2211.02058

Statistically optimal observables for global SMEFT fits

5th General meeting of the LHC EFT WG
02/12/22

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

Nikhef

2211.02058

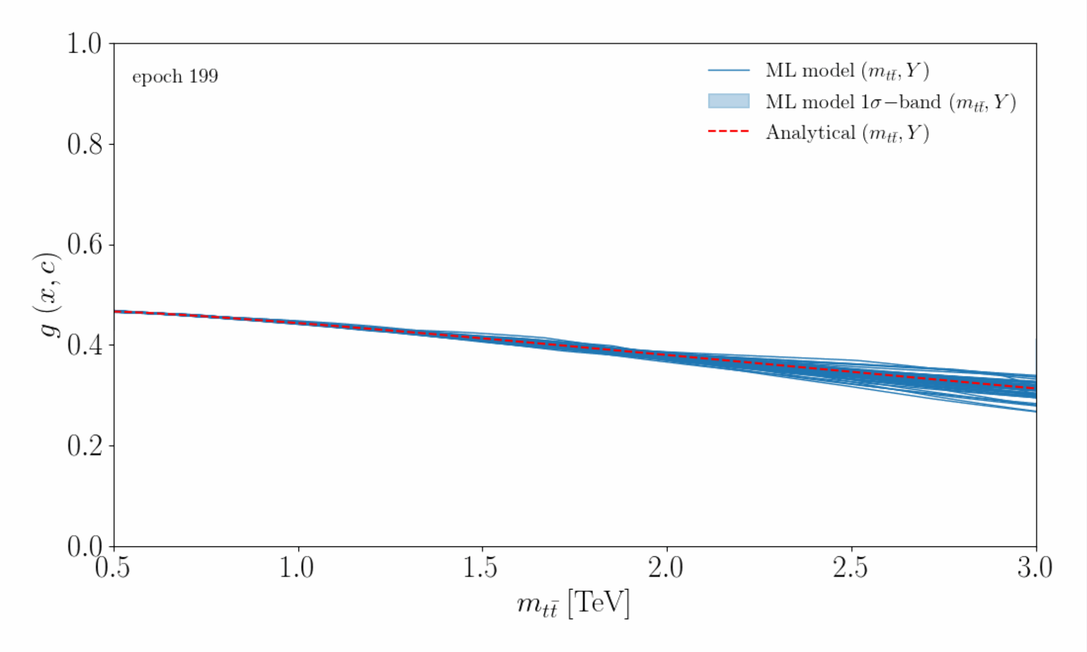
Unbinned multivariate observables

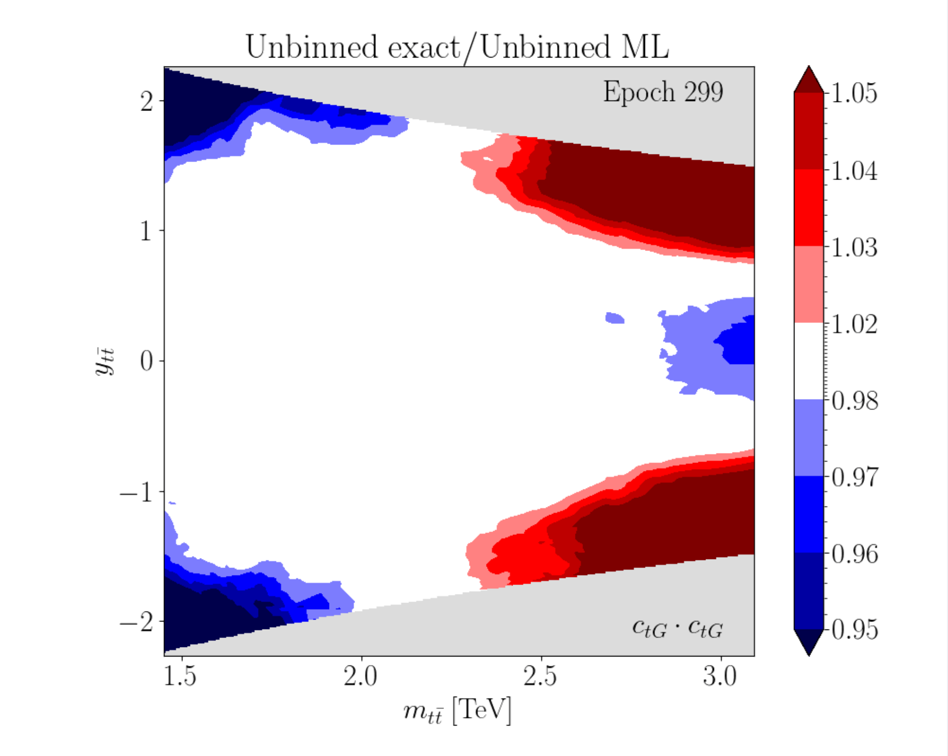
- Which measurement is the most sensitive to SMEFT operators?
  - Inclusive, single to multi-differential (which variables)
  - Binned or unbinned
  - Which binning?
- **`ML4EFT`**: open-source framework to integrate unbinned multivariate observables into **global** SMEFT fits
  - Provides optimal bounds on the EFT parameters
  - Useful diagnosis tool to assess information loss

