Effective Field Theories for Physics Beyond the Standard Model

ST JOHN'S COLLEGE
UNIVERSITY OF CAMBRIDGE

Maeve Una Madigan

Supervisor: Prof. B.C. Allanach

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This thesis is submitted for the degree of
Doctor of Philosophy

St. John's College August 2021

Declaration

This thesis is based on original research carried out while the author was a graduate student in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, from October 2018 to August 2021. This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared here and specified in the text.

The material in Chapter 3 and Appendices A, B and C is based on work done in collaboration with John Ellis, Ken Mimasu, Veronica Sanz and Tevong You and is published in Ref. [1]. All plots and calculations were produced using the `Fitmaker` public code, which can be viewed at the following `Gitlab` link: https://gitlab.com/kenmimasu/fitrepo. Contributions to this code formed a significant portion of the work of my PhD.

The material in Chapter 4 and Appendices D, E, F and G is based on work done in collaboration with Admir Greljo, Shayan Iranipour, Zahari Kassabov, James Moore, Juan Rojo, Maria Ubiali and Cameron Voisey and is published in Ref. [2]. My contribution to this work excludes the fitting of parton distribution functions; in particular, this includes the details presented in Section 4.3.3 and Appendices D, E and G, which are included in this thesis only for completeness.

The material in Chapter 5 is based on work done in collaboration with B. C. Allanach and Tyler Corbett and is published in Ref. [3]. In this Chapter, I performed all calculations and simulations and created all plots.

Chapters 1 and 2 contain introductory material. These chapters are written by me based on review articles as specified in the text, and adapted from Ref. [3], co-authored with B.C. Allanach and Tyler Corbett.

This thesis is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared here and specified in the text. The material in Ref. [2] and Chapter 4 has previously been submitted for a degree of doctor of philosophy by Cameron Voisey at the University of Cambridge. No other part of this work has been submitted, or is being concurrently submitted, for a

degree or other qualification at the University of Cambridge or any other university or similar institution.

Maeve Una Madigan
August 2021

Abstract

Effective Field Theories for Physics Beyond the Standard Model
Maeve Una Madigan

The wealth of data gathered by the LHC has tested the Standard Model to an unprecedented level, with searches for physics beyond the Standard Model placing strong constraints on new physics. As these constraints become more stringent, it is increasingly likely that physics beyond the Standard Model will be heavy. An effective field theory (EFT) provides a powerful framework for capturing the indirect effects of heavy new physics on low-energy observables. This thesis will explore the use of EFTs in the search for new physics beyond the Standard Model, taking three different perspectives.

Firstly, we will perform a global analysis of the dimension-6 Standard Model Effective Field Theory (SMEFT) using data from the Higgs, top, diboson and electroweak sectors, combining these four data sectors in a global fit for the first time. We will assess the interplay between these sectors, shedding light on the interplay between the Higgs and top sectors in particular. By comparing with previous fits, we will highlight the increase in sensitivity to the SMEFT provided by new data from Run II of the LHC. We will assess the impact of our fit on UV completions of the SMEFT by considering simple single-field extensions of the Standard Model, matched to the SMEFT. Finally, we will scan through all 2-, 3-, 4- and 5-parameter combinations of SMEFT Wilson coefficients, calculating the pull from the Standard Model of each combination. Doing so produces a broad and model-independent search for signs of new physics, complementing our global fit.

Secondly, we will produce a simultaneous determination of the parton distribution functions (PDFs) and the SMEFT using data from the high-mass tails of Drell-Yan distributions. These observables provide crucial constraints on energy-growing four-fermion EFT operators, while also constraining light quark PDFs in the large-x region. By neglecting this overlap and fitting PDFs using Standard Model assumptions, we may be reabsorbing new physics effects into the PDFs. By performing a simultaneous

fit we will assess the interplay between the PDF and EFT effects. Doing so using both LHC data as well as projections for the High Luminosity LHC will allow us to quantify the impact of a consistent joint determination on the EFT constraints, both now and in the high-luminosity regime.

Finally, we will turn to future hadron colliders in the hope that these may be able to directly detect physics beyond the Standard Model. The weak effective field theory has been used to parametrise hints at indirect evidence for new physics from measurements of rare B meson decays at experiments such as LHCb. The resulting shifts of the Wilson coefficients of the weak effective theory can be reproduced by extending the SM by a scalar leptoquark with flavour non-universal couplings to quarks and leptons. We will create projections for such a leptoquark, addressing the question: if a scalar leptoquark exists and can provide a solution to the neutral current B anomalies, could it be detected at future hadron colliders? By creating projections for the discovery potential of the High-Luminosity LHC and Future Circular Collider, we hope to strengthen the case for these future colliders.

Acknowledgements

To begin, I would like to thank Ben Allanach for his guidance as my PhD supervisor. His advice and insights into physics and academia in general have been invaluable, and I am grateful for his words of support and motivation throughout my PhD.

Thank you to Maria Ubiali for her guidance and for involving me in her research - working together has been a highlight of my PhD, and I look forward to more work with the PBSP team.

Over the course of my PhD I have been lucky to work with many incredible people. A special thank you to John Ellis, Ken Mimasu, Veronica Sanz and Tevong You for guiding me into the world of effective field theories: working together has been both inspiring and fun. Thank you to everyone who worked on Ref. [2]: Admir Greljo, Shayan Iranipour, Zahari Kassabov, James Moore, Juan Rojo, Maria Ubiali and Cameron Voisey; in particular to Cameron Voisey for his patience while benchmarking k-factors. Finally, thank you to Tyler Corbett and Joseph Tooby-Smith for sharing their knowledge and insights during our work together.

My time as a PhD student has been shaped by my experience at DAMTP and the Cavendish Laboratory. Thank you to everyone in the Cambridge Pheno Working Group for their advice and interesting discussions each week, including: Hannah Banks, Herschel Chawdhry, Joe Davighi, Ben Gripaios, Ward Haddadin, Chris Lester, Scott Melville, Mathieu Pellen, René Poncelet, Andrei Popescu, Rupert Tombs, Joseph Tooby-Smith, Bryan Webber and many others. Thank you to Manda Stagg for her work in the HEP group, in particular in giving Pavilion B its warm and welcoming atmosphere. A similar thank you to all of the students and staff at DAMTP - I won't try to name you all, but I am grateful for the lunches, coffees, seminars and treasure hunts we've spent together. Thanks to João Melo for comments on the renormalisability of this thesis, and to Philip Clarke for proof reading.

A huge thank you to all of the friends who've supported me and made these last few years so enjoyable. Thank you to all of the friends I've met through the SBR at St John's College, in particular to my old housemates Alva Casey, Matt Coulter and Manuela Zimmerman for entertaining me through a few lockdowns. Thank you

to Joseph Tooby-Smith and Ward Haddadin for making me laugh each Wednesday morning; to Roxana Rosca-Mead for making Christmas happen; to Christophe Gillain for his words of wisdom; to João Melo for great advice; to Julia Kenny for the music and smiles; to Philip Clarke for endless support and long walks; to Sarah McManus for her wholesome words of kindness.

Finally, thank you to my family for their constant support, especially in this last year. Thank you to Kieran, Eoin and Jo for brightening my days with good humour and dog photos, and to Joan and Eamonn whose love, support and kindness have been so important to me during my time at Cambridge.

Table of contents

List of figures xiii

List of tables xxv

Nomenclature xxxi

1 Introduction 1

1.1 Thesis layout . . . 2

2 Effective Field Theories 3

2.1 The Effective Field Theory Formalism . . . 3

2.1.1 Top-down . . . 5

2.1.2 Matching . . . 6

2.1.3 Renormalisation . . . 8

2.2 The Standard Model Effective Field Theory . . . 9

2.2.1 SMEFT operators and bases . . . 10

2.2.2 Dimension-5 . . . 12

2.2.3 Dimension-6 . . . 13

2.3 Global SMEFT fits . . . 17

2.3.1 Electroweak input scheme . . . 20

2.3.2 Linear vs quadratic . . . 20

2.3.3 Statistical methods . . . 21

2.4 New physics at high-p_T . . . 25

2.4.1 Direct searches . . . 25

2.4.2 Indirect searches . . . 26

2.4.3 Parton distribution functions . . . 27

2.5 Evidence for new physics in B meson decays . . . 29

2.5.1 Neutral current B anomalies . . . 29

2.5.2 Effective field theory interpretation . . . 31

2.5.3 Leptoquarks 32

3 Top, Diboson, Higgs and Electroweak fit to the SMEFT 37
3.1 Introduction 37
3.2 Fitting procedure 40
3.3 Data 42
3.3.1 Dataset description 42
3.3.2 Signal strengths 44
3.4 SMEFT Framework 45
3.4.1 Flavour-universal scenario 45
3.4.2 Top-specific flavour scenario 46
3.5 SMEFT Predictions 49
3.5.1 General strategy 49
3.5.2 Higgs production 52
3.5.3 Gauge boson and Higgs decays 54
3.5.4 Diboson data 56
3.5.5 Top data 56
3.6 Global results 57
3.6.1 Higgs, Diboson and Electroweak fit 57
3.6.2 Top fit 63
3.6.3 Combined Top, Higgs, Diboson and Electroweak Fit 66
3.7 Constraints on UV Completions 82
3.7.1 Simple tree-level-induced SMEFTs 82
3.7.2 Tree-level SMEFT patterns 87
3.7.3 R-parity-conserving stop squarks at the 1-loop level 91
3.7.4 Survey of combinations of multiple operators 92
3.8 Concluding discussion 95

4 Parton Distributions in the SMEFT from high-energy Drell-Yan tails 99
4.1 Introduction 99
4.2 SMEFT benchmark scenarios 101
4.2.1 Benchmark I: oblique corrections $\hat{W}$ and $\hat{Y}$ 101
4.2.2 Benchmark II: left-handed muon-philic lepton-quark interactions 104
4.3 Experimental data, theory predictions, and fit settings 105
4.3.1 Experimental data 105
4.3.2 Theoretical predictions 108
4.3.3 Methodology for the simultaneous PDF and EFT fits 113

4.4 Results . . . 119
4.4.1 PDF and EFT interplay in current high-mass Drell-Yan data . . 120
4.4.2 EFT constraints on scenario II from current high-mass Drell-Yan data . . . 126
4.4.3 On the EFT interpretation of high-mass dilepton searches . . . 128
4.4.4 Overview of current constraints . . . 132
4.5 Projections for the High-Luminosity LHC . . . 133
4.5.1 Generation of HL-LHC pseudo-data . . . 134
4.5.2 Impact on PDF uncertainties . . . 137
4.5.3 PDF and EFT interplay at the HL-LHC . . . 138
4.6 Concluding discussion . . . 146

5 Leptoquarks at Future Hadron Colliders 149
5.1 Introduction . . . 149
5.2 Standard Model Backgrounds . . . 151
5.2.1 Methodology . . . 151
5.2.2 Validation of Standard Model background simulations . . . 156
5.2.3 Future collider Standard Model backgrounds . . . 158
5.3 Projections of collider sensitivity . . . 162
5.3.1 Leptoquark signal simulations . . . 162
5.3.2 Statistical analysis . . . 163
5.3.3 Validation . . . 165
5.4 Results . . . 166
5.4.1 Wide resonances . . . 168
5.5 Concluding discussion . . . 170

6 Conclusions and future directions 175

Appendix A Datasets 179

Appendix B Numerical fits with nested sampling 185

Appendix C Tables of numerical fit results 189

Appendix D Detailed SM PDF comparisons 193

Appendix E Fit quality for SM and SMEFT PDFs 197

Appendix F Validation of the SMEFT K-factors 203

Appendix G Flavour dependence of the SMEFT PDFs **211**

References **213**

List of figures

2.1.1 Muon decay $\mu^- \to e^- \nu_\mu \bar{\nu}_\mu$ takes place through a $W^\pm$ boson interaction in the SM (left). In the limit of $p^2 \ll m_W^2$ this can be approximated by a four-fermion interaction (right). 7

2.2.1 Higgs production through associated vector production (VH) and vector boson fusion (VBF) can be modified by nonzero values of the Wilson coefficients C_{HD} and $C_{H\Box}$. The black circles denote their insertion into the Z-Z-H vertex. 15

2.2.2 Higgs production through gluon fusion can be modified by the operator $\mathcal{O}_{HG}$. The presence of this operator means that the leading contribution in the SMEFT is at tree-level, as shown by the diagram on the left. The leading contribution in the SM is shown on the right and proceeds through a top quark loop. 15

2.2.3 On the left, Higgs production in association with a $t\bar{t}$ pair can be modified by the operator $\mathcal{O}_{uH}$. On the right, the coupling of Z to a l^+l^- pair is modified by the insertion of operators such as $\mathcal{O}_{Hl}^{(1)}$ and $\mathcal{O}_{Hl}^{(3)}$. 17

2.3.1 Many operators in the dimension-6 SMEFT contribute to various LHC observables including Higgs data, top quark data, diboson production and EWPO and will benefit from a global fit incorporating these datasets simultaneously. This diagram is taken from Ref. [1] and depicts the dimension-6 operators constrained in the global fit of Chapter 3. 18

2.3.2 Constraints on the SMEFT Wilson coefficients C_{HWB}, C_{HD} (where $\Lambda = 1$ TeV) from Higgs data and electroweak precision observables. These constraints are produced using the `Fitmaker` code, which will be introduced in Chapter 3. 19

2.4.1 The leading contribution to the Drell-Yan $pp \to l^+l^-$ process. The cross section factorises into the non-perturbative parton distribution functions describing the proton constituents, and the hard scattering process $q\bar{q} \to l^+l^-$. 28
2.5.1 Diagrams showing the decay $B^0 \to K^0\mu^+\mu^-$ in the SM (left) and through the insertion of a dimension-6 operator of the weak effective theory (right). The leading contribution from the SM is at 1-loop, whereas a dimension-6 operator allows this process to take place at tree-level. 30
2.5.2 Leptoquark S_3 mediating a tree-level $b \to s\mu^+\mu^-$ transition. 33
2.5.3 Leptoquark production mechanisms at hadron colliders. The dominant production mechanism is pair production (left) through gluon-gluon fusion. This mechanism is independent of the couplings of the leptoquark to leptons and quarks. In contrast, single production (right) always occurs in association with a lepton. 34

3.5.1 Illustration of the effects of selected SMEFT operators on representative gluon fusion simplified template cross-sections σ_i relative to the corresponding SM cross-sections σ_{SM}, for $C_i/\Lambda = 1$ TeV^{-2}. The various Higgs p_T and m_{jj} bins provide complementary sensitivities and hence discriminating power between the operators. 53
3.6.1 Constraints on the indicated individual and marginalised operator coefficients $C_i(1\text{ TeV})^2/\Lambda^2$ (top and third panels) and the corresponding scales Λ for the indicated values of the C_i at the 95% confidence level (second and bottom panels), from a combined linear fit to the Higgs, diboson and electroweak precision observables. In the top two panels, the bars show the 95% CL ranges from the LHC Run 1 and early Run 2 data (green), current data without using the STXS measurements (grey), and current data including STXS using either the on-shell vector boson approximation or the full $1 \to 4$ matrix elements for the 4-fermion Higgs decay modes taken from Ref. [4] (blue). In the bottom two panels, the corresponding marginalised results are indicated by yellow, orange and red bars, respectively. We also show in purple in the individual case (grey in the marginalised case) the effect of dropping the ATLAS Z+jets measurement. 61

3.6.2 Constraints on the indicated individual and marginalised operator coefficients at the 95% confidence level (upper and lower figures, respectively), from a combined linear fit to the top data and electroweak precision observables. The impact of $t\bar{t}$ data is highlighted by the evolution of the constraints starting from no $t\bar{t}$ data (green) adding Run 1 $t\bar{t}$ total and differential cross-section data (grey/yellow), the corresponding Run 2 $t\bar{t}$ data (blue/orange), and finally $t\bar{t}$ asymmetry measurements A_{FB} from the Tevatron and A_C from the LHC (purple/red). 62

3.6.3 Upper panels: Breakdown of the impacts of various top datasets on the 2-dimensional subspace of the four-fermion operators $(C_{Qq}^{1,8}, C_{tq}^{8})$, setting all other operator coefficients to zero. Lower panels: Similar breakdown of the impacts of various top datasets on the 2-dimensional subspace of the four-fermion operators (C_{Qu}^{8}, C_{Qd}^{8}), setting all other operator coefficients to zero. Each panel on the RHS zooms in on the combined constraints of the LHS. 63

3.6.4 Results from the global fit to the electroweak, diboson, Higgs and top data in the top-specific $SU(2)^2 \times SU(3)^3$ scenario. Top two panels: fits to the individual operators, showing the 95% CL ranges for the operator coefficients C_i normalising the new physics scale Λ to 1 TeV, and the ranges for the scales Λ for different values of the C_i. Bottom two panels: similar, but marginalising over the other operators. In the top panels, fit results found using only electroweak, diboson and Higgs data are shown in green, those found in the combination with the top data are shown in blue, and those using only top data are shown in purple. In the bottom panels, the corresponding marginalised fit results are shown in yellow, red and pink. 68

3.6.5 Constraints on the indicated pairs of operator coefficients at the 95% confidence level, marginalised over the remaining degrees of freedom in the 'Higgs only' operator set. The green and mauve shaded areas correspond to combined linear fits to Higgs data and Higgs + top data, respectively. The blue ellipses indicate the marginalised constraints from Higgs + top data after introducing top-quark four-fermion operators into the fit, and the yellow ellipses are obtained from a fit dropping the $t\bar{t}H$ data. 70

3.6.6 Constraints on the indicated pairs of operator coefficients at the 95% confidence level, setting the other operator coefficients to zero. The shaded regions correspond to linear fits to Higgs signal strengths and 0 jet STXS bins (blue), $t\bar{t}H$ signal strengths (mauve), ≥ 1 jet STXS bins (orange) $t\bar{t}$ data (green), $t\bar{t}V$ data (red) and their combination (grey). The dashed ellipses show the constraints obtained by marginalising over the remaining Wilson coefficients of the full fit. 71

3.6.7 The measurement of the $t\bar{t}$ invariant mass distribution in the lepton+jets channel at 13 TeV by the CMS experiment [5] compared to to the SM prediction at the NNLO QCD + NLO EW level [6]. Also shown are predictions corresponding to the best-fit values for $\hat{C}_G$ (green upward triangles), $\hat{C}_{tG}$ (red downward triangles), their combination (purple circles) and the global best-fit point in the full parameter space (orange stars). 73

3.6.8 Same as Fig 3.6.5, but setting $C_G = 0$. 75

3.6.9 Comparison of the constraints on the indicated marginalised operator coefficients $C_i(1\text{ TeV})^2/\Lambda^2$ (top and third panels) and the corresponding scales Λ for the indicated values of the C_i at the 95% confidence level (second and bottom panels), found in a combined linear fit to the Higgs, diboson and electroweak precision observables (top two panels) and including in addition top data (bottom two panels), including C_G in the fit (orange) and setting $C_G = 0$ (red). Also displayed in yellow in the top two panels is a fit without LEP (EWPO and WW) measurements. 76

3.6.10 The full 34×34 correlation matrix for the marginalised top-specific fit. The operators are grouped into those affecting primarily electroweak precision observables, bosonic observables, Yukawa measurements and top electroweak measurements, as well as top-quark four-fermion operators. The entries in the correlation matrix are colour-coded according to the indicated magnitudes of the correlation coefficients. 78

3.6.11 Constraints on the eigenvectors of the global likelihood function. The top panel shows the strengths of the global constraints on the eigenvectors. The rows in the centre panel correspond to the different operator coefficients, the rows correspond to the different eigenvectors, and the colour-coded squares represent the moduli-squared of the operator components to the eigenvectors. The bottom panel shows the constraining powers of the electroweak precision data, LEP diboson data, Higgs coupling strength measurements from Runs 1 and 2, STXS , LHC diboson and Zjj, $t\bar{t}$, single-top and $t\bar{t}W/Z/\gamma$ measurements, respectively. Instances where there is no significant constraining power are indicated by "-". Explicit expressions for the eigenvectors are given in Table C.0.3 81

3.7.1 The horizontal bars show the mass limits (in TeV) at the 95% CL for the models described in Table 3.7.1, setting the corresponding couplings to unity. The coupling limits obtained when setting the mass to 1 TeV are shown in grey boxes. We also note in light blue the pulls that exceed 1-σ. 86

3.7.2 Constraints at the 95% CL on $(C_{H\Box}, C_{tH})$ in the *boson-specific scenario* (upper two panels) and $(C_{Hl}^{(1)}, C_{He})$ in the *lepton-specific scenario* (bottom panel). The lines correspond to the $2-\sigma$ limits obtained when we restrict the operators to the relations generated by integrating out the indicated single-field extensions of the SM. 89

3.7.3 Constraints at the 95% CL on $(C_{Hq}^{(1)}, C_{Hu})$ in the *quark-specific scenario* and $((C_{Hq}^{(1)})_{33}, C_{tH})$ in the *top-specific scenario*. The lines correspond to the $2-\sigma$ limits obtained when the operator coefficients are restricted to the relations generated by integrating out single-field extensions of the SM. 90

3.7.4 Limits from the global fit in the stop parameter plane, $(\frac{X_t}{m_{\tilde{t}}}, m_{\tilde{t}},)$. The two panels correspond to the low and high $\tan\beta$ choices, 1 and 20 respectively. 91

3.7.5 Stacked histograms of the distribution of pulls obtained in fits to 2 (upper left), 3 (upper right), 4 (lower left) and 5 (lower right) parameter subsets. The subsets have been split into three categories: those that include only operators that affect $t\bar{t}$ production (blue), those that do not include operators that affect $t\bar{t}$ production (orange), and the rest (green). The dashed vertical lines mark the expected 95% ranges for the pull distributions. 93

3.7.6 The percentage occurrence of a given operator in the combinations ranked in the top 10% of pulls found in fits to 2 (light blue), 3 (orange), 4 (green) and 5 (blue) parameter subsets. 93

4.3.1 Comparison of the CMS Drell-Yan 13 TeV data with the corresponding theoretical calculations at different perturbative orders as a function of the dilepton invariant mass $m_{\ell\ell}$ in the dielectron (left) and dimuon (right panel) final states. The bottom panels display the ratio of the theory calculations to the central value of the experimental data. We display the sum in quadrature of the experimental uncertainties, and the error band in the theory predictions correspond to the one-sigma PDF uncertainties. 110

4.3.2 Comparison between the (relative) experimental uncertainties and the corresponding EFT corrections, $K_{\rm EFT}(\hat{W},\hat{Y})-1$ in Eq. (4.3.9), for the ATLAS 7 TeV, CMS 8 TeV, and CMS 13 TeV Drell-Yan $m_{\ell\ell}$ distributions, for two representative values of $\hat{W}$ and $\hat{Y}$. 114

4.3.3 Comparison between the baseline SM PDF set of this work, labelled "DIS+DY", with the corresponding fit without high-mass DY data. We show results at $Q = 100$ GeV for PDFs normalised to the central value of the baseline (upper) and for the relative PDF uncertainties (lower panels). In the latter case, we also display the PDF uncertainties from the DIS-only fit. 116

4.4.1 The values of $\Delta\chi^2$, Eq. (4.3.13), obtained for the SMEFT PDFs (thus using the $\chi^2_{\rm eftp}(\mathbf{c}_i)$ values) as a function of $\hat{W}_i$ (left) and $\hat{Y}_i$ (right panel) in the sampling ranges of Eq. (4.4.1) together with the corresponding parabolic fits. The error bars indicate the finite-size uncertainties and the horizontal line corresponds to the $\Delta\chi^2 = 4$ condition defining the 95% CL intervals. The red cross indicates the SM expectation, $\hat{W} = \hat{Y} = 0$. 121

4.4.2 Comparison between the results of the parabolic fits to $\Delta\chi^2$, Eq. (4.4.2), for the $\hat{W}$ (left) and $\hat{Y}$ (right panel) parameters for either the SMEFT PDFs ($\chi^2_{\rm eftp}$, already displayed in Fig. 4.4.1) or the SM PDFs (hence with $\chi^2_{\rm smp}$). The insets zoom on the region close to $\Delta\chi^2 \simeq 0$. 122

4.4.3 Comparison between the SM PDF luminosities with their SMEFT counterparts, displayed as ratios to the central value of the SM luminosities, for representative values of the $\hat{W}$ (upper) and $\hat{Y}$ (lower panel) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in Table 4.4.1. 125

4.4.4 The R_{χ^2} estimator, Eq. (4.4.5), normalised to its SM value, as a function for $m_{\ell\ell}^{(\max)}$ for representative values of $\hat{W}$ (left) and $\hat{Y}$ (right panel). We display the results obtained both with SM and SMEFT PDFs, with the horizontal line indicating the reference SM value of R_{χ^2}. 127
4.4.5 The results of polynomial fits to $\chi^2\left(\mathbf{C}_{33}^{Dl}\right)$, Eq. (4.4.2), in scenario II. This χ^2 includes only the contributions from the two DY measurements in the dimuon final state. We display results for fits based on cross sections that account only for the linear, only for the quadratic, and for both linear and quadratic terms in the EFT expansion, in all cases using the baseline SM PDF set. The inset displays the fit to the linear EFT values with an enlarged x-axis range. 128
4.4.6 The data (number of events per 10 GeV bin) from the ATLAS Z' search from [7] in the di-electron (left) and di-muon (right) channels. We also display the theoretical predictions associated to the contributions from Drell-Yan and from the rest of the backgrounds, taken from the ATLAS publication. 129
4.4.7 Comparison between the results of the parabolic fits to the ATLAS search data [7] for $\hat{W}$ (left) and $\hat{Y}$ (right panel) when using the SMEFT PDFs ($\chi^2_{\rm eftp}$) as compared to the SM PDF baseline ($\chi^2_{\rm smp}$). The insets zoom on the region close to $\lambda(\hat{W}), \lambda(\hat{Y}) \simeq 0$. 131
4.4.8 Overview of the results obtained in this section concerning the EFT parameters $\hat{W}$ and $\hat{Y}$ (in scenario I) and $\mathbf{C}_{33}^{Dl}$ (in scenario II). We compare the 95% CL bounds derived in [8] with those obtained in this work from the high-mass DY cross section measurements (Table 4.4.1) and from the ATLAS Z' search data (Table 4.4.2), in both cases displaying the results obtained with either the SM or the SMEFT PDFs. In the former case, we indicate the results that account for PDF uncertainties; these are included by construction for the SMEFT PDFs. 133
4.5.1 Top panels: comparison of the projected HL-LHC pseudo-data for high-mass neutral-current Drell-Yan in the dielectron (left) and dimuon (right) final states as a function of $m_{\ell\ell}$ with the corresponding theory predictions obtained from the SM PDF baseline. The theoretical predictions, generated according to Eq. (4.5.3), are accompanied by their corresponding PDF uncertainties (green bars). Lower panels: the percentage statistical and systematic uncertainty in each $m_{\ell\ell}$ bin of the HL-LHC pseudo-data. 135

4.5.2 Same as Fig. 4.5.1 for charged-current Drell-Yan in bins of the transverse mass m_T. 135
4.5.3 Impact of the HL-LHC pseudo-data on the quark-antiquark luminosity $\mathcal{L}_{q\bar{q}}$ of the SM PDF baseline fit as a function of m_X. Left: the luminosities $\mathcal{L}_{q\bar{q}}$ for the DIS+DY baseline and the corresponding fits including the HL-LHC pseudo-data, either only NC or also with CC cross sections, presented as a ratio to the central value of the former. Right: the relative PDF uncertainty in $\mathcal{L}_{q\bar{q}}$ (with the central value of the DIS+DY baseline as reference) for the same fits. 137
4.5.4 The 95% confidence level contours in the $(\hat{W},\hat{Y})$ plane obtained from the DIS+DY fits that include the high-mass Drell-Yan HL-LHC pseudo-data (both in the NC and CC channels) when using either SM PDFs (blue) or conservative SM PDFs (green). In both cases the ellipses are obtained by performing a parabolic fit to $\chi^2_{\rm smp}$ with fixed PDFs. PDF uncertainties are included in the solid lines and not included in the dashed lines. The results are compared to those obtained in a simultaneous fit, namely with SMEFT PDFs (orange). In this case, the parabolic fit is performed to $\chi^2_{\rm eftp}$ by varying simultaneously the Wilson Coefficients and the PDFs. The crosses indicate the best fits in the three cases discussed in the text. 141
4.5.5 Same as Fig. 4.5.3, now comparing the quark-antiquark SM PDF luminosity in the fits including the HL-LHC pseudo-data with those obtained in the SMEFT PDF fits for representative values of the $\hat{W}$ and $\hat{Y}$ parameters. The corresponding comparison in the case of fits to available Drell-Yan data was shown in Fig. 4.4.3. 142
4.5.6 Same as Fig. 4.4.4 now for the fits including the HL-LHC pseudo-data. Note that in this case the $m_{\ell\ell}^{(\rm max)}$ cut refers to the transverse mass m_T for the charged-current distributions. 143
4.5.7 Left: the values of $\Delta\chi^2$ obtained for the SMEFT PDFs as a function of $\mathbf{C}_{33}^{D\mu}$ from the fits including the HL-LHC pseudo-data, together with the corresponding quartic polynomial fit. Right: comparison of the polynomial fit obtained with SMEFT PDFs and displayed in the left panel with its counterpart based on SM PDFs. 144
4.5.8 Same as Fig. 4.5.5 now in the case of the simultaneous fits of the PDFs and the $\mathbf{C}_{33}^{D\mu}$ EFT parameter taking into account HL-LHC pseudo-data. 145

4.6.1 Overview of the 95% CL bounds obtained in this work on the EFT parameters $\hat{W}$ (left), $\hat{Y}$ (middle), and $\mathbf{C}_{33}^{D\mu}$ (right panel) based on either the SM PDFs or the SMEFT PDFs. Both PDF and methodological uncertainties are accounted for in the bounds, when available. 146

5.1.1 A scalar leptoquark will contribute to $B_s - \bar{B}_s$ mixing at one-loop order. 150
5.2.1 There are four contributions to the SM background of the $\mu^+\mu^- jj$ signal. Here we show a representative Feynman diagram for each process. . . . 153
5.2.2 Validation plot showing our simulations of the $M_{\mu\mu}$ distribution of the SM backgrounds in the search for the pair production of second generation LQs in the $\mu^+\mu^- jj$ channel at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} in the preselection region. We compare our simulations to the ATLAS simulations and data for validation. 157
5.2.3 Validation plot showing our simulations of the $m_{\min}(\mu, j)$ distribution of the SM backgrounds in the search for the pair production of second generation LQs in the $\mu^+\mu^- jj$ channel at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} in the signal region. We compare our simulations to the ATLAS simulations and data for validation. 158
5.2.4 Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 13 TeV LHC with full Run II integrated luminosity of $\mathcal{L} = 140$ fb^{-1}. The LQ signal corresponds to $m_{\text{LQ}} = 1.2$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 5σ relative to the SM background. 160
5.2.5 Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 14 TeV HL-LHC with $\mathcal{L} = 3$ ab^{-1}. The LQ signal corresponds to $m_{\text{LQ}} = 1.8$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 7σ relative to the SM background. 160
5.2.6 Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 27 TeV HE-LHC with $\mathcal{L} = 15$ ab^{-1}. The LQ signal corresponds to $m_{\text{LQ}} = 3.5$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 6σ relative to the SM background. 161

5.2.7 Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 100 TeV FCC-hh with $\mathcal{L} = 20$ ab^{-1}. The LQ signal corresponds to $m_{\text{LQ}} = 11$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 3σ relative to the SM background. 161
5.3.1 Validation plot comparing our expected limits at 95% CL on $\sigma \times$ BR for LQ pair production and decay into a $\mu^+\mu^- jj$ final state at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} to the expected limits obtained by ATLAS. 165
5.4.1 Expected limits at 95% CL on $\sigma \times$BR for the pair production of LQs and decay into a $\mu^+\mu^- jj$ final state at the LHC with full Run II integrated luminosity (left) and the HL-LHC (right). Theory curves $\sigma_{\text{NLO}} \times$ BR are calculated for narrow width LQs with couplings chosen to fit the NCBAs. 166
5.4.2 Expected limits at 95% CL on $\sigma \times$BR for the pair production of LQs and decay into a $\mu^+\mu^- jj$ final state at the HE-LHC (left) and the FCC-hh (right). Theory curves $\sigma_{\text{NLO}} \times$ BR are calculated for narrow width LQs with couplings chosen to fit the NCBAs. 167
5.4.3 Predicted 5σ discovery reach (left) and mass exclusion at 95% CL (right) of the HL-LHC, HE-LHC and FCC-hh. Points correspond to the design integrated luminosities of each future collider of $\mathcal{L} = 3, 15, 20$ ab^{-1} respectively. 167
5.4.4 Comparison of the predicted $m_{\min}(\mu, j)$ distribution of LQ signal events for a LQ with large decay width Γ (left), and expected limits at 95% CL on $\sigma \times$ BR for the pair production of wide LQs decaying into a $\mu^+\mu^- jj$ final state at $\sqrt{s} = 27$ TeV, $\mathcal{L} = 15$ ab^{-1} (right). 169
5.4.5 Expected limits at 95% CL on $\sigma \times$ BR for the pair production of LQs with large decay width Γ, decaying into a $\mu+\mu-jj$ final state at $\sqrt{s} = 14$ TeV, $\mathcal{L} = 3$ ab^{-1} (left) and $\sqrt{s} = 100$ TeV, $\mathcal{L} = 20$ ab^{-1} (right). 169
5.5.1 We summarise constraints on the LQ mass and the product of couplings $|(Y_{de})_{32}(Y_{de})_{22}|$ from searches for LQ single and pair production by ATLAS and CMS, as well as constraints from perturbative unitarity. We overlay a curve indicating the LQ parameters which best-fit the NCBAs, taking the $\pm 1\sigma$ range of Eq. 2.5.9. These constraints are compared to our projections for the mass sensitivity of the HL-LHC, HE-LHC and FCC-hh. 173

B.0.1 Samples produced by nested sampling, projected onto 2- and 1-dimensional subspaces of the 7-dimensional parameter space. Along the diagonal we compare the distributions found with and without quadratic SMEFT contributions in blue and purple, respectively. 186
B.0.2 Marginalised 95 % credible intervals for each of the indicated operator coefficients resulting from a fit to Higgs signal strength data using nested sampling. We compare the effects of including quadratic contributions from dimension-6 operators (blue) to the case of linearised SMEFT predictions (purple). 187

D.0.1 Same as Fig. 4.3.3 comparing the baseline SM PDF set used in this work (labelled "DIS+DY") with the global `NNPDF3.1_str` determination. 194
D.0.2 Same as the upper panels of Fig. 4.3.3 comparing now the baseline SM PDF set used in this work (labelled "DIS+DY") with the corresponding DIS-only fit. Note that the comparison between the PDF uncertainties in these two fits was already displayed in the lower panels of Fig. 4.3.3. 196
D.0.3 The gluon-gluon, quark-antiquark, and quark-quark luminosities at $\sqrt{s} = 14$ TeV as a function of m_X for the DIS+DY baseline and for the same fit excluding the all the datapoints in the high-mass DY experimental sets listed in Table. 4.3.2. The top panels display the ratio of luminosities to the central value of the DIS+DY baseline, while the bottom panels compare the relative PDF uncertainties in each case. 196

F.0.1 Left: comparison of the SMEFT K-factors, Eq. (4.3.10), linearised in the EFT parameters, between the numerical and the analytical approaches for $\hat{W} = -10^{-3}$ and the kinematics of the ATLAS 7 TeV DY data. The label "cuts" indicates that we impose acceptance requirements of $p_T^\ell \geq 25$ GeV and $|\eta_\ell| \leq 2.5$ in the numerical (`SMEFTsim`) calculation. Right: relative difference in the K-factors shown in the left panel taking the analytical calculation as a reference. 205
F.0.2 Same as Fig. F.0.1 (also for $\hat{W} = 10^{-3}$) for the EFT K-factors for the double-differential Drell-Yan cross sections in $m_{\ell\ell}$ and $|y_{\ell\ell}|$, for the highest $m_{\ell\ell}$ bin of the CMS 7 TeV DY measurement. We only compute K-factors for $|y_{\ell\ell}|$ satisfying LO kinematics. 206

F.0.3 Left: the EFT K-factor for high-mass charged-current DY production at the HL-LHC as a function of m_T, the transverse mass of the neutrino-lepton pair, for $\hat{W} = -10^{-3}$. Right: the same quantity, now compared to the corresponding k-factors provided by the authors of [9]. 207
F.0.4 The impact of the SMEFT PDFs on R_W (above) and the SMEFT K-factor $K = 1 + \hat{W} R_W$ (below) calculated at the Scenario I benchmark points $\hat{W} = \pm 4 \cdot 10^{-5}$. 208
F.0.5 Same as Fig. F.0.4, in the case of $\hat{Y}$ parameter. 209
F.0.6 The impact of the SMEFT PDFs on $R^{\rm lin}$, $R^{\rm quad}$ (above) and the SMEFT K-factor $K = 1 + C_{33}^{D\mu} R^{\rm lin} + (C_{33}^{D\mu})^2 R^{\rm quad}$ (below) calculated at the Scenario II benchmark points $C_{33}^{D\mu} = \pm 0.014$. 210

G.0.1 Comparison between the SM and the SMEFT PDFs at $Q = 100$ GeV, displayed as ratios to the central value of the SM PDFs, for representative values of the $\hat{W}$ (upper) and of $\hat{Y}$ (lower panels) parameters. We show the gluon, the total quark singlet, the up quark and antiquark, and the down quark and antiquark PDFs. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 95% CL intervals reported in Table 4.4.1. 212

List of tables

2.2.1 Conventions for the field content of the SM. We use indices $i, j \in \{1, 2, 3\}$ to denote families of the leptons and quarks respectively, and L,R to denote left and right-handed fermions. 10

2.2.2 Dimension 6 operators in the Warsaw basis, adapted from Refs. [1] and [10]. Indices $p, q, r, s, t \in \{1, 2, 3\}$ label fermion generations. The operators are grouped by their field content, where X denotes the gauge field strengths, D denotes the covariant derivative, ψ denotes a fermion and L, R describe the chirality of the fermionic operators. Labels L, R of the left and right-handed fermions are suppressed for clarity. The shaded cells indicate operators that break the flavour symmetry $SU(3)^5$. Note that this division of operators into categories based on the fields and covariant derivatives present is basis-dependent. 14

2.4.1 Design centre of mass energies and integrated luminosities of the LHC Run II and future hadron colliders, adapted from Ref. [3]. 26

3.4.1 Four-fermion operators containing at least one third-generation bilinear in the 'dim6top' basis [11] assuming an $SU(2)^2 \times SU(3)^3$ flavour symmetry. The relations of the corresponding Wilson coefficients with those of the Warsaw basis are also shown. The shaded entries indicate operators that are not included in our analysis because significant constraints cannot be obtained from the chosen dataset at leading order and linear level in the EFT expansion, as discussed in the text. 49

3.7.1 Single-field extensions of the SM constrained by our analysis. 83

3.7.2 Operators generated at tree level by the single-field extensions listed in the first column. Each extension depends on a single coupling (see Table 3.7.1) as well as a new physics mass-scale M. The coefficients of the operators are each proportional to the squares of the corresponding coupling λ by the corresponding entry in the Table and divided by M^2. y_t, y_b and y_τ denote the top, bottom and tau Yukawa couplings respectively, v denotes the electroweak scale and α_s denotes the strong coupling. 84

3.7.3 Operator coefficients generated by the tree-level single-field models B, W and TB, which each depend on two couplings a and b, with y_t, y_b and y_τ denoting the top, bottom and tau Yukawa couplings respectively, v denoting the electroweak scale and α_s denoting the strong coupling. The coefficients of all operators are proportional to the corresponding entries in the Table and divided by M^2. 85

4.3.1 The low-mass and on-shell Drell-Yan datasets used in the present study. For each dataset we indicate the experiment, the centre-of-mass energy $\sqrt{s}$, the publication reference, the physical observable, and the number of data points 107

4.3.2 Same as Table 4.3.1 for the neutral-current high-mass Drell-Yan datasets considered in this work. We also indicate the final-state, whether the distribution is 1D (which are differential in the invariant mass, $m_{\ell\ell}$, of the final-state leptons) or 2D (which are differential in both the invariant mass of the leptons, $m_{\ell\ell}$, and in their rapidity, $y_{\ell\ell}$), and the values of $m_{\ell\ell}$ for the most energetic bin. Datasets indicated with **(*)** are used for the first time in this analysis in comparison with [12]. 108

4.3.3 The values of the χ^2 per data point evaluated for the high-mass DY datasets listed in Table 4.3.2, using theoretical predictions computed at different perturbative accuracy. The PDF sets used here are `NNPDF31_nlo_as_0118`, `NNPDF31_nnlo_as_0118` and `NNPDF31_nnlo_as_0118_luxqed` for the NLO QCD, NNLO QCD and NNLO QCD + NLO EW predictions respectively. For CMS 13 TeV, where different final states are available, we indicate the χ^2 values for each of them. For the ATLAS 8 TeV data, we only evaluated the combined NNLO QCD + NLO EW correction, and hence the pure NNLO QCD result is not given. 112

4.4.1 The 68% CL and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the $\Delta\chi^2$ values calculated from either the SM or the the SMEFT PDFs. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction, while the methodological (finite-size) uncertainty is included according to the approached described in §4.3.3. The fourth and fifth column indicate the absolute shift in best-fit values, Eq. (4.4.3) and the percentage broadening of the EFT parameter uncertainties, Eq. (4.4.4), when the SMEFT PDFs are consistently used instead of the SM PDFs. 123

4.4.2 The 68% and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the ATLAS search data of [7] when using either the SMEFT PDFs or their SM counterparts. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction. The fourth and fifth column indicate the absolute shift in best-fit values and the percentage broadening of the EFT parameter uncertainties when the SMEFT PDFs are consistently used instead of the SM PDFs. 132

4.5.1 Same as Table 4.4.1 for the 68% CL and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional ($\hat{W}$,$\hat{Y}$) fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. As in Table 4.4.1, for the SM PDFs we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for. 139

4.5.2 Same as Table 4.5.1 for the 68% and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional ($\hat{W}$,$\hat{Y}$) fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. The input PDF set for the analysis done using fixed SM PDFs (corresponding to the results displayed in the column "SM cons. PDFs") is a conservative PDF set that does not include any of the high-mass distributions or the HL-LHC projections nor the Run I and Run II high-mass dataset listed in Table 4.3.2. The limits obtained from the simultaneous fit of PDFs and Wilson coefficients (corresponding to the results displayed on the column "SMEFT PDFs") are the same as those in Table 4.5.1. 140

4.5.3 Same as Table 4.5.1, now for the $\mathbf{C}_{33}^{Du}$ parameter from EFT benchmark scenario II. 144

5.2.1 Phase space cuts defining the preselection and signal regions at $\sqrt{s} = 13$ TeV. All cuts are applied in the analysis after parton showering and detector simulation. 156
5.2.2 Cuts applied at parton-level to efficiently simulate events at $\sqrt{s} = 13$ TeV. 156
5.2.3 Phase space cuts defining the signal regions in simulations of the 13 TeV LHC and future colliders. All cuts are applied in the analysis after parton showering and detector simulation. 158
5.2.4 Phase space cuts applied at parton-level in `Madgraph5` to efficiently simulate events in the signal region for the 13 TeV LHC and future colliders. 159
5.2.5 Values of the jet cut-off parameter *xqcut* in GeV used in MLM matching of multi-jet events at the 13 TeV LHC and future colliders. All jet matching parameters are found in the presence of the generator cuts summarised in Table 5.2.4. 159
5.5.1 Summary of the expected 5σ discovery sensitivity and expected 95% CL exclusion sensitivity to S_3 for hadron colliders, from LQ pair production. 171

A.0.1 Diboson measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3. 180
A.0.2 Electroweak precision measurements included in the Fitmaker database. 180
A.0.3 Higgs Run 1 measurements included in the Fitmaker database. 181
A.0.4 Higgs Run 2 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3. 181
A.0.5 Top Run 1 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3. 182
A.0.6 Top Run 2 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3. 183
A.0.7 Top Run 2 measurements (continued) included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3. 184

C.0.1 Table of numerical results in Fig. 3.6.4 from the global fit to the electroweak, diboson, Higgs and top data in the top-specific $SU(2)^2 \times SU(3)^3$ scenario. 190
C.0.2 Relative constraining power in percent of different datasets on each coefficient of the global fit individually. Entries below 1% are not displayed. 'SS', $W_{\text{hel.}}$ and tX refer to Higgs signal strength, W-helicity fraction and single top data, respectively. 191

C.0.3 Components of the eigenvectors found in the principal component analysis of the global fit displayed in Fig. 3.6.11. Components with coefficients of magnitude less than 0.05 are omitted. 192

E.0.1 The kinematic cuts applied to the high-mass Drell-Yan datasets listed in Table 4.3.2, compared to those used in NNPDF3.1. 197

E.0.2 The values of the χ^2 per data point for the baseline SM PDF fit, labelled "DIS+DY", and for the corresponding fits based on reduced datasets. Here Eq. (4.3.5) is evaluated using the t_0 prescription. We also include the results obtained using `NNPDF3.1_str` with the kinematic cuts used in this work and summarised in Table E.0.1. Values in italics indicate datasets that do not enter the corresponding fit. 199

E.0.3 Same as Table E.0.2, now comparing the χ^2 values (computed using the t_0 prescription, as above) of the SM PDFs with those of the SMEFT PDFs for different values of the $\hat{W}$ and $\hat{Y}$ parameters in benchmark scenario I, specifically those displayed in Figs. 4.4.3 and G.0.1. 201

Nomenclature

Acronyms / Abbreviations

BSM	Beyond the Standard Model
CC	Charged Current
CKM	Cabibbo–Kobayashi–Maskawa
CL	Confidence Level
DIS	Deep Inelastic Scattering
DY	Drell-Yan
EFT	Effective Field Theory
EW	Electroweak
EWPO	Electroweak Precision Observables
EWSB	Electroweak Symmetry Breaking
FCC-hh	Future Circular Collider
FCNC	Flavour Changing Neutral Currents
FSR	Final State Radiation
ggF	Gluon fusion production mode of the Higgs
GIM	Glashow-Iliopoulos-Maiani
HE-LHC	High Energy mode of the Large Hadron Collider
HEFT	Higgs Effective Field Theory

HL-LHC	High Luminosity mode of the Large Hadron Collider
HM	High-Mass
IR	Infrared
LFU	Lepton Flavour Universality
LQ	Leptoquark
MC	Monte Carlo
NCBA	Neutral Current B Anomalies
NC	Neutral Current
NLO	Next to Leading Order
NWA	Narrow Width Approximation
PDF	Parton Distribution Function
PMNS	Pontecorvo–Maki–Nakagawa–Sakata
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QFT	Quantum Field Theory
SMEFT	Standard Model Effective Field Theory
STXS	Simplified Template Cross Section
UV	Ultraviolet
VBF	Vector Boson Fusion
VEV	Vacuum Expectation Value
WET	Weak Effective Theory
WH	Higgs production in association with a W boson
ZH	Higgs production in association with a Z boson

Chapter 1

Introduction

The Standard Model has been tested to an unprecedented level over Run I and Run II of the LHC. New measurements, such as the first observation of $t\bar{t}H$ production [13], have furthered our knowledge of the Standard Model couplings. Improvements in theoretical calculations and precision measurements have led to the use of the LHC as a precision machine, producing a high-precision measurement of the W boson mass [14], for example. These measurements add to the wealth of existing precision measurements from previous experiments including LEP and the Tevatron. The message we have received again and again from these is that the Standard Model provides an excellent description of physics, at least at the energy scales accessible to the LHC. If new physics beyond the Standard Model (BSM) is to exist, it must be associated with energy scales $\Lambda \gg v$, where $v = 246.22$ GeV denotes the electroweak scale.

