Unbinned measurements in global EFT fits from machine learning

Juan Rojo
VU Amsterdam & Theory group, Nikhef

CMS EFT Working Group
Zoom, 30/06/2022

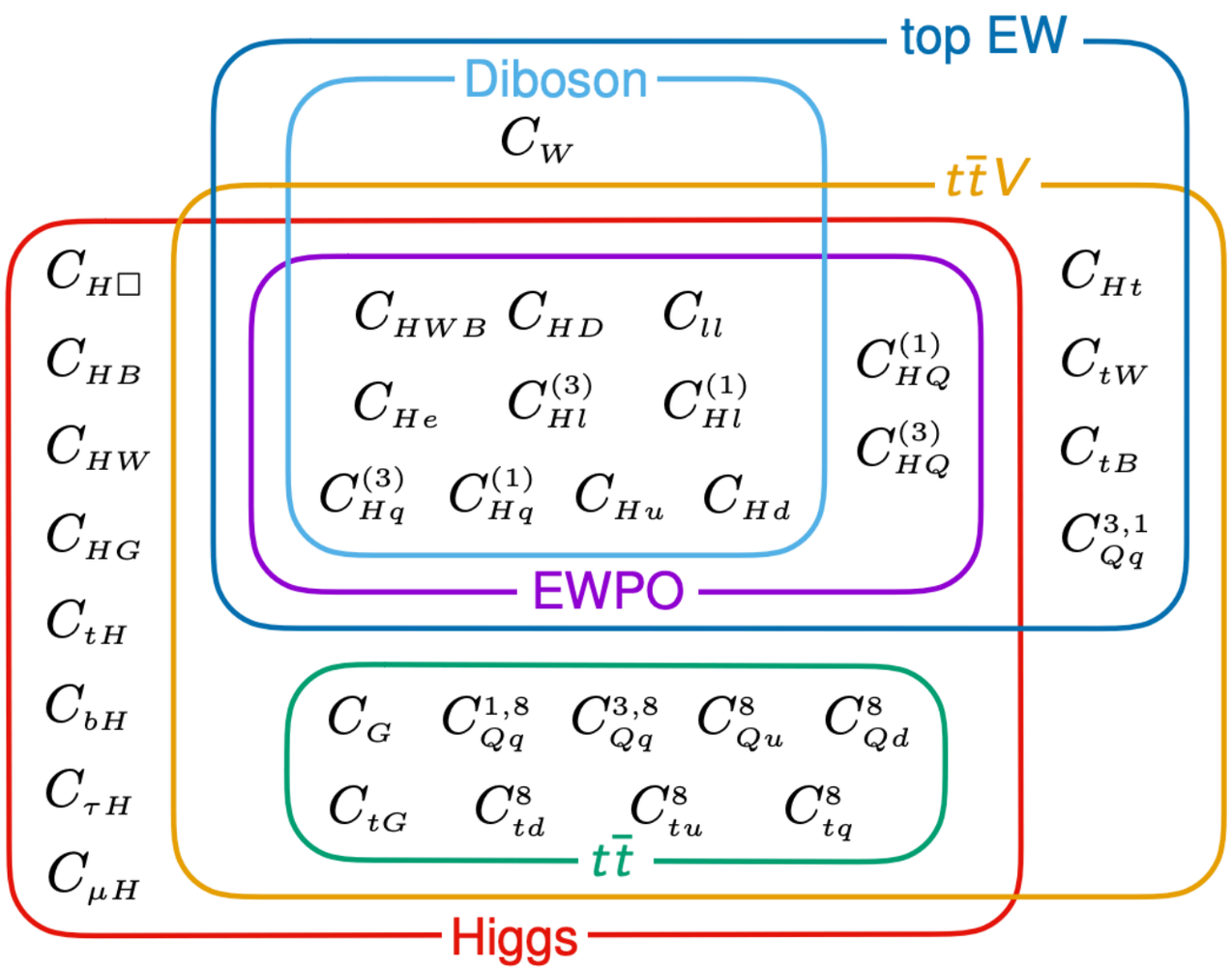
In collaboration with Raquel Gomez-Ambrosio, Jaco ter Hoeve, Maeve Madigan, and Veronica Sanz

The Standard Model EFT

Extend SM Lagrangian with **complete, non-redundant basis** of higher dimensional operators

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i^{N_{d5}} \frac{a_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_i^{N_{d7}} \frac{d_i}{\Lambda^3} \mathcal{O}_i^{(7)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^2} \mathcal{O}_j^{(8)} + \ldots$$

N_{d6} = 59 (2499) for one (three) flavour generations



FitMaker, 2012.02779

Low-energy limit of generic UV-complete theories; systematic parametrisation of BSM effects; fully renormalizable QFT; matched to a large number of BSM models that reduce to the SM; increased sensitivity at LHC from xsec growth in CoM energy ….

Fulfilling the potential of the SMEFT framework demands global analyses based on **a wide range of process and data** to cover all relevant **directions in the EFT parameter space**

The Standard Model EFT

from Lagrangian …

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_{m=1}^{N_6} \frac{c_m}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_{n=1}^{N_8} \frac{b_j}{\Lambda^4} \mathcal{O}_i^{(8)} + \ldots$$

SM **EFT$_{d6}$** **EFT$_{d8}$**

Linear EFT cross-sections: interference SM-EFT$_{d6}$

Quadratic EFT cross-sections: squares EFT$_{d6}$

to cross-sections ….

$$\sigma_{\text{SMEFT}}\left(\boldsymbol{c}, \boldsymbol{\Lambda}\right) \simeq \sigma_{\text{SM}} \times \left(1 + \sum_{m=1}^{N_6} \frac{c_m}{\Lambda^2} \sigma_m^{(\text{eft})} + \sum_{m,n=1}^{N_6} \frac{c_m c_n}{\Lambda^4} \sigma_{m,n}^{(\text{eft})}\right)$$

evaluate at (N)NLO QCD + NLO EW

evaluate at NLO QCD with **SMEFT@NLO**

The Standard Model EFT

… to constraints on the EFT parameters

$$\chi^2(\boldsymbol{c}, \Lambda) = \frac{1}{n_{\rm dat}} \sum_{i,j=1}^{n_{\rm dat}} \left(\sigma_{\rm i,SMEFT}(\boldsymbol{c}, \Lambda) - \sigma_{\rm i,exp}\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(\sigma_{\rm j,SMEFT}(\boldsymbol{c}, \Lambda) - \sigma_{\rm j,exp}\right)$$

log-likelihood minimisation

(multi-gaussian likelihood)

All existing global EFT analyses are based on SM measurements ``re-interpreted'' in EFT framework

Linear EFT cross-sections:

interference SM-EFT$_{d6}$

Quadratic EFT cross-sections:

squares EFT$_{d6}$

to cross-sections ….