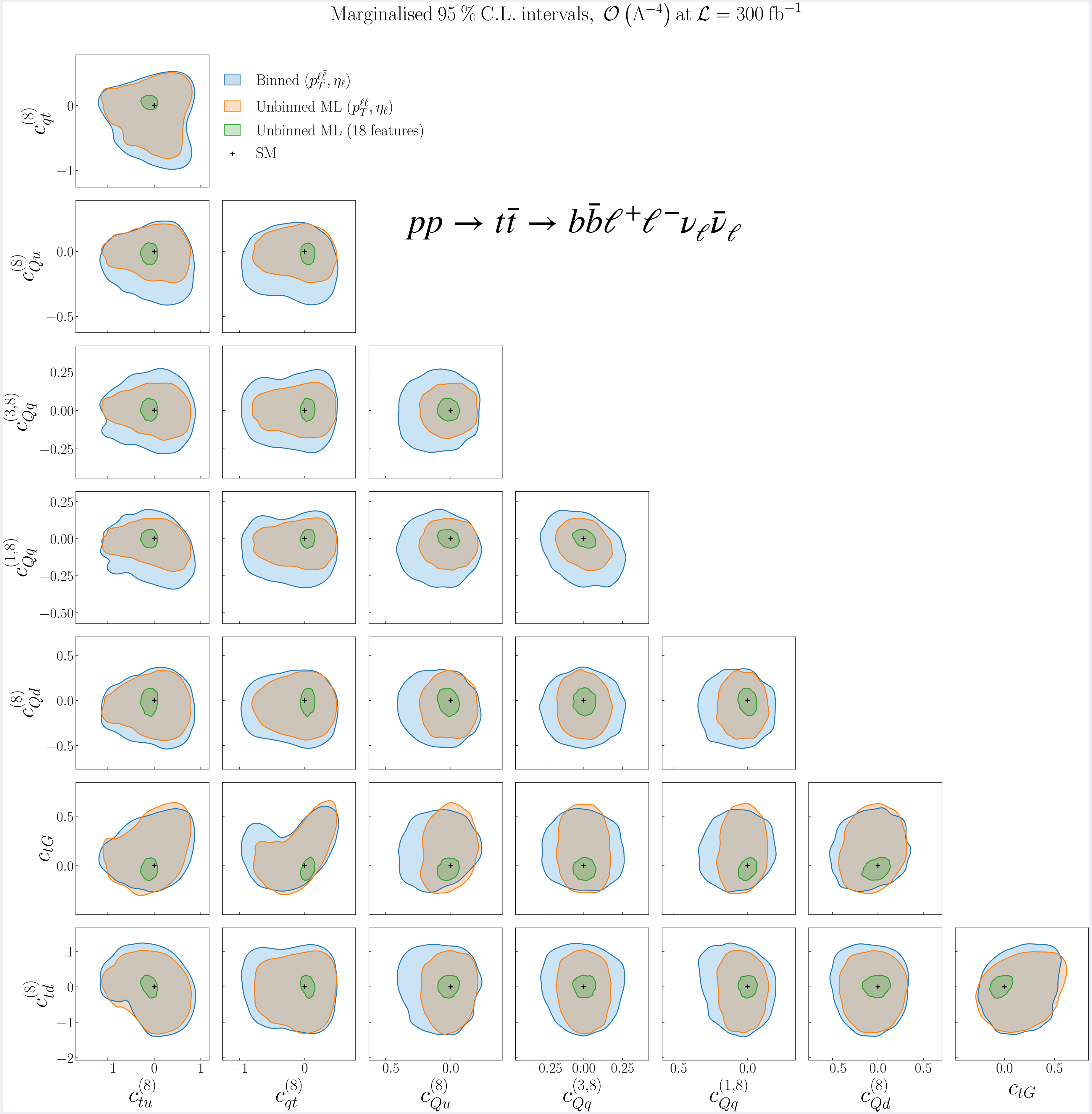
2211.02058

Unbinned multivariate observables

- Combines machine learning regression and classification to parameterise **high-dimensional** likelihood-ratios
- Accounts for **methodological uncertainties** by means of the Monte Carlo replica method
- Scales quadratically with the number of EFT parameters, and can be **fully parallelised**

