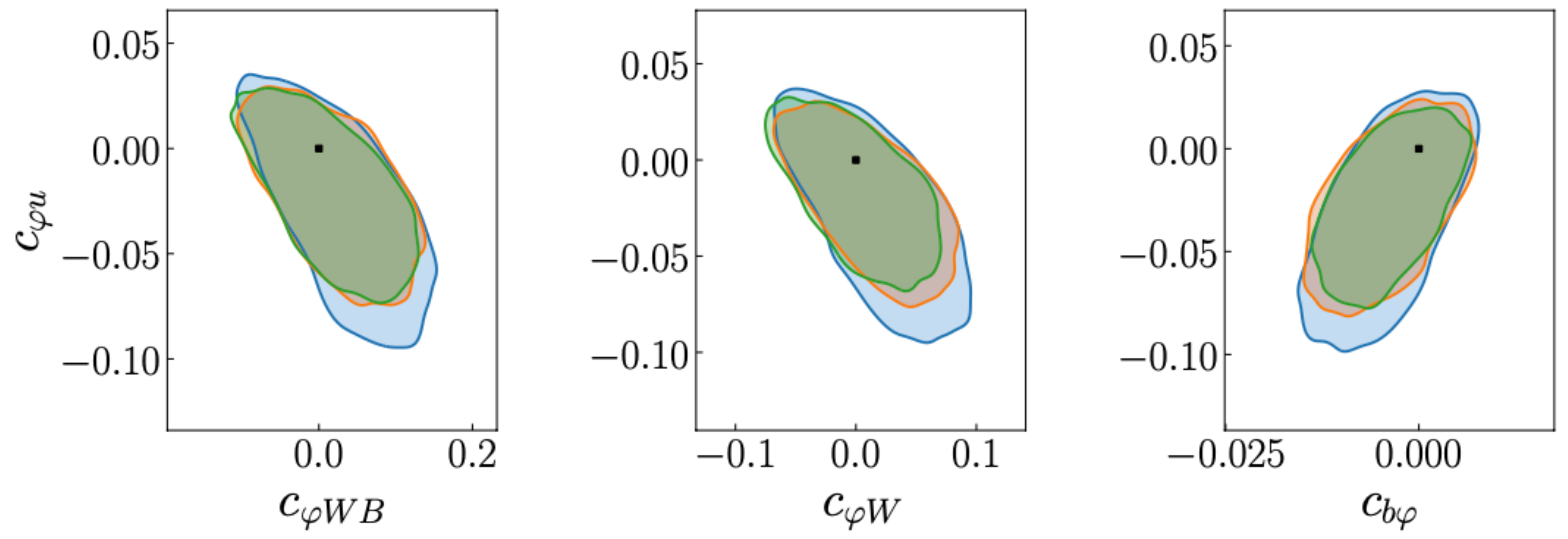


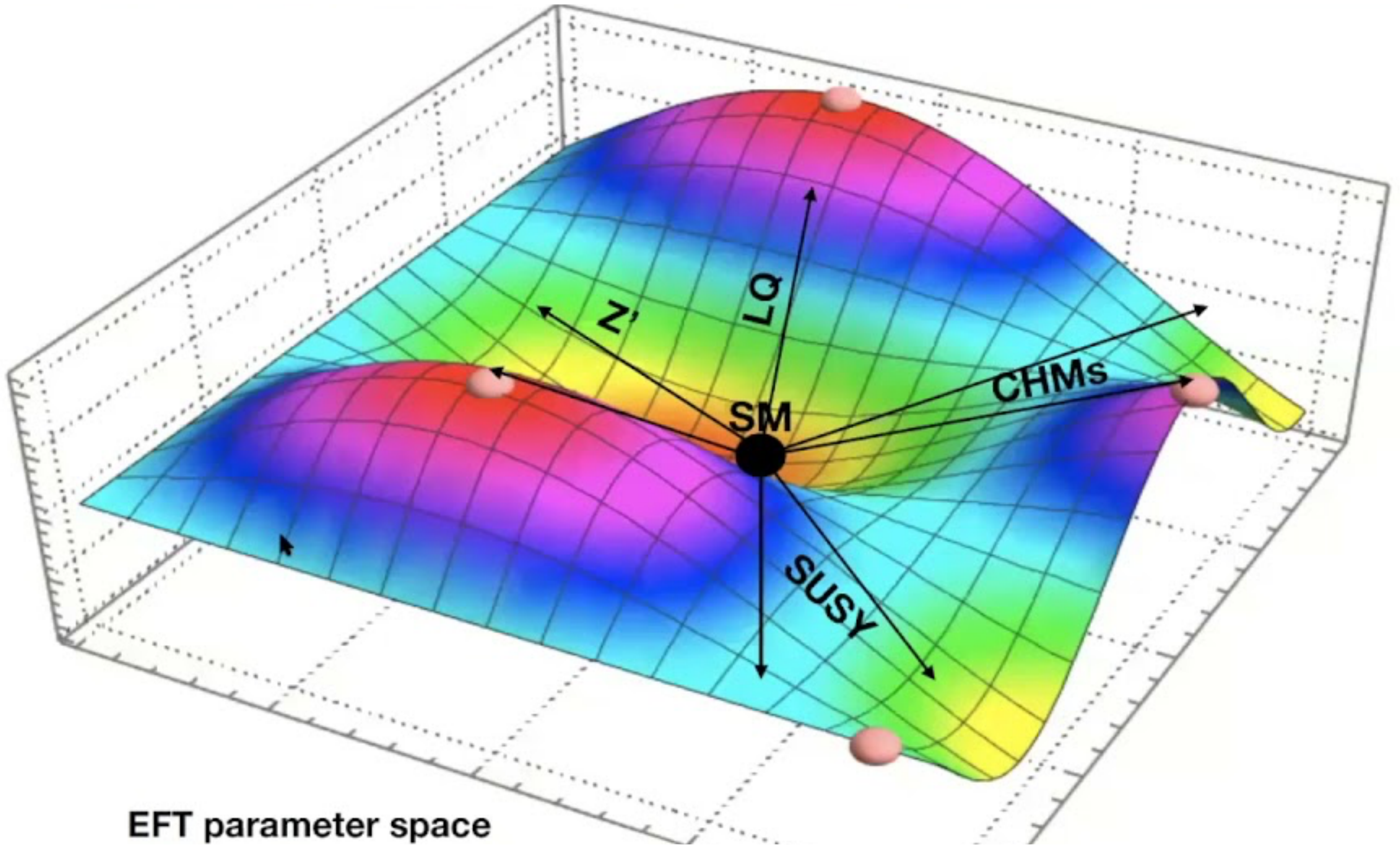


Image credit: V. Sanz

Towards an optimal global SMEFT fit

Machine Learning at GGI
Florence, September 2022

Jaco ter Hoeve
VU Amsterdam & Theory Group, Nikhef

Raquel Gomez-Ambrosio, Maeve Madigan, Juan Rojo, Veronica Sanz

Nikhef

VU VRIJE UNIVERSITEIT AMSTERDAM

Our team

Raquel Gomez-Ambrosio

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

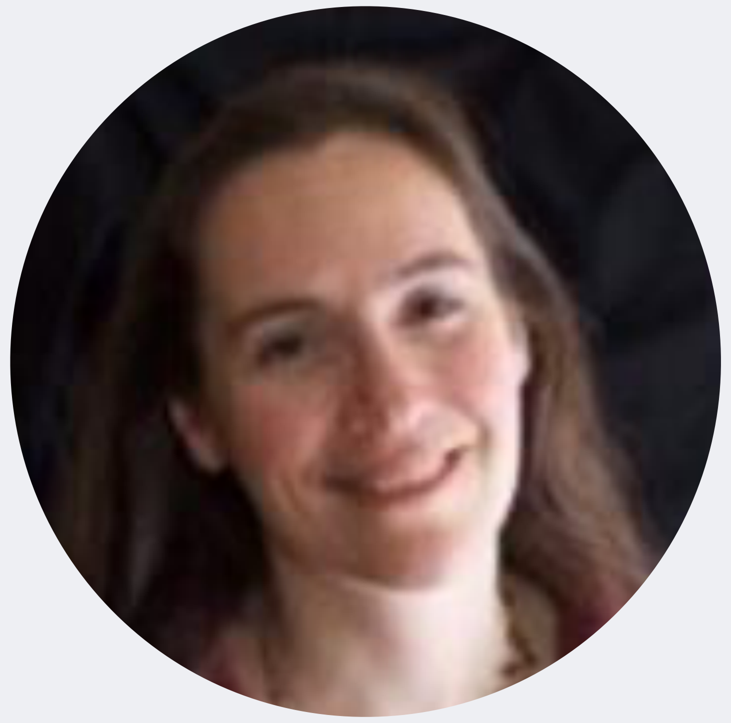
Maeve Madigan

UNIVERSITY OF CAMBRIDGE

Juan Rojo

Nikhef

VU VRIJE UNIVERSITEIT AMSTERDAM



Veronica Sanz

US UNIVERSITY OF SUSSEX

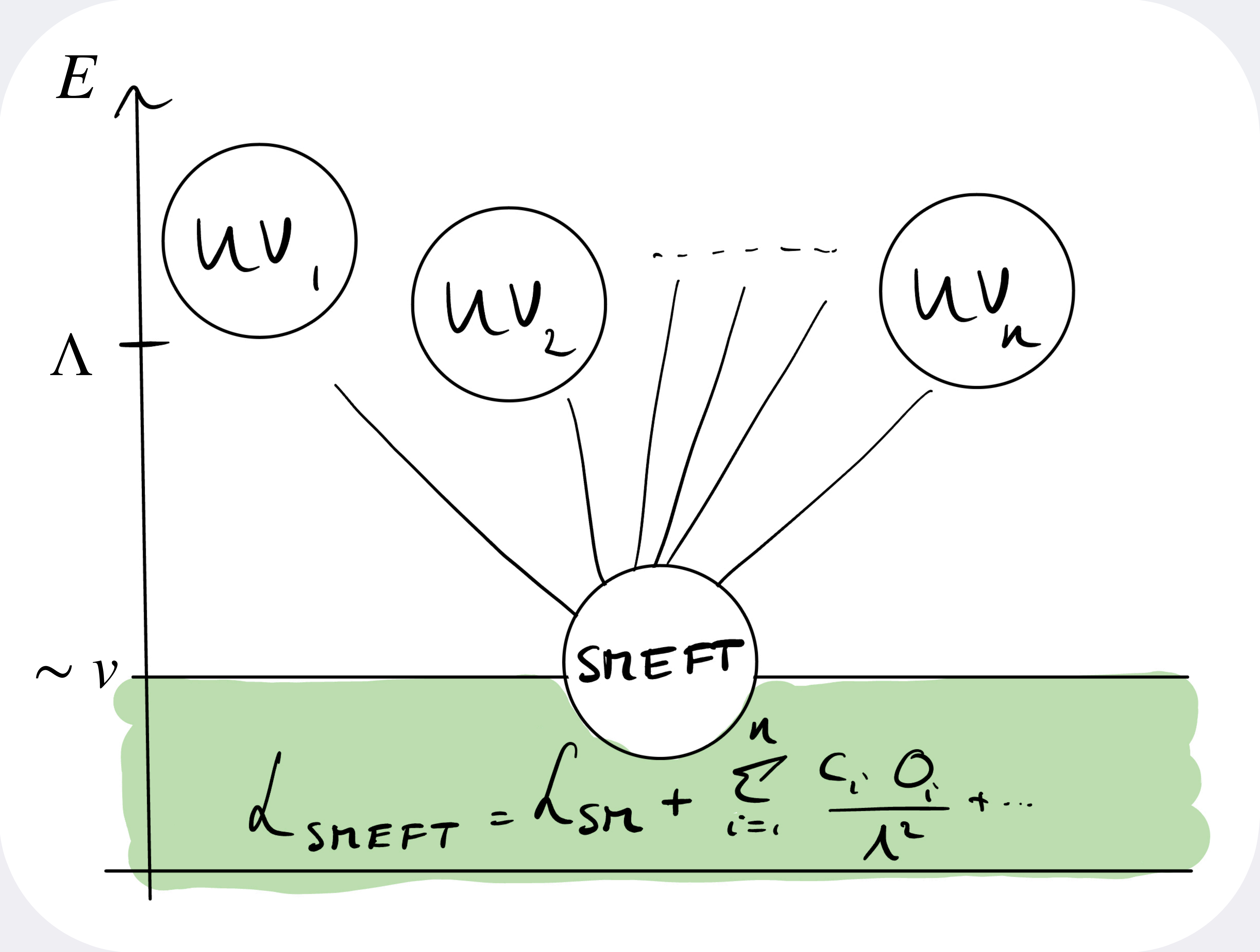
Jaco ter Hoeve

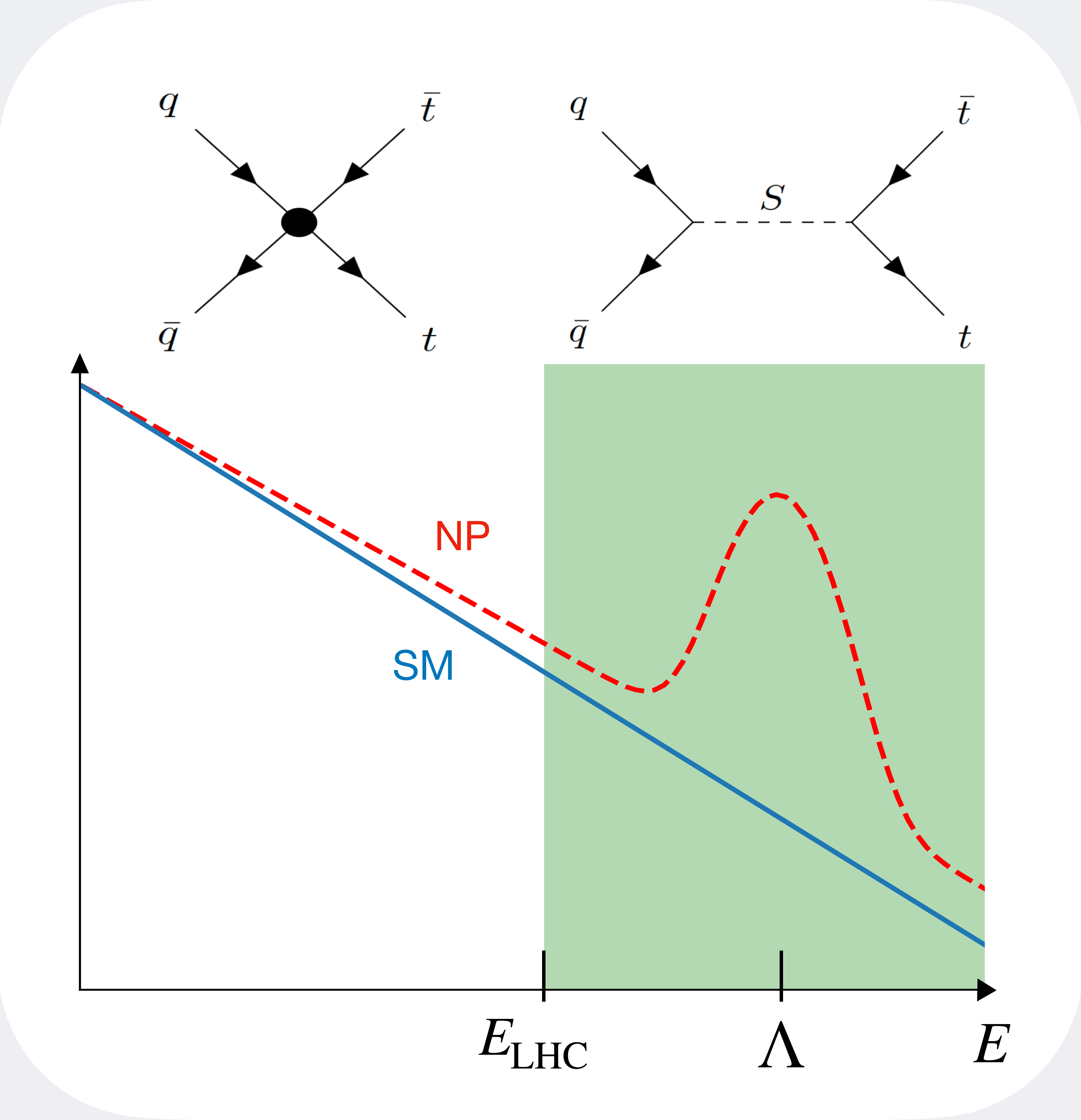
Nikhef

VU VRIJE UNIVERSITEIT AMSTERDAM

The Standard Model as an EFT

Effective Field Theories have become a **main tool** to search for new physics in a “model independent way”