$$\sigma_{\rm SMEFT}\left(\boldsymbol{c}, \boldsymbol{\Lambda}\right) \simeq \sigma_{\rm SM} \times \left(1 + \sum_{m=1}^{N_6} \frac{c_m}{\Lambda^2} \sigma_m^{\rm (eft)} + \sum_{m,n=1}^{N_6} \frac{c_m c_n}{\Lambda^4} \sigma_{m,n}^{\rm (eft)}\right)$$

evaluate at (N)NLO QCD + NLO EW

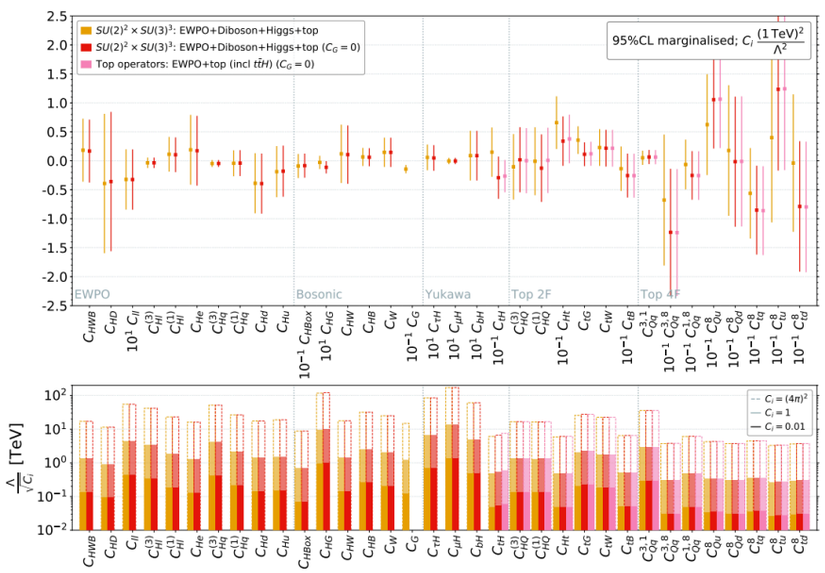
evaluate at NLO QCD

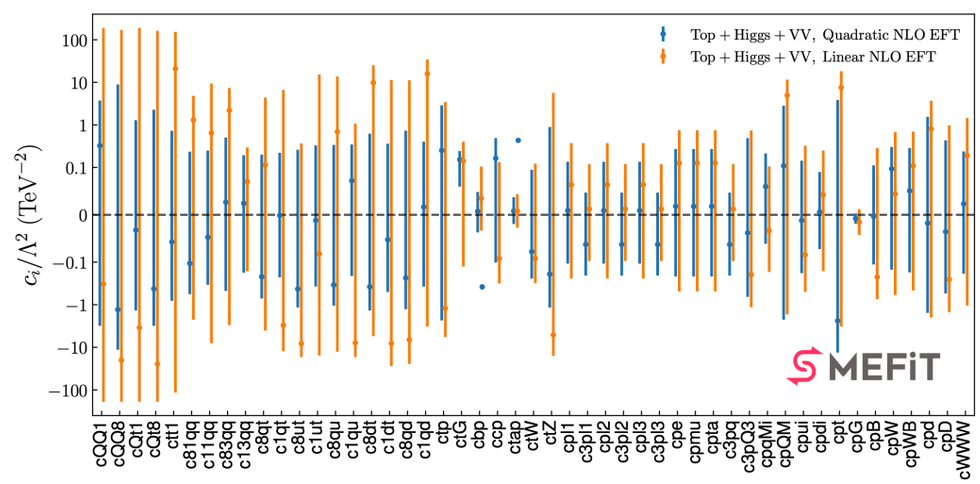
with **SMEFT@NLO**

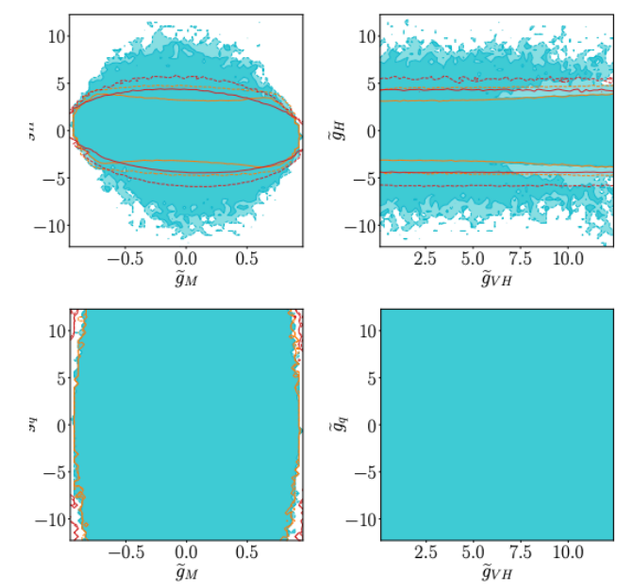
Global EFT analyses

- Various groups have presented EFT analysis combining **data from different processes**, hence sensitive to a **large number of EFT directions**, and their interpretation in terms of UV models

e.g. SMEFiT combines 317 cross-section measurements to constrain 50 dim-6 EFT coefficients







Fitmaker

SMEFiT

SFitter

arXiv: 2012.02779

arXiv: 2105.00006

arXiv: 2108.01094

- None of these measurements used has been **optimised for EFT studies**, can we do better?

From binned to unbinned likelihoods

- Most measurements used for EFT fits are presented in terms of **multi-Gaussian likelihoods**

EFT parameters — *# of bins* — *# of events per bin (exp)* — *# of events per bin (pred)*

$$\mathcal{L}(\boldsymbol{n};\boldsymbol{\nu}(\boldsymbol{c})) = \prod_{i=1}^{N_b} \exp\left[-\frac{1}{2}\frac{(n_i - \nu_i(\boldsymbol{c}))^2}{\nu_i(\boldsymbol{c})}\right]$$

(only stat uncertainties here)

- What is the **optimal number of bins** for a subsequent EFT analysis? How much information do we gain/lose by measuring a cross-section **differential in additional kinematic variables**?

- This multi-gaussian likelihood can be extended to a **Poissonian likelihood** to account for distributions where some bins have a small # of events (*e.g.* **high energy tails**)

$$\mathcal{L}(\boldsymbol{n};\boldsymbol{\nu}(\boldsymbol{c})) = \prod_{i=1}^{N_b} \frac{\nu_i^{n_i}(\boldsymbol{c})}{n_i!} e^{-\nu_i(\boldsymbol{c})}$$

- Likelihood maximisation: derive CL internals for the EFT coefficients, given the observed data

are we losing information due to specific choice of binning and final-state kinematic distributions?