Physics beyond the Standard Model must exist. Evidence for this fact comes from observations of dark matter and the baryon asymmetry of the universe, as well as measurements of neutrino oscillations, none of which can be described by the Standard Model. This point has been reinforced in recent years, with hints at BSM physics appearing in measurements of rare B meson decays (the neutral current B anomalies [15]) and in measurements of the muon magnetic moment (the muon $g-2$ [16]). Each a subtle discrepancy with yet-unknown origins, these deviations add fire to the search for new physics.

While the LHC is famous for its discovery of the Higgs boson, this discovery did not come out of the blue: it was preceded by measurements at LEP [17] and the Tevatron [18] which featured subtle excesses in the number of measured events, suggesting the existence of a Higgs boson [19]. This path to discovery may be relevant to the search for new physics: before a *direct* detection in a traditional 'bump-hunt' experiment, we may see subtle deviations from the Standard Model in *indirect* searches

for new physics. An effective field theory provides a powerful theoretical framework for capturing these subtle deviations in a model-independent fashion. Let E denote the energy scale of an observable at the LHC. The effect of a BSM particle at scales $\Lambda \gg E$ on this observable can be well-approximated by integrating out the particle and using an effective field theory to capture its indirect effects.

This thesis will explore the use of effective field theories in the search for new physics. We will investigate the power of LHC data from the Higgs, top, diboson and electroweak sectors in constraining the Wilson coefficients of the SMEFT. As we will see, the constraints we obtain suggest small values of Wilson coefficients, confirming the fact that the type of deviation we are searching for is subtle. Confidence in our theoretical inputs is crucial in such a case, and we will shed light on the question of whether our use of parton distribution functions (PDFs) in computing theoretical predictions may be clouding the search for new physics. In particular, we will investigate the interplay of PDFs and the SMEFT by performing a simultaneous fit of both. Finally, while an effective field theory can successfully provide indirect evidence for a BSM particle, characterising the particle relies on its direct detection at a collider experiment. We will create projections for the discovery of a leptoquark, estimating its discovery potential at future colliders such as the High Luminosity LHC and the Future Circular Collider.

1.1 Thesis layout

The structure of this thesis is as follows. We will begin in Chapter 2 with an introduction to the effective field theory formalism and the SMEFT, followed by an overview of the technical details required by this thesis. In Chapter 3 we will present a global fit of the dimension-6 SMEFT. Furthermore, we will analyse the impact of this global fit on various UV completions of the SMEFT. In Chapter 4 we will continue to focus on EFT fits, this time considering their interplay with PDFs. We will present a simultaneous determination of the PDFs and EFT using a combination of deep inelastic scattering data and Drell-Yan data, including the high-mass tails of Drell-Yan invariant mass distributions. In Chapter 5 we will focus on new physics in the form of a leptoquark, motivated by signs of new physics in rare B meson decays. Finally, in Chapter 6 we will conclude, summarising this thesis and highlighting directions for future work. Appendices A, B and C contain supplementary material associated with Chapter 3, and Appendices D, E, F and G contain supplementary material associated with Chapter 4.

Chapter 2

Effective Field Theories

2.1 The Effective Field Theory Formalism

We will begin with a brief introduction to the effective field theory (EFT) formalism, following the reviews of Refs. [20–27]. An EFT is a powerful tool for describing physical systems which are valid in a well-defined limit. We begin by specifying the dynamical degrees of freedom relevant to the system. These are a set of fields which, along with their derivatives, will form the building blocks of the Lagrangian density $\mathcal{L}_{eff}$. Secondly, the EFT is defined by specifying symmetries. With these in hand we can construct all possible local operators $\mathcal{O}_i$ compatible with our symmetries, forming the Lagrangian as the sum of these with coefficients g_i:

$$\mathcal{L}_{eff} = \sum_i g_i \mathcal{O}_i. \tag{2.1.1}$$

Equation 2.1.1 is a sum of an infinite number of operators. This is where an EFT differs from the picture of a quantum field theory (QFT) usually taught in graduate courses. In constructing a QFT such as the SM, the standard procedure is to remove all operators of dimension $D > 4$; otherwise an infinite number of couplings is required to renormalise the theory.[1] If our theory depends on an infinite number of coefficients, how can it be predictive?

The final ingredient is the power-counting of the EFT. By construction, an EFT provides an *effective* description of physics valid in a particular region of phase space. We will define this region by first defining a cutoff scale Λ, assuming our theory is valid at energies $E \ll \Lambda$. The beauty of the EFT framework can be seen when we treat the

[1] The resulting QFT is sometimes called a *renormalisable QFT* even though, as we will discuss, an EFT is also renormalisable.

EFT as an expansion in the ratio $\delta = E/\Lambda$ and truncate Eq. 2.1.1 at fixed order in the expansion parameter δ. By doing so we form a theory containing a finite number of operators and depending on a finite number of coefficients.

Consider the limit in which $\delta \ll 1$. In this limit, some operators $\mathcal{O}_i$ have a more significant impact on observables than others. A power-counting scheme assigns a fixed order of δ to each operator. Let D_i denote the dimension of operator $\mathcal{O}_i$. To produce a dimensionless action we require $[\mathcal{L}_{eff}] = 4$, implying that the coefficients g_i must have dimension $[g_i] = 4 - D_i$. We define dimensionless coefficients C_i, known as Wilson coefficients, by factoring out the mass dimension of g_i into powers of the scale Λ, giving

$$\mathcal{L}_{eff} = \sum_i \frac{C_i}{\Lambda^{D_i-4}} \mathcal{O}_i. \tag{2.1.2}$$

Consider the insertion of a single operator $\mathcal{O}_i$ into a tree-level Feynman diagram. The operator $\mathcal{O}_i$ will always be accompanied by a factor of $\frac{1}{\Lambda^{D_i-4}}$, and thus the contribution of $\mathcal{O}_i$ to the amplitude will have order $n = D_i - 4$. Similarly, if we have multiple insertions of various operators $\{\mathcal{O}_i\}$, their collective contribution to the amplitude produces a dependence on δ of order

$$n = \sum_i (D_i - 4). \tag{2.1.3}$$

With the power-counting formula Eq. 2.1.3 at our disposal, we can now determine how far in the expansion of Eq. 2.1.2 we must go to perform calculations to order n. If we wish to calculate amplitudes to leading order, $n = 0$, we retain only the operators of dimension $D \leq 4$. In contrast, to work to order $n = 2$ we may expand $\mathcal{L}_{eff}$ up to dimension $D = 6$. At this order, any amplitude may consist of up to two insertions of operators of dimension $D = 5$, or a single $D = 6$ insertion.

Finally, we will rewrite the expansion of $\mathcal{L}_{eff}$ in Eq. 2.1.2 as

$$\mathcal{L}_{eff} = \mathcal{L}_{D\leq 4} + \sum_i \frac{C_i^5}{\Lambda} \mathcal{O}_i^5 + \sum_i \frac{C_i^6}{\Lambda^2} \mathcal{O}_i^6 + ... \tag{2.1.4}$$

The first term, $\mathcal{L}_{D\leq 4}$, consists of all operators contributing at leading order in δ. These are simply the *relevant* and *marginal* operators of the renormalisable QFT Lagrangian. The remaining operators are subleading in n and are considered *irrelevant* operators: as energy E lowers, these operators are suppressed by powers of δ and become less and less important. While the irrelevant operators are forbidden in a renormalisable QFT, their inclusion here makes clear one of the advantages of working with an EFT: it provides

a theory which is valid in a well-defined region, while systematically parametrising the effects of neglecting physics outside of this region. Although in this expression for $\mathcal{L}_{eff}$ we have neglected anything at $\mathcal{O}(\delta^3)$ and above, it is within our power to include these higher order corrections if and when this level of precision is required.

This captures the *bottom-up* EFT construction. We are agnostic as to what lies beyond the region of validity of our EFT, and the Wilson coefficients are the unknown dimensionless coefficients of the operators $\mathcal{O}_i$. This is the approach taken in the construction of the Standard Model Effective Field Theory (SMEFT), as we do not know what particles, if any, lie at high mass scales.

2.1.1 Top-down

We could take an alternative *top-down* perspective and treat the EFT as the limit of a well-known renormalisable QFT in the low-energy regime. We will see that this is how the weak effective theory is defined, taking a low-energy limit of the SM. As a simple example, consider a theory of a scalar field ϕ in the region $E \ll \Lambda$. We can separate ϕ into hard and soft modes ϕ_H and ϕ_S by writing $\phi = \phi_H + \phi_S$. By construction, the low-energy physics is described by interactions of the soft modes ϕ_S only, and we can calculate physical quantities from n-point correlation functions derived from the generating functional

$$Z[J_S] = \int \mathcal{D}\phi_S \mathcal{D}\phi_H \exp\Big(iS(\phi_S, \phi_H) + i\int d^4x J_S(x)\phi_S(x)\Big), \tag{2.1.5}$$

where $S(\phi_S, \phi_H)$ is the action and $J_S(x)$ denotes an external current. We define the effective action $S_{eff}(\phi_S)$ by 'integrating out' the hard modes,

$$\int \mathcal{D}\phi_H \exp\Big(iS(\phi_S, \phi_H)\Big) \equiv \exp\Big(iS_{eff}(\phi_S)\Big), \tag{2.1.6}$$

implicitly defining the effective Lagrangian $\mathcal{L}_{eff}$ as

$$S_{eff}(\phi_S) = \int d^4x \mathcal{L}_{eff}(\phi_S). \tag{2.1.7}$$

In practice, integrating over the hard modes will result in a non-local effective Lagrangian, with non-locality over scales of order $\Delta x \sim \frac{1}{\Lambda}$. Non-localities are removed by expanding $\mathcal{L}_{eff}$ around $\Delta x = 0$, resulting in an expansion in $\frac{1}{\Lambda}$. The final expression

for $\mathcal{L}_{eff}$ is a sum of local operators $\mathcal{O}_i$ constructed from ϕ_S, with coefficients $g_i(\Lambda)$:

$$\mathcal{L}_{eff}(\phi_S) = \sum_i g_i(\Lambda)\mathcal{O}_i(\phi_S). \tag{2.1.8}$$

As in the bottom-up construction, we may define Wilson coefficients C_i by factoring out the Λ-dependence of $g_i(\Lambda)$ to reproduce Eq. 2.1.2.

By taking a top-down perspective, we gain a new insight into the impact on the EFT of physics outside of the region of validity. We see that although the LHS of Eq. 2.1.6 depends on physics above the cutoff Λ, the effective Lagrangian contains operators constructed only from ϕ_S and its derivatives. All dependence on the hard modes and physics above the scale Λ is captured by the Wilson coefficients. By learning about the Wilson coefficients of an EFT, we can expect to learn something about physics outside of the region of validity of the EFT.

2.1.2 Matching

This point can be illustrated by matching a UV-complete QFT to the corresponding low-energy EFT. Consider the SM at energy E below the mass of the $W^\pm$ boson, $E \ll m_W$. In this limit the SM is well-approximated by a low-energy EFT in which the $W^\pm$ boson is no longer a degree of freedom. The resulting EFT is the Fermi theory of low-energy weak interactions.

Consider the decay of a muon: $\mu^- \to e^-\bar{\nu}_e\nu_\mu$. Fig. 2.1.1 shows the SM diagram for this process in which the muon decays through a $W^\pm$ boson. We can describe the interaction by the Lagrangian

$$\mathcal{L}_{\text{SM}} \supset -\frac{g_L}{\sqrt{2}}(\bar{\nu}_{e_L,i}\gamma^\rho e_{L,i}W^+_\rho + \bar{e}_{L,i}\gamma^\rho \nu_{e_L,i}W^-_\rho), \tag{2.1.9}$$

where $i \in \{1,2,3\}$ denotes lepton flavour indices and g_L denotes the $SU(2)_L$ coupling strength. We calculate the SM amplitude to be

$$\mathcal{M}_{\text{SM}} = -\frac{g_L^2}{2(p^2 - m_W^2)}(\bar{\nu}_{e_L}\gamma^\rho e_L)(\bar{\mu}_L\gamma_\rho\nu_{\mu_L}). \tag{2.1.10}$$

In the EFT, the $W^\pm$ boson is no longer a dynamical degree of freedom. Instead, this interaction takes place through a four-fermion interaction, shown on the right of Fig. 2.1.1. This interaction is expressed in the EFT Lagrangian in terms of Wilson

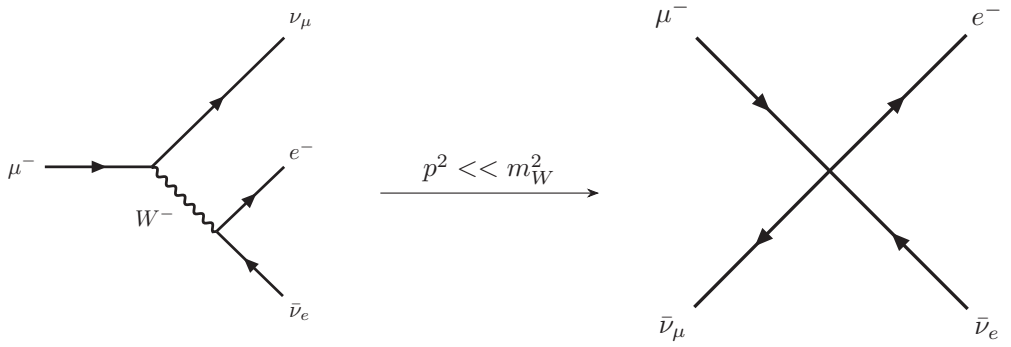


Fig. 2.1.1. Muon decay $\mu^- \to e^- \nu_\mu \bar{\nu}_\mu$ takes place through a $W^\pm$ boson interaction in the SM (left). In the limit of $p^2 \ll m_W^2$ this can be approximated by a four-fermion interaction (right).

coefficient C and new physics scale m_W, resulting in the amplitude $\mathcal{M}_{\text{EFT}}$:

$$\begin{aligned} \mathcal{L}_{\text{EFT}} &\supset \frac{C}{m_W^2}(\bar{\nu}_{e_L}\gamma^\rho e_L)(\bar{\mu}_L\gamma_\rho\nu_{\mu_L}) + \mathcal{O}(\frac{p^4}{m_W^4}) \\ \Rightarrow \mathcal{M}_{\text{EFT}} &= \frac{C}{m_W^2}(\bar{\nu}_{e_L}\gamma^\rho e_L)(\bar{\mu}_L\gamma_\rho\nu_{\mu_L}) + \mathcal{O}(\frac{p^4}{m_W^4}). \end{aligned} \tag{2.1.11}$$

As expected, $\mathcal{M}_{\text{SM}}$ and $\mathcal{M}_{\text{EFT}}$ differ. In the high-p^2 regime we expect the EFT to break down, whereas the SM captures the effects of an on-shell $W^\pm$ boson. In the IR limit, however, we expect these expressions to agree. We can test this explicitly by expanding the $W^\pm$ boson propagator of $\mathcal{M}_{\text{SM}}$ in the limit $p^2 \ll m_W^2$, to produce

$$\mathcal{M}_{\text{SM}} \to \frac{g_L^2}{2m_W^2}(\bar{\nu}_{e_L}\gamma^\rho e_L)(\bar{\mu}_L\gamma_\rho\nu_{\mu_L}) + \mathcal{O}(\frac{p^4}{m_W^4}). \tag{2.1.12}$$

Comparing Eq's 2.1.12 and 2.1.11 we see that the amplitudes agree as long as we fix C/m_W^2 to satisfy

$$\frac{C}{m_W^2} = -\frac{g_L^2}{2m_W^2}. \tag{2.1.13}$$

Historically, the Wilson coefficient of the EFT is expressed in terms of the Fermi constant G_F as $\frac{C}{m_W^2} = -\frac{4G_F}{\sqrt{2}}$, reducing Eq. 2.1.13 to the well-known relation

$$G_F = \frac{1}{\sqrt{2}v^2}, \tag{2.1.14}$$

where v is the Higgs vacuum expectation value (VEV) and we have made use of the relation $m_W^2 = \frac{1}{2}v^2 g_L^2$.

The process of comparing the EFT and full theory in the IR limit to connect the Wilson coefficients to the full theory is known as *matching*. This calculation is an example of tree-level matching, in which only the tree-level contributions to the amplitudes $\mathcal{M}_{\text{SM}}$, $\mathcal{M}_{\text{EFT}}$ are calculated and compared.

Through this example we see that by performing measurements of the Wilson coefficient G_F, we learn about parameters of the UV model g_L^2/m_W^2. Further measurements are required to disentangle g_L and m_W.

2.1.3 Renormalisation

We initially referred to the higher-dimensional interactions of the EFT in Eq. 2.1.4 as non-renormalisable. This was motivated by the fact that for each possible amplitude constructed from the operators in the EFT, although a counterterm can always be found, an infinite number of operators is required to fully renormalise the theory in this way. For example, renormalising a loop diagram consisting of two dimension-5 operators will require a dimension-6 counterterm. Renormalising loop diagrams consisting of these dimension-6 operators will then require further higher-dimensional operators, and so forth. However, once we truncate at a fixed order in the EFT expansion, this is no longer an issue. For example, truncating at order E/Λ forbids the loop diagrams consisting of two dimension-5 operators, avoiding the need to include dimension-6 operators to renormalise these terms.

Once equipped with a renormalisation scheme, an EFT forms a well-defined QFT. It is convenient to work with a mass-independent regularisation and renormalisation scheme. An example is the combination of dimensional regularisation and $\overline{\text{MS}}$, which is often used for the SMEFT. With this definition we can in principle calculate observables in the SMEFT at any order in α_s, α_{EW}, just as in the SM.

Regularising using a mass-independent scheme, such as dimensional regularisation, allows us to maintain the dimensional arguments used to form the power-counting formula of Eq. 2.1.3. Consider regularising loop integrals using instead a UV cutoff Λ_c. For each operator $\mathcal{O}_i$ we may only produce factors of $1/\Lambda^{D_i-4}$ through the insertion of the operator into a diagram vertex, just as in the case of tree-level diagrams. However, factors of Λ_c may arise in the numerator through loops. The convergence of the EFT expansion now depends on the size of Λ_c: if $\Lambda_c > \Lambda$, convergence is destroyed. Working instead with a mass-independent regularisation scheme removes this ambiguity.

The use of a mass-independent scheme also affects the renormalisation group evolution of the EFT parameters. The dependence of a Wilson coefficient C_i on the renormalisation scale μ is given by

$$\mu \frac{dC_i}{d\mu} = \gamma_{ij} C_j. \tag{2.1.15}$$

γ_{ij} denotes the anomalous dimension matrix, and has been fully calculated for the dimension-6 operators of the SMEFT [28–30]. This tells us that arbitrarily setting $C_i = 0$ is a scale-dependent statement, and C_i may become nonzero once μ is varied unless forbidden by a symmetry. In a mass-independent renormalisation scheme the β functions and anomalous dimension matrix γ_{ij} are independent of the mass parameters in our theory. As a result, as we evolve the renormalisation scale μ downwards, heavy particles do not decouple from the theory. For example in the SMEFT, if we evolve μ down to a scale below the mass of the top quark $\mu < m_t$, we must insert the decoupling of the top quark by hand.

2.2 The Standard Model Effective Field Theory

We began our discussion of EFTs by observing that while the SM currently provides an excellent description of measurements at the electroweak scale $v \sim 246$ GeV, physics beyond the standard model (BSM) may consist of heavy new particles at scales $\Lambda \gg v$. Integrating a high-mass particle out of a UV theory will result in operators of dimension $D \geq 5$ consisting of interactions of the SM degrees of freedom. We will use the SMEFT to parametrise the effects of such high-mass fields on observables at energies accessible to the LHC [31, 20, 23, 10, 26, 32, 33].

To construct the SMEFT, we assume that the SM gauge symmetries $SU(3)_c \times SU(2)_L \times U(1)_Y$ continue to hold. In particular, we assume that $SU(2)_L \times U(1)_Y$ is linearly realised, with the Higgs field H in a doublet representation of $SU(2)_L$.[2] The degrees of freedom of the SMEFT are the fields of the SM, and we will always assume three generations of fermions. Their conventions are shown in Table 2.2.1. We will use H to denote the Higgs doublet, and G_μ, W_μ, B_μ to denote the gauge fields of $SU(3)_c$, $SU(2)_L$ and $U(1)_Y$ respectively.

[2]The Higgs Effective Field Theory is formed when this assumption is relaxed and a Higgs boson is instead included in the theory as a scalar singlet [34].

Field	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
$l_i = \{\nu_{L,i}, e_{L,i}\}$	1	2	$-1/2$
$e_{R,i}$	1	1	-1
$q_j = \{u_{L,j}, d_{L,j}\}$	3	2	$1/6$
$u_{R,j}$	3	1	$2/3$
$d_{R,j}$	3	1	$-1/3$
H	1	2	$1/2$
G	8	1	0
W	1	3	0
B	1	1	0

Table 2.2.1. Conventions for the field content of the SM. We use indices $i, j \in \{1, 2, 3\}$ to denote families of the leptons and quarks respectively, and L,R to denote left and right-handed fermions.

We take our expansion parameter to be the ratio of energy scales E/Λ and express the Lagrangian as follows:

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{C^5}{\Lambda}\mathcal{O}^5 + \sum_i \frac{C_i^6}{\Lambda^2}\mathcal{O}_i^6 + ... \tag{2.2.1}$$

In the SMEFT, the SM Lagrangian is simply the leading approximation of a series of operators of increasing power of E/Λ. At each order in E/Λ, we open up a new set of interactions of the SM fields and introduce a new set of coefficients into the theory.

2.2.1 SMEFT operators and bases

The operators of the SMEFT can be formed by constructing all Lorentz invariant operators consisting of SM fields which respect the SM gauge symmetries. The set of operators at each dimension d forms a vector space, V_d. However, there are a number of equivalences between the operators in V_d under which their contributions to the S-matrix (and therefore to physical observables of interest) are equal. The following is based on Refs. [35, 23, 36, 27].

Firstly, operators may be equal up to a total derivative. In perturbation theory the contribution from the total derivative is simply a surface term, which will not contribute to the S-matrix. Secondly, operators may be related by the equations of motion: any linear combination of operators which vanishes when the equations of motion are satisfied will not contribute to the S-matrix. Finally, operators may be

related by mathematical identities such as the Fierz identity and Schouten identity, implying a redundancy between these operators.

Define U_d to be the subspace of V_d formed from the set of operators which do not contribute to the S-matrix as a result of the above equivalences. U_d will contain linear combinations of elements of V_d which vanish when the equations of motion are satisfied, taking the form

$$\sum_i k_i \mathcal{O}_i = \sum_\phi F_\phi \frac{\delta S}{\delta \phi}\,, \tag{2.2.2}$$

where $\mathcal{O}_i \in V_d$ and k_i denote arbitrary coefficients. On the RHS we sum over SM fields ϕ where S denotes the action and F_ϕ is some polynomial of the SM fields and their covariant derivatives [35]. Furthermore, U_d will contain total derivatives and linear combinations of elements of V_d which vanish under Fierz and Schouten identites.

The quotient space $M_d = V_d/U_d$ is defined as follows. Consider operators $\mathcal{O}_1$, $\mathcal{O}_2 \in V_d$. We define $\mathcal{O}_1$ and $\mathcal{O}_2$ to be equivalent if $\mathcal{O}_1 - \mathcal{O}_2 \in U_d$ *i.e.* their difference does not contribute to the S-matrix. Under this definition of equivalence, we can associate each $\mathcal{O} \in V_d$ with an equivalence class $[\mathcal{O}]$ of operators. The quotient space M_d is then the set of all such equivalence classes. Note that the quotient space M_d is itself a vector space. A linear combination of equivalence classes produces another equivalence class, and the zero vector $[0]$ can be identified with U_d, the set of all zero contributions to the S-matrix. Finally, to calculate observables explicitly we must first choose a basis of operators in which to express these observables. We do so by selecting, for each equivalence class $[\mathcal{O}]$, a representative $\mathcal{O}_1 \in [\mathcal{O}]$. Different choices of representative $\mathcal{O}_1$ result in different choices of basis.

The dimension of M_d increases rapidly with d. We will assume 3 generations of fermions when counting this dimension in the SMEFT, unless otherwise specified. At dimension-5 in the SMEFT only 12 operators are needed. We will see these explicitly in §2.2.2. At dimension-6 numerous bases exist and are popular in the literature, including the Warsaw basis [10], the SILH basis [37, 38] and the HISZ basis [39]. Transformations between these are automated by programs including `DEFT` and `Rosetta` [36, 40]. We will make use of the Warsaw basis throughout this thesis, and will discuss this basis in more detail in §2.2.3. The operators of the Warsaw basis are given in Table 2.2.2, consisting of 2499 operators (see Ref. [30] for an explicit counting). Hilbert series methods have been used to count the number of operators up to dimension-12 [41–44] for arbitrary numbers of fermion generations. In particular, these provide a counting and classification of the types of operators appearing at dimension 7 and 8, finding a total of 1542 operators at dimension-7 and 44807 at dimension-8. Explicit forms of

bases for dimension-7 operators are given in Ref. [45, 46], and similarly for dimension-8 in Ref. [47, 48]. We will not consider operators at dimensions higher than dimension-6 in this thesis.

Before concluding this section, we will return to the concept of equivalences and redundancies among operators. These may be illustrated by considering an example from Ref. [27], in which we start with the dimension-6 operator $\mathcal{O} = (H^\dagger H) D_\mu H^\dagger D^\mu H$, which does not appear in the Warsaw basis. Integrating by parts, this can be written as

$$\mathcal{O} = (H^\dagger H) D_\mu H^\dagger D^\mu H = \frac{1}{2}\Big(\mathcal{O}_{H\Box} - (H^\dagger H)(H^\dagger \Box H + \Box H^\dagger H)\Big). \tag{2.2.3}$$

Terms involving $\Box H$ can be eliminated by the Higgs equation of motion

$$\Box H = \lambda v^2 H - 2\lambda (H^\dagger H) H - j_H \tag{2.2.4}$$

where λ denotes the dimensionless coupling of the Higgs potential. The current j_H is given by

$$j_H = \bar{q}^i Y_d^{ij} d_R^j + \bar{l}^i Y_e^{ij} e_R^j - \tilde{q}^i Y_u^{ij} u_R^j \tag{2.2.5}$$

where Y_d, Y_e and Y_u denote 3×3 matrices of Yukawa couplings, and $\tilde{q}^a = \epsilon^{ab} q^b$ with fundamental $SU(2)_L$ indices a, b. Substituting these into $\mathcal{O}$, we find it can be written as a sum of Warsaw basis dimension-6 operators and terms appearing in the SM Lagrangian:

$$\mathcal{O} = \frac{1}{2}\mathcal{O}_{H\Box} + 2\lambda \mathcal{O}_H - \lambda v^2 (H^\dagger H)^2 + \frac{1}{2}\mathcal{O}_{\psi H}, \tag{2.2.6}$$

where fermionic terms such as $(H^\dagger H) j_H H$ have been grouped into the operator $\mathcal{O}_{\psi H}$. We see that although $\mathcal{O}$ depends only on the SM degrees of freedom and respects the SM gauge symmetries, this operator is redundant when working in the Warsaw basis.

2.2.2 Dimension-5

In this and the following subsection we will focus on the operators of dimension-5 and dimension-6 respectively. For each we will discuss in detail the physically inequivalent operators *i.e.* the operators which, at each dimension, contribute to the S-matrix. We will choose a fixed basis for each discussion.

At dimension-5, only the Weinberg operator contributes to the S-matrix [49], given by

$$\mathcal{O}_{ij}^5 = (\bar{l}_i^c \tilde{H}^*)(\tilde{H}^\dagger l_j) + h.c, \tag{2.2.7}$$

where l^c denotes the charge-conjugated LH leptons, $\tilde{H} = i\sigma^2 H^*$ and i, j denote lepton family indices. $\mathcal{O}^5_{ij}$ is symmetric in i, j and is not Hermitian, implying that this operator consists of 12 independent components [41]. The Weinberg operator $\mathcal{O}^5$ violates lepton number L by $\Delta L = 2$ units, a reminder that the accidental symmetries of the SM are not required to hold in the SMEFT.

This operator is responsible for producing neutrino masses. Schematically making the replacement $H \to \frac{1}{\sqrt{2}}v$ after electroweak symmetry breaking (EWSB) reduces $\mathcal{O}^5$ to a Majorana mass term of the form

$$\mathcal{O}^5 \sim \frac{v^2}{2}\bar{\nu^c}_L \nu_L. \tag{2.2.8}$$

The detection of neutrino oscillations [50] provided the first evidence for nonzero neutrino mass differences Δm^2, implying that at least two neutrinos must have nonzero mass. These measurements place limits on the neutrino masses which can be interpreted as constraints on the ratio C^5/Λ, where C^5 is the 3×3 matrix of Wilson coefficients, indicating $\Lambda \gtrsim 10^{15}$ GeV for $C^5 \sim \mathcal{O}(1)$ [27].

2.2.3 Dimension-6

Table 2.2.2 shows the operators at dimension-6 expressed in the Warsaw basis [10], forming a total of 2499 operators (see Ref. [30] for an explicit counting). The operators are divided into categories reflecting the fields and number of derivatives present. We will discuss the phenomenology of each category in turn. It is important to note that this division of operators into categories is basis-dependent. As we saw in the example of §2.2.1 in Eq. 2.2.6, an operator which consists purely of the Higgs doublet H and its covariant derivative in one basis may generate four-fermion operators as well as Higgs operators when rotated to the Warsaw basis. For consistency we will work with only the Warsaw basis throughout this thesis.

$\boldsymbol{X^3}$: X denotes the gauge field strengths $W_{\mu\nu}$, $G_{\mu\nu}$, $B_{\mu\nu}$. Operators $\mathcal{O}_{\widetilde{G}}$, $\mathcal{O}_{\widetilde{W}}$ are CP-violating, whereas $\mathcal{O}_G$ and $\mathcal{O}_W$ conserve CP. Nonzero values of the Wilson coefficient C_G will contribute to processes involving jets. These are plentiful in LHC measurements, and we will see in Chapter 3 that this leads to strong correlations between C_G and other dimension-6 coefficients.

X^3		H^6 and H^4D^2		ψ^2H^3	
$\mathcal{O}_G$	$f^{ABC}G^{A\nu}_{\mu}G^{B\rho}_{\nu}G^{C\mu}_{\rho}$	$\mathcal{O}_H$	$(H^\dagger H)^3$	$\mathcal{O}_{eH}$	$(H^\dagger H)(\bar{l}_p e_r H)$
$\mathcal{O}_{\widetilde{G}}$	$f^{ABC}\widetilde{G}^{A\nu}_{\mu}G^{B\rho}_{\nu}G^{C\mu}_{\rho}$	$\mathcal{O}_{H\Box}$	$(H^\dagger H)\Box(H^\dagger H)$	$\mathcal{O}_{uH}$	$(H^\dagger H)(\bar{q}_p u_r \widetilde{H})$
$\mathcal{O}_W$	$\varepsilon^{IJK}W^{I\nu}_{\mu}W^{J\rho}_{\nu}W^{K\mu}_{\rho}$	$\mathcal{O}_{HD}$	$\left(H^\dagger D^\mu H\right)^\star \left(H^\dagger D_\mu H\right)$	$\mathcal{O}_{dH}$	$(H^\dagger H)(\bar{q}_p d_r H)$
$\mathcal{O}_{\widetilde{W}}$	$\varepsilon^{IJK}\widetilde{W}^{I\nu}_{\mu}W^{J\rho}_{\nu}W^{K\mu}_{\rho}$				

X^2H^2		ψ^2XH		ψ^2H^2D	
$\mathcal{O}_{HG}$	$H^\dagger H\, G^A_{\mu\nu}G^{A\mu\nu}$	$\mathcal{O}_{eW}$	$(\bar{l}_p\sigma^{\mu\nu}e_r)\tau^I H W^I_{\mu\nu}$	$\mathcal{O}^{(1)}_{Hl}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{l}_p\gamma^\mu l_r)$
$\mathcal{O}_{H\widetilde{G}}$	$H^\dagger H\, \widetilde{G}^A_{\mu\nu}G^{A\mu\nu}$	$\mathcal{O}_{eB}$	$(\bar{l}_p\sigma^{\mu\nu}e_r)H B_{\mu\nu}$	$\mathcal{O}^{(3)}_{Hl}$	$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\bar{l}_p\tau^I\gamma^\mu l_r)$
$\mathcal{O}_{HW}$	$H^\dagger H\, W^I_{\mu\nu}W^{I\mu\nu}$	$\mathcal{O}_{uG}$	$(\bar{q}_p\sigma^{\mu\nu}T^A u_r)\widetilde{H}\, G^A_{\mu\nu}$	$\mathcal{O}_{He}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{e}_p\gamma^\mu e_r)$
$\mathcal{O}_{H\widetilde{W}}$	$H^\dagger H\, \widetilde{W}^I_{\mu\nu}W^{I\mu\nu}$	$\mathcal{O}_{uW}$	$(\bar{q}_p\sigma^{\mu\nu}u_r)\tau^I\widetilde{H}\, W^I_{\mu\nu}$	$\mathcal{O}^{(1)}_{Hq}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{q}_p\gamma^\mu q_r)$
$\mathcal{O}_{HB}$	$H^\dagger H\, B_{\mu\nu}B^{\mu\nu}$	$\mathcal{O}_{uB}$	$(\bar{q}_p\sigma^{\mu\nu}u_r)\widetilde{H}\, B_{\mu\nu}$	$\mathcal{O}^{(3)}_{Hq}$	$(H^\dagger i\overleftrightarrow{D}^I_\mu H)(\bar{q}_p\tau^I\gamma^\mu q_r)$
$\mathcal{O}_{H\widetilde{B}}$	$H^\dagger H\, \widetilde{B}_{\mu\nu}B^{\mu\nu}$	$\mathcal{O}_{dG}$	$(\bar{q}_p\sigma^{\mu\nu}T^A d_r)H\, G^A_{\mu\nu}$	$\mathcal{O}_{Hu}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{u}_p\gamma^\mu u_r)$
$\mathcal{O}_{HWB}$	$H^\dagger\tau^I H\, W^I_{\mu\nu}B^{\mu\nu}$	$\mathcal{O}_{dW}$	$(\bar{q}_p\sigma^{\mu\nu}d_r)\tau^I H\, W^I_{\mu\nu}$	$\mathcal{O}_{Hd}$	$(H^\dagger i\overleftrightarrow{D}_\mu H)(\bar{d}_p\gamma^\mu d_r)$
$\mathcal{O}_{H\widetilde{W}B}$	$H^\dagger\tau^I H\, \widetilde{W}^I_{\mu\nu}B^{\mu\nu}$	$\mathcal{O}_{dB}$	$(\bar{q}_p\sigma^{\mu\nu}d_r)H\, B_{\mu\nu}$	$\mathcal{O}_{Hud}$	$i(\widetilde{H}^\dagger D_\mu H)(\bar{u}_p\gamma^\mu d_r)$

$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}L)(\bar{R}R)$	
$\mathcal{O}_{ll}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	$\mathcal{O}_{ee}$	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	$\mathcal{O}_{le}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{e}_s\gamma^\mu e_t)$
$\mathcal{O}^{(1)}_{qq}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma^\mu q_t)$	$\mathcal{O}_{uu}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma^\mu u_t)$	$\mathcal{O}_{lu}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma^\mu u_t)$
$\mathcal{O}^{(3)}_{qq}$	$(\bar{q}_p\gamma_\mu\tau^I q_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	$\mathcal{O}_{dd}$	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma^\mu d_t)$	$\mathcal{O}_{ld}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma^\mu d_t)$
$\mathcal{O}^{(1)}_{lq}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	$\mathcal{O}_{eu}$	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$	$\mathcal{O}_{qe}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$
$\mathcal{O}^{(3)}_{lq}$	$(\bar{l}_p\gamma_\mu\tau^I l_r)(\bar{q}_s\gamma^\mu\tau^I q_t)$	$\mathcal{O}_{ed}$	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$	$\mathcal{O}^{(1)}_{qu}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$\mathcal{O}^{(1)}_{ud}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$\mathcal{O}^{(8)}_{qu}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{u}_s\gamma^\mu T^A u_t)$
		$\mathcal{O}^{(8)}_{ud}$	$(\bar{u}_p\gamma_\mu T^A u_r)(\bar{d}_s\gamma^\mu T^A d_t)$	$\mathcal{O}^{(1)}_{qd}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$\mathcal{O}^{(8)}_{qd}$	$(\bar{q}_p\gamma_\mu T^A q_r)(\bar{d}_s\gamma^\mu T^A d_t)$

$(\bar{L}R)(\bar{R}L)$ and $(\bar{L}R)(\bar{L}R)$		B-violating	
$\mathcal{O}_{ledq}$	$(\bar{l}^j_p e_r)(\bar{d}_s q^j_t)$	$\mathcal{O}_{duq}$	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(d^\alpha_p)^T C u^\beta_r\right]\left[(q^{\gamma j}_s)^T C l^k_t\right]$
$\mathcal{O}^{(1)}_{quqd}$	$(\bar{q}^j_p u_r)\varepsilon_{jk}(\bar{q}^k_s d_t)$	$\mathcal{O}_{qqu}$	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jk}\left[(q^{\alpha j}_p)^T C q^{\beta k}_r\right]\left[(u^\gamma_s)^T C e_t\right]$
$\mathcal{O}^{(8)}_{quqd}$	$(\bar{q}^j_p T^A u_r)\varepsilon_{jk}(\bar{q}^k_s T^A d_t)$	$\mathcal{O}_{qqq}$	$\varepsilon^{\alpha\beta\gamma}\varepsilon_{jn}\varepsilon_{km}\left[(q^{\alpha j}_p)^T C q^{\beta k}_r\right]\left[(q^{\gamma m}_s)^T C l^n_t\right]$
$\mathcal{O}^{(1)}_{lequ}$	$(\bar{l}^j_p e_r)\varepsilon_{jk}(\bar{q}^k_s u_t)$	$\mathcal{O}_{duu}$	$\varepsilon^{\alpha\beta\gamma}\left[(d^\alpha_p)^T C u^\beta_r\right]\left[(u^\gamma_s)^T C e_t\right]$
$\mathcal{O}^{(3)}_{lequ}$	$(\bar{l}^j_p\sigma_{\mu\nu}e_r)\varepsilon_{jk}(\bar{q}^k_s\sigma^{\mu\nu}u_t)$		

Table 2.2.2. Dimension 6 operators in the Warsaw basis, adapted from Refs. [1] and [10]. Indices $p, q, r, s, t \in \{1, 2, 3\}$ label fermion generations. The operators are grouped by their field content, where X denotes the gauge field strengths, D denotes the covariant derivative, ψ denotes a fermion and L, R describe the chirality of the fermionic operators. Labels L, R of the left and right-handed fermions are suppressed for clarity. The shaded cells indicate operators that break the flavour symmetry $SU(3)^5$. Note that this division of operators into categories based on the fields and covariant derivatives present is basis-dependent.

