sigma(X, c)} \frac{d\sigma(x, c)}{dx}$$

is parameterised by a feed-forward NN trained to Monte-Carlo simulations, **benchmarked** with analytical calculations



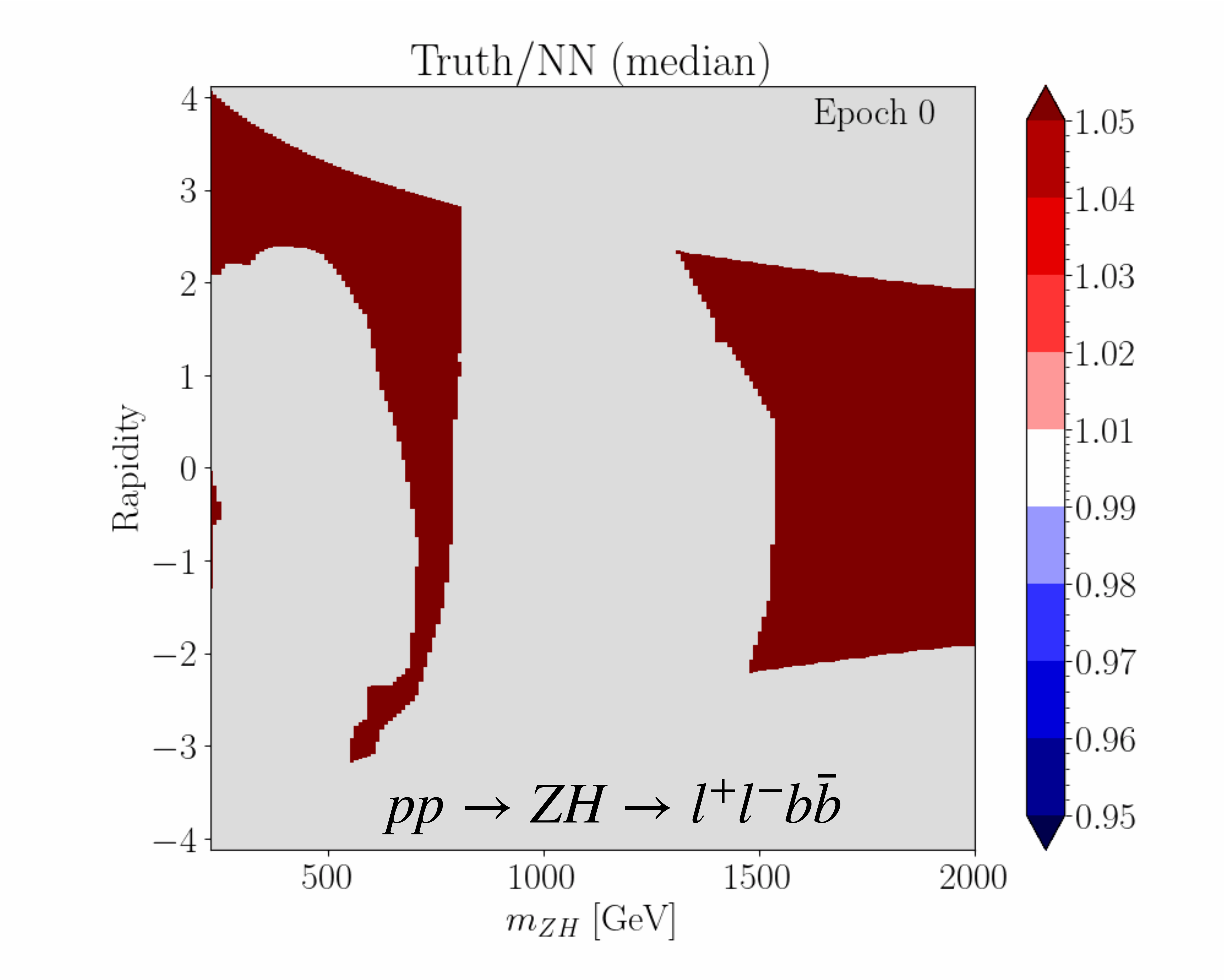
*Actually the likelihood ratio $r(x, c) = d\sigma(x, c)/d\sigma(x,0)$

Statistically optimal observables from ML

Dependence of cross-section* on all (independent) kinematic variables and all EFT coefficients

$$g(x_i, c) \equiv \frac{1}{\sigma(X, c)} \frac{d\sigma(x, c)}{dx}$$

is parameterised by a feed-forward NN trained to Monte-Carlo simulations, **benchmarked** with analytical calculations



Challenge: correctly describing tails of distributions (low stats)

*Actually the likelihood ratio $r(x, c) = d\sigma(x, c)/d\sigma(x, 0)$

Training formalism

- Finding the likelihood ratio between hypotheses H_0 and H_1 can be formulated in terms of an **optimisation problem** on two balanced training sets S_0 and S_1

$$\boxed{L[f(x)] = -\sum_{e\in S_0} w_e \log(1-f(x_e)) - \sum_{e\in S_1} w_e \log f(x_e)}$$

EFT events

SM events

$$x_e = \{p_T, Y, m_{ZH}, \ldots\}$$

Training formalism

- Finding the likelihood ratio between hypotheses H_0 and H_1 can be formulated in terms of an **optimisation problem** on two balanced training sets S_0 and S_1

$$\boxed{L[f(x)] = -\int dx \underbrace{\frac{d\sigma_0}{dx}}_{\text{EFT events}} \log(1-f) - \int dx \underbrace{\frac{d\sigma_1}{dx}}_{\text{SM events}} \log f}$$

Training formalism

- Finding the likelihood ratio between hypotheses H_0 and H_1 can be formulated in terms of an **optimisation problem** on two balanced training sets S_0 and S_1

$$L[f(x)] = -\int dx \underbrace{\frac{d\sigma_0}{dx}}_{\text{EFT events}} \log(1-f) - \int dx \underbrace{\frac{d\sigma_1}{dx}}_{\text{SM events}} \log f$$

$$\frac{\delta L}{\delta f} = 0 \implies \hat{f} = \frac{1}{1 + d\sigma_0/d\sigma_1}$$

Training formalism

- Finding the likelihood ratio between hypotheses H_0 and H_1 can be formulated in terms of an **optimisation problem** on two balanced training sets S_0 and S_1

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

EFT events

SM events

$$\boxed{\frac{\delta L}{\delta f} = 0 \implies \hat{f} = \frac{1}{1 + d\sigma_0/d\sigma_1}}$$

- A perfectly trained binary classifier is **one to one** with the likelihood ratio

Training formalism

- Finding the likelihood ratio between hypotheses H_0 and H_1 can be formulated in terms of an **optimisation problem** on two balanced training sets S_0 and S_1

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

EFT events

SM events

$$\boxed{\frac{\delta L}{\delta f} = 0 \implies \hat{f} = \frac{1}{1 + d\sigma_0/d\sigma_1}}$$

- A perfectly trained binary classifier is **one to one** with the likelihood ratio

- Equivalent formulations exist, e.g. in terms of a quadratic loss

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

Training formalism

- Exploit the polynomial structure of the EFT coefficients to separate the learning problem

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha}$$

1. Linear term

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha + c^2 n_\beta}$$

2. Quadratic term

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = \ldots + c_1 c_2 n_\gamma}$$

3. Cross term

- Cross section **positivity** can be enforced through either Lagrange multipliers or a final ReLU

Scaling behaviour

Our method scales efficiently with the number of EFT coefficients and can be parallelised

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(1)}}$$

Trained in parallel

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(n)}}$$

1. Linear terms

Scaling behaviour

Our method scales efficiently with the number of EFT coefficients and can be parallelised

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(1)}}$$

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(1)} + c^2 n_\beta^{(1)}}$$

Trained in parallel

Trained in parallel

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(n)}}$$

$$\boxed{\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(n)} + c^2 n_\beta^{(n)}}$$

1. Linear terms

2. Quadratic terms

Scaling behaviour

Our method scales efficiently with the number of EFT coefficients and can be parallelised

$$\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(1)}$$

Trained in parallel

$$\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(n)}$$

1. Linear terms

$$\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(1)} + c^2 n_\beta^{(1)}$$

Trained in parallel

$$\frac{d\sigma_0}{d\sigma_1} = 1 + c \cdot n_\alpha^{(n)} + c^2 n_\beta^{(n)}$$

2. Quadratic terms

$$\frac{d\sigma_0}{d\sigma_1} = \ldots + c_1 c_2 n_\gamma^{(1)}$$

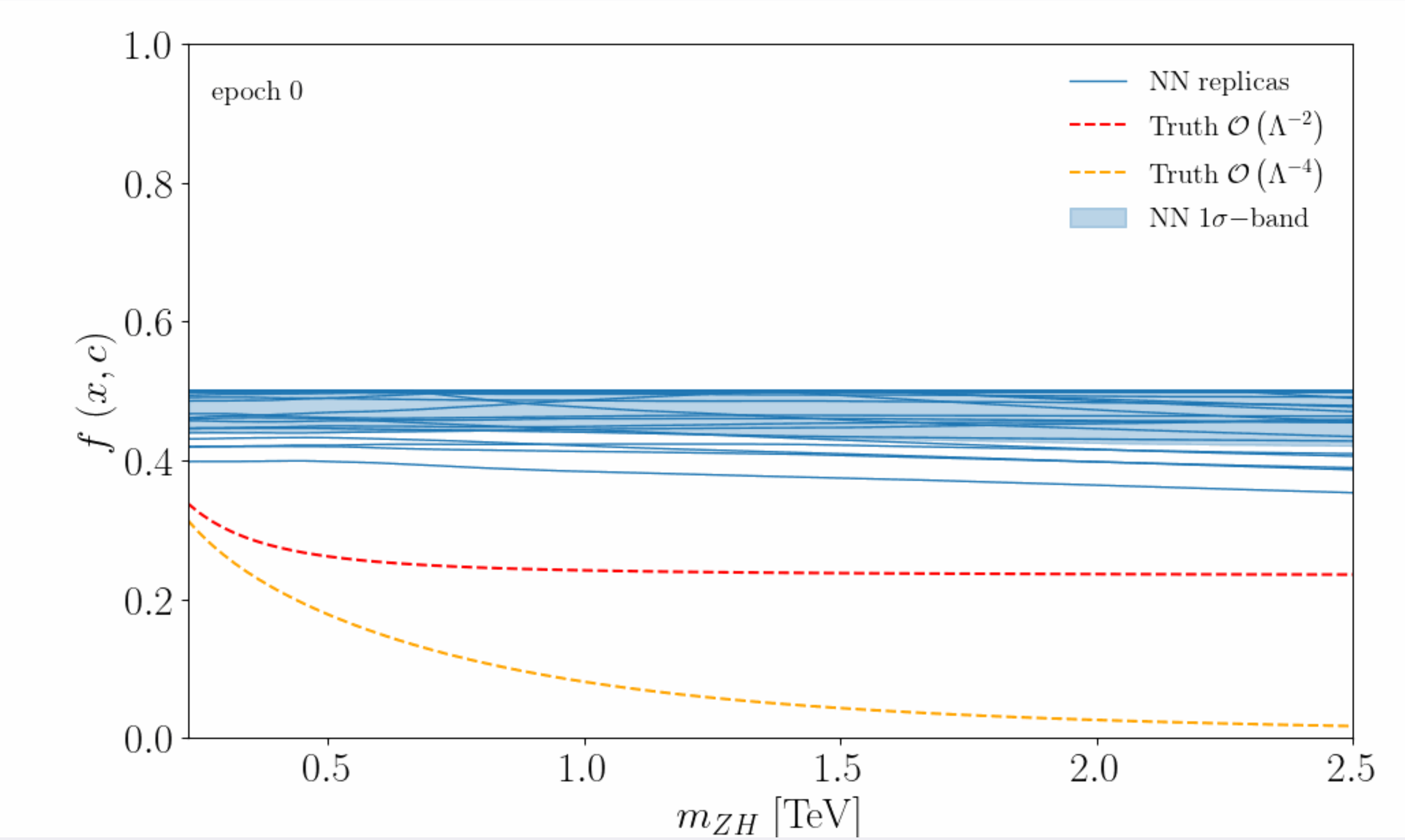
Trained in parallel

$$\frac{d\sigma_0}{d\sigma_1} = \ldots + c_1 c_2 n_\gamma^{(n_{\text{pairs}})}$$

3. Cross terms

The Monte Carlo replica method

We train a collection of 50 NN instances on independent MC datasets to estimate the **NN uncertainty**