From binned to unbinned likelihoods

main goal: construct unbinned observables from ML and assess their relevance for **global EFT fits,** by comparing their impact with those of ``traditional" binned observables

$$\mathcal{L}(\boldsymbol{c}) = \frac{\nu_{\rm tot}^{N_{\rm ev}}(\boldsymbol{c})}{N_{\rm ev}!} e^{-\nu_{\rm tot}(\boldsymbol{c})} \prod_{i=1}^{N_{\rm ev}} f_\sigma\left(\boldsymbol{x}_i, \boldsymbol{c}\right)$$

unbinned likelihood

EFT parameters

number of events (exp)

number of events (pred)

probability distribution in the final-state kinematics

$$f_\sigma\left(\boldsymbol{x}, \boldsymbol{c}\right) = \frac{1}{\sigma(X, \boldsymbol{c})} \frac{d\sigma(\boldsymbol{x}, \boldsymbol{c})}{d\boldsymbol{x}}$$

fully differential cross-section

e.g. in $pp \to hW \to b\bar{b}\ell\nu_\ell$

$$\boldsymbol{x} = \left(p_T^\ell, y_\ell, m_{b\bar{b}}, y_b, p_T^b, \ldots\right)$$

by construction, this unbinned likelihood **contains all the information** from the observed events

Challenge: evaluation of unbinned likelihood very costly, specially for **particle & detector level final states**

Solution: parametrise the diff xsecs in terms of **deep neural networks** and train them on Monte Carlo data

From binned to unbinned likelihoods

For an analysis sensitive to n_p EFT coefficients, we can express the **SMEFT cross-section** as

$$f_\sigma(\boldsymbol{x},\boldsymbol{c}) = f_\sigma(\boldsymbol{x},\boldsymbol{0}) + \sum_{j=1}^{n_p} f_\sigma^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_p}\sum_{k\geq j}^{n_p} f_\sigma^{(j,k)}(\boldsymbol{x})c_j c_k$$

SMEFT cross-section, differential in all variables included in x

SM cross-section

linear EFT xsec

quadratic EFT xsec

Technically it is advantageous to parametrise the **likelihood ratio to the SM**

$$r_\sigma(\boldsymbol{x},\boldsymbol{c}) = \frac{f_\sigma(\boldsymbol{x},\boldsymbol{c})}{f_\sigma(\boldsymbol{x},\boldsymbol{0})} = 1 + \sum_{j=1}^{n_p} r_\sigma^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_p}\sum_{k\geq j}^{n_p} r_\sigma^{(j,k)}(\boldsymbol{x})c_j c_k$$

Once unbinned observables have been constructed, one can **extend the global EFT fit likelihood** to account also for the information they provide

$$\mathcal{L}(\boldsymbol{c}) = \prod_{k=1}^{N_\mathcal{D}} \mathcal{L}_k(\boldsymbol{c}) = \prod_{k=1}^{N_\mathcal{D}^{(\mathrm{ub})}} \mathcal{L}_k^{(\mathrm{ub})}(\boldsymbol{c}) \prod_{j=1}^{N_\mathcal{D}^{(\mathrm{bp})}} \mathcal{L}_j^{(\mathrm{bp})}(\boldsymbol{c}) \prod_{\ell=1}^{N_\mathcal{D}^{(\mathrm{bg})}} \mathcal{L}_\ell^{(\mathrm{bg})}(\boldsymbol{c})$$

Full likelihood global EFT fit ***unbinned obs*** ***binned, Poisson obs*** ***binned, Gaussian obs***

Cross-section ML parametrisation

We need hence to **parametrise** the dependence of the distribution ratios on the **kinematics**

$$r_\sigma(\boldsymbol{x},\boldsymbol{c}) = \frac{f_\sigma(\boldsymbol{x},\boldsymbol{c})}{f_\sigma(\boldsymbol{x},\boldsymbol{0})} = 1 + \sum_{j=1}^{n_p} r_\sigma^{(j)}(\boldsymbol{x})c_j + \sum_{j=1}^{n_p}\sum_{k\geq j}^{n_p} r_\sigma^{(j,k)}(\boldsymbol{x})c_j c_k$$

e.g. for Higgs associate production, we may need to **parametrise a 5D function**

$$f_\sigma\left(\boldsymbol{x},\boldsymbol{c}\right) = \frac{1}{\sigma(X,\boldsymbol{c})}\frac{d\sigma(\boldsymbol{x},\boldsymbol{c})}{d\boldsymbol{x}}$$

$$pp \to hW \to b\bar{b}\ell\nu_\ell$$

$$\boldsymbol{x} = \left(p_T^\ell, y_\ell, m_{b\bar{b}}, y_b, p_T^b, \right)$$

$$r_\sigma^{(j)}(\boldsymbol{x}) = r_\sigma^{(j)}\left(p_T^\ell, y_\ell, m_{b\bar{b}}, y_b, p_T^b\right)$$

Adopt deep neural networks as **universal unbiased interpolants** and bypass the need to evaluate the differential cross-section on a event-by-event basis

The structure of EFT corrections (linear + quadratic) make possible **parallelizing** the training

related work by Chen at al 2007.10356, Tito d'Agnolo et al 1912.12155, Brehmer et al 1805.00013 + many others …

Cross-section ML parametrisation

Exploit EFT structure: **sequential parametrisation** of EFT cross-sections with neural nets

generate **MC data** with single EFT coefficient activated + SM baseline

$$\mathcal{D}_{\text{eft}}(\boldsymbol{c} = (0, \ldots, 0, c_j^{(\text{tr})}, 0, \ldots, 0))$$

parametrise **linear term** in the cross-section ratio

$$r_\sigma(\boldsymbol{x}, c_j^{(\text{tr})}) = 1 + c_j^{(\text{tr})} \text{NN}^{(j)}(\boldsymbol{x})$$

determine the **decision boundary *g(x,c)*** used to classify events into SM or EFT by minimizing cross-entropy

$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\int d\boldsymbol{x} \frac{d\sigma(\boldsymbol{x}, \boldsymbol{c})}{d\boldsymbol{x}} \log(1 - g(\boldsymbol{x}, \boldsymbol{c})) - \int d\boldsymbol{x} \frac{d\sigma(\boldsymbol{x}, \boldsymbol{0})}{d\boldsymbol{x}} \log g(\boldsymbol{x}, \boldsymbol{c})$$

g -> 0 for EFT ***g -> 1 for SM***

$$g(\boldsymbol{x}, \boldsymbol{c}) = \left(1 + \frac{d\sigma(\boldsymbol{x}, \boldsymbol{c})}{d\boldsymbol{x}} \Big/ \frac{d\sigma(\boldsymbol{x}, \boldsymbol{0})}{d\boldsymbol{x}}\right)^{-1} = \frac{1}{1 + r_\sigma(\boldsymbol{x}, \boldsymbol{c})}$$

outcome of classification problem: **NN parametrisation of the linear EFT cross-section**

Cross-section ML parametrisation

Exploit EFT structure: **sequential parametrisation** of EFT cross-sections with neural nets

Repeat for each individual EFT coefficient, obtain all the **linear EFT cross-sections**