$\boldsymbol{H^6}$ **and** $\boldsymbol{H^4D^2}$**:** $\mathcal{O}_H$ supplements the SM with new self-interactions of the Higgs boson. The operators $\mathcal{O}_{H\Box}$ and $\mathcal{O}_{HD}$ modify the Higgs couplings to the electroweak gauge bosons. An example is shown in Fig. 2.2.1, where these operators contribute to Higgs production via associated vector boson production (VH) and vector boson fusion (VBF) through their modifications to the H-Z-Z coupling.

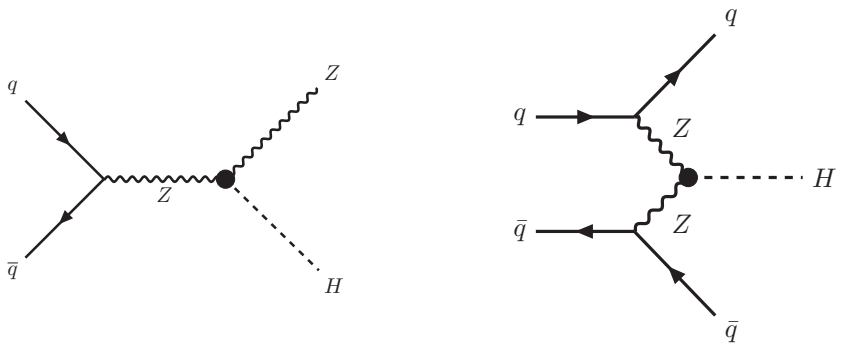


Fig. 2.2.1. Higgs production through associated vector production (VH) and vector boson fusion (VBF) can be modified by nonzero values of the Wilson coefficients C_{HD} and $C_{H\Box}$. The black circles denote their insertion into the Z-Z-H vertex.

$\boldsymbol{X^2H^2}$: similarly, operators in this category couple the gauge field strengths to the Higgs boson. C_{HG} contributes to the gluon fusion (ggF) production mode of the Higgs through the insertion shown in Fig. 2.2.2. In contrast to the SMEFT, the leading contribution to ggF in the SM occurs at 1-loop, as shown by the right diagram of Fig. 2.2.2.

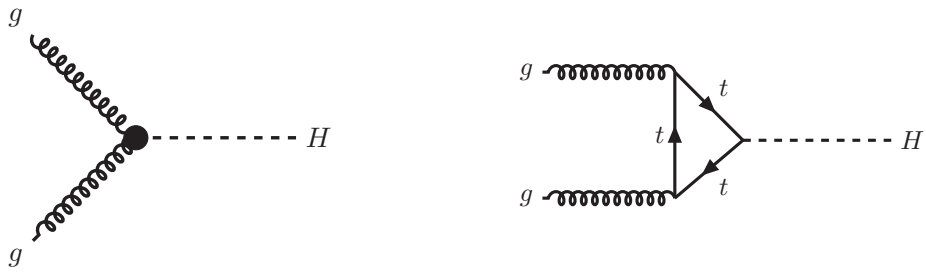


Fig. 2.2.2. Higgs production through gluon fusion can be modified by the operator $\mathcal{O}_{HG}$. The presence of this operator means that the leading contribution in the SMEFT is at tree-level, as shown by the diagram on the left. The leading contribution in the SM is shown on the right and proceeds through a top quark loop.

Next we consider operators containing fermions. We will denote these by ψ in Table. 2.2.2. The flavour indices of the quarks and leptons, denoted by p, r, s and t in Table 2.2.2, are largely responsible for the vast number of operators at dimension-6. A common approach to deal with such a large number of operators is to impose a symmetry such that some are forbidden or made redundant, such as the $SU(3)^5$ flavour symmetry.

In the absence of the Yukawa interactions, the SM Lagrangian has a $U(3)^5$ symmetry [51, 52] which performs unitary transformations of q_i, u_i, d_i, l_i and e_i, for example

$$q_i \to (U_q)_{ij} q_j \quad \text{where} \quad U_q \in U(3)_q. \tag{2.2.9}$$

This symmetry can be decomposed into five $SU(3)$ flavour rotations and five $U(1)$ symmetries, two of which are baryon number B and lepton number L:

$$U(3)^5 \to SU(3)_q \times SU(3)_u \times SU(3)_d \times SU(3)_l \times SU(3)_e \times U(1)^5 \,. \tag{2.2.10}$$

If we impose the $SU(3)^5$ symmetry on the SMEFT, we reduce it to only the unshaded operators in Table 2.2.2. As expected, this has no effect on the bosonic operators. The number of fermionic operators is reduced, and those which are allowed are required to have common Wilson coefficients and no off-diagonal terms, *i.e.* we have a single Wilson coefficient for each unshaded operator in the table.

In total we find that the $SU(3)^5$ symmetry reduces the number of operators to a total of 42 in the Warsaw basis. This can be seen from Table 2.2.2 by counting the number of unshaded entries. Note that this number differs from the number of operators in the SMEFT with one generation of fermions, calculated to be 84 by Ref. [41]. The differences arises because chirality flipping operators, such as the up-type Yukawa $\mathcal{O}_{uH} = (H^\dagger H)(\bar{q} u H)$, are included in the one-generation-fermion SMEFT but forbidden by flavour universality.

$\psi^2 H^3$: these operators are a product of a $(H^\dagger H)$ factor multiplied by the dimension-4 Yukawa operators of the SM. Replacing the $(H^\dagger H)$ term with a factor of $v^2/2$, we see that contributions from these operators may lead to a shift in the Yukawa interactions between the Higgs and fermions. LHC measurements are gaining sensitivity to these couplings: for example, the recent observation of $t\bar{t}H$ production at the LHC [13] provides a measurement which is sensitive to $\mathcal{O}_{uH}$ through the insertion shown in Fig. 2.2.3. These operators violate $SU(3)^5$.

$\boldsymbol{\psi^2 XH}$ **:** similarly, these operators violate $SU(3)^5$. They are known as dipole operators as they produce dipole interactions after the Higgs acquires a VEV. A nonzero value of C_{eB}, for example, would contribute to the well-constrained electron dipole moment [53].

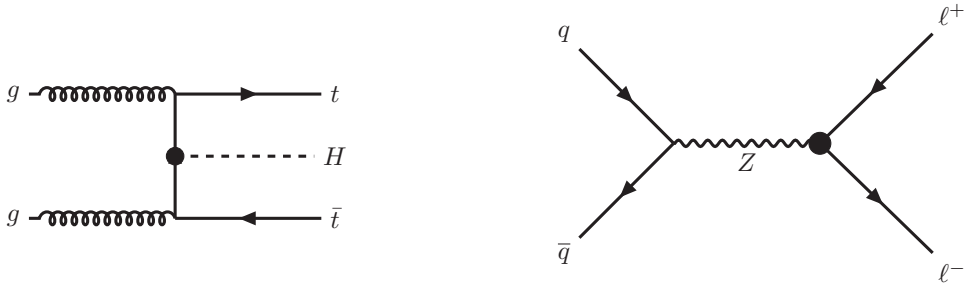


Fig. 2.2.3. On the left, Higgs production in association with a $t\bar{t}$ pair can be modified by the operator $\mathcal{O}_{uH}$. On the right, the coupling of Z to a l^+l^- pair is modified by the insertion of operators such as $\mathcal{O}_{Hl}^{(1)}$ and $\mathcal{O}_{Hl}^{(3)}$.

$\boldsymbol{\psi^2 H^2 D}$**:** operators in this class contribute to the couplings of electroweak gauge bosons to the fermions. For example, the singlet and triplet operators $\mathcal{O}_{Hl}^{(1)}$ and $\mathcal{O}_{Hl}^{(3)}$ parametrise new physics contributions to $Z \to ll$ and $W \to l\nu_l$, as shown in Fig. 2.2.3.

$\boldsymbol{\psi^4}$**:** The operators in the lower half of Table 2.2.2 are the four-fermion interactions, categorised by their chirality L or R. Four of these are B-violating, and as a result their Wilson coefficients are highly constrained by measurements of proton decay. Interactions with mixed-chirality violate $SU(3)^5$, even in the flavour diagonal entries. Operators in the $(\bar{R}R)(\bar{R}R)$, $(\bar{L}L)(\bar{R}R)$ and $(\bar{L}L)(\bar{L}L)$ categories contribute significantly to the Drell-Yan ($pp \to l^+l^-$) process.

2.3 Global SMEFT fits

The price we pay for the model independence of an EFT is the large number of Wilson coefficients which are *a priori* unknown, and must be determined from measurements. This is particularly true of the SMEFT, depending on 2499 Wilson coefficients at dimension-6. The goal of a global fit is to constrain the Wilson coefficients, fitting them to the wealth of data available from the LHC and other particle physics experiments. We should not arbitrarily set a subset of coefficients to zero to simplify this fit. Not only does this spoil the model independence and basis independence of the EFT, but renormalisation group running of the Wilson coefficients implies that this is a

scale-dependent statement, as discussed in §2.1.3. As such, when fitting to a set of observables, we should consider all operators relevant to the processes in question.

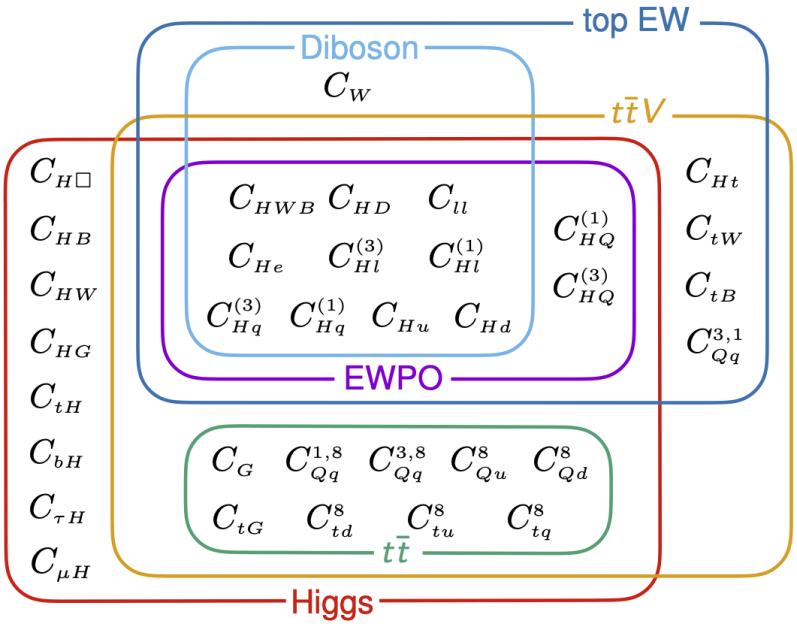


Fig. 2.3.1. Many operators in the dimension-6 SMEFT contribute to various LHC observables including Higgs data, top quark data, diboson production and EWPO and will benefit from a global fit incorporating these datasets simultaneously. This diagram is taken from Ref. [1] and depicts the dimension-6 operators constrained in the global fit of Chapter 3.

Moreover, each Wilson coefficient may contribute to multiple sets of data. This can be seen explicitly in Fig. 2.3.1, taken from Ref. [1], in which we depict the contribution of various coefficients of the dimension-6 SMEFT to datasets considered in this thesis including data describing the Higgs, top quark and diboson production as well as electroweak precision observables (EWPO). Notice that operators in the $\psi^2 H^2 D$ category such as $C_{Hl}^{(1)}$ and $C_{Hl}^{(3)}$ contribute to a wide range of observables, and as a result, including all data at our disposal will most effectively constrain these operators.

As an example, consider Fig. 2.3.2 demonstrating the constraints on the SMEFT Wilson coefficients C_{HD}, C_{HWB} from Higgs data and EWPO.[3] We see that the Higgs data and EWPO separately provide important constraints on these parameters. The orange ellipse indicates that C_{HWB} and C_{HD} are anticorrelated in a fit to EWPO, while the blue band indicates that Higgs data provides strong constraints on C_{HWB}. However, the green ellipse indicates that the combination of Higgs and EWPO significantly

[3] See Appendix A for details of the datasets.

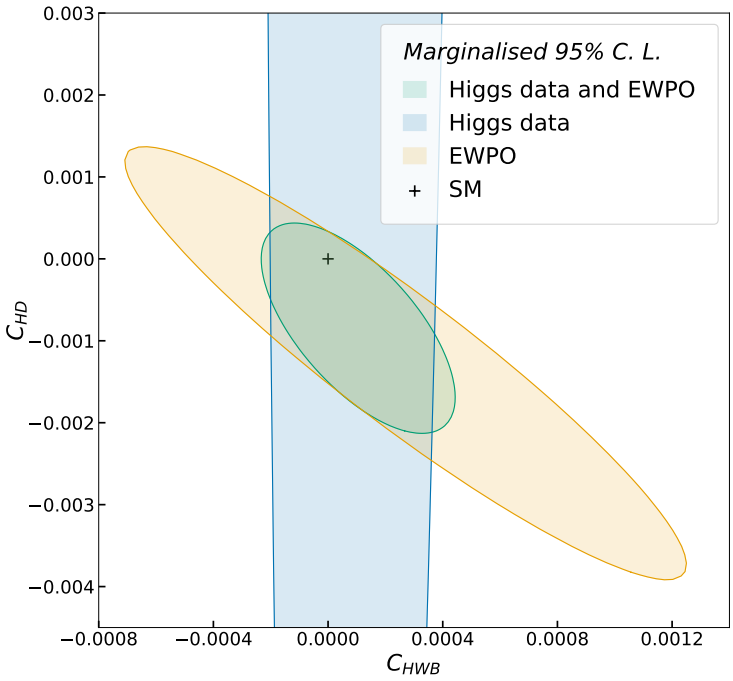


Fig. 2.3.2. Constraints on the SMEFT Wilson coefficients C_{HWB}, C_{HD} (where $\Lambda = 1$ TeV) from Higgs data and electroweak precision observables. These constraints are produced using the `Fitmaker` code, which will be introduced in Chapter 3.

improves the constraints, removing the nearly-flat direction of the fit to Higgs data, and reducing the range allowed by the EWPO.

Fig. 2.3.2 also demonstrates the importance of allowing all operators to vary simultaneously in the fit. By simultaneously fitting C_{HWB} and C_{HD} to Higgs data and EWPO we learn that these parameters are anticorrelated, and as a result, UV models which predict positive correlations between these coefficients will be highly constrained. We would not learn about this correlation by fitting C_{HWB} and C_{HD} one parameter at a time.

The first goal of this thesis is to introduce and discuss the results of Ref. [1] combining top, Higgs, diboson and electroweak data in a global fit to the dimension-6 SMEFT for the first time. This has been preceded by numerous global fits, including fits incorporating Higgs, diboson and electroweak data [54–59] as well as fits to the top sector [60–66]. The `SMEFiT` collaboration have recently performed a similar fit to the combination of top, Higgs and diboson data [67].

2.3.1 Electroweak input scheme

To perform a global fit, a number of choices must be made which impact the results of the fit. One of these is the choice of *electroweak input scheme*. The SM has 19 free parameters, three of which are relevant to the electroweak sector and must be provided as inputs before calculating observables in the electroweak sector for our global fit. We will work with the $\{\alpha_{EW}, G_F, m_Z\}$ scheme. The remaining electroweak parameters are determined by their tree-level relation to $\{\alpha_{EW}, G_F, m_Z\}$. For example, in the SM the Higgs VEV v is related to G_F by $G_F = \frac{1}{\sqrt{2}v^2}$. If we include dimension-6 operators in the Lagrangian, this relationship receives corrections. In the Warsaw basis these take the form

$$G_F = \frac{1}{\sqrt{2}v^2}\Big(1 - \frac{1}{\sqrt{2}}\frac{v^2}{\Lambda^2}(C_{ll} - 2C_{Hl}^{(3)}\Big). \tag{2.3.1}$$

Therefore working in this input scheme will indirectly result in constraints on the operators $\mathcal{O}_{ll}$, $\mathcal{O}_{Hl}^{(3)}$ even when these operators are not directly relevant to the measured observables. This is highlighted by the inclusion of the coefficient C_{ll} in most data sectors shown in Fig. 2.3.1 (all except the $t\bar{t}$ sector, a QCD process which does not depend on electroweak input parameters), despite these processes not having any direct dependence on four-lepton interactions.

2.3.2 Linear vs quadratic

A second choice that must be made prior to performing a global SMEFT fit is the order in the EFT expansion at which we truncate observables. At dimension-6 we may write an amplitude as

$$\mathcal{A} = \mathcal{A}_{SM} + \sum_i \frac{C_i}{\Lambda^2}\mathcal{A}_i \tag{2.3.2}$$

where $\mathcal{A}_i$ denotes the contribution from operator $\mathcal{O}_i$ to $\mathcal{A}$. Observables are computed from $|\mathcal{A}|^2$ given by

$$|\mathcal{A}|^2 = |\mathcal{A}_{SM}|^2 + \sum_i \frac{C_i}{\Lambda^2}\mathcal{A}_{SM}^\dagger\mathcal{A}_i + \sum_{i,j}\frac{C_iC_j}{\Lambda^4}\mathcal{A}_i^\dagger\mathcal{A}_j. \tag{2.3.3}$$

The first term depends only on the SM. The second term is *linear* in Wilson coefficients and $\mathcal{O}(E^2/\Lambda^2)$ in the EFT expansion. In contrast, the third term is *quadratic* in Wilson coefficients and $\mathcal{O}(E^4/\Lambda^4)$ in the EFT expansion. Throughout this thesis we will use the terms linear and quadratic to refer to the dependence of the *observables* on the Wilson coefficients.

The EFT approximation of Eq. 2.3.2 relies on small values of $C_i E^2/\Lambda^2$, where E is the characteristic energy scale of the observable. In this limit we expect the quadratic, $\mathcal{O}(E^4/\Lambda^4)$, terms to be subdominant, and thus can neglect them as a leading approximation. Neglecting these quadratic terms is further motivated by the fact that these terms enter into the observable at the same order in E/Λ as the dimension-8 operators, which we will not include in this thesis. In Chapter 3 we will take this perspective, working to linear order in Wilson coefficients.

However, the assumption that the linear, $\mathcal{O}(E^2/\Lambda^2)$, terms form the leading contribution rests on the assumption that the BSM amplitudes $\mathcal{A}_i$ will interfere appreciably with the SM. This is not always the case. Consider the four-fermion SMEFT operator $(\mathcal{O}_{lq}^{(1)})_{2233}$, defined as

$$(\mathcal{O}_{ql}^{(1)})_{2233} = (\bar{l}_2\gamma_\nu l_2)(\bar{q}_3\gamma^\nu q_3). \tag{2.3.4}$$

In Chapter 4 we will constrain this operator using LHC Drell-Yan data. The contribution to Drell-Yan from this operator is suppressed by the small b-quark parton distribution function (PDF). As a result of this, constraining this operator using only the linear contribution produces very weak constraints. These weak constraints are improved by including also the quadratic term, which benefits from the high-energy scaling of the amplitude $\sim E^2/\Lambda^2$.

Through this example we see that in some cases, the inclusion of the quadratic terms may produce the best constraints on the SMEFT. However, these constraints should be viewed with caution: as mentioned above, the quadratic terms enter into the observable at the same order in E/Λ as the dimension-8 operators of the SMEFT, which are beyond the scope of this thesis. By including the quadratic terms, and not the dimension-8 operators, the constraints produced may not be accurate. For the majority of this thesis we will avoid this issue by including only the linear terms. We will make use of the fact that the strong constraints produced by working to quadratic order are a subset of the less stringent constraints obtained by working instead to linear order. By retaining only the terms linear in Wilson coefficients, we gain a conservative estimate of the true constraining power of the data. We will discuss this viewpoint in more detail in Chapter 3 in the context of top quark data.

2.3.3 Statistical methods

A global fit of the SMEFT relies on a choice of fitting methodology. In each of Chapters 3 and 4 we will perform a fit, determining the values of Wilson coefficients excluded by the available data and therefore constraining the Wilson coefficients to

a range consistent with our observations. In each case, the fitting methodology used will be discussed. Here we will give a brief overview and introduction to the necessary statistical concepts, following Refs. [68–73]. We will use a frequentist interpretation of probability throughout this subsection, except where we explicitly discuss Bayesian parameter estimation.

The likelihood

Consider a dataset D. We will denote the i^{th} datapoint in our dataset by y_i^{exp}, and use $y(\vec{C},\theta)_i$ to denote the corresponding theoretical prediction, where $\vec{C}$ denotes a vector of Wilson coefficients. The parameter θ is a nuisance parameter and can be used, for example, to parametrise the effect of systematic uncertainties on our data. The likelihood $\mathcal{L}(D|\vec{C},\theta)$ is the probability of the observed dataset D, given the value of Wilson coefficients $\vec{C}$ and θ. We will suppress the dependence on D and denote the likelihood by $\mathcal{L}(\vec{C},\theta)$ in the following.

We will make use of a Gaussian likelihood,

$$L(\vec{C},\theta) = \mathcal{N}\exp\Big(-\frac{1}{2}(y^{\text{exp}} - y(\vec{C},\theta))_i^T (V^{-1})_{ij}(y^{\text{exp}} - y(\vec{C},\theta))_j\Big) \tag{2.3.5}$$

where V_{ij} denotes the covariance matrix, $\mathcal{N}$ is a normalisation factor and summation over i,j is implied. LHC measurements consist of measurements of events in N bins of a histogram, denoted by n_i where $i = 1,..,N$. The Gaussian likelihood makes the assumption that we have a large number of events per bin *i.e.* that terms of order $\sim \mathcal{O}(1/\sqrt{n_i})$ can be neglected [73].

The Gaussian likelihood takes into account the statistical and systematic uncertainties on each datapoint as well as the correlations between datapoints via the covariance matrix V_{ij}. The diagonal entry V_{ii} is the variance of y_i^{exp}. The off-diagonal entry V_{ij} is related to the correlation ρ_{ij} between y_i^{exp} and y_j^{exp} by

$$\rho_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}}\sqrt{V_{jj}}} \tag{2.3.6}$$

where there is no summation over indices i, j. Note, however, that we cannot account for asymmetric uncertainties. This implies that if an experiment publishes a value of y_i^{exp} with uncertainties $\Delta(y_i^{exp})_{lower}$ and $\Delta(y_i^{exp})_{upper}$, we must symmetrise the upper and lower uncertainties before incorporating them into the covariance matrix, resulting in a loss of information when constructing the likelihood.

Alternative choices of likelihood exist. These include the Variable Gaussian likelihood of Ref. [74] which is capable of accounting for asymmetric experimental uncertainties, and the Poisson likelihood which is well-suited to datasets in which the large-n_i approximation may not hold, for example high-p_T searches for BSM particles.

The χ^2 test statistic

In order to constrain the Wilson coefficients $\vec{C}$, we must quantify their compatibility with the observed data. We define the test statistic $t(\vec{C})$ as

$$t(\vec{C}) = -2\ln\left(\frac{\mathcal{L}(\vec{C},\hat{\hat{\theta}})}{\mathcal{L}(\hat{\vec{C}},\hat{\theta})}\right). \tag{2.3.7}$$

The best-fit values of $\vec{C}$ and θ are given by their maximum likelihood estimators denoted by $\hat{\vec{C}}$ and $\hat{\theta}$ respectively. We denote the maximum likelihood estimator of θ at fixed value of $\vec{C}$ by $\hat{\hat{\theta}}$. Wilk's theorem [75] states that in the limit of a large number of events n_i in each bin, $t(\vec{C})$ follows a χ^2 distribution with number of degrees of freedom given by the number of Wilson coefficients in our fit. In the case of the Gaussian likelihood this can be readily seen by expressing $t(\vec{C})$ in the form of a difference of two χ^2 functions as

$$\begin{aligned} t(\vec{C}) &= (y^{\text{exp}} - y(\vec{C},\hat{\hat{\theta}}))_i^T (V^{-1})_{ij} (y^{\text{exp}} - y(\vec{C},\hat{\hat{\theta}}))_j \\ &\quad - (y^{\text{exp}} - y(\hat{\vec{C}},\hat{\theta}))_i^T (V^{-1})_{ij} (y^{\text{exp}} - y(\hat{\vec{C}},\hat{\theta}))_j \\ &= \chi^2(C,\hat{\hat{\theta}}) - \chi^2(\hat{C},\hat{\theta}). \end{aligned} \tag{2.3.8}$$

Compatibility between data and theory is quantified by the p-value. This is defined as the probability of observing data as or more incompatible with $\vec{C}$ than the data we have observed. The p-value is calculated from $t(\vec{C})$ as

$$p(t(\vec{C})) = \int_{t(\vec{C})}^{\infty} f(t(\vec{C}')|\vec{C}) dt(\vec{C}') = 1 - F(t(\vec{C})|\vec{C}) \tag{2.3.9}$$

where $f(t(\vec{C}')|\vec{C})$ denotes the probability distribution of $t(\vec{C}')$ under the assumption of data distributed according to the SMEFT at $\vec{C}'$, and $F(t(\vec{C})|\vec{C})$ is the corresponding cumulative distribution function.

We will place constraints on $\vec{C}$ by computing confidence intervals at 95 % confidence level (CL). These confidence intervals are defined as the range of $\vec{C}$ compatible with the data at a significance of $\alpha = 0.05$. We obtain confidence intervals on $\vec{C}$ by solving

$p(\vec{C}_{lim}) = 0.05$ for $\vec{C}_{lim}$ to obtain the limiting values of $\vec{C}$ consistent with our data. In practice we make use of the cumulative distribution function $F(t(\vec{C}), \vec{C})$, obtaining $\vec{C}_{lim}$ by solving $t(\vec{C}_{lim}) = F^{-1}(0.95)$. For example, suppose we wish to constrain two Wilson coefficients. The test statistic $t(C_1, C_2)$ then follows a χ^2 distribution with 2 degrees of freedom. This tells us that F is the cumulative distribution of a χ^2 with 2 degrees of freedom, in which case $F^{-1}(0.95)$ takes the value $F^{-1}(0.95) \approx 5.99$[4].

Finally, we note that we will often deal with observables $y(\vec{C})_i$ which are linear in $\vec{C}$, as discussed in §2.3.2. In this case, $t(\vec{C})$ is a quadratic function of $\vec{C}$. The solutions $\vec{C}_{lim}$ to the equation $t(\vec{C}_{lim}) = F^{-1}(0.95)$ will take the form of an ellipse, or more generally, an n-dimensional ellipsoid, where n is the dimension of $\vec{C}$. An example of two-dimensional elliptic constraints can be seen in Fig. 2.3.2.

Individual and profiled limits

Consider a fit to n Wilson coefficients C_i, $i = 1, ..., n$. Suppose we are interested only in the constraints on C_1. We will use *individual limits* to refer to the constraints obtained on C_1 from the test statistic $t(C_1, 0, ..., 0)$ *i.e.* in the limit that all other Wilson coefficients are set to zero. Setting the subset $\{C_2, ..., C_n\}$ of Wilson coefficients to zero destroys the model independence and basis independence of the effective field theory, however. As a result, although individual limits provide a rough guide of the constraining power of our dataset, there are not realistic constraints on the Wilson coeffient C_1.

A more realistic alternative is provided by *profiled limits*. The profiled limits on C_1 are obtained from the test statistic $t(C_1, \hat{C}_2, ..., \hat{C}_n)$ *i.e.* we set all remaining Wilson coefficients to their maximum likelihood estimates, allowing only C_1 to vary. As we will discuss in the next subsection, profiled limits are often referred to as *marginalised limits*, in particular when using the Gaussian likelihood.

Bayesian parameter estimation

We may take an alternative perspective on constraining $\vec{C}$ by using a Bayesian framework. Let $p(\vec{C}|D)$ denote the probability of $\vec{C}$, given the data set D. Bayes' theorem states that $p(\vec{C}|D)$ is related to the likelihood function $\mathcal{L}(\vec{C})$ by

$$p(\vec{C}|D) = \frac{\mathcal{L}(\vec{C})\pi(\vec{C})}{\int \mathcal{L}(\vec{C})\pi(\vec{C})d^nC} \tag{2.3.10}$$

[4]Values of the cumulative distribution function of the χ^2 distribution can be found in Ref. [76] or from the Python package Scipy [77].

where here we have neglected nuisance parameters θ. $\pi(\vec{C})$ is the prior probability distribution function, which takes into account any prior knowledge of the distribution of $\vec{C}$. If we could calculate the probability distribution $p(\vec{C}|D)$ from Eq. 2.3.10, constraints on $\vec{C}$ could then be derived from this. In practice calculating $p(\vec{C}|D)$ from Eq. 2.3.10 is often nontrivial, due to complicated likelihood functions and/or difficulties calculating the denominator of Eq. 2.3.10. Instead, we infer information about $p(\vec{C}|D)$ by sampling from $p(\vec{C}|D)$, using methods such as Markov Chain Monte Carlo (MCMC) or Nested Sampling [68]. In this thesis we will make use of the method of Nested Sampling, outlined in more detail in Appendix B.

Suppose we perform a fit to n Wilson coefficients $\vec{C}$, but are interested only in the constraints on C_1. In the Bayesian framework, to obtain constraints on C_1 only, we *marginalise* over the coefficients $C_2, ..., C_n$:

$$p(C_1|D) = \int d^{n-1}C p(\vec{C}|D)\,, \tag{2.3.11}$$

resulting in $p(C_1|D)$, the probability distribution for the coefficient of interest C_1. It can be shown that when working with the Gaussian likelihood, and assuming a constant prior ($\pi(\vec{C}) \propto const$), the result of this marginalisation coincides with the profiled limits discussed above [68]. As a result, the phrases *profiled* and *marginalised* are often used interchangeably in the literature, including within this thesis. In general, however, marginalisation and profiling produce different results.

2.4 New physics at high-p_T

2.4.1 Direct searches

While Chapter 3 will focus on a global SMEFT fit, in Chapters 4 and 5 we will turn to measurements in the high-p_T regime in search of new physics. In this context, hadron colliders are often known as *discovery machines*. While the clean backgrounds of lepton colliders allow them to perform precise measurements, hadron colliders benefit from the high proton mass, reaching high centre of mass energies and directly producing on-shell heavy new physics. Exciting prospects for these direct searches come from proposals for the future circular collider (FCC-hh): a collider of 100km circumference, expected to run with centre of mass energy $\sqrt{s} = 100$ TeV and collect an integrated luminosity of $\mathcal{L} = 20$ ab^{-1}. In the nearer future we await the High Luminosity LHC (HL-LHC), expected to run at 14 TeV and collect an integrated luminosity of $\mathcal{L} = 3$ ab^{-1}. At the time of writing Ref. [3], proposals also existed for a high energy hadron

collider (HE-LHC) with centre of mass energy $\sqrt{s} = 27$ TeV and luminosity $\mathcal{L} = 15$ ab^{-1}. We summarise these details in Table. 2.4.1.

	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [ab^{-1}]
LHC	13	0.14
HL-LHC	14	3
HE-LHC	27	15
FCC-hh	100	20

Table 2.4.1. Design centre of mass energies and integrated luminosities of the LHC Run II and future hadron colliders, adapted from Ref. [3].

It is important that we quantify the expected performance of each future collider. In particular, projections of future collider sensitivities to new physics provide crucial motivation for these projects. A benchmarking study was performed in Ref. [78], estimating the potential of the future colliders of Table 2.4.1 to uncover new physics through direct searches. Only loose assumptions about BSM particles were made, studying a range of two-body final states ($t\bar{t}$,jj,l^+l^-,W^+W^-) into which many BSM particles could decay. In Chapter 5, to further estimate the discovery potential of future colliders, we will focus on a specific new physics scenario: the scalar leptoquark. As we will discuss in §2.5, these BSM particles are well-motivated by measurements of discrepancies from the SM in rare B meson decays. Collider sensitivity to their discovery, however, is limited by SM background rates. The estimation of such background rates is therefore of vital importance to our estimates of the sensitivity. For each future collider, accurate simulations of the SM backgrounds, including the detector response, will be produced. By specialising the event selection, signal region, cuts and SM background to the leptoquark scenario we will provide a realistic estimate of the sensitivity of future colliders to the discovery of a leptoquark through direct searches.

2.4.2 Indirect searches

Hadron colliders are not limited to direct searches for new physics, however, and even measurements of high-p_T observables can indirectly constrain new physics. In Ref. [8] it was shown that measurements of the high-mass region of Drell-Yan invariant mass distributions have the potential to produce constraints on SMEFT operators which are competitive with those of the low-energy precision measurements from LEP. This is counterintuitive: hadron colliders are not generally associated with precision tests of new physics due to their complicated SM backgrounds. Schematically, we can

understand why this is so by expressing the leading contribution from dimension-6 operators to an amplitude $\mathcal{A}$ as

$$\begin{aligned}\mathcal{A} &\sim \mathcal{A}_{\text{SM}} + C\frac{E^2}{\Lambda^2} \\ &\sim \mathcal{A}_{\text{SM}} + C\frac{v^2}{\Lambda^2} + C\frac{vp}{\Lambda^2} + C\frac{p^2}{\Lambda^2},\end{aligned} \tag{2.4.1}$$

where v denotes the electroweak scale and p is the typical momentum transfer of the process. Contributions of order $\mathcal{O}(v^2/\Lambda^2)$ come from operators such as $\mathcal{O}_{dH} = (H^\dagger H)(\bar{q}_p d_r H)$ as discussed in §2.2.3. In contrast, four-fermion operators will produce contributions to the amplitude which scale as p^2/Λ^2. This quadratic scaling leads to an enhancement of the contribution from the EFT to the high-mass Drell-Yan tails, resulting in powerful constraints despite the large statistical uncertainties in these measurements.

As a result, the SMEFT has become the tool of choice for precision searches for new physics, even for high-mass, high-p_T measurements. ATLAS and CMS have used high-mass Drell-Yan measurements to constrain a set of four-fermion operators of the SMEFT [79, 80], while high-p_T searches have been recast to constrain four-fermion operators relevant to the neutral current B anomalies [81–84].

In Chapter 4 we will study the effects of four-fermion dimension-6 SMEFT operators on the high-mass region of Drell-Yan distributions. From the energy scaling of Eq. 2.4.1, we expect these operators to smoothly rescale the tail of the invariant mass distribution. To detect this subtle sign of new physics, we must have confidence in our theoretical predictions.

2.4.3 Parton distribution functions

Parton distribution functions (PDFs) form a key input into the calculation of LHC observables. Consider the diagram in Fig. 2.4.1 showing the leading contribution to the Drell-Yan process, mediated by a Z boson or photon. This process occurs when quark and antiquark constituents of the colliding protons interact with a high momentum transfer. In this high-energy regime, the partonic cross section $\hat{\sigma}\Big(q_f(x_1p_2)q_{\bar{f}}(x_2p_2) \to l^+l^-\Big)$ can be calculated in perturbation theory. The leading contribution in α_s is given by

$$\hat{\sigma}\Big(q_f(x_1p_2)q_{\bar{f}}(x_2p_2) \to l^+l^-\Big) = \frac{Q_f^2}{3}\frac{4\pi\alpha_{EW}^2}{3s} \tag{2.4.2}$$

where Q_f is the electric charge of the initial state quark and $s = (p_1 + p_2)^2$.

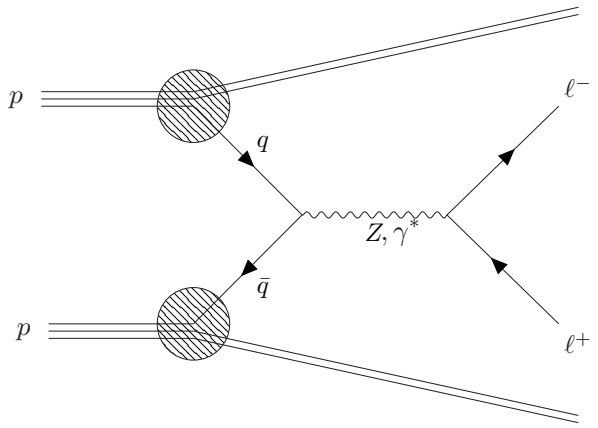


Fig. 2.4.1. The leading contribution to the Drell-Yan $pp \to l^+l^-$ process. The cross section factorises into the non-perturbative parton distribution functions describing the proton constituents, and the hard scattering process $q\bar{q} \to l^+l^-$.

To calculate the total cross section for the Drell-Yan process from this, we need the probability that quark f carries a fraction x of the incoming proton momentum. This is determined by the parton distribution function (PDF) $f_f(x, Q^2)$, where Q^2 denotes the energy scale at which the process takes place [85–87]. The total cross section is then given by

$$\sigma(p(p_1)p(p_2) \to l^+l^-) = \int_0^1 dx_1 \int_0^1 dx_2 \sum_f f_f(x_1, Q^2) f_{\bar{f}}(x_2, Q^2) \hat{\sigma}(q_f(x_1 p_2) q_{\bar{f}}(x_2 p_2) \to l^+l^-) + \mathcal{O}\left(\frac{\Lambda^2_{\text{QCD}}}{Q^2}\right). \tag{2.4.3}$$

The cross section is subject to the *Factorisation Theorem* of QCD [88]: it factorises into a short-distance piece, the hard scattering cross section $\hat{\sigma}$, and a long-distance piece governed by the PDFs where we sum over initial state quarks and antiquarks denoted by f and $\bar{f}$ respectively. This factorisation is not exact, however, and corrections of order $\mathcal{O}\left(\frac{\Lambda^2_{\text{QCD}}}{Q^2}\right)$ arise due to the underlying soft interactions of the remaining quark and gluon constituents of the proton [87].

The long-distance physics described by the PDFs is non-perturbative, and as a result the PDFs cannot be calculated in perturbation theory from first principles. Instead they are determined from fits to data by collaborations such as NNPDF [89], MMHT [90] and CTEQ [91]. The fitting procedure is made easier by two features of QCD. Firstly, perturbative QCD determines the Q^2 dependence of the PDF through

the DGLAP evolution equations. Therefore by determining $f(x, Q_0^2)$, the PDF can then be evolved to $f(x, Q_1^2)$ where $Q_1^2 \neq Q_0^2$. Secondly, as a result of the factorisation theorem, the PDFs are universal: once determined, the PDF $f(x, Q^2)$ can be used to calculate many observables.

An inconsistency arises, however, when we compare PDF fits and SMEFT fits. When we perform a SMEFT fit we compare a measured observable, σ_{meas}, with a theory prediction σ_{th} calculated by convoluting the partonic cross section $\hat{\sigma}_{th}$ with the PDF as in Eq. 2.4.3. We calculate $\hat{\sigma}_{th}$ assuming the SMEFT. In contrast, in PDF fits we assume only the SM, *not the SMEFT*. By doing so, we provide the SMEFT fit with an inconsistency in our theoretical predictions. The problem worsens when we consider the data incorporated into our SMEFT and PDF fits: often there is an overlap. For example, data involving jets are used in constraining the SMEFT in Ref. [92] and top-quark data are used in the global fit of Ref. [67], while NNPDF3.1 [89] use both of these types of data in their PDF fits.

These inconsistencies may lead to the reabsorption of new physics effects into the PDF parametrisation, particularly the subtle effects of new physics in high-p_T tails. Addressing this issue forms the second goal of this thesis. In Chapter 4 we will tackle this challenge by quantifying the interplay between PDF and EFT effects in high-energy Drell-Yan tails. Our work is an extension of a previous study [93] investigating this interplay in Deep Inelastic Scattering (DIS) observables, where a subtle interplay was observed. However, although DIS data provide important constraints on the PDFs, sensitivity to EFT effects is limited. By adding DY data to this study we will investigate whether this interplay remains when the EFT is subject to more stringent constraints.

