JHEP 03 (2023) 033

2211.02058

Statistically optimal observables for global SMEFT fits

HEFT 2023

20/06/23

Jaco ter Hoeve
VU Amsterdam & Nikhef theory group

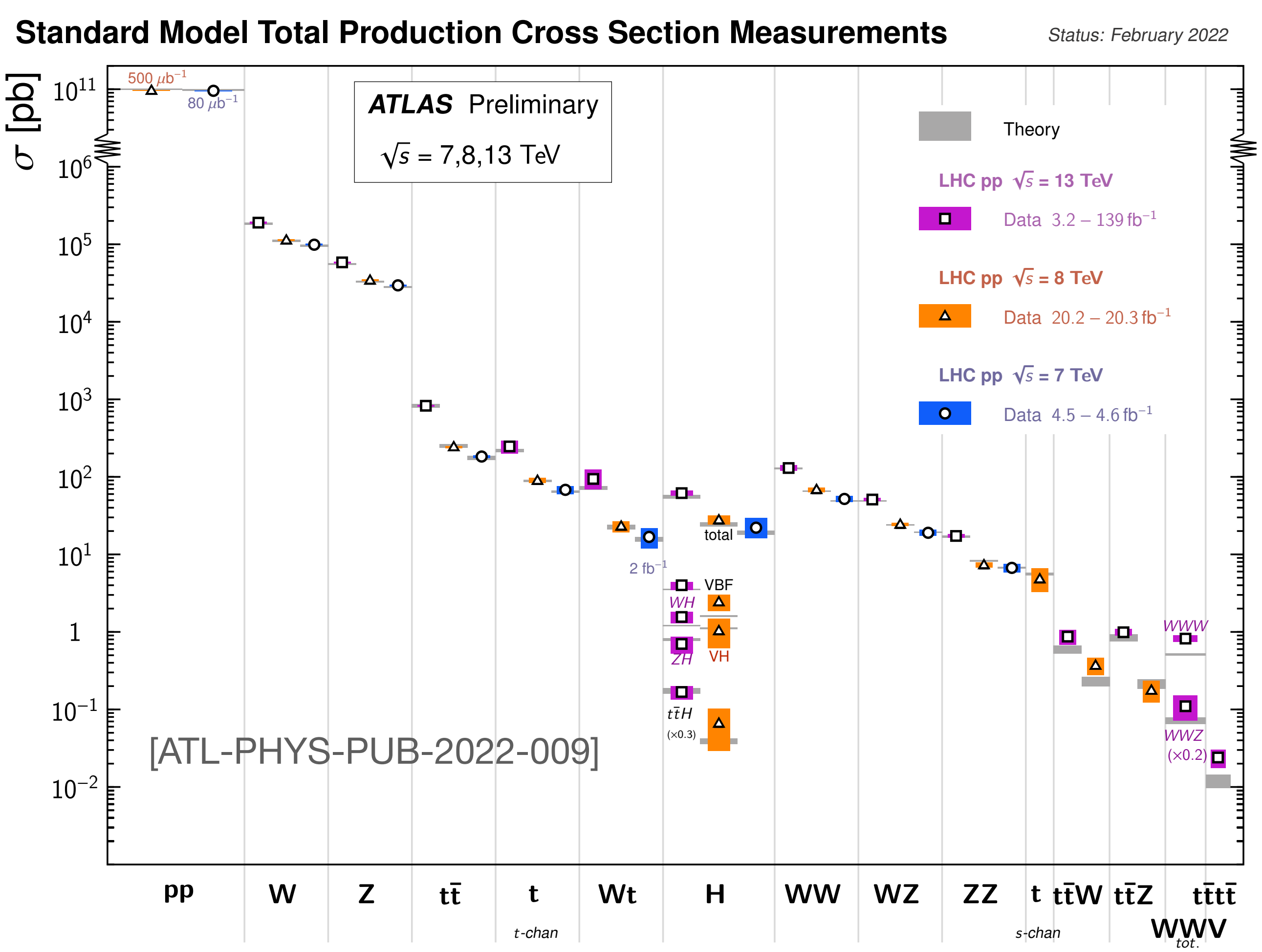
in collaboration with R. G. Ambrosio,
M. Madigan, J. Rojo and V.Sanz

VU VRIJE UNIVERSITEIT AMSTERDAM

NWO

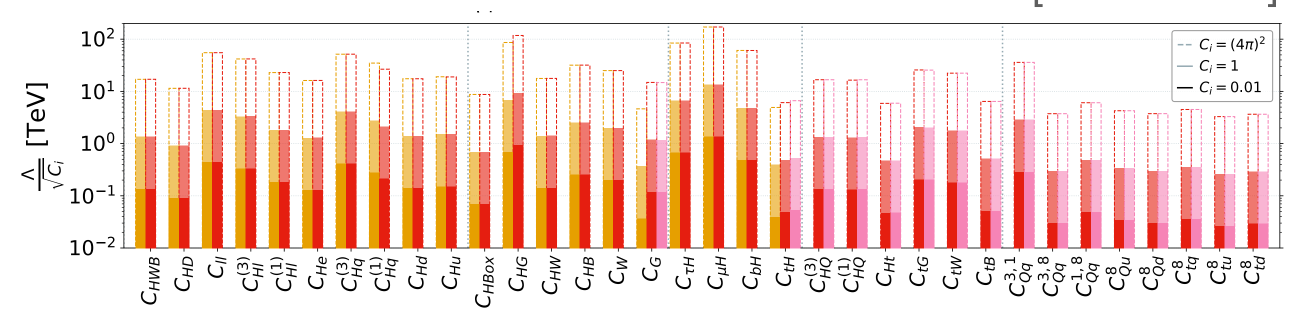
Nikhef

Traditional observables in the SMEFT

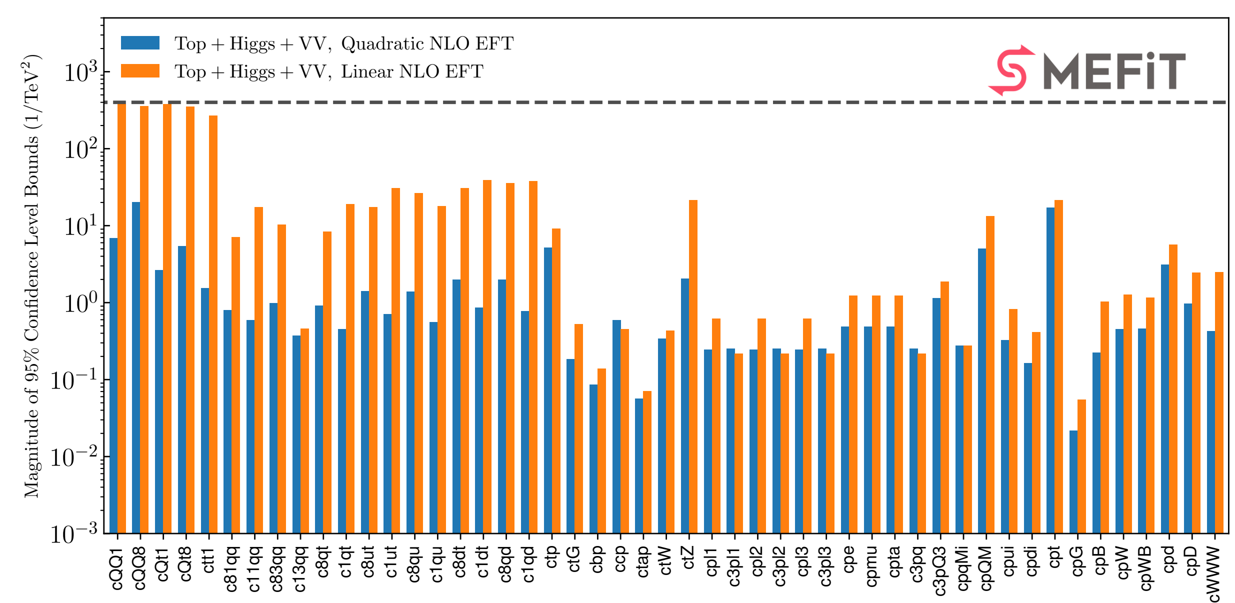


$$\chi^2 \sim \left(\sigma_i(c) - \sigma_{i,\text{exp}}\right)\left(\text{cov}^{-1}\right)_{ij}\left(\sigma_j(c) - \sigma_{j,\text{exp}}\right)$$

[2012.02779]



[2302.0660]



$$\sigma(c) = \sigma_{\text{SM}}\left(1 + \sum_i^{N_{d6}} \kappa_i c_i + \sum_{i<j}^{N_{d6}} \tilde{\kappa}_{ij} c_i \cdot c_j\right)$$

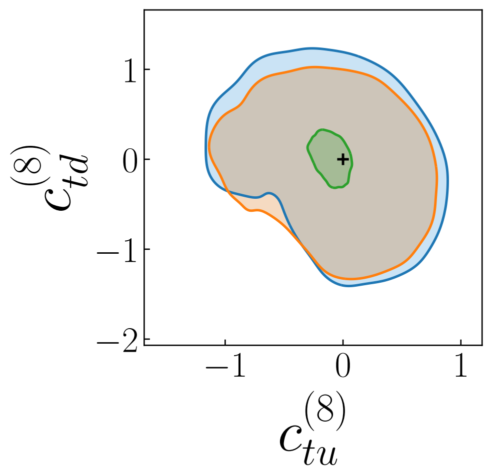
See next talk by T. Giani!

In this talk

- Global efforts **reinterpret** "SM measurements” in an EFT context
- Which measurement is the most **sensitive** to EFT parameters?
  - Inclusive, single to multi-differential (which variables)
  - Binned or unbinned, which binning?

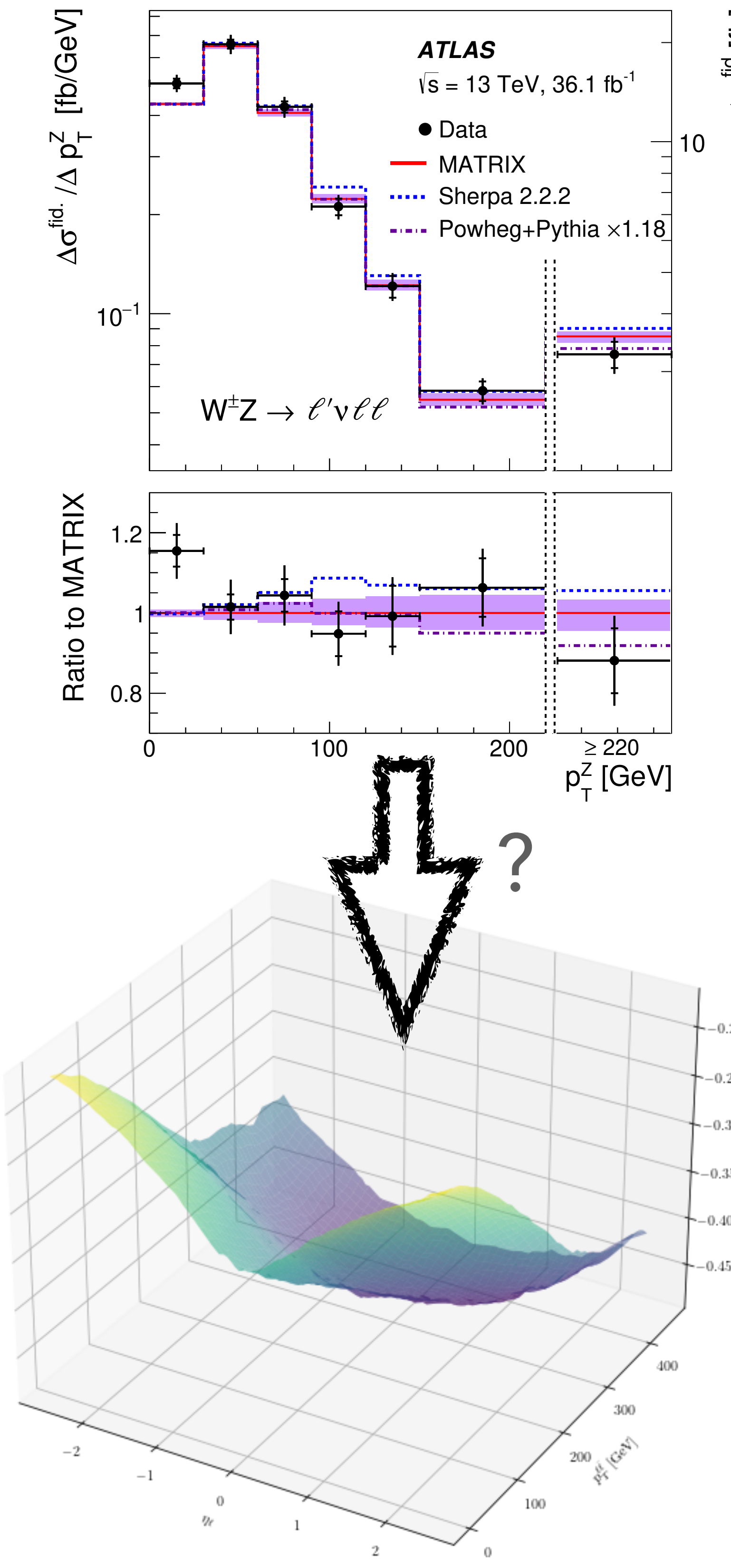
Framework needed to integrate unbinned multivariate observables into **global** SMEFT fits

- **Optimal** bounds on the EFT parameters
- Useful **diagnosis tool** to assess information loss



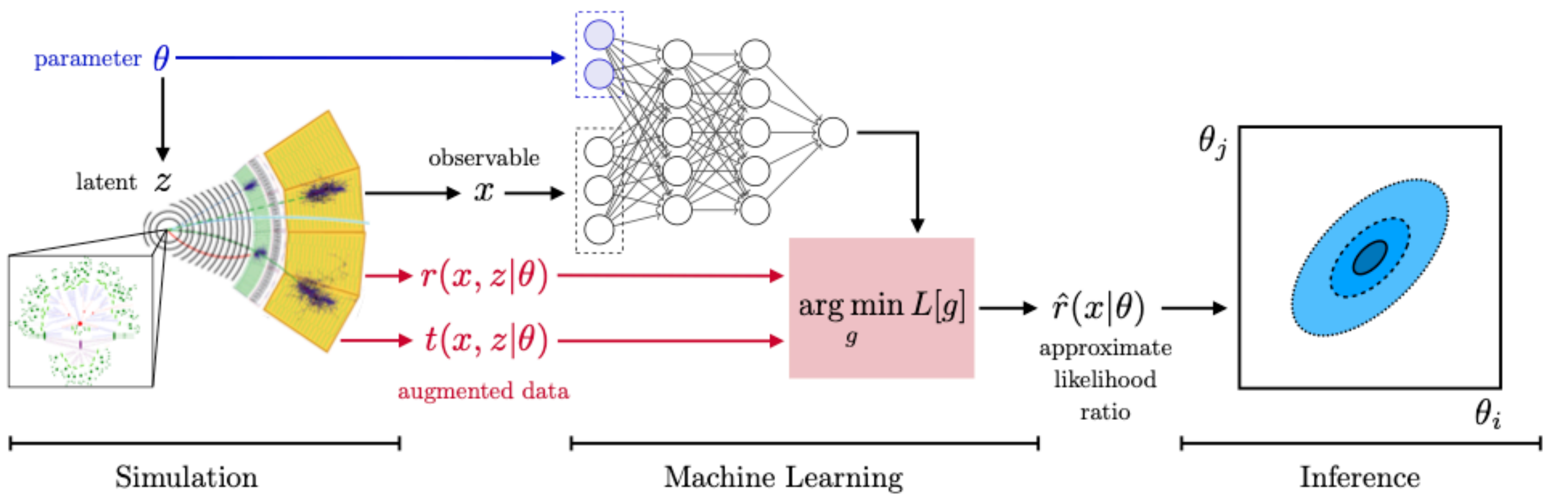
The ML4EFT framework

- Constructs unbinned multivariate observables by means of **classification and regression** techniques
- The number of NN to be trained grows quadratically with the number of EFT parameters, yet is **fully parallelisable**
- Accounts for **methodological uncertainties** by means of the Monte Carlo replica method

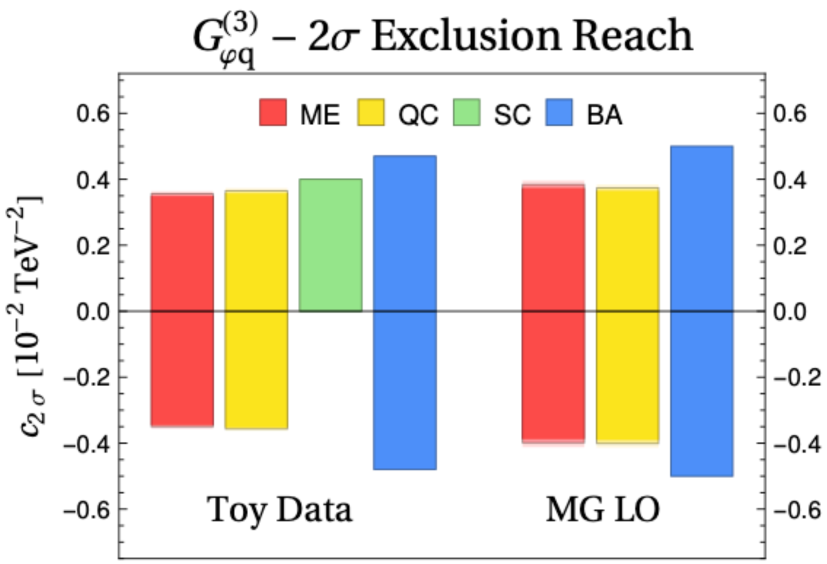


Related work

- Madminer series (J.Brehmer, K.Cranmer, G.Louppe et al.) [1907.10621, 1805.00020, …]
- Parameterized classifiers for SMEFT (A. Wulzer et al.) [2007.10356]
- Learning the EFT likelihood with tree boosting (R. Schöfbeck et al) [2205.12976]
- Back to the Formula (A. Butter, T. Plehn et al) [2109.10414]



[2010.06439]



[2007.10356]

Statistically optimal observables from ML

- Starting from two balanced datasets $\mathscr{D}_{\mathrm{SM}}$ and $\mathscr{D}_{\mathrm{EFT}}$ drawn from $f(\boldsymbol{x}\,|\,\mathrm{SM})$ and $f(\boldsymbol{x}\,|\,\mathrm{EFT})$, we minimise the cross-entropy loss

$$L[g(\boldsymbol{x})] = -\frac{1}{N}\sum_{e\in\mathscr{D}_{\mathrm{SM}}} w_e \log(1-g(\boldsymbol{x}_e)) - \frac{1}{N}\sum_{\mathscr{D}_{\mathrm{EFT}}} w_e \log g(\boldsymbol{x}_e)$$

Event weights

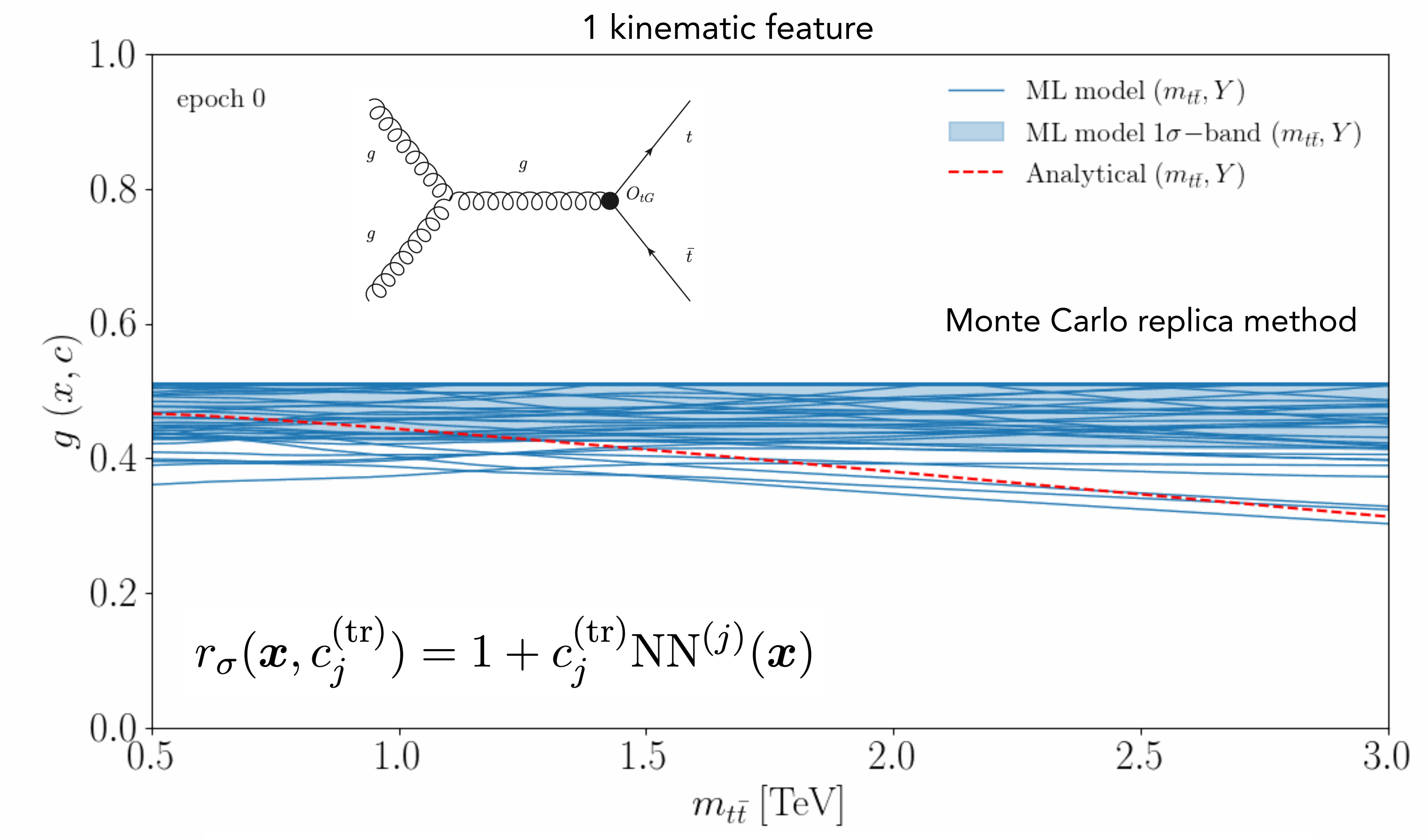
$\{m_{t\bar{t}}, \eta_l, \Delta\phi, \ldots\}$