Similar approach can be used to parametrise the **quadratic EFT terms**

$$r_\sigma(\boldsymbol{x}, c_j^{(\mathrm{tr})}) = \left\{1 + c_j^{(\mathrm{tr})} \mathrm{NN}^{(j)}(\boldsymbol{x})\right\} + \left(c_j^{(\mathrm{tr})}\right)^2 \mathrm{NN}^{(j,j)}(\boldsymbol{x})$$

already determined

from cross-entropy minimisation

finally, parametrise the crossed quadratic terms

$$r_\sigma(\boldsymbol{x}, \boldsymbol{c}) = \Big\{1 + c_{j,\mathrm{train}} \cdot \mathrm{NN}^{(j)}(\boldsymbol{x}) + c_{k,\mathrm{train}} \cdot \mathrm{NN}^{(k)}(\boldsymbol{x}) + c_{j,\mathrm{train}}^2 \cdot \mathrm{NN}^{(j,j)}(\boldsymbol{x}) + c_{k,\mathrm{train}}^2 \cdot \mathrm{NN}^{(k,k)}(\boldsymbol{x})\Big\} + c_{j,\mathrm{train}} c_{k,\mathrm{train}} \cdot \mathrm{NN}^{(j,k)}(\boldsymbol{x})$$

already determined

from cross-entropy minimisation

 # processors required scales as ***O(n$_p^2$)***, making our approach amenable to **parallelisation** and suitable to explore **large EFT parameter spaces**

Neural network training

$pp \to hW$

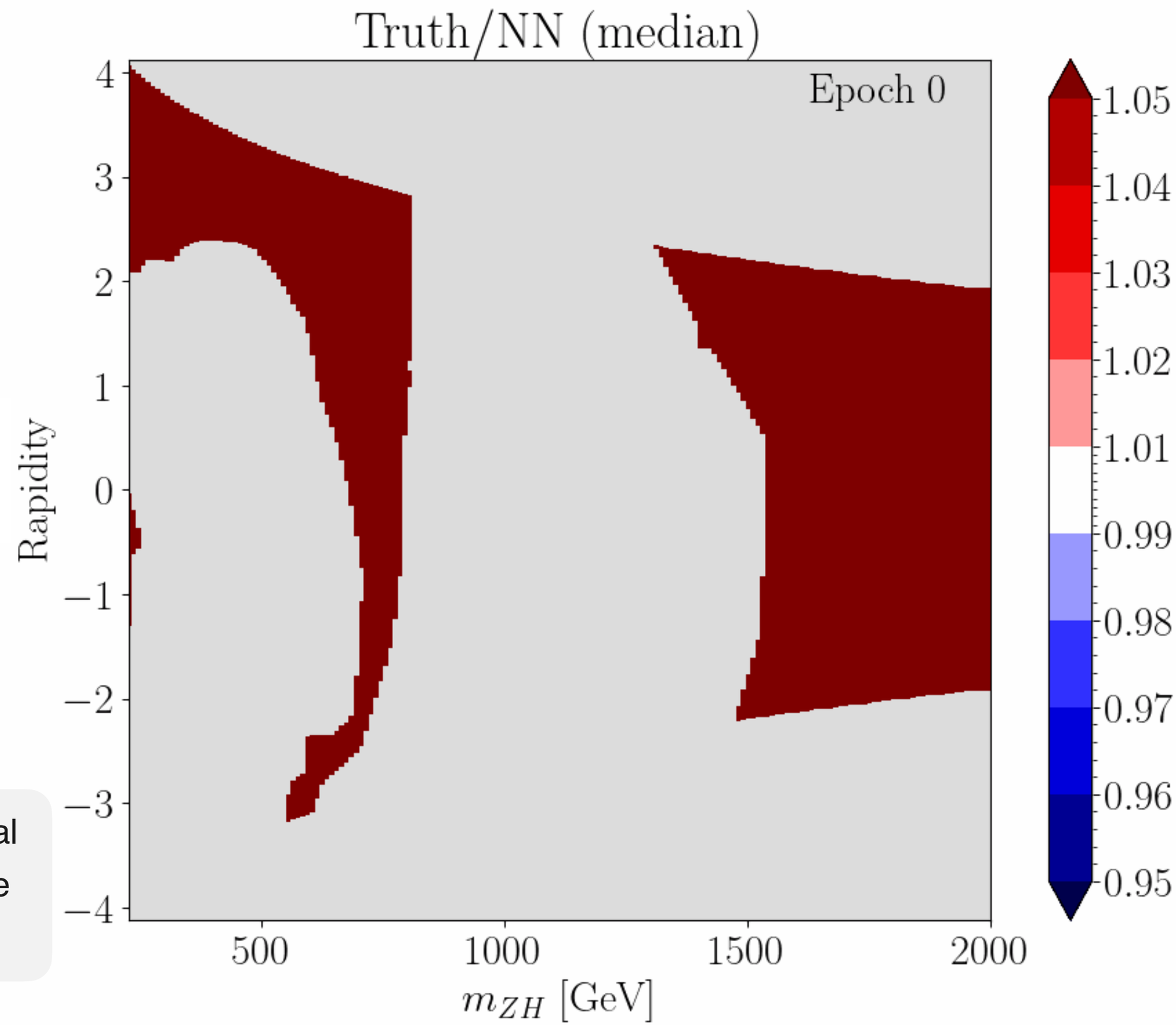
linear EFT parametrisation

$$\frac{d^2\sigma\left(m_{Zh}, y_Z; \boldsymbol{c}\right)}{dm_{Zh}dy_Z}$$

$$\mathcal{O}_{Hq}^{(3)} = (H^\dagger i \overleftrightarrow{D}^i_\mu H)(Q\sigma^i\gamma^\mu Q)$$
$$\mathcal{O}_{HW} = H^\dagger H W^{i\mu\nu} W_{i\mu\nu}$$