2.5 Evidence for new physics in B meson decays

2.5.1 Neutral current B anomalies

Finally, we return to low energies and consider evidence for new physics in rare B meson decays. Flavour changing neutral currents (FCNC) are highly suppressed in the SM. Lepton flavour universality (LFU) forbids these processes from taking place at tree-level, and the leading loop-level diagrams are often further suppressed by small CKM elements and the GIM mechanism [94]. New physics may lead to significant contributions to these observables, particularly if it contributes at tree-level. Measurements of FCNC, therefore, provide an important low-energy probe of BSM physics [95, 96].

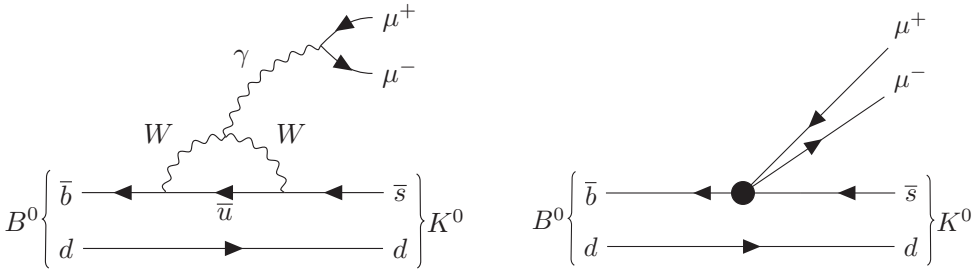


Fig. 2.5.1. Diagrams showing the decay $B^0 \to K^0\mu^+\mu^-$ in the SM (left) and through the insertion of a dimension-6 operator of the weak effective theory (right). The leading contribution from the SM is at 1-loop, whereas a dimension-6 operator allows this process to take place at tree-level.

An example of this is the branching ratio of the decay $B^0 \to K^0\mu^+\mu^-$. In the SM the leading diagram is at 1-loop, and one of the contributing diagrams is shown in Fig. 2.5.1 (left). LHCb have measured BR($B^0 \to K^0\mu^+\mu^-$) to be lower than the theoretical prediction, indicating a small discrepancy from the SM [97]. This is also true of measurements of the branching ratios BR($B_s \to \mu^+\mu^-$), BR($B^+ \to K^+\mu^+\mu^-$), BR($B^+ \to K^{+*}\mu^+\mu^-$), BR($B_s^0 \to \phi\mu^+\mu^-$) and BR($\lambda_b^0 \to \lambda^0\mu^+\mu^-$) [97–103]. Similar discrepancies from the SM are seen in measurements of the angular distributions of B decays [104–107] and in measurements of the ratios R_K and $R_{K^{(*)}}$, defined as the ratios of branching fractions

$$R_{K^{(*)}} \equiv \frac{BR(B \to K^{(*)}\mu^+\mu^-)}{BR(B \to K^{(*)}e^+e^-)}. \tag{2.5.1}$$

Measurements of R_K and R_{K^*} in particular have gathered attention. Calculations of most of the observables listed above include large hadronic uncertainties. This is not true of R_K and R_{K^*}: LFU in the SM implies that the hadronic uncertainties associated with the numerator and denominator cancel one another, making these theoretically 'clean' observables. Recently, LHCb measured R_K using 9 fb^{-1} of data, finding $R_K = 0.846^{+0.042+0.013}_{-0.039-0.012}$ in the range 1.1 GeV^2 $< q^2 <$ 6 GeV^2 where q^2 is the dilepton invariant mass squared [15]. This is a discrepancy of 3.1σ from the SM predictions. A previous analysis combining Run I and 2 fb^{-1} of LHCb Run II data measured $R_K = 0.846^{+0.06+0.016}_{-0.054-0.014}$ in the same range, a deviation of approximately 2.5σ [108, 109]. Similarly, R_{K^*} has been measured by LHCb as $R_{K^*}[0.045\text{GeV}^2 < \text{q}^2 < 1.1\text{GeV}^2] = 0.66^{+0.11}_{-0.07} \pm 0.03$ and $R_{K^*}[0.045\text{GeV}^2 < \text{q}^2 < 1.1\text{GeV}^2] = 0.69^{+0.11}_{-0.07} \pm 0.05$.

Each measurement shows a discrepancy of approximately 2.5σ, with the deviation following the same pattern as other measurements, suggesting values lower than the SM predictions. Measurements of R_{K^*} by Belle are less intriguing: these are compatible with both the SM and LHCb measurements due to larger uncertainties: $R_{K^*}[15\text{GeV}^2 < \text{q}^2 < 19\text{GeV}^2] = 1.18^{+0.52}_{-0.32} \pm 0.1$ and $R_{K^*}[0.1\text{GeV}^2 < \text{q}^2 < 8\text{GeV}^2] = 0.9^{+0.27}_{-0.21} \pm 0.1$ [110].

The branching ratio $\text{BR}(B_s \to \mu^+\mu^-)$ is also considered a theoretically clean observable. This process has a purely leptonic final state, and hadronic uncertainties enter only through the B_s decay constant which is calculated to high precision by lattice QCD methods [111, 112]. The most recent measurement of $\text{BR}(B_s \to \mu^+\mu^-)$ by LHCb found a result lower than the SM prediction [100], a discrepancy with a significance of 2.5 σ. Similar discrepancies were found in previous measurements by ATLAS and CMS [98, 99]. Ref. [113] combined these measurements with measurements of R_K and R_{K^*} in a fit to the SM, and found that collectively these theoretically clean observables are discrepant from the SM at a statistical significance of 4 σ.

This suggests that although individual discrepancies from the SM may be small, collectively they point towards new physics in $b \to sl^+l^-$ transitions. We will refer to these discrepancies as the *neutral current B anomalies* (NCBAs).

2.5.2 Effective field theory interpretation

Taken individually, the statistical significance of each of these measurements is too small to claim a discovery. However, incorporating them into a global fit can aid in their interpretation, investigating whether BSM physics is capable of improving the global description of the data. A number of global analyses exist: we will make use of the results of Refs. [114, 115].

As discussed in §2.3, the EFT framework provides a model-independent setting in which to interpret signs of new physics. In the case of low-energy flavour physics data, the weak effective theory is most relevant [116, 117]. We will use this EFT to describe physics at the scale $\mu \sim m_b$ where m_b denotes the b-quark mass. The weak effective theory depends on fewer degrees of freedom than the SMEFT: the $W^\pm$ and Z electroweak gauge bosons have been integrated out, as have the Higgs boson and the top quark. The effective Lagrangian contains the terms

$$\mathcal{L}_{eff}^{\text{NP}} = \frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \frac{e^2}{16\pi^2} \sum_i C_i O_i + h.c. \tag{2.5.2}$$

where O_i and C_i are the dimension-6 operators and their associated Wilson coefficients. A prefactor of $\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \frac{e^2}{16\pi^2}$ has been factored out of the Wilson coefficients, where V_{tb}

and V_{ts} denote elements of the CKM matrix. At the scale m_b, the SM gauge symmetry $SU(2)_L \times U(1)_Y$ has been broken down to $U(1)_{\text{EM}}$. In this regime we can treat the left-handed up and down-type quarks $u_{L,i}$, $d_{L,i}$ on a separate footing, allowing us to perform a rotation to the mass eigenbasis. The operators $\mathcal{O}_i$ of the weak effective theory are constructed from quarks in the mass eigenbasis. For the remainder of this Chapter we will denote the weak eigenbasis using primes, and the mass eigenbasis will be represented by unprimed notation.

Consider the subset of operators describing a 4-fermion interaction of two leptons and the b and s quarks. It is found in Ref. [115] that by performing a one-parameter fit of the data to each Wilson coefficient separately, new physics in the muon sector is favoured. A preference is found for new physics in the coefficients C_9 and C_{10}, describing the following vector and axial interactions:

$$\begin{aligned} O_9 &= (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \mu), \\ O_{10} &= (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu \gamma^5 \mu), \end{aligned} \tag{2.5.3}$$

with a pull of 5.7σ and 4.8σ from the SM respectively. Similarly, the combination $C_9 = -C_{10}$ is favoured over the SM with a pull of 5.9σ, indicating that new physics may couple preferentially to left-handed quarks and muons through the operator $\mathcal{O}_{LL}$:

$$\mathcal{O}_{LL} = \frac{1}{2}(\mathcal{O}_9 - \mathcal{O}_{10}) = (\bar{s}\gamma_\mu P_L b)(\bar{\mu}\gamma^\mu P_L \mu). \tag{2.5.4}$$

Fig. 2.5.1 (right) shows an example of how the insertion of one of these operators may modify the leading order diagram describing the $B^0 \to K^0\mu^+\mu^-$ process, allowing it to take place at tree-level.

2.5.3 Leptoquarks

The operators $\mathcal{O}_9$, $\mathcal{O}_{10}$ and their combination $\mathcal{O}_{LL}$ may arise due to heavy new physics mediating an interaction between muons and b,s quarks. An interesting BSM candidate is the leptoquark. A number of scalar leptoquark models have been proposed to solve the NCBAs, as well as models of vector leptoquarks and Z' bosons, often arising due to a $U(1)'$ extension of the SM gauge group [118–129].

We will consider a simple model of a scalar leptoquark with couplings to left-handed quarks b_L, s_L and muons μ_L. Following the notation of [130], a complex $SU(2)_L$ triplet S_3, with quantum numbers $(\bar{3}, 3, \frac{1}{3})$ under the standard model gauge group can produce this interaction. Fig. 2.5.2 shows how this leptoquark may mediate the $b \to s\mu^+\mu^-$

transition at tree-level. The Yukawa couplings of the leptoquark to the i^{th}-family

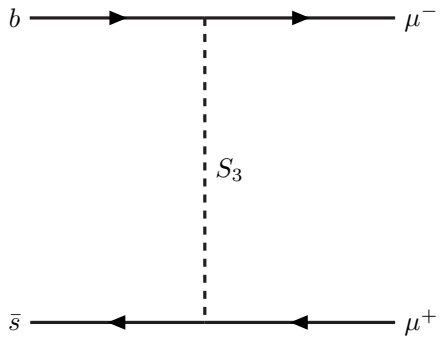


Fig. 2.5.2. Leptoquark S_3 mediating a tree-level $b \to s\mu^+\mu^-$ transition.

SM quark (q'_i) and lepton (l'_i) $SU(2)_L$ doublets are given by (in the primed weak eigenbasis) [126, 127]:

$$\mathcal{L}_{\text{Yukawa}} = (Y_l)_{ij}\overline{q^{C\prime}}_{i,a}\epsilon_{ab}\tau^k_{bc}l'_{j,c}S^k_3 + (Y_q)_{ij}\overline{q^{C\prime}}_{i,a}\epsilon_{ab}\tau^k_{bc}q'_{j,c}(S^k_3)^\dagger + h.c., \tag{2.5.5}$$

where we have suppressed QCD gauge indices, $i, j \in \{1,2,3\}$ are family indices, $a, b, c \in \{1,2\}$ are fundamental $SU(2)_L$ indices, $k \in \{1,2,3\}$ is an adjoint $SU(2)_L$ index, the superscript C denotes a charge conjugated fermion, ϵ_{ab} is the Levi Civita symbol, the τ^k_{ab} are the Pauli matrices and Y_l and Y_q are 3×3 matrices of complex dimensionless Yukawa couplings.

In order to avoid proton instability we assume that baryon number is conserved, setting $(Y_q)_{ij}$ to zero. After EWSB S_3 becomes ($S^{-2/3}$, $S^{+1/3}$, $S^{+4/3}$) where the superscript denotes electric charge. The left-handed quarks and leptons mix according to

$$\mathbf{P}'^T = V_P\mathbf{P}^T, \tag{2.5.6}$$

where $P \in \{u_L,\ d_L,\ e_L,\ \nu_L\}$ and bold face denotes a 3-vector in family space. V_P are then unitary dimensionless 3 by 3 matrices, being experimentally constrained via the CKM combination $V_{\text{CKM}} = V^\dagger_{u_L}V_{d_L}$ and the PMNS combination $U_{\text{PMNS}} = V^\dagger_{\nu_L}V_{e_L}$. With these definitions we arrive at the couplings of the leptoquark to the leptons and quarks in the mass eigenbasis,

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} = &-\sqrt{2}\,\overline{\mathbf{d}^C_L}Y_{de}\mathbf{e}_L S^{+4/3} - \overline{\mathbf{u}^C_L}Y_{ue}\mathbf{e}_L S^{+1/3} \\ &- \overline{\mathbf{d}^C_L}Y_{d\nu}\nu_L S^{+1/3} + \sqrt{2}\,\overline{\mathbf{u}^C_\mathbf{L}}Y_{u\nu}\nu_L S^{-2/3} + h.c.,\end{aligned} \tag{2.5.7}$$

where $Y_{de} = V_{d_L}^T Y_l V_{e_L}$, $Y_{ue} = V_{u_L}^T Y_l V_{e_L}$, $Y_{d\nu} = V_{d_L}^T Y_l V_{\nu_L}$, and $Y_{u\nu} = V_{u_L}^T Y_l V_{\nu_L}$.

Assuming the leptoquark has a large mass m_{LQ}, we can describe the interaction of Fig. 2.5.2 using the operators of the weak effective theory. Integrating the leptoquark out of the theory and comparing the resulting interaction to the weak effective theory Lagrangian of Eq. 2.5.2, we find the following relation:

$$\begin{aligned} \frac{(Y_{de})_{32}(Y_{de}^*)_{22}}{m_{LQ}^2} &= C_{LL}\frac{4G_F}{\sqrt{2}}V_{tb}V_{ts}^*\frac{e^2}{16\pi^2} \\ &= \frac{C_{LL}}{(36\text{TeV})^2}, \end{aligned} \tag{2.5.8}$$

indicating that we need nonzero $(Y_{de})_{32}$ and $(Y_{de})_{22}$ to explain the NCBAs. Taking the value of C_{LL} to be $C_{LL} = -1.06 \pm 0.16$ from the best-fit point of the global fit of Ref. [114], we constrain the ratio of leptoquark couplings and mass to take the value

$$\frac{(Y_{de})_{32}(Y_{de}^*)_{22}}{m_{LQ}^2} = (8.2 \pm 1.2) \times 10^{-4}\text{TeV}^{-2}. \tag{2.5.9}$$

We will typically use the central value from this fit in order to fix $(Y_{de})_{32}(Y_{de}^*)_{22}$ for a given value of m_{LQ}. Although in general Y_{de} are complex, we will here take real values for simplicity and because we are not considering CP-violating observables.

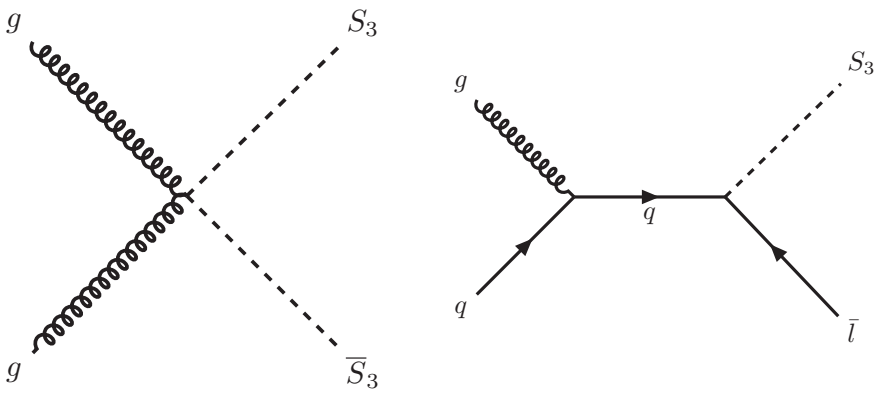


Fig. 2.5.3. Leptoquark production mechanisms at hadron colliders. The dominant production mechanism is pair production (left) through gluon-gluon fusion. This mechanism is independent of the couplings of the leptoquark to leptons and quarks. In contrast, single production (right) always occurs in association with a lepton.

As we saw in §2.4.1, direct searches aim to discover BSM particles through their direct production. Fig. 2.5.3 (left) shows the dominant production mechanism, gluon-gluon fusion, through which the pair production of leptoquarks at hadron colliders

may take place. Since by definition a leptoquark couples to a quark and a lepton, it must carry colour to preserve $SU(3)$ and therefore, by $SU(3)$ gauge symmetry, must couple to gluons via the QCD coupling constant. In contrast, single production of a leptoquark is always dependent on the leptoquark coupling Y_{ql} as shown by Fig. 2.5.3 (right). Assuming the size of Y_{ql} required to explain the NCBAs in Eq. 2.5.9, and taking into consideration the constraints on Y_{ql} from previous searches [131, 132], this production mechanism is expected to be subdominant. Turning to the leptoquark decay, the NCBAs predict that there should be couplings between S_3 and $\bar{s}$, μ or $\bar{b}$, μ from the first term in Eq. 2.5.7. Thus we expect the decay channels $S_3 \to \bar{b}\mu$ and $S_3 \to \bar{s}\mu$ to be open, suggesting that searches for scalar leptoquarks in the $\mu^+\mu^- jj$ channel may provide an insight into the NCBAs.

The most recent search for the pair production of scalar leptoquarks by ATLAS uses 139 fb^{-1} of Run II 13 TeV data and excludes leptoquarks in the $\mu^+\mu^- jj$ channel with masses up to $m_{LQ} = 1.7$ TeV [133]. In this analysis, the bottom quark remains untagged and is counted merely as a jet. We can therefore apply these constraints to leptoquarks solving the NCBAs. Looking to the future, the hadron colliders described in §2.4.1 will provide new information on scalar leptoquarks as well as on the NCBAs themselves [134], concurrently with Belle II [135, 136]. This brings us to the final goal of this thesis. In Chapter 5 we will estimate the sensitivity of future hadron colliders to scalar leptoquarks in the $\mu^+\mu^- jj$ channel. Assuming these leptoquarks solve the NCBAs, we will address the question of whether limits on leptoquark masses will be increased by future hadron colliders. Or, more optimistically, we will ask: *if the NCBAs are signs of new physics and scalar leptoquarks are responsible, could these leptoquarks be discovered in the near future?*

This Chapter began with an introduction to effective field theories and the Standard Model Effective Field Theory, followed by an overview of how these theoretical frameworks are crucial to the search for new physics: in measuring their Wilson coefficients in fits to data, in parametrising new physics in parton distribution functions and in their use to interpret anomalous measurements in flavour physics data. We will now turn to the first of these, focusing our attention on a global fit to Higgs, electroweak, diboson and top data.

Chapter 3

Top, Diboson, Higgs and Electroweak fit to the SMEFT

3.1 Introduction

The goal of this Chapter is to present a fit of the SMEFT to data from the top, Higgs, diboson and electroweak sectors. This is the first global dimension-6 SMEFT analysis to include top data and operators in a simultaneous combination of the constraints from the Higgs, electroweak, diboson and top sectors. This Chapter is based on Ref. [1] and all results are produced using a newly-developed public code called `Fitmaker`.

The global analysis we present here is performed at linear order in the dimension-6 SMEFT operator coefficients. Our SMEFT predictions for the observables are computed to linear order in the SMEFT operator coefficients, while quadratic corrections from dimension-6 operators are neglected. The theoretical prediction for a cross section σ takes the form

$$\sigma = \sigma_{\rm SM} + \sum_i a_i \frac{C_i}{\Lambda^2} + \mathcal{O}(\frac{1}{\Lambda^4}) \tag{3.1.1}$$

where $\sigma_{\rm SM}$ denotes the contribution from the SM Lagrangian, and a_i denotes the contribution from the operator associated with Wilson coefficient C_i. For operators where quadratic contributions are non-negligible, such as in the top sector as shown by Ref. [63], the linear constraints can be viewed as provisional. Nevertheless, keeping those operators in the global fit typically yields conservative marginalised limits and allows one to assess where the impact on other operators can be significant to a first approximation. As discussed in §2.3.2, the inclusion of quadratic contributions should be done with caution as they are generically of the same order as linear interference terms with coefficients of dimension-8 operators, though exceptions exist in some

specific UV completions [137]. For example, in a UV completion with a single particle and a single coupling the quadratic dimension-6 contributions can be larger than the linear dimension-6 ones in the strong-coupling regime, though not more generally. The importance of their effect is therefore a model-dependent question. We note also that sensitivity of a linear fit to quadratic contributions is an indicator for a possible breakdown in the regime of validity of the SMEFT, so care must be taken in the interpretation of the fit.

We will formulate the dimension-6 SMEFT in terms of the Warsaw basis. As we saw in Chapter 2, the representation of the Warsaw basis given in Table 2.2.2 involves a vast number of operators. We will reduce this number by imposing a symmetry such that some operators are forbidden or made redundant, firstly making use of the flavour universal symmetry introduced in §2.2.3 to reduce the number of operators relevant to our fit to a total of 20. This symmetry enforces universality among fermion flavours, placing the right-handed up and top quarks on the same footing, for example. This is not ideal when dealing with top quark data: it would be preferable to single out the top quark as separate from the remaining quarks. To do so, we will relax the requirement of flavour universality and impose a less restrictive symmetry, the 'top-specific' symmetry, the details of which will be outlined in §3.4.2. This scenario increases the number of operators in our fit to a total of 34.

The `Fitmaker` code incorporates two types of fitting methodology. The first of these is a frequentist fitting methodology which involves the calculation and minimisation of a χ^2 test statistic. We perform a χ^2 fit for a vector of observables, $\vec{y}$, with covariance matrix, $\mathbf{V}$, and theory predictions for those observables, $\vec{\mu}(C_i)$, depending on Wilson coefficient C_i, using a χ^2 function defined as

$$\chi^2(C_i) = (\vec{y} - \vec{\mu}(C_i))^T \, \mathbf{V}^{-1} \, (\vec{y} - \vec{\mu}(C_i)) \ . \tag{3.1.2}$$

The best-fit Wilson coefficients can then be found by a simple χ^2 minimisation, and confidence intervals can be calculated as outlined in §2.3.3. As we will discuss in more detail in §3.2, the procedure of finding the best-fit point and constraints can be performed analytically. As a result, this χ^2 fitting methodology is very fast and efficient, and we will see that the calculational speed of the χ^2 fitting methodology makes possible a broad-band search for BSM physics.

`Fitmaker` also incorporates a Bayesian method of parameter estimation, using which we can also perform fits of the SMEFT to data. This Bayesian method is the nested sampling algorithm [138], outlined in more detail in Appendix B. The nested sampling fitting framework is not limited in principle to linear order: we illustrate in

Appendix B, focusing on Higgs data, how our procedure can be used for quadratic fits as well, though a global analysis at quadratic order is beyond the scope of this work. Within this Chapter, the Bayesian method of nested sampling is used only once, in §3.7. We use this methodology to produce one-dimension constraints on single field extensions of the SM, making use of the fact that it allows us to place a prior on the model parameters to enforce initial conditions on the parameters. This will be outlined in more detail in §3.7. Aside from this, all results in this Chapter are produced using the frequentist χ^2 fitting methodology.

This analysis is an update to a previous study, Ref. [54]. Updates include improvements to the data incorporated in the fit: our analysis uses a full set of data from LHC Run 2, in particular the latest Higgs Simplified Template Cross Section (STXS) measurements, differential distributions in WW diboson and Zjj measurements, and updated top observables including kinematic distributions, $t\bar{t}$, single-top and $t\bar{t}W/Z$ production. In addition to expanding our dataset, updates include improvements in the SMEFT predictions. We compute the dimension-6 contributions to Higgs gluon-gluon fusion in the STXS bins using `SMEFT@NLO` [139], including the interference between the loop diagram of the SM with tree-level diagrams and loop diagrams of the SMEFT. We incorporate the full SMEFT dependence in off-shell Higgs to 4 lepton decays [4]. We also provide a self-consistent treatment of the triple-gluon operator $\mathcal{O}_G$, including it at linear order. This operator had been omitted from previous fits [140] on the basis of strong constraints at quadratic order [141–143].

Our principal results are displayed in Fig. 3.6.4, in which we show the individual and marginalised constraints on 34 dimension-6 operators of the SMEFT from a fit to a combination of top, Higgs, diboson and electroweak data. Numerical results for these plots are presented in Table C.0.1. We do not find any significant discrepancy with the SM. We find a small preference for a non-zero value for the coefficient C_G of the triple-gluon operator $\mathcal{O}_G$. This discrepancy is driven by discrepancies observed in top data and, as we will discuss in §3.6.3, the discrepancy is mirrored by trends in the coefficients of other operators affecting the tt cross section. We await future theoretical developments and measurements in the top sector to provide further understanding of these discrepancies. In §3.6.3 we also analyse the interplay between the top and Higgs sectors, including a dedicated discussion of sensitivities in 'Higgs-only'[1] operator planes that illustrate the impact of the top data and of the triple-gluon operator coefficient

[1]We stress that naming these operators as 'Higgs-only' is a basis-dependent statement. We work in the Warsaw basis. However, given another choice of basis for the dimension-6 operators of the SMEFT, these operators may not necessarily depend on the Higgs boson, or may incorporate additional SM fields.

C_G. We analyse correlations between the coefficients of the 34 operators and perform a principal component analysis, identifying the most and least constrained combinations of SMEFT operators. We find that the top and Higgs sectors do have an influence on one another, but that the interplay is moderate at present.

The layout of this Chapter is as follows. We will begin in §3.2 by setting out our fitting procedure. This is described in more detail in Appendix B, including a nested sampling calculational method that we illustrate in an analysis of Higgs signal strengths to quadratic order in the dimension-6 operator coefficients. §3.3 summarises the dataset that we use in the global fit, which is described in more detail in Appendix A. In §3.4 we will review the SMEFT framework we use, and §3.5 describes how theory predictions are calculated. The results from our global linear fit are presented in §3.6, where we display results from the electroweak, Higgs and top sectors separately and in combination. We derive constraints for all individual operators and also constraints including all dimension-6 operators and marginalising. Applications of our analysis to some specific BSM scenarios are presented in §3.7, including single-field extensions of the SM and a light-stop scenario, and a survey of fits with contributions from any combination of two, three, four or five dimension-6 operators, none of which provide any significant evidence for physics beyond the SM. Our conclusions are summarised in §3.8.

3.2 Fitting procedure

Our global analysis is produced using a frequentist fitting methodology. We perform a χ^2 fit for a vector of observables, $\vec{y}$, with covariance matrix,[2] $\mathbf{V}$, and theory predictions for those observables, $\vec{\mu}(C_i)$, using a χ^2 function defined as

$$\chi^2(C_i) = (\vec{y} - \vec{\mu}(C_i))^T \, \mathbf{V}^{-1} \, (\vec{y} - \vec{\mu}(C_i)) \; . \tag{3.2.1}$$

The predictions are functions of the dimension-6 operator coefficients C_i, and are truncated at the linear level so as to include only the interference term with the SM. The least-squares estimators $\hat{C}_i$ that extremise the χ^2 function can be obtained analytically in the case of a linear fit (see, e.g., Ref. [144] for a review). We may write the linear theory prediction in terms of a matrix $\mathbf{H}$ that characterises the modification

[2] In certain cases the experimental data is provided with non-symmetric uncertainties, i.e. $\sigma = \sigma_{\text{exp}} + \Delta\sigma_1 - \Delta\sigma_2$. We define a symmetrised uncertainty as $\Delta\sigma = \frac{1}{2}(\Delta\sigma_1 + \Delta\sigma_2)$. The diagonal entries of the covariance matrix $\mathbf{V}$ are then computed from $\Delta\sigma$: schematically, we can write this as $V_{ii} = (\Delta\sigma)^2$.

of the SM predictions $\vec{\mu}^{\rm SM}$ at linear order:

$$\mu_\alpha(C_i) = \mu_\alpha^{\rm SM} + \mathbf{H}_{\alpha i} C_i \,. \tag{3.2.2}$$

A summation over repeated indices is implied; the index α ranges over the number of observables and i ranges over the number of dimension-6 coefficients. Solving $\partial\chi^2/\partial C_i = 0$ gives the best-fit values as

$$\hat{\vec{C}} = \left(\mathbf{H}^T\mathbf{V}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{V}^{-1}(\vec{y} - \vec{\mu}^{\rm SM}) \equiv \mathbf{F}^{-1}\vec{\omega}\,. \tag{3.2.3}$$

It is convenient to define the symmetric Hessian matrix $\mathbf{F}$, also known as the Fisher information matrix, and the χ^2 gradient vector $\vec{\omega}$ as

$$\mathbf{F} \equiv \mathbf{H}^T\mathbf{V}^{-1}\mathbf{H} \quad , \quad \vec{\omega} \equiv \mathbf{H}^T\mathbf{V}^{-1}(\vec{y} - \vec{\mu}^{\rm SM})\,, \tag{3.2.4}$$

in terms of which the χ^2 function Eq. (4.3.5) can be written as

$$\chi^2(C_i) = \chi^2_{\rm SM} - 2\vec{C}^T\vec{\omega} + \vec{C}^T\mathbf{F}\vec{C} = \chi^2_{\rm min} + \left(\vec{C} - \hat{\vec{C}}\right)^T\mathbf{F}\left(\vec{C} - \hat{\vec{C}}\right), \tag{3.2.5}$$

where $\mathbf{F} \equiv \mathbf{U^{-1}}$ is the inverse of the covariance matrix of the least-squares estimators, $\hat{\vec{C}}$. From this we may calculate the pull P:

$$P \;\equiv\; \sqrt{\chi^2_{\rm SM} - \chi^2(C_i)}\,. \tag{3.2.6}$$

Splitting the coefficients into $\vec{C} = \{\vec{C}_A, \vec{C}_B\}$, we may profile over a subset of coefficients $\vec{C}_A$ to obtain the least-squares estimators $\hat{\vec{C}}_B$ for the remaining coefficients $\vec{C}_B$. For this purpose, the Fisher information matrix may be decomposed into the sub-matrices

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_A & \mathbf{F}_{AB} \\ \mathbf{F}_{AB}^T & \mathbf{F}_B \end{pmatrix}, \tag{3.2.7}$$

and the gradient vectors as $\vec{\omega} = \{\vec{\omega}_A, \vec{\omega}_B\}$. The profiled best-fit values are then given by

$$\hat{\vec{C}}_B = \left(\mathbf{F}_B - \mathbf{F}_{AB}^T\mathbf{F}_A^{-1}\mathbf{F}_{AB}\right)^{-1}\left(\vec{\omega}_B - \mathbf{F}_{AB}^T\mathbf{F}_A^{-1}\vec{\omega}_A\right). \tag{3.2.8}$$

In cases where a prior on the coefficients needs to be imposed, for example when the magnitude-squared of couplings cannot go negative as when matching to specific UV models in §3.7, or when including quadratic dependences on the coefficients, this analytic method may no longer be used. A Bayesian method of parameter estimation,

using the numerical nested sampling algorithm and the `MultiNest` code, has therefore also been implemented in `Fitmaker`, as described in Appendix B.

3.3 Data

In this Section we describe the data used in our global fit. We summarise in §3.3.1 the main categories of data and refer the reader to Appendix A for a complete list of the observables that have been implemented in `Fitmaker`, together with their source references. In §3.3.2 we then discuss how these datapoints are incorporated into the fit.

The most precise electroweak measurements, other than the W mass, remain those from LEP and the SLC.[3] The Higgs boson discovery at the LHC enabled the possibility of a closed global SMEFT fit to a complete set of dimension-6 operators for the first time. Higgs physics has since progressed rapidly to include more channels and sub-categories beyond signal strengths. In particular, the STXS categorisations of the various Higgs production sub-channels provide further sensitivity to different directions in the parameter space, as illustrated, for example, in Fig. 3.5.1 for the case of gluon fusion and described further in the next Section.

The higher energies at the LHC also allow certain measurements of diboson and dilepton final states to become competitive with LEP [148, 149, 8, 150–152, 9], enable complementary probes of higher-dimensional operators [153, 154], and, moreover, give access to top physics with higher statistics than ever before, including the previously unreachable $t\bar{t}W/Z/H$ and other, rare production processes such as four-top production [155, 156, 63, 157]. More operators, under less restrictive flavour assumptions, can then be included in a global SMEFT fit. This is particularly motivated since the top quark is often expected to be more sensitive to BSM physics.

3.3.1 Dataset description

The following is a summary of the different categories of observables that we consider—see Appendix A for more details and references. The data fall into four categories, as follows. We provide a counting of our measurements where each datapoint is counted as a separate measurement. For example, each bin of a binned distribution is individually counted.

- The set of electroweak precision observables (EWPOs) includes the pseudo-observables measured on the Z resonance by LEP and SLD, together with the

[3] We note that global SMEFT fits would benefit greatly from a future Z-pole run [145–147].

W boson mass measurements by CDF and D0 at the Tevatron and ATLAS at the LHC:

$$\text{EWPO:} \quad \{\Gamma_Z, \sigma^0_{\text{had.}}, R^0_l, A^l_{FB}, A_l, R^0_b, R^0_c, A^b_{FB}, A^c_{FB}, A_b, A_c, M_W\}\,. \tag{3.3.1}$$

See Table A.0.2 for more details of the measurements of EWPO. *We include a total of 14 electroweak measurements.*

- For diboson measurements, we include the W^+W^- measurements of total cross-sections at different energies and angular distributions at LEP, the fiducial differential cross-section in leading lepton p_T by ATLAS at the LHC, and ATLAS and CMS fiducial differential cross-section measurements of the Z-boson p_T in leptonic $W^\pm Z$ production. We also incorporate the differential distribution in $\Delta\phi_{jj}$ for the Zjj measurement given by ATLAS, which we include in the diboson category because it is sensitive to related physics. Further details of the diboson measurements are included in Table A.0.1. *We include a total of 118 diboson measurements.*

- The Higgs dataset at the LHC includes the combination of Higgs signal strengths by ATLAS and CMS for Run 1, and for Run 2 both signal strengths and STXS measurements are used. ATLAS in particular provide the combined stage 1.0 STXS for $4l, \gamma\gamma, WW^*, \tau^+\tau^-$ and $b\bar{b}$, while for CMS we use the signal strengths of $4l$, $\gamma\gamma$, WW^*, $\tau^+\tau^-$, $b\bar{b}$ and $\mu^+\mu^-$. We also include the $Z\gamma$ signal strength from ATLAS and a differential WW^* cross-section measurement from CMS. See Tables A.0.3 and A.0.4 for more details of the Higgs measurements. *We include a total of 72 Higgs measurements.*

- The top data consists of differential distributions in various $t\bar{t}$ channels and cross-section measurements of top pair production in association with a W/Z boson or a photon (the $t\bar{t}V$ dataset), as well as various single top differential and inclusive cross-section measurements, for both Runs 1 and 2. Tables A.0.5, A.0.6 and A.0.7 contain further details of the top measurements. *We include a total of 137 top measurements.*

Overall, we include a total of 341 measurements in our analysis.

3.3.2 Signal strengths

We build our selected dataset by combining statistically independent measurements, including correlation information by means of published covariance/correlation matrices, when available.[4] In general, for LHC data, this amounts to a single ATLAS and CMS measurement of a particular final state for each LHC run. When multiple measurements, e.g., differential distributions, are reported, a single distribution is chosen to avoid statistical overlap between events, as information about the statistical correlation between distributions is generally not provided. This choice of differential distribution is made based on maximising the sensitivity of our fit. If not already given in such a form, each measurement is converted into a corresponding 'signal strength', μ, defined as the ratio of the observed value to the best available theory prediction,

$$\mu \equiv \frac{\sigma_{\text{exp}}}{\sigma_{\text{SM}}}\,, \tag{3.3.2}$$

where σ_{exp} denotes the measured value of the observable σ and σ_{SM} denotes the SM theory prediction for the observable.

Values for σ_{exp} are taken from the publications and their associated entries on `HEPdata`, where available. In some cases the data consist of differential measurements, provided either in the form of absolute differential cross sections ($\frac{d\sigma}{dx}$) or normalised differential cross sections ($\frac{1}{\sigma}\frac{d\sigma}{dx}$, where σ denotes the measured total cross section and x denotes an arbitrary kinematic variable). The differential data are taken from the publication and its associated entry in `HEPdata`, where available, using absolute differential cross section measurements, $\vec{\sigma}_{\text{abs}}$, and their associated covariance matrices, $\mathbf{\Sigma}_{\text{abs}}$. If only normalised differential cross sections ($\vec{\sigma}_{\text{norm}}$) are published, they are converted to absolute ones using the best available measurement of the inclusive cross section for that process in the same channel, denoted by $\sigma_{\text{tot}} \pm \delta\sigma_{\text{tot}}$. Covariance matrices are then updated to reflect the correlations between the bins induced by the common rescaling by the total cross section. The absolute differential measurement and its covariance matrix are then

$$\begin{aligned}
\vec{\sigma}_{\text{abs}} &= \vec{\sigma}_{\text{norm}}\sigma_{\text{tot}}\,, \\
\mathbf{\Sigma}_{\text{abs}} &= \mathbf{\Sigma}_{\text{norm}}\sigma_{\text{tot}}^2 + \delta\sigma_{\text{tot}}^2\vec{\sigma}_{\text{norm}} \otimes \vec{\sigma}_{\text{norm}}\,.
\end{aligned}$$

We can then proceed to use Eq. 3.3.2 to convert σ_{exp} or $\vec{\sigma}_{\text{abs}}$ into a signal strength. We take the value for σ_{SM} from the numbers quoted in the experimental publications:

[4] See Ref. [158] for a study of the impact of correlations in global fits.

in most cases this represents the best available prediction for $\sigma_{\rm SM}$, accounting for higher order QCD and EW corrections where appropriate. Where available, `fastnlo` tables [159–161] were used to obtain NNLO QCD predictions for the differential $t\bar{t}$ data. The SM theoretical errors are taken to be uncorrelated and the relative signal strength covariance matrix is obtained by adding the relative experimental and theoretical covariances as follows:

$$\mathbf{\Sigma}_\mu = \left(\frac{\mathbf{\Sigma}_{\rm exp}}{\vec{\sigma}_{\rm exp} \otimes \vec{\sigma}_{\rm exp}} + {\rm diag}(\vec{\delta}_{\rm th}/\vec{\sigma}_{\rm th})^2 \right) (\vec{\mu} \otimes \vec{\mu}), \quad \text{where} \quad \vec{\mu} \equiv \frac{\vec{\sigma}_{\rm exp}}{\vec{\sigma}_{\rm th}} \,. \tag{3.3.3}$$

This corresponds to adding the relative experimental and theory uncertainties in quadrature. The observables are stored in the `Fitmaker` database in `json` format, together with metadata and information about how each signal strength was obtained.

3.4 SMEFT Framework

Our global analysis is performed at dimension-6 in the SMEFT, assuming the Warsaw basis. We neglect CP-violating operators as well as interactions violating L and B. Two flavour scenarios are imposed on the SMEFT to reduce the number of degrees of freedom in our fit. The first is the flavour-universal $SU(3)^5$ symmetry, which we utilise in a fit to electroweak precision observables, Higgs data and diboson data. We relax this symmetry to form a *top-specific* flavour symmetry when including top quark data in our global fit. The top-specific flavour symmetry breaks flavour universality in the up-type quark sector, singling out operators involving the top quark and allowing us the flexibility to fit these operators to top-quark data. We will begin by reviewing these symmetries and introducing the operators constrained by our analysis.

3.4.1 Flavour-universal scenario

Assuming $SU(3)^5$ symmetry reduces the Warsaw basis to the flavour-universal scenario in which only the operators in cells not shaded in Table 2.2.2 are allowed, with common flavour-diagonal Wilson coefficients and no off-diagonal entries. Neglecting CP-violating interactions, one is left with 31 degrees of freedom, of which 16 are relevant for a leading-order fit to electroweak precision, diboson and Higgs data. To this we also add 4 operators that explicitly break the flavour-universal symmetry. These operators affect Higgs physics through a shift of the tau, muon, b-quark and top-quark Yukawa couplings. These are included as they are well constrained by data in our fit,

in particular measurements in the $H \to \tau\bar{\tau}$, $\mu\bar{\mu}$, $b\bar{b}$, and $t\bar{t}$ decay channels. The 20 operators in our flavour-universal scenario are then the following:

$$\begin{aligned} &\text{EWPO:} && \mathcal{O}_{HWB}\,,\ \mathcal{O}_{HD}\,,\ \mathcal{O}_{ll}\,,\ \mathcal{O}_{Hl}^{(3)}\,,\ \mathcal{O}_{Hl}^{(1)}\,,\ \mathcal{O}_{He}\,,\ \mathcal{O}_{Hq}^{(3)}\,,\ \mathcal{O}_{Hq}^{(1)}\,,\ \mathcal{O}_{Hd}\,,\ \mathcal{O}_{Hu}\,, \\ &\text{Bosonic:} && \mathcal{O}_{H\Box}\,,\ \mathcal{O}_{HG}\,,\ \mathcal{O}_{HW}\,,\ \mathcal{O}_{HB}\,,\ \mathcal{O}_{W}\,,\ \mathcal{O}_{G}\,, \\ &\text{Yukawa:} && \mathcal{O}_{\tau H}\,,\ \mathcal{O}_{\mu H}\,,\ \mathcal{O}_{bH}\,,\ \mathcal{O}_{tH}\,. \end{aligned} \tag{3.4.1}$$

We have categorised these operators roughly, into sets that are mostly constrained by electroweak precision observables (EWPO), those that can only be constrained at tree-level by Higgs and diboson measurements (bosonic), and operators that induce shifts in the Yukawa couplings (Yukawa).