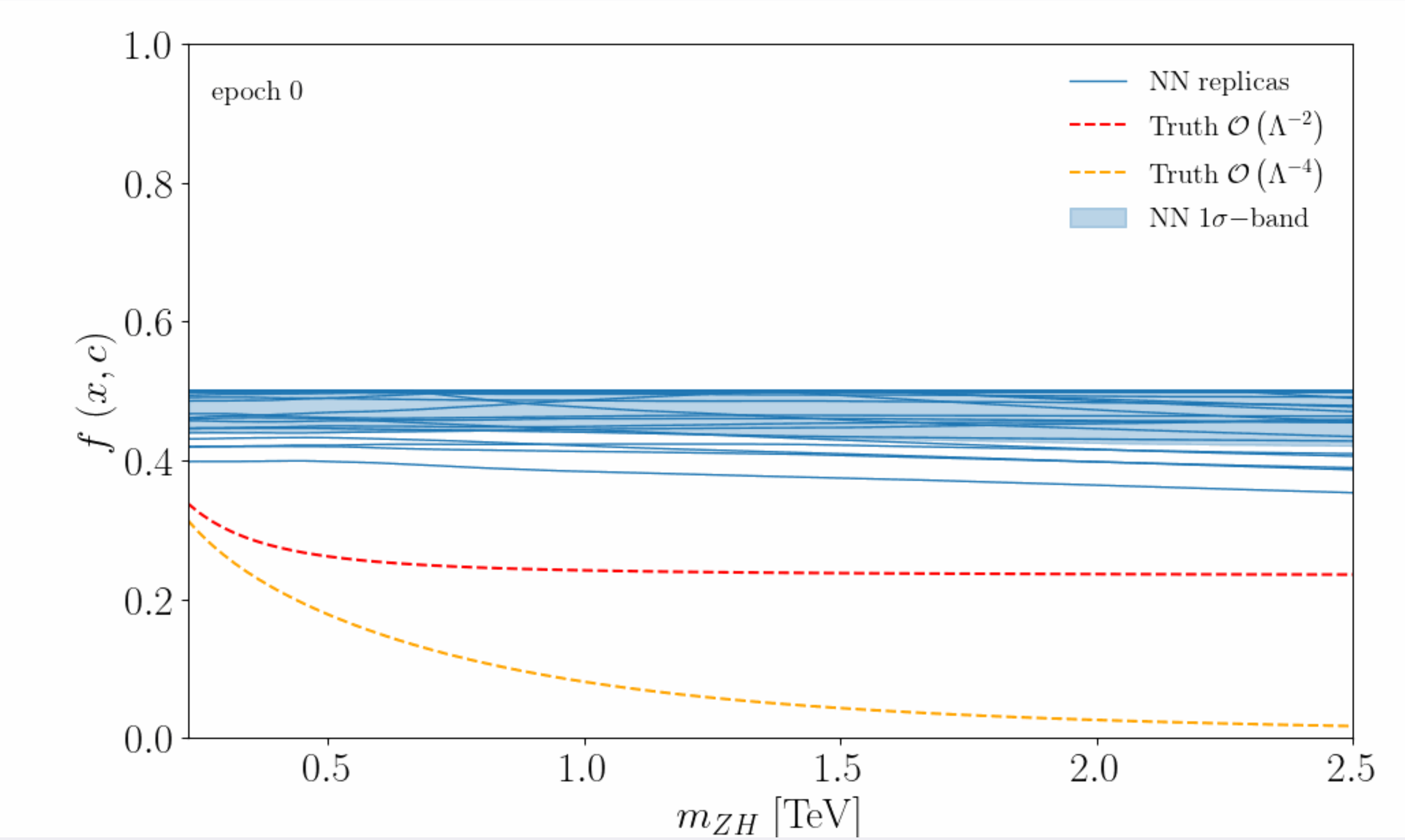


Training settings

- LO QCD MC dataset (100K)
- PyTorch Implementation
- Validation loss is monitored to avoid overfitting
- Training takes ~30 min per core (1 replica/core)
- 5 hidden layers (30 units)
- ReLU activation functions

The Monte Carlo replica method

We train a collection of 50 NN instances on independent MC datasets to estimate the **NN uncertainty**

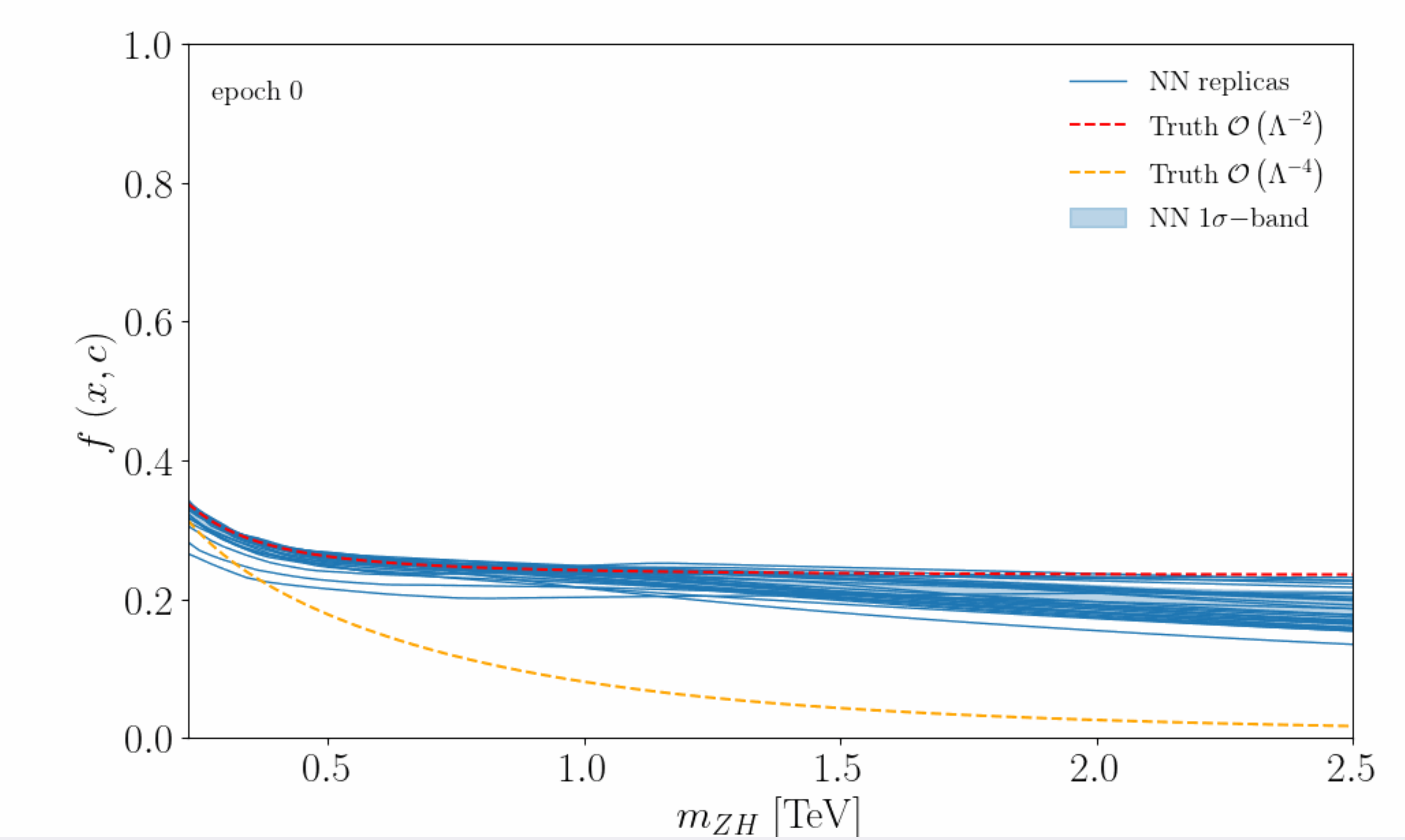


Training settings

- LO QCD MC dataset (100K)
- PyTorch Implementation
- Validation loss is monitored to avoid overfitting
- Training takes ~30 min per core (1 replica/core)
- 5 hidden layers (30 units)
- ReLU activation functions

The Monte Carlo replica method

We train a collection of 50 NN instances on independent MC datasets to estimate the **NN uncertainty**

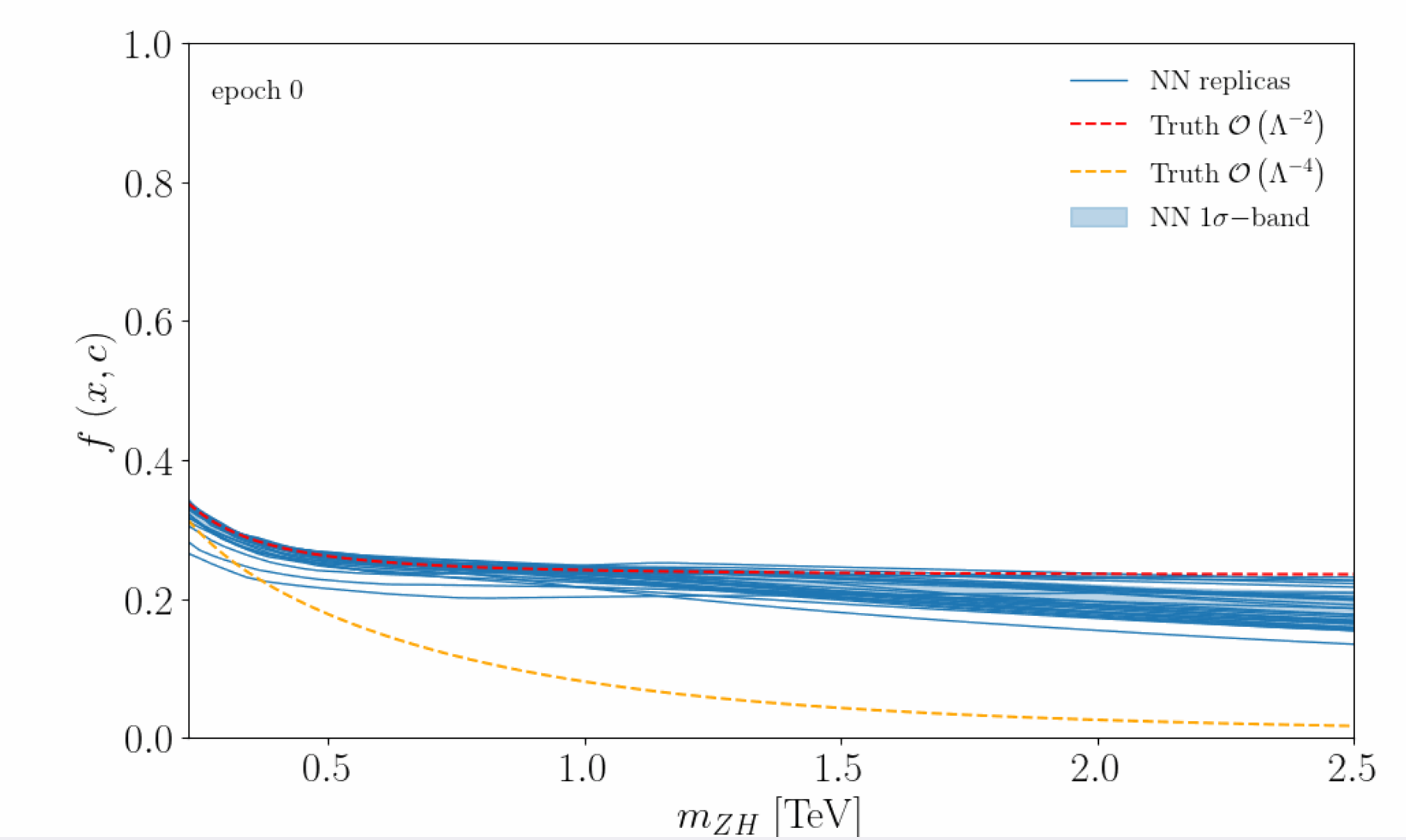


Training settings

- LO QCD MC dataset (100K)
- PyTorch Implementation
- Validation loss is monitored to avoid overfitting
- Training takes ~30 min per core (1 replica/core)
- 5 hidden layers (30 units)
- ReLU activation functions