validate NN outcome with
exact analytical result

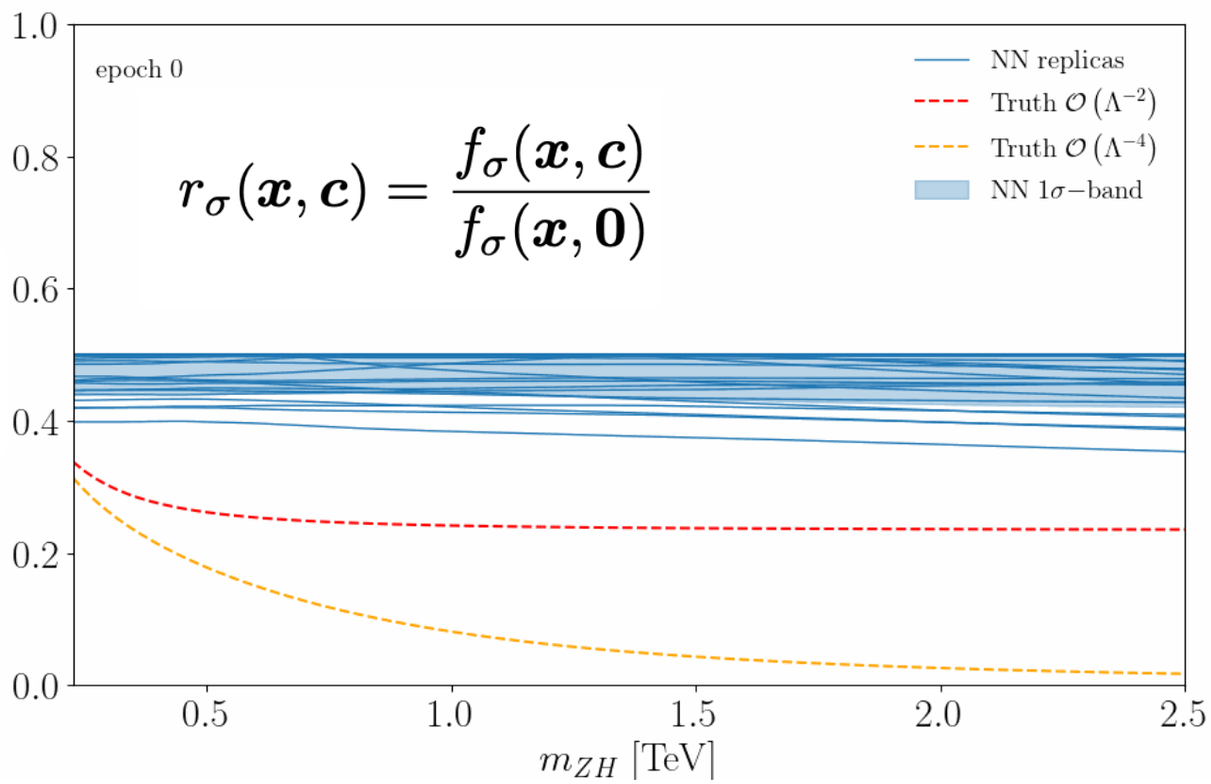
good description of differential EFT cross-section, challenge is the low-stat tails



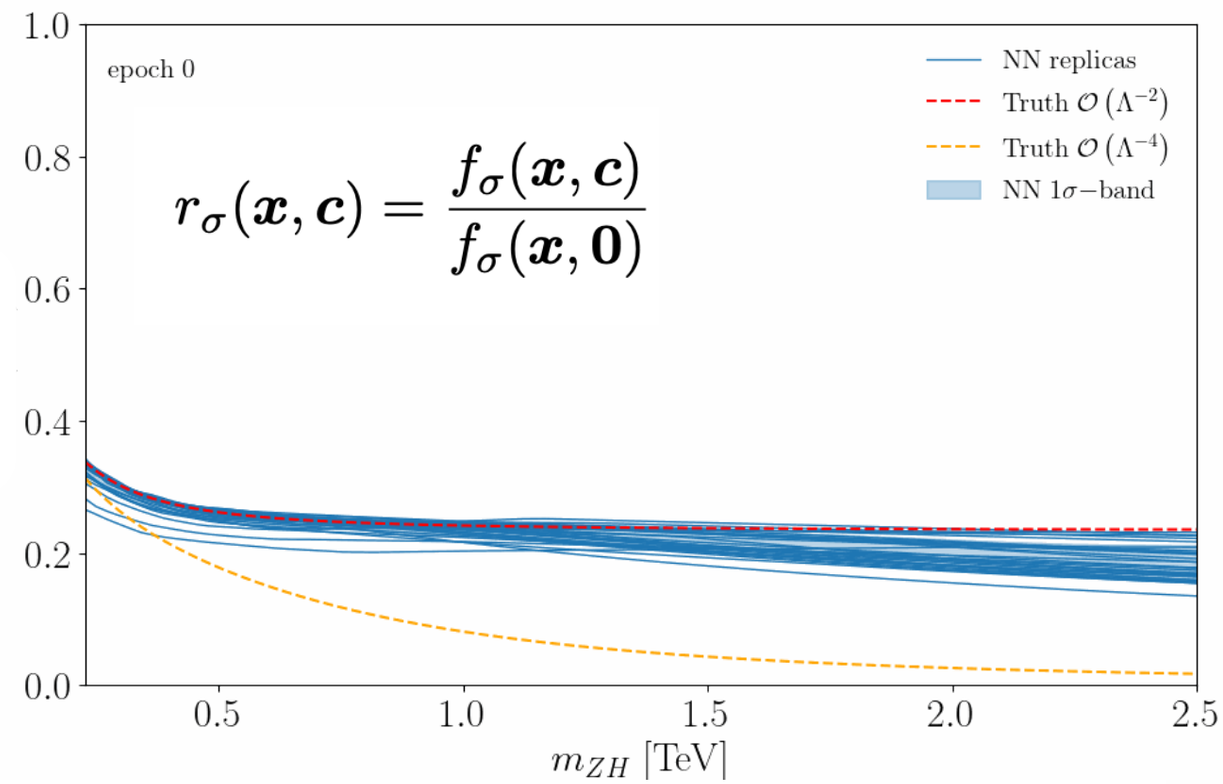
Neural network training

Model uncertainties estimated by means of the **Monte Carlo replica method:** train a large number of different replicas to distinct, statistically equivalent MC datasets

step 1: learn linear EFT cross-sections, repeat for many MC replicas



step 2: learn quadratic EFT cross-sections (having learnt already linear terms)



fully keeps into account correlations

The use of the MC replica method makes possible **propagating model and methodological uncertainties** to the final results: bounds on the EFT parameter space

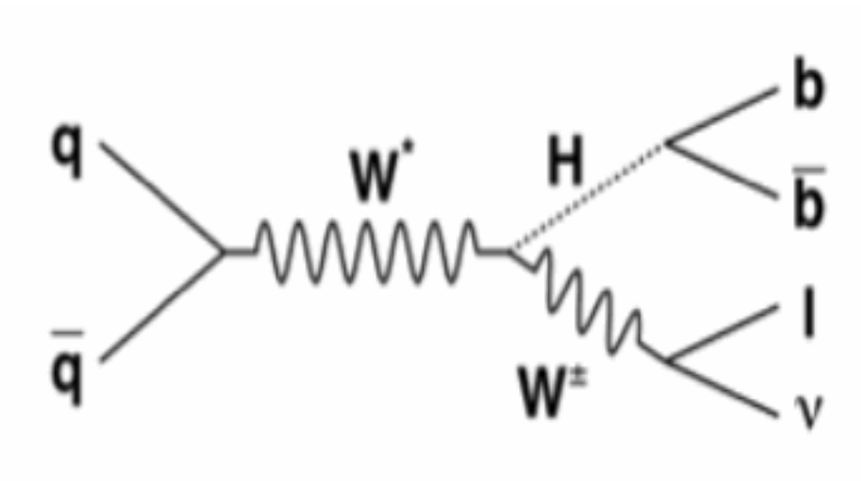
same approach as in the NNPDF framework!