3.4.2 Top-specific flavour scenario

The minimal flavour scenario that singles out top-quark couplings relaxes the $SU(3)^5$ symmetry as follows [11]:

$$\begin{aligned} SU(3)^5 &\to SU(2)^2 \times SU(3)^3 \\ &= SU(2)_q \times SU(2)_u \times SU(3)_d \times SU(3)_l \times SU(3)_e\,. \end{aligned}$$

By relaxing $SU(3)_q \to SU(2)_q$ we single out the top and bottom quark $SU(2)_L$ doublet and enforce flavour universality only on the first two generations of quark doublets. Similarly, relaxing $SU(3)_u \to SU(2)_u$ singles out the top quark singlet, enforcing flavour universality only among the right-handed up-type fields. This allows chirality-flipping interaction terms involving the third-generation quark doublet and right-handed up-type fields, notably the top-quark Yukawa interaction. The following three additional dimension-6 operators in the SMEFT are now allowed:[5]

$$[\mathcal{O}_{uG}]^{33} = \mathcal{O}_{tG}\,, \quad [\mathcal{O}_{uB}]^{33} = \mathcal{O}_{tB}\,, \quad [\mathcal{O}_{uW}]^{33} = \mathcal{O}_{tW}\,. \tag{3.4.2}$$

[5]The analogous operator $[\mathcal{O}_{uH}]^{33} = \mathcal{O}_{tH}$ is already included in the set of Yukawa operators listed in Eq. (3.4.1).

The flavour-universality conditions on operators in the $\psi^2 H^2 D$, $\bar{L}L\bar{L}L$, $\bar{R}R\bar{R}R$ and $\bar{L}L\bar{R}R$ classes that contain q or u are also relaxed. Schematically,

$$C_{\text{univ.}} \sum_{i=1,2,3} \mathcal{K}^\mu \bar{f}_i \gamma_\mu f_i \to \begin{cases} C_3 \mathcal{K}^\mu \bar{f}_3 \gamma_\mu f_3 \,, \\ C_{\text{univ.}} \sum_{j=1,2(,3)} \mathcal{K}^\mu \bar{f}_j \gamma_\mu f_j \,, \end{cases} \tag{3.4.3}$$

where $\mathcal{K}^\mu$ is a combination of other fields. Here a choice must be made in the second line of Eq. 3.4.3 whether to split the degrees of freedom into a fully-universal operator that preserves the full flavour symmetry or an operator that respects only the reduced symmetry $SU(2)^2 \times SU(3)^3$, corresponding to keeping or removing the index in red, respectively. The two are related by a basis rotation. We adopt the second option, since it better separates the degrees of freedom that affect only top measurements from the rest. The $\psi^2 H^2 D$ class grows to

$$\begin{aligned} [\mathcal{O}_{Hq}^{(1)}] &\to [\mathcal{O}_{Hq}^{(1)}]^{j,j} \text{ and } [\mathcal{O}_{Hq}^{(1)}]^{3,3} = \{\mathcal{O}_{Hq_i}^{(1)}, \mathcal{O}_{HQ}^{(1)}\} \,, \\ [\mathcal{O}_{Hq}^{(3)}] &\to [\mathcal{O}_{Hq}^{(3)}]^{j,j} \text{ and } [\mathcal{O}_{Hq}^{(3)}]^{3,3} = \{\mathcal{O}_{Hq_i}^{(3)}, \mathcal{O}_{HQ}^{(3)}\} \,, \\ [\mathcal{O}_{Hu}] &\to [\mathcal{O}_{Hu}]^{j,j} \text{ and } [\mathcal{O}_{Hu}]^{3,3} = \{\mathcal{O}_{Hu_i}, \mathcal{O}_{Ht}\} \,, \end{aligned} \tag{3.4.4}$$

where Q denotes here the third-generation quark doublet. Four-fermion operators are split generically into 'four-light', 'two-heavy-two-light' and 'four-heavy' flavour components, where 'light' and 'heavy' denote the first two and the third generations, respectively. The four-light degrees of freedom are the same as the four-fermion operators of the $SU(3)^5$ scenario, except that they are flavour-universal over three generations for d, l and e, and only the first two generations for q and u. The classification of the additional four-fermion operators under this generalisation is slightly more involved, due to the permutation symmetries on the flavour indices as well as the equivalence of certain degrees of freedom via Fierz identities. This is discussed in Ref. [11], where a 'dim6top' basis is chosen for the operators involving top fields with LHC top physics observables in mind. The new operators are shown in Table 3.4.1 with their definitions in terms of the Warsaw basis coefficients.

A total of 31 new CP-conserving degrees of freedom are introduced by this relaxation of the flavour symmetry. However, our analysis is only sensitive to a subset of these, for two main reasons. First, our chosen dataset does not constrain a number of the operators allowed by this flavour assumption. These include all flavour-universal four-light fermion operators, which are constrained by electron-positron collider data and numerous low-energy scattering and decay experiments (see Ref. [162] for a recent compilation of constraints), as well as high-energy Drell-Yan and dijet observables at

hadron colliders [163, 8, 150–152, 9]. Furthermore, it is also effectively blind to the two-heavy quark two-lepton class of operators listed in Table 3.4.1. Although many of them mediate the same final states as those selected by, *e.g.*, $t\bar{t}V$ measurements, they do so in the absence of a resonant intermediate W or Z boson decaying into the lepton pair. So far, searches have implemented selections to enhance this resonant contribution, and are therefore not sensitive to the non-resonant phase space populated by the operators in question. The second important feature of our analysis is that it is restricted to the linear, $\mathcal{O}(1/\Lambda^2)$, level in the EFT expansion. This restricts the sensitivity to the set of operators that interfere appreciably with the dominant SM amplitudes for the processes of interest. This is not the case for the six neutral-current mediating, two-heavy two-light operators in the upper left section of Table 3.4.1. These operators mediate $q\bar{q} \to t\bar{t}$ production in the colour-singlet channel, which does not interfere at LO with the strongly-dominant SM QCD contribution. In contrast, the corresponding charged-current operator affects single-top quark production that, being an electroweak process in the SM, does have such an interference term. Finally, we also omit the four-heavy operators that would mainly be constrained by $t\bar{t}b\bar{b}$ measurements and four-top production searches. These data have been shown to be largely sensitive, at present, to the quadratic EFT contributions [155, 156, 63, 157], and our analysis would not yield meaningful bounds in these directions. We therefore only include in our analysis 8 two-heavy two-light quark degrees of freedom: the colour-singlet, charged-current operator and seven neutral, colour-octet operators.

To summarise, the 34 operators relevant for our leading-order, linear fit in the top-specific flavour scenario are the 20 listed in Eqs. (3.4.1) plus the 14 discussed above (three in (3.4.2), three more in (3.4.4) and eight two-light two-heavy quark operators):

$$\begin{aligned}
&\text{EWPO:} && \mathcal{O}_{HWB}\,,\ \mathcal{O}_{HD}\,,\ \mathcal{O}_{ll}\,,\ \mathcal{O}_{Hl}^{(3)}\,,\ \mathcal{O}_{Hl}^{(1)}\,,\ \mathcal{O}_{He}\,,\ \mathcal{O}_{Hq}^{(3)}\,,\ \mathcal{O}_{Hq}^{(1)}\,,\ \mathcal{O}_{Hd}\,,\ \mathcal{O}_{Hu}\,,\\
&\text{Bosonic:} && \mathcal{O}_{H\Box}\,,\ \mathcal{O}_{HG}\,,\ \mathcal{O}_{HW}\,,\ \mathcal{O}_{HB}\,,\ \mathcal{O}_{W}\,,\ \mathcal{O}_{G}\,,\\
&\text{Yukawa:} && \mathcal{O}_{\tau H}\,,\ \mathcal{O}_{\mu H}\,,\ \mathcal{O}_{bH}\,,\ \mathcal{O}_{tH}\,,\\
&\text{Top 2F:} && \mathcal{O}_{HQ}^{(3)}\,,\ \mathcal{O}_{HQ}^{(1)}\,,\ \mathcal{O}_{Ht}\,,\ \mathcal{O}_{tG}\,,\ \mathcal{O}_{tW}\,,\ \mathcal{O}_{tB}\,,\\
&\text{Top 4F:} && \mathcal{O}_{Qq}^{3,1}\,,\ \mathcal{O}_{Qq}^{3,8}\,,\ \mathcal{O}_{Qq}^{1,8}\,,\ \mathcal{O}_{Qu}^{8}\,,\ \mathcal{O}_{Qd}^{8}\,,\ \mathcal{O}_{tQ}^{8}\,,\ \mathcal{O}_{tu}^{8}\,,\ \mathcal{O}_{td}^{8}\,.
\end{aligned} \tag{3.4.5}$$

These are grouped into top operators involving two (top 2F) and four (top 4F) heavy fermions, respectively. As in the flavour-universal scenario, here we include the 4

operators affecting the tau, muon, b-quark and top-quark Yukawa couplings as they are well-constrained by our dataset. The tau, muon and b-quark Yukawa operators violate the top-specific flavour symmetry, whereas this symmetry is preserved by the top Yukawa operator.

$\mathcal{O}_i$	C_i Definition	$\mathcal{O}_i$	C_i Definition
4 quark (2 heavy 2 light)			
$\mathcal{O}_{Qq}^{1,1}$	$\sum_{i=1,2}\left([C_{qq}^{(1)}]^{ii33}+\frac{1}{6}[C_{qq}^{(1)}]^{i33i}+\frac{1}{2}[C_{qq}^{(3)}]^{i33i}\right)$	$\mathcal{O}_{Qq}^{1,8}$	$\sum_{i=1,2}\left([C_{qq}^{(1)}]^{i33i}+3[C_{qq}^{(3)}]^{i33i}\right)$
$\mathcal{O}_{Qq}^{3,1}$	$\sum_{i=1,2}\left([C_{qq}^{(3)}]^{ii33}+\frac{1}{6}[C_{qq}^{(1)}]^{i33i}-\frac{1}{6}[C_{qq}^{(3)}]^{i33i}\right)$	$\mathcal{O}_{Qq}^{3,8}$	$\sum_{i=1,2}\left([C_{qq}^{(1)}]^{i33i}-[C_{qq}^{(3)}]^{i33i}\right)$
$\mathcal{O}_{tu}^{1}$	$\sum_{i=1,2}\left([C_{uu}]^{ii33}+\frac{1}{3}[C_{uu}]^{i33i}\right)$	$\mathcal{O}_{tu}^{8}$	$\sum_{i=1,2}2[C_{uu}]^{i33i}$
$\mathcal{O}_{td}^{1}$	$\sum_{i=1,2(,3)}[C_{ud}^{(1)}]^{33ii}$	$\mathcal{O}_{td}^{8}$	$\sum_{i=1,2(,3)}[C_{ud}^{(8)}]^{33ii}$
$\mathcal{O}_{tq}^{1}$	$\sum_{i=1,2}[C_{qu}^{(1)}]^{ii33}$	$\mathcal{O}_{tq}^{8}$	$\sum_{i=1,2}[C_{qu}^{(8)}]^{ii33}$
$\mathcal{O}_{Qu}^{1}$	$\sum_{i=1,2}[C_{qu}^{(1)}]^{33ii}$	$\mathcal{O}_{Qu}^{8}$	$\sum_{i=1,2}[C_{qu}^{(8)}]^{33ii}$
$\mathcal{O}_{Qd}^{1}$	$\sum_{i=1,2(,3)}[C_{qd}^{(1)}]^{33ii}$	$\mathcal{O}_{Qd}^{8}$	$\sum_{i=1,2(,3)}[C_{qd}^{(8)}]^{33ii}$
4 quark (4 heavy)			
$\mathcal{O}_{QQ}^{1}$	$2[C_{qq}^{(1)}]^{3333}-\frac{2}{3}[C_{qq}^{(3)}]^{3333}$	$\mathcal{O}_{QQ}^{8}$	$8[C_{qq}^{(3)}]^{3333}$
$\mathcal{O}_{Qt}^{1}$	$[C_{qu}^{(1)}]^{3333}$	$\mathcal{O}_{Qt}^{8}$	$[C_{qu}^{(8)}]^{3333}$
$\mathcal{O}_{tt}$	$[C_{uu}^{(1)}]^{3333}$		
2 heavy 2 lepton			
$\mathcal{O}_{Ql}^{-(1)}$	$\sum_{i=1,2,3}[C_{lq}^{1}]^{ii33}-[C_{lq}^{3}]^{ii33}$	$\mathcal{O}_{tl}^{(1)}$	$\sum_{i=1,2,3}[C_{lu}]^{ii33}$
$\mathcal{O}_{Ql}^{3(1)}$	$\sum_{i=1,2,3}[C_{lq}^{3}]^{ii33}$	$\mathcal{O}_{te}^{(1)}$	$\sum_{i=1,2,3}[C_{eu}]^{ii33}$
$\mathcal{O}_{Qe}^{(1)}$	$\sum_{i=1,2,3}[C_{eQ}]^{ii33}$		

Table 3.4.1. Four-fermion operators containing at least one third-generation bilinear in the 'dim6top' basis [11] assuming an $SU(2)^2 \times SU(3)^3$ flavour symmetry. The relations of the corresponding Wilson coefficients with those of the Warsaw basis are also shown. The shaded entries indicate operators that are not included in our analysis because significant constraints cannot be obtained from the chosen dataset at leading order and linear level in the EFT expansion, as discussed in the text.

3.5 SMEFT Predictions

3.5.1 General strategy

The previous section introduced the operators of the dimension-6 SMEFT which will be used in the global fit. Here, we will determine our theoretical predictions for the

dependence of each observable on these SMEFT operators. We will begin by discussing the general strategy for how these predictions are computed. Following this, we will outline how these methods are specialised to Higgs observables in §3.5.2, boson decay rates in §3.5.3, diboson observables in §3.5.4 and top quark observables in §3.5.5.

Recall from §3.3 that all data are incorporated into the fitting methodology in the form of a signal strength μ_X,

$$\mu_X \equiv \frac{X}{X_{SM}} = 1 + \sum_i a_i^X \frac{C_i}{\Lambda^2} + \mathcal{O}\left(\frac{1}{\Lambda^4}\right). \tag{3.5.1}$$

where a_i^X denotes the linear contribution of a given Wilson coefficient, C_i, to a physical quantity, X, such as a production cross-section, partial or total decay width, or asymmetry. Both the SMEFT predictions X and SM predictions X_{SM} for all of the included observables were computed at leading order (LO) in perturbation theory, using the code `MadGraph5_aMC@NLO` together with the `SMEFTsim` [164] and/or the `SMEFT@NLO` [139] UFO models. In almost all cases, the LO computations correspond to tree-level computations. The exception to this is the gluon-gluon fusion process, which is a 1-loop process in the SM. This calculation will be discussed in detail in §3.5.2. Unless stated otherwise, we generate and analyse our events at the parton level and apply analysis-specific selection criteria to obtain the relevant fiducial regions of phase space.

In most cases the calculation of X in the SMEFT is limited to LO in perturbation theory by the UFO models. In principle, the SM predictions X_{SM} could be calculated to higher order, however. We choose to limit X_{SM} to LO for consistency between X and X_{SM}. Note that the value for X_{SM} will then in general differ from the values of σ_{SM} used to convert published experimental observables into signal strength format, as discussed in §3.3: while X_{SM} is calculated at LO, σ_{SM} is taken from the experimental publication and consists of the best available theoretical prediction, including QCD and/or EW corrections.

Once we have calculated X and X_{SM}, their ratio is calculated to obtain μ_X as in Eq. 3.5.1, from which we can then extract a_i^X. The a_i^X can usually be obtained with a single, high statistics Monte Carlo (MC) run for each coefficient. In some cases, non-linear contributions from Wilson coefficients can arise in MC predictions due to the W-mass shift and modifications of total widths of intermediate particles. The former is a consequence of using the electroweak $\{\alpha_{EW}, G_F, m_Z\}$ input scheme, while the latter may modify branching ratios of narrow resonances such as the Higgs, W, Z or top. When such states are produced on shell, factorising a given process into

production and decay via the narrow-width approximation (NWA) allows the decay contributions to be computed separately, then added to the prediction. If these effects cannot be factorised, as in the case of W-mass modifications or off-shell vector bosons in Higgs decays, the a_i^X are obtained by generating the predictions over a range of values of the coefficient C_i, numerically fitting the resulting μ_X values to a polynomial function in C_i, and extracting the linear dependence. This is done, in particular, for Higgs production in association with a W-boson (WH), vector boson fusion (VBF), $pp \to WW \to \ell\nu\ell\nu$ and $pp \to W^\pm Z \to \ell^+\ell^-\ell^\pm\nu$.

We use the following values in the aforementioned electroweak input scheme:

$$\begin{aligned} \alpha_{EW}^{-1} = 127.95, \quad G_F = 1.16638 \times 10^{-5}\,\text{GeV}^{-2}\,, \\ m_Z = 91.1876\,\text{GeV}, \quad m_H = 125.09\,\text{GeV}, \quad m_t = 173.2\,\text{GeV}\,. \end{aligned} \tag{3.5.2}$$

All other fermions are taken to be massless, which implies the use of five-flavour scheme PDFs, for which we use the default `NNPDF23_nlo_as_0119` sets [165] provided by `MadGraph5_aMC@NLO`. The one exception is when the lighter fermions appear as Higgs boson decay products, when we assume that they interact with the Higgs via their Yukawa couplings, $y_f \equiv 2^{\frac{3}{2}} m_f \sqrt{G_F}$, with masses taken to be

$$m_\mu = 0.106\,\text{GeV}\,, \quad m_\tau = 1.77\,\text{GeV}, \quad m_c = 0.907\,\text{GeV}, \quad \text{and} \quad m_b = 3.237\,\text{GeV}\,. \tag{3.5.3}$$

The latter two have been run up to the Higgs mass scale. In some cases, the top-specific operators independently modify b-quark initiated contributions to electroweak processes such as VBF and diboson. We do not take these SMEFT modifications into account as they are highly suppressed by the b PDFs.

We do not assign a theory uncertainty to our predictions of the SMEFT contributions, assuming that they will be subdominant with respect to other uncertainties such as the baseline SM theory predictions. We also neglect other theoretical uncertainties inherent to the SMEFT framework itself, such as omitting quadratic dimension-6 or linear dimension-8 contributions and other higher-order effects (see, e.g., Refs. [166–171] for discussions of these and related uncertainties). We note that operator mixing from RGE running and loops can also induce extra constraints [172, 173, 62, 174] and the effects of including SMEFT operators in parton distribution functions are also starting to be investigated [93, 2].

3.5.2 Higgs production

We computed LO predictions in the full parameter space of our basis for the five main Higgs production modes: gluon fusion (ggF), VBF, associated production with a W or Z (WH, ZH), and associated production with a top quark pair, $t\bar{t}H$. Some results are taken from Ref. [54], after being cross checked by independent computations. Predictions in STXS bins are compared to and found to be in agreement with the predictions presented in Refs. [58, 175]. The only one-loop calculations that we employ in our analysis involve the Higgs coupling to gluons, for which the LO contribution in the SM arises at one-loop level. Despite being a loop-induced coupling it mediates the $gg \to H$ Higgs boson production mode, which is dominant at hadron colliders. Since many Higgs measurements are very sensitive to this production mechanism and the associated, $gg \to H(+\text{jets})$ processes, we include as leading effects in the SMEFT both the tree-level contribution from C_{HG} and the leading effects from the operators that modify the top-loop contribution to the SM coupling: C_{tH}, C_{tG}, and $C_{H\Box}$. One final operator, C_G, modifies the gluon self-interaction allowing for a contribution to gluon fusion Higgs production in association with one or more jets, which we also include at one-loop order in this channel for the first time. The computations of linear contributions involve extracting the interference between loop diagrams of the SM with tree-level diagrams from C_{HG} as well as loop diagrams with a single operator insertion, and are made possible by `SMEFT@NLO`. We use a fixed renormalisation and factorisation scale of m_H in all such computations.

Figure 3.5.1 illustrates the predictions we obtain for a selection of stage 1.1 STXS gluon fusion bins [176], highlighting the potential additional discriminating power offered by the inclusion of H+jet(s). These predictions were obtained by parton-level generation of Higgs production in association with one or two additional jets. Specific predictions for, e.g., the 0-jet gluon fusion bin with Higgs $p_T > 10$ GeV or the '3-jet like' $\geq$ 2-jet bins with $p_T(Hjj) > 25$ GeV would require a matching/merging procedure interfaced with parton showering that goes beyond the level of sophistication of our analysis. Instead, we take the same dependence on the coefficients as we find for the associated parton-level bin ($p_T^H = 0$ and $p_T(Hjj) = 0$). This corresponds to assuming that the main effect on the population of the non-zero p_T bins will come from the parton shower, which does not depend on the EFT coefficients. Comparing our results with merged sample analysis of Ref. [58] (Tables 10-14), we find this assumption to be excellent for the two 0-jet bin, which have almost identical linear coefficients. Furthermore, the relevant 2-jet bins are compatible with the 'parton shower only' assumption within about 10%. We note that the comparison of individual coefficients

between the two analyses is not completely possible due to the different STXS binnings and electroweak input schemes used; additional information on the MC generation would be needed for a detailed cross-check. Nevertheless, we compare numbers for each operator where reasonable, and find that our C_{HG} contributions agree within 10–20%, while C_{tG} displays larger differences on the order of 20–60%. The other operators do not induce kinematics-dependent effects, contributing overall rescalings for which we also find good agreement with Ref. [58].

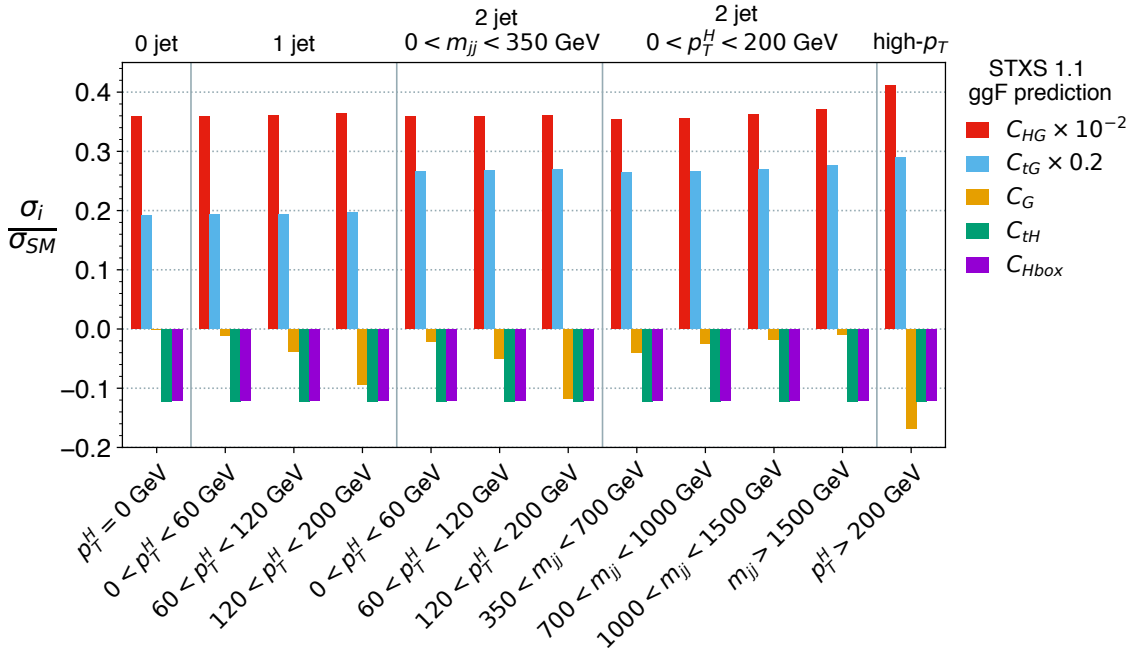


Fig. 3.5.1. Illustration of the effects of selected SMEFT operators on representative gluon fusion simplified template cross-sections σ_i relative to the corresponding SM cross-sections σ_{SM}, for $C_i/\Lambda = 1$ TeV^{-2}. The various Higgs p_T and m_{jj} bins provide complementary sensitivities and hence discriminating power between the operators.

If one were to use only Higgs signal strength data, they would be limited to the sensitivity shown in the left-most '0-jet' entry, and measurements of gluon fusion would exhibit a degeneracy in the four relevant coefficients. Instead, allowing for associated production with jets offers sensitivity to C_G and, most importantly, breaks the aforementioned degeneracy by exploiting the different energy dependences of the five operators. Besides $\mathcal{O}_{H\Box}$ and $\mathcal{O}_{tH}$, there are 3 additional operators that globally rescale gluon fusion rates by shifting the top Yukawa interaction: C_{HD}, $C_{Hl}^{(3)}$ and C_{ll}. We do not show their impact because their contributions can simply be obtained from that of $\mathcal{O}_{H\Box}$ by multiplicative factors of $-\frac{1}{4}$, -1 and $\frac{1}{2}$, respectively. Furthermore, they

are severely constrained by electroweak precision observables, to the point where they are not expected to affect gluon fusion (or $t\bar{t}H$).

Another relevant loop-induced process in Higgs production is $gg \to ZH$, which, while being formally an NNLO QCD contribution to ZH, accounts for a significant portion of the inclusive cross section [177], and has a harder p_T spectrum, which could be especially relevant for EFT interpretations [178]. Recent STXS definitions include dedicated bins for this contribution, although we are not aware of any explicit extraction of the cross sections. Notably, this process was shown to be sensitive to top quark interactions with the Z, which are relatively poorly measured elsewhere [179, 180]. However, we leave to future work the inclusion of this loop-induced sensitivity to electroweak top operators, also present in other processes such as $gg \to VV$ [181], and electroweak Higgs production and decay [182].

3.5.3 Gauge boson and Higgs decays

Many of the measurements included in our fit involve the on-shell production and decay of SM gauge bosons and/or the Higgs boson. As mentioned above, we employ the NWA to factorise production and decay, such that, at the linear level in the SMEFT, the modification to the cross section of a given process is a combination of the a_i^X for production, the partial width to the decay channel and the total width of the parent particle. The total width shifts of the W and Z bosons, have been determined as follows, where Λ has been fixed to 1 TeV:

$$\begin{aligned}\frac{\Gamma_Z}{\Gamma_Z^{SM}} = 1 &- 0.05\, C_{HWB} - 0.041\, C_{HD} + 0.082\, C_{ll} - 0.13\, C_{Hl}^{(3)} - 0.012\, C_{Hl}^{(1)} - 0.012\, C_{He} \\ &+ 0.077\, C_{Hq}^{(3)} + 0.0078\, C_{Hq}^{(1)} + 0.016\, C_{Hu} - 0.012\, C_{Hd} + 0.021\, C_{HQ}^{(3)} + 0.021\, C_{HQ}^{(1)}\,,\end{aligned} \tag{3.5.4}$$

$$\frac{\Gamma_W}{\Gamma_W^{SM}} = 1 - 0.14\, C_{HWB} - 0.065\, C_{HD} + 0.10\, C_{ll} - 0.16\, C_{Hl}^{(3)} + 0.081\, C_{Hq}^{(3)}\,. \tag{3.5.5}$$

Higgs boson decays present a richer structure, due to the importance of four-fermion decay modes that are mediated in the SM by the H couplings to the W and Z. Although the Higgs mass is too small for both gauge bosons to be on-shell, it is often assumed that at least one of the gauge bosons is, simplifying the decay to a three-body, $H \to V f' \bar{f}$, final state with the NWA applied to on-shell vector boson, $V = W, Z$. This neglects certain interference effects between, e.g., neutral-current and charged-current mediated four-fermion decays. The SMEFT introduces tree-level $H\gamma\gamma$, $H\gamma Z$ and Hgg interactions, all of which can contribute to four-fermion decay modes, and degrade the

accuracy of the NWA. In practice, however, experiments often make invariant mass cuts around m_Z in, e.g., $h \to 4\ell$ analyses, that could largely mitigate this effect.

Ref. [4] performs an in-depth calculation and analysis of Higgs decays to four fermions beyond the NWA, from which we take our predictions. We also include the contributions to $h \to gg$ mediated by operators that modify the top-quark loop contribution. These have the same relative impact as they do on $gg \to H$, discussed in the previous Section.[6]

The results of Ref. [4] are given for the flavour-universal $SU(3)^5$ scenario. This is more restrictive than our top-specific one. In the flavour-universal scenario, the operators $\mathcal{O}_{Hq}^{(1)}$ and $\mathcal{O}_{Hq}^{(3)}$ treat all three fermion generations on the same footing. As discussed in §3.4.2, in the top-specific scenario, we use $\mathcal{O}_{Hq}^{(1)}$ and $\mathcal{O}_{Hq}^{(3)}$ to represent the operators which are universal in the first two quark generations and do not involve third generation quarks. In this flavour scenario we single out operators involving the top quark. In particular, we single out the operators $\mathcal{O}_{HQ}^{(1)}$ and $\mathcal{O}_{HQ}^{(3)}$, which affect the left-handed b-quark couplings to the Z. In order to account for four-fermion decays involving b quarks, it was necessary to adapt the results of Ref. [4] such that they are expressed in the top-specific framework rather than the flavour-universal framework. We do this by using `SMEFT@NLO` to generate predictions specifically for the operators $\mathcal{O}_{Hq}^{(1)}$, $\mathcal{O}_{Hq}^{(3)}$, $\mathcal{O}_{HQ}^{(1)}$ and $\mathcal{O}_{HQ}^{(3)}$. All contributions to the total Higgs width were found to be

$$\begin{aligned}\frac{\Gamma_H}{\Gamma_H^{SM}} = 1 &+ 0.18\, C_{HWB} + 0.018\, C_{HD} + 0.045\, C_{ll} + 0.11\, C_{H\Box} + 1.7\, C_{HG} - 0.075\, C_{HW} \\ &- 0.093\, C_{HB} - 0.089\, C_{Hl}^{(3)} - 0.000059\, C_{Hl}^{(1)} - 0.000039\, C_{He} - 0.000051\, C_{Hd} \\ &+ 0.0037\, C_{Hq}^{(3)} - 0.00025 C_{Hq}^{(1)} + 0.000055\, C_{Hu} - 0.00014\, C_{HQ}^{(3)} + 0.00038\, C_{HQ}^{(1)} \\ &- 0.73\, C_{\tau H} - 0.0057\, C_{tH} - 4.0\, C_{bH} - 0.043\, C_{\mu H} + 0.044\, C_{tG}\,.\end{aligned} \tag{3.5.6}$$

In any case, the only experimentally accessible four-fermion decay modes are the leptonic ones, so quark current operators are only practically relevant via their effect on the total width, which is clearly very small. For comparison, we computed the Higgs decays using the NWA for $h \to VV^* \to 4f$ processes, noting that the two approaches can have significantly different predictions for some operators. However, we do not find any appreciable differences in the results of the global fit, indicating that Higgs

[6] However, for reasons of consistency, we do not include the contributions to $h \to \gamma\gamma$ and $Z\gamma$ mediated by operators that modify the top-quark loop contribution, which are formally of the same order as other NLO electroweak corrections that we do not include in general, and would be similarly relevant for, *e.g.*, Z-pole data.

decays are not the primary source of constraints for the operators whose predictions are sensitive to whether or not this approximation is used. Finally, we note that a recent EFT interpretation of Higgs production measurements in the $h \to 4\ell$ channel [183] pointed out significant acceptance differences in the kinematical selection between the SM and the EFT due to the additional off-shell photon contributions from C_{HW}, C_{HB} and C_{HWB}. We do not take any acceptance corrections into account in our analysis, but also do not expect these to have a large impact on the global fit, given the fact that our results are unaffected by whether or not we use the NWA.

3.5.4 Diboson data

We have obtained predictions for the fiducial signal strengths in bins of leading-lepton p_T for the ATLAS WW [184] analysis, the Z-boson p_T for the CMS and ATLAS WZ analyses [185, 186], and in bins of $\Delta\phi_{jj}$ for the ATLAS Zjj analysis [187], following the general strategy outlined above. For the WW and WZ analyses, we find good agreement of the total SM fiducial cross-section, whereas in the Zjj cases we have validated our analysis by comparison with the binned signal strengths for the operator coefficients reported by ATLAS and CMS. In particular, as pointed out in Ref. [187], we find that this channel is the most sensitive to the interference term that is linear in the triple gauge boson operator $\mathcal{O}_W$. For LEP we used the WW results of Ref. [188] that are provided for the total and differential cross-sections at different centre of mass energies. We note that this analysis uses a restricted set of angular distribution bins to mitigate the effects of unknown correlations in those bins.

3.5.5 Top data

Since we only make use of parton-level unfolded data, all top predictions are generated with stable top quarks. The data assume SM-like decay chains for the top quarks, and we do not take into account the small modifications to the W boson branching fractions in top decays due to current operators. For top production in association with a gauge or Higgs boson, we do take into account the modified branching fraction of the associated boson in the measured decay channel (usually leptonic for W and Z). However, this choice is not expected to have a significant impact on the results, given that top data is mainly sensitive to top-related operators, with those that modify gauge or Higgs boson decays being better constrained elsewhere. Much of the data we use overlaps with the data used in the top sector fit of Ref. [63], in which predictions are also obtained with `SMEFT@NLO`, but including parton shower effects. We make use of the

linear, LO parts of these in our work[7] and generate new, parton-level samples for these measurements, as including parton showering was not found to be very significant in determining the a_i^X. Our work also includes some top quark asymmetry measurements, namely those of the forward-backward and charge asymmetries from the Tevatron and LHC, respectively. Splitting a measured cross section, σ, into two regions, σ_F and σ_B, the asymmetry, A, is defined as $A \equiv (\sigma_F - \sigma_B)/(\sigma_F + \sigma_B)$. The linearised contribution of a given operator is proportional to the difference between its relative contributions to the cross sections in the two regions, $a_i^{\sigma_F}$ and $a_i^{\sigma_B}$,

$$a_A^i = \frac{1 - A_{SM}^2}{2A_{SM}}(a_i^{\sigma_F} - a_i^{\sigma_B}) \,, \tag{3.5.7}$$

highlighting how asymmetry measurements may be useful in breaking parameter space degeneracies in total cross section measurements that would, instead, be sensitive to the sum, $a_i^{\sigma_F} + a_i^{\sigma_B}$. We use the latest NNLO QCD + NLO EW theory predictions for the top asymmetries from Ref. [189]. The impact of the asymmetries is quantified in §3.6.2.

3.6 Global results

3.6.1 Higgs, Diboson and Electroweak fit

The main emphasis of this Section is on the improvements in the Higgs data since the Run 1 and early Run 2 data that were analysed in the SMEFT framework in [54]. However, as has been emphasised previously, e.g., in [190, 140], there is considerable overlap between the sets of operators whose coefficients are constrained by both electroweak and Higgs data,[8] as visualised in Fig. 2.3.1. Therefore we present in this Section results from a joint fit to the combined Higgs, diboson and electroweak data, including the 20 operators of relevance listed in Eq. 3.4.1. We recall that the latter are dominated by data from LEP, with the most important LHC contribution coming from an ATLAS measurement of M_W.

As already mentioned, this joint fit is carried out to linear order in the dimension-6 SMEFT operator coefficients, neglecting quadratic dimension-6 contributions to the LHC measurements and linear dimension-8 contributions, which, as discussed, are *a priori* of similar order in the scale Λ of high-mass BSM physics. Details of the

[7] We thank the authors of Ref. [63] for sharing the predictions with us.

[8] We show later in Fig. 3.6.9 the effect on the marginalised fit of removing the LEP EWPO and WW datasets.

fit procedure are described for the analytic method of computing a χ^2 in §3.2 and a numerical nested sampling method in Appendix B. All results in this Section are produced using the analytic method, but have verified in representative cases that fit results do not depend significantly on the method used. We emphasise again the importance of using, as well as total Higgs production and decay rates, kinematic measurements of Higgs production as encapsulated in STXS measurements, due to the different p_T dependences of dimension-6 contributions to production amplitudes, whose relative importances are generally enhanced at higher p_T [54], as seen in Fig. 3.5.1.

The top panel in Fig. 3.6.1 shows the 95% CL intervals for the coefficients $C_i(1\text{ TeV})^2/\Lambda^2$ of the 20 dimension-6 SMEFT operators contributing to the joint Higgs, diboson and precision electroweak data analysis when each operator is analysed individually. We note that certain coefficients have been scaled as indicated by the labels on the x-axis. In particular, the Yukawa operator C_{tH}, the triple-gauge boson operators C_W and C_G, and the Higgs-only operator $C_{H\Box}$ have all been rescaled by 2 orders of magnitude to appear on the same scale. Dotted grey vertical lines separate the sets of operators that contribute mostly to electroweak precision observables, that contribute mostly to Higgs and diboson measurements (Bosonic), and that modify the Yukawa couplings (Yukawa). These categories are indicated as guides to aid the reader; however, as discussed in more detail later, there are correlations between these sectors.

The two right-most bars (grey and blue) for each operator are from fits using all the available Higgs and diboson data from Run 2 of the LHC together with the precision electroweak constraints, differing only in whether they include the STXS constraints. The grey bars demonstrate the effect of replacing the combined ATLAS STXS dataset with the latest combined ATLAS signal strengths, in which case there are slightly weaker constraints for some of the operator coefficients. These results do not differ significantly whether predictions in the on-shell approximation are used, where the $H \to 4f$ process is assumed to originate from an underlying $H \to VV^* \to 4f$ decay with one gauge boson taken to be on-shell, or whether the SMEFT dependence of the full $H \to 4f$ matrix elements is used (see [4] and §3.2). For comparison, the left-most bar (green) is from a fit in which only the Higgs and diboson data analysed in [54] are used. In the case of C_W, we also show (in purple) the effect of dropping the ATLAS Zjj measurement [187], which does not impact significantly the constraints on the other operator coefficients.

In general, the individual fit ranges for the first 10 operator coefficients of Fig. 3.6.1 counting from the left, i.e., up to and including C_{Hu}, are very similar in fits including all the LHC Run 2 data to those found using the earlier set of Higgs, diboson and

electroweak data. This is because the precision electroweak data provide the strongest constraints on these individual operator coefficients. The impacts of the Higgs and diboson data are more apparent for the rest of the operator coefficients, i.e., from $C_{H\Box}$ rightwards, particularly for the Yukawa operators that have benefited from improved sensitivity of the Higgs couplings to the tau and bottom. The relative constraining power of datasets, as measured by the Fisher Information, is given in Table C.0.2 in Appendix C, and confirms the points discussed above. It also quantifies the importance of the Zjj data (84%) in pinning down C_W, compared to $W^{\pm}Z$ (13%) and LEP 2 W^+W^- (3%)

Using the same colours, the bars in the second panel show the 95% CL lower limits on the Λ_i on logarithmic scales in units of TeV, for different values of C_i. These reaches are estimated by taking half the width of the 95% CL ranges of C_i/Λ_i^2 as the typical sensitivity of the measurement. Here and in subsequent analogous panels, the darker (lighter) coloured histograms are for $C_i = 0.01(1)$ and the histograms with dashed outlines are for the strong-coupling perturbativity limit $C_i = (4\pi)^2$. In general, the Λ_i scales would be modified by a factor $\sqrt{C_i}$, which would depend on whether the Wilson coefficient is induced by strongly- or weakly-coupled new physics, at tree- or loop-level.

The corresponding 95% CL constraints for the marginalised case, where we include simultaneously all operators in the analysis and then profile the likelihood over all coefficients except one, as described in §3.2, are shown in the lower pair of panels in Fig. 3.6.1. The yellow, orange and red bars from left to right are the fits to the old data, the new data without the STXS measurements, and including them, respectively. There is again no significant difference for different treatments of $h \to 4f$. The marginalised constraints are weaker overall than the individual constraints, as the fit is allowed to explore all possible variations in the space of coefficients. We also note that the STXS measurements play a key role in the marginalised constraints for some operators, e.g., C_{HG}, C_G and C_{tH}. As discussed in §3.5.2, removing them causes a degeneracy in the parameter space that prevents meaningful constraints in these directions. We show in dark brown the impact on C_W of dropping the Zjj constraint, which still does not significantly affect the other operators. The more traditional diboson measurements suffer from suppressed SM interference at high energy due to helicity selection [191]. This is particularly so for W^+W^-, while $W^{\pm}Z$ appears to retain some sensitivity. This is why the bound without Zjj changes significantly, becoming dominated by the $W^{\pm}Z$ and/or LEP 2 data, with the high mass W^+W^- distributions from the LHC yielding no significant improvement. The Zjj observable is therefore extremely useful

for constraining anomalous gauge boson self-interactions at linear level, overcoming the non-interference issue and accessing the leading contributions in the SMEFT expansion.

We see that most of the 95% CL ranges are reduced when the full Run 2 data are included, some quite substantially, the only exceptions being C_G and C_{HG}. This occurs despite the individual constraints improving in both parameter space directions. We attribute the slight worsening of the marginalised bounds to the presence of a highly boosted $H \to b\bar{b}$ measurement in the 2018 data [192] that selects Higgs $p_T > 450$ GeV, which is significantly higher than the highest, $p_T > 200$ GeV of the stage 1.0 STXS bins used in our 2020 dataset. Removing this observable degrades the 2018 bounds below our most recent results. We expect this sensitivity to be recovered once the stage 1.2 measurements, which probe a similarly high p_T region, are incorporated. We find $\chi^2/n_{\rm dof} = 0.94$ ($p = 0.72$) for our flavour-universal global fit, where $n_{\rm dof} = 204 - 20 = 184$ (subtracting the 20 SMEFT degrees of freedom from the 204 observables in the fit). This is to be compared with $\chi^2/n_{\rm dof} = 0.93$ ($p = 0.76$) for the SM, where $n_{\rm dof} = 204$. Among the 20 operators considered in the fit to the Higgs, diboson and electroweak data, the most weakly constrained operator coefficients are all constrained so that $\Lambda/\sqrt{C_i} \gtrsim 500\,(400)$ GeV in the individual (marginalised) fits, suggesting that the linear SMEFT treatment may be adequate in this sector [193].

Fig. 3.6.1. Constraints on the indicated individual and marginalised operator coefficients $C_i(1\text{ TeV})^2/\Lambda^2$ (top and third panels) and the corresponding scales Λ for the indicated values of the C_i at the 95% confidence level (second and bottom panels), from a combined linear fit to the Higgs, diboson and electroweak precision observables. In the top two panels, the bars show the 95% CL ranges from the LHC Run 1 and early Run 2 data (green), current data without using the STXS measurements (grey), and current data including STXS using either the on-shell vector boson approximation or the full $1 \to 4$ matrix elements for the 4-fermion Higgs decay modes taken from Ref. [4] (blue). In the bottom two panels, the corresponding marginalised results are indicated by yellow, orange and red bars, respectively. We also show in purple in the individual case (grey in the marginalised case) the effect of dropping the ATLAS $Z+$ jets measurement.