The Monte Carlo replica method

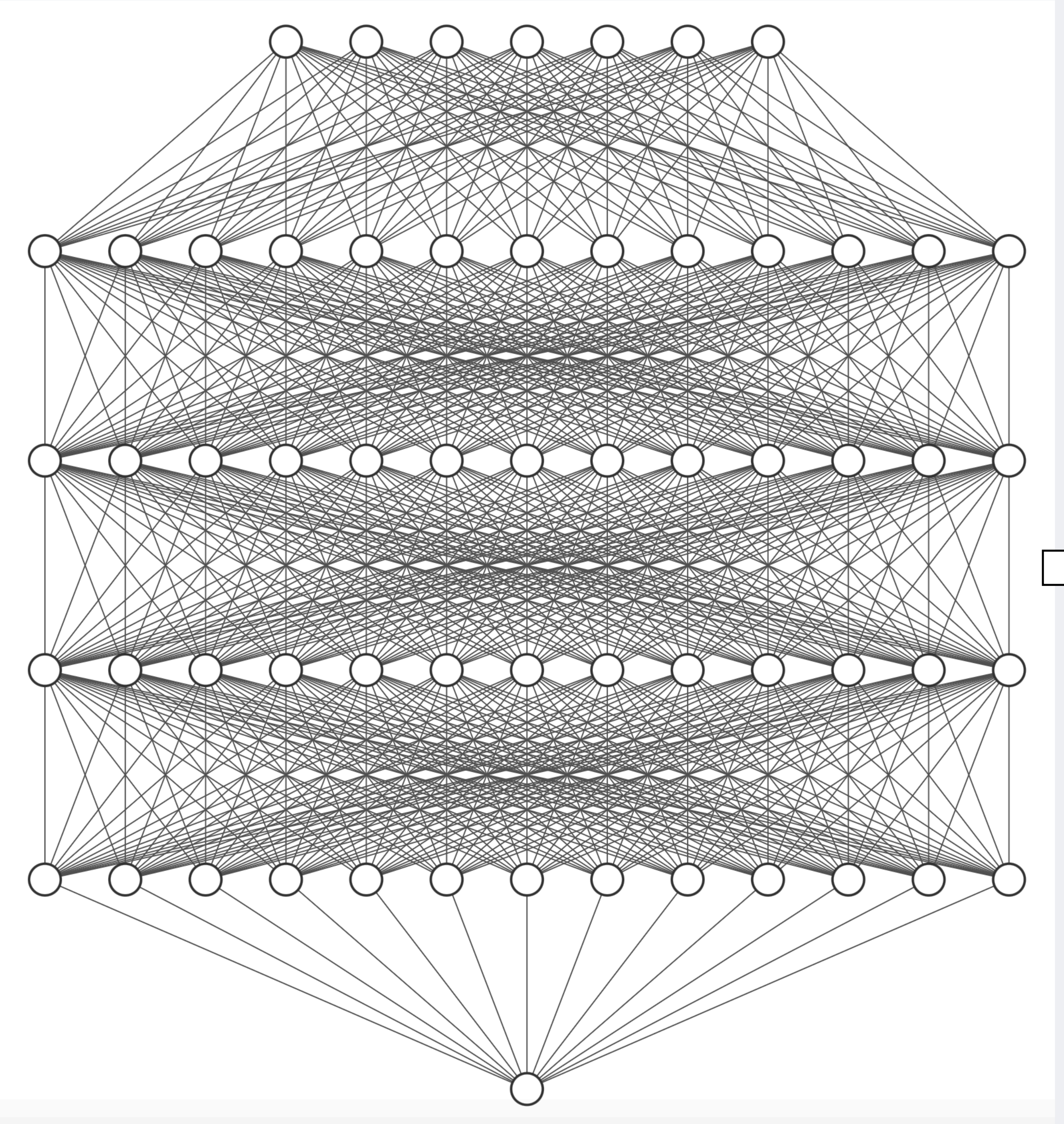
We train a collection of 50 NN instances on independent MC datasets to estimate the **NN uncertainty**



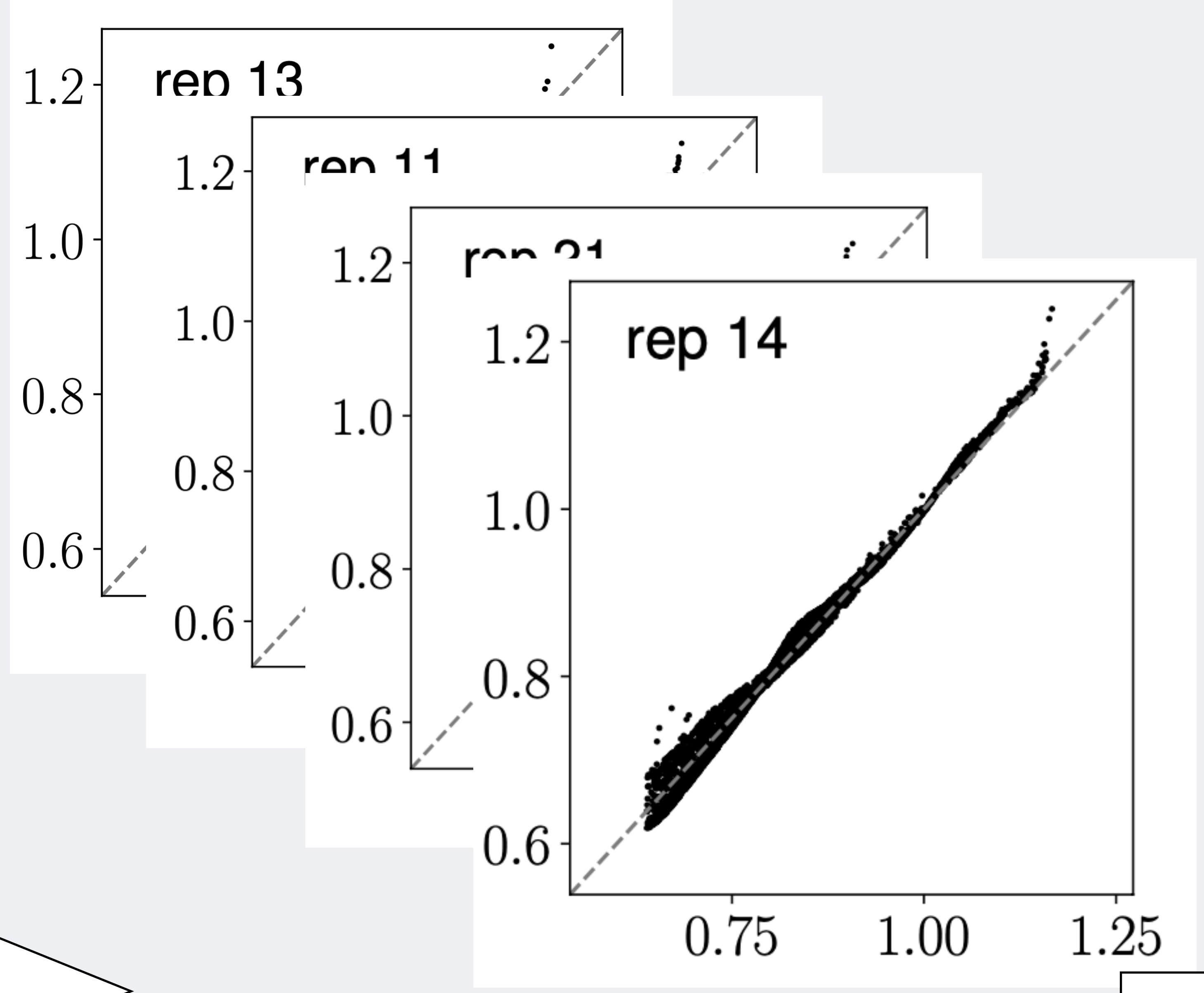
Training settings

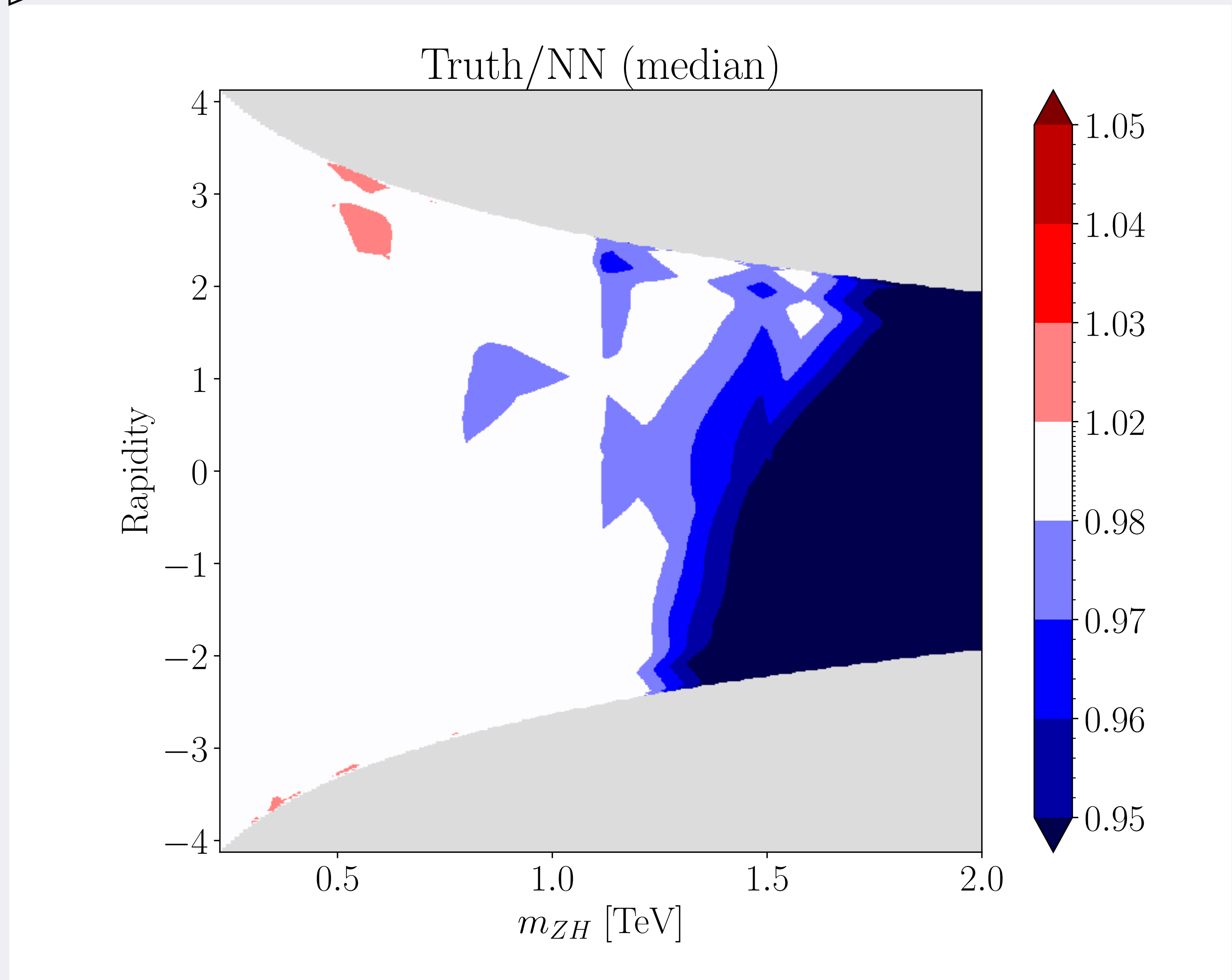
- LO QCD MC dataset (100K)
- PyTorch Implementation
- Validation loss is monitored to avoid overfitting
- Training takes ~30 min per core (1 replica/core)
- 5 hidden layers (30 units)
- ReLU activation functions