- The learned decision boundary $g(\boldsymbol{x})$ is one-to-one with the likelihood ratio (LR) as $N \to \infty$

$$\frac{\delta L}{\delta g} = 0 \implies \hat{g}(\boldsymbol{x}) = \left(1 + \frac{f(\boldsymbol{x}\,|\,\mathrm{EFT})}{f(\boldsymbol{x}\,|\,\mathrm{SM})}\right)^{-1} \equiv \frac{1}{1+r(\boldsymbol{x})}$$

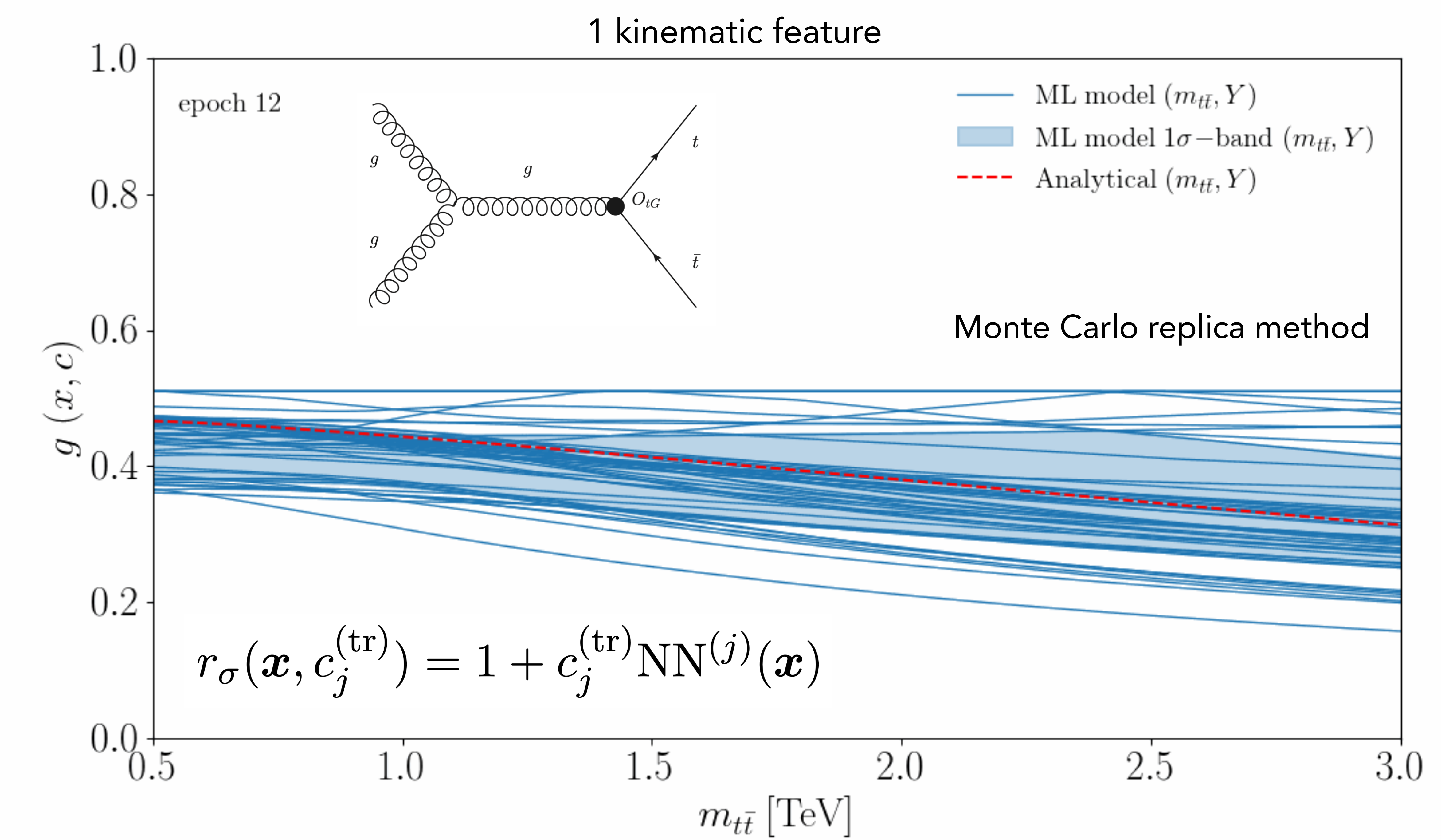
Parameterise with NNs

Learning unbinned likelihoods



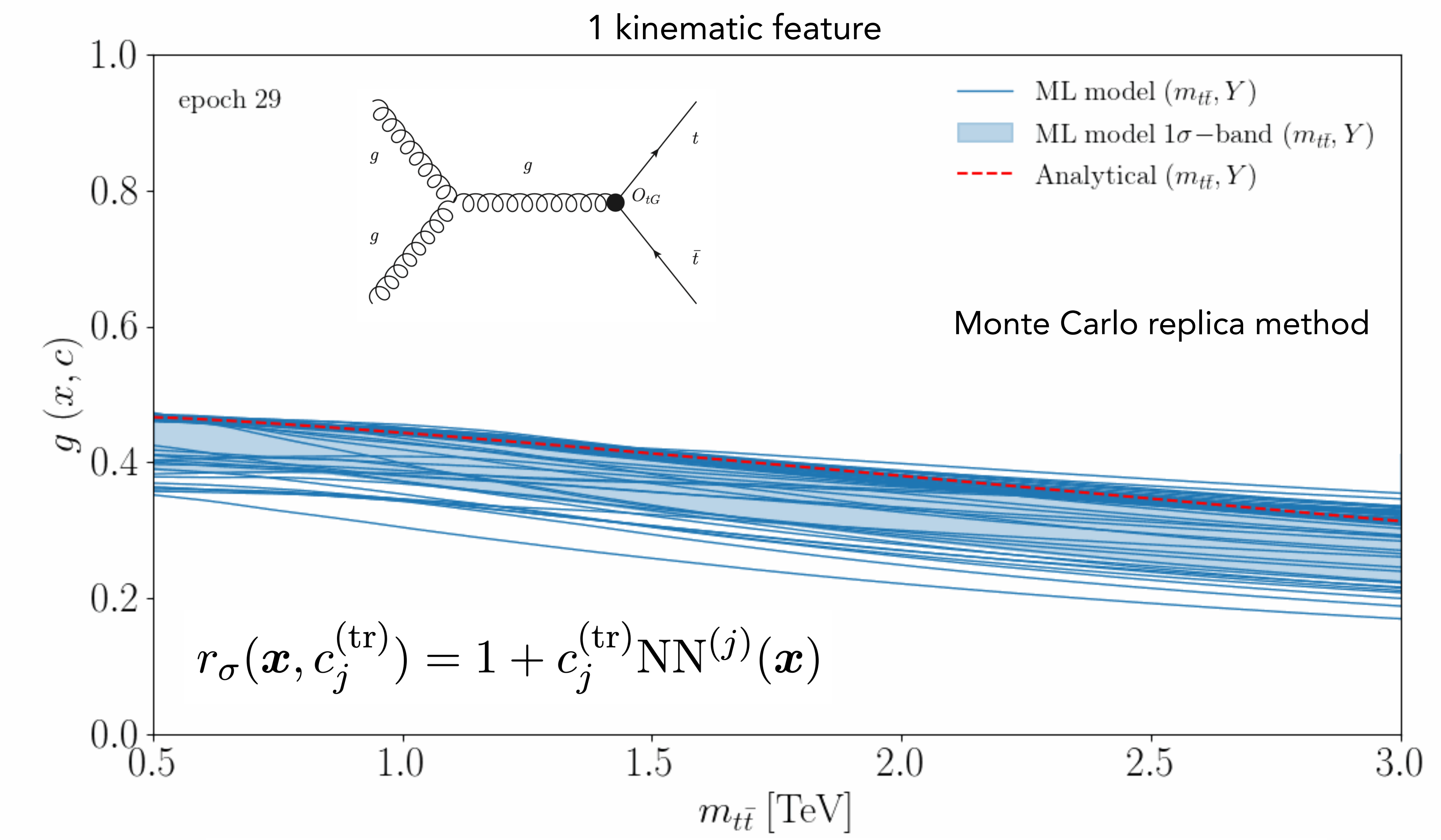
$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c}) \sum_{i=1}^{N_{\mathrm{ev}}} \log(1 - g(\boldsymbol{x}_i, \boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0}) \sum_{j=1}^{N_{\mathrm{ev}}} \log g(\boldsymbol{x}_j, \boldsymbol{c})$$

Learning unbinned likelihoods



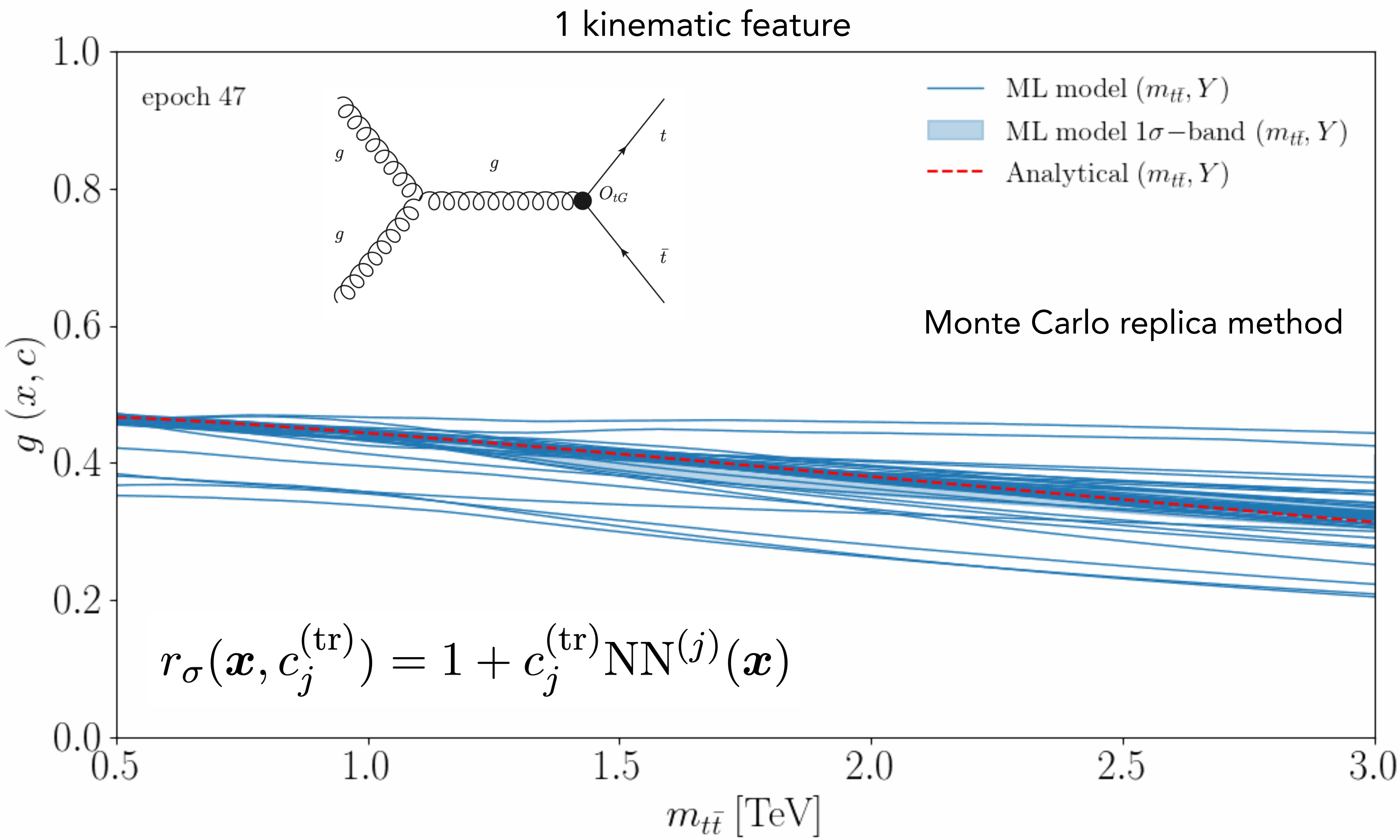
$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c}) \sum_{i=1}^{N_{\mathrm{ev}}} \log(1 - g(\boldsymbol{x}_i, \boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0}) \sum_{j=1}^{N_{\mathrm{ev}}} \log g(\boldsymbol{x}_j, \boldsymbol{c})$$

Learning unbinned likelihoods



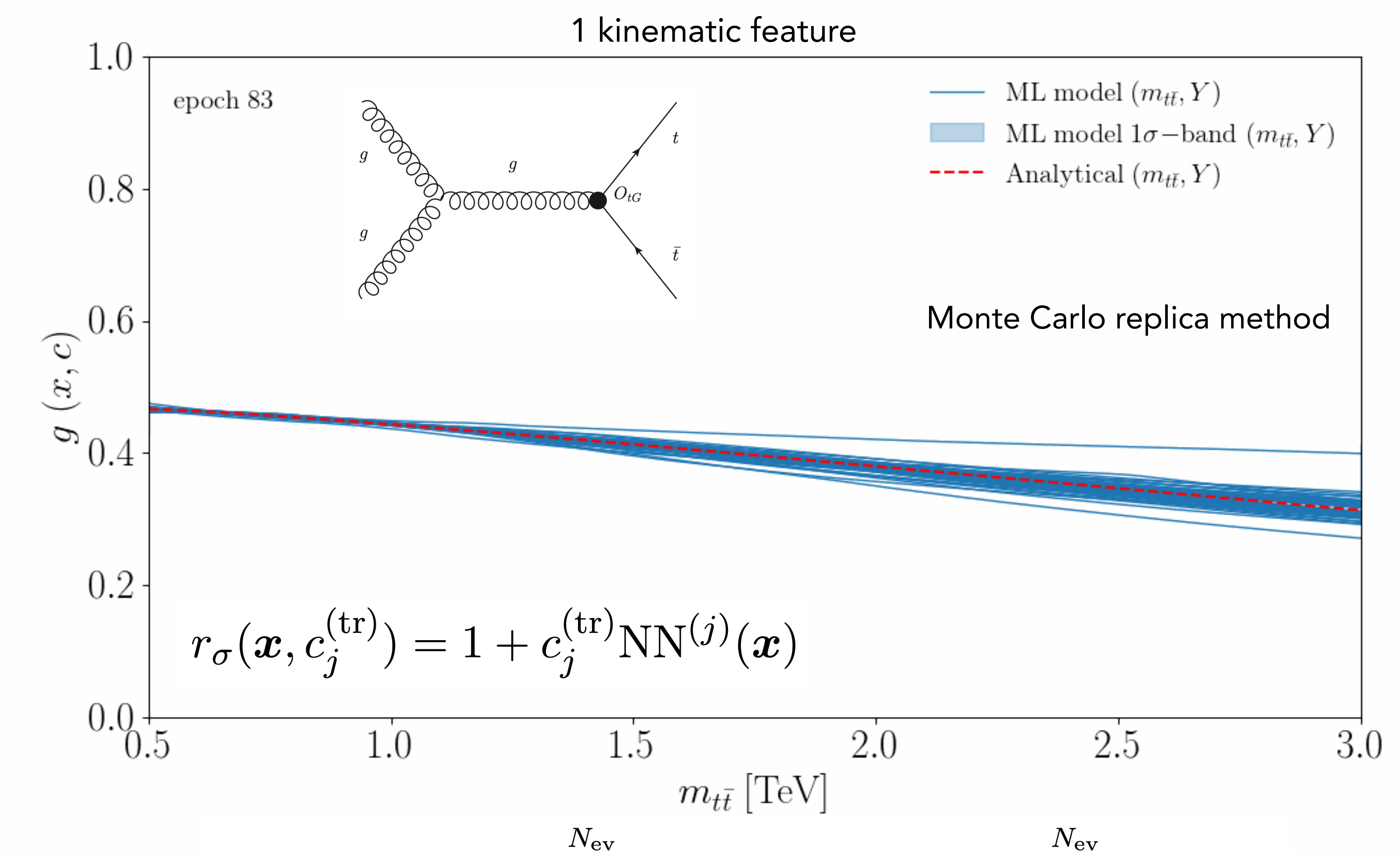
$$L[g(\boldsymbol{x},\boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c})\sum_{i=1}^{N_{\mathrm{ev}}}\log(1-g(\boldsymbol{x}_i,\boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0})\sum_{j=1}^{N_{\mathrm{ev}}}\log g(\boldsymbol{x}_j,\boldsymbol{c})$$

Learning unbinned likelihoods



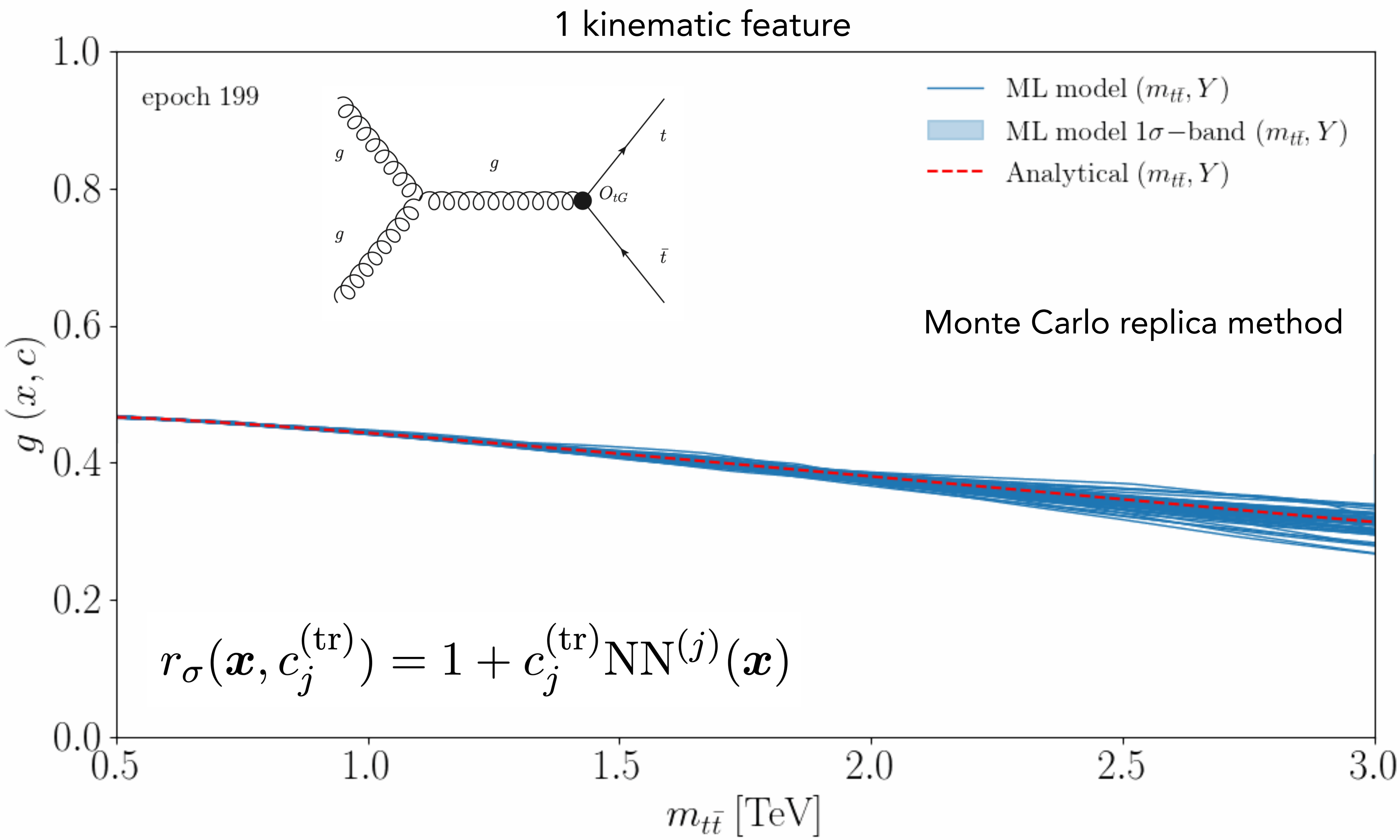
$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c}) \sum_{i=1}^{N_{\mathrm{ev}}} \log(1 - g(\boldsymbol{x}_i, \boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0}) \sum_{j=1}^{N_{\mathrm{ev}}} \log g(\boldsymbol{x}_j, \boldsymbol{c})$$

Learning unbinned likelihoods



$$L[g(\boldsymbol{x},\boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c})\sum_{i=1}^{N_{\mathrm{ev}}}\log(1-g(\boldsymbol{x}_i,\boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0})\sum_{j=1}^{N_{\mathrm{ev}}}\log g(\boldsymbol{x}_j,\boldsymbol{c})$$

