A Monte Carlo analysis of the SMEFT in the top quark sector

Emma Slade[*]

Rudolf Peierls Centre for Theoretical Physics, University of Oxford,
Clarendon Laboratory, Parks Road, Oxford OX1 3PU, United Kingdom

E-mail: emma.slade@physics.ox.ac.uk

We present a framework for carrying out global analyses of the Standard Model Effective Field Theory: SMEFiT. This approach is based on the Monte Carlo replica method, widely used in the case of NNPDF fits of the proton structure, for deriving a faithful estimate of the experimental and theoretical uncertainties. As a proof of concept of the SMEFiT methodology, we present a study of the constraints on the SMEFT provided by top quark production measurements from the LHC. We derive bounds for the 34 degrees of freedom relevant for the interpretation of the LHC top quark data and compare these bounds with previously reported constraints.

---

[*]Speaker.

Introduction The Large Hadron Collider (LHC) is pursuing an extensive program of direct searches for physics beyond the Standard Model (BSM) by exploiting its unique reach in energy. As well as searching for the direct production of new particles, one can also perform indirect searches, where precise measurements are compared to Standard Model (SM) predictions with the aim of discovering BSM effects via deviations in the tails of SM distributions.

A powerful framework with which we can parameterise these deviations with respect to the SM in a model-independent way is the Standard Model Effective Field Theory (SMEFT) [1, 2, 3]. In the SMEFT, the effects of BSM dynamics at high scales $E \simeq \Lambda$ are parametrised for $E \ll \Lambda$ in terms of higher-dimensional (irrelevant) operators built up from the SM fields and symmetries. Analysing experimental data in the SMEFT framework is non-trivial as even when only considering operators that conserve baryon and lepton number [3], one ends up with over 2000 operators at dimension-6 in the absence of flavour assumptions. This implies that global SMEFT analyses need to explore a complicated parameter space with a large number of degenerate ("flat") directions.

From the methodological point of view, a global fit of the SMEFT from LHC measurements requires combining state-of-the-art theoretical calculations (in the SM and in the SMEFT) with a wide variety of experimental cross-sections and distributions. In this work we develop a novel strategy for global SMEFT analyses. This approach [4], which we denote by SMEFiT, combines the generation of Monte Carlo (MC) replicas to estimate and propagate uncertainties, with methodological techniques such as closure testing and cross-validation.

As a proof of concept of the SMEFiT methodology, we apply it here for the first time to the study of top quark production at the LHC in the SMEFT framework at dimension-6. We include the NLO QCD corrections to the SMEFT contributions and compute both the linear ($\mathscr{O}(\Lambda^{-2})$) and the quadratic ($\mathscr{O}(\Lambda^{-4})$) contributions to the SMEFT predictions. By exploiting the SMEFiT methodology, we derive the probability distribution in the space of SMEFT Wilson coefficients.

The SMEFT framework Let us begin by reviewing the SMEFT formalism [2, 5], with emphasis on its description of the top quark sector. As mentioned above, the effects of new BSM particles with a mass scale $M \simeq \Lambda$ can be parametrised at lower energies $E \ll \Lambda$ in a model-independent way in terms of a basis of operators constructed from the SM fields and their symmetries. The resulting Lagrangian then admits the following expansion

$$\mathscr{L}_{\text{SMEFT}} = \mathscr{L}_{\text{SM}} + \sum_i^{N_{d6}} \frac{c_i}{\Lambda^2} \mathscr{O}_i^{(6)} + \sum_j^{N_{d8}} \frac{b_j}{\Lambda^4} \mathscr{O}_j^{(8)} + \dots, \tag{1}$$

where $\mathscr{L}_{\text{SM}}$ is the SM Lagrangian, and $\{\mathscr{O}_i^{(6)}\}$ and $\{\mathscr{O}_j^{(8)}\}$ stand for the elements of the operator basis of mass-dimension $d = 6$ and $d = 8$, respectively. Operators with $d = 5$ and $d = 7$, which violate lepton and/or baryon number conservation [6, 7], are not considered here. In this work we adopt the Warsaw basis for $\{\mathscr{O}_i^{(6)}\}$ [3], and neglect effects arising from operators with mass dimension $d \geq 8$.