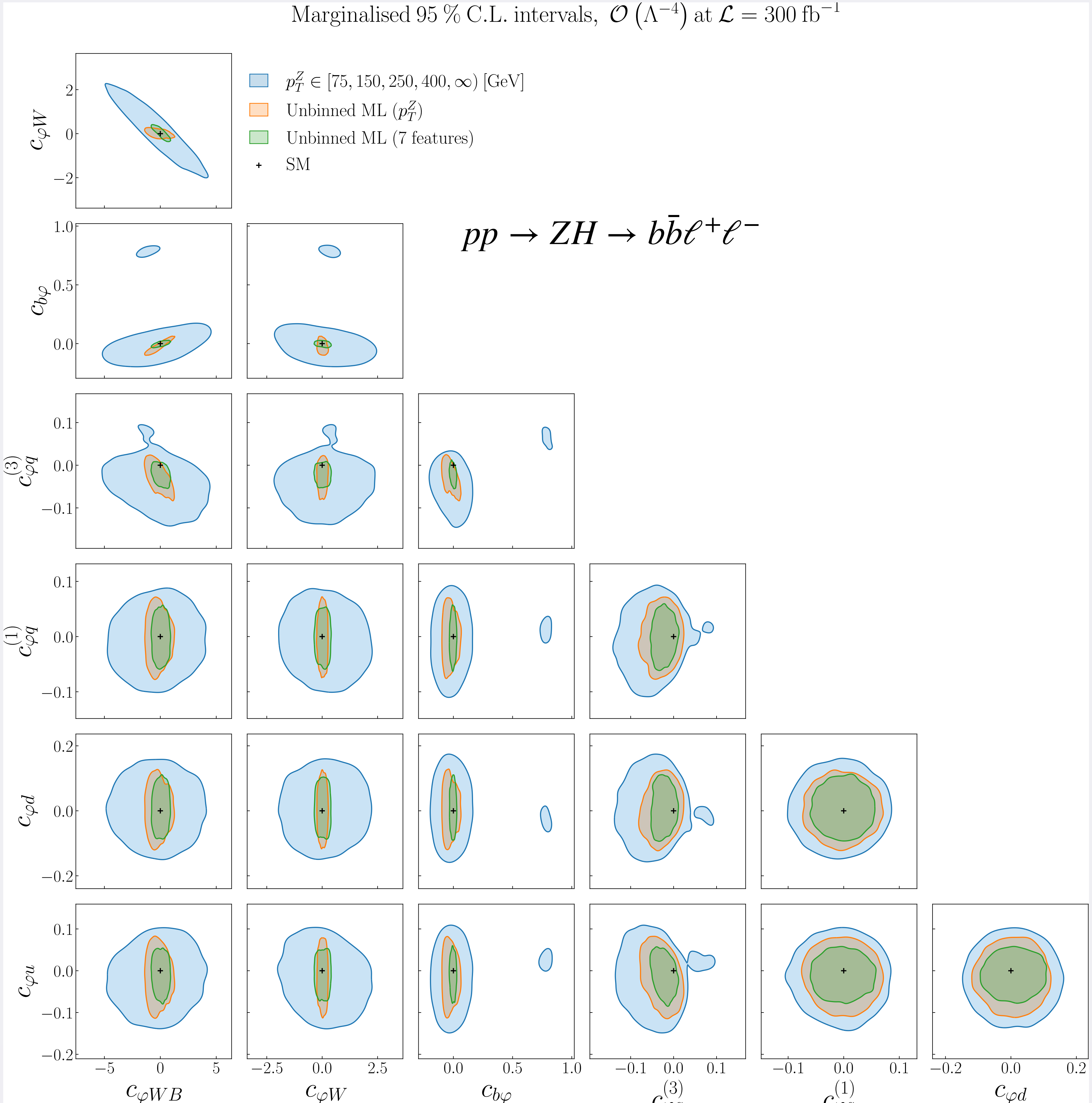






2211.02058

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



2211.02058

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

Action points

- Adopt the **`ML4EFT`** framework in LHC experiments to present to the community unbinned measurements that are **optimal** for global EFT fits
- Compare optimal unbinned observables to existing binned observables to **asses the information loss** incurred by binning
- **`ML4EFT`** is well documented and open-source: make it a community based effort
  - lhcfitnikhef.github.io/ML4EFT