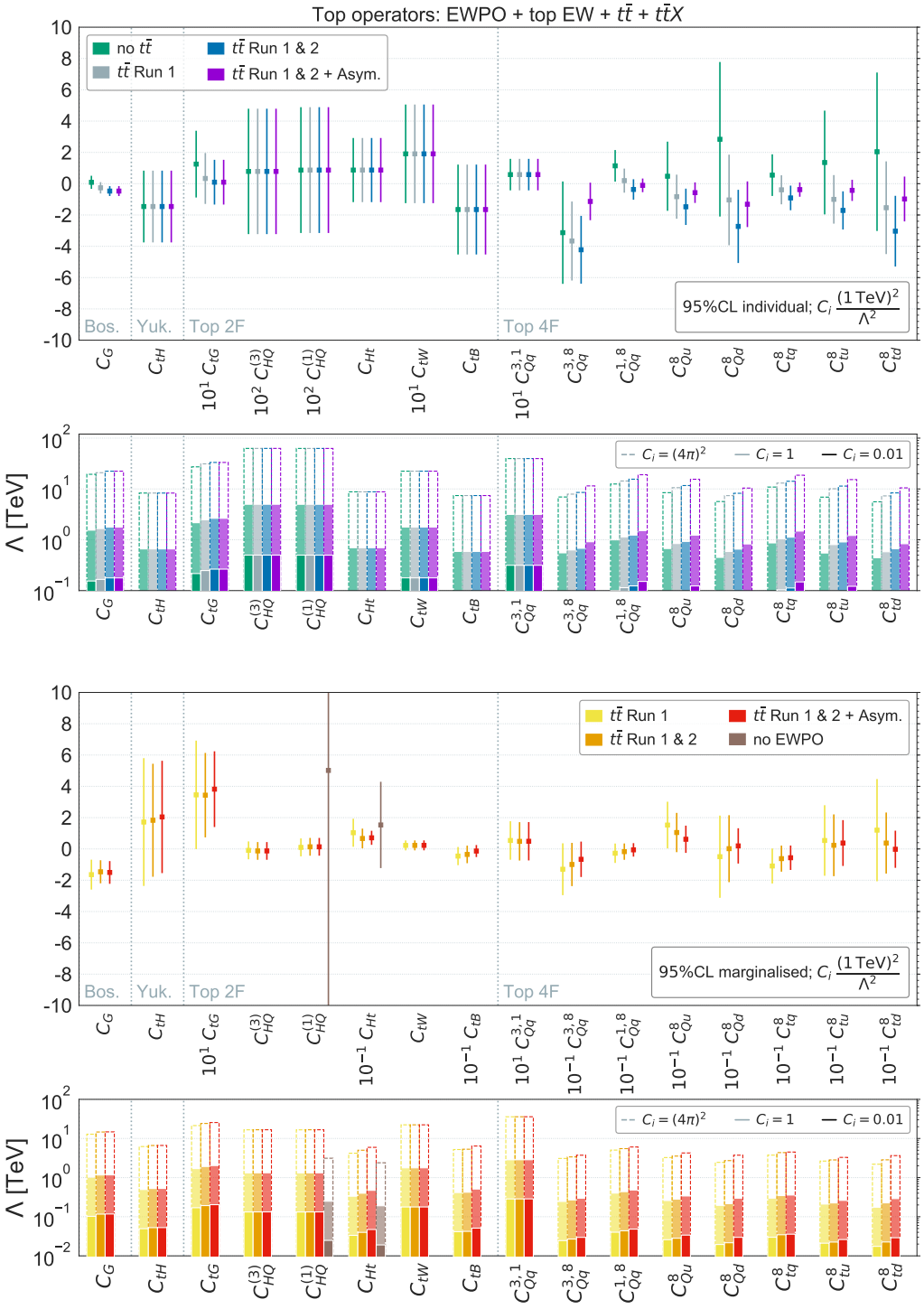


Fig. 3.6.2. Constraints on the indicated individual and marginalised operator coefficients at the 95% confidence level (upper and lower figures, respectively), from a combined linear fit to the top data and electroweak precision observables. The impact of $t\bar{t}$ data is highlighted by the evolution of the constraints starting from no $t\bar{t}$ data (green) adding Run 1 $t\bar{t}$ total and differential cross-section data (grey/yellow), the corresponding Run 2 $t\bar{t}$ data (blue/orange), and finally $t\bar{t}$ asymmetry measurements A_{FB} from the Tevatron and A_C from the LHC (purple/red).

3.6.2 Top fit

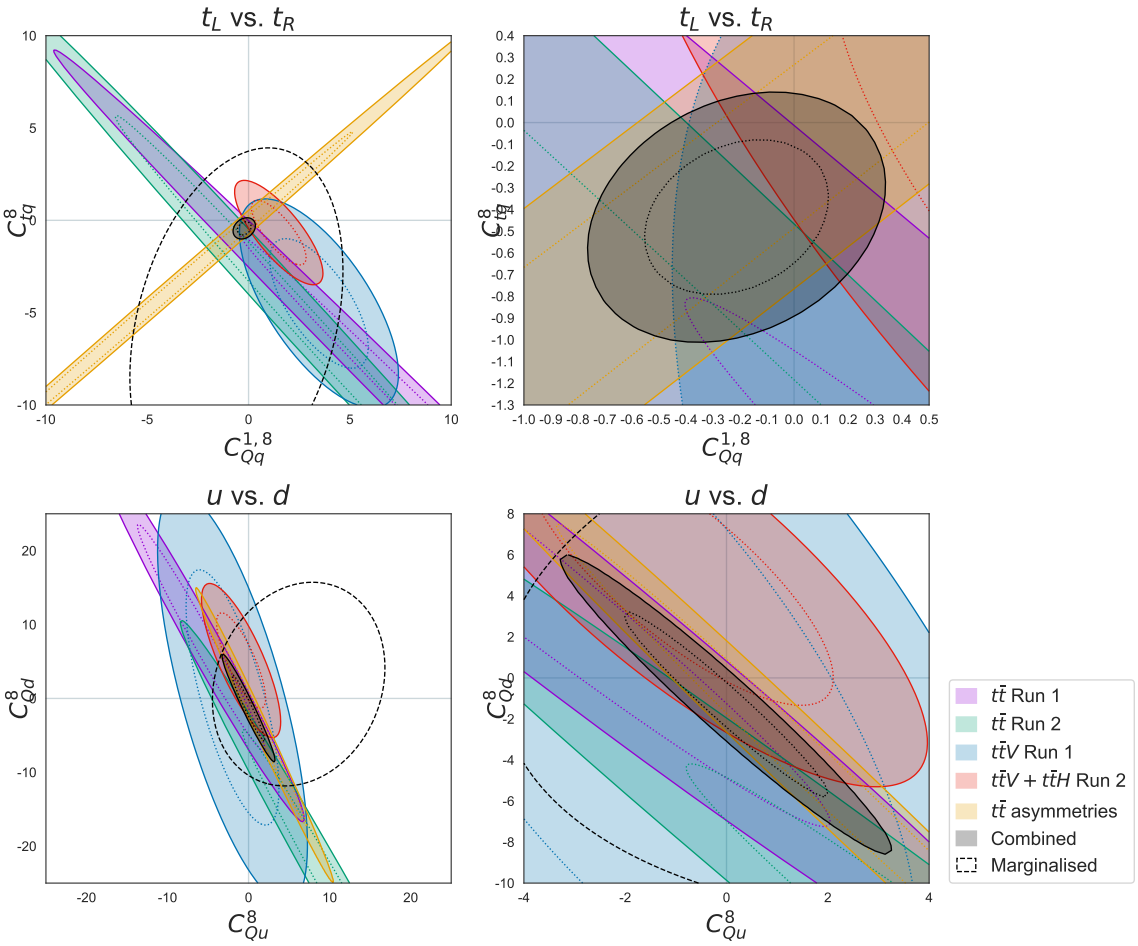


Fig. 3.6.3. Upper panels: Breakdown of the impacts of various top datasets on the 2-dimensional subspace of the four-fermion operators $(C_{Qq}^{1,8}, C_{tq}^8)$, setting all other operator coefficients to zero. Lower panels: Similar breakdown of the impacts of various top datasets on the 2-dimensional subspace of the four-fermion operators (C_{Qu}^8, C_{Qd}^8), setting all other operator coefficients to zero. Each panel on the RHS zooms in on the combined constraints of the LHS.

We consider now the information provided by top quark data, in a dedicated fit to a subset of relevant 'top quark operators', i.e., those involving at least one top quark field (the top Yukawa operator, and the top 2F and top 4F categories in Equation 3.4.5) plus the triple gluon operator, C_G, which globally affects most QCD-induced processes. We also include electroweak precision observables, which close a single blind direction $(C_{HQ}^{(1)} + C_{HQ}^{(3)})$ affecting the left-handed $Zb\bar{b}$ coupling, as well as the latest $t\bar{t}H$ signal strength measurements from the two CMS and ATLAS Higgs combination papers

included in the final fit, to constrain the top Yukawa operator. Overall, the $\chi^2/n_{\rm dof}$ for the SM is $\chi^2/n_{\rm dof} = 0.92$ ($p = 0.76$) where $n_{\rm dof} = 153$. However, there are a number of observables that exhibit some tension with the SM predictions. These include the recent 13 TeV measurement of the differential p_T distribution in t-channel single top production [194] where $\chi^2/n_{\rm dof} = 5.3$ and $p = 7.5 \times 10^{-5}$ ($n_{\rm dof} = 5$), and $t\bar{t}W$ cross section measurements by CMS at 8 TeV [195] with $\chi^2/n_{\rm dof} = 2.1$ and $p = 0.15$ ($n_{\rm dof} = 1$), ATLAS at 8 TeV [196] with $\chi^2/n_{\rm dof} = 2.6$ and $p = 0.11$ ($n_{\rm dof} = 1$) and ATLAS at 13 TeV [197] with $\chi^2/n_{\rm dof} = 1.9$ and $p = 0.17$ ($n_{\rm dof} = 1$). Furthermore, tensions are seen in the CMS $t\bar{t}$ differential distributions at 8 TeV ($\frac{d\sigma}{dm_{t\bar{t}}y_{t\bar{t}}}$ in the l+jets channel) with $\chi^2/n_{\rm dof} = 1.6$ and $p = 0.05$ ($n_{\rm dof} = 16$) and similarly in the CMS $t\bar{t}$ differential distributions at 13 TeV ($\frac{d\sigma}{dm_{t\bar{t}}}$ in the dilepton channel), with $\chi^2/n_{\rm dof} = 1.6$ and $p = 0.1$ ($n_{\rm dof} = 10$). These tensions may lead to a preference for non-zero Wilson coefficients, though this depends on whether other, more consistent observables also constrain the operators in question.

Fig. 3.6.2 shows the 95% CL intervals for the coefficients individually (top two panels) and after marginalisation (bottom two panels). The green bars in the top two panels are obtained without using any $t\bar{t}$ data, the grey bars are obtained including the Run 1 $t\bar{t}$ total and differential cross-section data, the blue bars include also the corresponding Run 2 $t\bar{t}$ data, and finally the purple bars include also the $t\bar{t}$ asymmetry measurements A_{FB} from the Tevatron and A_C from the LHC. When removing the $t\bar{t}$ data entirely, a closed fit is not possible, so marginalised constraints do not exist. The corresponding colours for the latter three sets of data (Run 1 $t\bar{t}$, including Run 2, and including A_{FB} asymmetries) in the bottom two panels are yellow, orange and red. We also show in dark brown the impact of removing electroweak precision observables for the two operators most affected, $C_{HQ}^{(1)}$ and C_{Ht}. As previously, the constraints on the scales $\Lambda/\sqrt{C_i}$ are estimated by taking half the width of the 95% CL ranges of C_i/Λ_i^2 for $C = 1$ as the typical sensitivity of the measurement. We note that many of the constraints on the quantities $C_i(1\text{TeV})^2/\Lambda^2$ in the individual and marginalised cases differ by an order of magnitude, as shown in the top and third panels of Fig. 3.6.2, and that the corresponding constraints on the scales $\Lambda/\sqrt{C_i}$ are typically a factor ~ 3 stronger in the individual analysis, as seen in the second and bottom panels.

The impact of the tension with the SM for the aforementioned observables can be seen from the individual constraints of Fig. 3.6.2. We see that C_{tH} and the 'top 2F' category appear consistent with the SM. Although it disagrees the most with the SM, the single top differential p_T data does not lead to significant deviations for the operators that can affect it, $C_{Qq}^{3,1}$, C_{HQ}^{3} and C_{tW}. This is due to the fact that single operators are

not able to improve the fit significantly, given that the disagreement predominantly comes from the lowest p_T bin, combined with the relatively good agreement of most other single-top data. Instead, the effect of the $t\bar{t}W$ data can be seen in the individual constraints without $t\bar{t}$ data, that cause the deviation of the best-fit values for $C_{Qq}^{1,8}$ and $C_{Qq}^{3,8}$ away from zero in the positive and negative directions, respectively. This is consistent with a relative minus sign in the linear dependence of the $t\bar{t}W$ process on these two operators. On the other hand, $t\bar{t}$ data depend on 9 of the operators in question, namely all of the 'top 4F' category except $C_{Qq}^{3,1}$, as well as C_{tG} and C_G. Gradually adding the differential $t\bar{t}$ data draws the coefficients towards negative values, resulting in particularly large standard deviations, especially for C_G. Finally, the $t\bar{t}$ asymmetry observables add an orthogonal constraint on the four-fermion operators that restore consistency with the SM in this sector. Since C_G does not produce an angular asymmetry in the $t\bar{t}$ matrix element, its significant, non-zero best-fit value remains. Considering now the marginalised results, we see that more $t\bar{t}$ data and the inclusion of asymmetries has a significant impact on the global sensitivity, even indirectly affecting sensitivity to electroweak top quark couplings C_{Ht} and C_{tB} by constraining the allowed four-fermion contributions to $t\bar{t}Z/\gamma$. We also show the importance of electroweak precision observables in closing the parameter space for neutral top quark couplings, by noting the large significant degradation of the limits in this space when removing them from the fit. All of the potentially large standard deviations are washed out by the marginalisation, except for C_G and C_{tG}, which lie 4.1 and 3.2 standard deviations away from zero, respectively.

The interplay between different $t\bar{t}$ measurements in constraining the four fermion sector is shown in Fig. 3.6.3, inspired by a similar analysis in Ref. [61]. Two pairs of operators are selected, setting all other coefficients to zero: $C_{Qq}^{1,8}$ and C_{tq}^{8}, which couple to left- and right-handed top quarks, respectively, in the left panel, and $C_{Qu}^{1,8}$ and C_{Qd}^{8}, which couple left handed top quarks to the up and down quarks, respectively, in the right panel. We see explicitly the complementarity between the $t\bar{t}$ cross-section measurements, which strongly constrain one linear combination of $C_{Qq}^{1,8}$ and C_{tq}^{8}, and the asymmetry measurements, which constrain an orthogonal direction. The combination of these measurements constrains each of $C_{Qq}^{1,8}$ and C_{tq}^{8} quite tightly, though less so in the marginalised case, indicated by the dashed lines. The fact that forward-backward asymmetries are sensitive to the chiral structure of the $t\bar{t}$ matrix element explains why they excel at distinguishing modified interactions for left- and right-handed tops but less so the isospin of the initial state quark (u or d). In this case the combination of $t\bar{t}$ measurements constrains a highly correlated combination of the coefficients, though

each is only weakly constrained in the marginalised fit, as seen in the bottom two panels of Fig. 3.6.2. The large differences between their individual and marginalised limits that we observe indicate strong correlations among the top quark four fermion operators constraints and that the overall marginalised sensitivity is set by the less precise $t\bar{t}X$ data.

Overall, the best-constrained coefficients in Fig. 3.6.2 are C_{tG}, $C_{HQ}^{(3)}$, $C_{HQ}^{(1)}$, $C_{Qq}^{3,1}$ and C_G, with one direction being driven by electroweak precision observables. The large negative values of C_G can be traced back to $t\bar{t}$ differential cross-section measurements, and we discuss a specific example in §3.6.3 below, where we also consider the possibility that C_G is very small, as suggested on the basis of a quadratic analysis of multijet data. Four-fermion operator coefficients are less well constrained, with scales Λ between $800 - 1500$ GeV in the individual analysis and $300 - 500$ GeV in the marginalised analysis when the corresponding $C_i = 1$, in which case the validity of the global SMEFT interpretation for these operators could be questioned for weakly-coupled UV completions, given that some of the $t\bar{t}$ data extends up to TeV energies. We therefore expect the differences between our top data analysis and those performed at quadratic level to be especially significant for the top quark four-fermion operators, as shown in Ref. [63].

The neutral top quark operator coefficients C_{Ht}, C_{tB} are also particularly hard to constrain. Production of $t\bar{t}Z/\gamma$ and tZ are the main handles we have on these couplings, and these are still not so well measured and only beginning to produce differential data. The $t\bar{t}\gamma$ differential distributions in photon p_T turn out to provide the best handle on C_{tB}. Unfortunately, C_{Ht} does not predict any effects that grow in energy in either of these processes, and C_{tB} has a suppressed interference with the SM, meaning that one does not expect spectacular gains from differential measurements, especially in a linear analysis [179, 198]. Other rare electroweak top processes, such as $t\bar{t}Wj$ and tWZ have been shown to be sensitive to such unitarity-violating behaviour and will therefore provide useful constraining power, once they are measured at the LHC [199, 200]. However, in all cases we see that the dashed histograms extend beyond a TeV even in the marginalised case, indicating that the SMEFT analysis should be a good approximation in the strong-coupling limit $C_i = (4\pi)^2$.

3.6.3 Combined Top, Higgs, Diboson and Electroweak Fit

As we discussed in §3.4, there are different possible treatments of the flavour degrees of freedom within the SMEFT framework. Specifically, in this Chapter we assume either an $SU(3)^5$ symmetry in the operator coefficients (broken by the Yukawa operators) or

allow this symmetry to be broken to an $SU(2)^2 \times SU(3)^3$ symmetry by the coefficients of third-generation fermions in a top-specific flavour scenario.

We find that, among the 20 flavour-universal operators in the $SU(3)^5$ scenario, only the marginalised constraints on C_G, C_{HG} and C_{tH} are improved significantly by the additional top measurements beyond the constraints that are already provided by the electroweak, diboson and Higgs data, while the sensitivity to $C_{Hq}^{(1)}$ and $C_{Hq}^{(3)}$ decreases, since their third-generation flavour components have been separated into separately constrained degrees of freedom. This indicates the robustness of the fit despite the increase in the number of parameters. However, there are significant correlations between the datasets. Accordingly, in the following we focus attention on the results from a simultaneous global fit to the 34 operators of the top-specific flavour scenario that was described in §3.4.2.

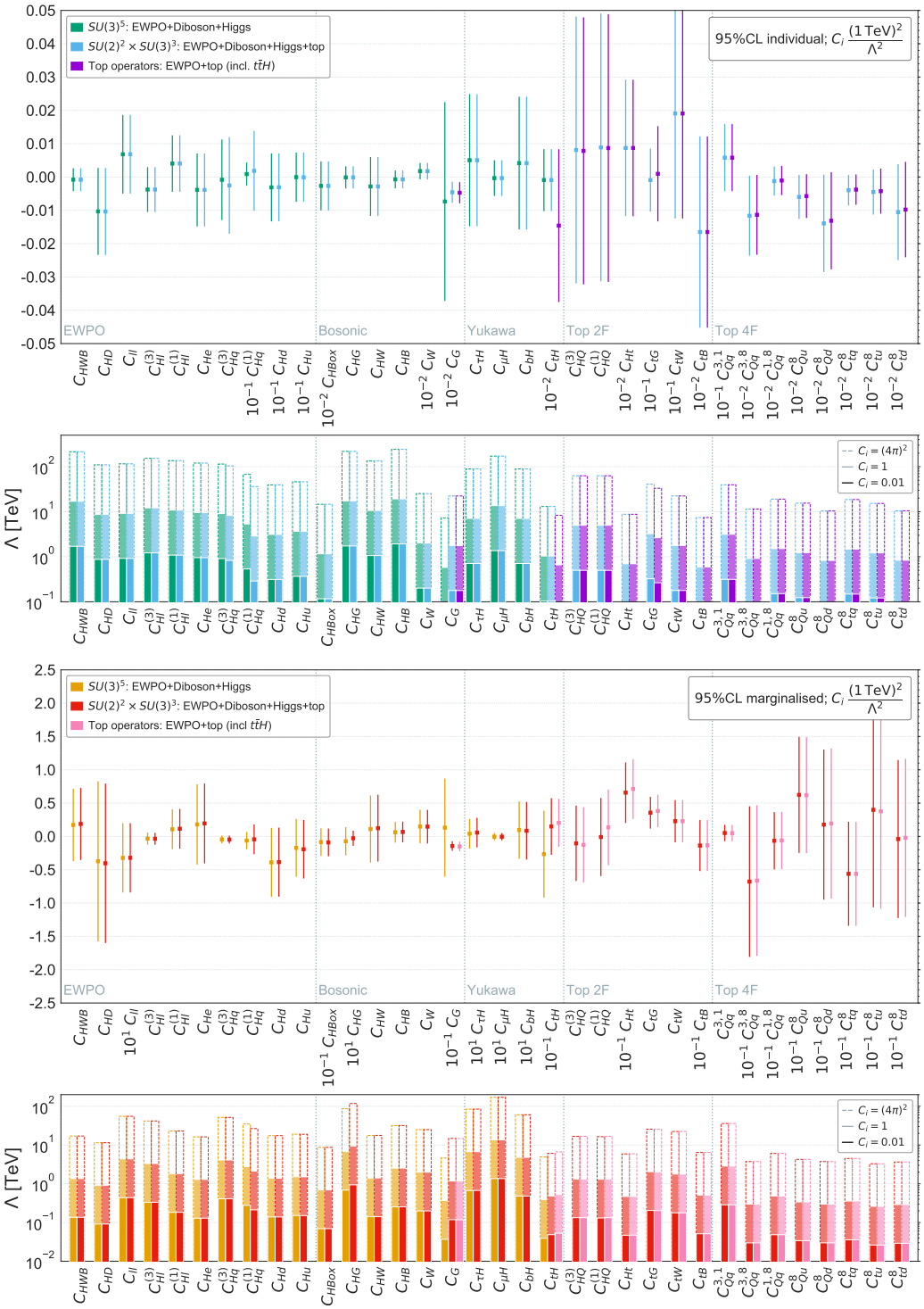


Fig. 3.6.4. Results from the global fit to the electroweak, diboson, Higgs and top data in the top-specific $SU(2)^2 \times SU(3)^3$ scenario. Top two panels: fits to the individual operators, showing the 95% CL ranges for the operator coefficients C_i normalising the new physics scale Λ to 1 TeV, and the ranges for the scales Λ for different values of the C_i. Bottom two panels: similar, but marginalising over the other operators. In the top panels, fit results found using only electroweak, diboson and Higgs data are shown in green, those found in the combination with the top data are shown in blue, and those using only top data are shown in purple. In the bottom panels, the corresponding marginalised fit results are shown in yellow, red and pink.

Operator sensitivities in individual and marginalised fits

Figs. 3.6.4 shows the results from this combined fit to all the available Higgs, electroweak, diboson and top data, switching on one operator at a time (top two panels) and marginalising over all other coefficients (bottom two panels), respectively. In each case, the upper panel shows the 95% CL ranges for the operator coefficients C_i normalising the corresponding new physics scales to 1 TeV. As indicated on the x-axis labels, certain operator coefficients have been rescaled for the sake of convenience. The bars in the lower panels show the 95% CL lower limits on the Λ_i on logarithmic scales in units of TeV. As previously, these reaches are estimated by taking half the width of the 95% CL ranges of C_i/Λ_i^2 for $C = 1$ as the typical sensitivity of the measurement.

The differences in the constraints on the 20 operators entering the flavour-independent $SU(3)^5$ fit between including top data or not are small in the individual case (purple vs blue), except for C_G. For a more detailed breakdown of the relative constraining power of different datasets on each individual coefficient, we refer the reader to Table C.0.2. Marginalising widens the ranges allowed by the fit, but the effect of marginalising over a larger set of coefficients - 34 compared to 20 in the top-specific flavour scenario compared to the flavour-universal case (red vs yellow) - introduces noticeable differences for only a few operator coefficients, namely C_G, C_{HG} and C_{tH}. The differences in the constraints on the top operators (shown in purple in the individual case and pink in the marginalised case) when the electroweak, diboson and Higgs data are included in the fit are generally small, apart from in the case of the Yukawa operator, C_{tH} and the top chromomagnetic dipole operator, C_{tG}. The loop-level constraints from Higgs production via gluon fusion are clearly very powerful in constraining these operators individually, but this sensitivity is diluted by marginalisation, which allows the other operators affecting this process to float.

Overall, the data are sufficient for a closed fit with no flat directions, and we find $\chi^2/\text{dof} = 0.81$ ($p = 0.99$, $n_{\text{dof}} = 321$) for our top-specific global fit. The sensitivities to the scale of new physics in the operator coefficients in the individual case are generally several hundred GeV or more for $C_i = 1$. The scale sensitivities in the marginalised case still reach a TeV for $C_i = 1$ for most of the electroweak precision observables set and some of the operators in the bosonic and top categories, falling to ~ 300 GeV for some of the other top operators. We emphasise, however, that specific UV completions each generate only a subset of operators, for which they have correspondingly improved reaches. The individual and marginalised fits can therefore be taken as optimistic and pessimistic sensitivity estimates, respectively, with realistic cases living somewhere between the two.

Sensitivities in 'Higgs-only' operator planes

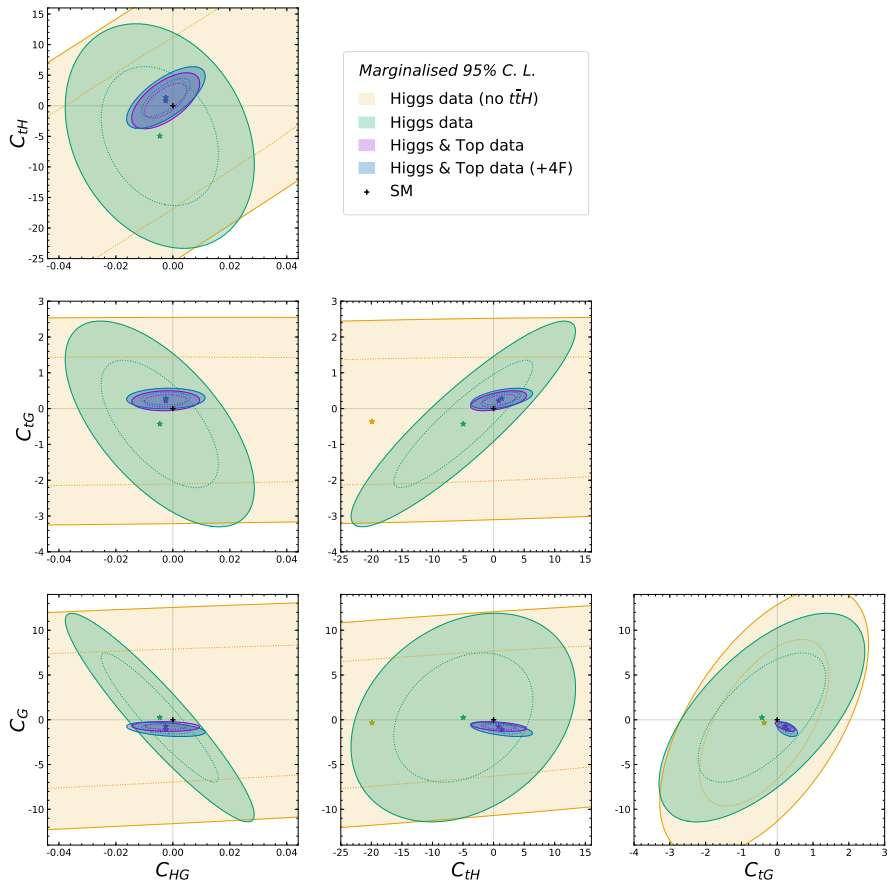


Fig. 3.6.5. Constraints on the indicated pairs of operator coefficients at the 95% confidence level, marginalised over the remaining degrees of freedom in the 'Higgs only' operator set. The green and mauve shaded areas correspond to combined linear fits to Higgs data and Higgs + top data, respectively. The blue ellipses indicate the marginalised constraints from Higgs + top data after introducing top-quark four-fermion operators into the fit, and the yellow ellipses are obtained from a fit dropping the $t\bar{t}H$ data.

In order to assess the potential impact of the interplay between top and Higgs data, we may consider the following subset of 'Higgs-only' operators:

$$\{C_{H\Box}, C_{HG}, C_{HW}, C_{HB}, C_{tH}, C_{bH}, C_{\tau H}, C_{\mu H}\} \tag{3.6.1}$$

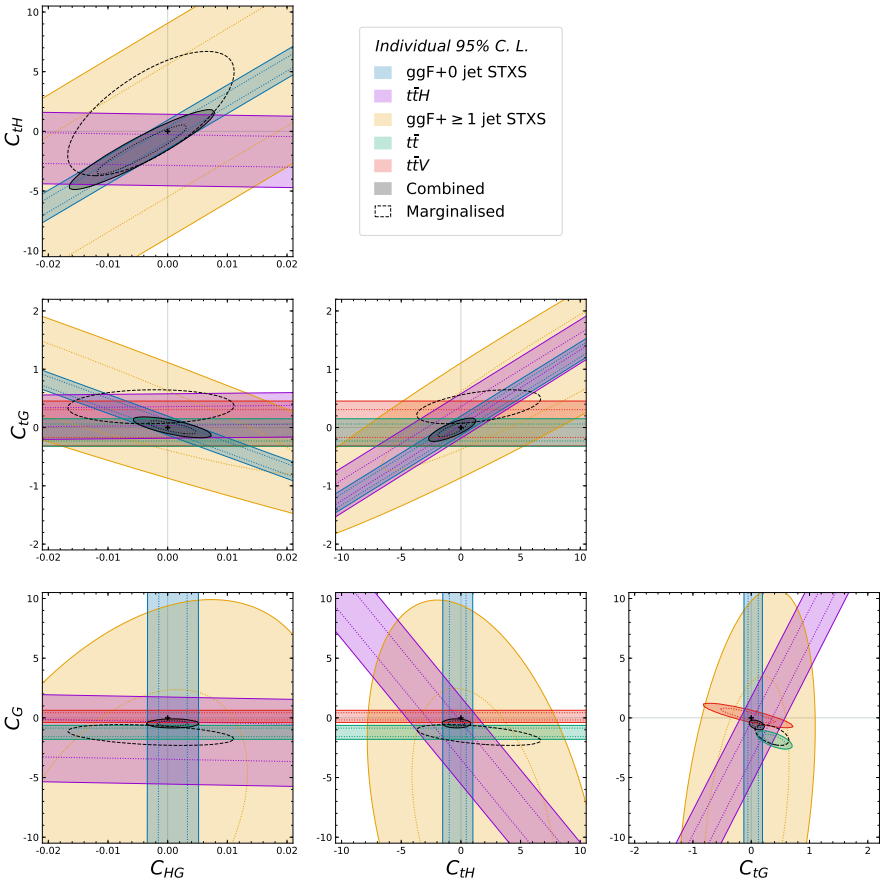


Fig. 3.6.6. Constraints on the indicated pairs of operator coefficients at the 95% confidence level, setting the other operator coefficients to zero. The shaded regions correspond to linear fits to Higgs signal strengths and 0 jet STXS bins (blue), $t\bar{t}H$ signal strengths (mauve), ≥ 1 jet STXS bins (orange) $t\bar{t}$ data (green), $t\bar{t}V$ data (red) and their combination (grey). The dashed ellipses show the constraints obtained by marginalising over the remaining Wilson coefficients of the full fit.

together with C_G and C_{tG}, which do not modify Higgs interactions directly but can impact gluon fusion. Performing a fit to this subset, Fig. 3.6.5 displays the result for the 95% CL constraints when top data are combined with Higgs data in planes showing different pairs of the operator coefficients C_{HG}, C_{tG}, C_{tH} and C_G, marginalised over the other coefficients in (3.6.1). This is the relevant set of operators in which the interplay between Higgs and top physics is most evident, taking place in the gluon fusion and $t\bar{t}$ associated Higgs production processes. It is well known that there is a degeneracy

in gluon fusion between C_{HG} and C_{tH} that prevents it from being used as a robust indirect constraint on the top Yukawa coupling, or conversely, heavy coloured particles that couple to the Higgs. This strongly motivates the direct measurement of the top Yukawa, via $t\bar{t}H$. In Fig. 3.6.5, the yellow ellipses show results for Higgs data, without $t\bar{t}H$, while the green ellipses show the sensitivity for Higgs data including it, indicating how a relative flat direction in this plane (top left panel) is lifted by the inclusion of $t\bar{t}H$. However, despite not being directly related to Higgs couplings, both C_{tG} and C_G can also affect Higgs(+jet) production in gluon fusion. Thankfully, these can be constrained by top data, particularly $t\bar{t}$ (and multi-jet data at order Λ^{-4}, for the latter). The results for the combination of Higgs with top data are shown as mauve ellipses, where we see in each plane a very substantial reduction of the area allowed at the 95% CL. The difference between the two sensitivities underlines the strength and importance of this data in indirectly pinning down BSM interactions of the Higgs, where now the C_{tG} and C_G directions are squeezed down by an order of magnitude. We also see that several (anti)correlations between pairs of operator coefficients are suppressed when top data are included, most noticeably in the (C_{HG}, C_G) plane. However, using top data to constrain only two operators is not in keeping with the global spirit of SMEFT interpretations, especially given the large number of degrees of freedom discussed in Sec. 3.4.2 that could potentially dilute its power to bound C_{tG} and C_G. We address this question by increasing our operator subset to include the 7 four-fermion operators that impact $t\bar{t}$ production, with the new marginalised constraints shown as blue ellipses. Surprisingly, there is little further change when adding the four-fermion operators, indicating a very limited dilution effect and underlining the robustness of the complementarity of top data in indirectly constraining Higgs couplings. This is especially encouraging given the fact that, as discussed in §3.6.2, our constraints on this set of operators are significantly weakened by the linear approximation used in our analysis, allowing for larger marginalisation effects than a quadratic-level fit would.

Fig. 3.6.6 displays the constraints on the same pairs of operator coefficients at the 95% confidence level when the coefficients of other operators are set to zero, with more fine-grained information on the constraints provided by the different datasets. The shaded regions are the results of linear fits to Higgs signal strengths and 0 jet STXS bins (blue), $t\bar{t}H$ signal strengths (mauve), ≥ 1 jet STXS bins (orange), $t\bar{t}$ data (green), $t\bar{t}V$ data (red) and their combination (grey). The dashed ellipses show the constraints in the corresponding parameter planes when marginalising over the remaining Wilson coefficients of the full fit, as shown in Fig. 3.6.5. As was to be expected, the constraints obtained when the other operator coefficients are set to zero are significantly stronger.

The complementarity between ggF, $t\bar{t}H$ and $t\bar{t}$ is again evident, with $t\bar{t}V$ data also providing some additional information on C_G. We also see that ≥ 1 jet STXS bins have not yet reached the level of precision needed to offer significant complementary information in this parameter space. However, we expect this to improve as increasingly fine-grained STXS binnings are measured.

The triple-gluon operator C_G

The operator $\mathcal{O}_G$ consists of triple-gluon field strengths [201–203] and so affects any observables sensitive to jets [204, 142, 143, 141]. This includes many of the Higgs and top processes in our global fit, as shown in Fig. 3.6.10 that we discuss in the next Section, where we see sizeable correlations of C_G with the operators $C_{HG}, C_{tH}, C_{tG}, C_{Qq}^{3,8}, C_{Qq}^{1,8}, C_{Qu}^{8}, C_{tq}^{8}, C_{tu}^{8}$ and C_{td}^{8}, spanning both the Higgs and top sectors. Therefore, the gluonic operator complicates the SMEFT interpretation of the measurements in these sectors.

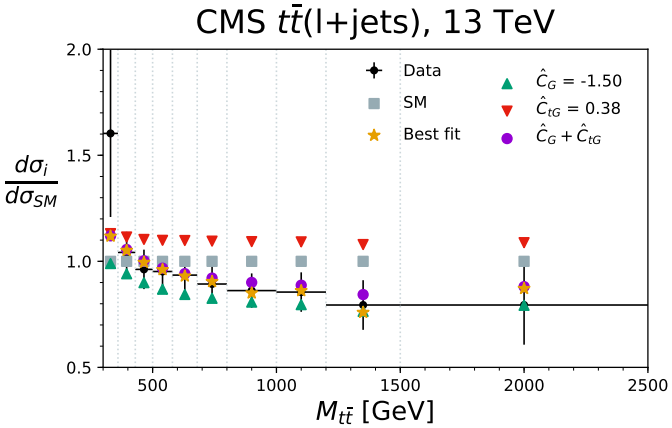


Fig. 3.6.7. The measurement of the $t\bar{t}$ invariant mass distribution in the lepton+jets channel at 13 TeV by the CMS experiment [5] compared to to the SM prediction at the NNLO QCD + NLO EW level [6]. Also shown are predictions corresponding to the best-fit values for $\hat{C}_G$ (green upward triangles), $\hat{C}_{tG}$ (red downward triangles), their combination (purple circles) and the global best-fit point in the full parameter space (orange stars).

It has been argued in Refs. [141–143] that a very strong constraint on C_G is provided by multijet data that are not included in our default data set, and that one can set $C_G = 0$ when analysing electroweak, Higgs or top data. However, this strong constraint relies on quadratic contributions to multijet observables, whereas our global fit is made to linear order. The linear contributions of C_G are small since the amplitude involving C_G in gluon-gluon scattering does not interfere with the SM, and in quark-gluon

scattering it does so only proportionally to the quark masses. This has made it a challenge to constrain in past studies [201–203]. However, the top sector, with its large quark mass, provides an opportunity to recover sensitivity at linear order to C_G, as studied most recently in Ref. [205]. Table C.0.2 indicates that $t\bar{t}$ (43%) and $t\bar{t}V$ data (56%) provide the entirety of the individual sensitivity to C_G. This is confirmed by comparing Figs. 3.6.1, 3.6.2 and 3.6.4, which also emphasise that these bounds are robust when marginalising over the other operators in both the Higgs and top sectors.

The C_G fit also shows the strongest pull away from zero, with a significance ~ 3 and $4\,\sigma$ for the best-fit in the individual and marginalised cases, respectively, as shown in Fig. 3.6.4. This effect is due to $t\bar{t}$ differential data, an example of which is given by the 13 TeV invariant mass distribution data shown in Fig. 3.6.7. We see there that the $m_{t\bar{t}}$ dependence of the cross-section data, normalised to the SM prediction, denoted by black points, differs quite significantly from that of the SM, represented by grey boxes at a value of 1, and it is this discrepancy that pulls C_G into negative territory, as shown by the green triangles, plotting the best-fit C_G contribution. We see that the agreement with the data is obtained by an interplay between C_G and C_{tG}, whose best-fit prediction, shown by the red triangles, improves the agreement in the low mass bins while not too significantly affecting the high mass region. The sum of the best-fit C_G and C_{tG} predictions, shown by the purple circles, coincides with the global best-fit prediction, shown by the orange stars, demonstrating that the fit to the data is obtained primarily by this interplay. We emphasise, however, that the significance of this effect could be reduced if there were some important contribution to the $m_{t\bar{t}}$ distribution close to threshold that has not been included in the SM calculation. Moreover, such a large pull away from the SM in this case is not meaningful, in view of the potentially important quadratic contributions from C_G. What the linear fit demonstrates is the size of the linear constraint on C_G, which is not known *a priori*, and its dependence on other operators. Significant indirect effects of C_G on the other operators may also then be questioned.

Accordingly, we have investigated the consequences of assuming that C_G can be better constrained by including dedicated QCD multi-jet data, and have analysed the effects of setting $C_G = 0$ in our study of the impact of top data on Higgs coupling measurements, as well as in our marginalised global fits in the flavour-universal $SU(3)^5$ and the top-specific $SU(2)^2(3)^3$ scenarios. Fig. 3.6.8 shows that the global space of constraints is changed compared to Fig. 3.6.5, especially when only including Higgs data (green ellipse), with improved sensitivity (see Fig 3.6.9) and a different pattern of correlations. Once the top data is included, the overall sensitivity is improved, and

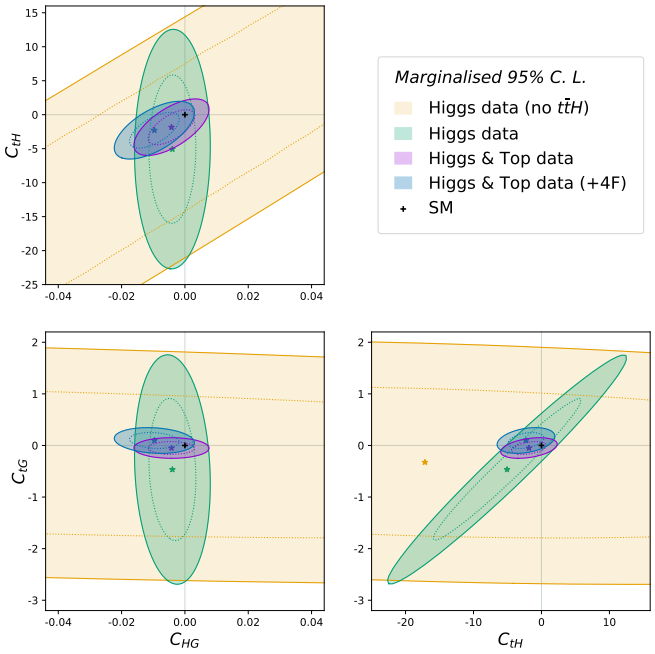


Fig. 3.6.8. Same as Fig 3.6.5, but setting $C_G = 0$.

the best-fit point before including the four-fermion operators (mauve ellipse) is closer to the SM. The relative impact of adding the four-fermion operators (blue ellipse) is more noticeable when $C_G = 0$ than when it is non-zero. When $C_G = 0$, the tension with the SM in the $t\bar{t}$ data pulls the four-fermion operators and C_{tG}. These operators also affect $t\bar{t}H$, and C_{tG} also modifies gluon fusion, leading to a cascade of shifts in the ranges of these operators. Moving to the global results, we see in the top two panels of Fig. 3.6.9 that this constraint has little effect on the 95% ranges we find in our marginalised $SU(3)^5$ fit, except that the ranges of C_{HG} and C_{tH} are reduced noticeably when we set $C_G = 0$,[9] as expected from the previous discussion of Fig. 3.6.8. In the $SU(2)^2 \times SU(3)^3$ (bottom two panels of Fig. 3.6.9) there are shifts in the central values of several top operator coefficients, with the four fermion operators moving further away from the SM to absorb the aforementioned discrepancies in top data. Overall, no significant reductions in the ranges of any operator coefficients are observed, and thus our fit results are relatively insensitive to the treatment of C_G.