Training

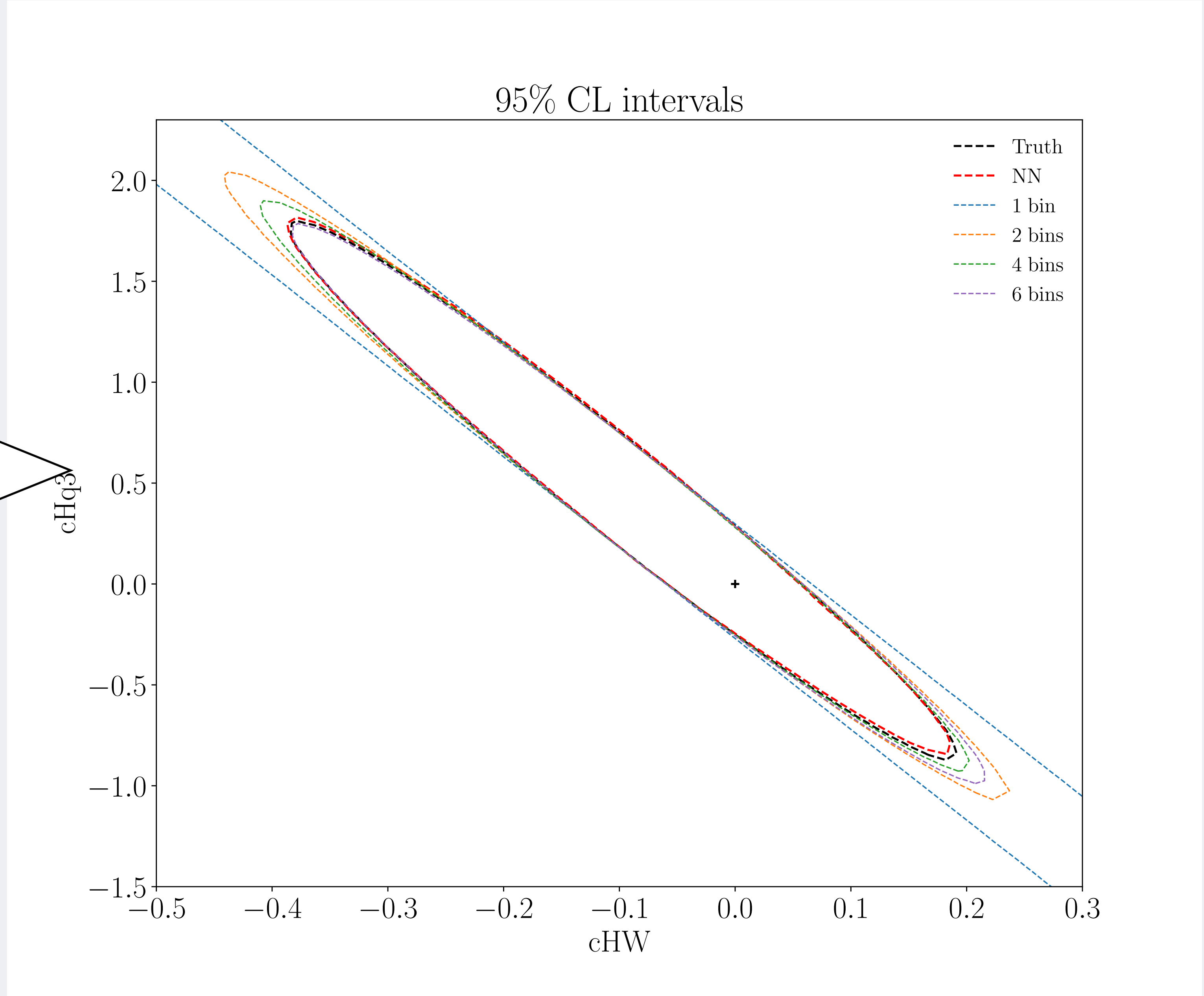


$$L[f(x)] = -\int dx \frac{d\sigma_0}{dx} \log(1-f) - \int dx \frac{d\sigma_1}{dx} \log f$$





Inference



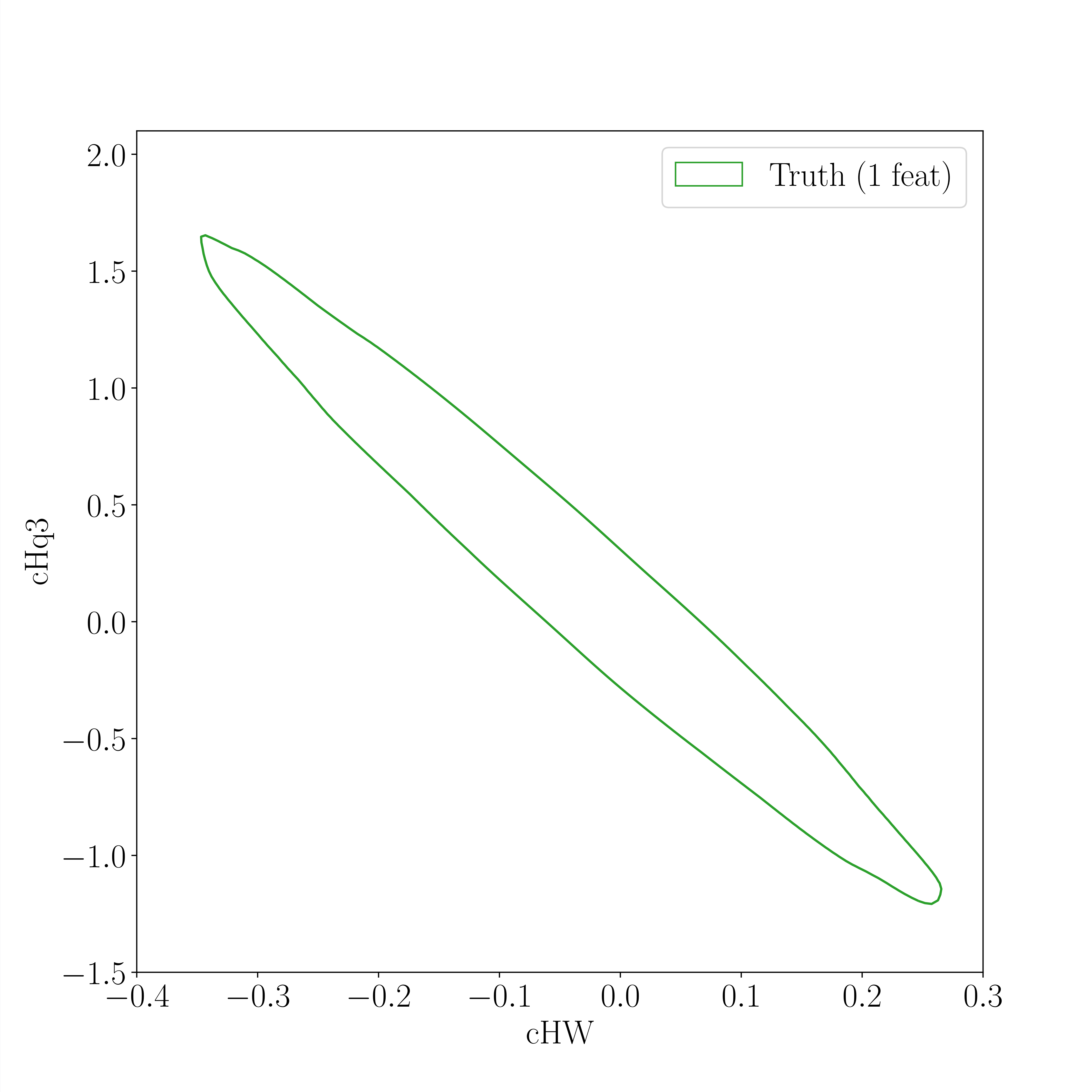
Profile Likelihood Ratio

$$q_c = 2\left[\nu(c) - \nu(\hat{c}) - \sum_i \log \frac{d\sigma(x_i, c)}{d\sigma(x_i, \hat{c})}\right]$$

Results

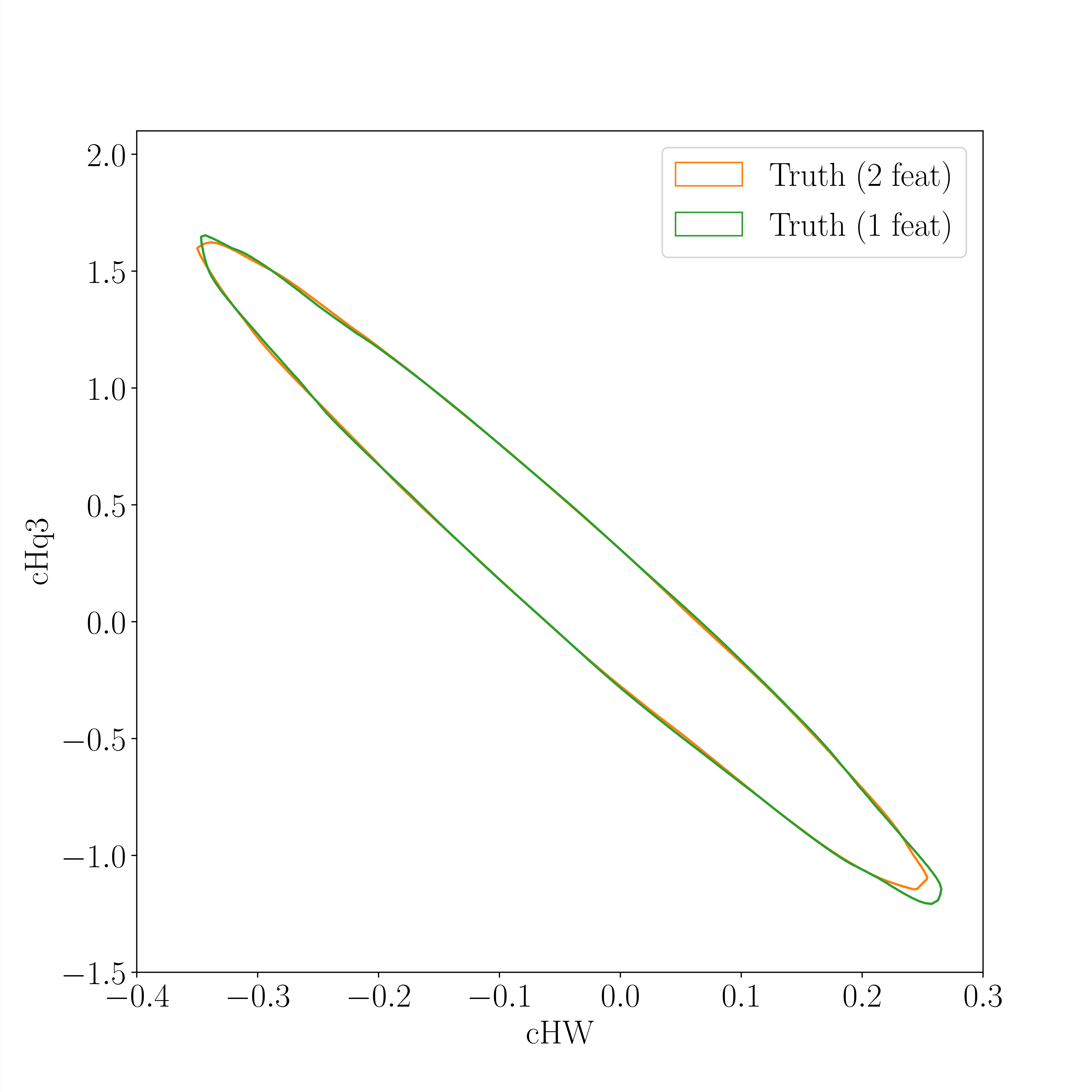
- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]

- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included



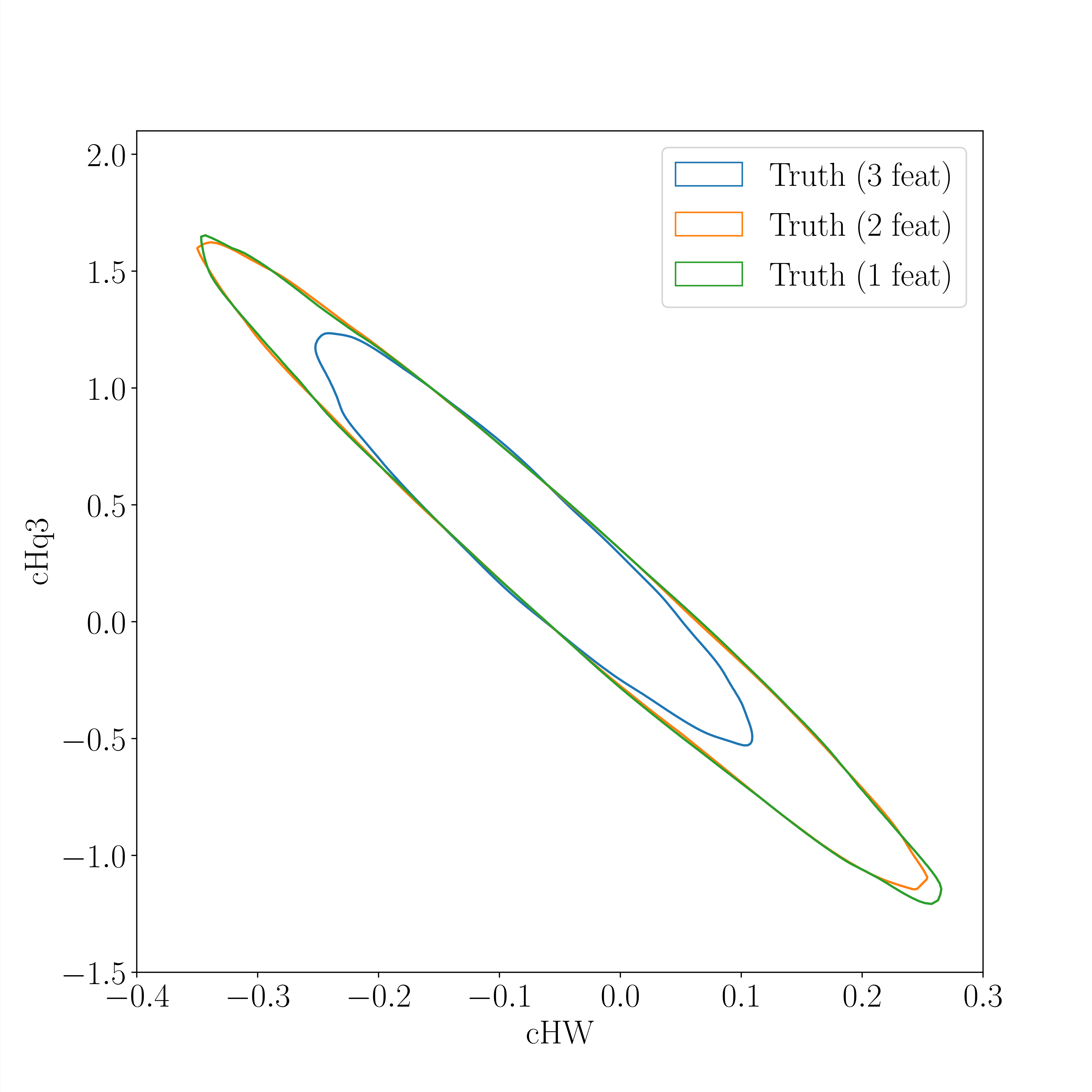
Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included



Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included

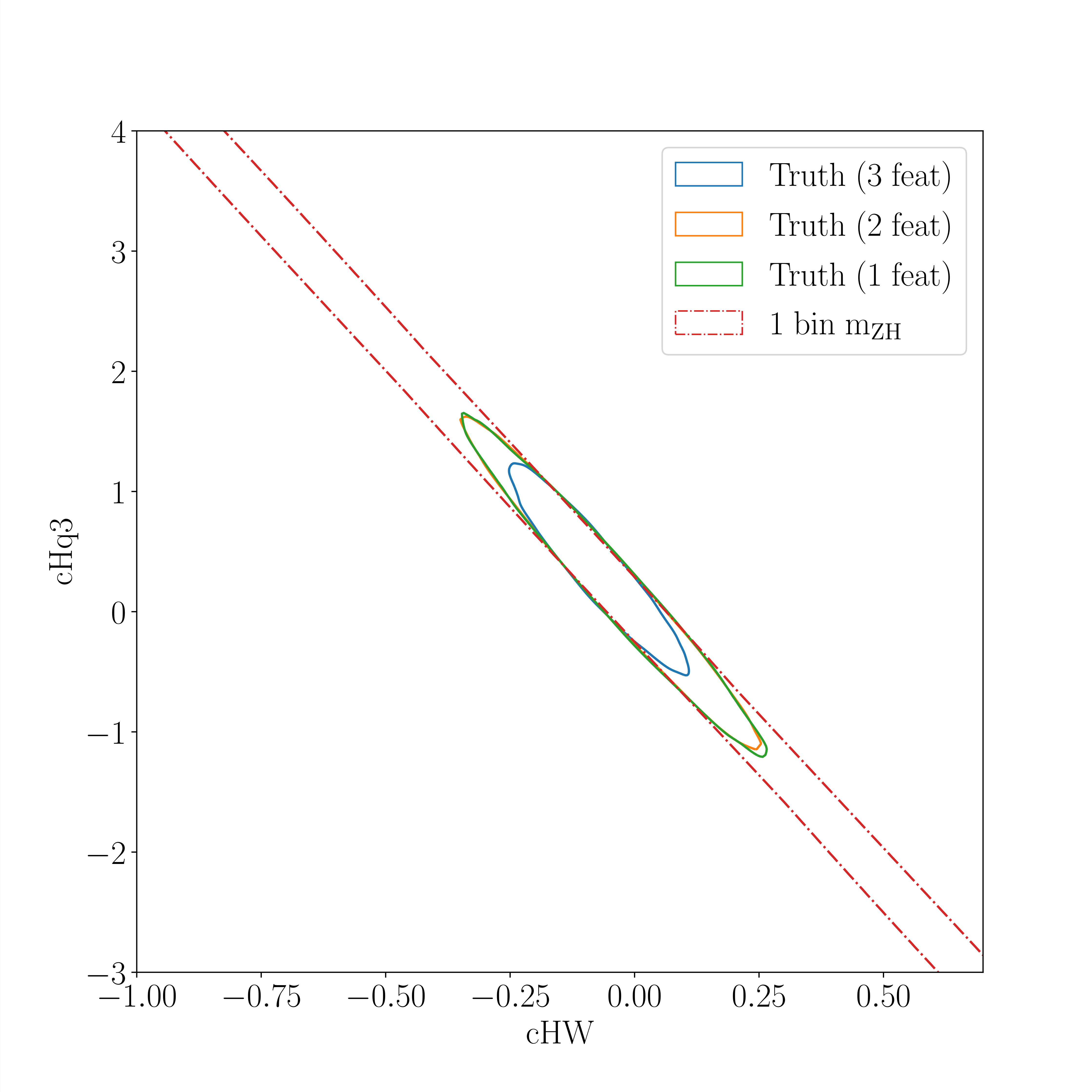


Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]

- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included

- Binned analyses eventually **saturate** to unbinned case

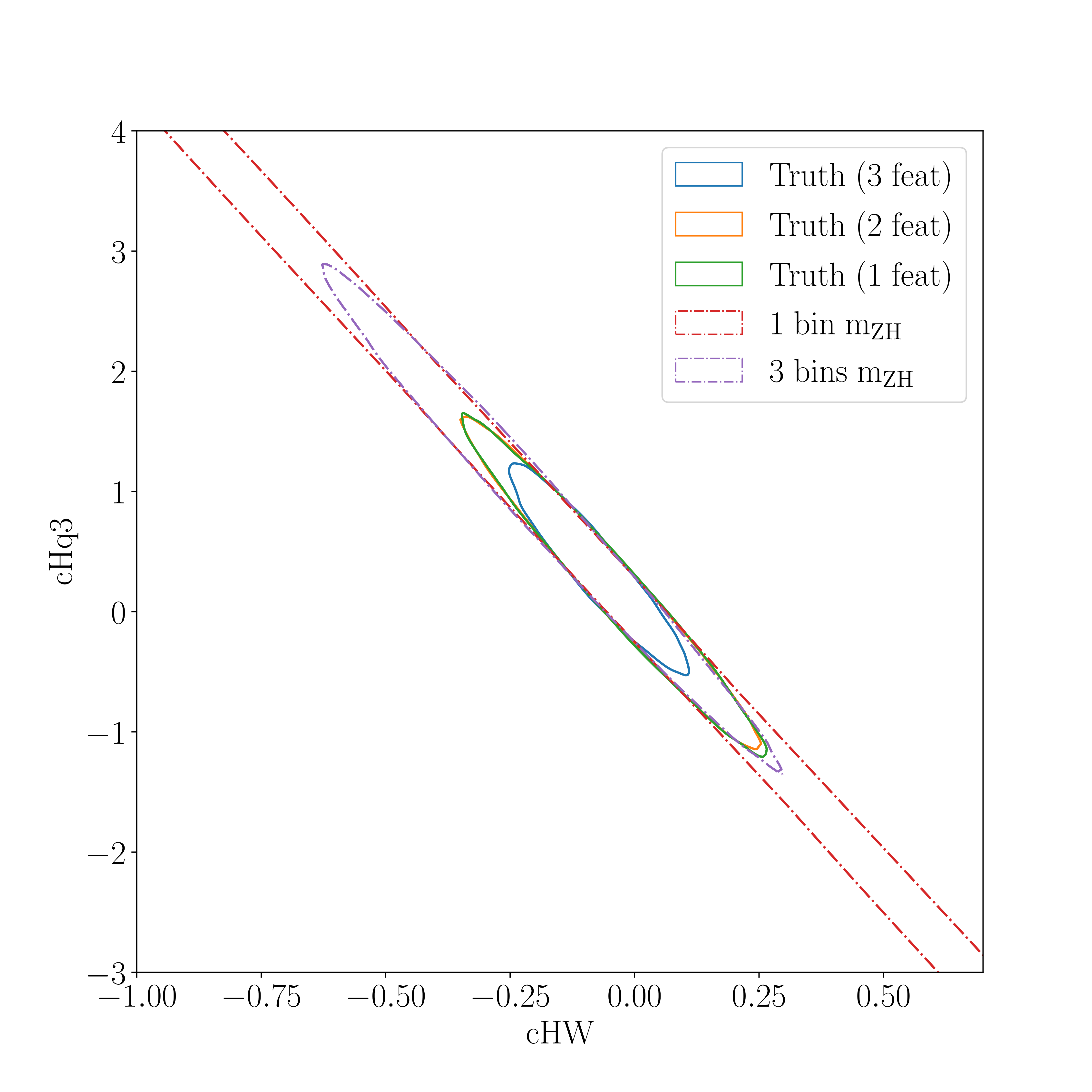


Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]

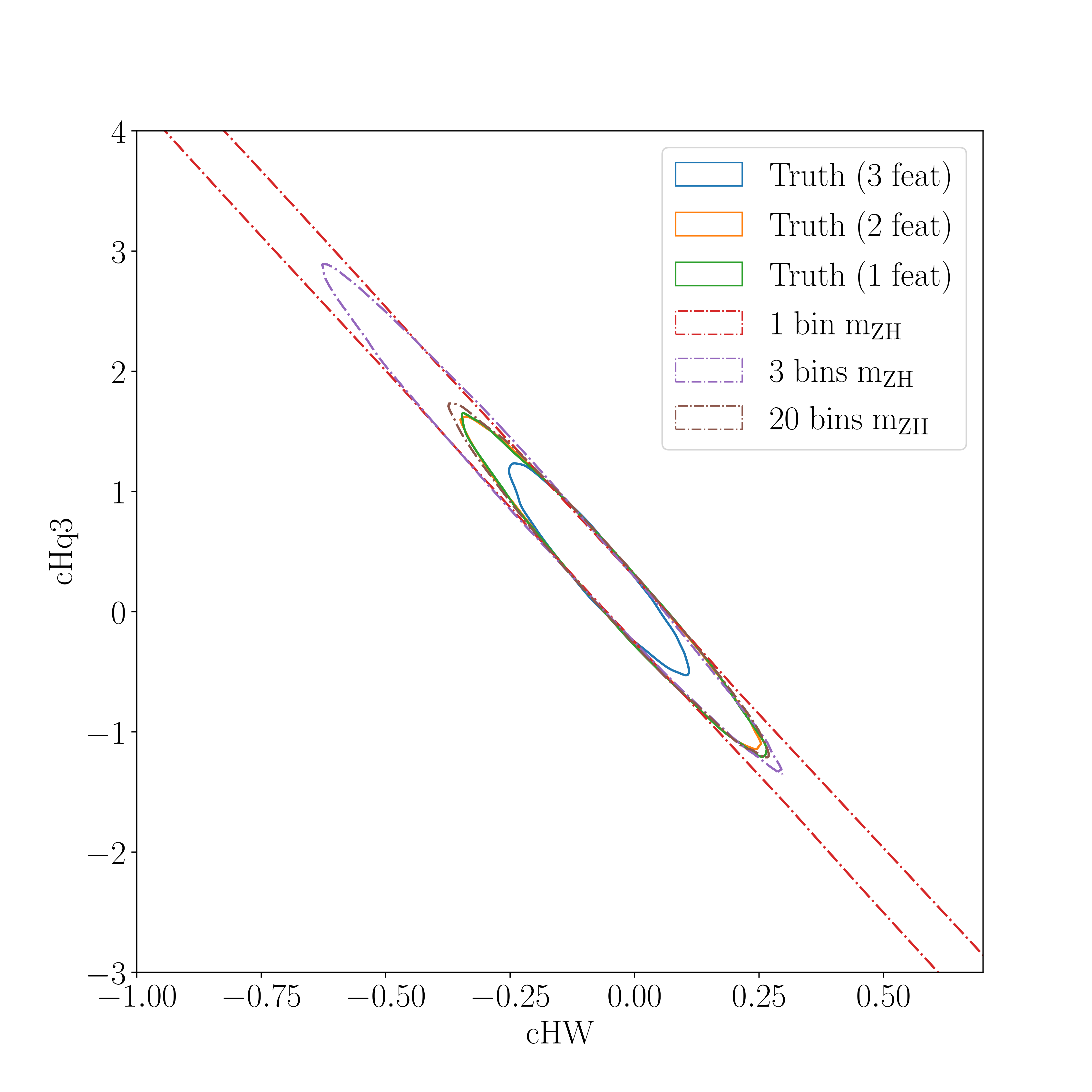
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included