Learning unbinned likelihoods

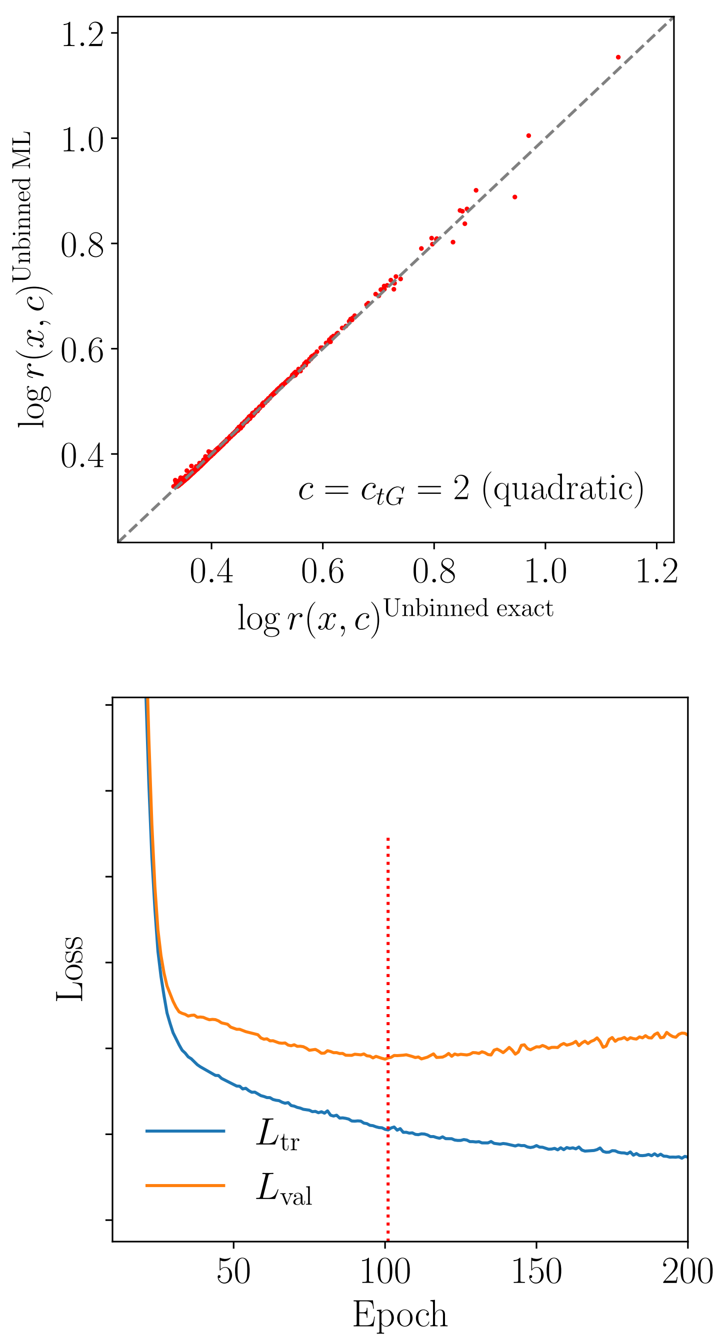


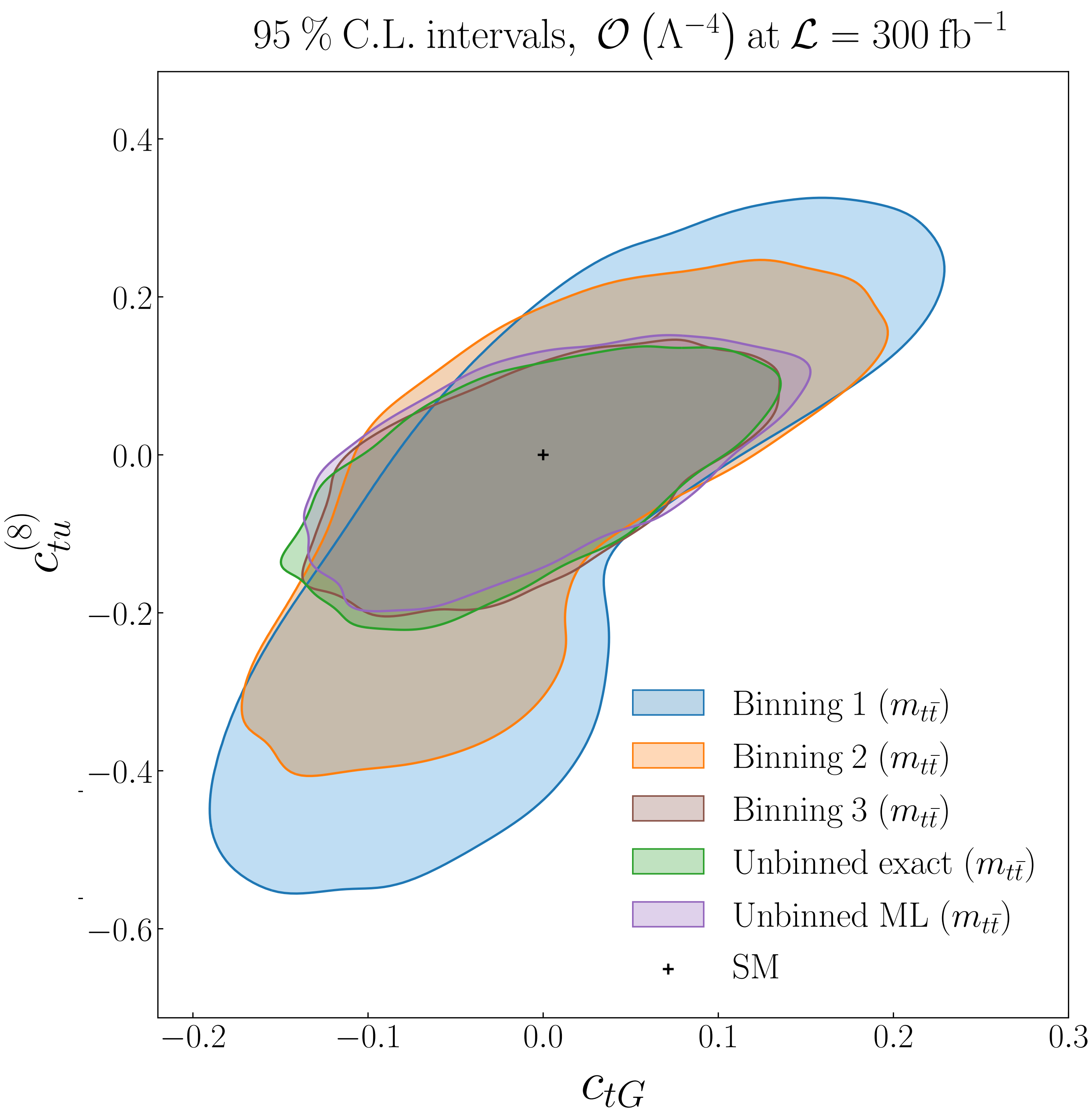
$$r_\sigma(\boldsymbol{x}, c_j^{(\mathrm{tr})}) = 1 + c_j^{(\mathrm{tr})} \mathrm{NN}^{(j)}(\boldsymbol{x})$$

$$L[g(\boldsymbol{x}, \boldsymbol{c})] = -\sigma_{\mathrm{fid}}(\boldsymbol{c}) \sum_{i=1}^{N_{\mathrm{ev}}} \log(1 - g(\boldsymbol{x}_i, \boldsymbol{c})) - \sigma_{\mathrm{fid}}(\boldsymbol{0}) \sum_{j=1}^{N_{\mathrm{ev}}} \log g(\boldsymbol{x}_j, \boldsymbol{c})$$

Learning unbinned likelihoods

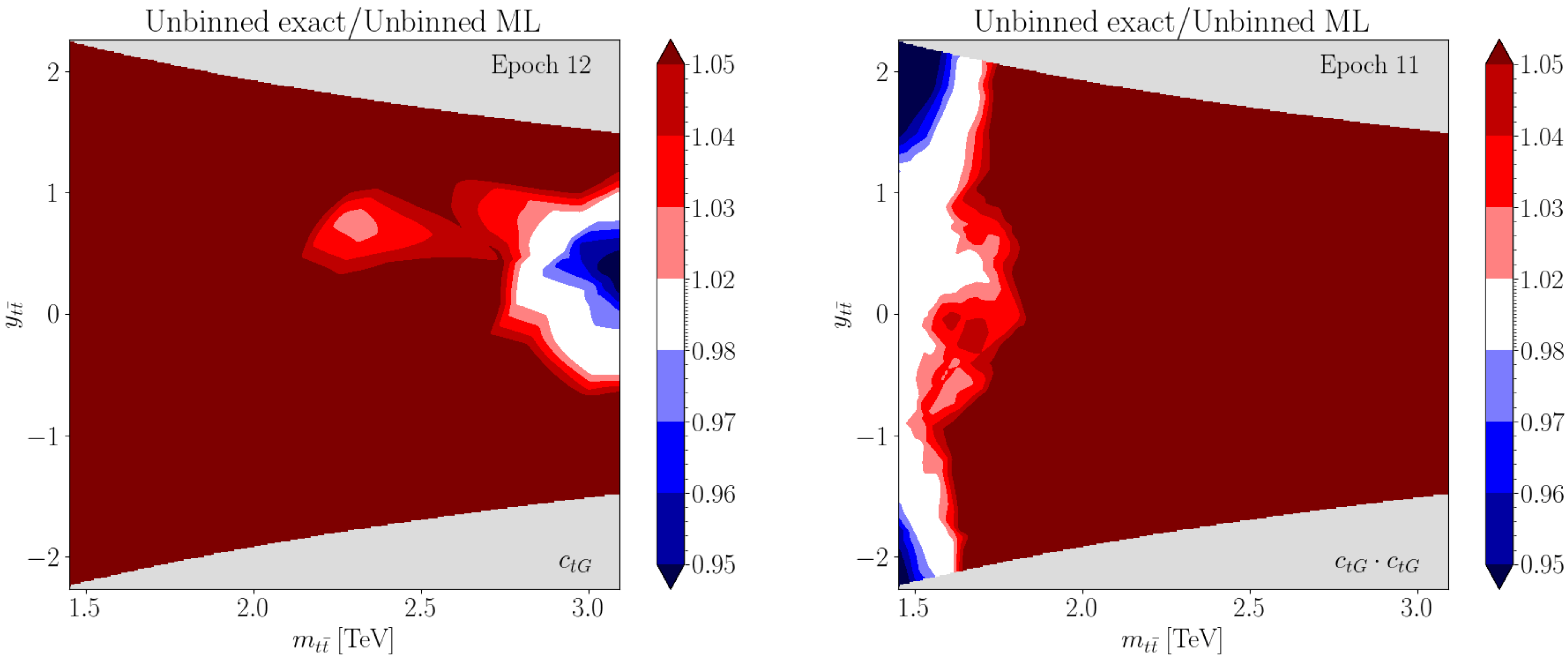
A fine binning **reproduces** the ML parameterisation