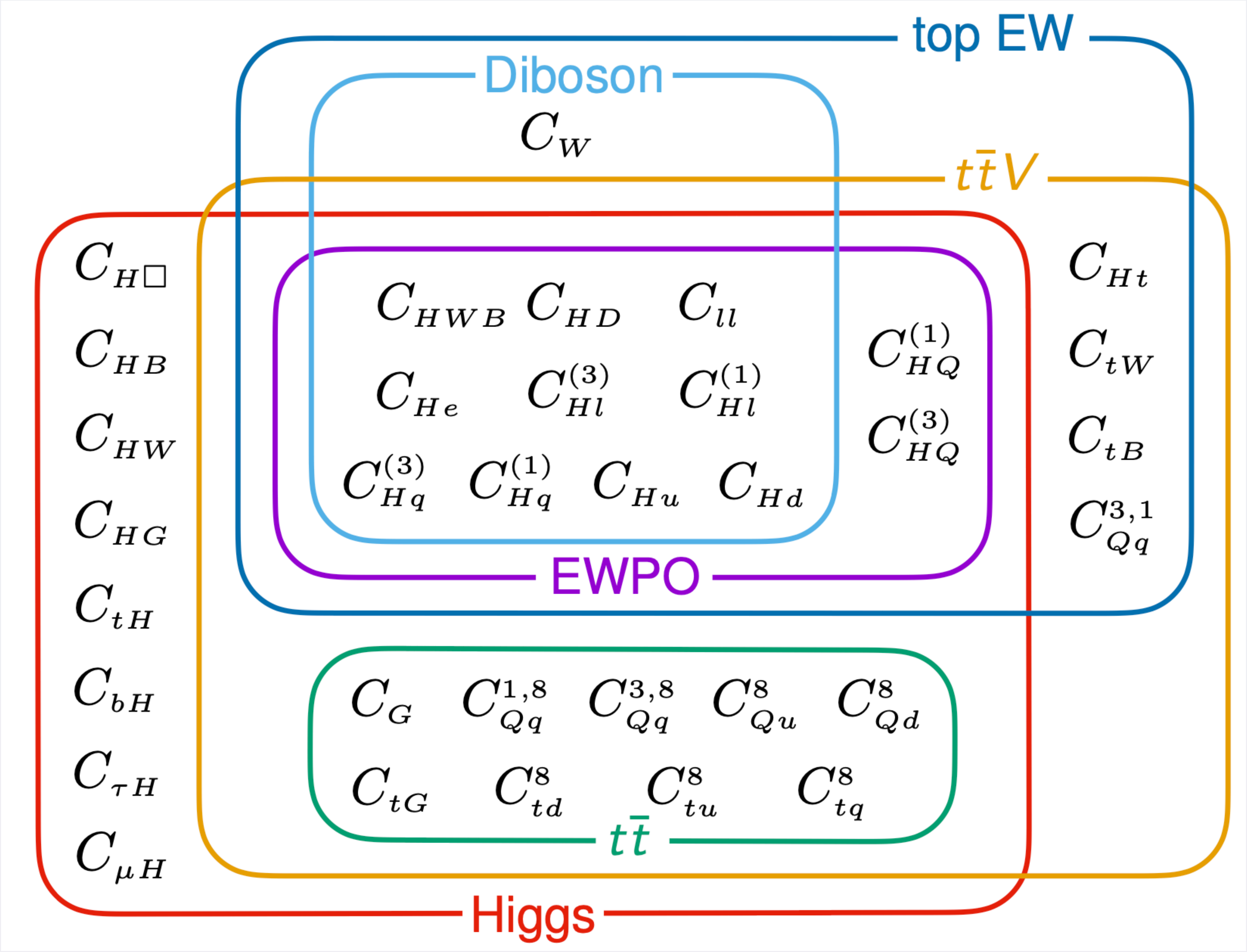




The Standard Model as an EFT

Extend the SM Lagrangian by higher dimensional operators

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i^{N_{d5}} \frac{c_i}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i^{N_{d7}} \frac{c_i}{\Lambda^3}\mathcal{O}_i^{(7)} + \sum_i^{N_{d8}} \frac{b_i}{\Lambda^4}\mathcal{O}_i^{(8)} + \ldots$$



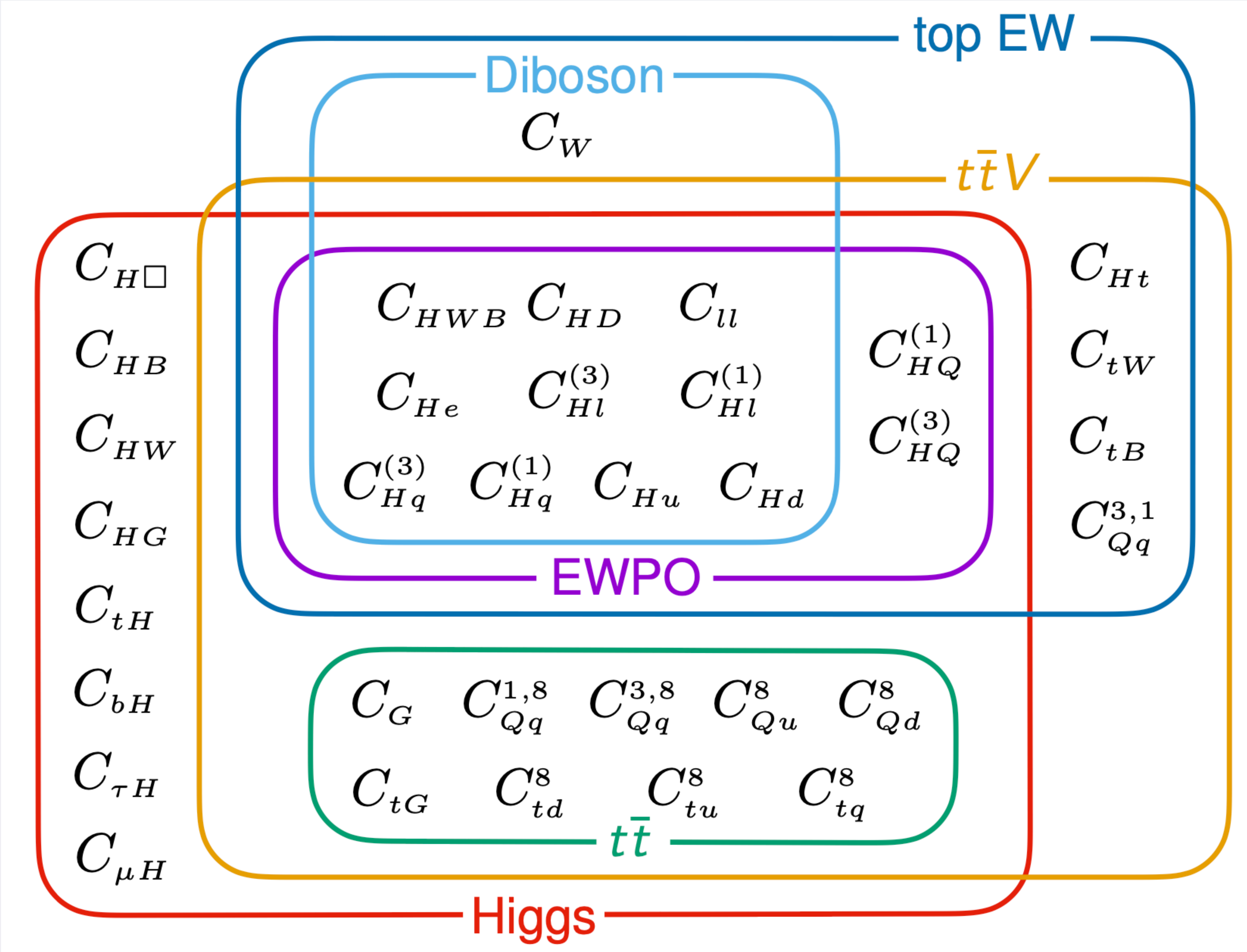
[Fitmaker, 2012.02779]

59 (2499) for one (three) flavour generations

The Standard Model as an EFT

Extend the SM Lagrangian by higher dimensional operators

$$\mathscr{L}_{\mathrm{SMEFT}} = \mathscr{L}_{\mathrm{SM}} + \sum_i^{N_{d5}} \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_i^{N_{d7}} \frac{c_i}{\Lambda^3} \mathcal{O}_i^{(7)} + \sum_i^{N_{d8}} \frac{b_i}{\Lambda^4} \mathcal{O}_i^{(8)} + \dots$$



[Fitmaker, 2012.02779]

59 (2499) for one (three) flavour generations

The Standard Model as an EFT

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i^{N_{d5}} \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \sum_i^{N_{d7}} \frac{c_i}{\Lambda^3} \mathcal{O}_i^{(7)} + \sum_i^{N_{d8}} \frac{b_i}{\Lambda^4} \mathcal{O}_i^{(8)} + \ldots$$

- **Systematic parameterisation** of the theory space in the vicinity of the SM
- **Low energy limit** of generic UV-complete theories at high energies
- Assumes the **SM field content and symmetries**
- **Complete basis** at any given mass dimension
- Fully **renormalizable** QFT
- Can be **matched to any BSM model** that reduces to the SM at low energies

The Standard Model as an EFT

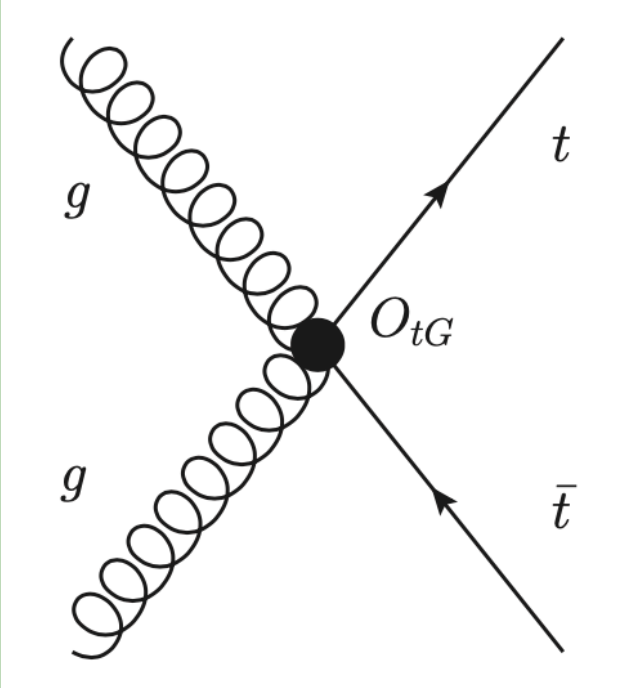
Example: From operator to modified cross section

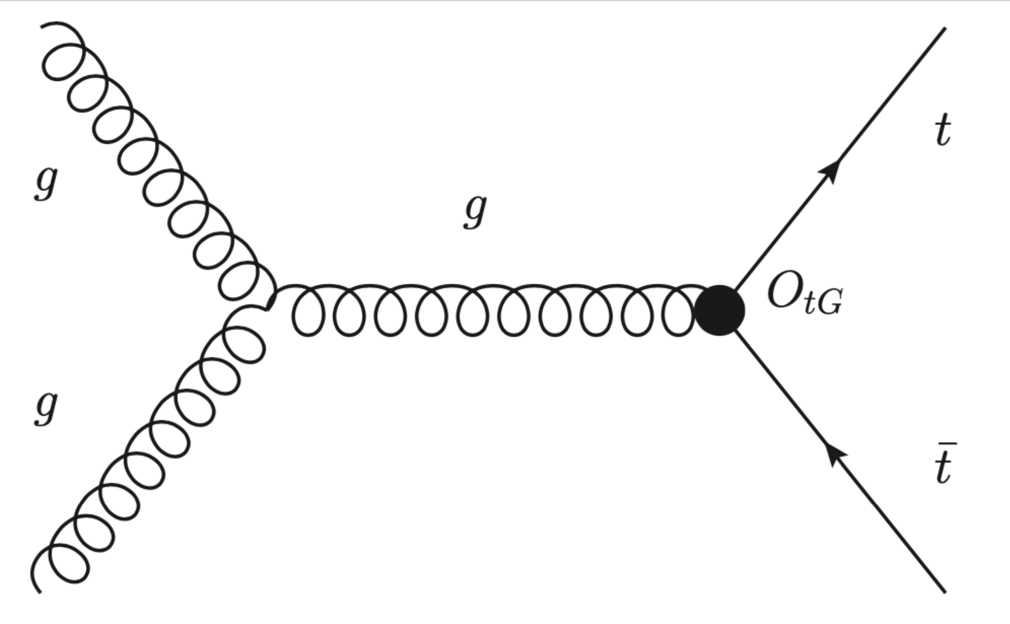
$$\mathcal{O}_{tG} = ig_s \left(\bar{Q}\tau^{\mu\nu}T_A t\right)\tilde{\varphi}G^A_{\mu\nu}$$

$$G^A_{\mu\nu} = \partial_\mu G^A_\nu - \partial_\nu G^A_\mu + g_s f^{ABC}G^B_\mu G^C_\nu$$

$$\bar{Q} = (t_L \quad b_L)$$

$$\varphi \sim \begin{pmatrix} 0 \\ v \end{pmatrix}$$





$$\sigma = \sigma_{\text{SM}} \times \left(1 + \frac{c_{tG}}{\Lambda^2}a + \frac{c_{tG}^2}{\Lambda^4}b\right)$$

Linear EFT corrections: interference SM-EFT$_{d6}$

Quadratic EFT corrections: EFT$_{d6}$-EFT$_{d6}$

The Standard Model as an EFT

from the SMEFT **Lagrangian** ....