In general, the effects of the dimension-6 operators can be written as follows:

$$\sigma = \sigma_{\text{SM}} + \sum_i^{N_{d6}} \kappa_i \frac{c_i}{\Lambda^2} + \sum_{i,j}^{N_{d6}} \widetilde{\kappa}_{ij} \frac{c_i c_j}{\Lambda^4}, \tag{2}$$

where σ_{SM} indicates the SM prediction and c_i are the Wilson coefficients we wish to fit.

In Eq. (2), the second term arises from operators interfering with the SM amplitude. The resulting $\mathcal{O}(\Lambda^{-2})$ corrections to the SM cross-sections represent formally the dominant correction, though in many cases they can be subleading. The third term in Eq. (2), representing $\mathcal{O}(\Lambda^{-4})$ effects, are from the squared amplitudes of the SMEFT operators. In principle, this term may not need to be included, depending on whether the truncation at $\mathcal{O}(\Lambda^{-2})$ order is done at the Lagrangian or the cross-section level, but in practice there are often valid reasons to include them in the calculation.

In this work we follow the strategy documented in the LHC Top Quark Working Group note [8]. We adopt the Minimal Flavour Violation (MFV) hypothesis [9] in the quark sector as the baseline scenario. We further assume that the CKM matrix is diagonal, such that the Yukawa couplings are non-zero only for the top and bottom quarks. In other words, we impose a $U(2)_q \times U(2)_u \times U(2)_d$ flavour symmetry in the first two generations. In addition, we restrict ourselves to the CP-even operators, and focus on those operators that induce modifications in the interactions of the top quark with other SM fields. Under these assumptions, we end up with 34 independent degrees of freedom to fit.

Methodology   In this work, we adopt the MC replica method, inspired by the NNPDF fits to parton densities (see Ref. [10] and references therein), to propagate the experimental uncertainties from experimental cross-sections to the fitted SMEFT coefficients $\{c_i\}$. The idea is to construct a sampling of the probability distribution in the space of the experimental data, which then translates into a sampling of the probability distribution in the space of the SMEFT coefficients. This strategy can be implemented by generating a large number of artificial replicas of the original data. The main advantage of the MC method is that it does not make any assumption about the probability distribution of the coefficients, and is not limited to Gaussian distributions. Moreover, it is suited to problems where the parameter space is large and complicated, with a large number of quasi-degenerate minima and flat directions.

In this work, we use as input to all our theory calculations the NNPDF3.1 NNLO no-top PDF set [10]. By removing the top data from the PDF fit, this prevents us double-counting the data both in the PDFs and the SMEFT fits. To account for the removal of the data in the PDF fit, we include PDF uncertainties in the covariance matrix. As we do not currently account for missing higher-order uncertainties, we use NNLO QCD predictions for all available SM processes, and NLO otherwise.

Results   In the following, we present the fit results for the central values $\langle c_i \rangle$, defined as

$$\langle c_l \rangle \equiv \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} c_l^{(k)}, \qquad (3)$$

and the corresponding 95% CL uncertainties, δc_i, for the 34 dimension-6 SMEFT degrees of freedom. In all cases, the number of replicas $N_{\text{rep}} = 1000$. We also study the cross-correlations between these degrees of freedom as they provide an important piece of information since these correlations might be large because of flat directions in the parameter space.

In Fig. 1 we display the best-fit values with the corresponding 95% confidence levels. The dashed line indicates the SM prediction as reference. In the right panel, we show the associated fit
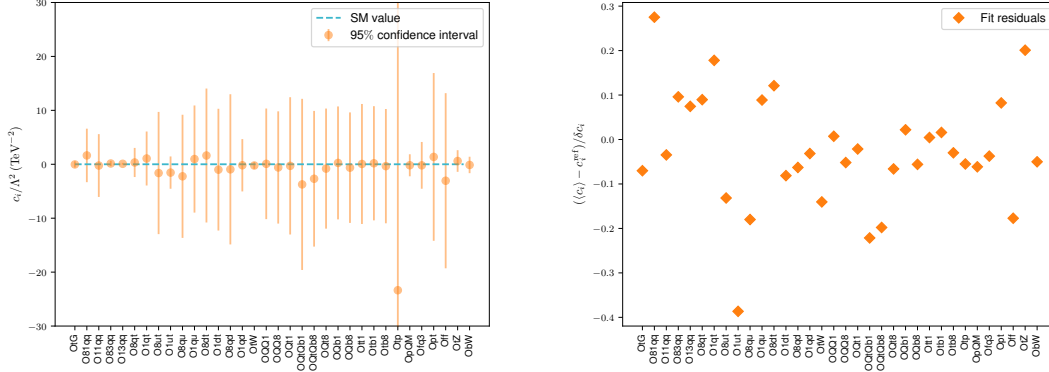
Figure 1: Left: the best-fit values with the corresponding 95% confidence intervals for the degrees of freedom considered in this analysis. Right plot: the associated fit residuals.