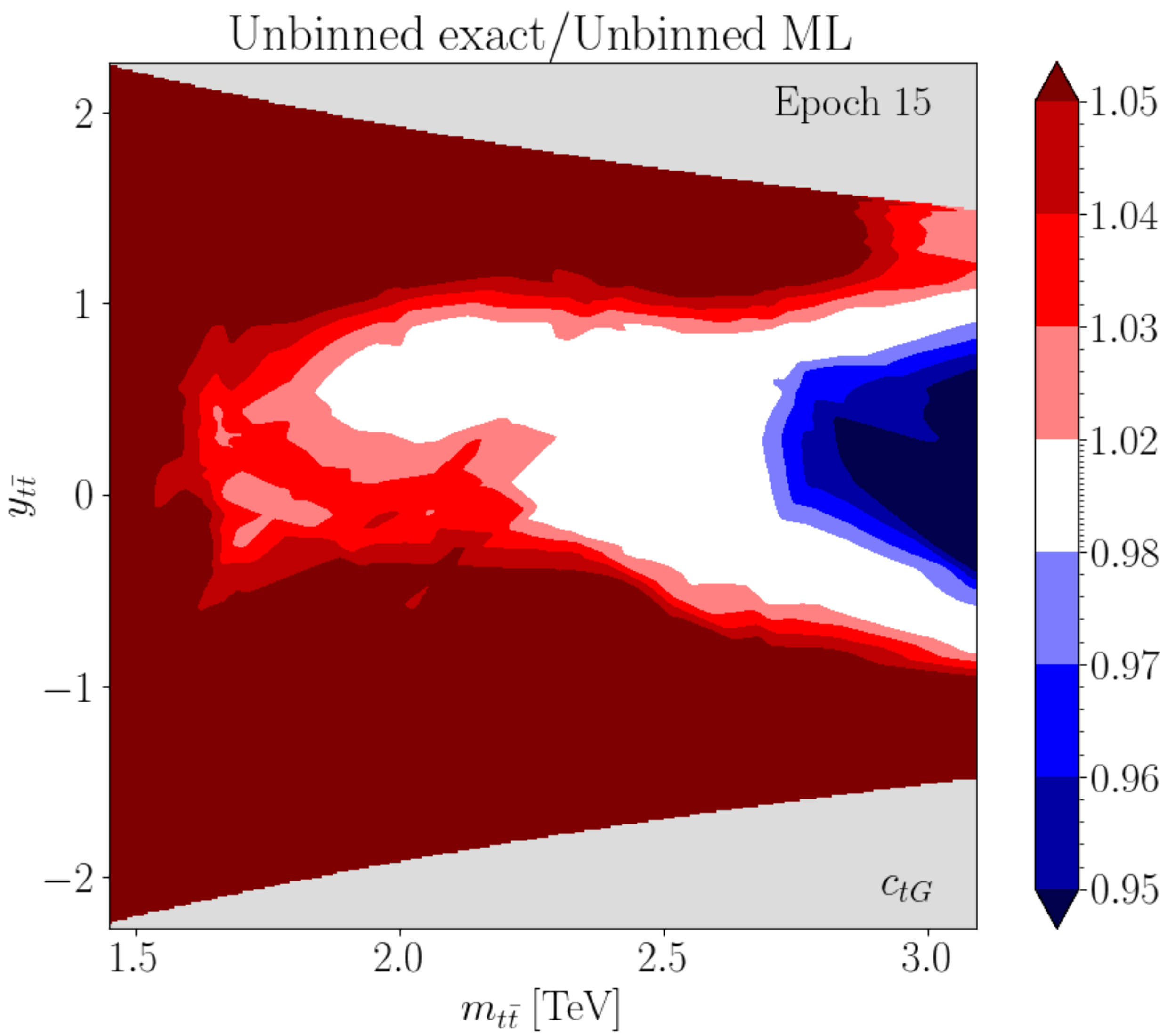
Learning unbinned likelihoods

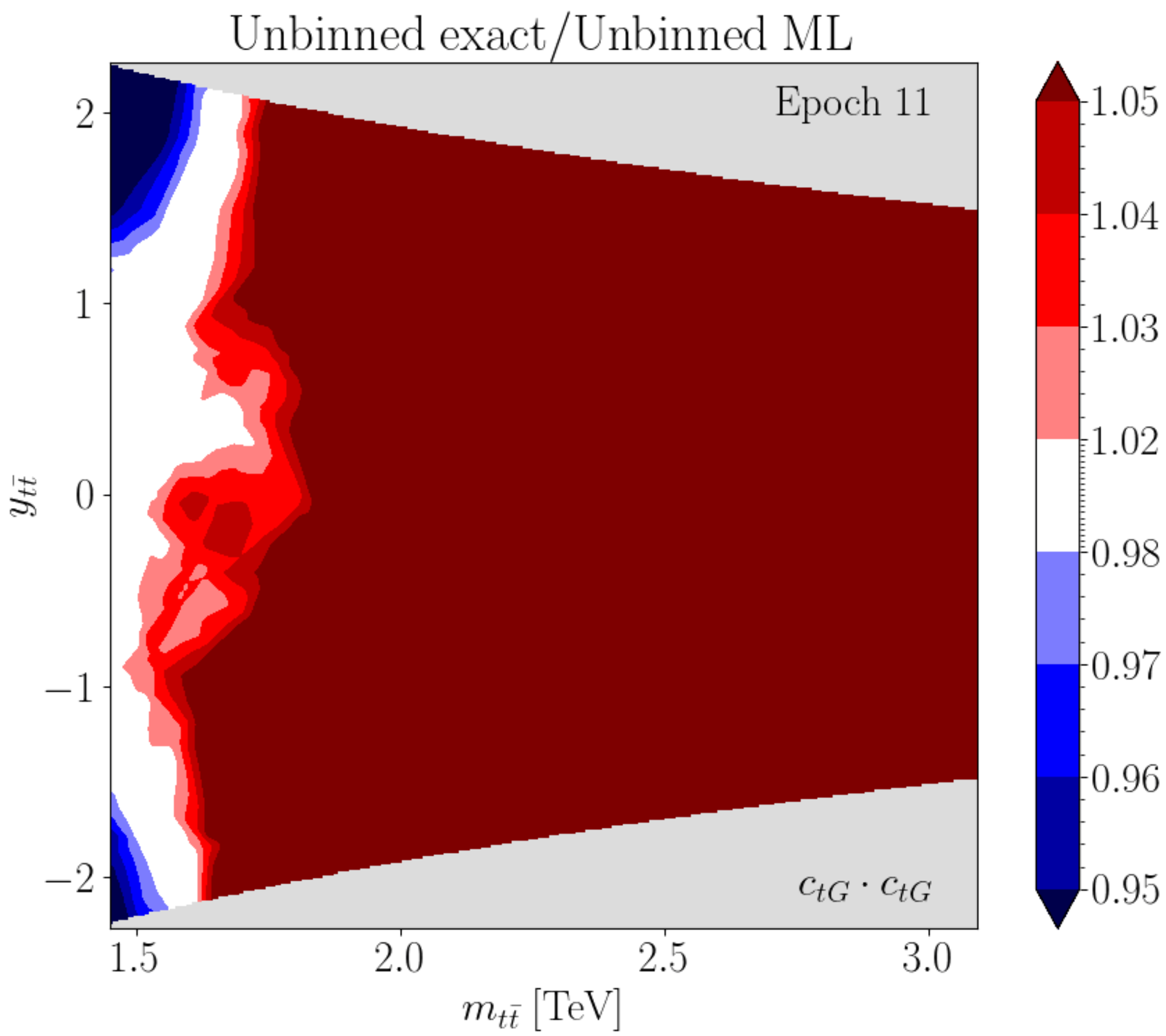
2 kinematic features



Learning unbinned likelihoods

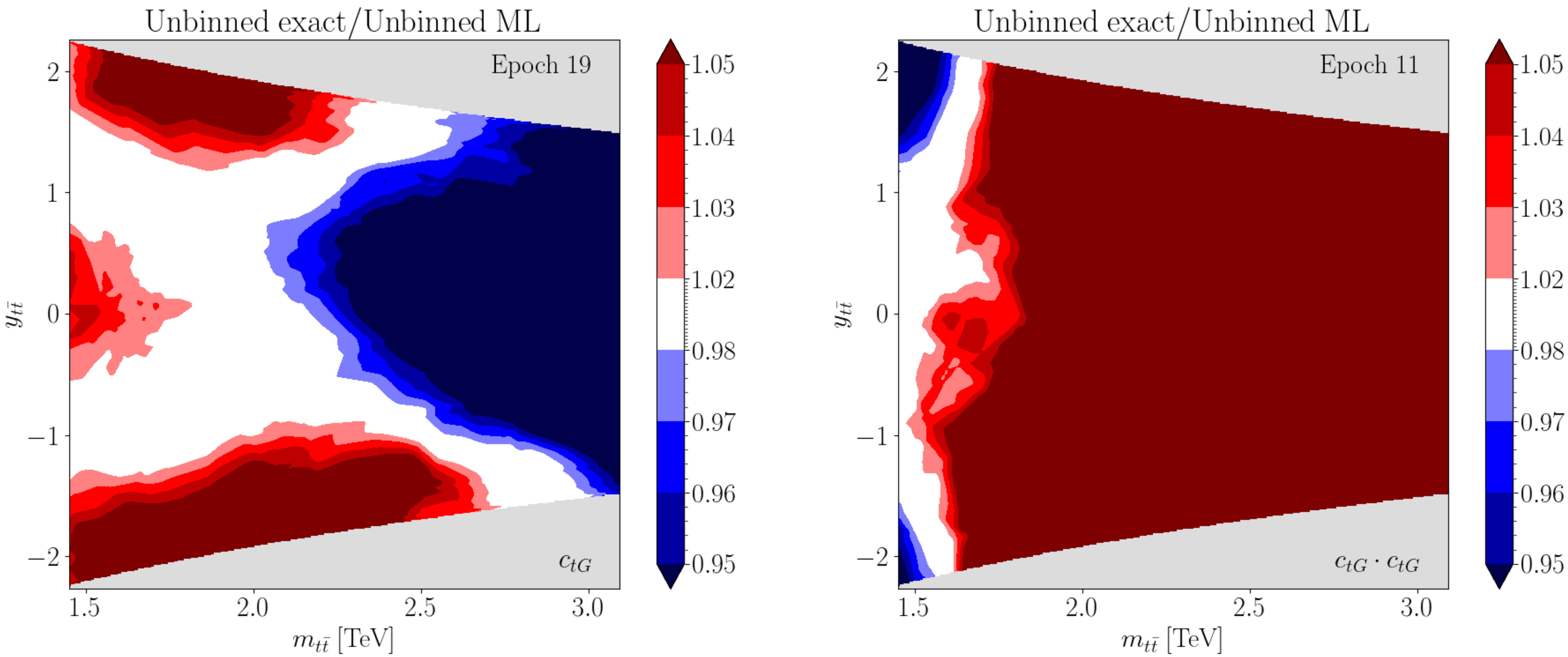
2 kinematic features





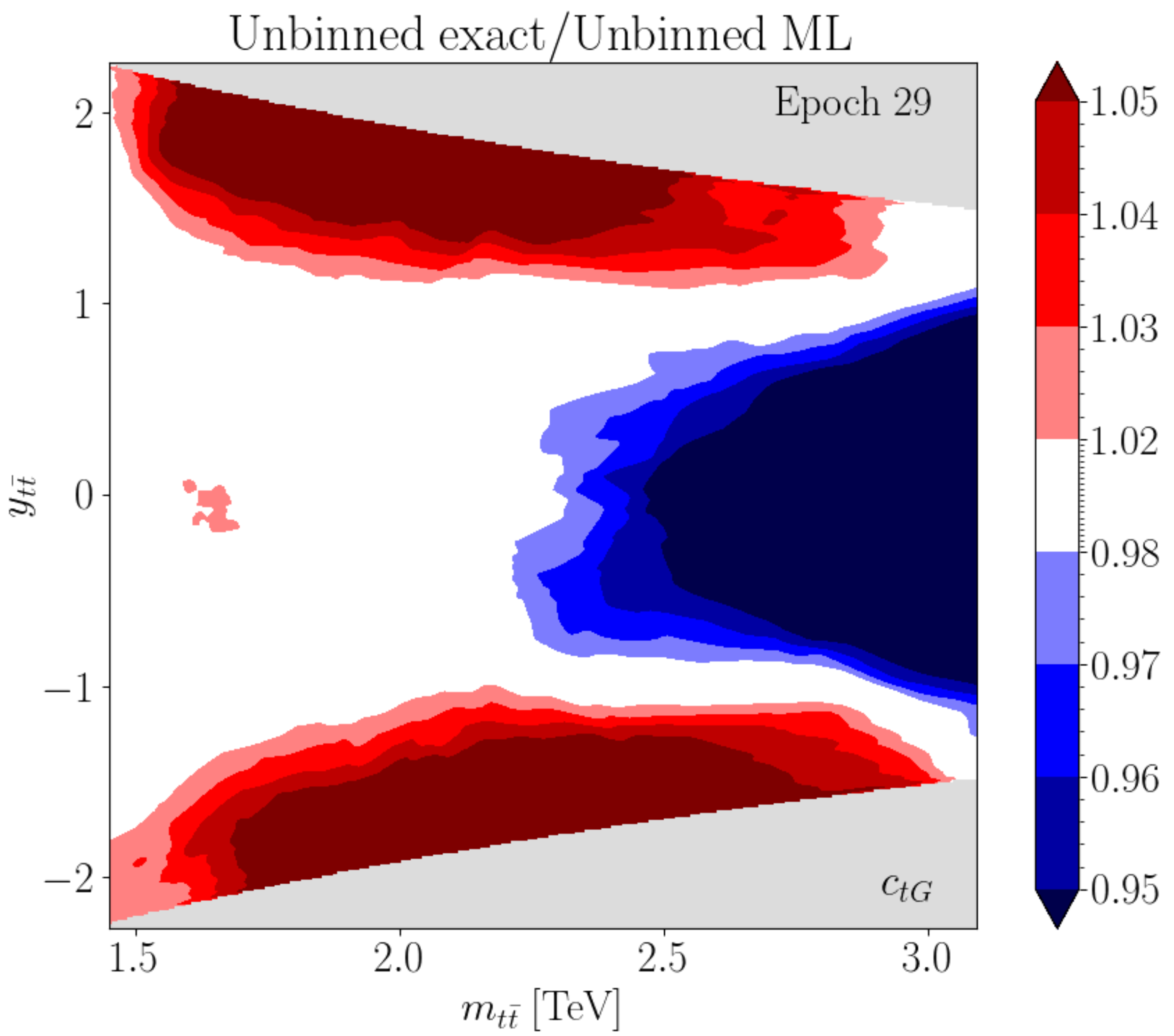
Learning unbinned likelihoods

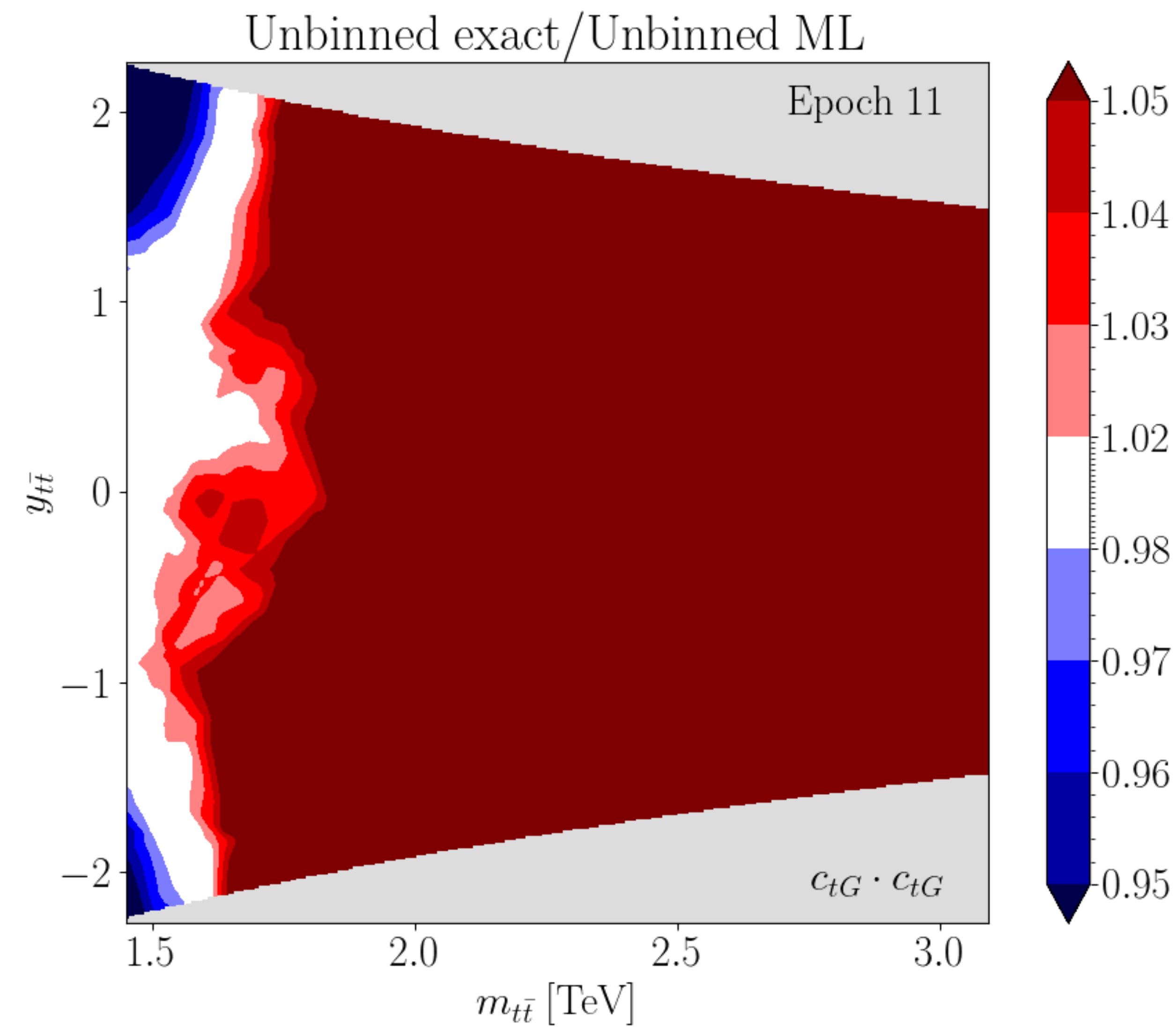
2 kinematic features



Learning unbinned likelihoods

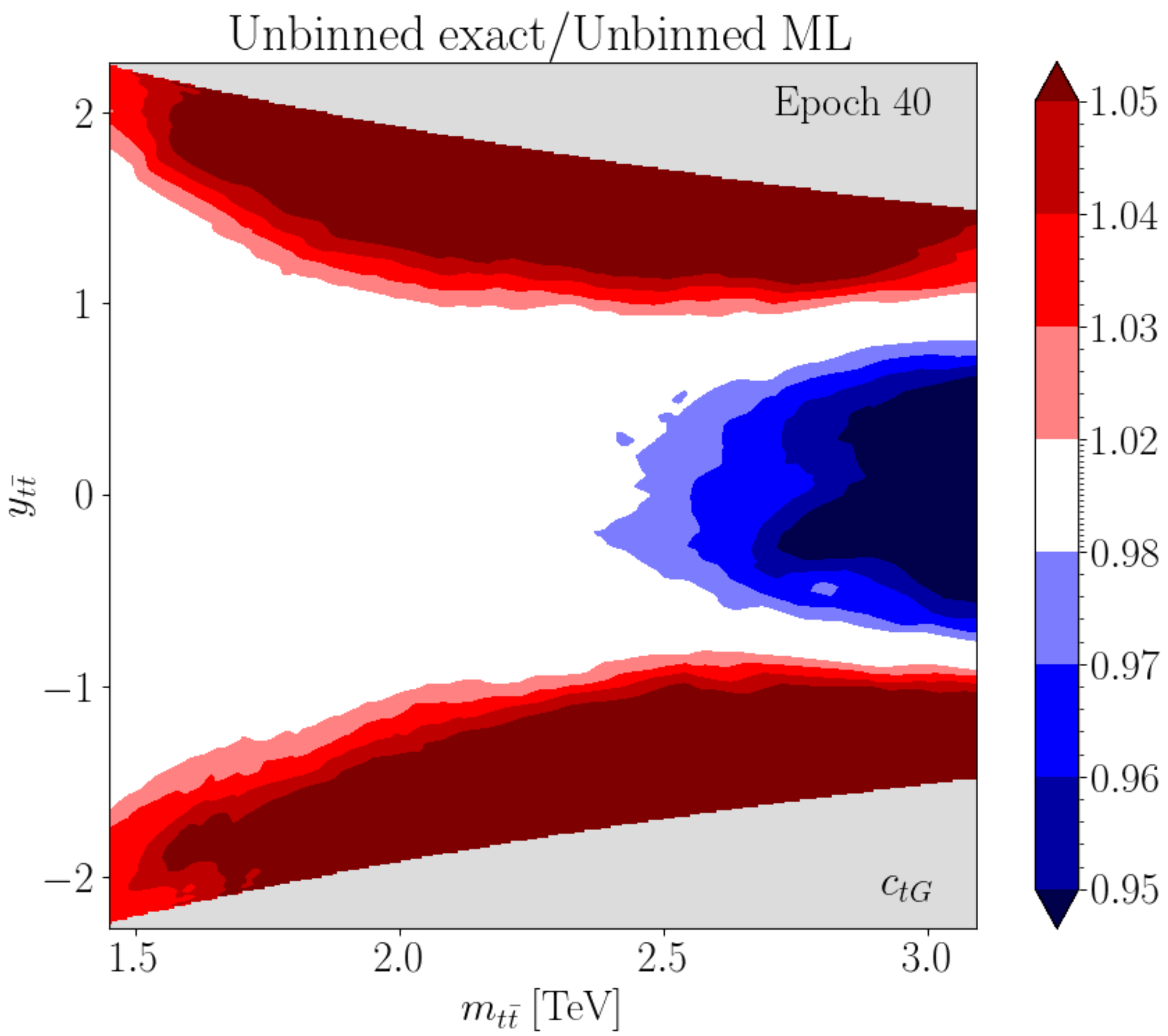
2 kinematic features

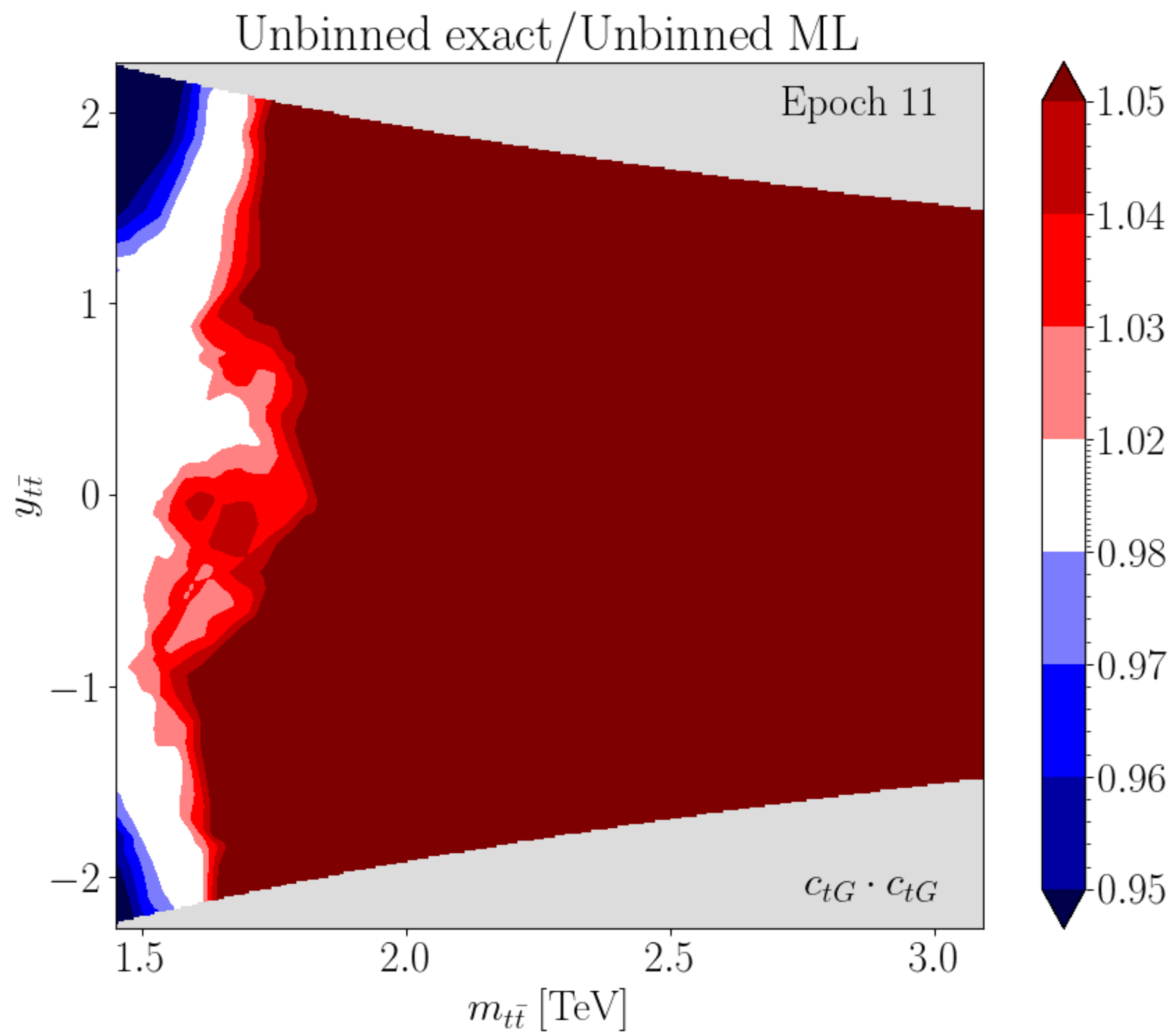




Learning unbinned likelihoods

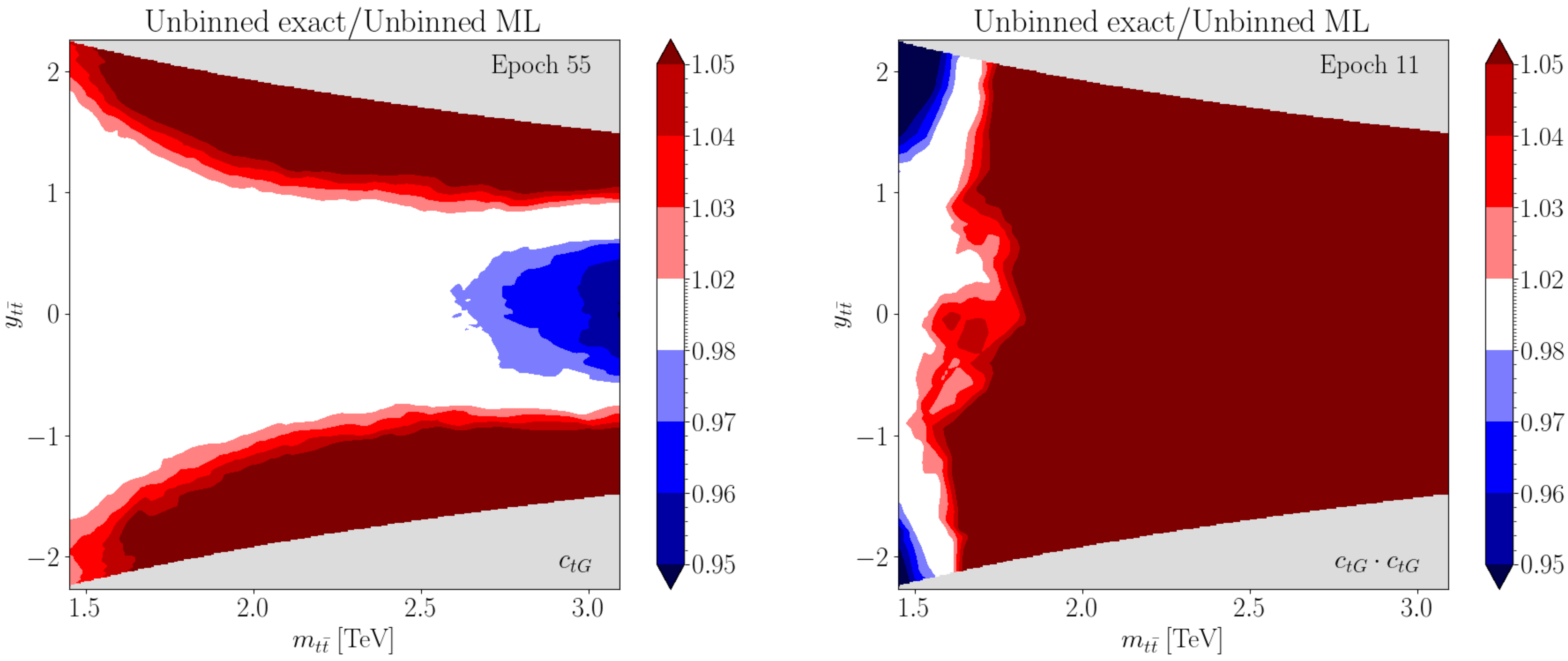
2 kinematic features





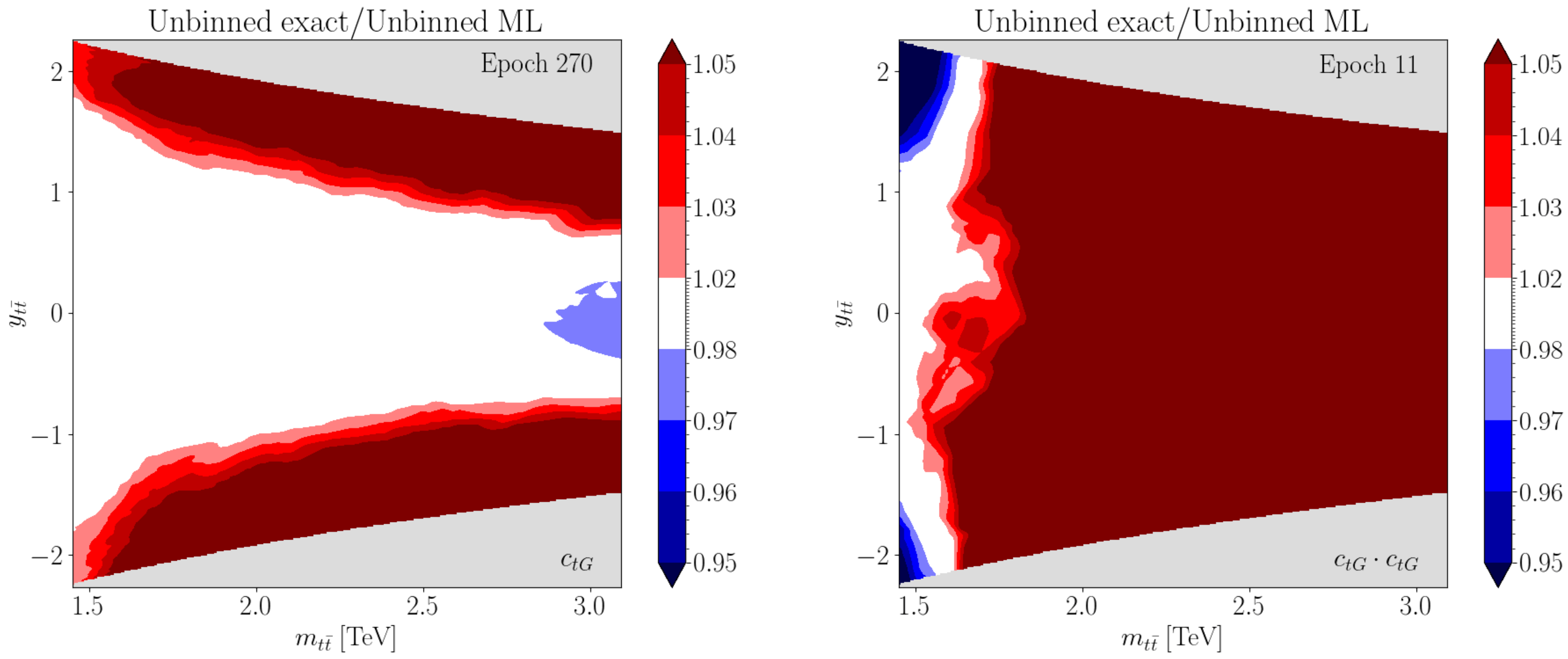
Learning unbinned likelihoods

2 kinematic features



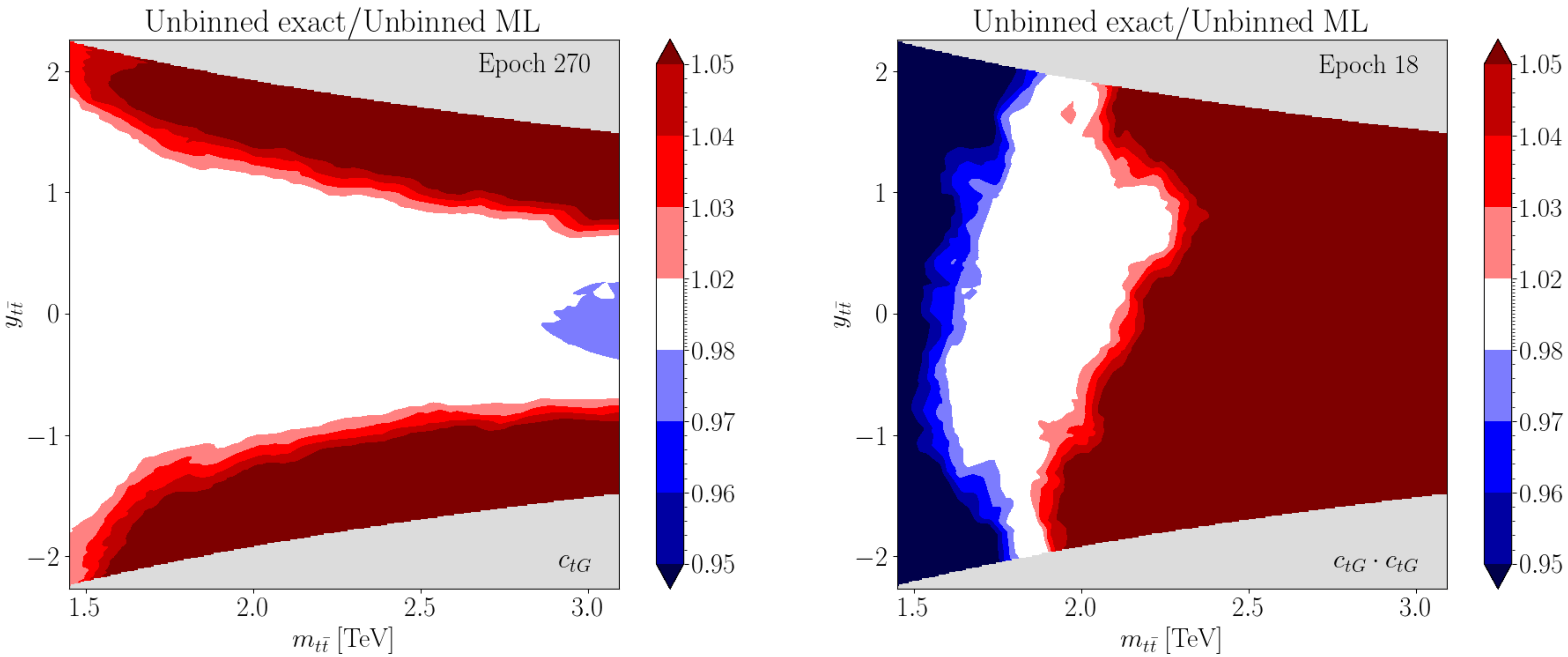
Learning unbinned likelihoods

2 kinematic features



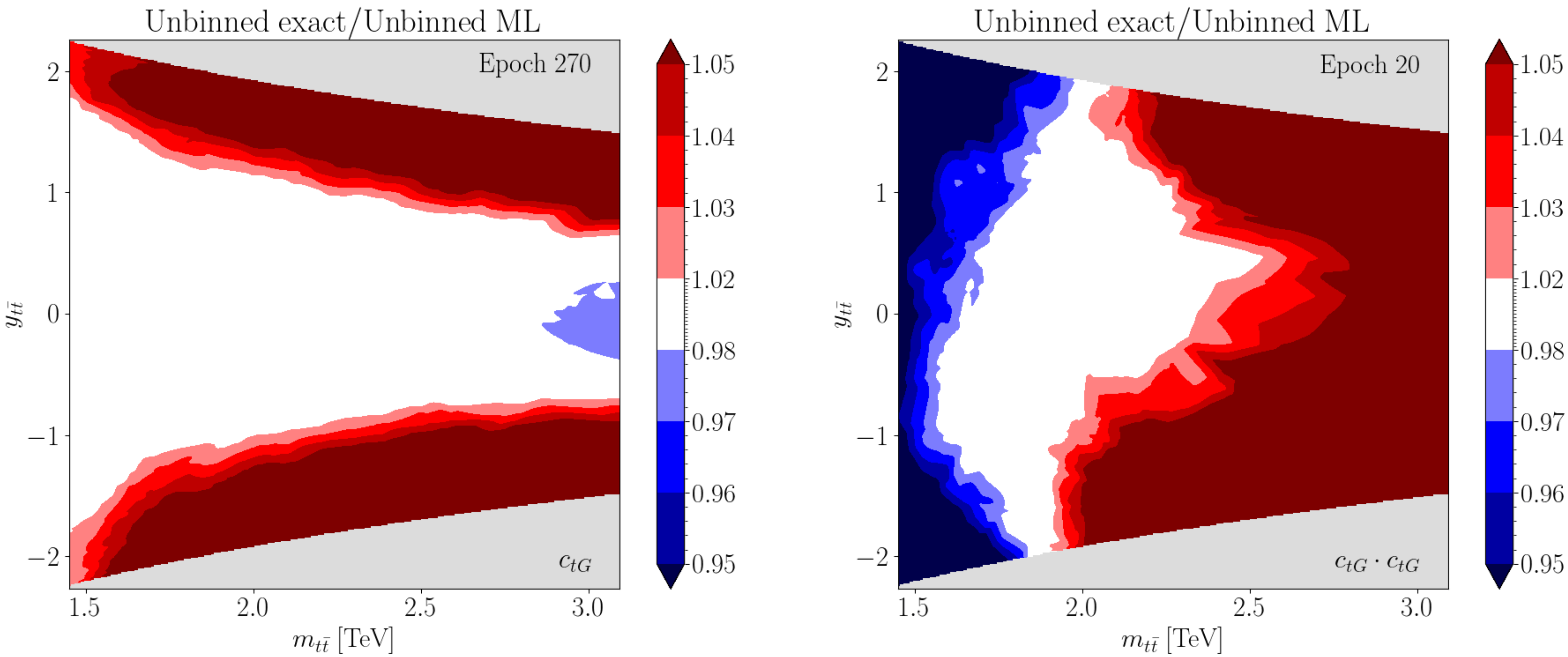
Learning unbinned likelihoods

2 kinematic features