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

SM

EFT$_{\text{d6}}$

EFT$_{\text{d8}}$

to **cross-sections**....

Linear EFT corrections:
interference SM-EFT$_{\text{d6}}$

Quadratic EFT corrections:
EFT$_{\text{d6}}$-EFT$_{\text{d6}}$

$$\sigma_{\text{SMEFT}} \approx \sigma_{\text{SM}} \times \left(1 + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \kappa_i + \sum_{i,j}^{N_{d6}} \frac{c_i c_j}{\Lambda^4} \tilde{\kappa}_{ij} \right)$$

Evaluate at (N)NLO QCD + NLO EW

Evaluate at NLO QCD with SMEFT@NLO

The Standard Model as an EFT

to **cross-sections**….

Linear EFT corrections:
interference SM-EFT$_{d6}$

Quadratic EFT corrections:
EFT$_{d6}$-EFT$_{d6}$

$$\sigma_{\rm SMEFT} \approx \sigma_{\rm SM} \times \left(1 + \sum_{i}^{N_{d6}} \frac{c_i}{\Lambda^2} \kappa_i + \sum_{i,j}^{N_{d6}} \frac{c_i c_j}{\Lambda^4} \tilde{\kappa}_{ij} \right)$$

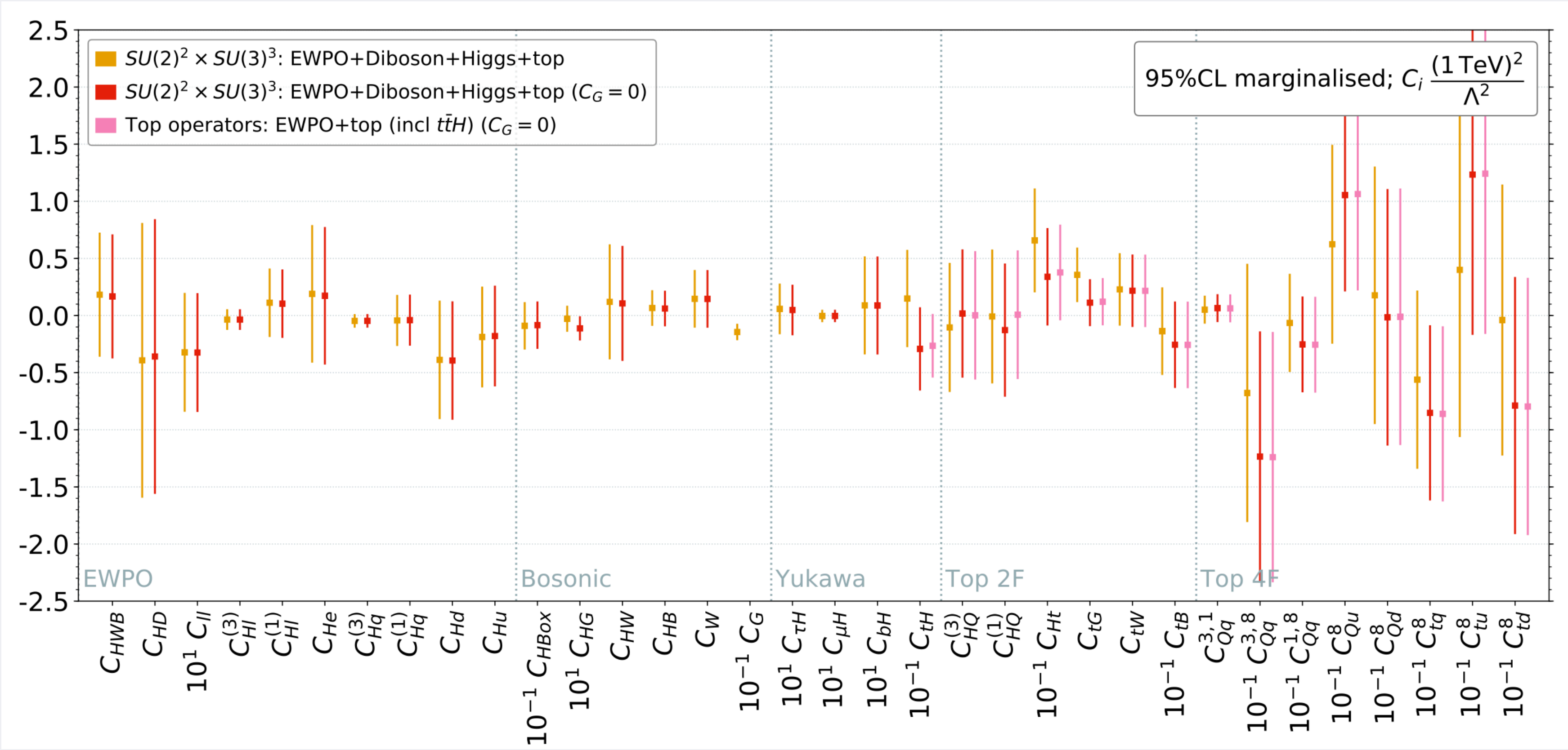
to **constraints** on the EFT parameters

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

Multi-Gaussian log likelihood optimisation problem

Global EFT analyses

Several groups have presented global EFT analyses combining data from **many different** processes (top, Higgs, diboson,…), yielding **sensitivity to many directions** in the EFT parameter space



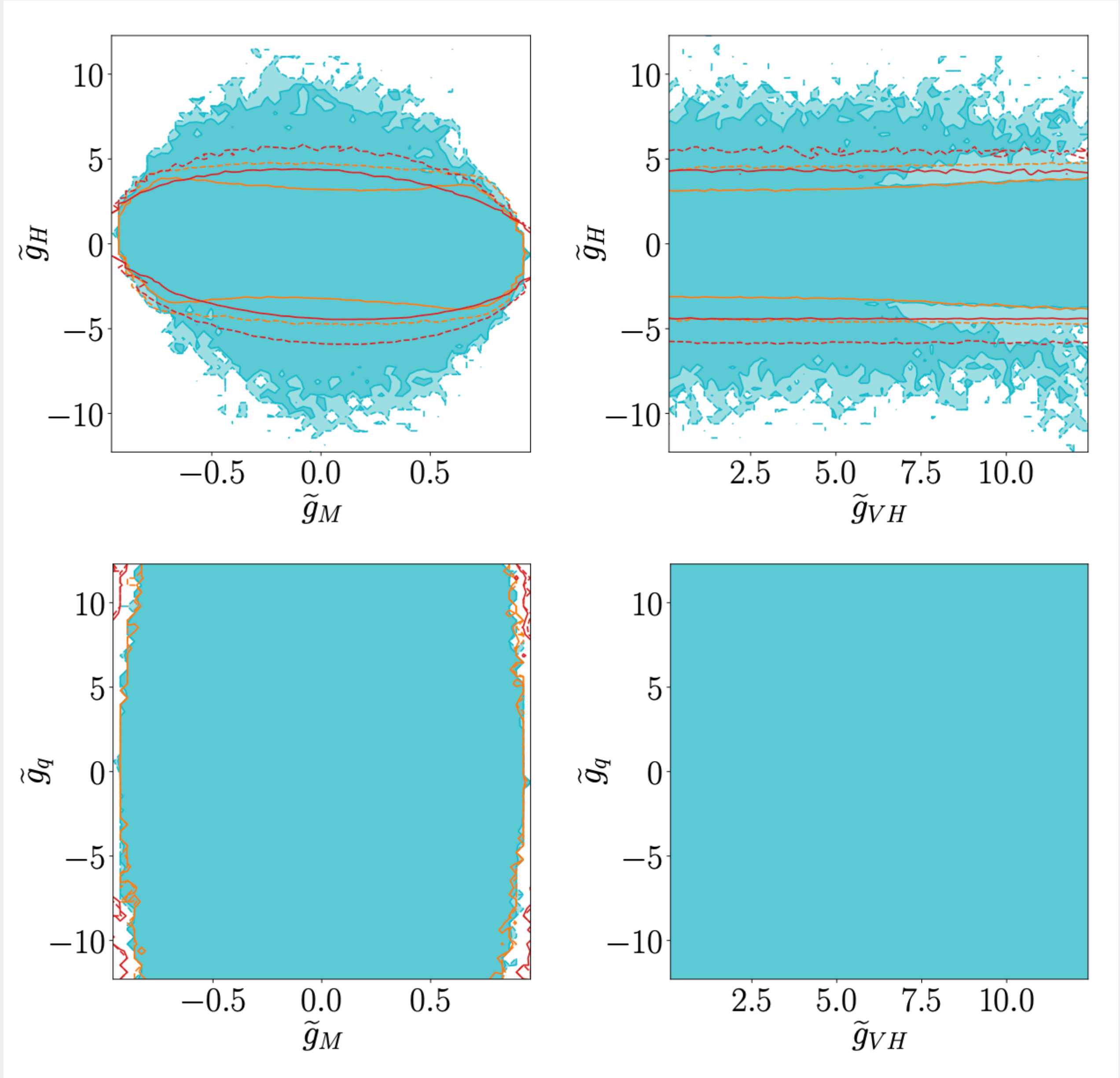
Fitmaker

[2012.02779]

SMEFiT

[2105.00006]

Combines 317 cross section measurements
to constrain 50 EFT parameters



SFitter

[2108.01094]

None of these measurements have been **optimised for EFT studies**, can one do better?

From binned to unbinned

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

Number of bins

Number of observed events

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

Number of expected events

Only statistical uncertainties here
No cross-correlations

Q1: What is the optimal number of bins?

Q2: How much information does one gain/lose by measuring the cross-section in additional kinematic variables?

- In the tails of the EFT distributions, the # of events can be small (i.e. < 30), and one must use the Poissonian likelihood:

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

From binned to unbinned

<u>**Main goal**</u>: construct unbinned observables from ML and assess their relevance for **global EFT fits** by comparing their impact to those of "traditional binned" observables

Unbinned extended likelihood

# of expected events

# of observed events

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

Probability distribution in the final state kinematics

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

Fully differential cross-section

e.g. in $ZH \to \ell^+\ell^- b\bar{b}$ $\quad \boldsymbol{x} = (p_T^Z, p_T^b, \Delta R_{b\bar{b}}, \Delta\phi_{l,b}, \ldots)$ **"Features"**

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

From binned to unbinned

- For n_{eft} EFT coefficients c_i, we can express the EFT cross-section as

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

- For the purpose of limit setting, it is sufficient to consider its ratio to some reference point (e.g. the SM)

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

- Extend the global EFT likelihood by the unbinned likelihood to assess their relevance (avoid **double counting**):

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

Cross section ML parametrisation

- We thus need to parameterise the dependence of the distribution ratio on the kinematics

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

- Adopt NN as **universal unbiased interpolants** as a proxy of the likelihood ratio that is fast to evaluate

- The structure of the EFT cross section makes it possible to **parallelise** the training to any arbitrary number of EFT parameters

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

- We thus need to parameterise the dependence of the distribution ratio on the kinematics

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

- Use the "Likelihood ratio" trick to obtain a decision boundary that is 1-1 with $r_\sigma(\boldsymbol{x}, \boldsymbol{c})$

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

$g \to 0$ for EFT

$g \to 1$ for SM

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

Cross section ML parametrisation

- Generate **Monte Carlo data** with single EFT parameter activated + SM baseline

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

to extract the coefficient function accompanying $c_j^{(\mathrm{tr})}$ at the **linear level**

$$r_\sigma(\boldsymbol{x}, c_j^{(\mathrm{tr})}) = 1 + c_j^{(\mathrm{tr})} \cdot \mathrm{NN}^{(j)}(\boldsymbol{x}) \qquad \mathrm{NN}^{(j)}(\boldsymbol{x}) \rightarrow r_\sigma^{(j)}(\boldsymbol{x})$$

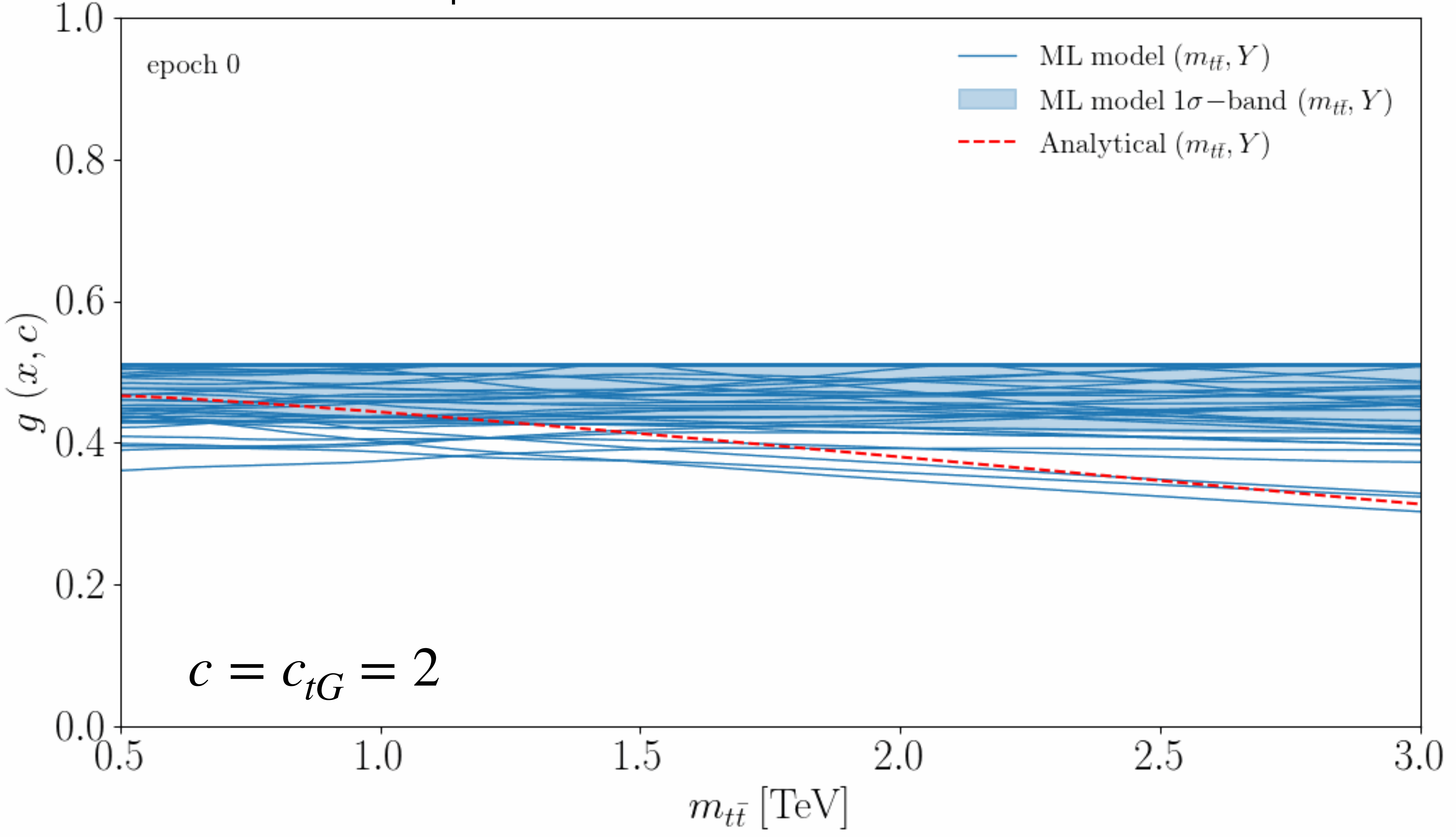
- The same logic holds at the **quadratic level**

$$\mathcal{D}_{\mathrm{eft}}(\boldsymbol{c} = (0,\ldots,0,c_j^{(\mathrm{tr})},0,\ldots,0,c_k^{(\mathrm{tr})},0,\ldots,0)) \qquad r_\sigma(\boldsymbol{x}, c_j^{(\mathrm{tr})}) = 1 + c_j^{(\mathrm{tr})}c_k^{(\mathrm{tr})} \cdot \mathrm{NN}^{(j,k)}(\boldsymbol{x})$$