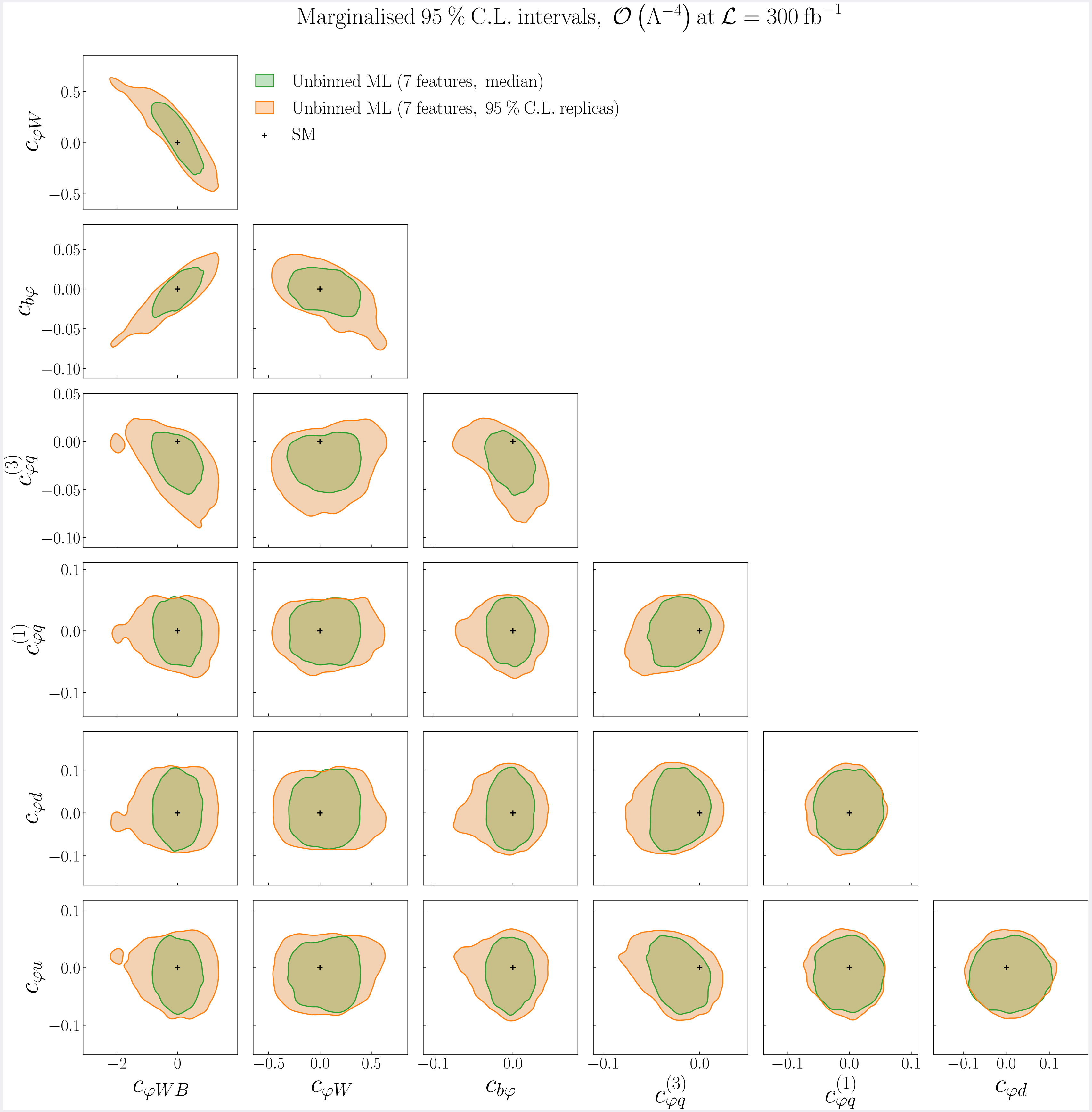
Thank you!

Backup

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

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

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.


Marginalised 95 % C.L. intervals, $\mathcal{O}(\Lambda^{-4})$ at $\mathcal{L} = 300$ fb^{-1}
Unbinned ML (7 features, median)
Unbinned ML (7 features, 95 % C.L. replicas)
SM
$c_{\varphi W}$
$c_{b\varphi}$
$c_{\varphi q}^{(3)}$
$c_{\varphi q}^{(1)}$
$c_{\varphi d}$
$c_{\varphi u}$
$c_{\varphi WB}$