Results

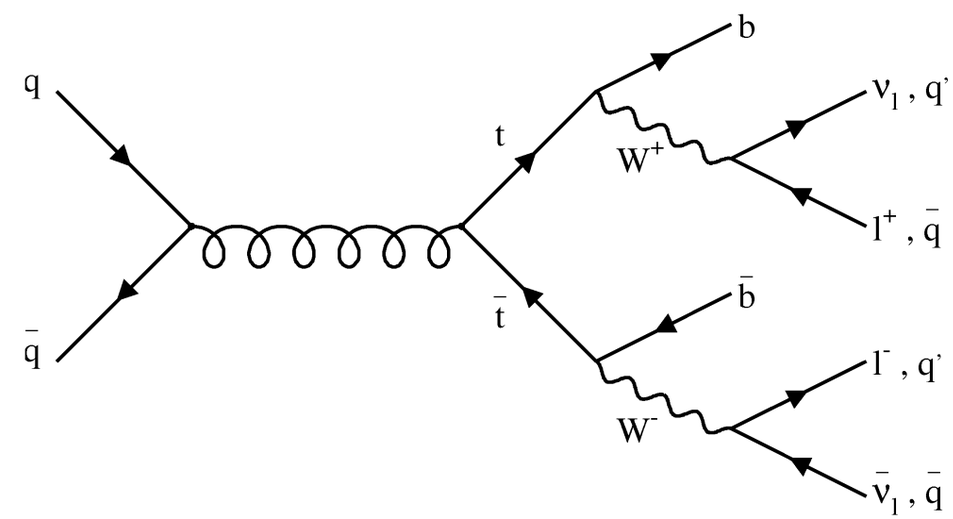
- As a proof of concept consider particle-level measurements of **Higgs associated production** and for **quark pair production** (already included in **SMEFiT global analysis**)

$$pp \to hW \to b\bar{b}\ell\nu_\ell$$

$$pp \to hZ \to b\bar{b}\ell^+\ell^-$$



$$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$$



- Same **theory settings**, operator basis, flavour assumptions as in SMEFiT to allow **integration**
- Using this framework, explore a number of crucial questions for global SMEFT fits

How much sensitivity to EFT coefficients we gain in **unbinned measurements**? How **differential** a measurement needs to be? Is **optimal binning** different in EFT interpretations?

Under which conditions an unbinned measurement is **justified**? Does the answer depend on whether or not one has a **global EFT fit**? What should we do at the **HL-LHC?**

Results

process:

$$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$$

SMEFT operators

$$\mathcal{O}_{tu}^{(1)} = (\bar{t}\gamma_\mu t)(\bar{u}\gamma^\mu u)$$

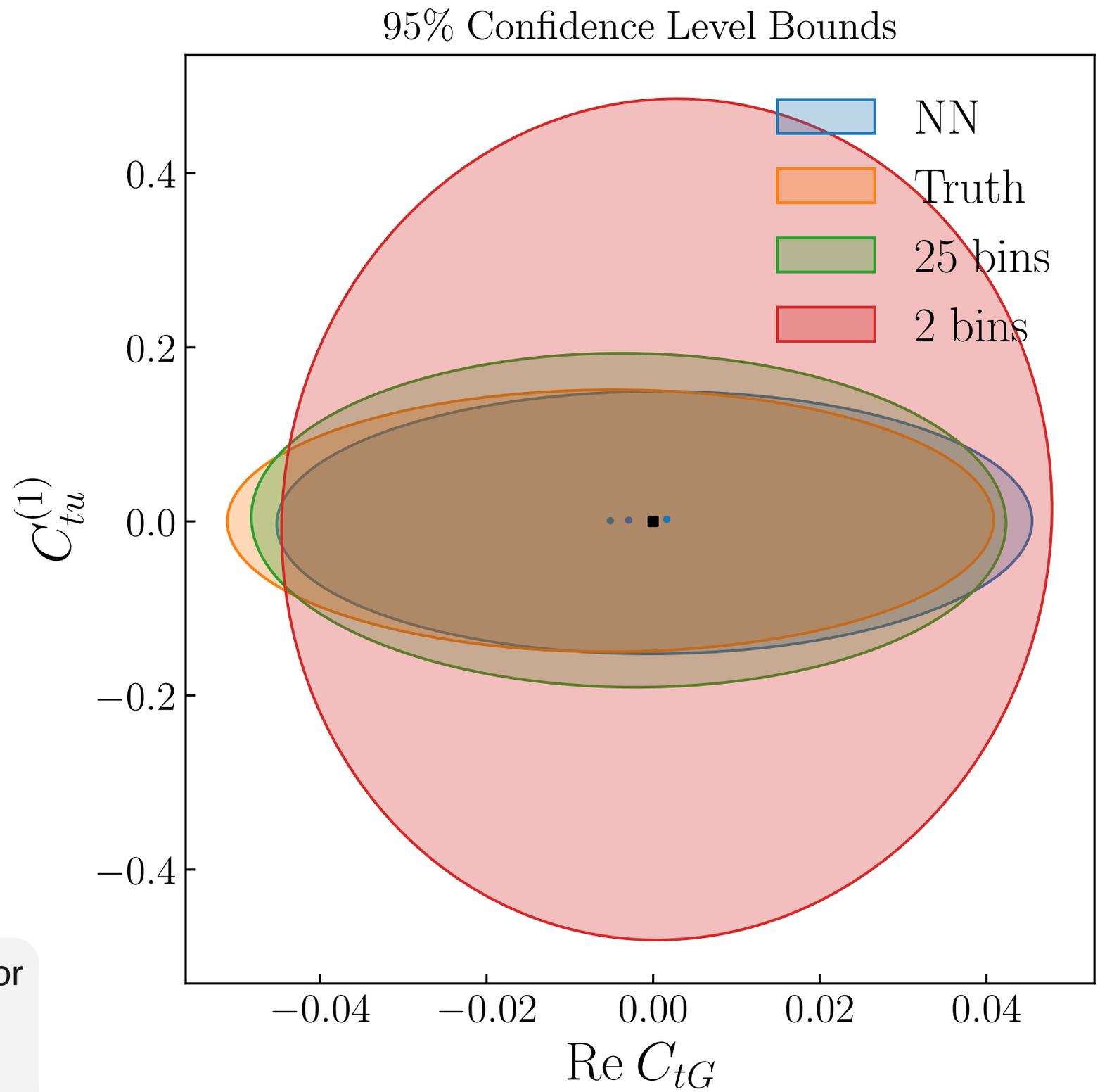
$$\mathcal{O}_{tG} = (\bar{Q}\sigma^{\mu\nu}T^a t)\tilde{H}G^a_{\mu\nu}.$$

Unbinned case: compare **analytical calculation** of the likelihood (`truth') with **ML interpolation** (`NN')

Binned case: compare coarse with fine binning

the ML-based observable reproduces the analytical result within replica uncertainties

since we consider only stat errors, for infinitely fine binning one should reproduce the unbinned result