Neural network training

Model uncertainties are estimated by means of the **Monte Carlo replica method:** train a collection of 50 NN instances on independent MC datasets

Animations are static in the pdf



Training settings

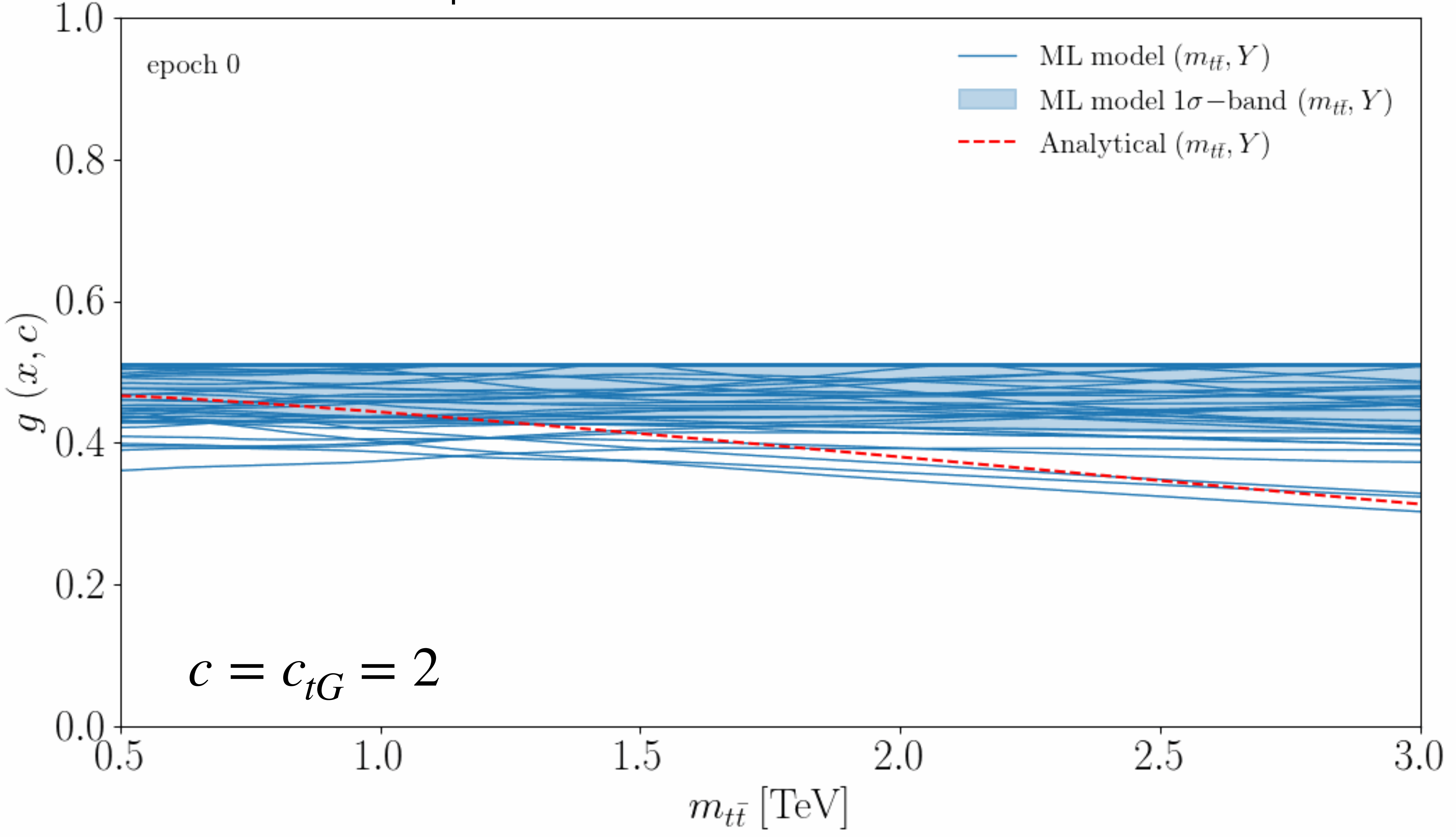
- LO QCD MC dataset (100K)
- Mini-batch gradient descent
- Training/Validation sets (80/20)
- Runtime ~20 min per core (1 replica/core)
- 3 hidden layers (100 units)
- ReLU activation functions

The MC replica method allows us to propagate model and methodological uncertainties to the space of EFT parameters! **(NNPDF approach)**

Neural network training

Model uncertainties are estimated by means of the **Monte Carlo replica method:** train a collection of 50 NN instances on independent MC datasets

Animations are static in the pdf



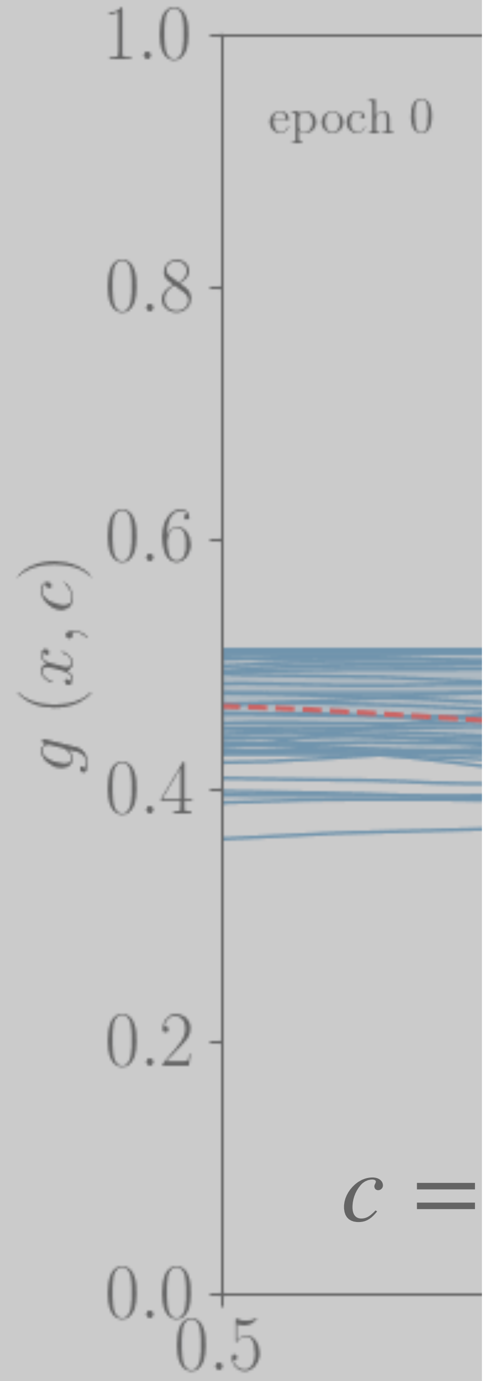
Training settings

- LO QCD MC dataset (100K)
- Mini-batch gradient descent
- Training/Validation sets (80/20)
- Runtime ~20 min per core (1 replica/core)
- 3 hidden layers (100 units)
- ReLU activation functions

The MC replica method allows us to propagate model and methodological uncertainties to the space of EFT parameters! **(NNPDF approach)**

Neural network training

Animations are s



Replicas

NNPDF Approach: Monte Carlo sample (empirical distribution)

Naïve idea: **uncertainty comes from data**, let's propagate the data distribution back through the *regularized inverse function*:

$$f_i(x) = \arg\min_f (\mathcal{L}(D_i)) = \mathcal{NN}(x \mid D_i)$$

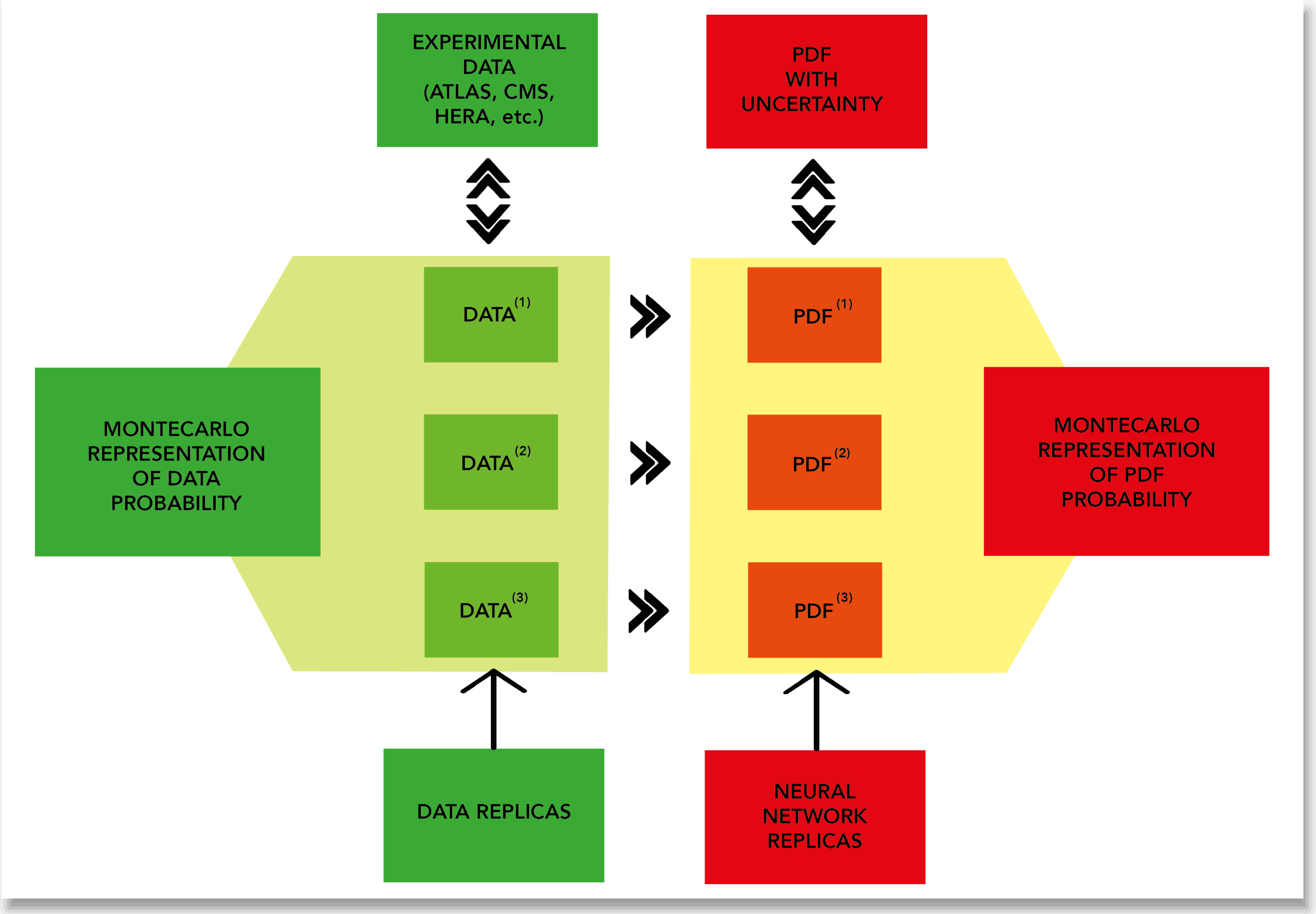
Loss contains the χ^2, but χ^2 is not the end of the story.

Very simple (possibly expensive[a]) usage: in order to obtain the **distribution of predictions**, apply your theoretical calculation to the element of the sample:

$$\mathcal{P}red\mathcal{S} = \{\mathcal{F}(f) \;\; \forall f \in \mathcal{P}ost\mathcal{S}\}$$