- Binned analyses eventually **saturate** to unbinned case



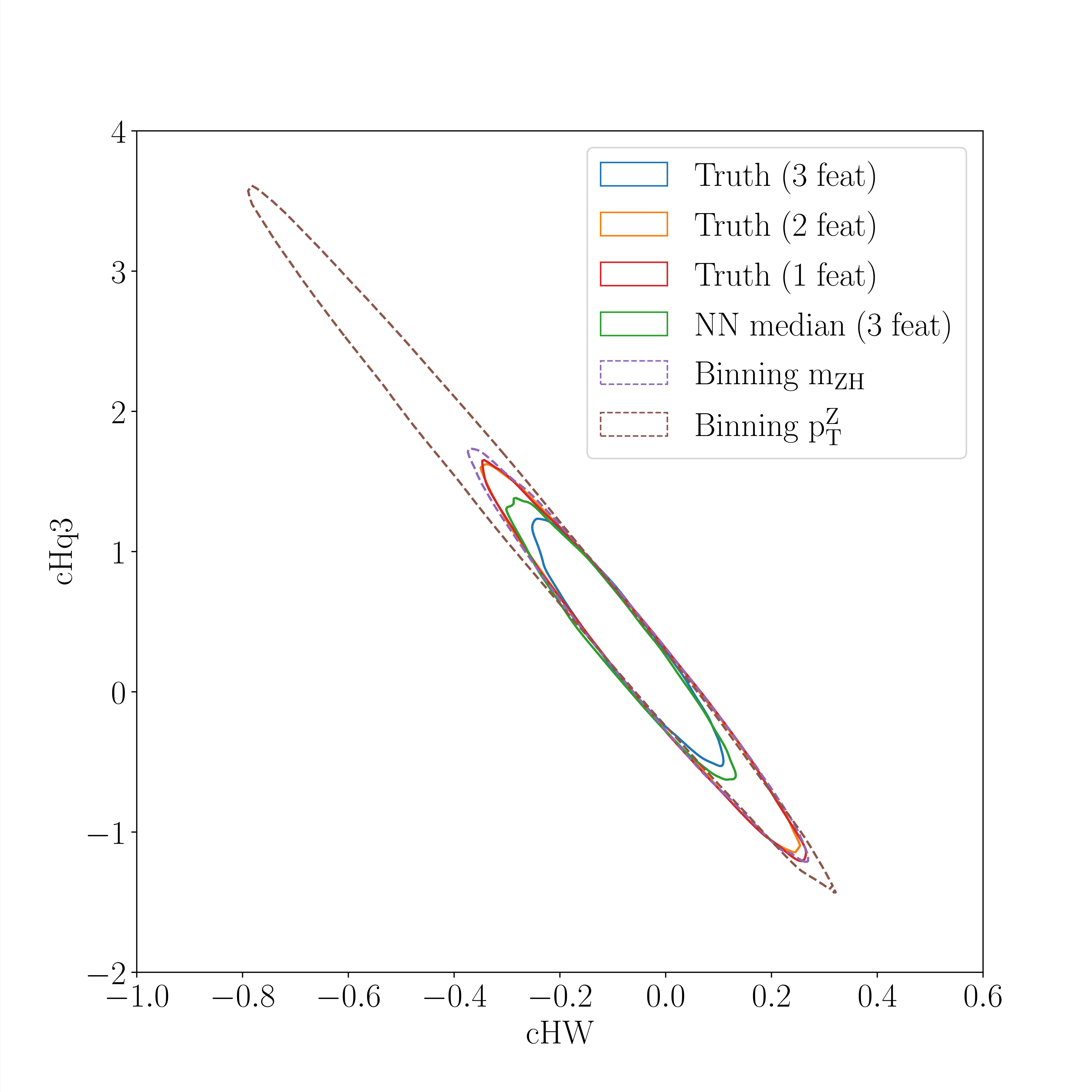
Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included
- Binned analyses eventually **saturate** to unbinned case



Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included
- Binned analyses eventually **saturate** to unbinned case
- Binning in p_T^Z is **suboptimal**, i.e. STXS



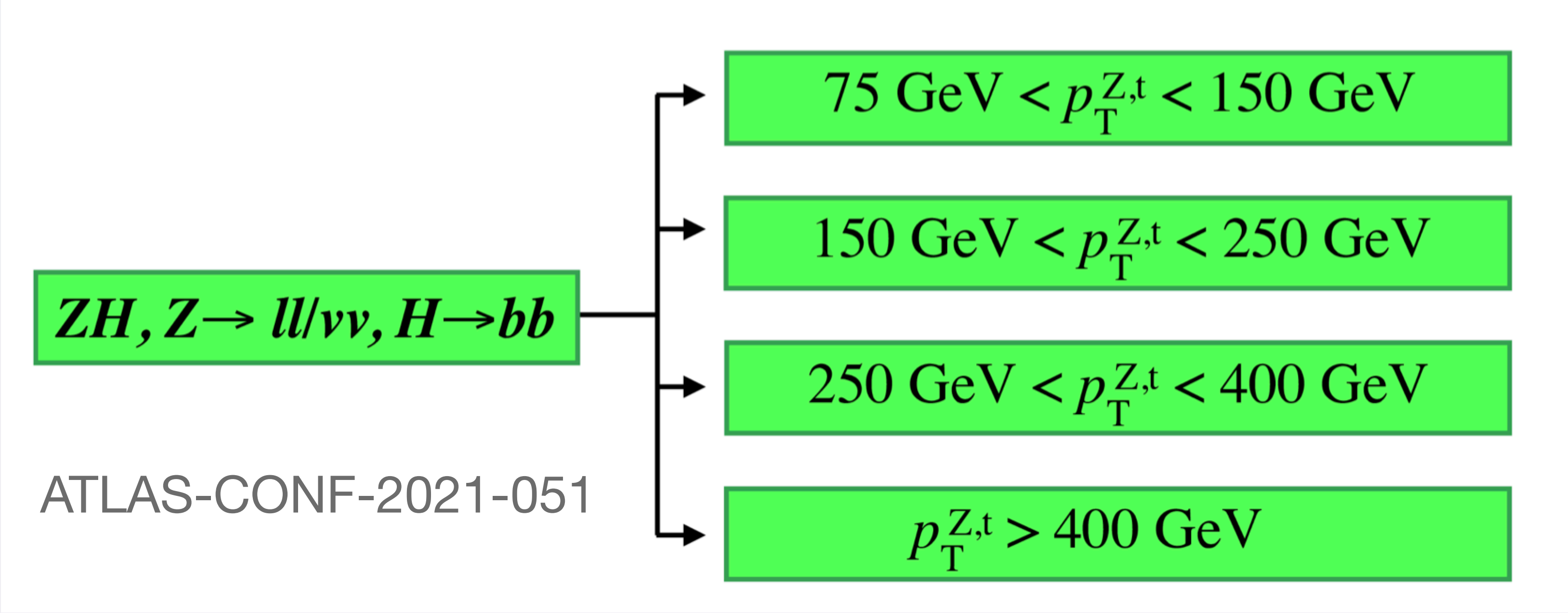
Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]

- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included

- Binned analyses eventually **saturate** to unbinned case

- Binning in p_T^Z is **suboptimal**, i.e. STXS



$ZH, Z \to ll/\nu\nu, H \to bb$

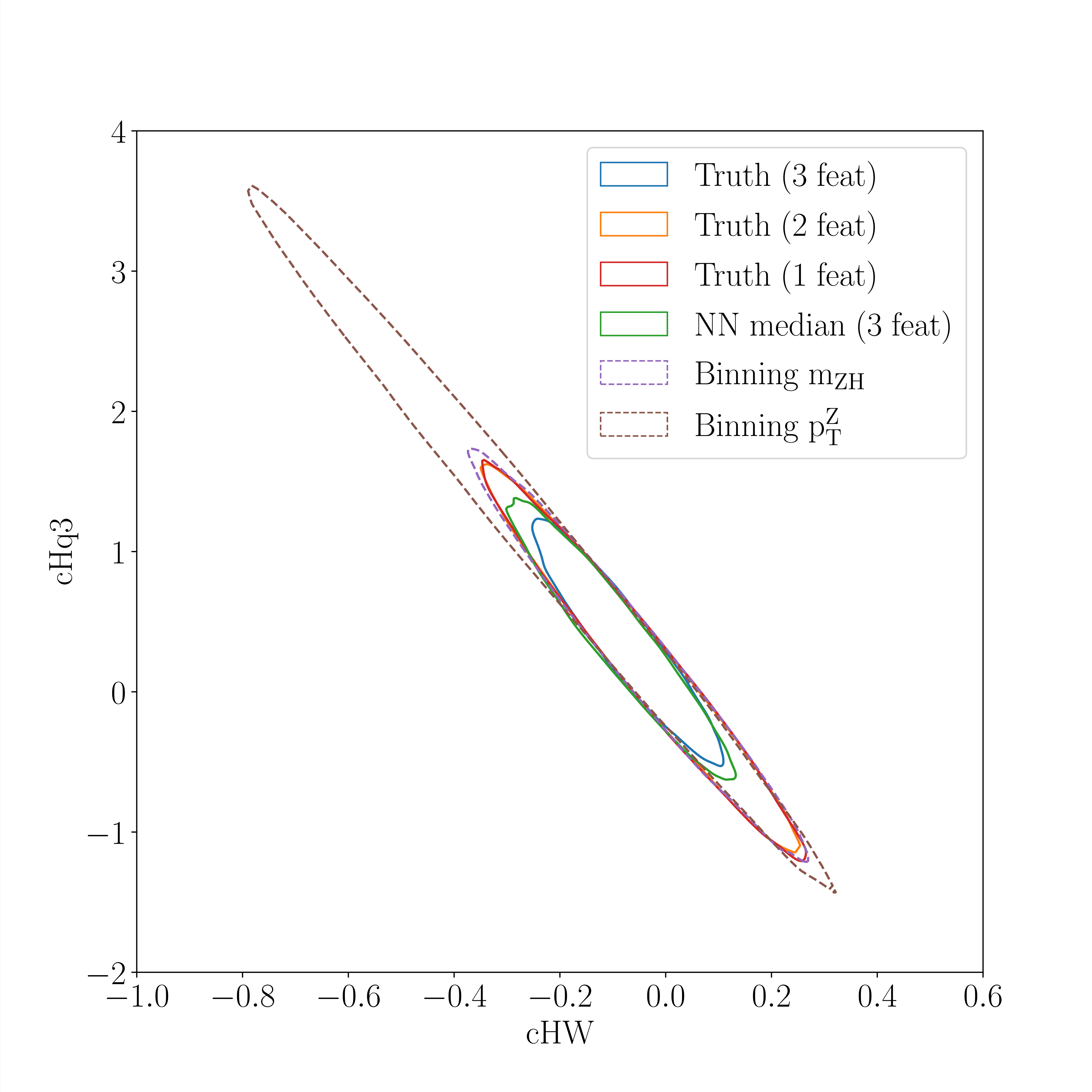
75 GeV $< p_T^{Z,t} <$ 150 GeV

150 GeV $< p_T^{Z,t} <$ 250 GeV

250 GeV $< p_T^{Z,t} <$ 400 GeV

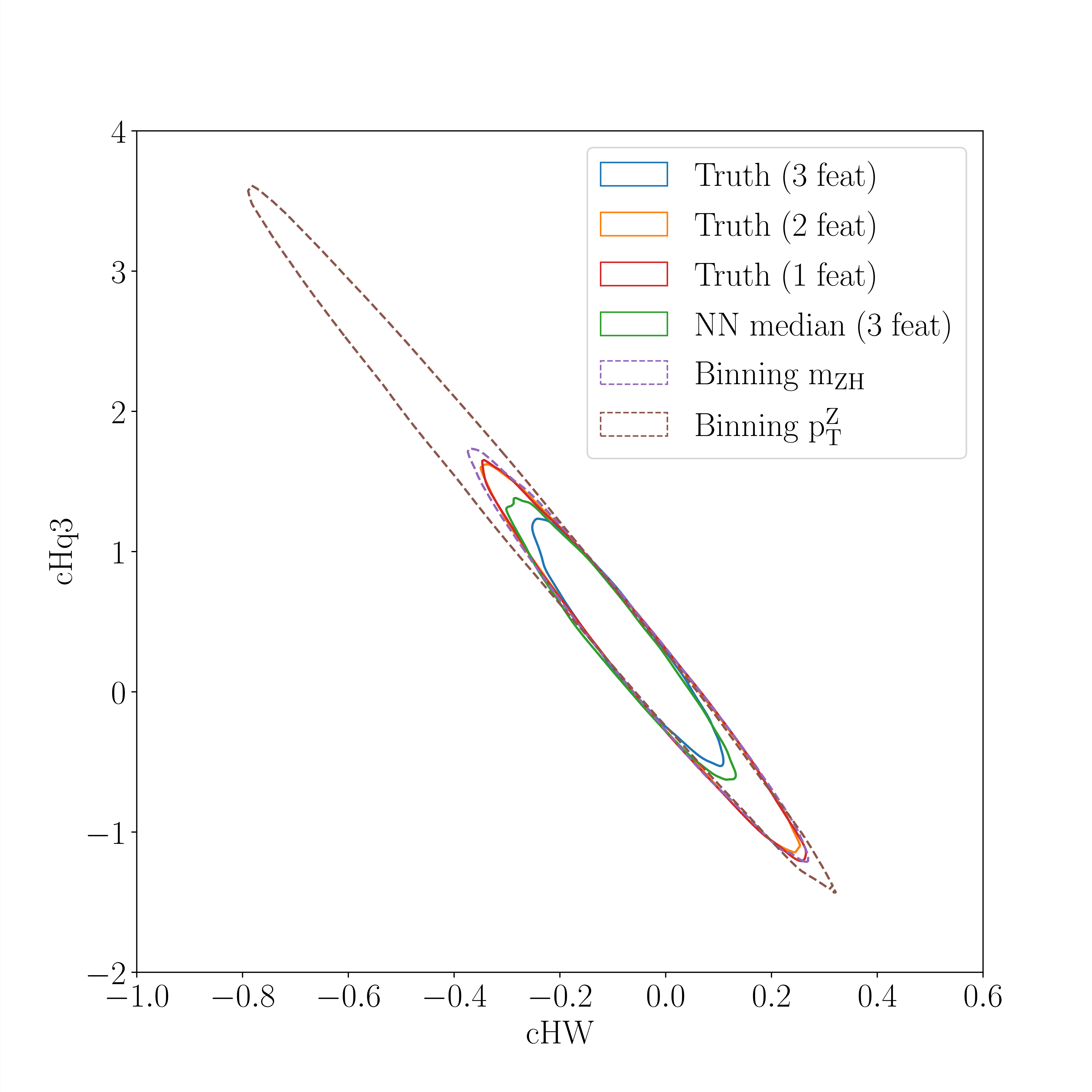
$p_T^{Z,t} >$ 400 GeV

ATLAS-CONF-2021-051



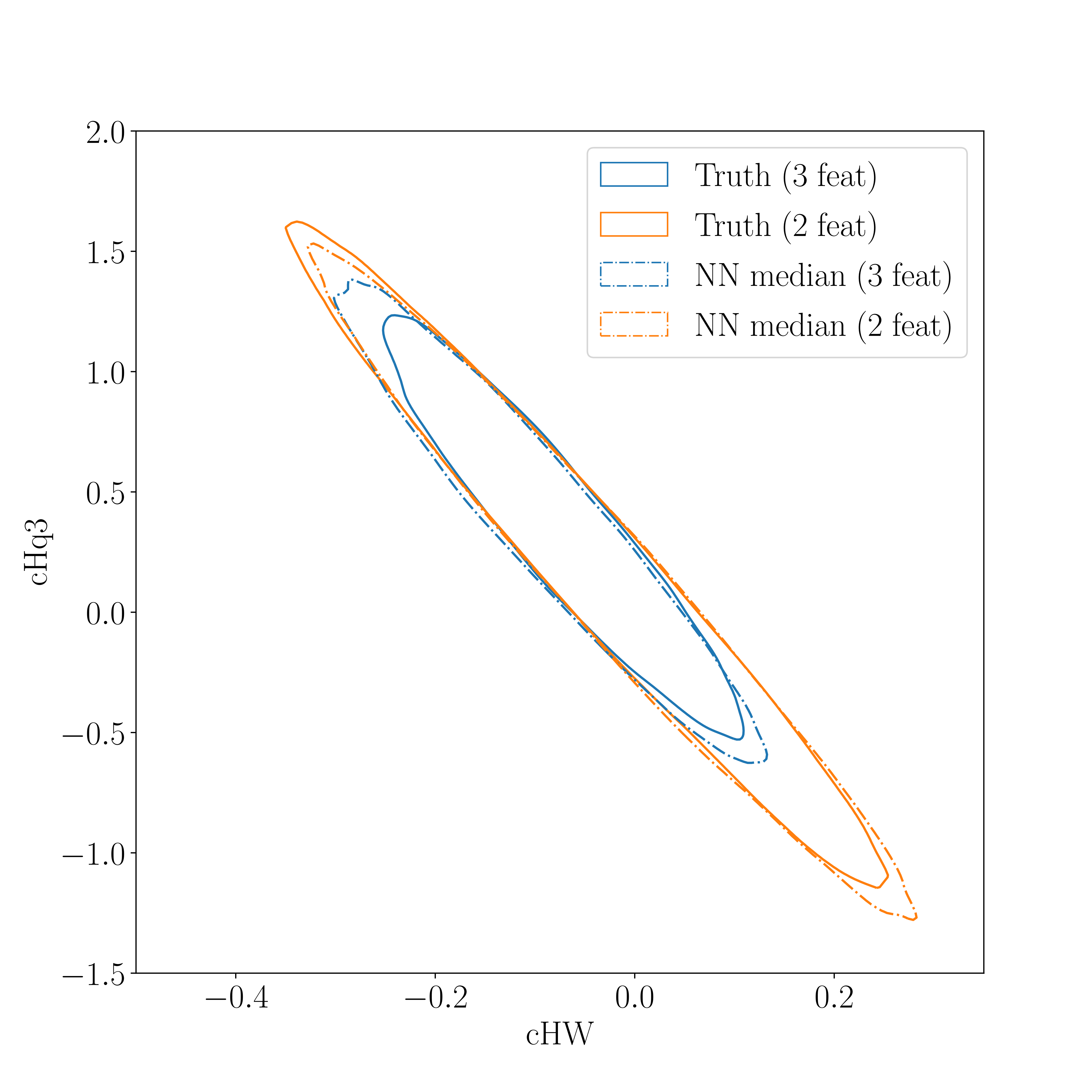
Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included
- Binned analyses eventually **saturate** to unbinned case
- Binning in p_T^Z is **suboptimal**, i.e. STXS



Results

- Use **Nested Sampling** to draw posterior samples F. Feroz et al. [1306.2144]
- A ground truth analysis with a combination of either m_{ZH} , Y or p_T^Z performs best with **all features** included
- Binned analyses eventually **saturate** to unbinned case
- Binning in p_T^Z is **suboptimal**, i.e. STXS
- NN reproduces the truth (**within uncertainties?**)

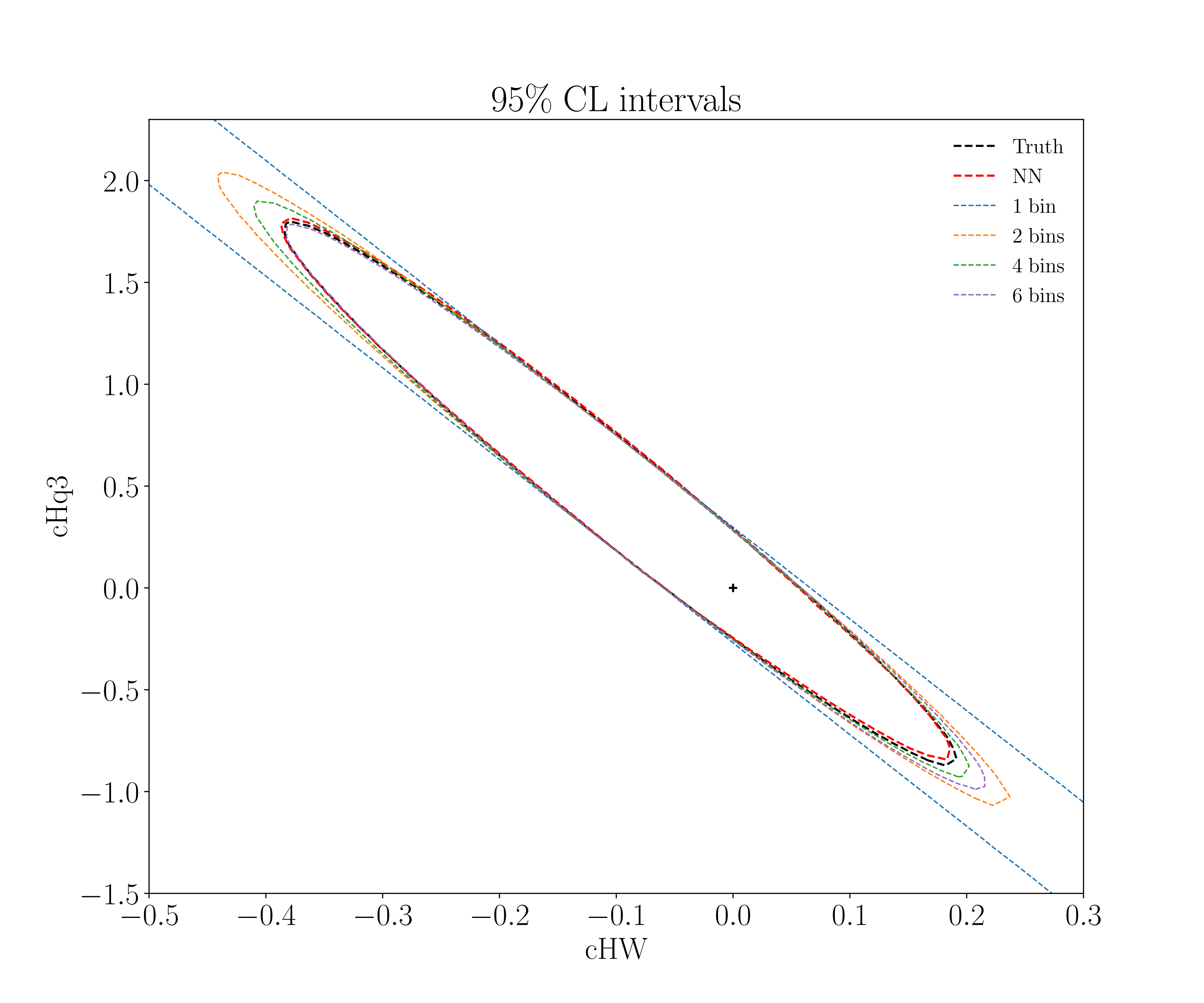


Conclusion and Outlook

- The SMEFT allows for a **model independent** framework to search for New Physics, taking **correlations** from different sectors into account