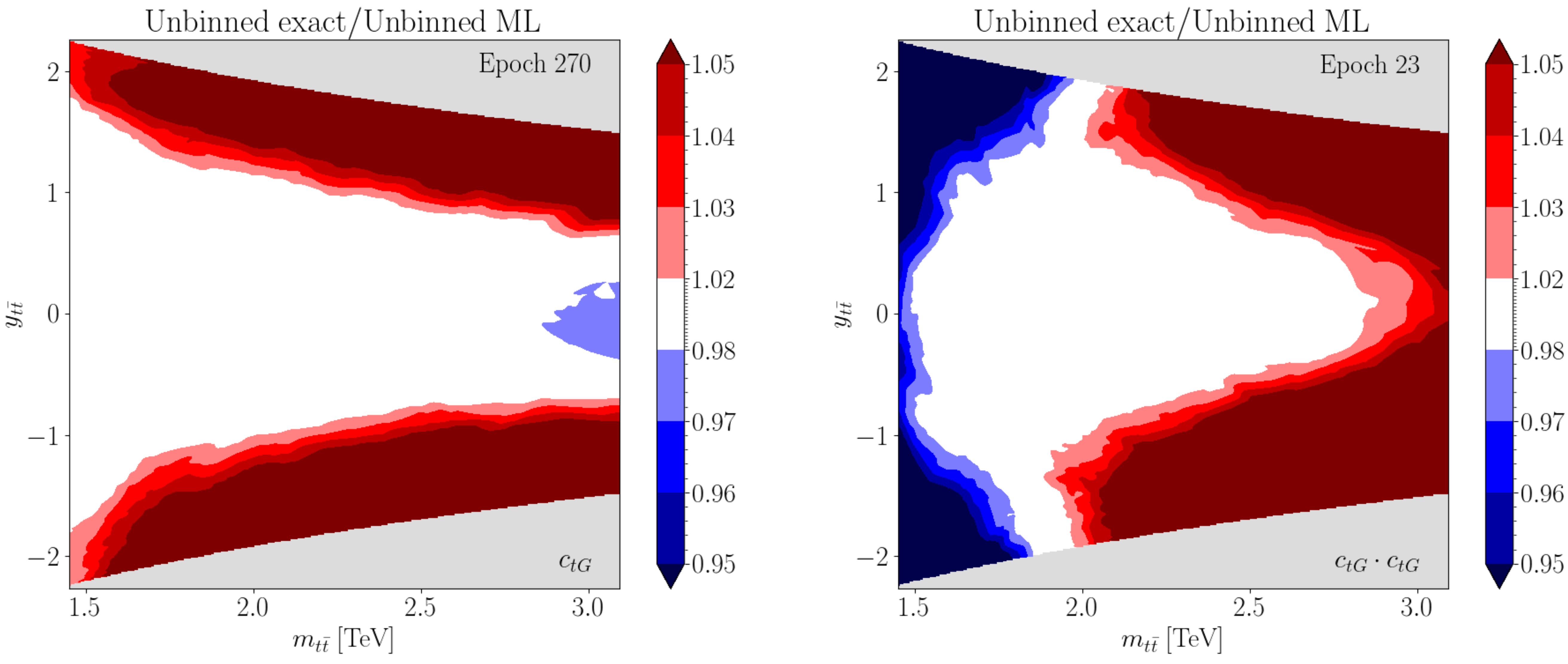
Learning unbinned likelihoods

2 kinematic features



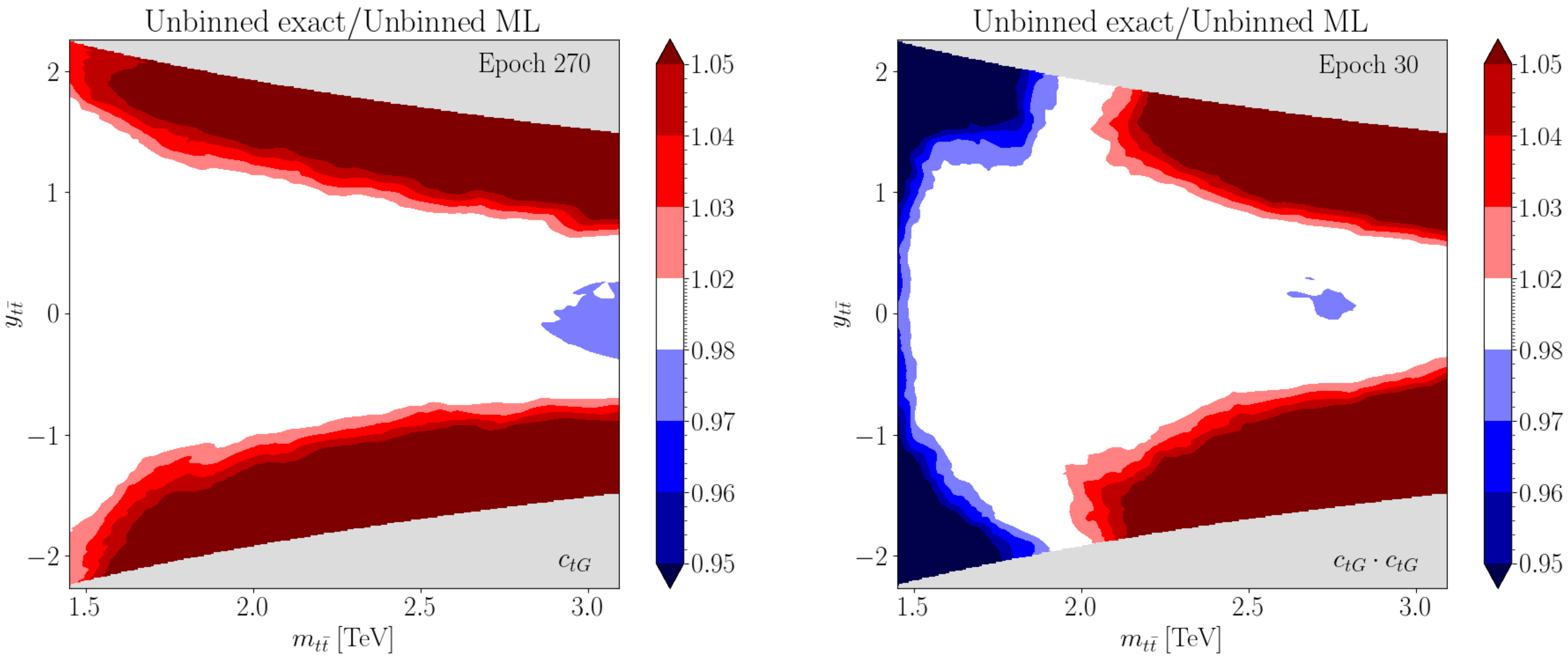
Learning unbinned likelihoods

2 kinematic features



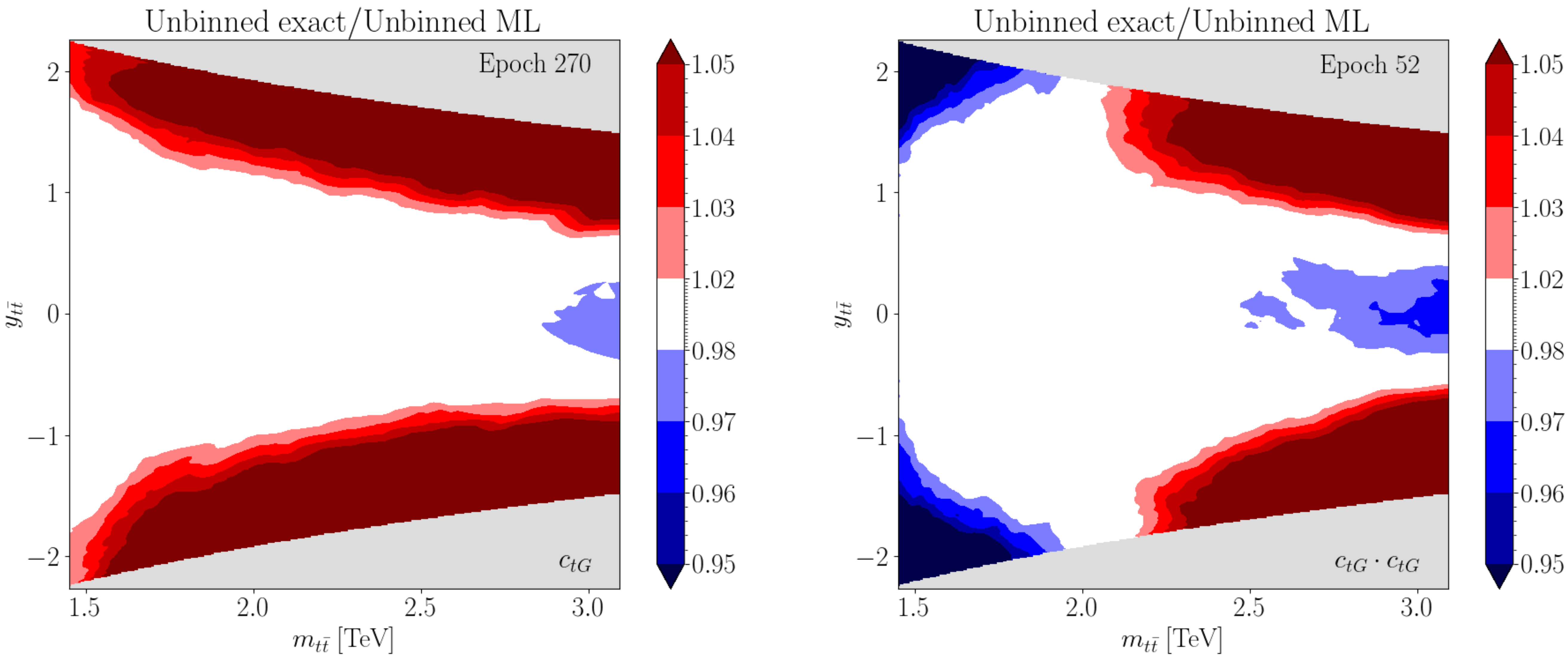
Learning unbinned likelihoods

2 kinematic features



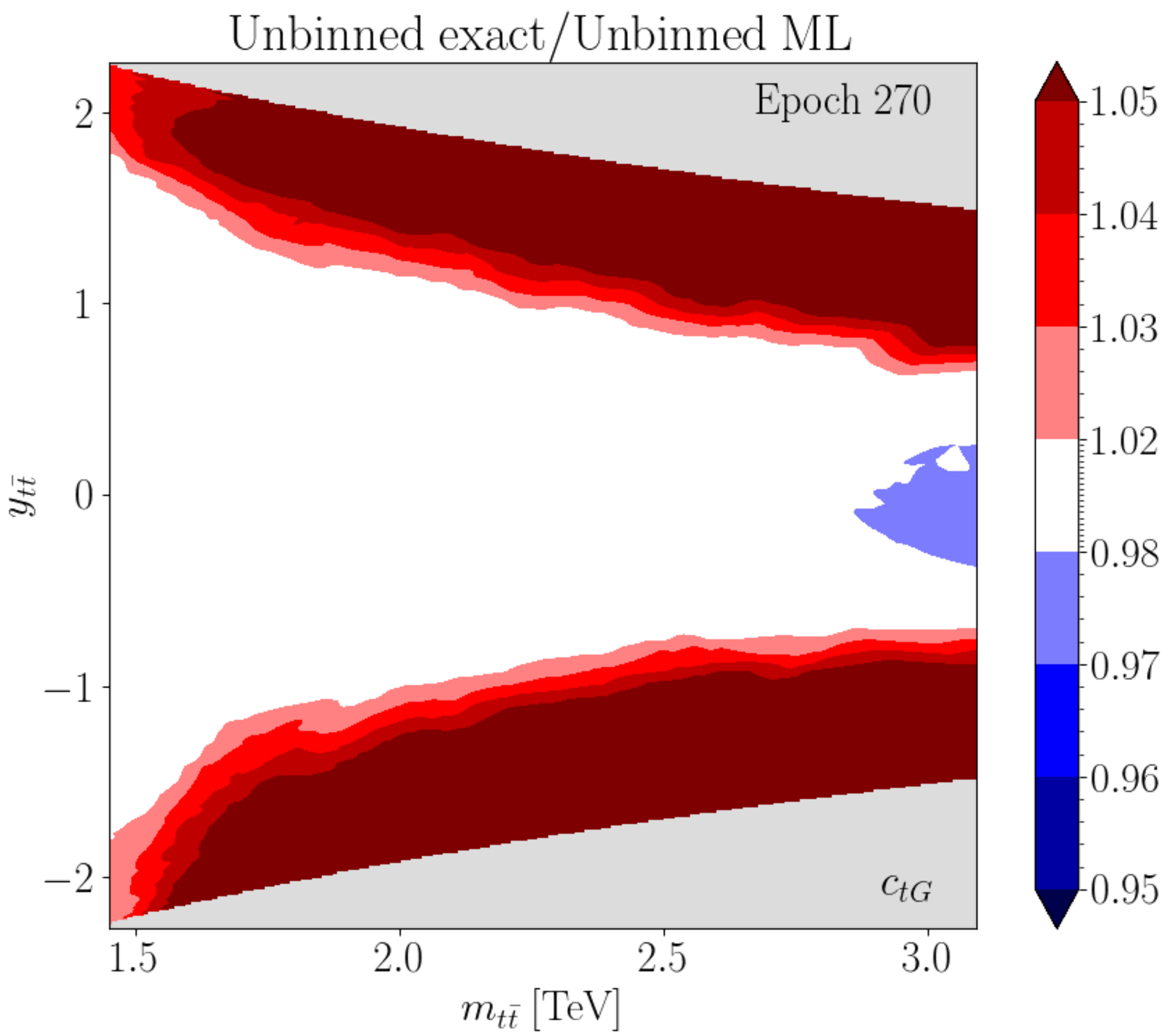
Learning unbinned likelihoods

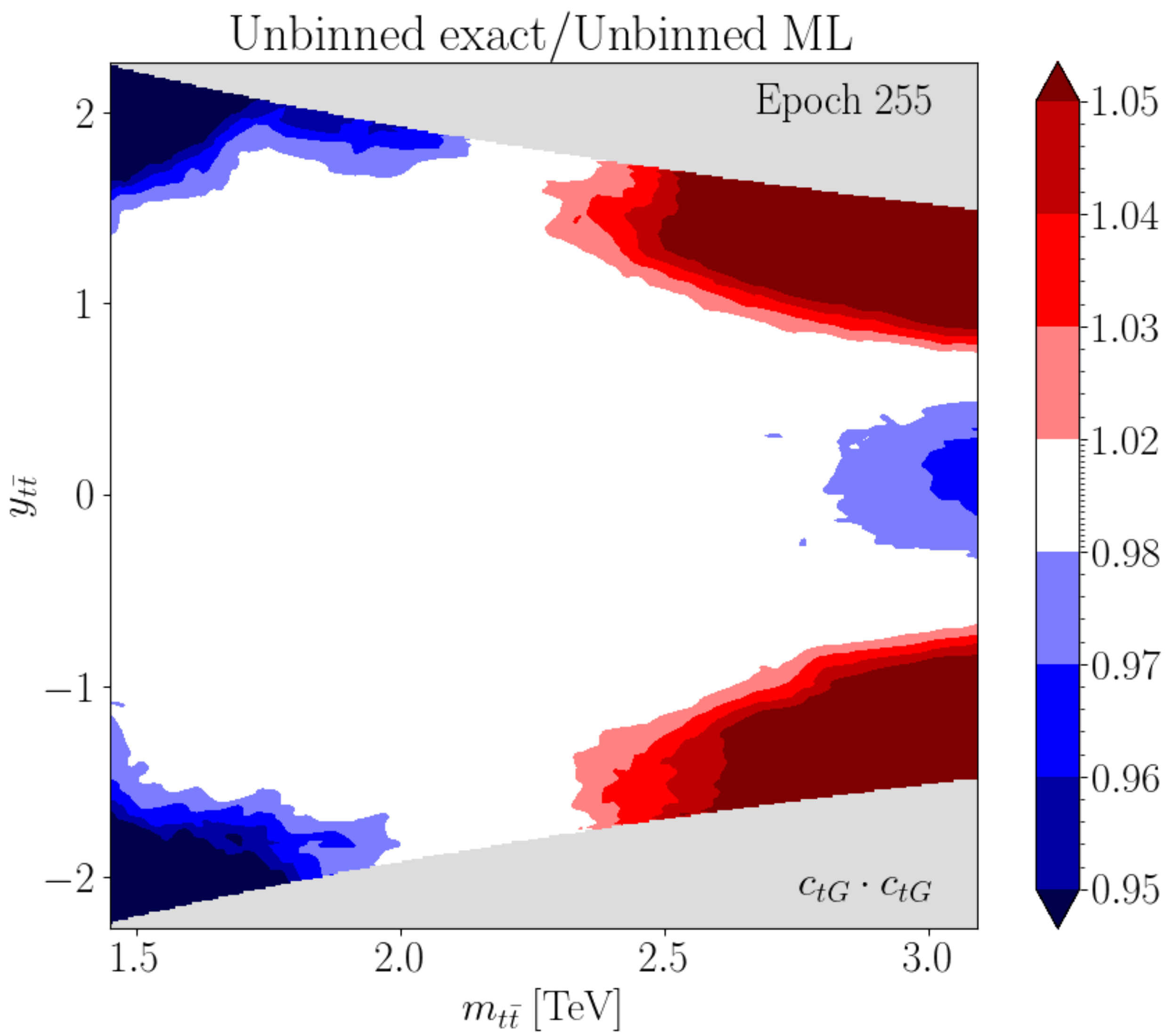
2 kinematic features



Learning unbinned likelihoods

2 kinematic features





Separating the learning problem

- In the SMEFT, the likelihood ratio to the SM takes the form

$$r(\boldsymbol{x}, \boldsymbol{c}) = 1 + \sum_{j=1}^{n_{\rm eft}} r^{(j)}(\boldsymbol{x}) c_j + \sum_{j=1}^{n_{\rm eft}} \sum_{k \geq j}^{n_{\rm eft}} r^{(j,k)}(\boldsymbol{x}) c_j c_k$$

- Learning building blocks $r^{(j)}(\boldsymbol{x})$ and $r^{(j,k)}(\boldsymbol{x})$ lets one walk through the **entire** EFT parameter space!