[a] To address this compressed sets are distributed along the main release. There is also a program to obtain specialized minimal sets for dedicated studies:

https://github.com/scarrazza/smpdf



Alessandro Candido's talk from Tuesday 11

tings

set (100K)

nt descent

n sets (80/20)

per core (1

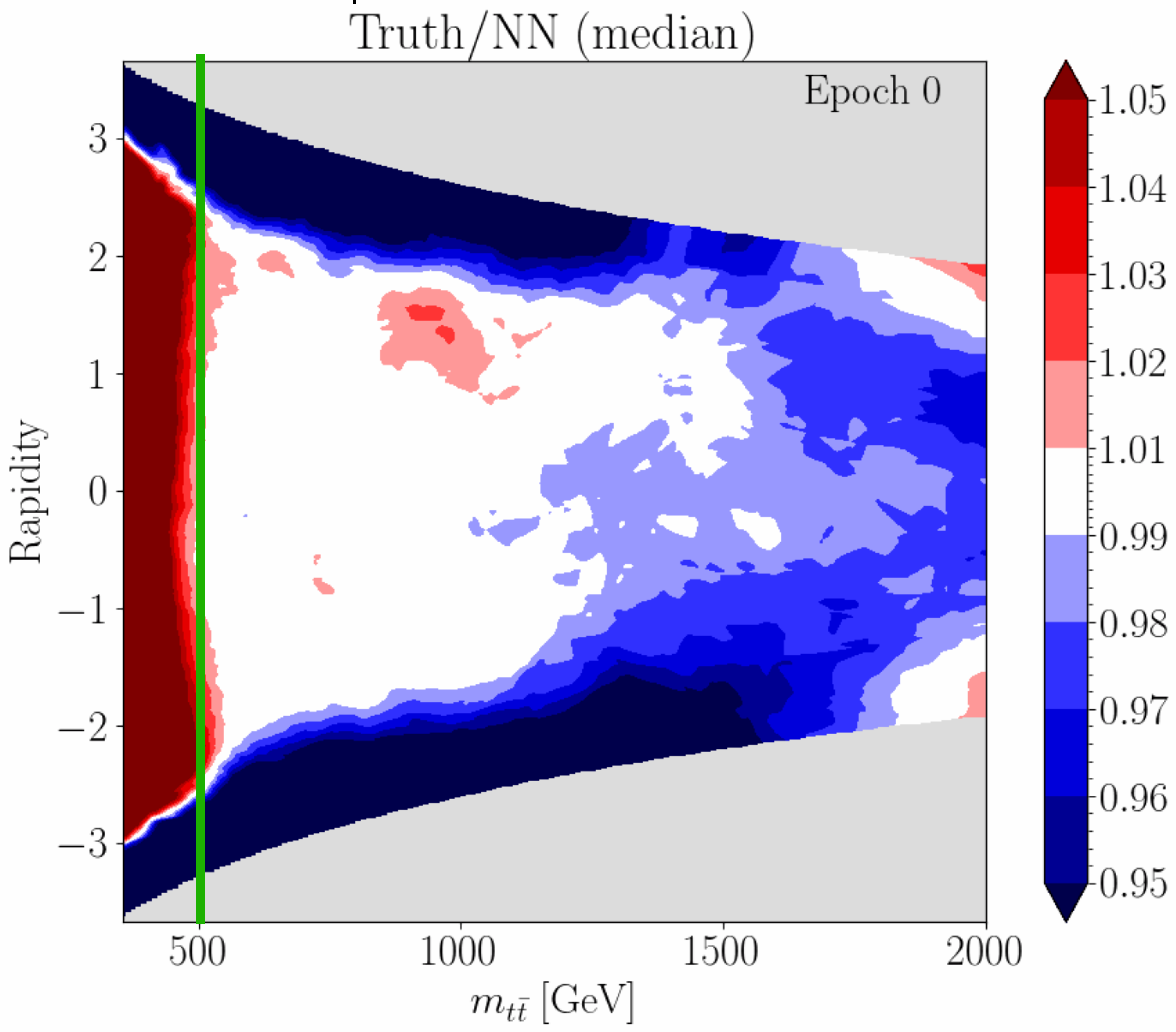
00 units)

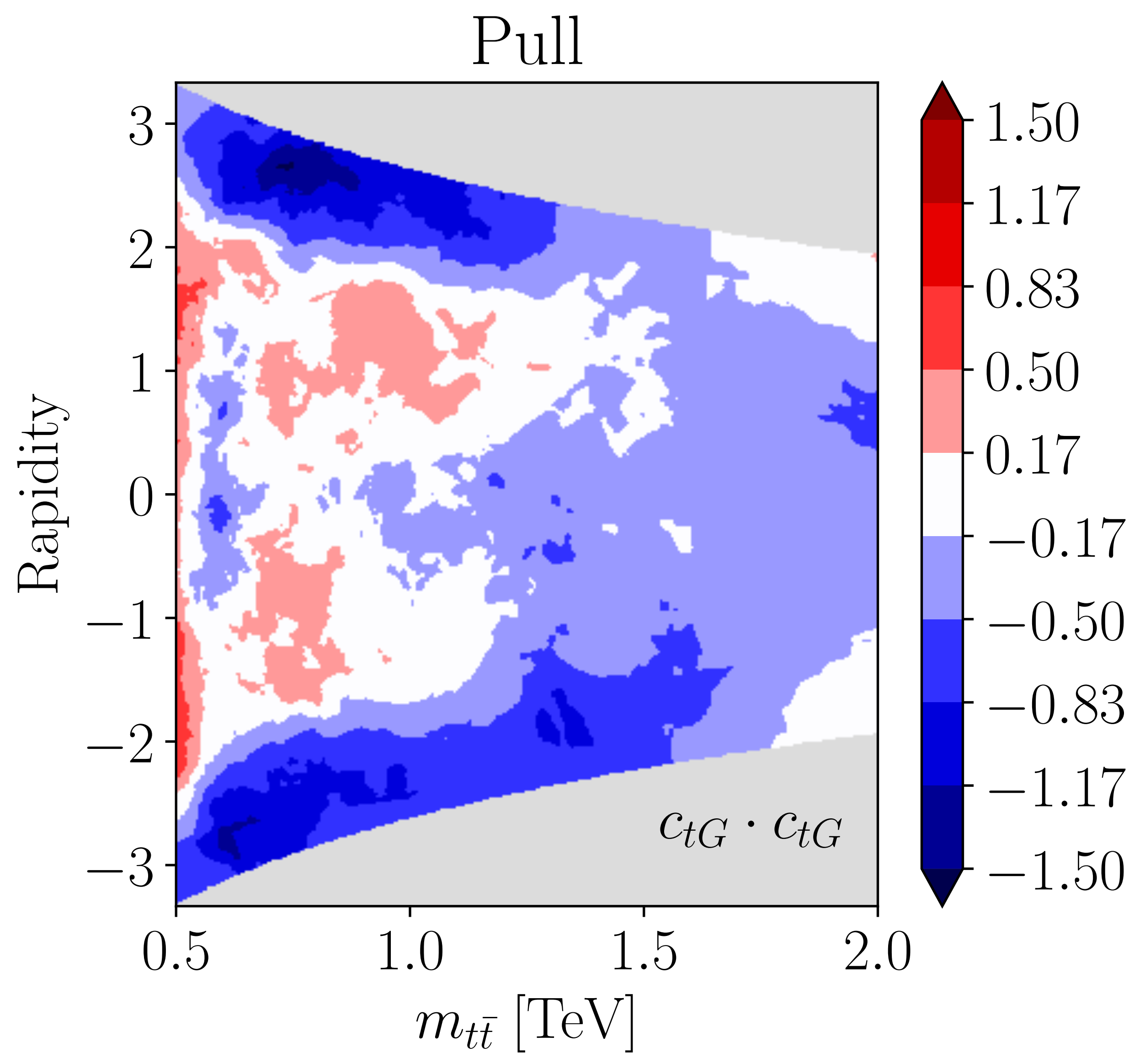
unctions

The MC replica method allows us to propagate model and methodological uncertainties to the space of EFT parameters! **(NNPDF approach)**