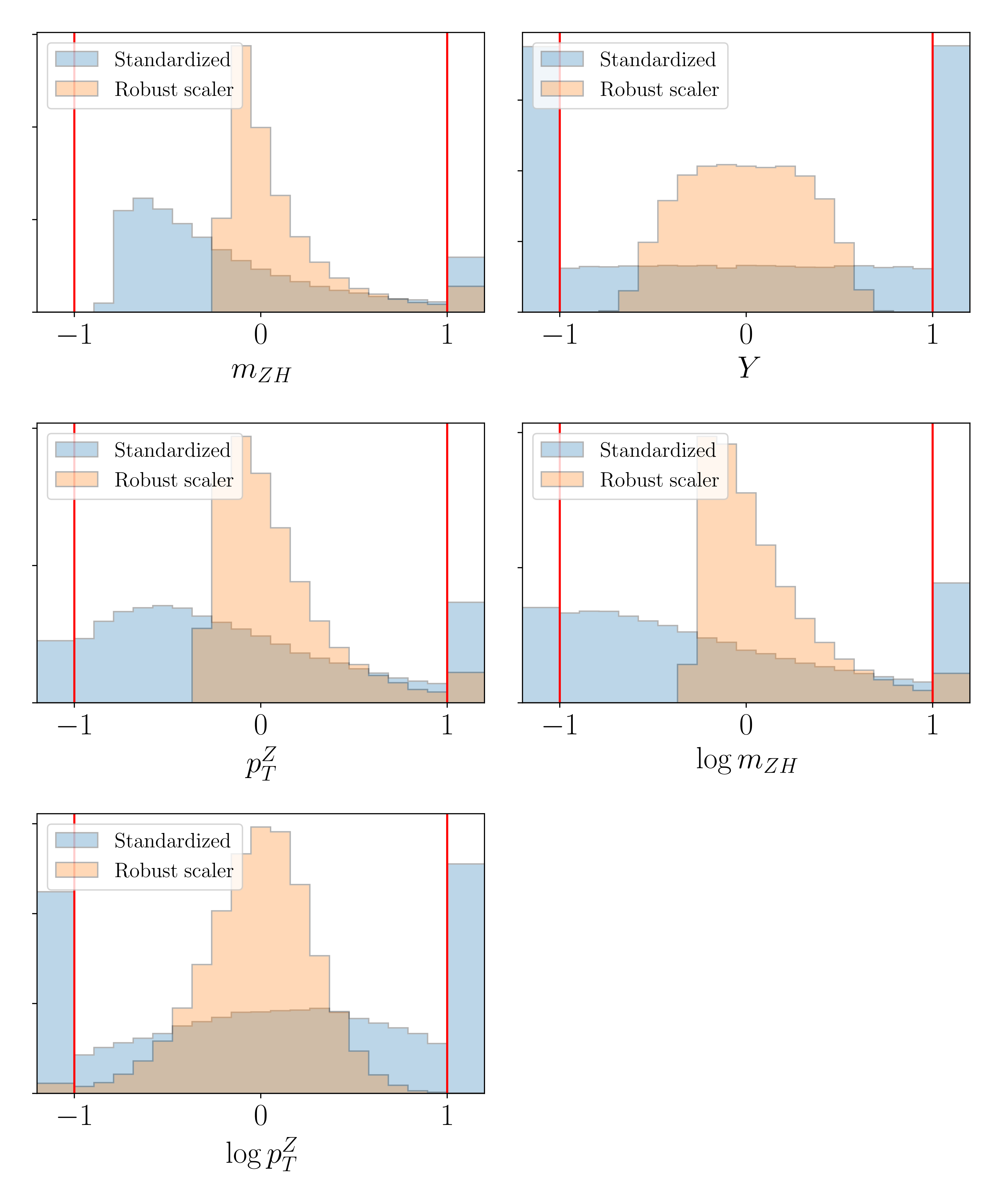
- Traditional SMEFT analyses can and should be **complemented** with **unbinned** measurements to optimise the constraining power on the Wilson coefficients

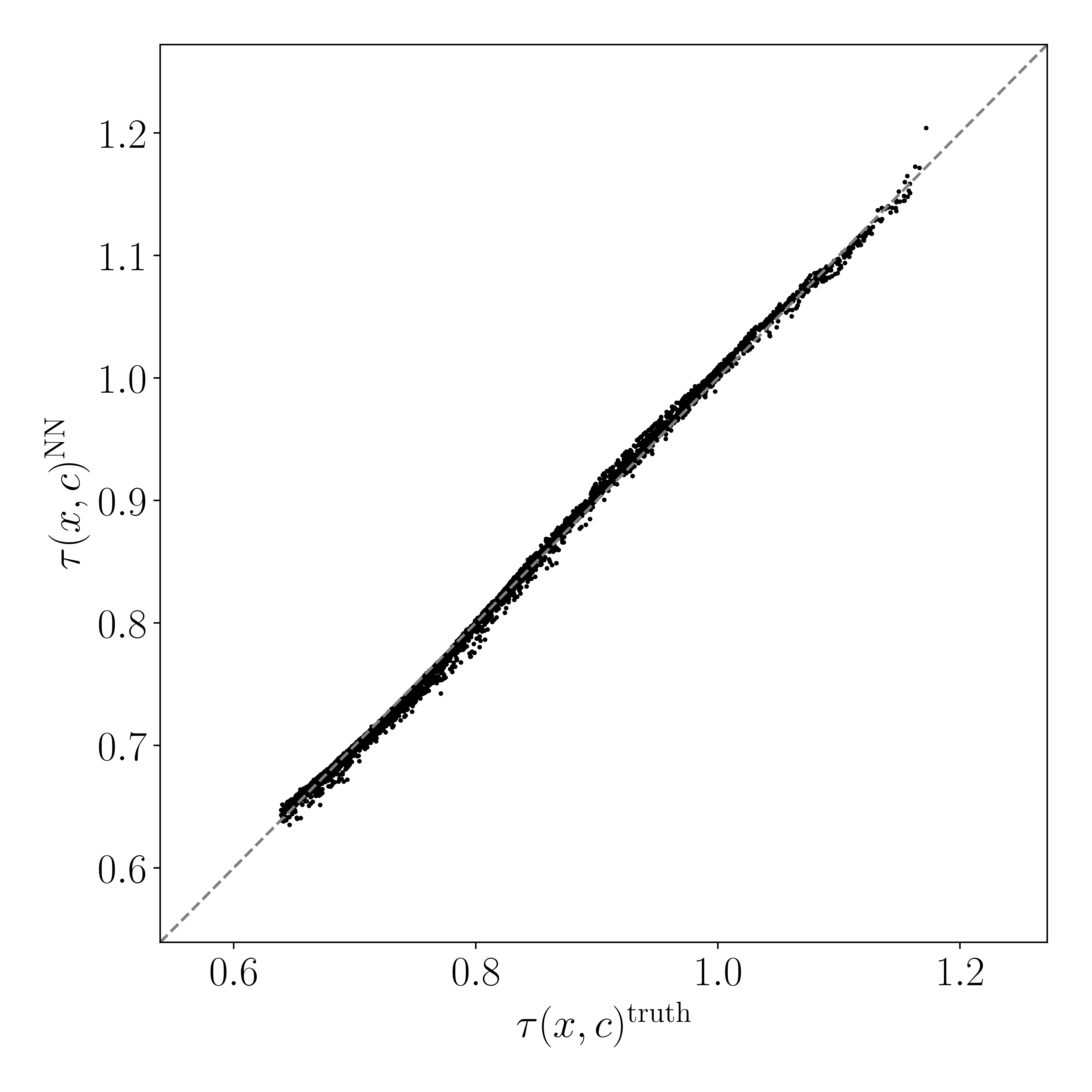
- Likelihood ratio parameterisations with NN are a **promising** way forward, and should be implemented in true global EFT fits

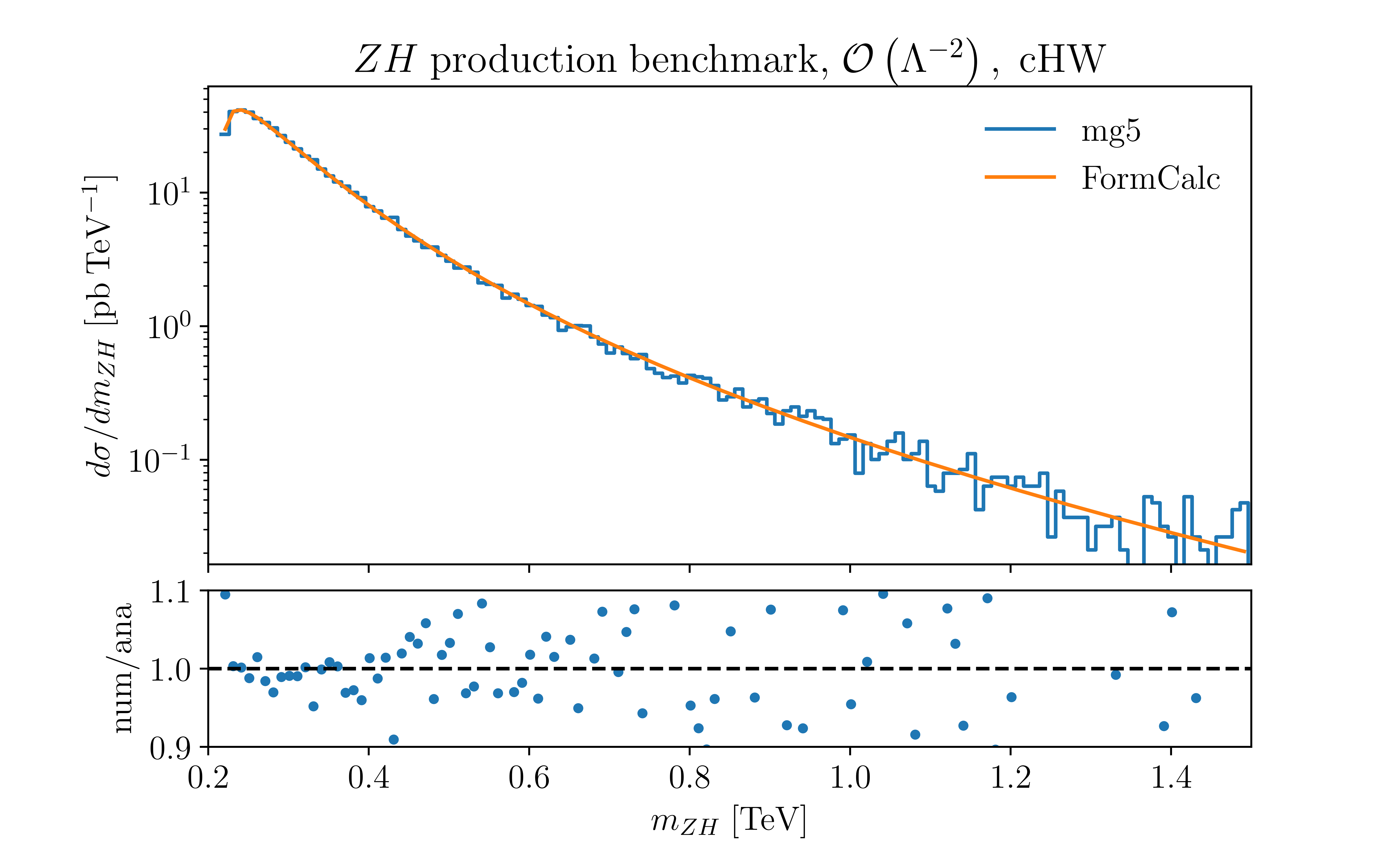


Thank you!
Questions?

Backup


Standardized
Robust scaler
−1
0
1
m_{ZH}
Y
p_T^Z
$\log m_{ZH}$
$\log p_T^Z$


1.2
1.1
1.0
0.9
0.8
0.7
0.6
$\tau(x,c)^{\mathrm{NN}}$
0.6
0.8
1.0
1.2
$\tau(x,c)^{\mathrm{truth}}$



ZH production benchmark, $\mathcal{O}\left(\Lambda^{-2}\right)$, cHW
mg5
FormCalc
$d\sigma/dm_{ZH}$ [pb TeV^{-1}]
10^{1}
10^{0}
10^{-1}
num/ana
1.1
1.0
0.9
0.2
0.4
0.6
0.8
1.0
1.2
1.4
m_{ZH} [TeV]