residuals r_i,

$$r_i \equiv \frac{(\langle c_i \rangle - c_i^{(\text{ref})})}{\delta c_i}, \tag{4}$$

which measure the deviation of the fit results with respect to the SM in units of the 95% CL uncertainties.

We find that the fit results are in good agreement with the SM within uncertainties. Note that, due to the correlations between the degrees of freedom, the size of the residuals are smaller than in the case where they are completely independent. We have explicitly verified this; see [4] for more details. We also observe that there are a wide range of values for the fit uncertainties for the different degrees of freedom. The origin of these differences is due to the fact that different degrees of freedom are constrained by different processes, and in each case the amount of experimental information varies quite considerably.

The interpretation of the 95% CL uncertainties shown in Fig. 1, requires some care. The reason is that, with the available experimental data, we are not able to fully seperate the independent directions in the parameter space. As a consequence, there will be in general correlations between the fit parameters. To quantify this, in Fig. 2 we show a heat map indicating the values of the correlation coefficient,

$$\rho(c_i, c_j) = \frac{\frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} c_i^{(k)} c_j^{(k)} - \langle c_i \rangle \langle c_j \rangle}{\delta c_i \delta c_j}. \tag{5}$$

between the degrees of freedom constrained from the fit. In this heat map, dark blue regions correspond to degrees of freedom that are significantly correlated, while light green regions are degrees of freedom that are significantly anti-correlated. The effects of such correlations are ignored in fits where these degrees of freedom are constrained individually rather than marginalised from the global fit results, and lead in general to artificially tighter constraints.

We finally show the comparison between our global fit results and the bounds reported in the LHC top WG EFT note, as well as with individual fit results in Fig. 3. We find that for some of the fitted degrees of freedom our bounds are stronger than those reported in previous studies, in some
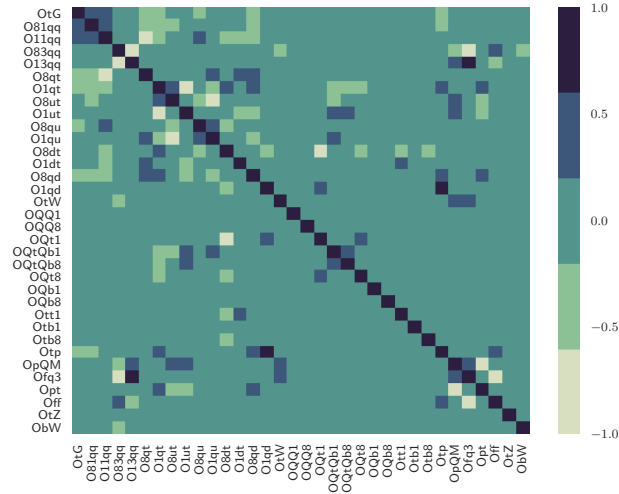
Figure 2: Heat map indicating the values of the correlation coefficient $\rho(c_i, c_j)$ between the 34 fitted coefficients shown in Fig. 1, see text for more details.

cases by nearly one order of magnitude. One can see how the individual bounds are in general tighter or at most comparable to the marginalised ones. This emphasises the point that individual fits to operators produce artifically tight bounds. It is only in global fits, where the correlations are accounted for, that we are able to produce realistic estimates.