Neural network training

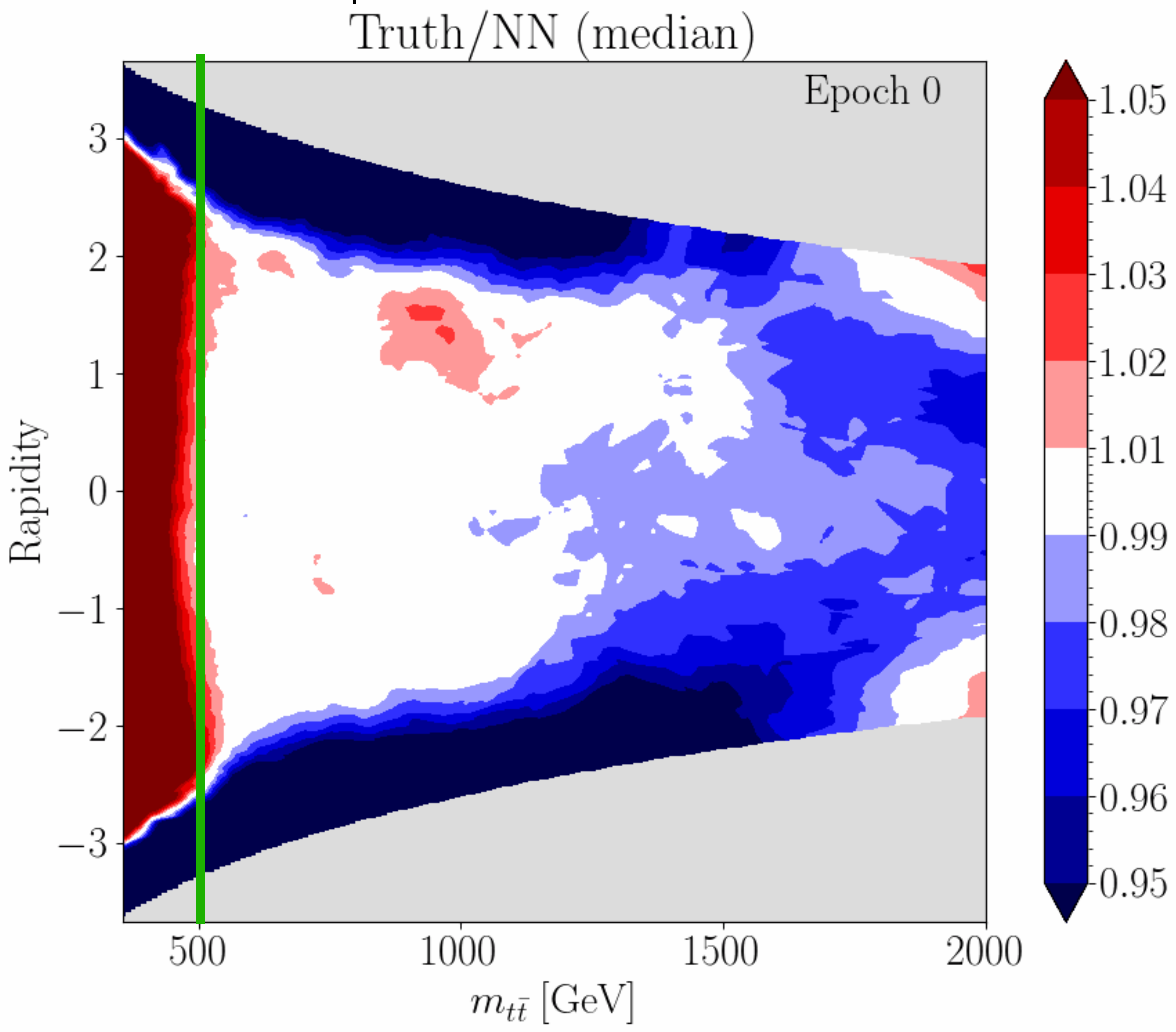
Animations are static in the pdf

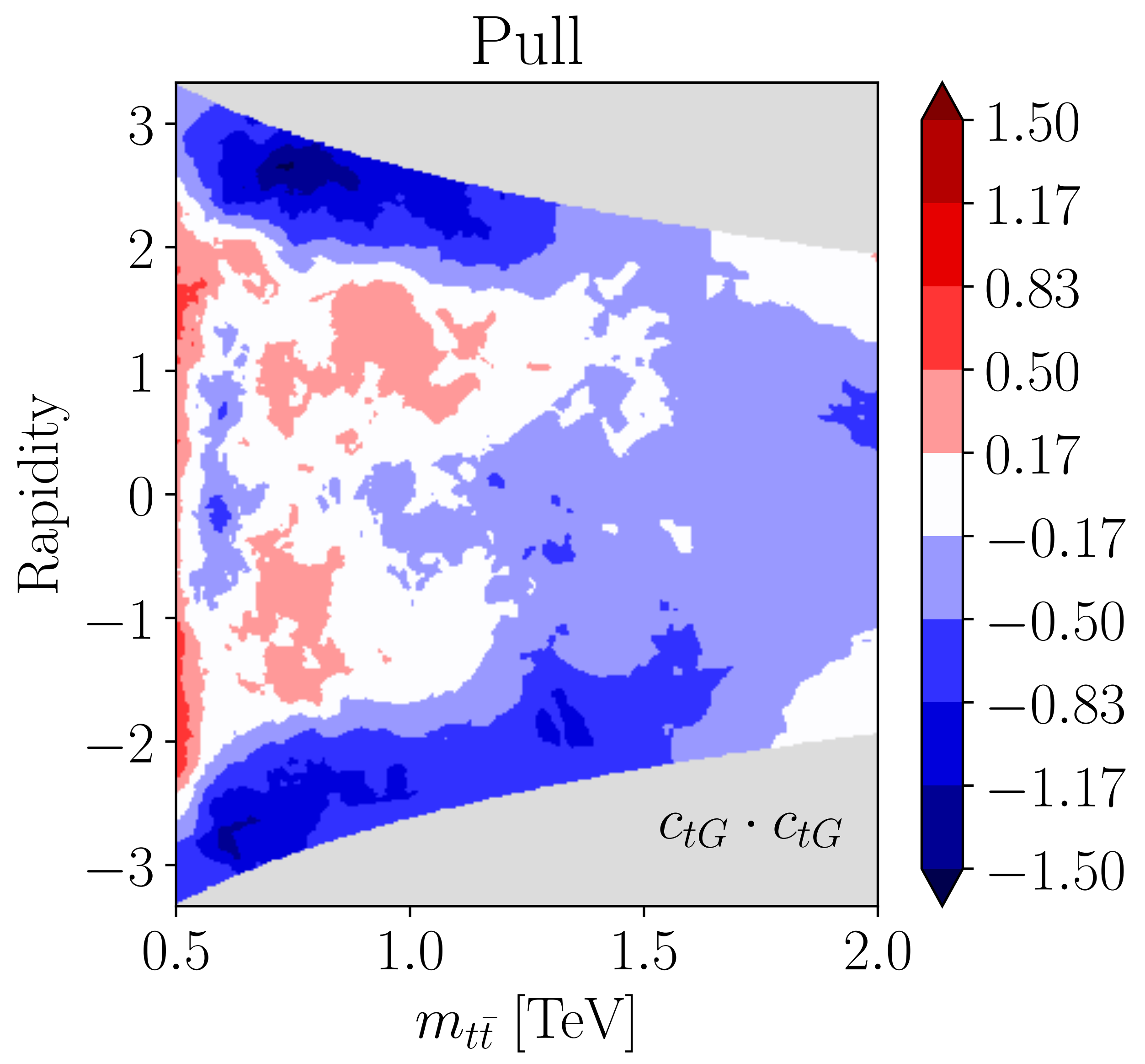




Neural network training

Animations are static in the pdf

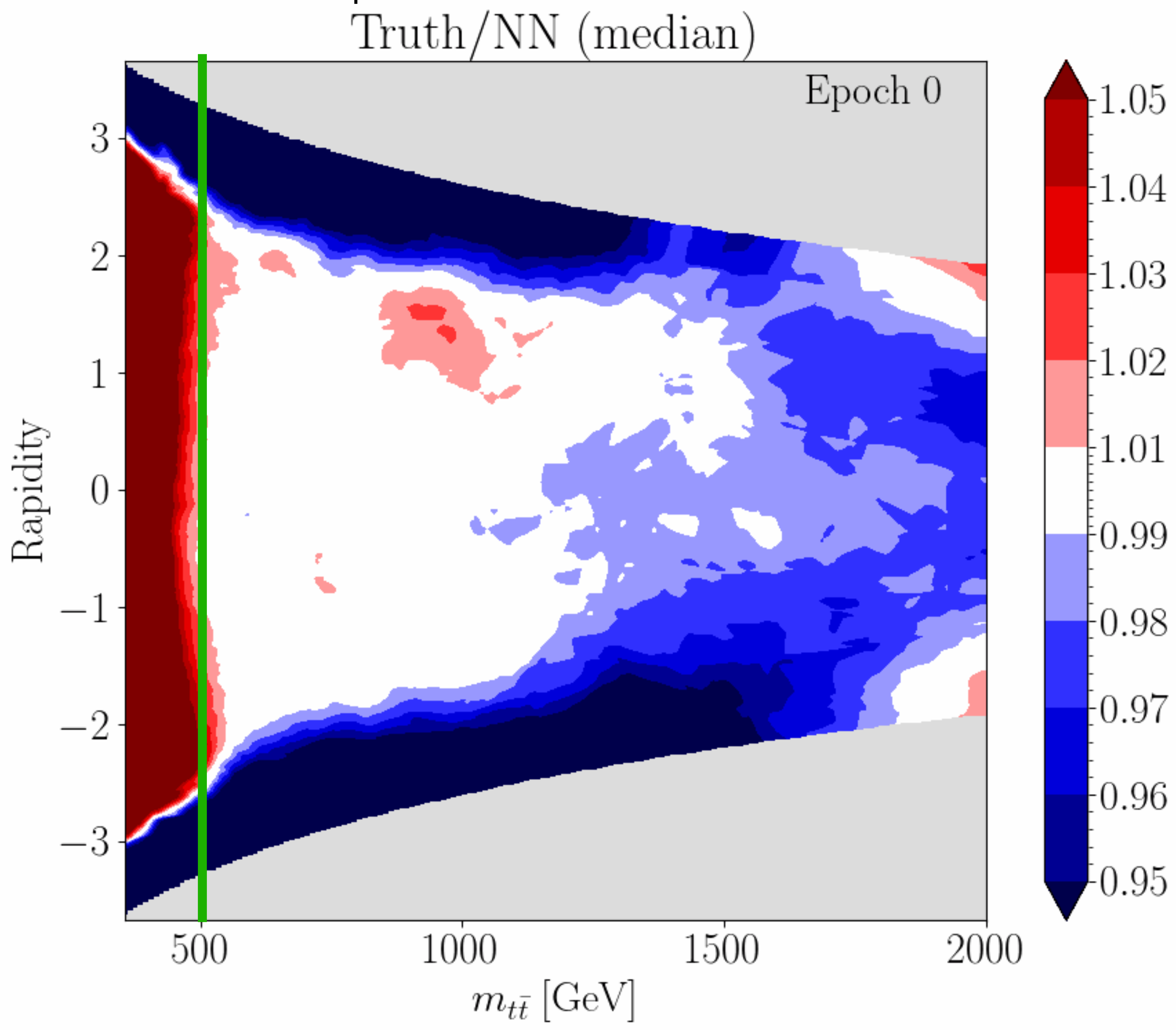


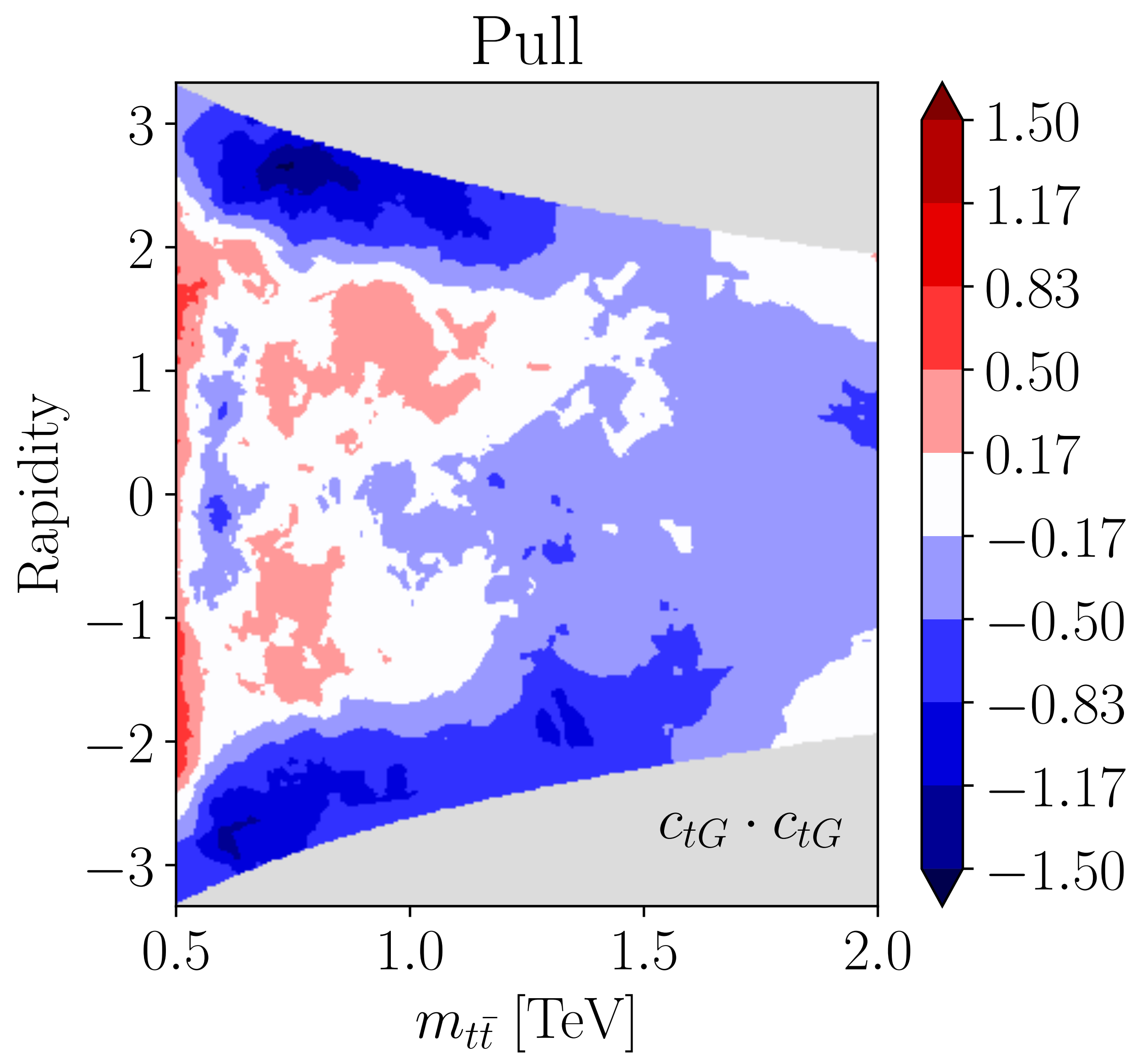


Challenge: correctly describing tails of distributions (low stats)

Neural network training

Animations are static in the pdf





Challenge: correctly describing tails of distributions (low stats)

Cross section ML parametrisation

- Cross section **positivity** can be enforced through either Lagrange multipliers ….

$$L[g] \to L[g] + \lambda \cdot \mathrm{ReLU}\left(\frac{g-1}{g}\right)$$

Non-zero whenever $r_\sigma(\boldsymbol{x}, \boldsymbol{c}) < 0$

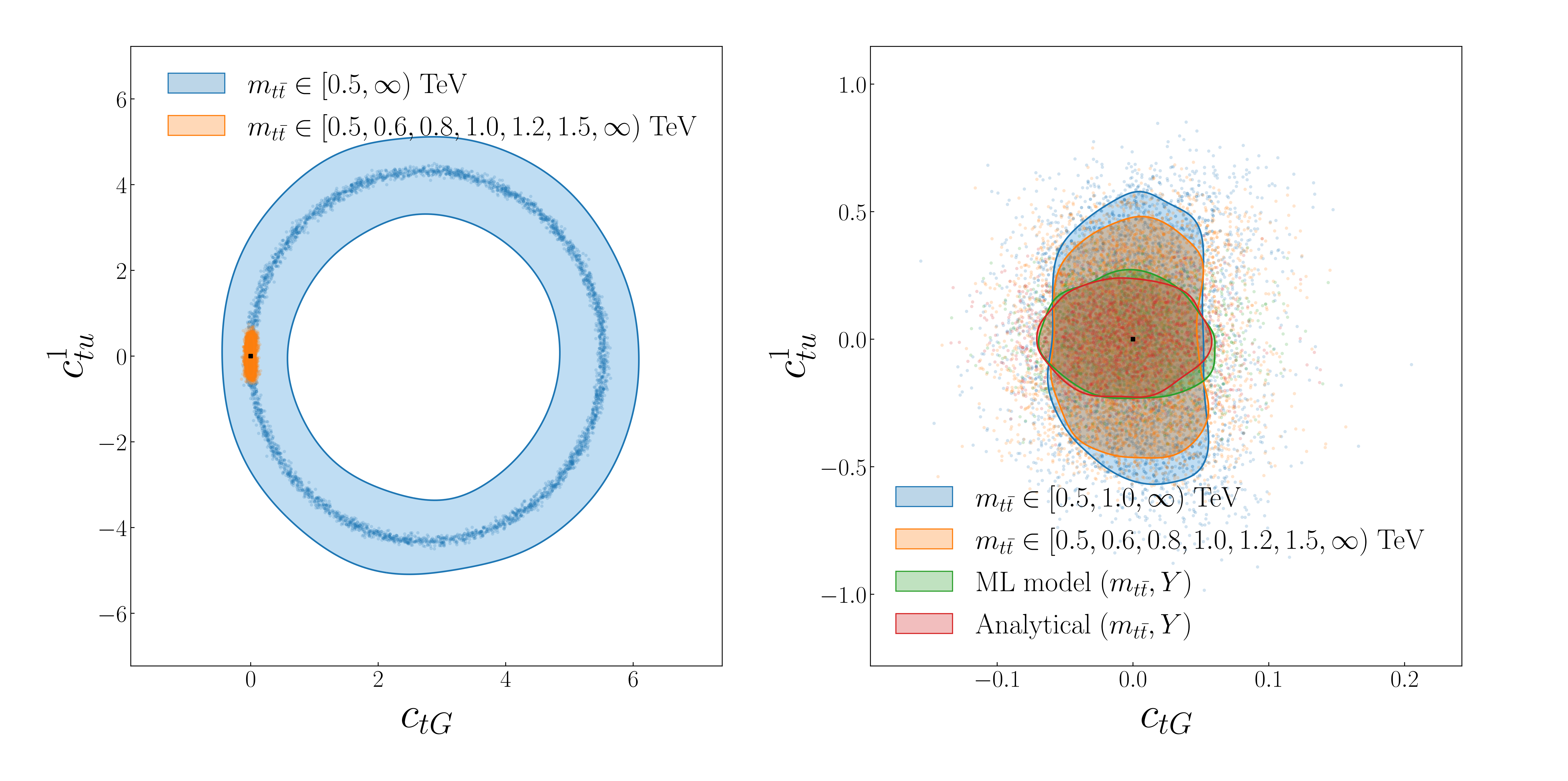
…. or a final ReLU depending on the value of $c_j^{(\mathrm{tr})}$

$$\mathrm{NN}^{(j)}(\boldsymbol{x}) \to \mathrm{NN}^{(j)}(\boldsymbol{x}; c_j^{(\mathrm{tr})}) = \begin{cases} \mathrm{ReLU}(\mathrm{NN}^{(j)}(\boldsymbol{x})) - 1/c_j^{(\mathrm{tr})}, & \text{if } c_j^{(\mathrm{tr})} > 0 \\ -\mathrm{ReLU}(\mathrm{NN}^{(j)}(\boldsymbol{x})) - 1/c_j^{(\mathrm{tr})}, & \text{if } c_j^{(\mathrm{tr})} < 0 \end{cases}$$

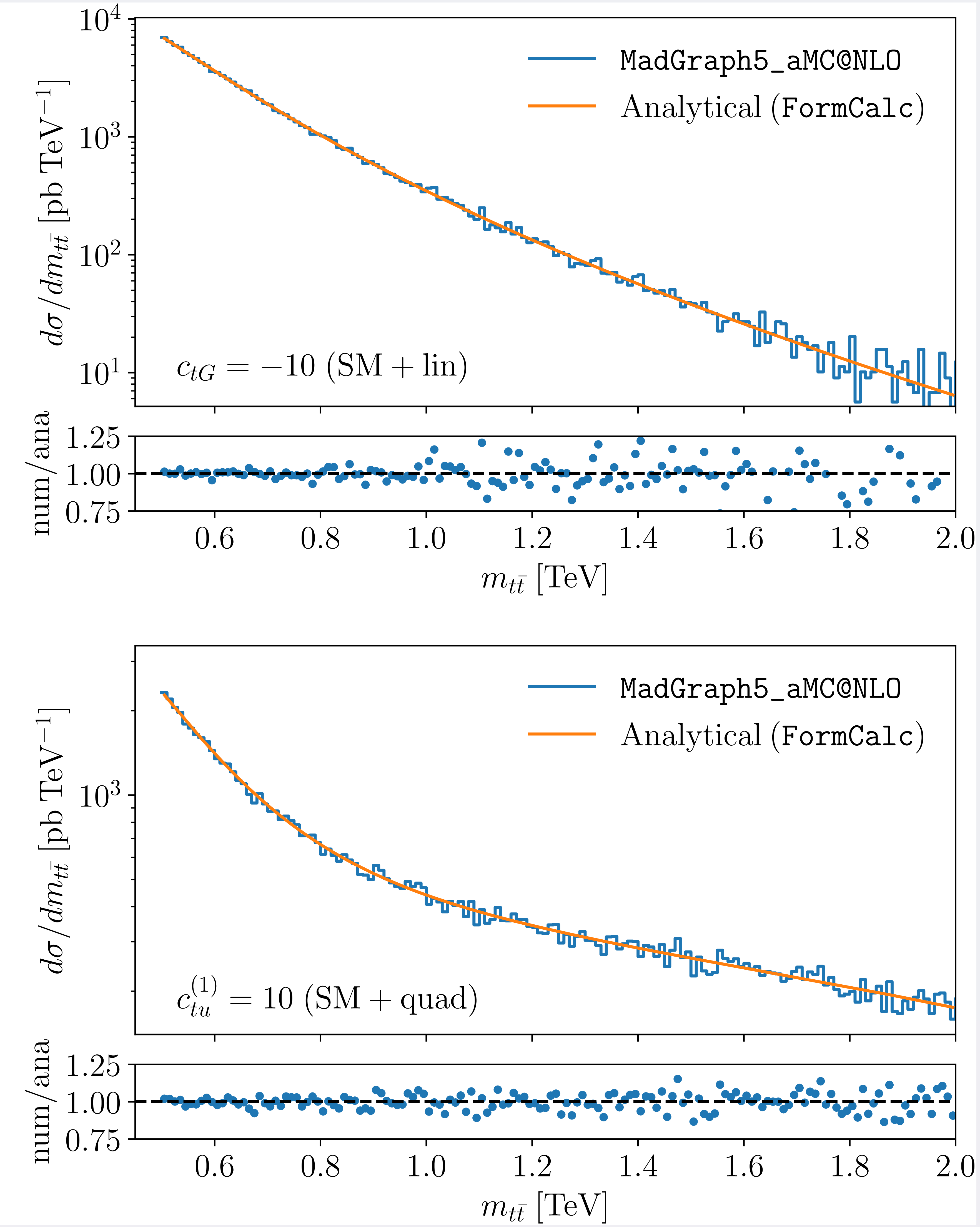
- # of processors scales like $\mathcal{O}(n_{\mathrm{eft}})$, making our approach **suitable** for a large EFT parameter space