Results

process:

$$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$$

SMEFT operators

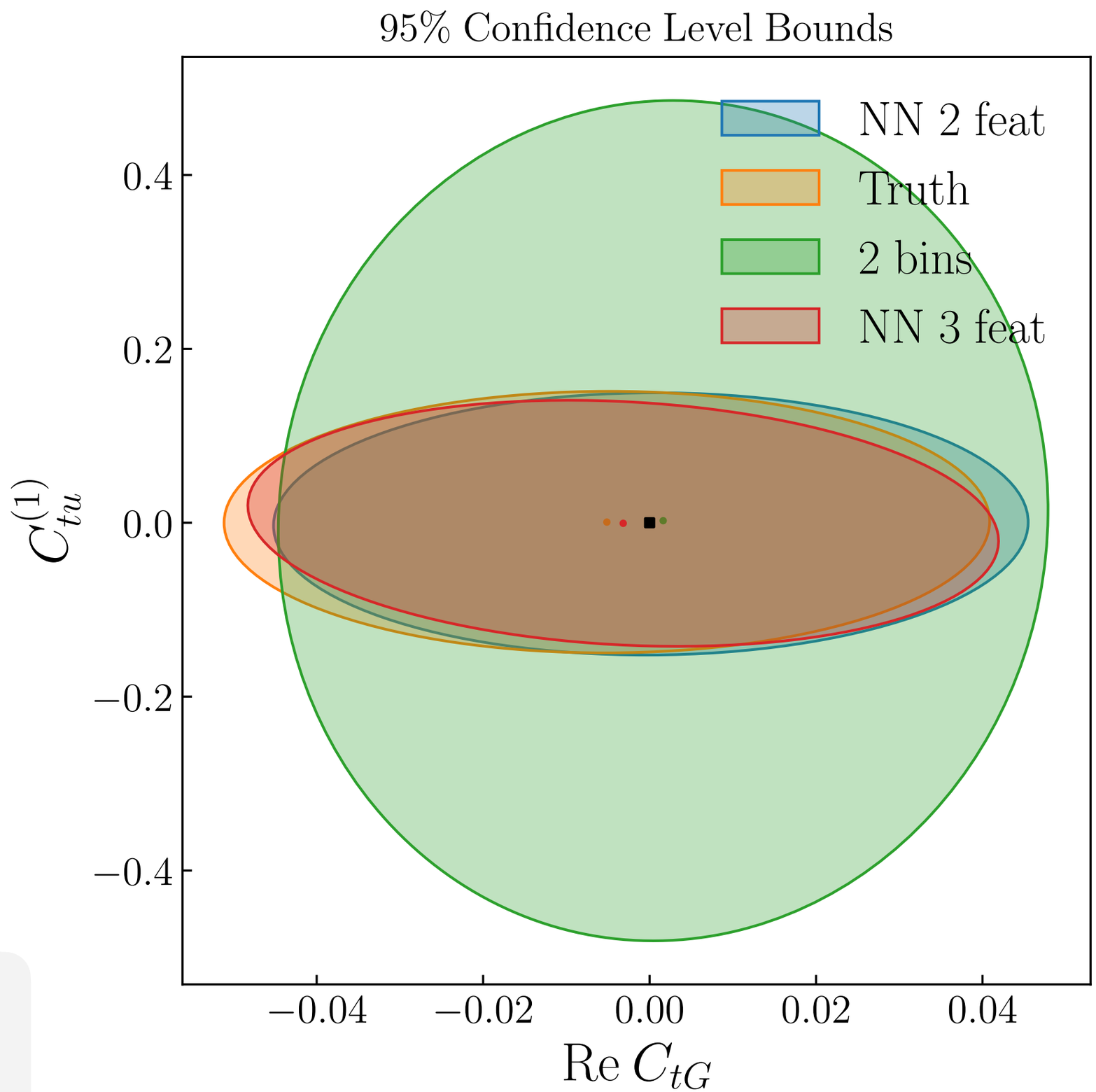
$$\mathcal{O}_{tu}^{(1)} = (\bar{t}\gamma_\mu t)(\bar{u}\gamma^\mu u)$$

$$\mathcal{O}_{tG} = (\bar{Q}\sigma^{\mu\nu}T^a t)\tilde{H}G^a_{\mu\nu}.$$

Unbinned case: compare **analytical calculation** of the likelihood (`truth') with **ML interpolation** (`NN'), the latter case with 2 or three features

the ML-based observable reproduces the analytical result within replica uncertainties

In this case, adding a third feature (extra kinematic variable) to the observable does not affect the bounds



Results

process:

$$pp \to hZ \to b\bar{b}\ell^+\ell^-$$

SMEFT operators

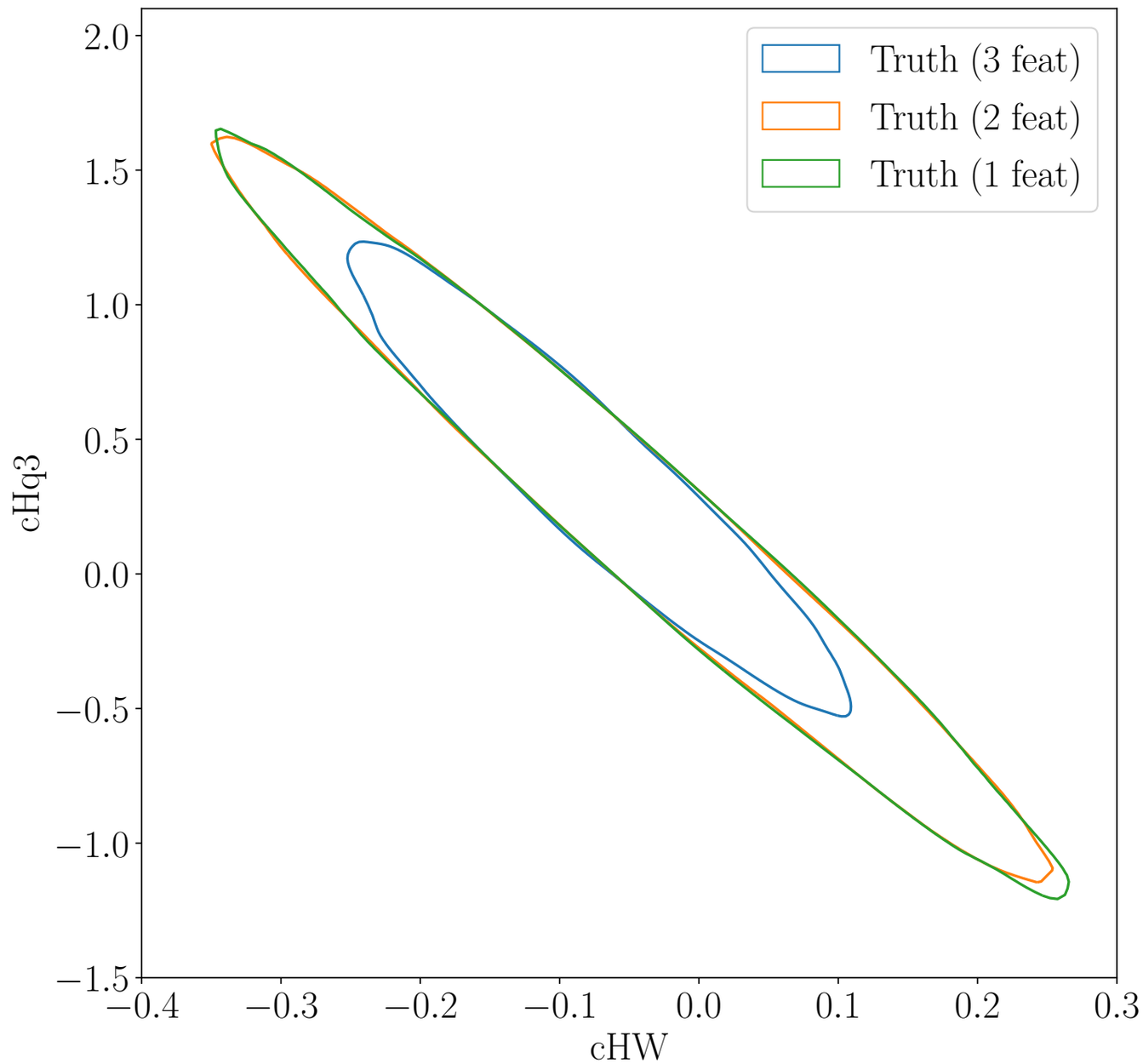
$$\mathcal{O}_{Hq}^{(3)} = (H^\dagger i \overleftrightarrow{D}^i_\mu H)(Q\sigma^i\gamma^\mu Q)$$