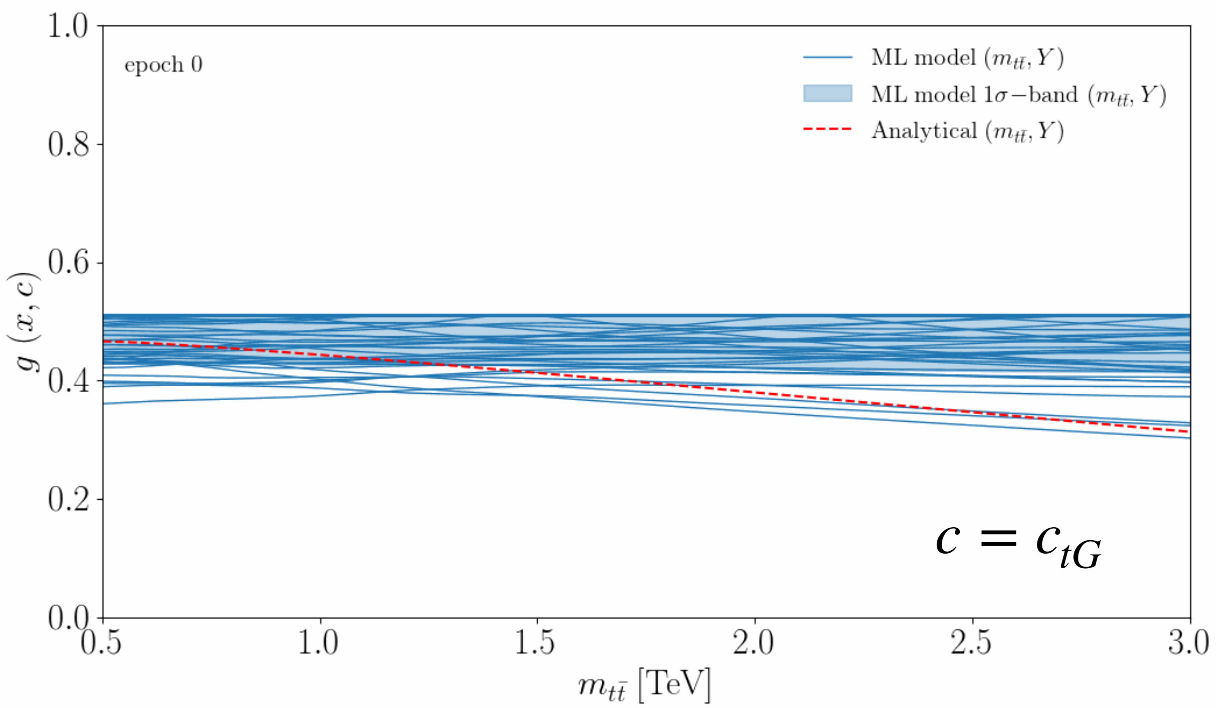
Example: to learn a single $r^{(j)}$, generate $\mathscr{D}_{\rm sm}$ and $\mathscr{D}_{\rm eft}$ at c_j up to $\mathcal{O}(\Lambda^{-2})$. Then $r(\boldsymbol{x}, \boldsymbol{c}) = 1 + r^{(j)}(\boldsymbol{x}) c_j^{(\rm tr)}$ and training means

$$g(\boldsymbol{x}, c_j^{(\rm tr)}) = \left(1 + \left[1 + c_j^{(\rm tr)} \cdot \mathrm{NN}^{(j)}(\boldsymbol{x})\right]\right)^{-1} \qquad \mathrm{NN}^{(j)}(\boldsymbol{x}) \to r^{(j)}(\boldsymbol{x})$$

Uncertainty treatment

MC replica method: propagate methodological uncertainties as well as finite training set effects to the space of models

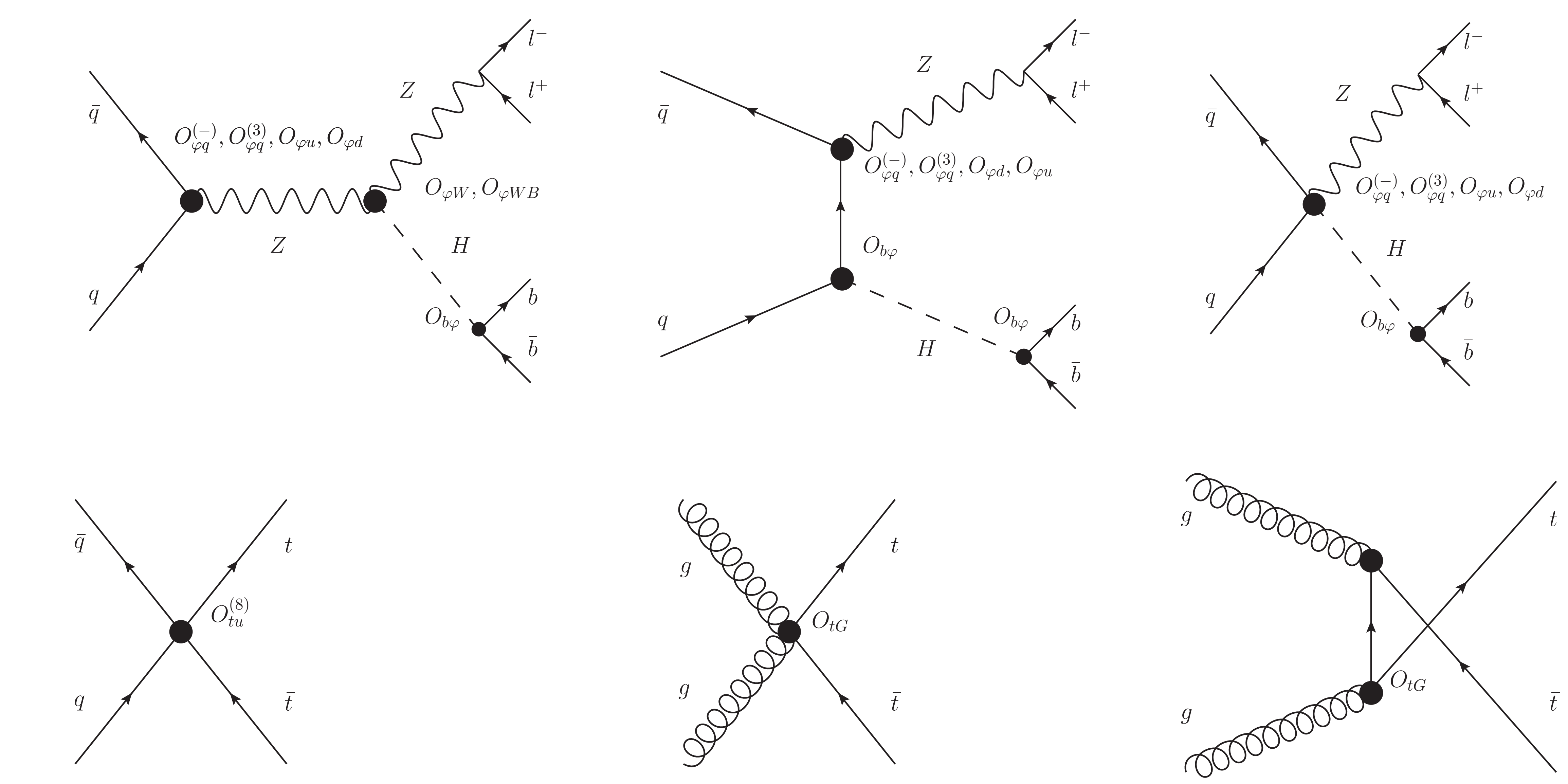
$$\hat{r}^{(i)}(\boldsymbol{x}, \boldsymbol{c}) \equiv 1 + \sum_{j=1}^{n_{\rm eft}} \mathrm{NN}_i^{(j)}(\boldsymbol{x}) c_j + \sum_{j=1}^{n_{\rm eft}} \sum_{k \geq j}^{n_{\rm eft}} \mathrm{NN}_i^{(j,k)}(\boldsymbol{x}) c_j c_k \,, \qquad i = 1, \ldots, N_{\rm rep}$$



Process	$N_{\rm rep}$	$\widetilde{N}_{\rm ev}$ (per replica)	$N_{\rm nn}$	#trainings
$pp \to t\bar{t}$	50	10^5	4	200
$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$	25	10^5	40	1000

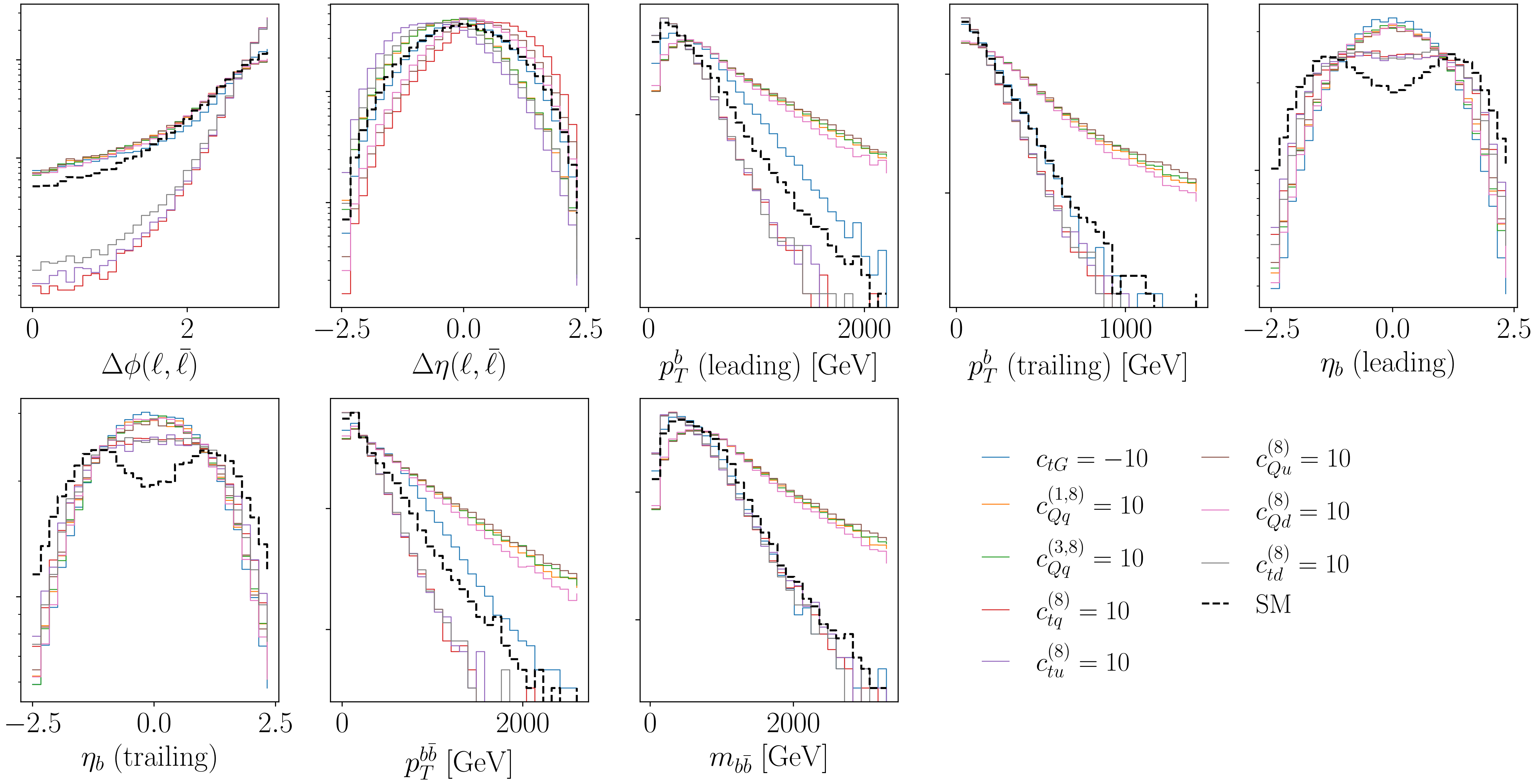
Let's go multivariate

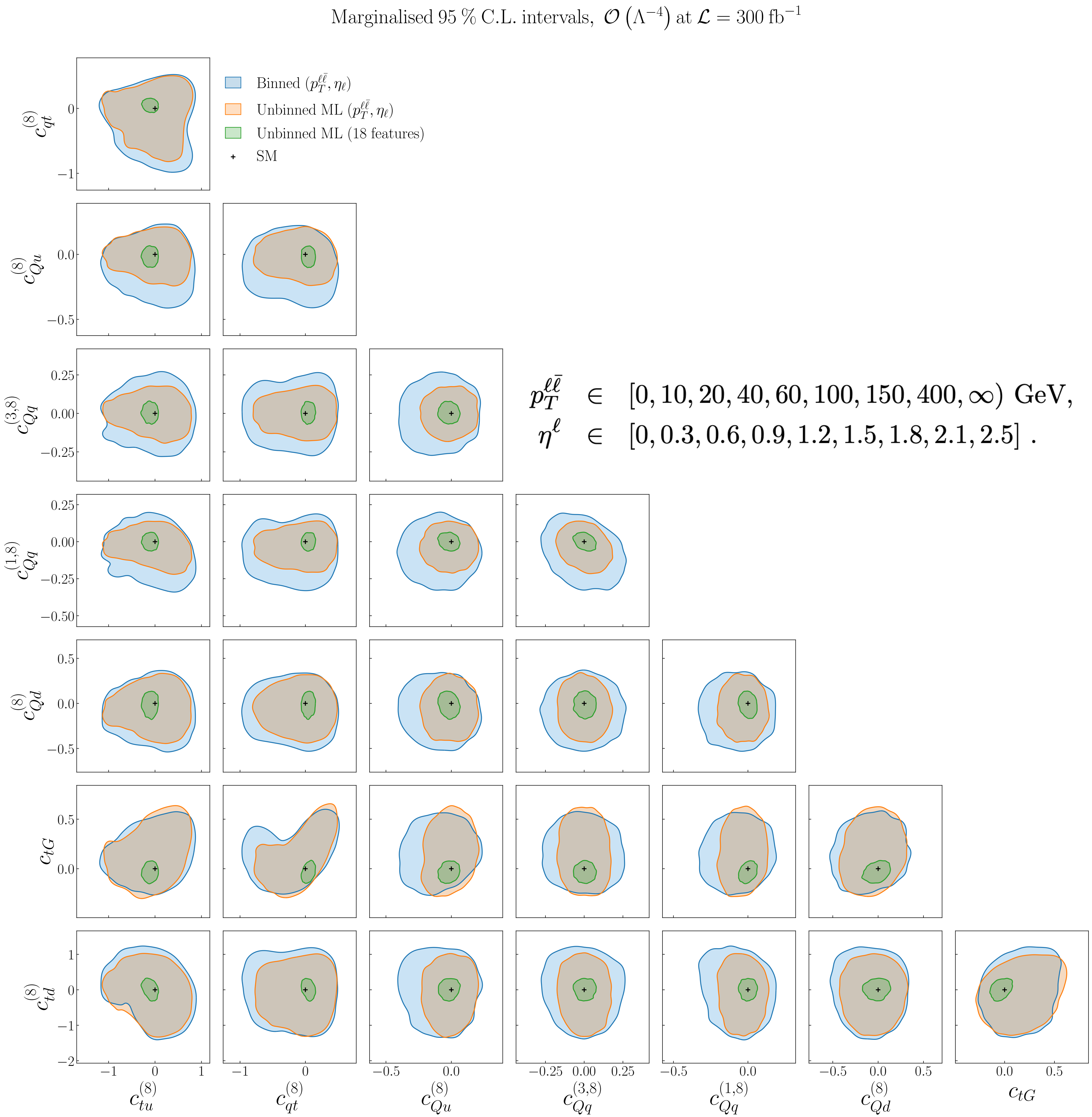
- $pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$: 18 features, 8 EFT coefficients
- $pp \to hZ \to b\bar{b}\ell^+\ell^-$: 7 features, 7 EFT coefficients



Let's go multivariate

- $pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$: 18 features, 8 EFT coefficients
- $pp \to hZ \to b\bar{b}\ell^+\ell^-$: 7 features, 7 EFT coefficients



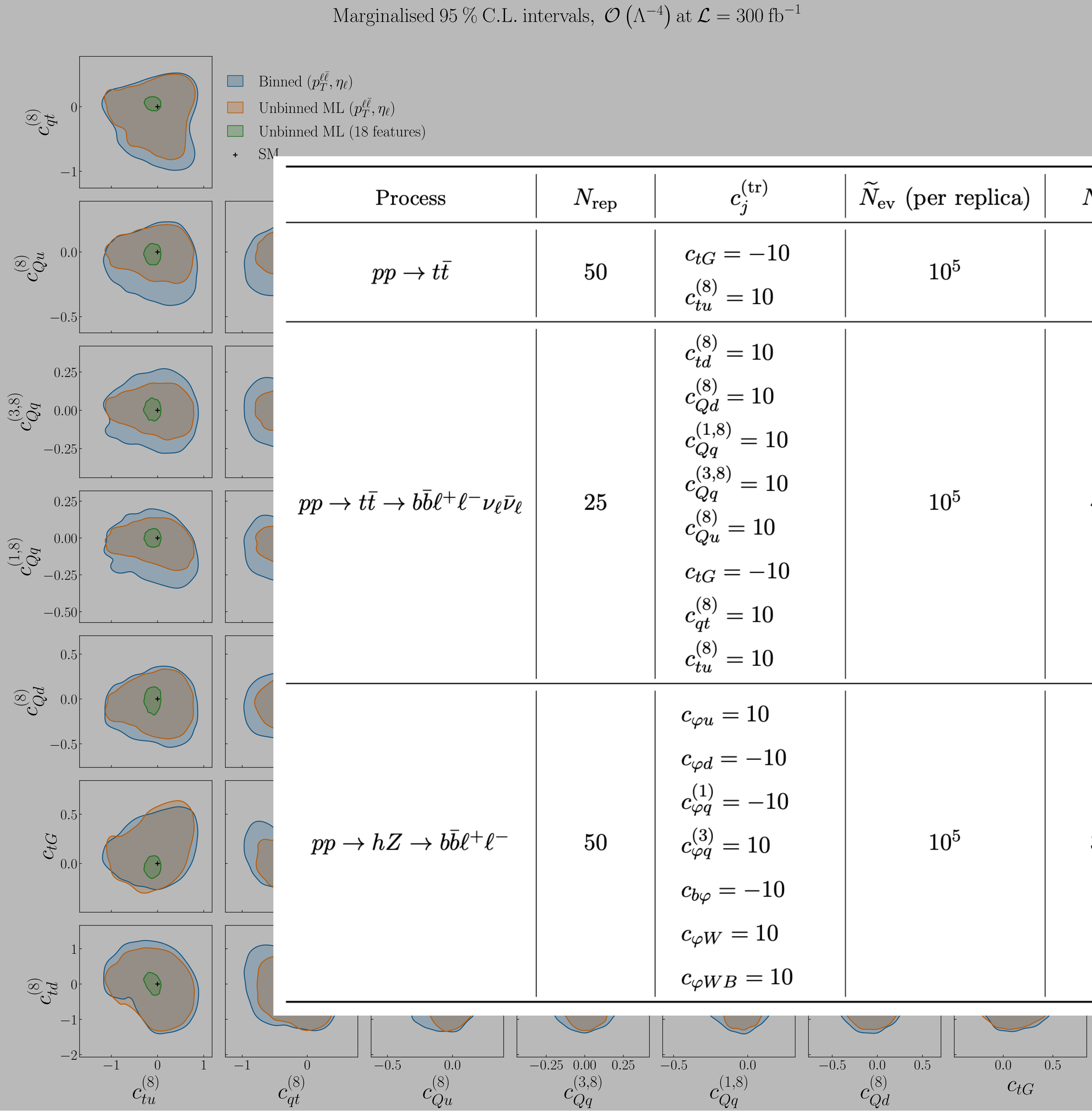


$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$

$$p_T^{\ell\bar{\ell}} \in [0, 10, 20, 40, 60, 100, 150, 400, \infty)\ \text{GeV},$$
$$\eta^\ell \in [0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.5]\ .$$