Results

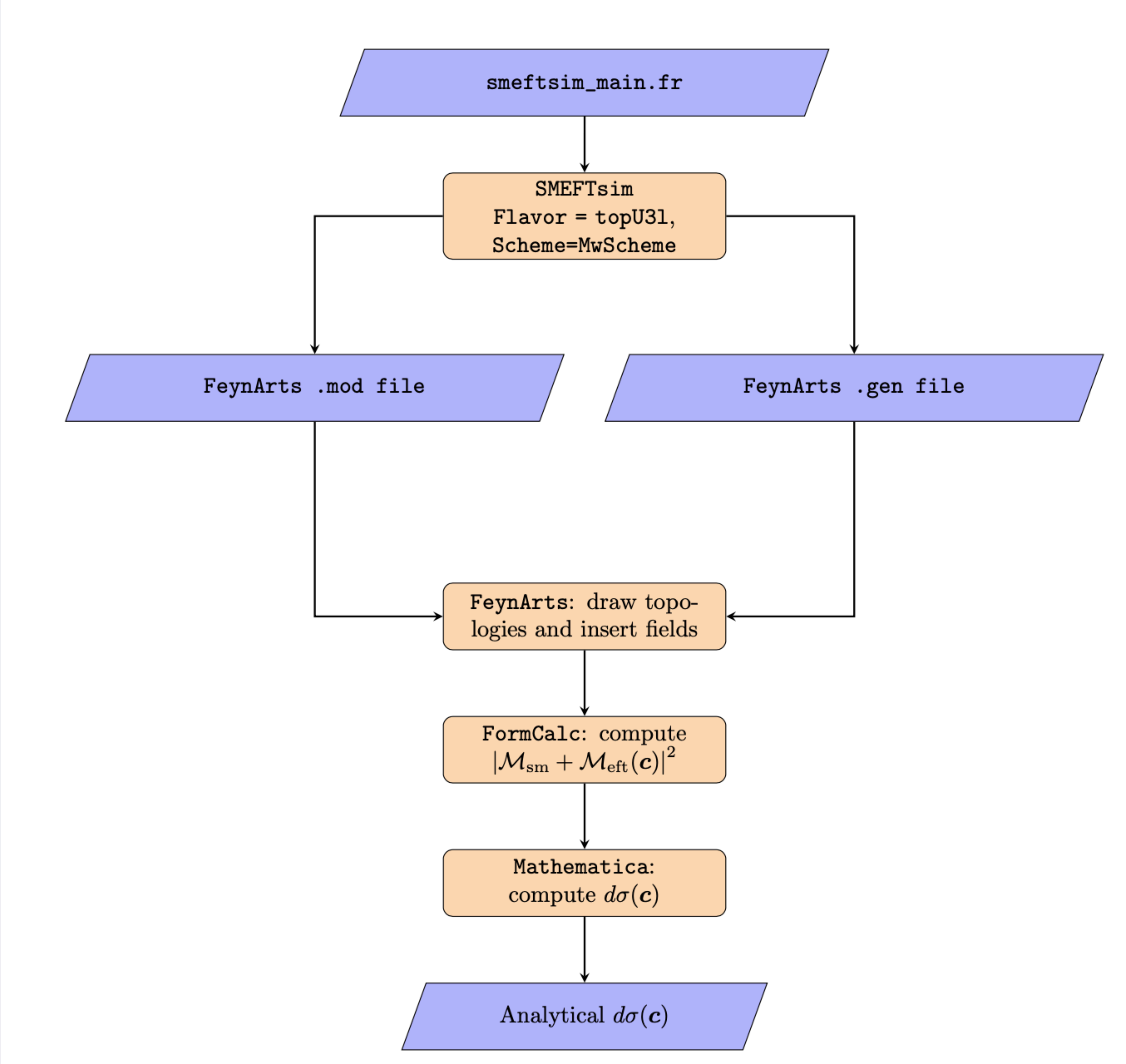
- To **validate** our methodology, we performed a study of top-quark pair production at the parton level and compared the outcome to the **analytical** result



Results



Use FormCalc to obtain analytical expressions T. Hahn [9807565]



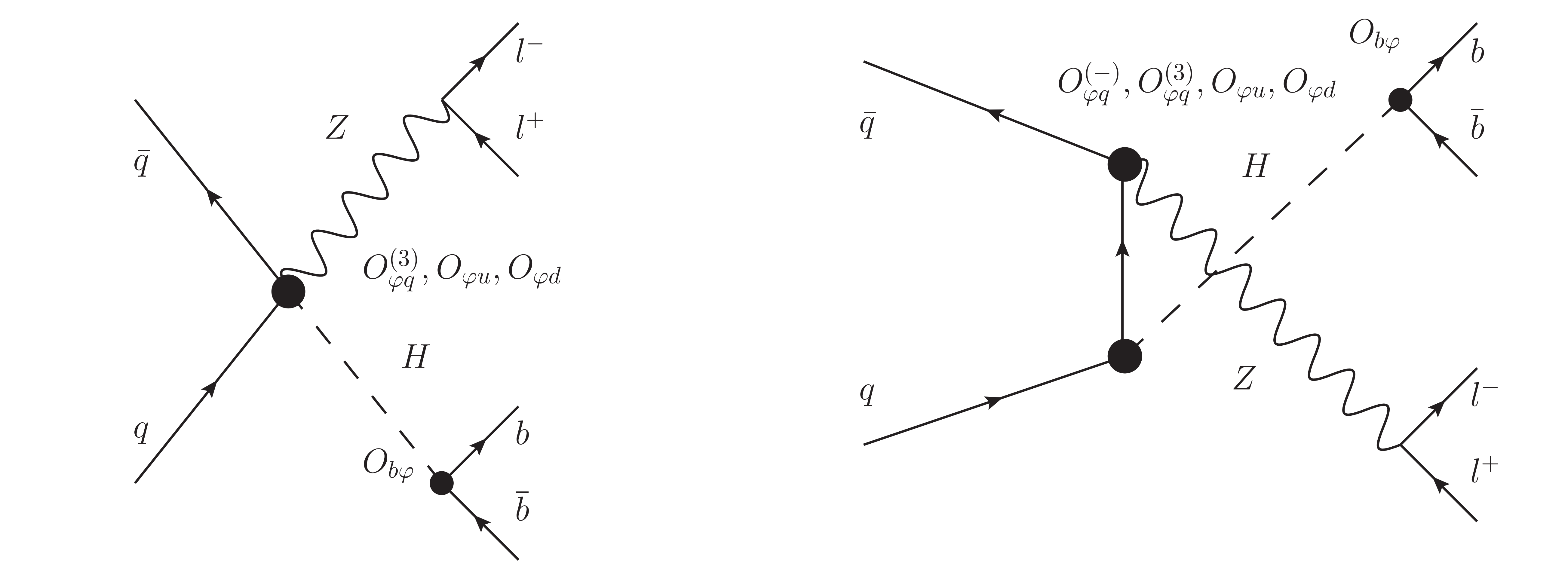
Results

- Next, we consider **Higgs associated production** and **top-quark pair production** (already included in the **SMEFiT global analysis)**
- We make sure to use the same operator basis, flavour assumptions as in SMEFiT to allow for **integration**

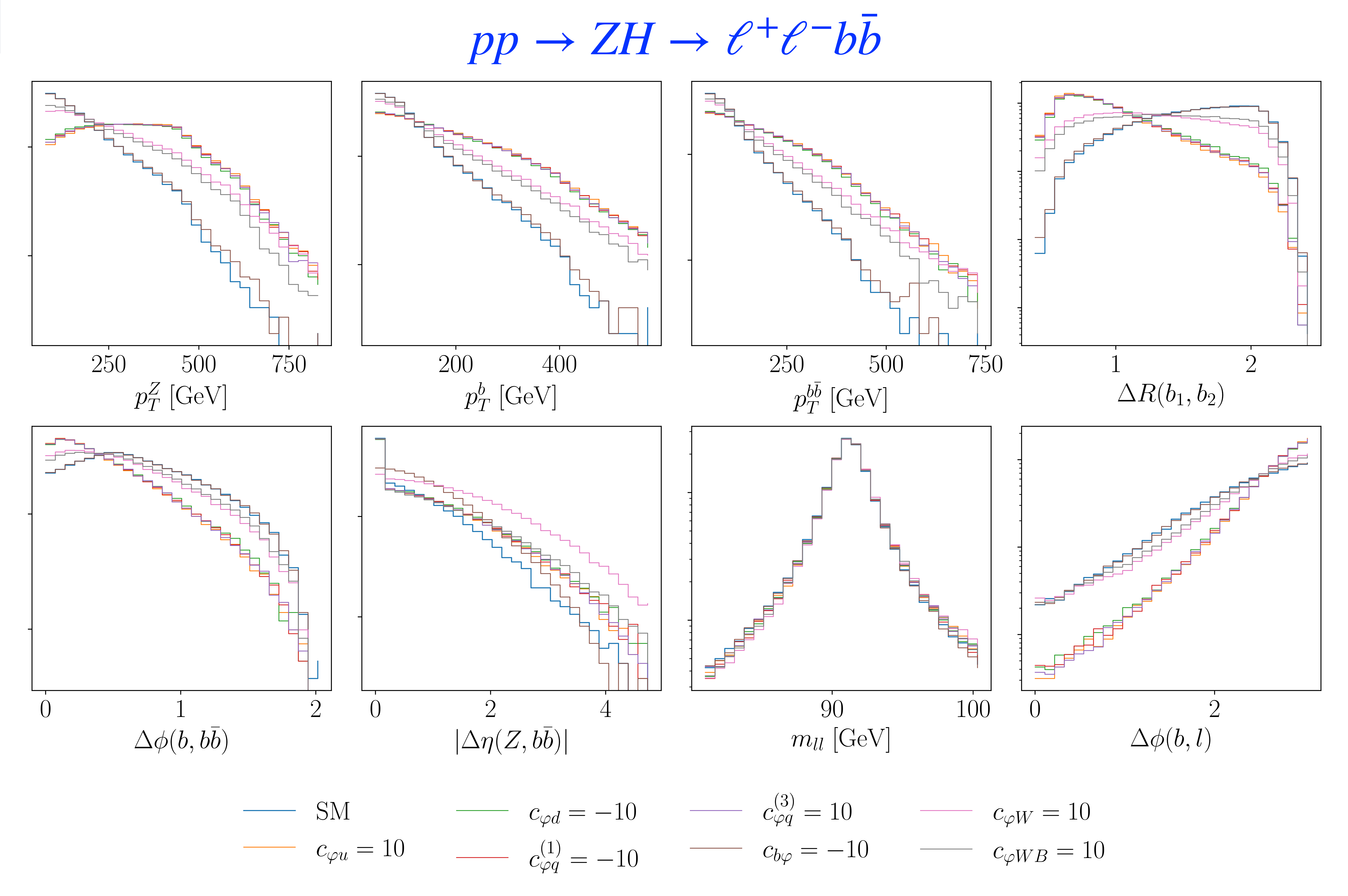
operator	SMEFTsim	SMEFiT	Definition
$\mathcal{O}_{\varphi u}$	cHu	cpui	$\sum_{i=1,2}(\varphi^\dagger i D_\mu \varphi)(\bar{u}_i\gamma^\mu u_i)$
$\mathcal{O}_{\varphi d}$	cHd	cpdi	$\sum_{i=1,2}(\varphi^\dagger i D_\mu \varphi)(\bar{d}_i\gamma^\mu d_i)$
$\mathcal{O}_{\varphi q}^{(1)}$	cHj1	–	$\sum_{i=1,2} i(\varphi^\dagger \overleftrightarrow{D}_\mu \varphi)(\bar{q}_i\gamma^\mu q_i)$
$\mathcal{O}_{\varphi q}^{(3)}$	cHj3	c3pq	$\sum_{i=1,2} i(\varphi^\dagger \overleftrightarrow{D}_\mu \tau_I \varphi)(\bar{q}_i\gamma^\mu \tau^I q_i)$
$\mathcal{O}_{\varphi q}^{(-)}$	cHj1 − cHj3	cpqMi	–
$\mathcal{O}_{b\varphi}$	cbHRe	cbp	$(\varphi^\dagger\varphi)\bar{Q}b\varphi + \text{h.c.}$
$\mathcal{O}_{\varphi W}$	cHW	cpW	$(\varphi^\dagger\varphi)W_I^{\mu\nu}W_{\mu\nu}^I$
$\mathcal{O}_{\varphi WB}$	cHWB	cpWB	$(\varphi^\dagger\tau_I\varphi)B^{\mu\nu}W_{\mu\nu}^I$

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

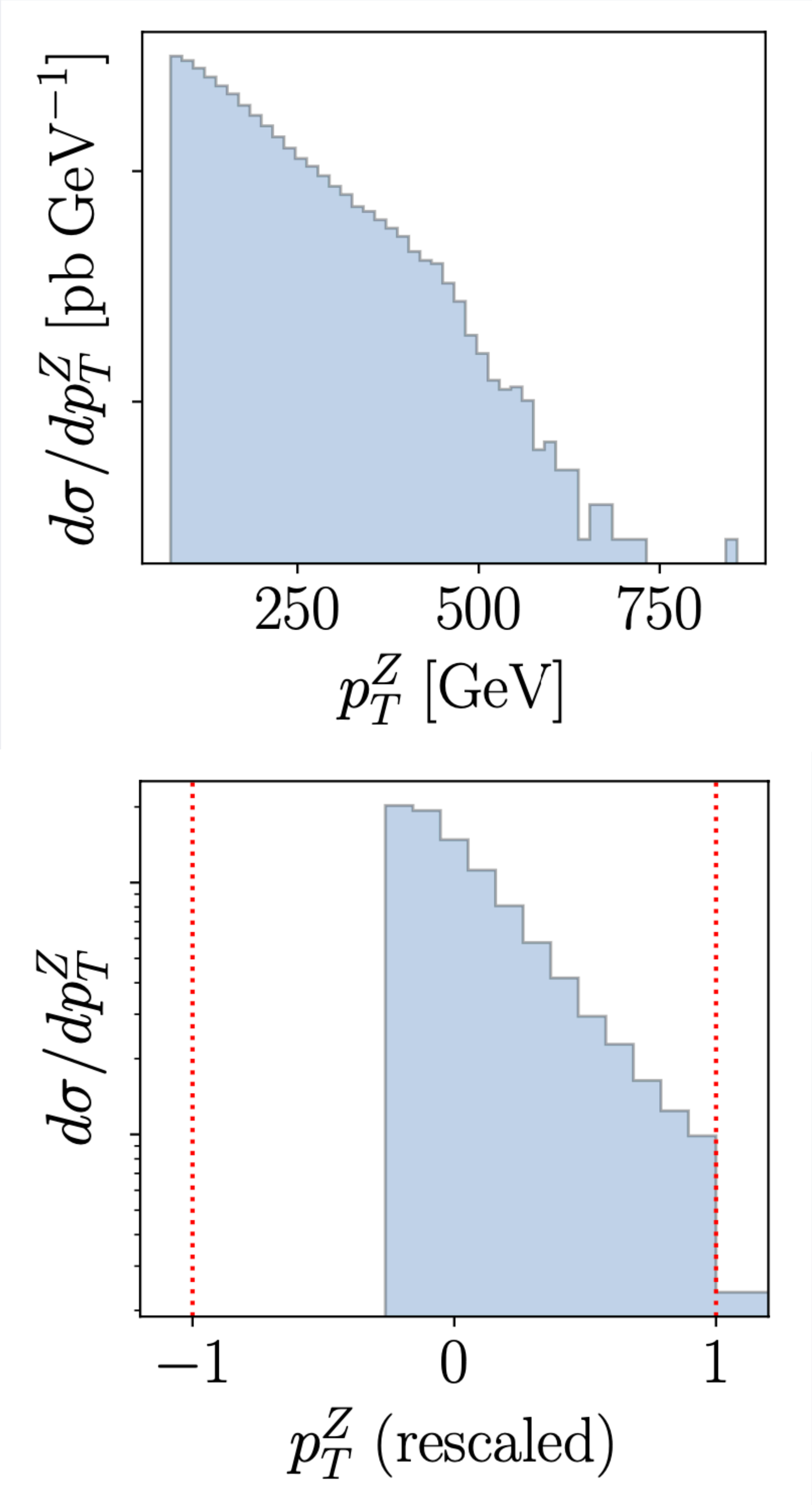
Use realistic **cuts** from arXiv:1708.03299