[9]For completeness, we also show in the top two panels of Fig. 3.6.9 the effects of dropping the electroweak precision observables (and LEP WW) from the marginalised $SU(3)^5$ fit (yellow bars). As could be expected, there are large effects on the constraints on the operators that contribute most to the electroweak precision observables, but quite small effects in the bosonic and Yukawa sectors.

Fig. 3.6.9. Comparison of the constraints on the indicated marginalised operator coefficients $C_i(1\text{ TeV})^2/\Lambda^2$ (top and third panels) and the corresponding scales Λ for the indicated values of the C_i at the 95% confidence level (second and bottom panels), found in a combined linear fit to the Higgs, diboson and electroweak precision observables (top two panels) and including in addition top data (bottom two panels), including C_G in the fit (orange) and setting $C_G = 0$ (red). Also displayed in yellow in the top two panels is a fit without LEP (EWPO and WW) measurements.

Correlation matrix and principal component analysis

The full correlation matrix for the top-specific marginalised fit is shown in Fig. 3.6.10, colour-coded and labelled in percentages. Some of these correlations can be explained intuitively by the simple fact that two operators contribute to the same observable, while others occur more indirectly, through a chain of dependencies that is difficult to trace through the inversion of the Fisher information matrix. We see that there are substantial (anti)correlations between the coefficients of operators in the EWPO and Bosonic sectors, and there are also many (anti)correlations within the top 4F sector. On the other hand, there is only one large off-diagonal entry in the top 2F sector, namely a negative correlation between $C_{HQ}^{(3)}$ and $C_{HQ}^{(1)}$. Along with C_{Ht}, that is mildly correlated with these two, these operators affect $Zt\bar{t}$ couplings. As previously discussed, one linear combination, $C_{HQ}^{(3)} + C_{HQ}^{(1)}$ modifies the $Zb\bar{b}$ coupling, and hence has a very strong LEP constraint, while C_{Ht} and the other combination of $C_{HQ}^{(3)}$ and $C_{HQ}^{(1)}$ can only be probed in electroweak top processes. The bottom Yukawa operator (C_{bH}) exhibits some correlations with those of the top (C_{tH}) and the tau ($C_{\tau H}$), and in the bosonic sector some moderate correlations are observed, notably between C_{HW}, C_{HB} and $C_{H\Box}$, and between C_{HG} and C_G (as expected from the discussion in Sec. 3.6.3). Turning to correlations between operators in different sectors, we note substantial (negative) correlations between C_{bH} in the Yukawa sector and $C_{HWB}, (C_{HD}), C_{Hl}^{(3)}, C_{Hq}^{(3)}$ and C_{He} in the EWPO sector, and $(C_{H\Box}), C_{HW}$ and C_{HB} in the bosonic sector respectively, as well as substantial positive (negative) correlations between C_{tH} in the Yukawa sector and $C_{H\Box}, C_G$ and (C_{HG}) in the Bosonic sector. Finally, there are several large (anti)correlations between operators in the top 4F and top 2F sectors, namely $C_{Qq}^{3,1}$ and $C_{HQ}^{(3)}$ (positive), $C_{Qq}^{3,1}$ and $C_{HQ}^{(1)}$ (negative), C_{Qu}^{8} and C_{Ht} (negative), and C_{tq}^{8} and C_{Ht} (positive). Overall, there are 22 correlation coefficients with magnitude ≥ 0.2 between operators in a top sector on the one hand and in a Yukawa, bosonic or electroweak sector on the other hand. These and the top sector may not be talking to each other very loudly, but they are starting to whisper to each other.

In the Gaussian approximation to the global likelihood that is used here, it is also informative to diagonalise the constraints on the operator coefficients in the orthonormal eigenvector basis and perform a principal component analysis. This tells us which directions are most constrained in the fit, and what operators contribute to those directions. We display the constraints on the eigenvectors graphically in Fig. 3.6.11: the rows in the centre panel correspond to the different operator coefficients, the columns correspond to the different eigenvectors, and the colour-coded squares represent the moduli-squared of the operator components in the eigenvectors. The

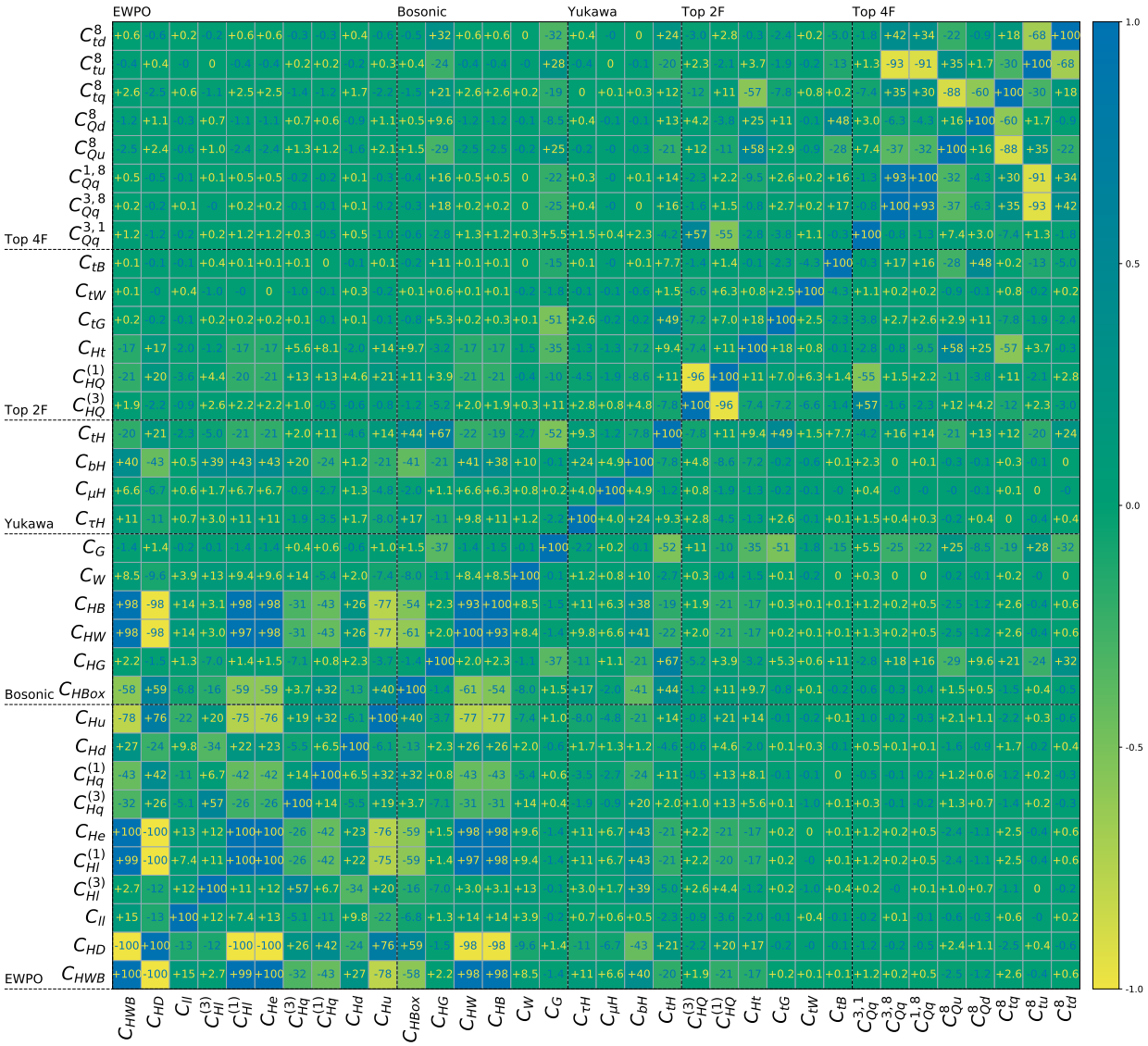


Fig. 3.6.10. The full 34×34 correlation matrix for the marginalised top-specific fit. The operators are grouped into those affecting primarily electroweak precision observables, bosonic observables, Yukawa measurements and top electroweak measurements, as well as top-quark four-fermion operators. The entries in the correlation matrix are colour-coded according to the indicated magnitudes of the correlation coefficients.

latter are ordered such that the strengths of the global constraints decrease from left to right, as seen in the top panel of Fig. 3.6.11, where the 95% CL bounds on the scales Λ are calculated assuming that the linear combination of operator coefficients making up that particular eigenvector is set to unity. The bottom panel tabulates the respective relative constraining powers of the electroweak precision data, LEP diboson data, Higgs coupling strength measurements from Runs 1 and 2, STXS measurements, LHC diboson and Zjj measurements, $t\bar{t}$ measurements, single top measurements and $t\bar{t}V$ measurements. These are defined as the relative contribution of each dataset to the corresponding entry of the diagonalised Fisher information. Entries where there is no significant constraining power are indicated by "-".

We see that the largest component of the best-constrained eigenvector is $\mathcal{O}_{HB}$, and that the limit on its scale exceeds 20 TeV, with the most important contribution coming from the STXS measurements, followed by Higgs signal strength measurements. The scales of four more eigenvectors are constrained beyond the 10 TeV level, with the most important contributions coming from the electroweak precision measurements as well as STXS and Higgs signal strength measurements. They can broadly be associated with the powerful sensitivity we have obtained in constraining the $H\gamma\gamma$ and Hgg interactions. The first 12 eigenvectors are constrained by a mixture of EWPO and Higgs data, showing that these two sets are providing complementary and competitive bounds in the multi-TeV range.

We next see a particularly strong constraint coming entirely from single top data on $C_{Qq}^{3,1}$ alone. We note that other operators contribute along that eigenvector direction, which is given by $-0.98C_{Qq}^{3,1} - 0.17C_{HW} + 0.08C_{HQ}^{(3)}$, but with too small a magnitude to be visibly coloured.[10] However, this is partly responsible for the large (anti-)correlations between $C_{Qq}^{3,1}$ and $C_{HQ}^{(3)}$ ($C_{HQ}^{(1)}$) shown in Fig. 3.6.10, and $C_{Qq}^{3,1}$ also appears in other directions with a small contribution. Several other examples of relatively isolated operators can be found across the figure, identified by the columns dominated by a single, very dark spot. Here the eigen-directions nearly coincide with a particular operator, such that the rest of the fit should be relatively independent of whether these are included or not. Specific examples are $C_{\mu H}$, which is constrained in isolation by the $H \to \mu\mu$ signal strength, and C_{tW}, which is constrained mainly by W-helicity fraction measurements in $t\bar{t}$ data. We also see that Zjj mostly constrains C_W with not much effect on the rest of the fit. The relation between measurements and constraints on operators can be indirect, illustrating the complementarity between the different datasets: for example, the LHC WW and WZ diboson data are responsible for 46% of

[10] For completeness, the numerical values of the eigenvector components are provided in Table C.0.3.

the constraining power along the eigenvector direction principally aligned with C_{HW}, despite the lack of C_{HW} dependence in diboson data. However, their inclusion helps to close directions of limited sensitivity in the fit to Higgs and electroweak data that are then better able to constrain *e.g.* C_{HW}.

The least strongly constrained eigenvector is predominantly $\mathcal{O}_{tu}^{8}$, with a scale bounded just above 200 GeV when the operator coefficients are normalised to unity, mainly by $t\bar{t}$ data. It is followed by three more eigenvectors with scales ~ 300 GeV, whose principal components are top operators. The most important constraints on these eigenvectors are in the top sector, principally from the $t\bar{t}$ and $t\bar{t}V$ data. As discussed in §3.6.2, while the validity of the SMEFT may be questioned when the operator coefficients are normalised to unity, it should be reliable for all top operators in the strong-coupling limit. Also, we expect the SMEFT to be valid for the better-constrained eigenvectors even for unit-normalised coefficients, since these eigenvectors have relatively small top operator components, as seen in the middle panel of Fig. 3.6.10.

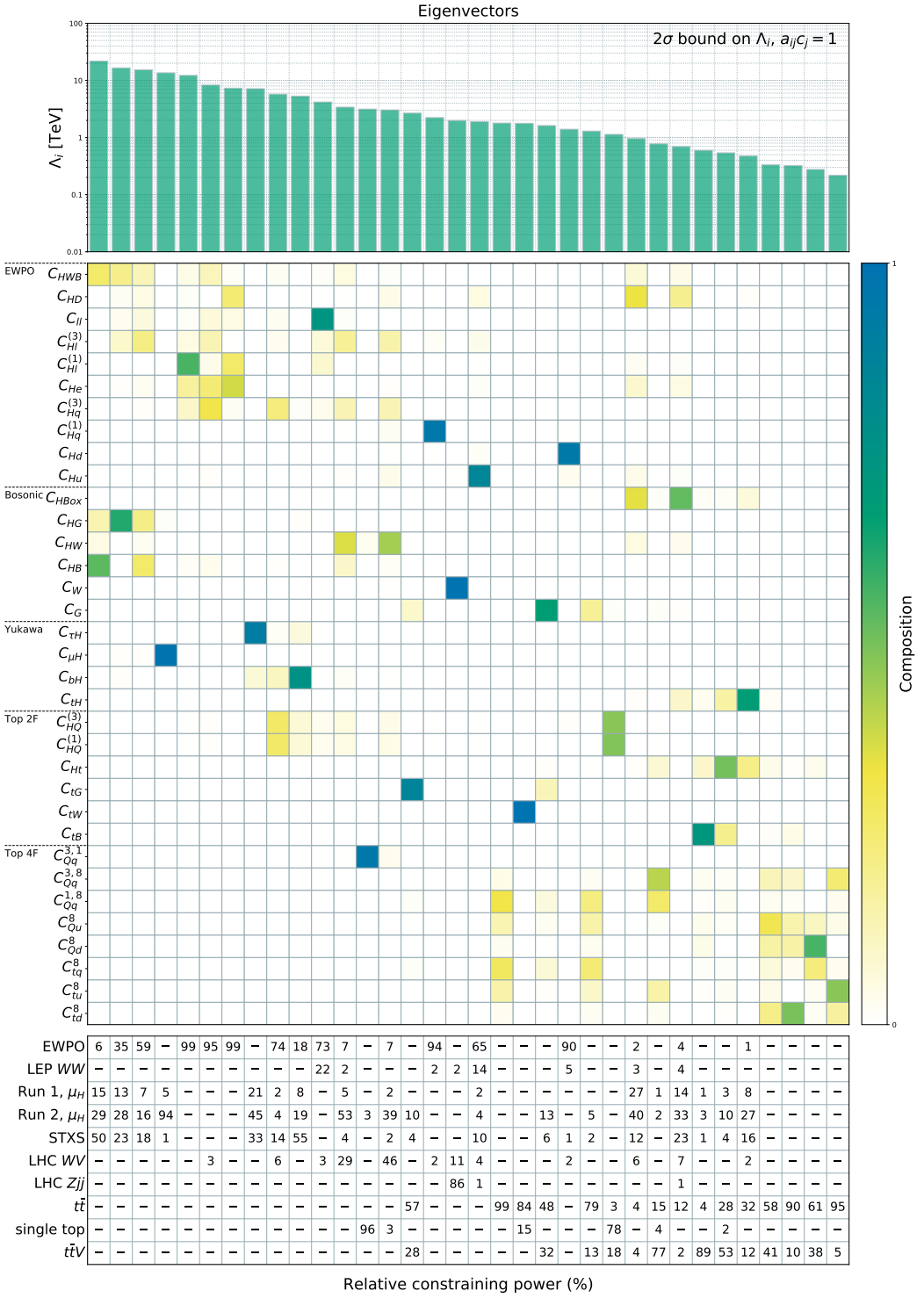


Fig. 3.6.11. Constraints on the eigenvectors of the global likelihood function. The top panel shows the strengths of the global constraints on the eigenvectors. The rows in the centre panel correspond to the different operator coefficients, the rows correspond to the different eigenvectors, and the colour-coded squares represent the moduli-squared of the operator components to the eigenvectors. The bottom panel shows the constraining powers of the electroweak precision data, LEP diboson data, Higgs coupling strength measurements from Runs 1 and 2, STXS , LHC diboson and Zjj, $t\bar{t}$, single-top and $t\bar{t}W/Z/\gamma$ measurements, respectively. Instances where there is no significant constraining power are indicated by "-". Explicit expressions for the eigenvectors are given in Table C.0.3

3.7 Constraints on UV Completions

So far our approach has been to use the SMEFT framework to combine all relevant data inputs and perform a global fit to the coefficients of the dimension-6 operators characterising possible modifications of the SM Lagrangian at leading order in the momentum expansion. The result of this fit, summarised in Fig. 3.6.4, provides a model-independent way to evaluate the compatibility of the SM with the available experimental data. This SMEFT analysis provides information on the level of new physics contributions compatible with the current data. We presented in the bottom panel of Fig. 3.6.4 limits on the dimensionful parameters $\Lambda/\sqrt{C_i}$, which can be interpreted as constraints on the possible scale of new physics compatible with current measurements. The individual limits correspond to bounds on a single operator assuming all others are zero, while the marginalised limits allow all other coefficients to vary. These can respectively be taken as optimistic and pessimistic estimates of the sensitivity, and we expect realistic models to generate some intermediate subset.

In this Section we go a step further in the interpretation of our fit and explore how specific UV completions of the SM Lagrangian are constrained by current measurements. In any given model, the global analysis we have presented is often not directly applicable, as typical models may generate more than just one of the dimension-6 SMEFT operators, but not all of them. Moreover, when a specific model contributes to more than a single operator, these contributions are often related, corresponding to a smaller subset of independent parameters.

To illustrate these model-dependent effects we have considered several model interpretations. In §3.7.1 we discuss models in which SMEFT operators are induced at the tree level, and in §3.7.2 we discuss classes of UV completions that share similar SMEFT patterns. We then analyse supersymmetric models with TeV-scale stops in which SMEFT operators are induced at the one-loop level in §3.7.3. Finally, in §3.7.4 we present results from a survey of the pulls for all fits with non-vanishing coefficients for combinations of 2, 3, 4 and 5 operators.

3.7.1 Simple tree-level-induced SMEFTs

We first study the implications of our analysis for single-field extensions of the SM that contribute to the SMEFT at tree level. This exercise updates the one presented in [54] and is based on the dictionary provided in [206].[11]

[11] We note that these one-parameter extensions of the SM have been included among the BSM benchmark proposals made by the LHC Higgs Working Group [207].

Name	Spin	SU(3)	SU(2)	U(1)	Param.	Name	Spin	SU(3)	SU(2)	U(1)	Param.
S	0	1	1	0	$(M_S,\, \kappa_S)$	Δ_1	$\frac{1}{2}$	1	2	$-\frac{1}{2}$	$(M_{\Delta_1},\lambda_{\Delta_1})$
S_1	0	1	1	1	(M_{S_1},y_{S_1})	Δ_3	$\frac{1}{2}$	1	2	$-\frac{1}{2}$	$(M_{\Delta_3},\lambda_{\Delta_3})$
φ	0	1	2	$\frac{1}{2}$	$(M_\varphi,Z_6\cos\beta)$	Σ	$\frac{1}{2}$	1	3	0	$(M_\Sigma,\lambda_\Sigma)$
Ξ	0	1	3	0	(M_Ξ,κ_Ξ)	Σ_1	$\frac{1}{2}$	1	3	-1	$(M_{\Sigma_1},\lambda_{\Sigma_1})$
Ξ_1	0	1	3	1	$(M_{\Xi_1},\kappa_{\Xi_1})$	U	$\frac{1}{2}$	3	1	$\frac{2}{3}$	(M_U,λ_U)
B	1	1	1	0	$(M_B,\hat{g}_H^B)$	D	$\frac{1}{2}$	3	1	$-\frac{1}{3}$	(M_D,λ_D)
B_1	1	1	1	1	(M_{B_1},g_{B_1})	Q_1	$\frac{1}{2}$	3	2	$\frac{1}{6}$	(M_{Q_1},λ_{Q_1})
W	1	1	3	0	$(M_W,\hat{g}_H^W)$	Q_5	$\frac{1}{2}$	3	2	$-\frac{5}{6}$	(M_{Q_5},λ_{Q_5})
W_1	1	1	3	1	$(M_{W_1},\hat{g}_{W_1}^\phi)$	Q_7	$\frac{1}{2}$	3	2	$\frac{7}{6}$	(M_{Q_7},λ_{Q_7})
N	$\frac{1}{2}$	1	1	0	(M_N,λ_N)	T_1	$\frac{1}{2}$	3	3	$-\frac{1}{3}$	(M_{T_1},λ_{T_1})
E	$\frac{1}{2}$	1	1	-1	(M_E,λ_E)	T_2	$\frac{1}{2}$	3	3	$\frac{2}{3}$	(M_{T_2},λ_{T_2})
T	$\frac{1}{2}$	3	1	$\frac{2}{3}$	(M_T,s_L^t)	TB	$\frac{1}{2}$	3	2	$\frac{1}{6}$	$(M_{TB},s_L^{t,b})$

Table 3.7.1. Single-field extensions of the SM constrained by our analysis.

We list in Table 3.7.1 the $SU(3) \times SU(2) \times U(1)$ quantum numbers and couplings of the new fields considered here. We assume flavour-universal couplings in all cases except T and TB, in which the new fields couple only to the third generation. These two models are taken from [208]. We consider only the renormalisable contributions from each single field extension. In the case of the models B and W (a Z' and W' respectively) we consider only their couplings to the Higgs doublet and set all fermion couplings to zero. In addition to evaluating the constraints on these single-field extensions, we also consider the following two combinations of the fields in Table 3.7.1 that yield single-parameter models via cancellations (see Ref. [207]), i.e., models that depend on only a single coupling (λ or g_H), as well as a mass M: *1)* Quark bidoublet model: $\{Q_1, Q_7\}$ with equal masses M and equal couplings λ to the top quark, and *2)* Vector-singlet pair model: $\{B, B_1\}$ with equal masses M and Higgs couplings proportional to g_H. We exhibit in Tables 3.7.2 and 3.7.3 the contributions made at tree level by exchanges of each of these fields to the SMEFT coefficients.[12] The numbers shown in the Tables should each be multiplied by the appropriate squared coupling factors and divided by the square of the mass scale M.[13]

[12] In general, the coloured, vector-like fermions contribute at one-loop order to C_{HG}. We include this contribution for the 2-parameter model TB, and verify in a representative example (the T field) that it has a negligible effect on the single-parameter model constraints.

[13] We do not provide limits on the two-parameter model Ξ_1, which has a complex coupling g_{Ξ_1}, but note that it behaves similarly to models Ξ and S_1.

Model	C_{HD}	C_{ll}	C^3_{Hl}	C^1_{Hl}	C_{He}	$C_{H\Box}$	$C_{\tau H}$	C_{tH}	C_{bH}
S						$-\frac{1}{2}$			
S_1		1							
Σ			$\frac{1}{16}$	$\frac{3}{16}$			$\frac{y_\tau}{4}$		
Σ_1			$-\frac{1}{16}$	$-\frac{3}{16}$			$\frac{y_\tau}{8}$		
N			$-\frac{1}{4}$	$\frac{1}{4}$					
E			$-\frac{1}{4}$	$-\frac{1}{4}$			$\frac{y_\tau}{2}$		
Δ_1					$\frac{1}{2}$		$\frac{y_\tau}{2}$		
Δ_3					$-\frac{1}{2}$		$\frac{y_\tau}{2}$		
B_1	1					$-\frac{1}{2}$	$-\frac{y_\tau}{2}$	$-\frac{y_t}{2}$	$-\frac{y_b}{2}$
Ξ	-2					$\frac{1}{2}$	y_τ	y_t	y_b
W_1	$-\frac{1}{4}$					$-\frac{1}{8}$	$-\frac{y_\tau}{8}$	$-\frac{y_t}{8}$	$-\frac{y_b}{8}$
φ							$-y_\tau$	$-y_t$	$-y_b$
$\{B, B_1\}$						$-\frac{3}{2}$	$-\ y_\tau$	$-\ y_t$	$-\ y_b$
$\{Q_1, Q_7\}$								y_t	
Model	C^3_{Hq}	C^1_{Hq}	$(C^3_{Hq})_{33}$	$(C^1_{Hq})_{33}$	C_{Hu}	C_{Hd}	C_{tH}	C_{bH}	
U	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{4}$	$\frac{1}{4}$			$\frac{y_t}{2}$		
D	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$	$-\frac{1}{4}$				$\frac{y_b}{2}$	
Q_5						$-\frac{1}{2}$		$\frac{y_b}{2}$	
Q_7					$\frac{1}{2}$		$\frac{y_t}{2}$		
T_1	$-\frac{1}{16}$	$-\frac{3}{16}$	$-\frac{1}{16}$	$-\frac{3}{16}$			$\frac{y_t}{4}$	$\frac{y_b}{8}$	
T_2	$-\frac{1}{16}$	$\frac{3}{16}$	$-\frac{1}{16}$	$\frac{3}{16}$			$\frac{y_t}{8}$	$\frac{y_b}{4}$	
T			$-\frac{1}{2}\frac{M_T^2}{v^2}$	$\frac{1}{2}\frac{M_T^2}{v^2}$			$y_t\frac{M_T^2}{v^2}$		

Table 3.7.2. Operators generated at tree level by the single-field extensions listed in the first column. Each extension depends on a single coupling (see Table 3.7.1) as well as a new physics mass-scale M. The coefficients of the operators are each proportional to the squares of the corresponding coupling λ by the corresponding entry in the Table and divided by M^2. y_t, y_b and y_τ denote the top, bottom and tau Yukawa couplings respectively, v denotes the electroweak scale and α_s denotes the strong coupling.

We show in Fig. 3.7.1 the results from our global fit for all of the one-parameter single-field extensions of the SM. In each of these models we constrain a positive quantity: $|\lambda|^2$. The constraints in Fig. 3.7.1 are found using the numerical fitter described in Appendix B. This method allows us to incorporate the constraint $|\lambda|^2 > 0$ as a Heaviside prior $\pi(|\lambda|^2 < 0) = 0$. The 2-σ constraints on the mass scales in

Model	C_{HD}	$C_{H\Box}$	$C_{\tau H}$	C_{tH}	C_{bH}
B	$-2a^2$	$-\frac{1}{2}(a^2-b^2)$	$-aby_\tau$	$-aby_t$	$-aby_b$
W	$\frac{1}{2}b^2$	$-\frac{1}{8}(3a^2+b^2)$	$-\frac{1}{4}y_\tau(a+b)^2$	$-\frac{1}{4}y_t(a+b)^2$	$-\frac{1}{4}y_b(a+b)^2$
Model	C_{tH}	C_{bH}	C_{Ht}	C_{HG}	
TB	$\frac{M_{TB}^2}{v^2}y_t a^2$	$\frac{M_{TB}^2}{v^2}y_b b^2$	$-\frac{M_{TB}^2}{v^2}a^2$	$-\frac{M_{TB}^2}{v^2}\frac{\alpha_s(0.65)}{8\pi}b^2$	

Table 3.7.3. Operator coefficients generated by the tree-level single-field models B, W and TB, which each depend on two couplings a and b, with y_t, y_b and y_τ denoting the top, bottom and tau Yukawa couplings respectively, v denoting the electroweak scale and α_s denoting the strong coupling. The coefficients of all operators are proportional to the corresponding entries in the Table and divided by M^2.

TeV units, assuming that the corresponding couplings are set to unity, are shown as horizontal bars.[14] We note that most of these limits exceed 1 TeV for unit coupling and do not depend on kinematic distributions probing this region, in which case the SMEFT approach is self-consistent. The $SU(2)$-singlet VLQ top-partner model (T) and S are the most poorly constrained, with $m_{S,T}$ >900, 770 GeV. We also show in grey boxes the corresponding bounds on the squared couplings, assuming a mass scale of 1 TeV. Most of the bounds are < 1, justifying a tree-level treatment. We also list all the pulls that exceed 1-σ, which is significant.

We can compare the mass limits for these models with the naive scale limits shown in Fig. 3.6.4. In a model-independent SMEFT analysis, one allows all the EFT operators to vary simultaneously. On the other hand, in specific models not all EFT coefficients are generated and those that do appear are related to each other in such a way that the number of free parameters of the model is matched to the number of independent EFT operators generated by the model. As an example, we discuss the following set of single-parameter models:

$$\Sigma,\ \Sigma_1,\ N \text{ and } E \quad ,$$

which all span the same types of SMEFT operators. They are all characterised by non-zero values for the following set of operators involving electroweak precision lepton observables:

$$C_{H\ell}^{1,3} \neq 0 \quad ,$$

whereas the other operator coefficients are zero, or very mildly constrained (e.g., $C_{\tau H}$, which is $\propto y_\tau$). Interpreted in terms of these models, the global SMEFT fit leads to

[14] In the case of the T (vector-like quark) model, the mass limit has been obtained using the relation $s_L^t \simeq \lambda v/\sqrt{2}M_T$ and setting $\lambda = 1$ [208].

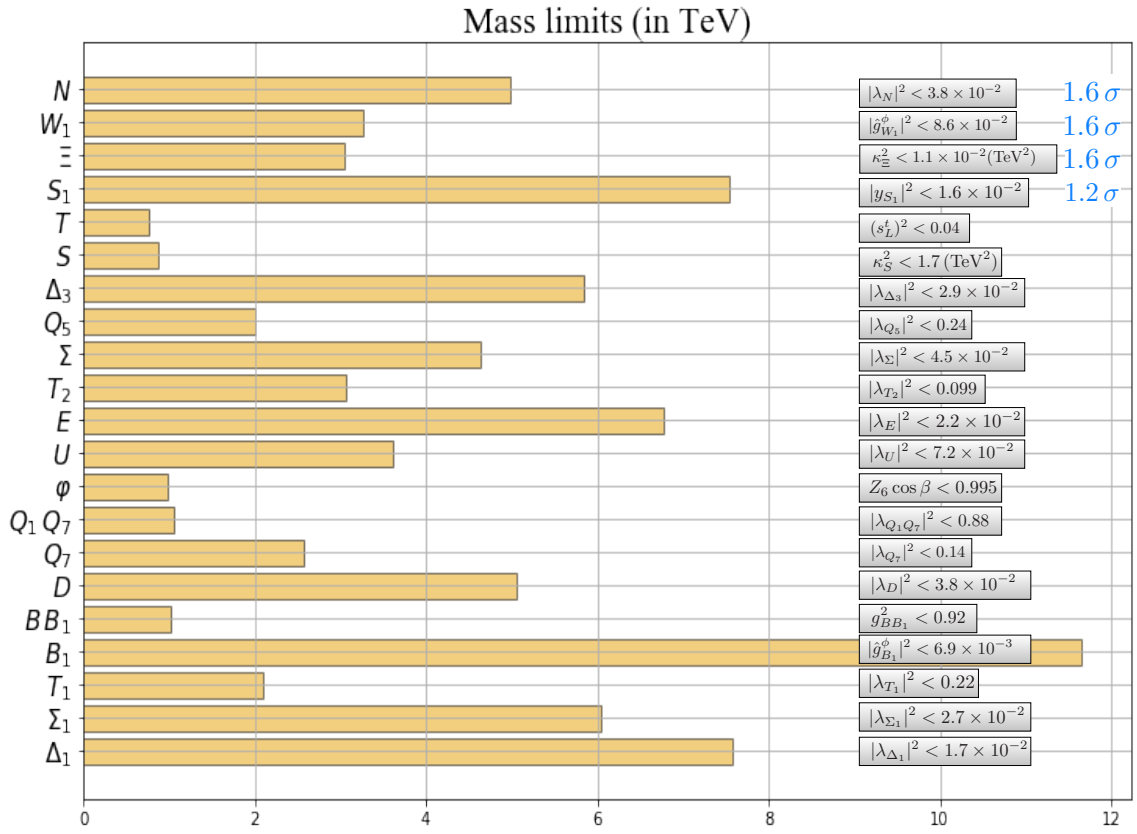


Fig. 3.7.1. The horizontal bars show the mass limits (in TeV) at the 95% CL for the models described in Table 3.7.1, setting the corresponding couplings to unity. The coupling limits obtained when setting the mass to 1 TeV are shown in grey boxes. We also note in light blue the pulls that exceed 1-σ.

mass limits of the order of 5 TeV for unit couplings, or corresponding coupling limits of $\mathcal{O}(10^{-1})$ for TeV resonances. Each of these particles generates a different relation between $C^1_{H\ell}$ and $C^3_{H\ell}$, leading to slightly different limits, e.g., for the model with a new neutral fermion N one expects $C^1_{H\ell} = -C^3_{H\ell} = |\lambda_N|^2/4m_N^2$. The sizes of the mass limits justify the SMEFT approach, and the small couplings for masses of a TeV justify working at tree level.

The constraints on particles beyond the SM would be weaker if their effects were not tree-level but loop-induced. These are typical of extensions of the SM where couplings to new states have to be in pairs, as would be the case if they carry a new conserved quantum number. We discuss in §3.7.3 one particularly interesting example with such loop-induced effects, namely stops in an R-parity-conserving supersymmetric model.

3.7.2 Tree-level SMEFT patterns

As already commented, and displayed in Table 3.7.2, simple extensions of the SM exhibit specific patterns in the operators they generate. Many operators have vanishing coefficients and those that are non-zero are often related. For example, model B_1 of Table 3.7.1 generates the bosonic operators $C_{H\Box}$, C_{HD} and Yukawa operators such as C_{tH}. These three operators are related by $C_{H\Box} = C_{tH} = -\frac{1}{2}C_{HD}$. Similarly, model W_1 generates a pattern $C_{H\Box} = C_{tH} = \frac{1}{2}C_{HD}$. Motivated by these patterns, we study the results of our fit in four subspaces of the SMEFT:

$$\begin{aligned}
\text{Boson-specific: } & (C_{HD},\, C_{H\Box},\, C_{tH}) \quad , \\
\text{Lepton-specific: } & (C_{He},\, C_{H\ell}^{(1,3)},\, C_{\ell\ell}) \quad , \\
\text{Quark-specific: } & (C_{Hu}, C_{Hd},\, C_{Hq}^{(1,3)},\, C_{tH}) \quad , \\
\text{Top-specific: } & ((C_{Hq}^{(1)})_{33},\, (C_{Hq}^{(3)})_{33},\, C_{HG},\, C_{bH},\, C_{tH}, C_{Ht}) \quad .
\end{aligned}$$

Results for the *boson-specific scenario* are shown in the top two panels of Fig. 3.7.2. We display 95% CL contours in the $(C_{H\Box}, C_{tH})$ plane as solid contours, marginalising over C_{HD}, and setting all other operator coefficients to zero. In the top panel we show how these operators are constrained in the cases of four specific UV models from Table 3.7.1: S, φ [which may be derived from a 2-Higgs Doublet Model (2HDM)], $Q_1\&Q_7$ and $B\&B_1$, showing a detail in the inset. As well as these 1-parameter models, we project two of the 2-parameter models shown in Table 3.7.3 onto the $(C_{H\Box}, C_{tH})$ plane : B and W. These cases are vector bosons with couplings $\hat{g}_H^B$ and $\hat{g}_H^W$ to the Higgs doublet, also known as Z' and W' bosons, respectively. Model B projects onto a line in the $(C_{H\Box}, C_{tH})$ plane, illustrated in the top panel for $r_B = 2, -2$, where $r_B = \text{Re}(\hat{g}_H^B)/\text{Im}(\hat{g}_H^B)$. Model W also projects onto a line, illustrated here for $r_W = -1, \frac{1}{3}$. While the slope of the line generated by model B is free to take any value, the line generated by model W is constrained to lie within the wedge bounded by the $r_W = -1, \frac{1}{3}$ lines shown.

In the middle panels of Fig. 3.7.2 we zoom in on the yellow ellipse shown in the top panel, so as to study the constraints in the $(C_{H\Box}, C_{tH})$ plane when $C_{H\Box} \propto C_{HD}$. This results in tighter constraints on $C_{H\Box}$ compared to when C_{HD} is treated as an independent parameter, as shown by the small ellipses in the left panel. The constants of proportionality $(-2, -4, 2)$ are those found in the patterns generated by models B_1, Ξ and W_1. Zooming in further in the right panel, these ellipses are squashed into near-vertical parallel lines showing the constraints on $C_{H\Box}$ and C_{tH} in the case of each of these 1-parameter models.

The *lepton-specific scenario* is shown in the bottom panel of Fig. 3.7.2, where the coefficients of the operators C_{ll}, $C_{Hl}^{(1)}$, $C_{Hl}^{(3)}$ and C_{He} are studied. The ellipses show how the data constrain $C_{Hl}^{(1)}$ and C_{He}, marginalising over $C_{Hl}^{(3)}$. In the yellow ellipse we also marginalise over C_{ll}, allowing for the nonzero values of C_{ll}. The constraints shrink when $C_{ll} = 0$, as in models Σ, Σ_1, N, E, Δ_1 and Δ_3.[15] The inset plot demonstrates how $C_{Hl}^{(1)}$ and C_{He} are constrained when we restrict these operators to the patterns associated with each of these 1-parameter models.

In Fig. 3.7.3 we turn to quark-Higgs interactions. The flavour-universal *quark-specific scenario* is shown in the upper panel, where we constrain $(C_{Hq}^{(1)}, C_{Hu})$, marginalising over the remaining quark-Higgs operators $C_{Hq}^{(3)}$, C_{Hd} and C_{tH}. Models that generate these operators often lead to the pattern $C_{Hq}^{(1)} \propto C_{Hq}^{(3)}$. The green and red ellipses show how such patterns narrow the constraints on C_{Hu} and $C_{Hq}^{(1)}$ for two examples: $C_{Hq}^{(3)} = \pm C_{Hq}^{(1)}$. Specialising to the 1-parameter models U, D, T_1, T_2 and Q_7 further restricts the operators, leading to the 1-dimensional constraints shown in the inset plots.

Finally, the lower panel of Fig. 3.7.3 considers the flavour-non-universal *top-specific scenario*, where we consider the operators $(C_{Hq}^{(1)})_{33}$, $(C_{Hq}^{(3)})_{33}$, C_{HG}, C_{bH}, C_{tH} and C_{Ht}. These operators are generated by the vector-like quark models T and TB with couplings to the third generation quarks only. Integrating out the $SU(2)_L$ singlet T generates the pattern $(C_{Hq}^{(3)})_{33} = -(C_{Hq}^{(1)})_{33}$, $C_{bH} = C_{Ht} = 0$. The green ellipse demonstrates how this pattern tightens the constraints on C_{tH}. In contrast, the $SU(2)_L$ doublet TB does not generate the $(C_{Hq}^{(3)})_{33}$ operator. Setting $(C_{Hq}^{(3)})_{33} = 0$ results in much narrower constraints in the $(C_{Hq}^{(1)})_{33}$ direction, as shown by the red ellipse.

These patterns, and the results shown in Figs. 3.7.2 and 3.7.3, may be considered as more general explorations of the model parameter space than in the two previous Sections. Readers exploring UV completions who are searching for the indirect LHC and LEP constraints on their models can match their scenario to the allowed ellipses in these figures. For example, models linked to neutrino physics could lead to the SMEFT pattern we have denoted as *lepton-specific*, whereas models with various additional scalars and gauge bosons would be contained among the *boson-specific scenarios*, and models with additional coloured particles could be included among the *quark-specific scenarios*.

[15] See also Ref. [209] for a more detailed fit to these six vector-like lepton models.