- Unbinned multivariate data is advantageous to constrain the EFT parameter space!
- Information loss incurred by binning can be quantified

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

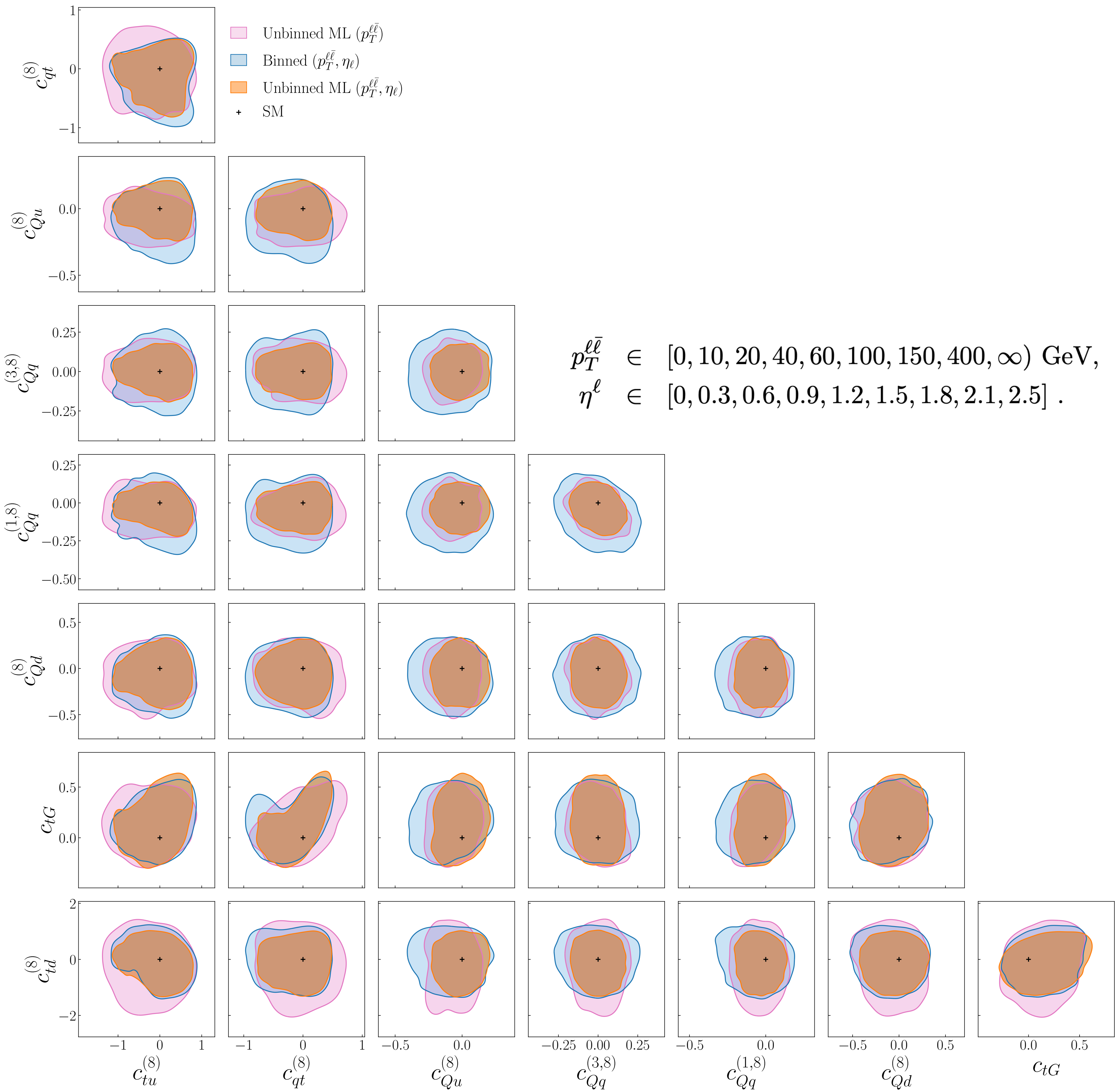


Process	$N_{\rm rep}$	$c_j^{(\rm tr)}$	$\widetilde{N}_{\rm ev}$ (per replica)	$N_{\rm nn}$	#trainings
$pp \to t\bar{t}$	50	$c_{tG} = -10$ $c_{tu}^{(8)} = 10$	10^5	4	200
$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$	25	$c_{td}^{(8)} = 10$ $c_{Qd}^{(8)} = 10$ $c_{Qq}^{(1,8)} = 10$ $c_{Qq}^{(3,8)} = 10$ $c_{Qu}^{(8)} = 10$ $c_{tG} = -10$ $c_{qt}^{(8)} = 10$ $c_{tu}^{(8)} = 10$	10^5	40	1000
$pp \to hZ \to b\bar{b}\ell^+\ell^-$	50	$c_{\varphi u} = 10$ $c_{\varphi d} = -10$ $c_{\varphi q}^{(1)} = -10$ $c_{\varphi q}^{(3)} = 10$ $c_{b\varphi} = -10$ $c_{\varphi W} = 10$ $c_{\varphi WB} = 10$	10^5	30	1500

ed
riate data is
ageous to
in the EFT
eter space!

tion loss
d by
can be
ed

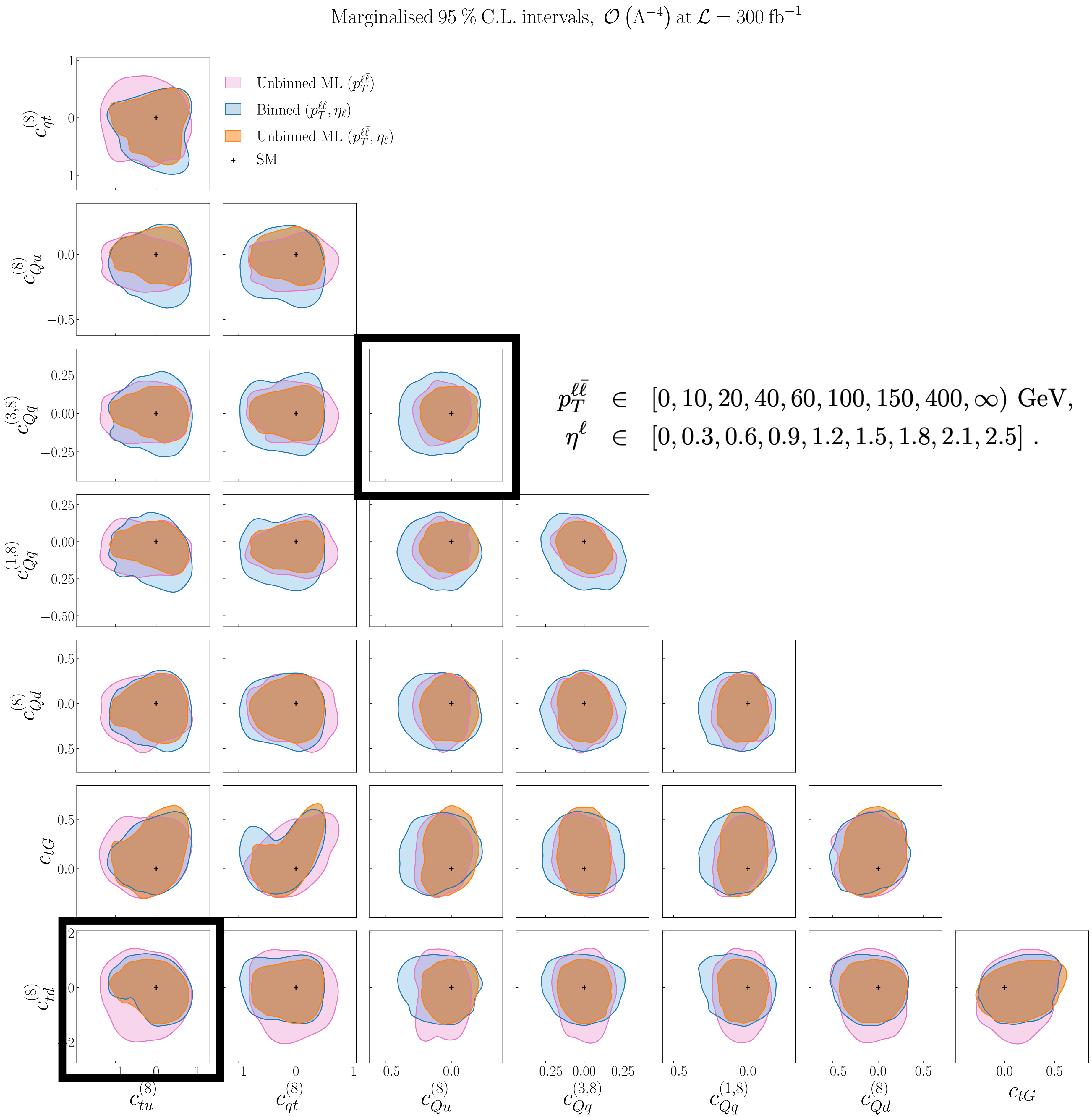
Marginalised 95 % C.L. intervals, $\mathcal{O}(\Lambda^{-4})$ at $\mathcal{L} = 300\ \text{fb}^{-1}$



$$p_T^{\ell\bar\ell} \in [0, 10, 20, 40, 60, 100, 150, 400, \infty)\ \text{GeV},$$
$$\eta^\ell \in [0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.5]\ .$$

$$pp \to t\bar t \to b\bar b \ell^+\ell^-\nu_\ell\bar\nu_\ell$$

- Unbinned multivariate data is advantageous to constrain the EFT parameter space!
- Adding extra features or going unbinned?

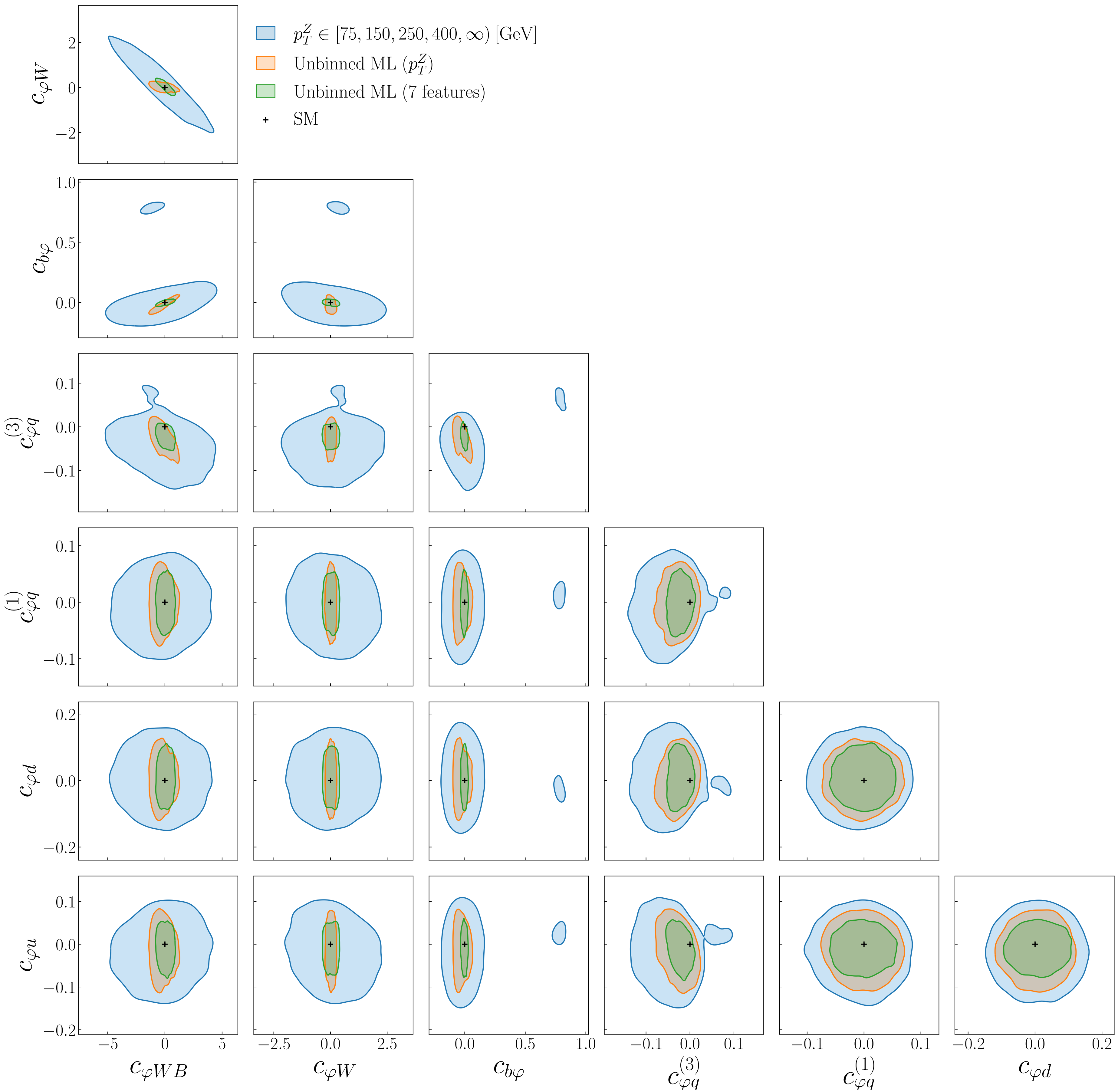


$$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$$

$$p_T^{\ell\bar{\ell}} \in [0, 10, 20, 40, 60, 100, 150, 400, \infty)\ \text{GeV},$$
$$\eta^\ell \in [0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.5]\ .$$

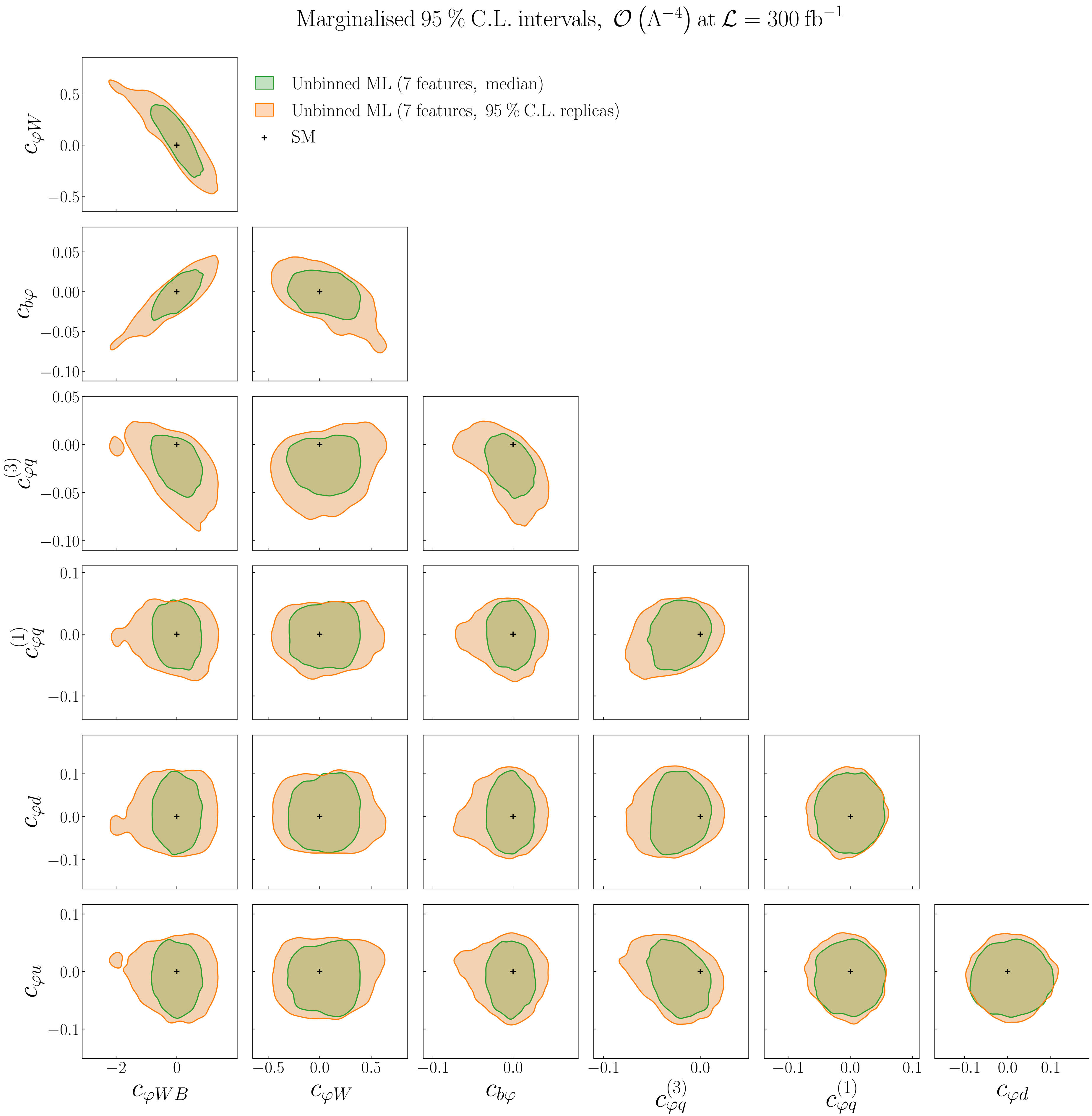
- Unbinned multivariate data is advantageous to constrain the EFT parameter space!
- Adding extra features or going unbinned?

Marginalised 95 % C.L. intervals, $\mathcal{O}(\Lambda^{-4})$ at $\mathcal{L} = 300\ \text{fb}^{-1}$



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

- Unbinned multivariate data is advantageous to constrain the EFT parameter space!
- **Degeneracies** get lifted

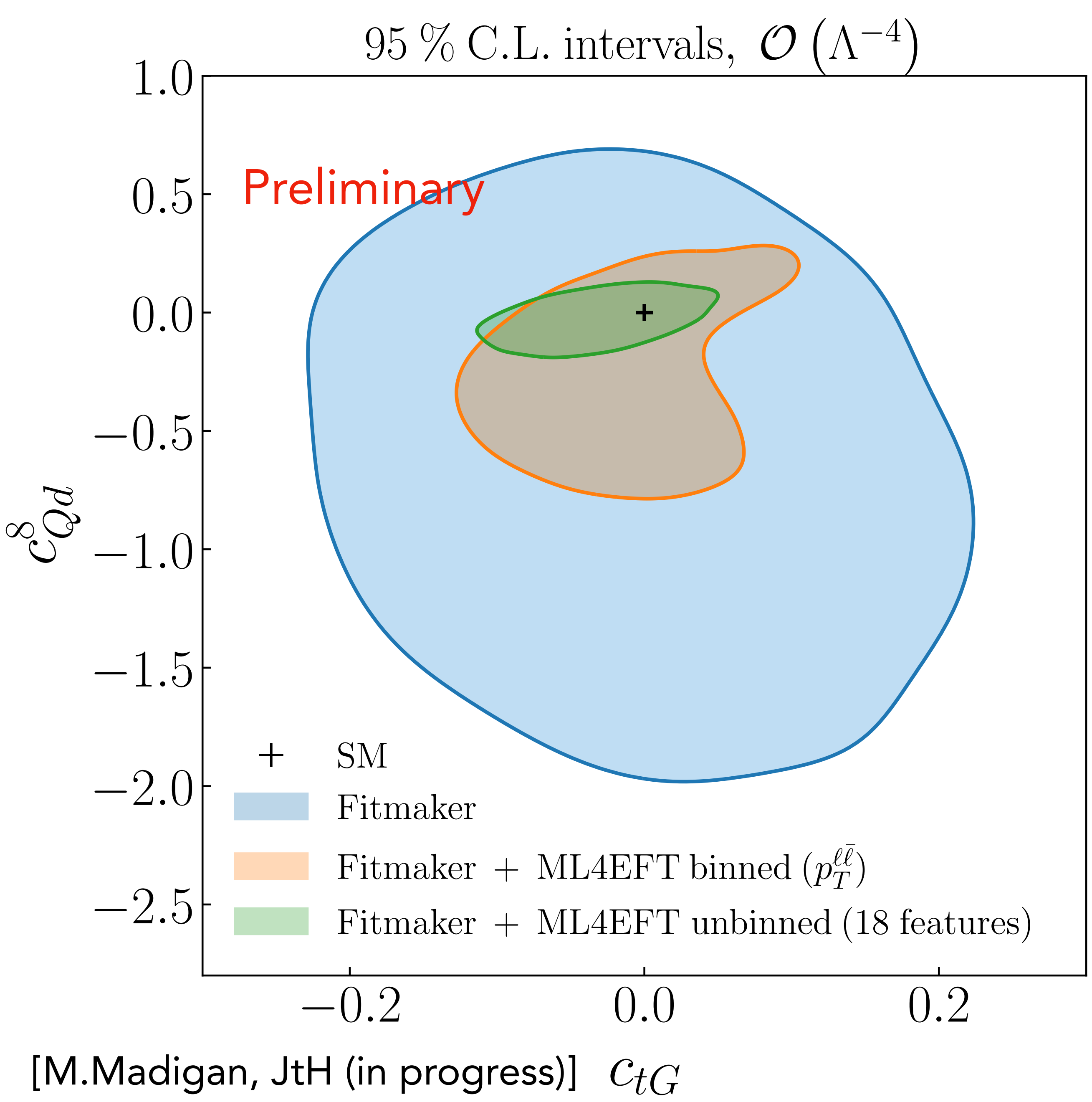


$pp \to hZ \to b\bar{b}\ell^+\ell^-$

- We account for the methodological uncertainties through the MC-replica method
- The qualitative picture does not change