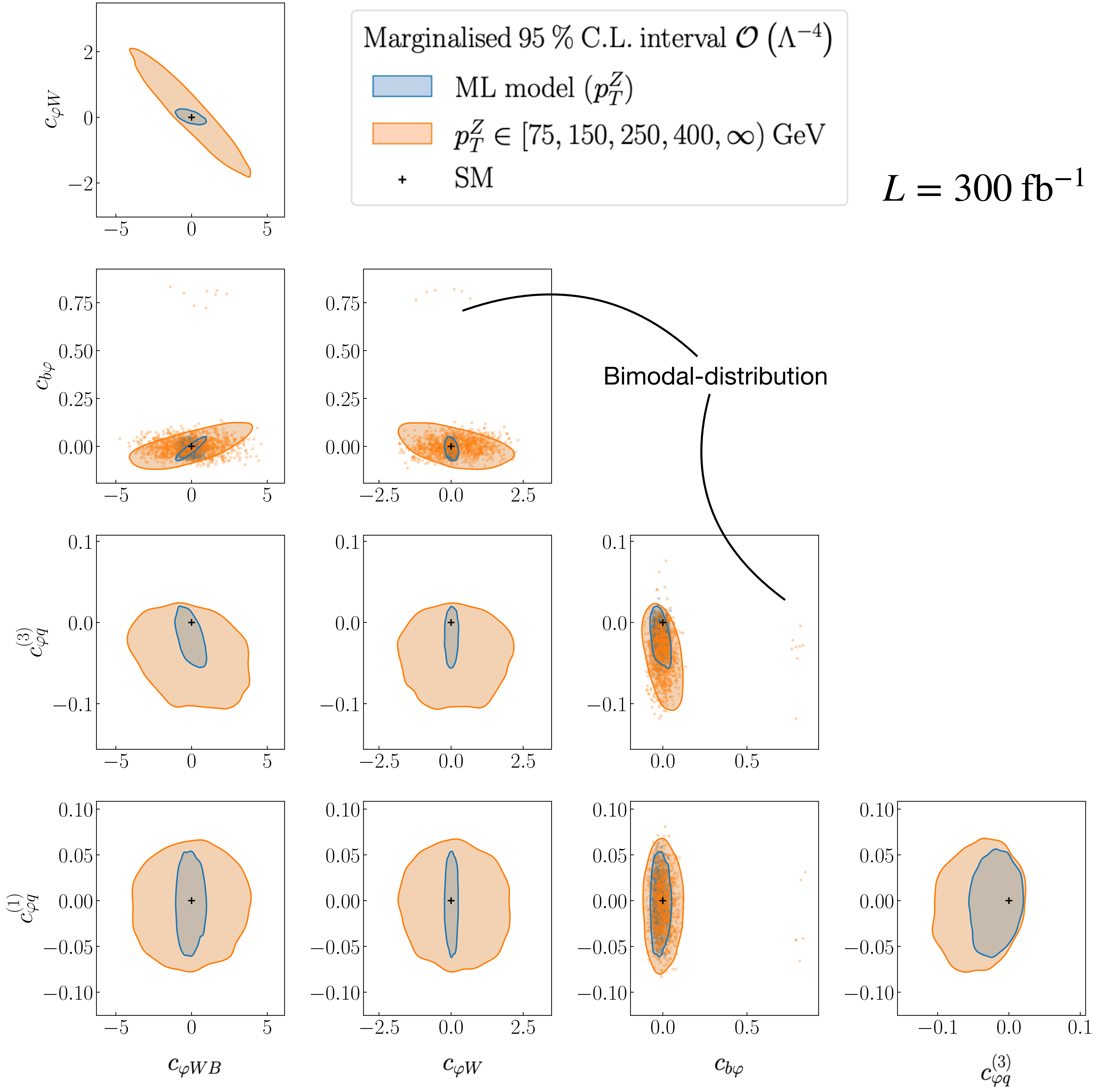


Results



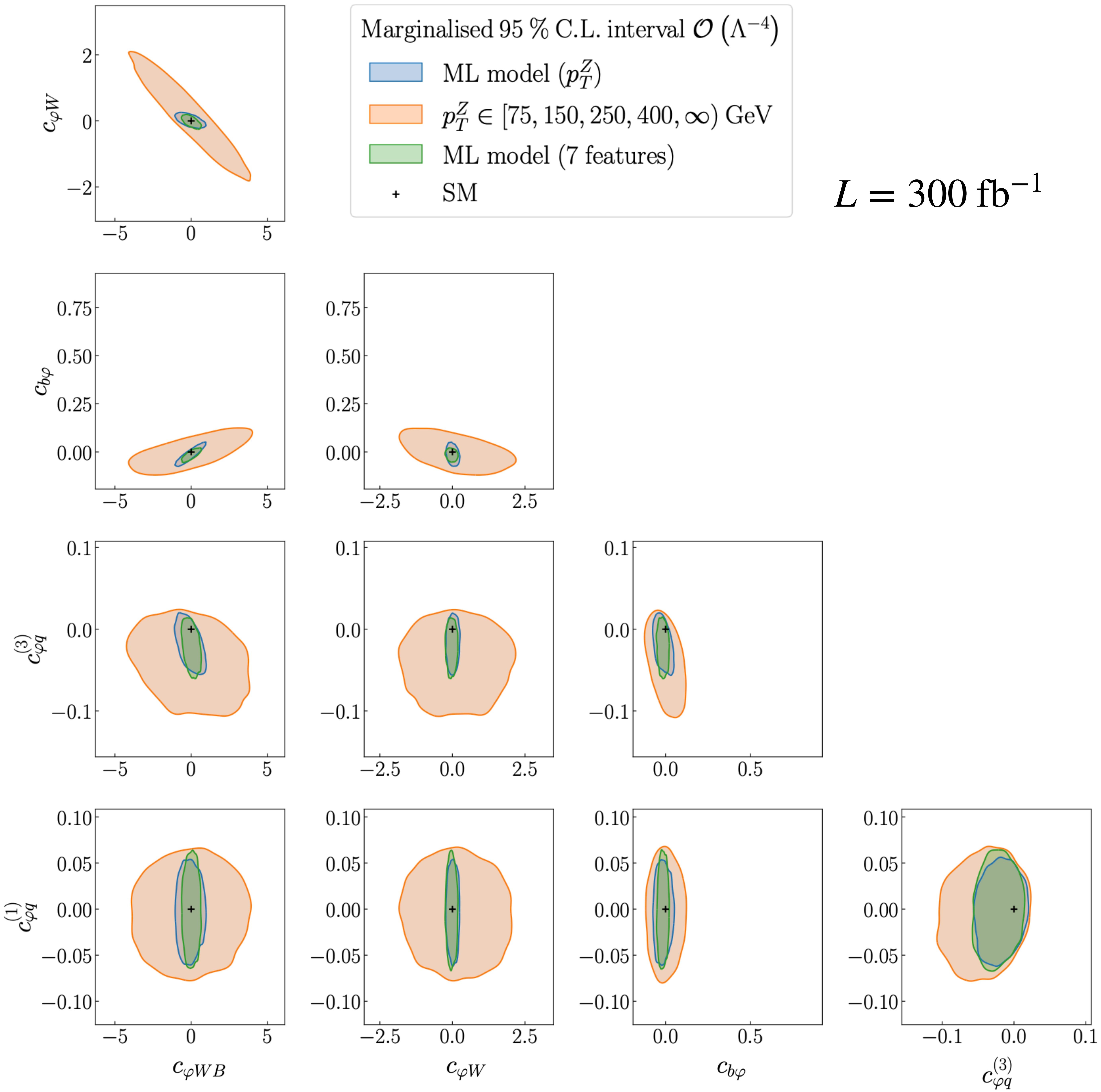
Rescale the features to the same scale





Results

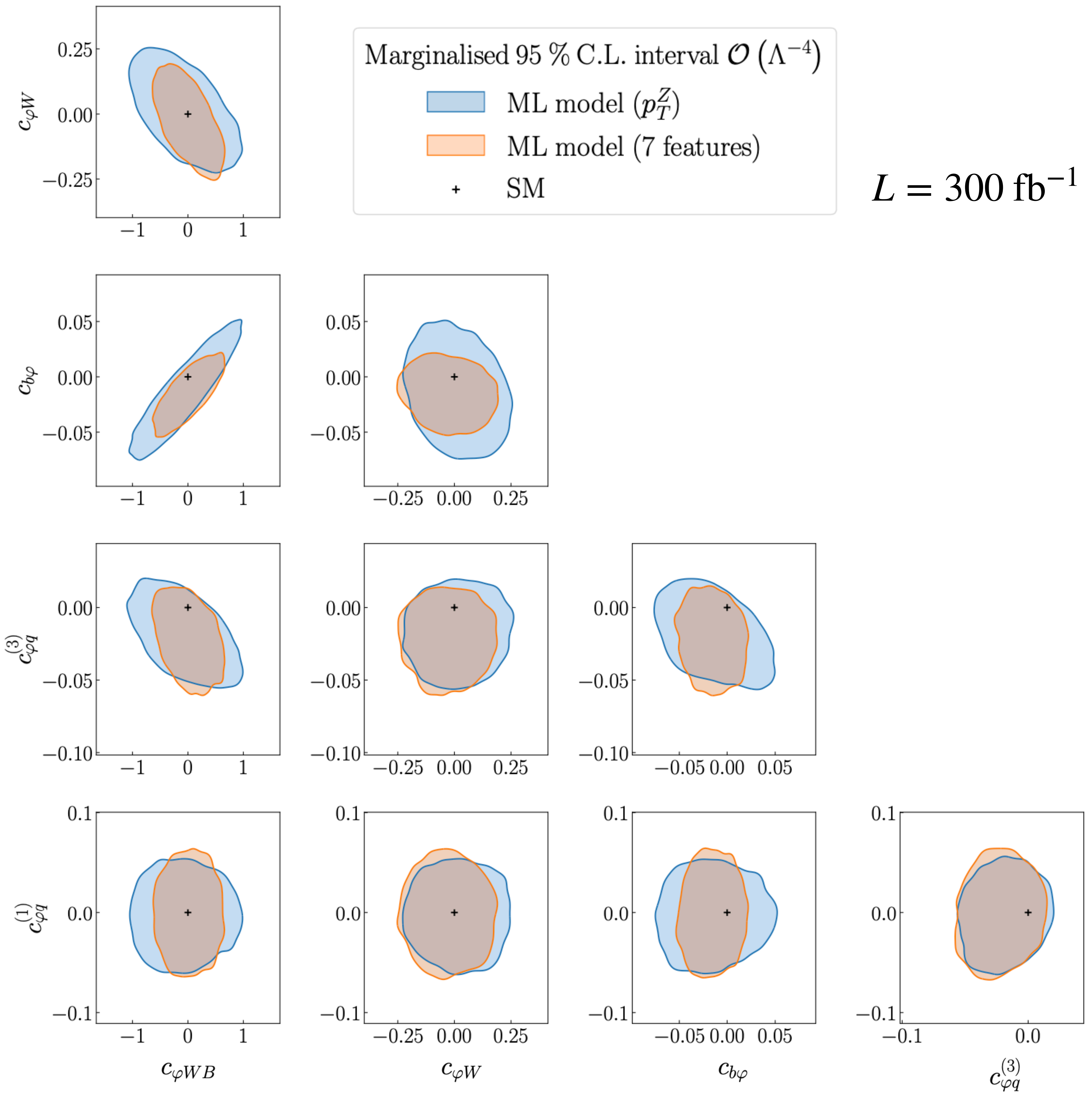
- Assess the **sensitivity** of the STXS binning in p_T^Z by comparing against the unbinned ML model trained on p_T^Z only
- This particular STXS binning is **suboptimal** for EFT studies
- The ML model provides a very **useful benchmark** for optimality studies



$L = 300\ \text{fb}^{-1}$

Results

- Enforce a lower bound on the (relative) statistical uncertainty of $1\,\%$
- Including all features leads to an **enhanced sensitivity** in most cases
- An unbinned multivariate analysis thus pays off, especially for $c_{\varphi WB}$ and $c_{b\varphi}$



Results

- Enforce a lower bound on the (relative) statistical uncertainty of $1\,\%$
- Including all features leads to an **enhanced sensitivity** in most cases
- An unbinned multivariate analysis thus pays off, especially for $c_{\varphi WB}$ and $c_{b\varphi}$

Conclusion and outlook

- The SMEFT provides a model independent framework to search for NP that fully takes into account **correlations** and **connects** new BSM phenomena

- A global fit will profit from unbinned measurements: we are extending SMEFiT based on a **ML likelihood (ratio) parameterisation**

- Our approach scales to an **arbitrary large number of EFT coefficients** and gives a faithful estimate of the **model uncertainties**

- Our tool can be used to determine the **optimal EFT sensitivity** of measurements

Conclusion and outlook

- The SMEFT provides a model independent framework to search for NP that fully takes into account **correlations** and **connects** new BSM phenomena

- A global fit will profit from unbinned measurements: we are extending SMEFiT based on a **ML likelihood (ratio) parameterisation**

- Our approach scales to an **arbitrary large number of EFT coefficients** and gives a faithful estimate of the **model uncertainties**

- Our tool can be used to determine the **optimal EFT sensitivity** of measurements

Thank you!