$$\mathcal{O}_{HW} = H^\dagger H W^{i\mu\nu} W_{i\mu\nu}$$

Unbinned case based on **analytical calculation** of the likelihood (`truth')

We can identify that adding a third kinematic feature brings in **extra sensitivity**

$$\boldsymbol{x} = \left(p_T^Z, y_Z, m_{hZ}\right)$$



Within our approach one can systematically identify when the sensitivity in the EFT coefficients on additional bins or kinematic variables ``**saturates**'', key input for experimental analyses

Results

process:

$$pp \to hZ \to b\bar{b}\ell^+\ell^-$$

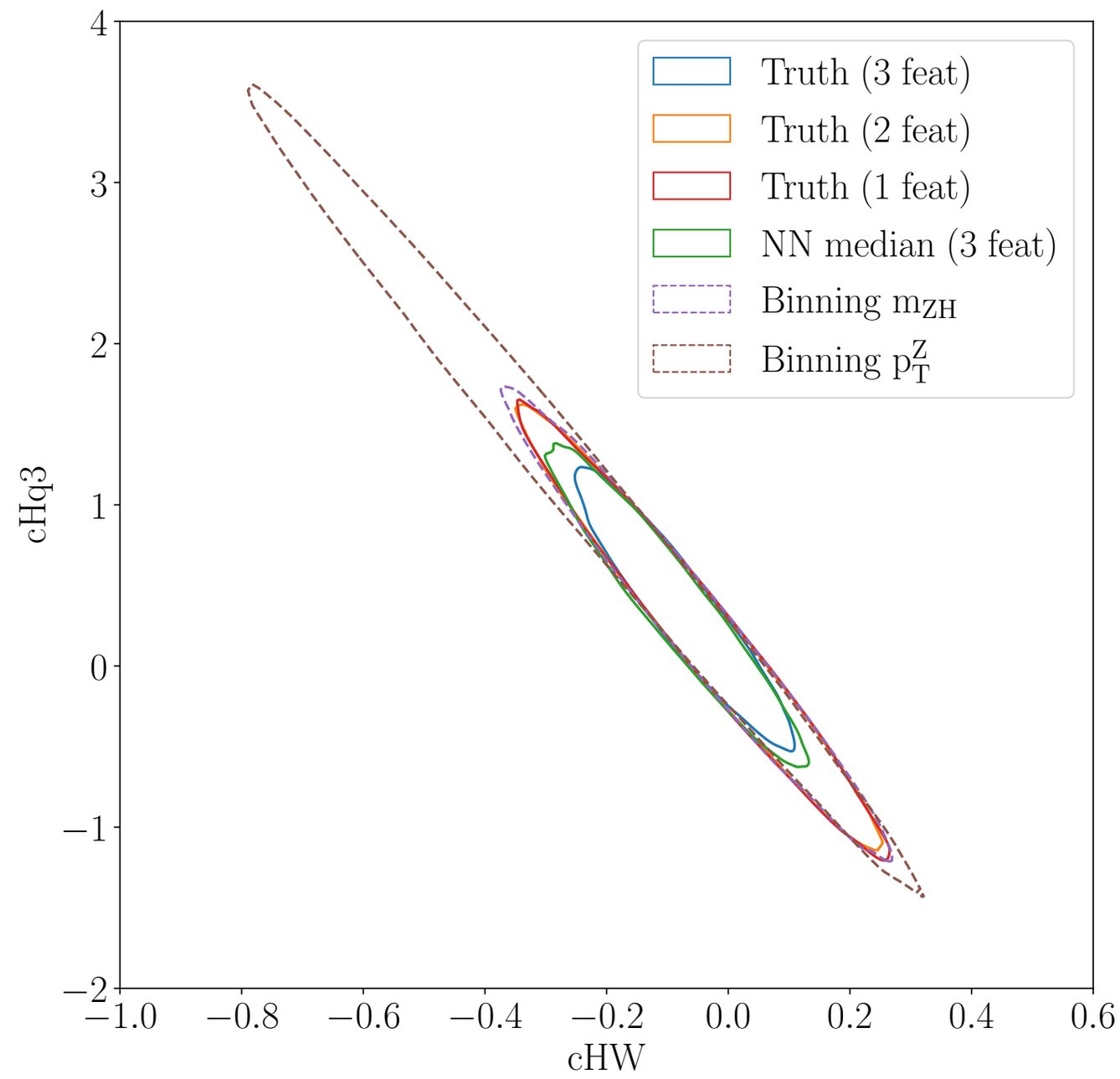
SMEFT operators

$$\mathcal{O}_{Hq}^{(3)} = (H^\dagger i \overleftrightarrow{D}^i_\mu H)(Q\sigma^i\gamma^\mu Q)$$

$$\mathcal{O}_{HW} = H^\dagger H W^{i\mu\nu} W_{i\mu\nu}$$

Unbinned case: compare **analytical calculation** of the likelihood (`truth') with **ML interpolation** (`NN')

Binned case: compare binning in two different kinematic variables



unbinned observables also provide a **robust benchmark** to optimise binned analyses

exact analytical calculation only available for simple processes, in general rely on MC generators

Summary and outlook

- The EFT framework provides a robust strategy to interpret particle physics data in a (mostly) **model-independent manner** of new BSM phenomena

- Only within a **global SMEFT interpretation** it is possible to compare with largest possible class of UV-complete theories and to reduce assumptions i.e. concerning flavour structure

- We are extending the SMEFiT framework to include **unbinned observables** to further constrain the EFT parameter space, based on a **ML likelihood (ratio) parametrisation**

- Our approach scales to an **arbitrary large number of EFT coefficients** and multi-dimensional observables and can be used to determine the **optimal EFT sensitivity** of measurements

- Ongoing work to carry out **global EFT fits based on both binned and unbinned observables**, to include systematic uncertainties, and to further improve the toolchain performance

public code suitable to include unbinned observables in global EFT fits will be released