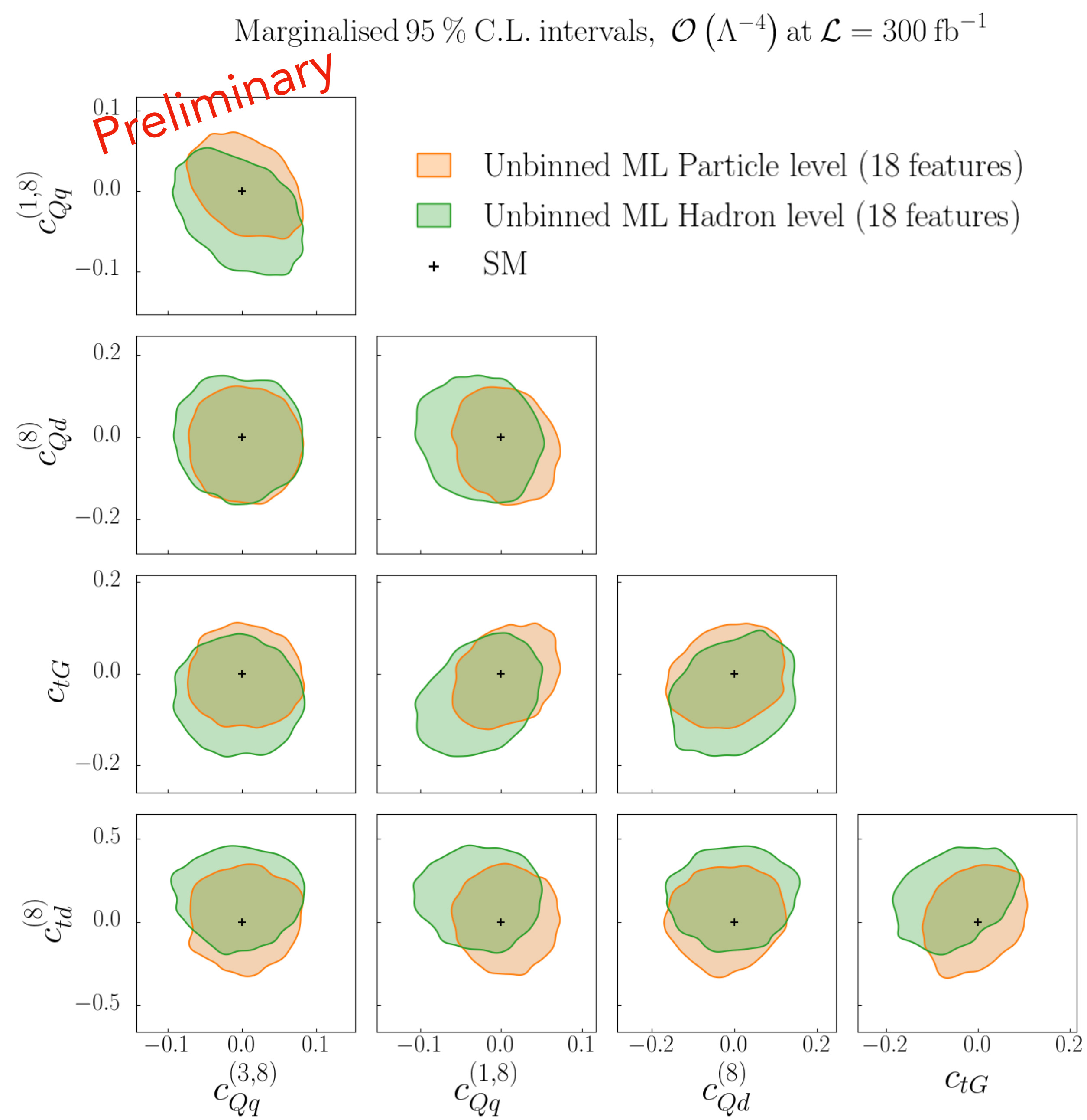
Ongoing efforts

1. Integration into global fits



[M.Madigan, JtH (in progress)]

2. Hadronized level



MSc project by Pim Herbschleb

Conclusion and outlook

- **`ML4EFT`** accounts for methodological uncertainties and has efficient scaling properties, as required for global EFT fits
- Global EFT fits based on unbinned observables **enhance** the sensitivity significantly
- **Integration** into existing global fit frameworks possible
- Please visit **`ML4EFT`** on GitHub (documentation + **tutorial**)

lhcfitnikhef.github.io/ML4EFT

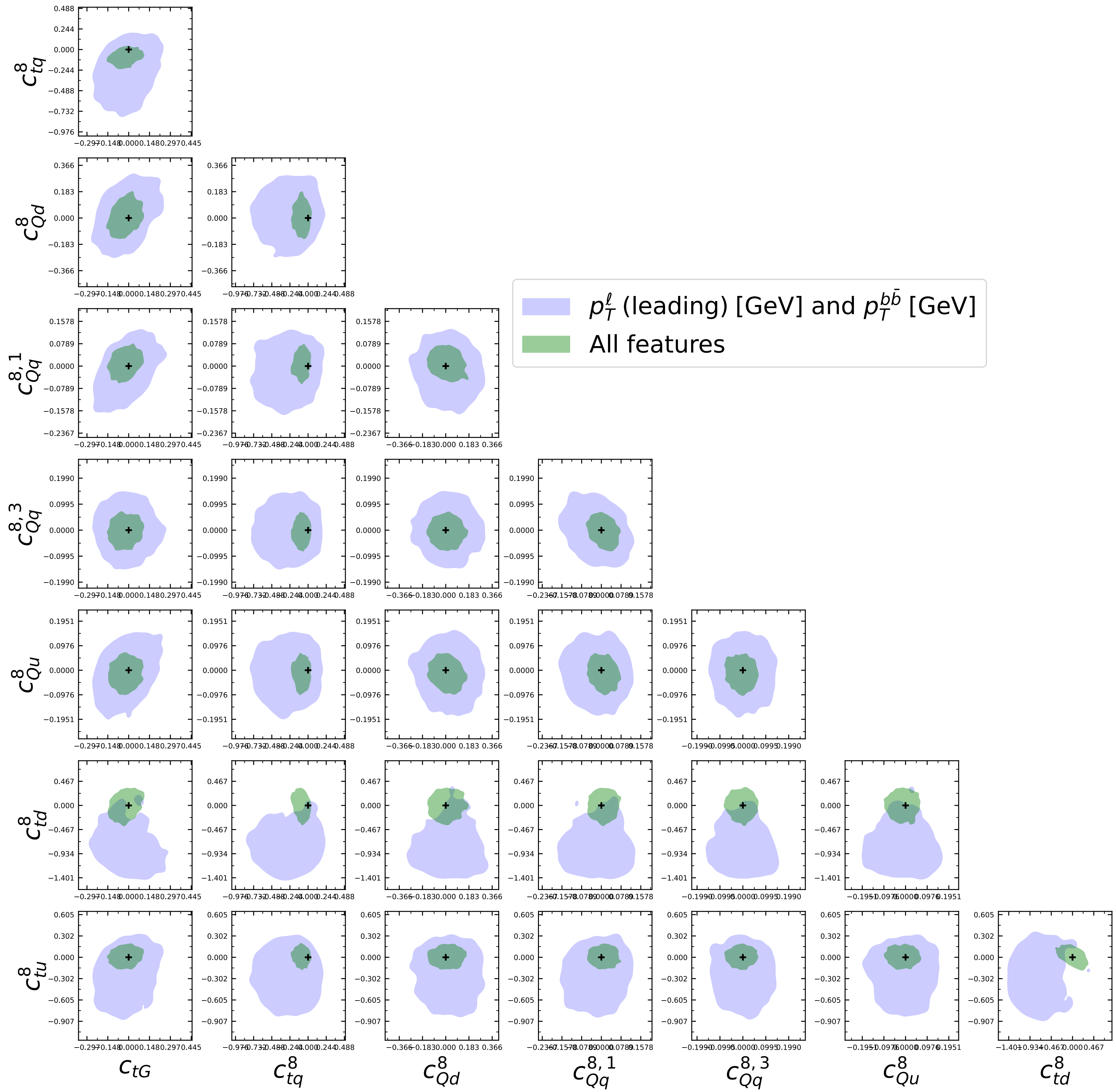
Conclusion and outlook

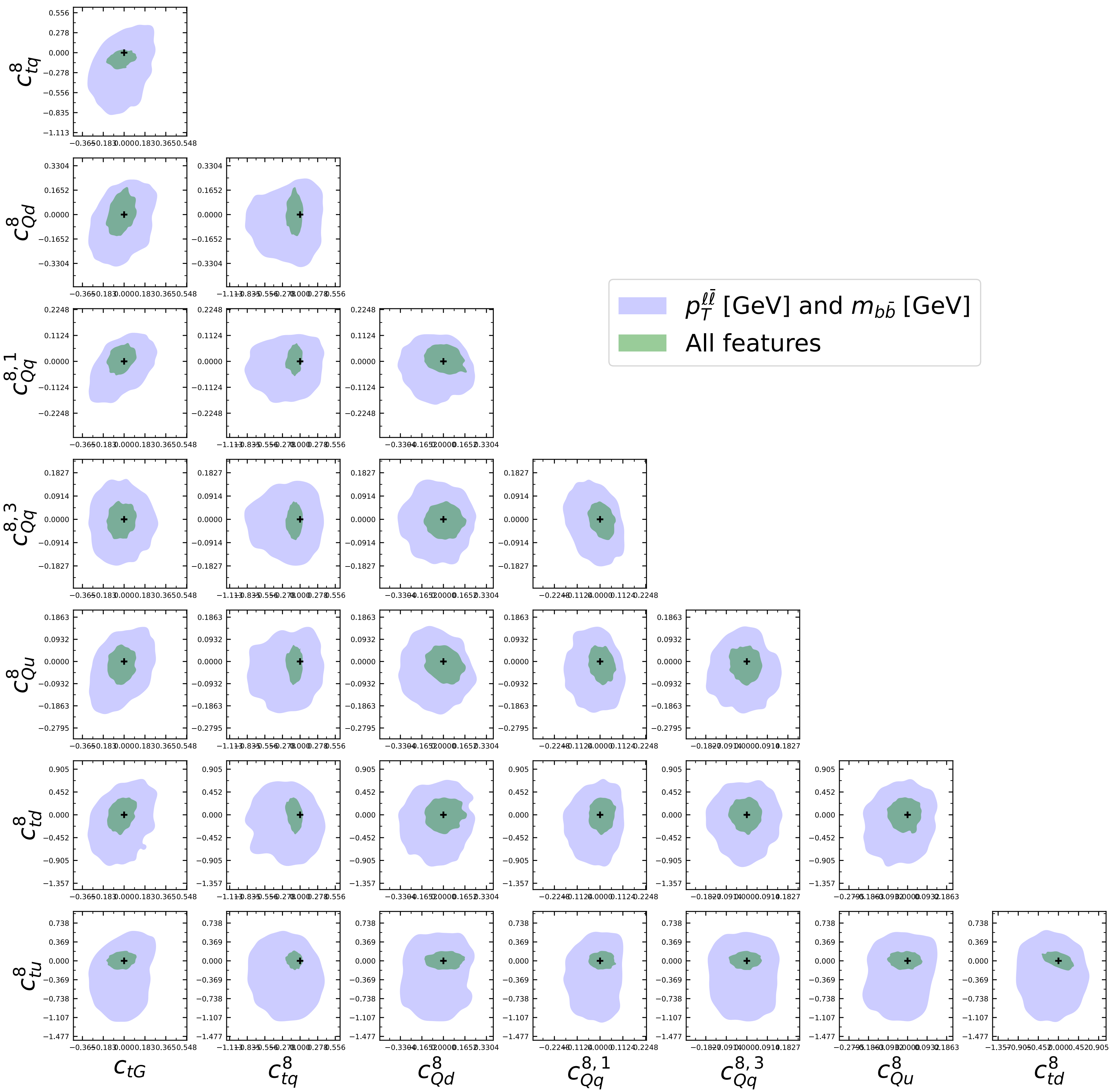
- `ML4EFT` accounts for methodological uncertainties and has efficient scaling properties, as required for global EFT fits
- Global EFT fits based on unbinned observables **enhance** the sensitivity significantly
- **Integration** into existing global fit frameworks possible
- Please visit `ML4EFT` on GitHub (documentation + **tutorial**)

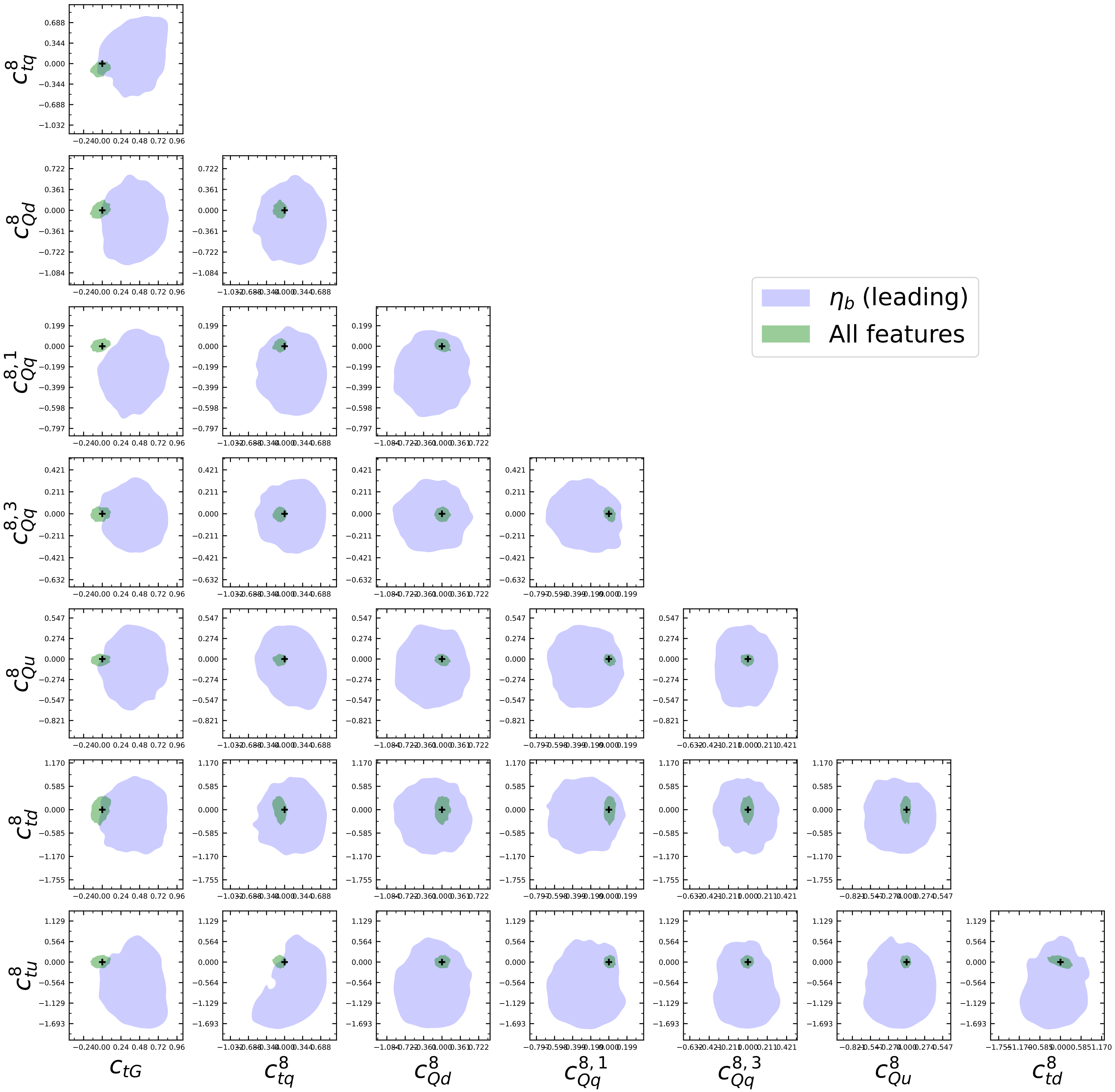
lhcfitnikhef.github.io/ML4EFT

Thank you!

Backup


p_T^ℓ (leading) [GeV] and $p_T^{b\bar{b}}$ [GeV]
All features
c_{tq}^8
c_{Qd}^8
$c_{Qq}^{8,1}$
$c_{Qq}^{8,3}$
c_{Qu}^8
c_{td}^8
c_{tu}^8
c_{tG}
c_{tq}^8
c_{Qd}^8
$c_{Qq}^{8,1}$
$c_{Qq}^{8,3}$
c_{Qu}^8
c_{td}^8


$p_T^{\ell\bar{\ell}}$ [GeV] and $m_{b\bar{b}}$ [GeV]
All features
c_{tG}
c_{tq}^8
c_{Qd}^8
$c_{Qq}^{8,1}$
$c_{Qq}^{8,3}$
c_{Qu}^8
c_{td}^8
c_{tu}^8



η_b (leading)
All features
c_{tq}^{8}
c_{Qd}^{8}
$c_{Qq}^{8,1}$
$c_{Qq}^{8,3}$
c_{Qu}^{8}
c_{td}^{8}
c_{tu}^{8}
c_{tG}

Training times and hyperparameters

process	features	hidden layers	learning rate	n_{batch}	time (min)
$pp \to t\bar{t}$	$m_{t\bar{t}}$	$25 \times 25 \times 25$	10^{-3}	5	17.3 ± 13.9
	$m_{t\bar{t}}, y_{t\bar{t}}$	$25 \times 25 \times 25$	10^{-3}	5	16.4 ± 12.7
$pp \to t\bar{t} \to b\bar{b}\ell^+\ell^-\nu_\ell\bar{\nu}_\ell$	$p_T^{\ell\bar{\ell}}$	$25 \times 25 \times 25$	10^{-3}	1	46.8 ± 35.0
	$p_T^{\ell\bar{\ell}}, \eta_\ell$	$25 \times 25 \times 25$	10^{-3}	1	53.7 ± 29.9
	18	$100 \times 100 \times 100$	10^{-4}	50	5.4 ± 2.7
$pp \to hZ \to b\bar{b}\ell^+\ell^-$	p_T^Z	$100 \times 100 \times 100$	10^{-3}	100	9.4 ± 9.0
	7	$100 \times 100 \times 100$	10^{-4}	50	14.1 ± 8.7

Scaling

- All the ingredients that make up the LR can be trained in parallel, making **`ML4EFT`** ideal for optimised global EFT fits

$$r(\boldsymbol{x}, \boldsymbol{c}) = 1 + r^{(i)}(\boldsymbol{x}) c_i^{(\mathrm{tr})}, \quad i = 1, \ldots n_{\mathrm{eft}}$$

$$r(\boldsymbol{x}, \boldsymbol{c}) = 1 + r^{(i,j)}(\boldsymbol{x}) c_i^{(\mathrm{tr})} c_j^{(\mathrm{tr})}, \quad i, j = 1, \ldots n_{\mathrm{eft}},$$

- Number of NNs scales quadratically with n_{op} and can be fully parallelised

$$\text{Training cost} \sim \mathcal{O}\left(\frac{n_{\mathrm{eft}}^2}{n_{\mathrm{proc}}}\right)$$

Unbinned extended likelihood

After training,

$$\hat{r}(\boldsymbol{x}, \boldsymbol{c}) = 1 + \sum_{j=1}^{n_{\rm eft}} \mathrm{NN}^{(j)}(\boldsymbol{x}) c_j + \sum_{j=1}^{n_{\rm eft}} \sum_{k \geq j}^{n_{\rm eft}} \mathrm{NN}^{(j,k)}(\boldsymbol{x}) c_j c_k$$

serves as input to the **unbinned** extended likelihood ratio

$$q_c = 2 \left[\nu_{\rm tot}(\boldsymbol{c}) - \nu_{\rm tot}(\hat{\boldsymbol{c}}) - \sum_{i=1}^{N_{\rm ev}} \log r(x_i, \boldsymbol{c}, \hat{\boldsymbol{c}}) \right]$$

Kinematic features

$t\bar{t}$

of kinematic features $\boldsymbol{x}$, it is composed of $n_k = 18$ features: p_T of the lepton p_T^{ℓ}, p_T of the antilepton $p_T^{\bar{\ell}}$, leading p_T^{ℓ}, trailing p_T^{ℓ}, lepton pseudorapidity η_{ℓ}, antilepton pseudorapidity $\eta_{\bar{\ell}}$, leading η_{ℓ}, trailing η_{ℓ}, p_T of the dilepton system $p_T^{\ell\bar{\ell}}$, invariant mass of the dilepton system $m_{\ell\bar{\ell}}$, absolute difference in azimuthal angle $|\Delta\phi(\ell,\bar{\ell})|$, difference in absolute rapidity $\Delta\eta(\ell,\bar{\ell})$, leading p_T of the b-jet, trailing p_T of the b-jet, pseudorapidity of the leading b-jet η_b, pseudorapidity of the trailing b-jet η_b, p_T of the $b\bar{b}$ system $p_T^{b\bar{b}}$, and invariant mass of the $b\bar{b}$ system $m_{b\bar{b}}$. These features are partially correlated among them, and hence maximal

ZH

with $\Delta R(b_1, b_2) < 3.0, 1.8, 1.2$ for $p_T^Z \in (75, 150]$ GeV, $(150, 200]$ GeV, and $(200, \infty)$ GeV respectively. The array of kinematic features $\boldsymbol{x}$ for this process is composed of the following $n_k = 7$ features: the transverse momentum of the Z boson p_T^Z, that of the b-quark p_T^b, that of the $b\bar{b}$ pair $p_T^{b\bar{b}}$, the angular separation $\Delta R_{b\bar{b}}$ of the b-quarks, their azimuthal angle separation $\Delta\phi_{b,b\bar{b}}$, the rapidity difference between the dilepton and the $b\bar{b}$ system $\Delta\eta_{Z,b\bar{b}}$, and the azimuthal angle separation $\Delta\phi_{\ell b}$. Again, most of these features are correlated among them and hence there will be a degree of redundancy in the analysis.

Kinematic features