Conclusion The study presented in this work is the first proof-of-principle application of the SMEFiT framework. As a proof-of-concept of the SMEFiT framework, we have presented an analysis of top quark production measurements at the LHC at 8 TeV and 13 TeV. Our results are in good agreement with the SM expectations: we find that all the fitted SMEFT degrees of freedom are consistent with the SM result at the 95% CL. In addition, SMEFiT is easily able to cope with enlarging the fitted parameter space by adding new experimental data.

It further allows for the easy addition of new data via Bayesian reweighting [11, 12] without having to rerun fits or have direct access to the SMEFiT code. An open-source reweighting code which allows one to study the impact of new data on the baseline fit presented here will be released soon.

References

[1] S. Weinberg, *Baryon and Lepton Nonconserving Processes*, *Phys. Rev. Lett.* **43** (1979) 1566–1570.

[2] W. Buchmuller and D. Wyler, *Effective Lagrangian Analysis of New Interactions and Flavor Conservation*, *Nucl. Phys.* **B268** (1986) 621–653.

[3] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [arXiv:1008.4884].
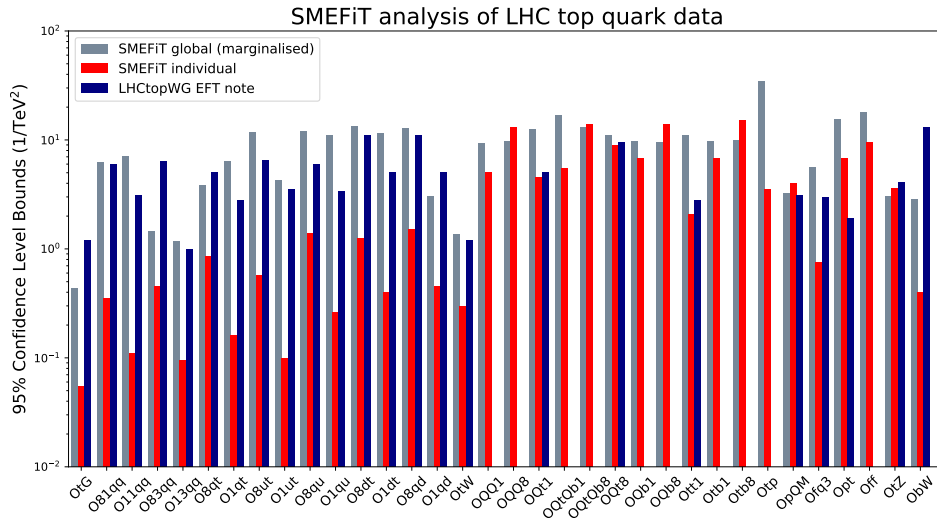
Figure 3: The 95% CL bounds on the 34 degrees of freedom included the present analysis, both in the marginalised (global) and in the individual fit cases, with the bounds reported in the LHC Top WG EFT note [8].

[4] N. P. Hartland, F. Maltoni, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector*, *JHEP* **04** (2019) 100, [arXiv:1901.05965].

[5] S. Weinberg, *Phenomenological Lagrangians*, *Physica* **A96** (1979) 327.

[6] C. Degrande, N. Greiner, W. Kilian, O. Mattelaer, H. Mebane, T. Stelzer, S. Willenbrock, and C. Zhang, *Effective Field Theory: A Modern Approach to Anomalous Couplings*, *Annals Phys.* **335** (2013) 21–32, [arXiv:1205.4231].

[7] A. Kobach, *Baryon Number, Lepton Number, and Operator Dimension in the Standard Model*, *Phys. Lett.* **B758** (2016) 455–457, [arXiv:1604.05726].

[8] J. A. Aguilar Saavedra et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory*, arXiv:1802.07237.

[9] G. D'Ambrosio, G. F. Giudice, G. Isidori, and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, *Nucl. Phys.* **B645** (2002) 155–187, [hep-ph/0207036].

[10] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J.* **C77** (2017), no. 10 663, [arXiv:1706.00428].

[11] **The NNPDF** Collaboration, R. D. Ball et al., *Reweighting NNPDFs: the W lepton asymmetry*, *Nucl. Phys.* **B849** (2011) 112–143, [arXiv:1012.0836].

[12] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data*, *Nucl.Phys.* **B855** (2012) 608–638, [arXiv:1108.1758].