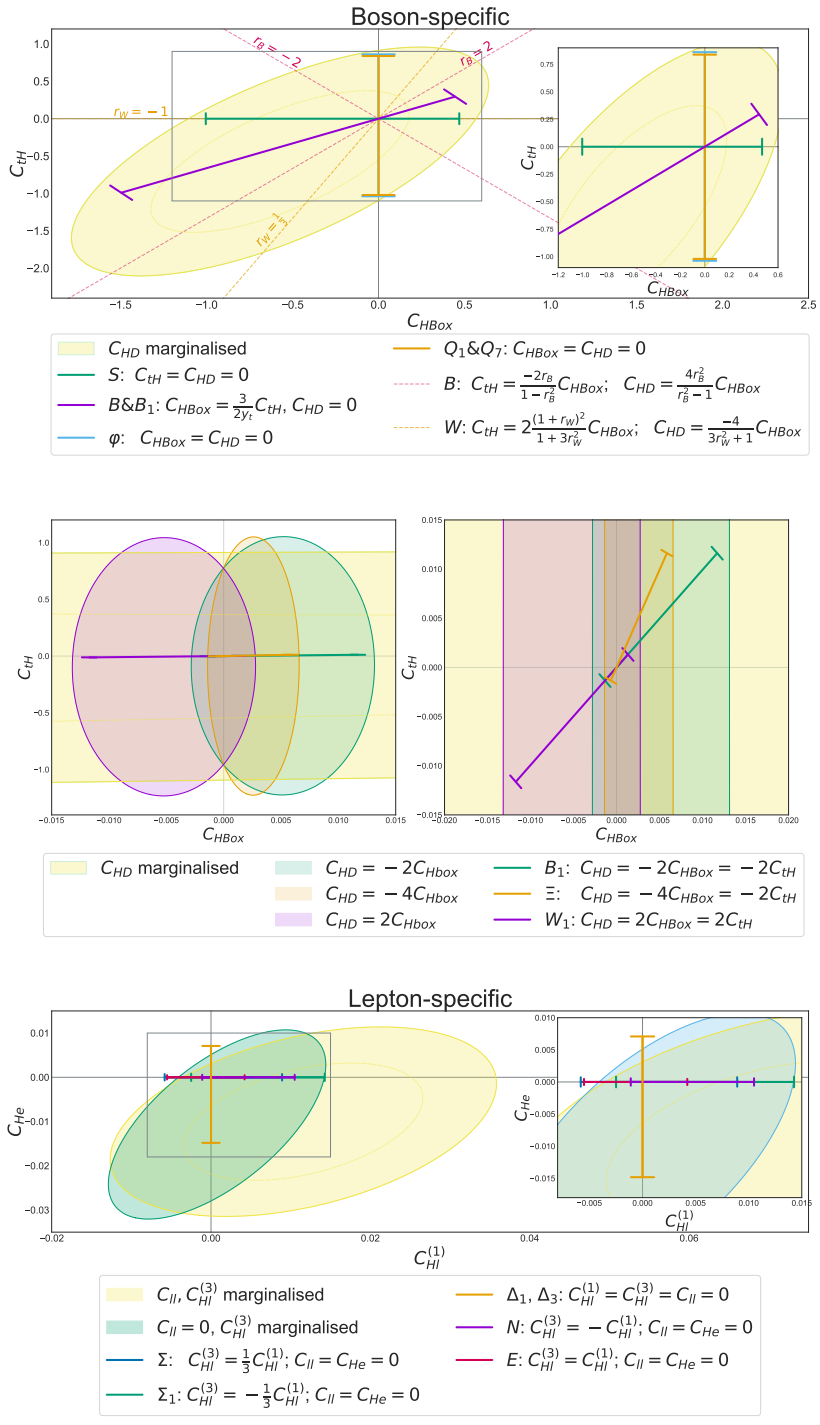


Fig. 3.7.2. Constraints at the 95% CL on $(C_{H\Box}, C_{tH})$ in the *boson-specific scenario* (upper two panels) and $(C_{Hl}^{(1)}, C_{He})$ in the *lepton-specific scenario* (bottom panel). The lines correspond to the $2-\sigma$ limits obtained when we restrict the operators to the relations generated by integrating out the indicated single-field extensions of the SM.

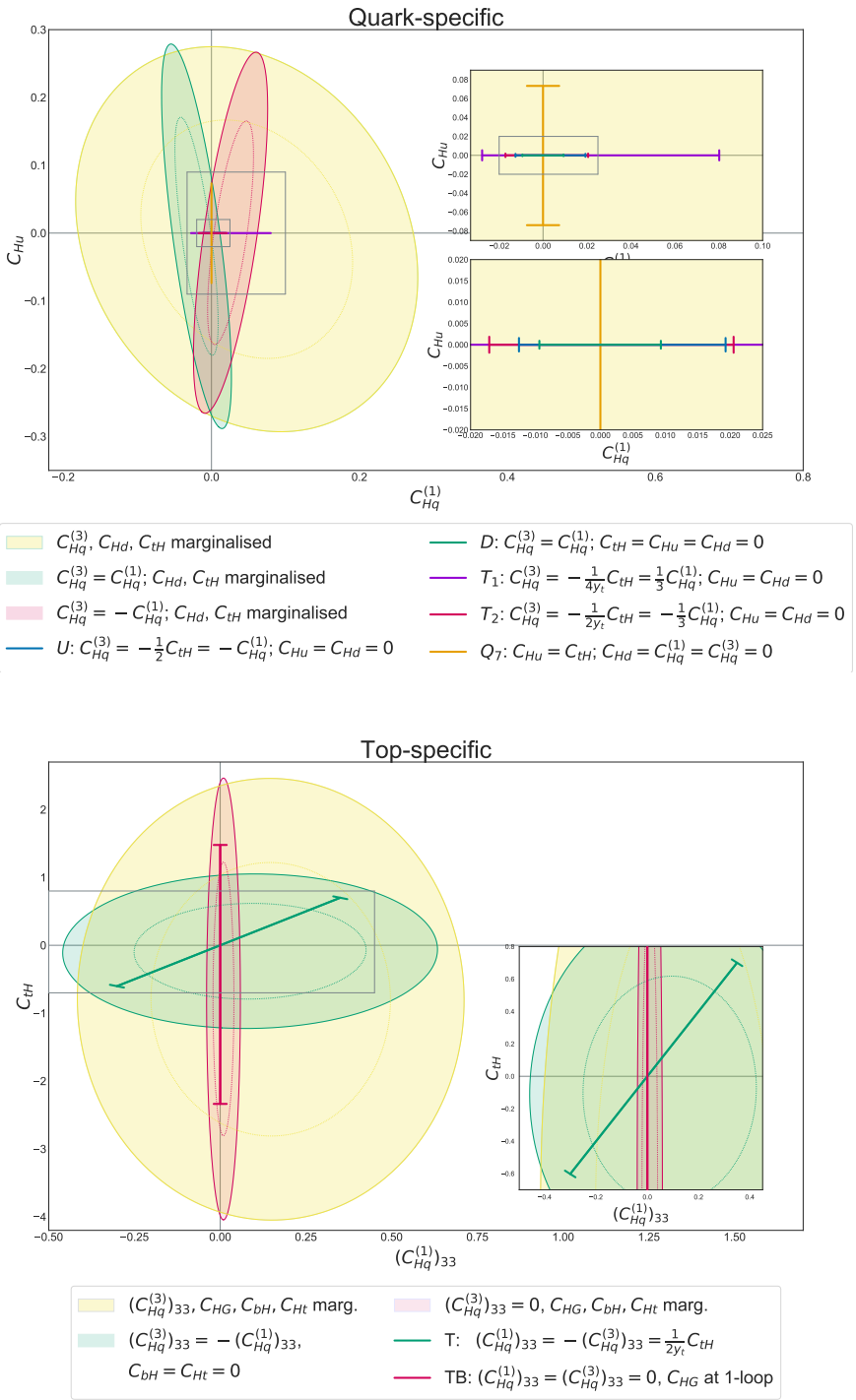


Fig. 3.7.3. Constraints at the 95% CL on $(C_{Hq}^{(1)}, C_{Hu})$ in the *quark-specific scenario* and $((C_{Hq}^{(1)})_{33}, C_{tH})$ in the *top-specific scenario*. The lines correspond to the $2-\sigma$ limits obtained when the operator coefficients are restricted to the relations generated by integrating out single-field extensions of the SM.

3.7.3 *R*-parity-conserving stop squarks at the 1-loop level

A particularly interesting loop-induced modification of the SM Lagrangian is *R*-parity-conserving supersymmetry with a light stop sector. Whereas the discussion in the previous Section of single-field tree-level models was motivated by simplicity, this scenario is motivated by the naturalness of the hierarchy between the electroweak scale and that of gravity or grand unification. A complete one-loop analysis of the light-stop scenario and a comparison with the SMEFT analysis was given in [210].

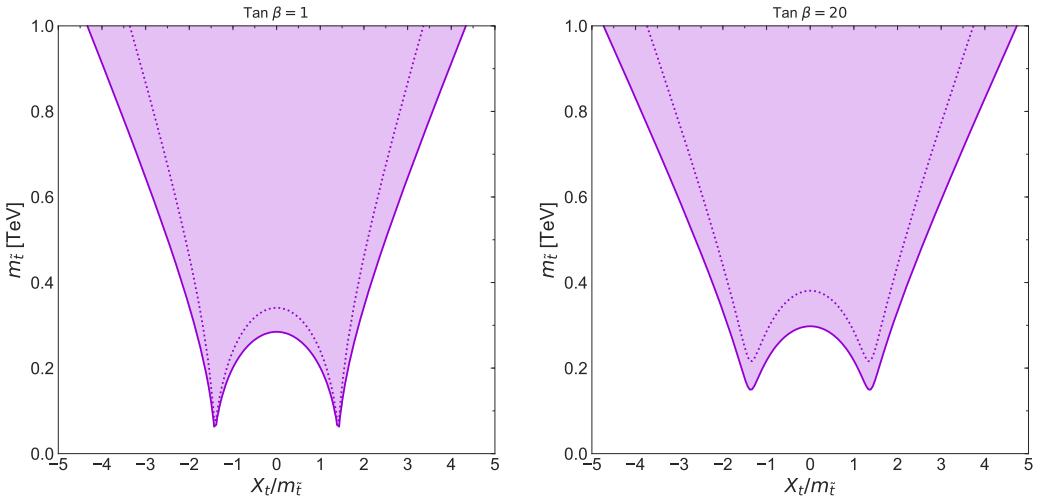


Fig. 3.7.4. Limits from the global fit in the stop parameter plane, $(\frac{X_t}{m_{\tilde{t}}}, m_{\tilde{t}},)$. The two panels correspond to the low and high $\tan\beta$ choices, 1 and 20 respectively.

In presenting our results, we follow [211], assuming a common diagonal mass term $m_{\tilde{t}}$ and denoting the stop mixing parameter by X_t. The constraints on degenerate stops are dominated by measurements of the $H \to gg$ and $H \to \gamma\gamma$ couplings, which constrain the dimension-6 operators C_{HG}, C_{HB}, C_{HW} and C_{HWB}. These constrain the stop parameters $m_{\tilde{t}}$ and X_t through the following relations:

$$
\begin{aligned}
C_{HG} &= \frac{g_s^2}{12}\frac{h_t^2}{(4\pi)^2}\left[(1+\frac{1}{12}\frac{c_{2\beta}g'^2}{h_t^2}) - \frac{1}{2}\frac{X_t^2}{m_{\tilde{t}}^2}\right], \\
C_{HB} &= \frac{17g'^2}{144}\frac{h_t^2}{(4\pi)^2}\left[(1+\frac{31}{102}\frac{c_{2\beta}g'^2}{h_t^2}) - \frac{38}{85}\frac{X_t^2}{m_{\tilde{t}}^2}\right], \\
C_{HW} &= \frac{g^2}{16}\frac{h_t^2}{(4\pi)^2}\left[(1-\frac{1}{6}\frac{c_{2\beta}g'^2}{h_t^2}) - \frac{2}{5}\frac{X_t^2}{m_{\tilde{t}}^2}\right], \\
C_{HWB} &= -\frac{gg'}{24}\frac{h_t^2}{(4\pi)^2}\left[(1+\frac{1}{2}\frac{c_{2\beta}g^2}{h_t^2}) - \frac{4}{5}\frac{X_t^2}{m_{\tilde{t}}^2}\right],
\end{aligned}
\tag{3.7.1}
$$

where $h_t \equiv \frac{m_t}{v}$, m_t denotes the top mass and β is related to the ratio of vacuum expectation values: $\tan\beta = \frac{\langle H_u \rangle}{\langle H_d \rangle}$. We calculate the constraints in the $(\frac{X_t}{m_{\tilde{t}}}, m_{\tilde{t}})$ plane for the representative values $\tan\beta = 1$ and 20, which are shown in the left and right panels of Fig. 3.7.4, respectively.

One sees in both panels of Fig. 3.7.4 that current LHC data constrain the stop mass scale to $\gtrsim 300$ GeV, except for $|\frac{X_t}{m_{\tilde{t}}}| \sim 1.5$ [212], where partial cancellations reduce the sensitivity to the stop mass scale below 200 GeV. In these regions the SMEFT analysis gives only qualitative results. These blind directions could be eliminated with future measurements of the H+jet differential distribution [213].

3.7.4 Survey of combinations of multiple operators

In general, new physics beyond the Standard Model could be expected to contribute to the SMEFT via exchanges of more than just a single massive particle, just as, e.g., W and Z exchanges both contribute to the Fermi 4-fermion EFT of the weak interactions, and various mesons including vectors ρ and scalars σ contribute to the low-energy pionic EFT of QCD. Another example is provided by supersymmetry, where there might be a pair of relatively light stops, which contribute to four different dimension-6 operator coefficients, as discussed in the previous subsection.

With this motivation, we have surveyed all fits with contributions from any combination of two, three, four or five dimension-6 operators, namely 561, 5984, 46736 and 278256 combinations, respectively. For each combination $\{\mathcal{O}_i\}$, we calculate the pull that the corresponding fit exerts, given by $P \equiv \sqrt{\chi^2_{\rm SM} - \chi^2_{\{\mathcal{O}_i\}}}$. Calculations of P for all of these combinations is possible only because, in the linear treatment that we have adopted in this Chapter, the calculations of the $\chi^2_{\{\mathcal{O}_i\}}$ are computationally undemanding.

Fig. 3.7.5 displays stacked histograms of the distributions of the pulls P obtained in fits to combinations of 2 (upper left), 3 (upper right), 4 (lower left) and 5 (lower right) operators $\{\mathcal{O}_i\}$. In each panel, the blue histogram is for combinations that include only operators that affect $t\bar{t}$ production, see Fig. 3.6.2, the orange histogram is for combinations that do not include any of these operators, and the green histogram is for the remaining combinations.

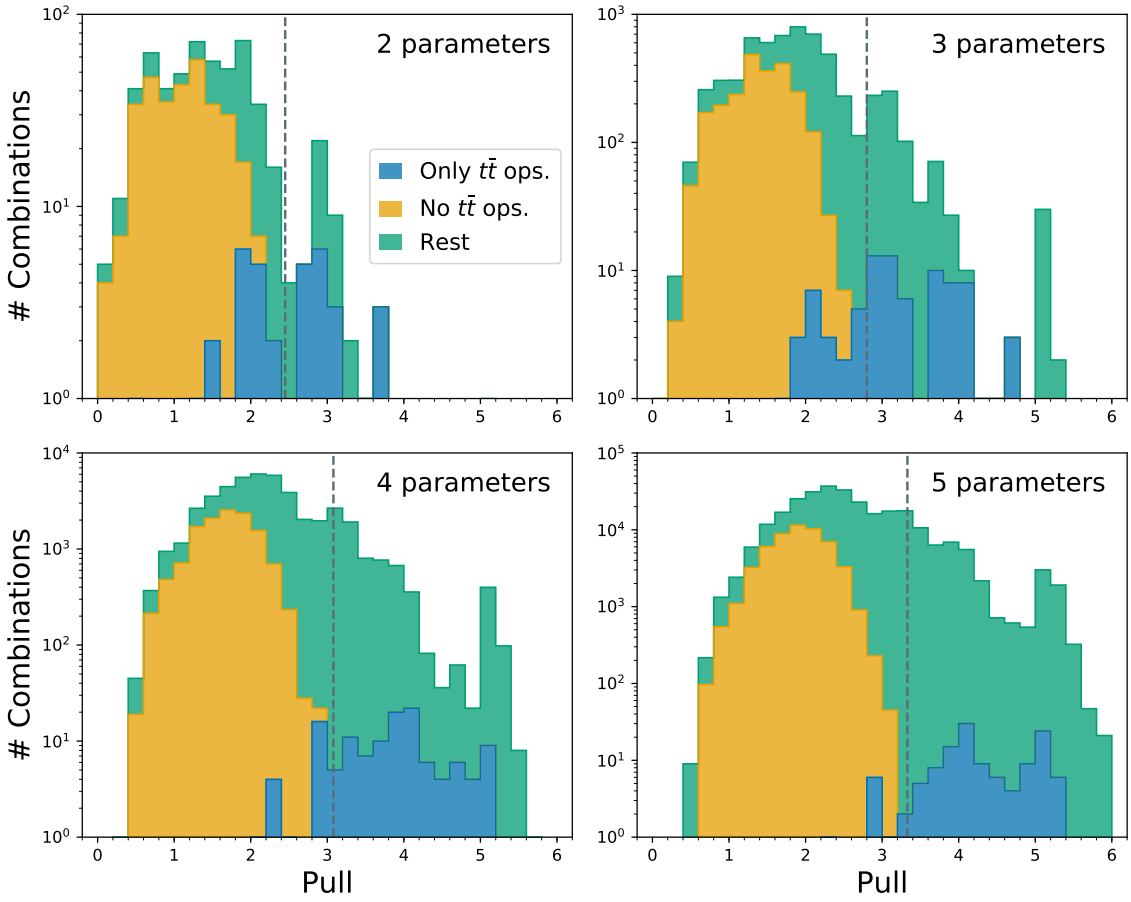


Fig. 3.7.5. Stacked histograms of the distribution of pulls obtained in fits to 2 (upper left), 3 (upper right), 4 (lower left) and 5 (lower right) parameter subsets. The subsets have been split into three categories: those that include only operators that affect $t\bar{t}$ production (blue), those that do not include operators that affect $t\bar{t}$ production (orange), and the rest (green). The dashed vertical lines mark the expected 95% ranges for the pull distributions.

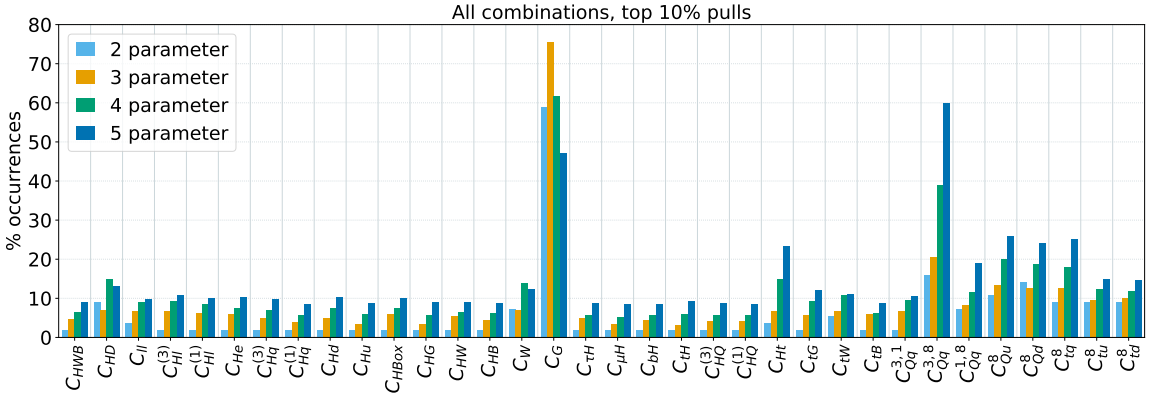


Fig. 3.7.6. The percentage occurrence of a given operator in the combinations ranked in the top 10% of pulls found in fits to 2 (light blue), 3 (orange), 4 (green) and 5 (blue) parameter subsets.

In all panels of Fig. 3.7.5 we see that the blue histograms for combinations that include operators affecting $t\bar{t}$ production exhibit peaks at $P > 2$ that move to > 4 for combinations of 5 operators. On the other hand, the orange histograms for combinations of operators that do not affect $t\bar{t}$ production are peaked at lower values of $P \lesssim 2$ and do not have long tails extending to large values of P, while the peaks of the green histograms are intermediate. The vertical lines in the panels of Fig. 3.7.5 mark the 95% ranges expected for Gaussian distributions of the pulls in the cases of 2, 3, 4 and 5 operators. We find that the percentage of combinations above the 95% marks range from 9% (2 operators) to 14% (5 operators) of the total number of combinations: more than expected for Gaussian distributions, but not excessive. Neither the orange nor the green histograms provide any indication of a significant deviation from the SM that can be described by any combination of 2, 3, 4 or 5 operators.

Fig. 3.7.6 shows how often any given operator appears in the combinations whose total pulls P ranked among the top 10% of those obtained in fits with combinations of 2 (light blue), 3 (orange), 4 (green) and 5 (dark blue) operators. The operators affecting $t\bar{t}$ production generally appear more often among these top 10% combinations, particularly $\mathcal{O}_G$. The prevalence of these operators in the high-pull combinations is due to the relatively poor quality of the global linear fit in the top sector that we discussed in §3.6.2.[16] Overall, we see that $\mathcal{O}_G$ appears in more than half of the top 10% of combinations of ≤ 4 operators, falling to somewhat over 40% for combinations of 5 operators. For comparison, almost 60% of the 5-operator combinations include $\mathcal{O}_{Qq}^{3,8}$, whereas this operator appears in smaller proportions of the 2-, 3- and 4-operator combinations.

This survey would suggest that the best prospects for BSM physics may be among the operators affecting $t\bar{t}$ production, particularly $\mathcal{O}_{Qq}^{3,8}$. However, we would emphasise that the data in the top sector are currently the least precise, and that there may be an issue with the $t\bar{t}$ mass distribution near threshold, as discussed in §3.6.2. As the constraints on the scales of the top operators are relatively weak, the resolution of this issue may require including quadratic contributions of these operators, which are not yet all available. We recall in this connection that an analysis of C_G including quadratic contributions to multijet production found a strong constraint restricting C_G to small values below the sensitivity of our analysis.

There are many physics scenarios that suggest the appearance of BSM physics in the top sector, such as the light stop scenario discussed in §3.7.3. However, as we see in

[16] As discussed there, this issue might be mitigated by including quadratic terms in the SMEFT expansion, but an analysis of this possibility lies beyond the scope of the present analysis.

Eq. (3.7.1), the operators $\{\mathcal{O}_{HG}, \mathcal{O}_{HB}, \mathcal{O}_{HW}, \mathcal{O}_{HWB}\}$ that are most constrained in this scenario contribute primarily in the electroweak and Higgs sectors, rather than the top sector. We find a pull $P = 1.9$ for the combination $\{\mathcal{O}_{HG}, \mathcal{O}_{HB}, \mathcal{O}_{HW}, \mathcal{O}_{HWB}\}$, which is typical for 4-operator combinations that do not include $t\bar{t}$ operators (the orange histogram in the lower left panel of Fig. 3.7.5), and the maximum pull when a fifth operator is included is $P = 3.2$, which occurs in the combination with $\mathcal{O}_G$, and is typical of the green histogram in the lower right panel of Fig. 3.7.5. So we find no hint of light supersymmetry in the current data.

That said, this type of broad-brush survey of operator combinations may be a useful way to help optimise the search for BSM physics using the SMEFT in the future.

3.8 Concluding discussion

We will conclude this Chapter by summarising the work presented here. We have made use of the `Fitmaker` code to produce a global analysis of the available top, Higgs, diboson and electroweak data in the framework of the SMEFT with dimension-6 operators included to linear order. We have presented results for fits including each operator individually, and also when marginalising over all other operators. In each case, we have presented results in an $SU(3)^5$ flavour-symmetric scenario and in an $SU(2)^2 \times SU(3)^3$ top-specific scenario. Our results are displayed in Fig. 3.6.4, with numerical results for the $SU(2)^2 \times SU(3)^3$ top-specific scenario presented in Table C.0.1 in Appendix C.

We find $\chi^2/\text{dof} = 0.94$ ($p = 0.72$, $n_{\text{dof}} = 184$) for our flavour-universal global fit, to be compared with $\chi^2/\text{dof} = 0.93$ ($p = 0.76$, $n_{\text{dof}} = 204$) in the SM. Similarly, we find $\chi^2/\text{dof} = 0.81$ ($p = 0.99$, $n_{\text{dof}} = 321$) for our top-specific fit, to be compared with $\chi^2/\text{dof} = 0.91$ ($p = 0.87$, $n_{\text{dof}} = 341$) in the SM. We do not find any significant discrepancy with the SM. However, the data in the top sector show a preference for a non-zero value for the coefficient C_G of the triple-gluon operator $\mathcal{O}_G$, which is mirrored by trends in the coefficients of other operators affecting $t\bar{t}$ measurements. This deviation can be traced back to the behaviour near threshold of the $t\bar{t}$ cross section, and we await with interest future experimental measurements and developments in their theoretical understanding. We note also that a fit to multijet data at quadratic order in C_G has constrained it to values so small that it would not contribute significantly to the measurements we consider. However, we have not included this constraint in our linear fit, for reasons of theoretical consistency.

We have analysed the constraints our results provide on single-field extensions of the SM that contribute to SMEFT operator coefficients at the tree level. Normalising to unit couplings, the lower limits on the corresponding BSM particle masses range between > 1 TeV to > 10 TeV. The largest pulls $P \equiv \sqrt{\chi^2_{\rm SM} - \chi^2_{\rm BSM}}$ are 1.6, and hence not significant. In some instances, a particular BSM particle may contribute to several operator coefficients, and we have analysed the constraints in boson-, lepton-, quark- and top-specific subspaces of the SMEFT. We have also analysed the constraints on low-mass stops, which contribute significantly to four operator coefficients at the one-loop level, finding they must weigh more than ~ 300 GeV when the stop mixing parameter satisfies $X_t = 0$.

Finally, we have surveyed the constraints on all possible 2-, 3-, 4- and 5-operator combinations, assuming that the other operators vanish. We find pulls that are insignificant for combinations that do not contain operators affecting $t\bar{t}$ measurements, whereas the pulls for combinations of $t\bar{t}$ operators are larger. However, the same caveats apply to their interpretation as to the discussion of the $t\bar{t}$ sector above. The full pull distributions including all operator combinations do not exhibit any significant features.

These examples indicate ways in which a global analysis of all the current data may be used to obtain the broadest possible, unbiased view on the nature of possible BSM physics within the assumptions of the SMEFT framework. If any specific model or pattern of SMEFT operators were to exert a significant pull, it would be a first indication of the direction of new physics, which could be followed up with a more focused study. As we saw in Chapter 2, a similar procedure of fitting combinations of operators of the weak effective theory to the NCBAs lead to an observation of a significant pull in the combination $C_9 = -C_{10}$, which we will follow up with a more detailed study in Chapter 5. Unfortunately, the dataset we have included in this SMEFT analysis provides no significant indication of possible BSM physics: the only operator for which a non-zero coefficient is preferred is $\mathcal{O}_G$, and this preference is not very convincing. It is driven, in particular, by the threshold behaviour of the $t\bar{t}$ production cross section, but an analysis of multijet data at quadratic order prefers much smaller values of C_G. More $t\bar{t}$ data and theoretical understanding may be needed to resolve this discrepancy.

Although such a global analysis of the SMEFT provides constraints on new physics which are unbiased by model dependence, biases may creep in through other sources. The parton distribution functions used in the computation of our theoretical predictions have been fit while assuming the SM, and may have unknowingly absorbed SMEFT

effects into the PDF parametrisation as a result. In a perfect world, a global analysis such as the one presented here would instead incorporate SMEFT-dependent PDFs, producing a truly global and simultaneous fit of both. Of course, this is computationally infeasible, at least at present. For now, turning to Chapter 4, we will focus on such a simultaneous fit in the context of only a select few observables: measurements of deep inelastic scattering and Drell-Yan distributions.

Chapter 4

Parton Distributions in the SMEFT from high-energy Drell-Yan tails

4.1 Introduction

In Chapter 2 we saw that when constraining the coefficients of an EFT, the use of parton distribution functions obtained from data while assuming the SM leads to an inconsistency in our theoretical predictions. One may ask why we focus on the PDFs, when this type of inconsistency may arise elsewhere. Parton showering, for example, is often used in generating the EFT predictions for LHC observables. Parton showering algorithms assume the SM, again leading to an inconsistency. We focus here on the PDFs because they are relevant to high-p_T observables. As we discussed in Chapter 2, PDFs receive important constraints from high-p_T observables, and the subtle deviations we are searching for in this regime provide an ideal setting for new physics to be swamped by our theoretical uncertainties and 'fit away' into the PDFs. This is not true of inputs such as the parton shower, which deal with much lower energy scales.

As the LHC moves to the HL-LHC, it important that we begin to shed light on this interplay between the PDF and EFT. The goal of this Chapter is to do so, presenting a study of the interplay between PDFs and EFT effects for high-mass Drell-Yan (DY) processes at the LHC, based on Ref. [2]. We will assess the impact of a simultaneous determination of the PDFs and EFT on the constraints we obtain on the Wilson coefficients, effectively answering the question: *have we been successfully constraining the SMEFT so far, despite this inconsistency?*

Drell-Yan processes in general, and high-mass measurements in particular, provide information on the light quark and anti-quark PDFs in a broad region of x representing an important ingredient in modern global PDF fits [214–217]. This fact, combined

with their power in constraining the SMEFT as outlined in Chapter 2, indicates that Drell-Yan processes are an ideal setting for the investigation into the interplay between PDF and EFT effects. Furthermore, high-mass Drell-Yan data will be instrumental at the High-Luminosity LHC (HL-LHC) to pin down the large-x PDFs [218]. We will therefore continue our study of the PDF and EFT interplay into the HL-LHC era, creating projections for the EFT bounds expected from a simultaneous fit to HL-LHC data. In total, our analysis accounts for available unfolded high-mass Drell-Yan cross section data, detector-level searches based on the full Run II luminosity, and dedicated HL-LHC projections. We do not consider other data beyond the DIS and DY processes, in order to ensure a theoretically consistent description of the parton distributions in the SMEFT. The price to pay for this reduced dataset is the information loss on some flavour combinations, specifically on the gluon PDF.

In order to interpret the Drell-Yan data in the EFT framework we formulate simple, yet motivated, BSM benchmark scenarios, which are chosen to represent a wide class of UV-complete theories. In the first scenario [8], we consider the $\hat{W}$ and $\hat{Y}$ electroweak parameters generated in universal theories that modify the electroweak gauge boson propagators and lead to flavour-universal deviations which grow with the invariant mass. In the second benchmark [83], we consider a flavour-specific scenario motivated by the NCBAs.

This Chapter will proceed as follows. First of all, in §4.2 we discuss the EFT benchmark scenarios that will be used in the fits. In §4.3 we summarise the datasets used in our analysis and the corresponding theoretical calculations, both in the SM and in the SMEFT, and discuss their impact on a PDF fit. We also describe the methodology we use to simultaneously fit PDFs and SMEFT coefficients, and how we deal with several sources of uncertainties in the fits. In §4.4 we present the results for the simultaneous determination of the SMEFT coefficients and the PDFs from the available high-mass DY data from LHC Run I and Run II in the two scenarios presented in §4.2, and assess how they modify the interpretation of BSM searches based on the SM PDFs. In §4.5 we present a summary of the constraints we find on the two scenarios considered and assess the outcome of a joint PDF and EFT analysis at the HL-LHC.

More technical discussions are collected in the appendices of this thesis, and include detailed comparisons of the SM PDF fits produced in this work with previous NNPDF global fits (App. D), the quantitative assessment of the fit quality to the various input datasets (App. E), a benchmarking study for the calculation of EFT cross sections (App. F), and a study of the flavour dependence of the SMEFT PDFs (App. G).

4.2 SMEFT benchmark scenarios

In this Section we present the two SMEFT benchmark scenarios that will be used in this work to interpret the LHC Drell-Yan processes. The first scenario belongs to the class of electroweak precision tests and is sensitive to a broad range of UV-complete theories proposed in the literature. The second benchmark represents a consistency check of the existing hints of LFU violation in rare B-meson decays reported by the LHCb collaboration. Both scenarios highlight the interplay between the PDFs and the EFT dynamics, illustrating in particular how the former changes and how constraints to the latter are modified.

4.2.1 Benchmark I: oblique corrections $\hat{W}$ and $\hat{Y}$

The oblique corrections, as originally proposed in [219, 220], play a key role in testing theories beyond the Standard Model. They parametrise the self-energy $\Pi_V(q^2)$ of the electroweak gauge bosons W^a_μ and B_μ, where $V = W^3W^3$, BB, W^3B, and W^+W^-. Truncating the momentum expansion at order q^4, while imposing proper normalisation and symmetry constraints, one concludes that there are only four oblique parameters which can be identified with dimension-6 operators in the SMEFT. These are the well-known $\hat{S}$, $\hat{T}$, $\hat{W}$, and $\hat{Y}$ parameters [221]. The parameters $\hat{S}$ and $\hat{T}$ are well constrained from precision LEP measurements [221] and grow slowly with q^2, while $\hat{W}$ and $\hat{Y}$ scale faster implying that their effects will be enhanced for the high-energy dilepton tails at the LHC [8]. While $\hat{T} = \mathcal{O}(q^0)$ and $\hat{S} = \mathcal{O}(q^2)$, instead one has that $\hat{W}, \hat{Y} = \mathcal{O}(q^4)$. In the universal basis [222], the $\hat{W}$ and $\hat{Y}$ parameters are the Wilson coefficients associated to the operators $\mathcal{O}_{2W}$ and $\mathcal{O}_{2B}$ respectively,

$$\mathcal{L}_{\text{SMEFT}} \supset \frac{\hat{W}}{m_W^2}\mathcal{O}_{2W} + \frac{\hat{Y}}{m_W^2}\mathcal{O}_{2B}. \tag{4.2.1}$$

Here m_W indicates the W-boson mass and

$$\begin{aligned} \mathcal{O}_{2W} &= -\frac{1}{2}(D^\mu W_{\mu\nu})^2\,, \\ \mathcal{O}_{2B} &= -\frac{1}{2}(\partial^\mu B_{\mu\nu})^2\,, \end{aligned} \tag{4.2.2}$$

where D^μ is the covariant derivative. The physical effects of these operators on the Drell-Yan process arise from the corrections to the propagators through the self-energy modifications, see Eq. (1) of Ref. [8].

Alternatively, using the equations of motion, $\mathcal{O}_{2W}$ and $\mathcal{O}_{2B}$ can be rotated to the Warsaw basis, in which the modifications to the Drell-Yan cross sections are instead captured by four-fermion contact interactions. The operators in the Warsaw basis relevant to the description of the Drell-Yan process are given by

$$\begin{aligned}
&\mathcal{O}_{ld} = (\bar{l}\gamma_\mu l)(\bar{d}\gamma^\mu d)\,, \qquad \mathcal{O}_{lu} = (\bar{l}\gamma_\mu l)(\bar{u}\gamma^\mu u)\,, \quad \mathcal{O}_{lq}^{(1)} = (\bar{l}\gamma^\mu l)(\bar{q}\gamma_\mu q)\,,\\
&\mathcal{O}_{ed} = (\bar{e}\gamma_\mu e)(\bar{d}\gamma^\mu d)\,, \qquad \mathcal{O}_{eu} = (\bar{e}\gamma_\mu e)(\bar{u}\gamma^\mu u)\,, \quad \mathcal{O}_{qe} = (\bar{q}\gamma_\mu q)(\bar{e}\gamma^\mu e)\,,\\
&\mathcal{O}_{lq}^{(3)} = (\bar{l}\sigma^a\gamma^\mu l)(\bar{q}\sigma^a\gamma_\mu q)\,.
\end{aligned} \tag{4.2.3}$$

The flavour indices are contracted within the brackets, for example $(\bar{l}\gamma_\mu l) \equiv (\bar{l}^1\gamma_\mu l^1 + \bar{l}^2\gamma_\mu l^2 + \bar{l}^3\gamma_\mu l^3)$. After a tranformation from the universal basis to the Warsaw basis, the modifications to the Drell-Yan cross sections are captured by the following Lagrangian,

$$\begin{aligned}
\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} - \frac{g_L^2\hat{W}}{4m_W^2}\mathcal{O}_{lq}^{(3)} - \frac{g_Y^2\hat{Y}}{m_W^2}\Big(&Y_lY_d\,\mathcal{O}_{ld} + Y_lY_u\,\mathcal{O}_{lu}\\
&+ Y_lY_q\,\mathcal{O}_{lq}^{(1)} + Y_eY_d\,\mathcal{O}_{ed} + Y_eY_u\,\mathcal{O}_{eu} + Y_eY_q\,\mathcal{O}_{qe}\Big)\,.
\end{aligned} \tag{4.2.4}$$

Here q, l are the SM quark and lepton left-handed doublets, while u, d, e are the right-handed singlets. Also, g_L and g_Y are the corresponding electroweak gauge couplings, while σ^a are the Pauli matrices, and the hypercharge $Y_f = 1/6,\ -1/2,\ 2/3,\ -1/3,$ and -1 for q, l, u, d, e, respectively. Summation over flavour indices is assumed, which implies that in this scenario the fermionic currents respect the $SU(3)^5$ global flavour symmetry.

To understand this basis rotation explicitly, consider the $\hat{W}$ parameter. Using the equations of motion and following [222], $\mathcal{O}_{2W}$ can be expressed as

$$\mathcal{O}_{2W} \to \frac{1}{2}g_L^2\lambda v^2|H|^4 - \frac{3}{8}g_L^2\mathcal{O}_{H\Box} - g_L^2\lambda\mathcal{O}_H - \frac{1}{4}g_L^2\mathcal{O}_y - \mathcal{O}_{HJW} - \frac{1}{2}\mathcal{O}_{2JW} \tag{4.2.5}$$

where v denotes the Higgs VEV and λ parametrises the quartic Higgs interaction in the SM. The operators $\mathcal{O}_{H\Box}$ and $\mathcal{O}_H$ are both operators in the Warsaw basis. $\mathcal{O}_y$ is a linear combination of Yukawa-like dimension-6 Warsaw basis operators:

$$\mathcal{O}_y = [y_u]_{ij}[\mathcal{O}_{uH}]_{ij} + [V_{\text{CKM}}y_d]_{ij}[\mathcal{O}_{dH}]_{ij} + [y_e]_{ij}[\mathcal{O}_{eH}]_{ij} + h.c. \tag{4.2.6}$$

where V_{CKM} denotes the CKM matrix and y_u, y_d and y_e are the Yukawa matrices of the SM. Similarly $\mathcal{O}_{2JW}$ is a linear combination of dimension-6 Warsaw basis operators

coupling the Higgs doublet to fermions,

$$\mathcal{O}_{HJW} = \frac{1}{4} g_L^2 ([\mathcal{O}_{Hq}^{(3)}]_{ii} + [\mathcal{O}_{Hl}^{(3)}]_{ii}), \tag{4.2.7}$$

where we sum over the diagonal entries. Finally $\mathcal{O}_{2JW}$ denotes the following linear combination of 4-fermion operators in the Warsaw basis:

$$\mathcal{O}_{2JW} = g_L^2 \Big(\frac{1}{4} [\mathcal{O}_{qq}^{(3)}]_{iijj} - \frac{1}{4} [\mathcal{O}_{ll}]_{iijj} + \frac{1}{2} [\mathcal{O}_{ll}]_{ijji} + \frac{1}{2} [\mathcal{O}_{lq}^{(3)}]_{iijj} \Big) \tag{4.2.8}$$

where summation over repeated indices is implied. Of the list of operators appearing in Eq. 4.2.5 and in Eqs. 4.2.6, 4.2.7 and 4.2.8, only $[\mathcal{O}_{lq}^{(3)}]_{iijj}$ is relevant to Drell-Yan cross sections when we consider only the contributions arising from the interference of the SMEFT and SM amplitudes at LO in QCD and EW. All of the remaining operators will either contribute to the Drell-Yan observables at order $\hat{W}^2$, or will not contribute at all. We can therefore ignore these operators and capture the effects of the $\hat{W}$ parameter using only the operators $[\mathcal{O}_{lq}^{(3)}]_{iijj}$, as given in Eq. 4.2.4. This change of basis will similarly rotate $\mathcal{O}_{2B}$ into a linear combination of Warsaw basis operators, from which only the 4-fermion operators of Eq. 4.2.4 are relevant when dealing with Drell-Yan cross sections at leading order in $\hat{W}$, $\hat{Y}$.

The parametrisation in Eq. (4.2.3) has been implemented using the `SMEFTsim` package [164] and cross-checked against the reweighting method used in Ref. [83] (see also [9]), as will be discussed in Sect. 4.3.2 and in Appendix F.

The analysis in Ref. [8] reports the following 95% confidence level intervals on $\hat{W}$ assuming $\hat{Y} = 0$,

$$\begin{aligned} &\hat{W} \in [-3, 15] \times 10^{-4} \text{ (ATLAS 8 TeV, 20.3 fb}^{-1} \text{ [223])} , \\ &\hat{W} \in [-5, 22] \times 10^{-4} \text{ (CMS 8 TeV, 19.7 fb}^{-1} \text{ [224])} , \end{aligned} \tag{4.2.9}$$

as well as, the 95% confidence level intervals for $\hat{Y}$ assuming $\hat{W} = 0$,

$$\begin{aligned} &\hat{Y} \in [-4, 24] \times 10^{-4} \text{ (ATLAS 8 TeV, 20.3 fb}^{-1} \text{ [223])} , \\ &\hat{Y} \in [-7, 41] \times 10^{-4} \text{ (CMS 8 TeV, 19.7 fb}^{-1} \text{ [224])} . \end{aligned} \tag{4.2.10}$$

These bound have been computed by assuming SM PDFs. In our analysis, for this benchmark scenario, we see how the limits based on SM PDFs are modified once a consistent determination of the SMEFT PDFs is done, requiring a simultaneous fit

of the PDFs together with the $\hat{W}$ and $\hat{Y}$ parameters from the high-mass Drell-Yan distributions.

4.2.2 Benchmark II: left-handed muon-philic lepton-quark interactions

Following Ref. [83], here we consider gauge invariant four-fermion operators built from the SM quark and lepton $SU(2)_L$ doublets. In the Warsaw basis, Eq. (4.2.3), these correspond to the $\mathcal{O}_{lq}^{(3)}$ and $\mathcal{O}_{lq}^{(1)}$ operators. Expanding the $SU(2)_L$ indices, we find that the SMEFT Lagrangian contains operators of the form

$$\mathcal{L}_{\rm SMEFT} \supset \frac{\mathbf{C}_{ij}^{U\mu}}{v^2}(\bar{u}_L^i\gamma_\mu u_L^j)(\bar{\mu}_L\gamma^\mu\mu_L) + \frac{\mathbf{C}_{ij}^{D\mu}}{v^2}(\bar{d}_L^i\gamma_\mu d_L^j)(\bar{\mu}_L\gamma^\mu\mu_L)\ , \tag{4.2.11}$$

where $v \approx 246$ GeV is the Higgs vacuum expectation value and $\mathbf{C}_{ij}^{U\mu}$ and $\mathbf{C}_{ij}^{D\mu}$ represent matrices of Wilson coefficients. In Eq. (4.2.11), $i,j = 1,2,3$ indicate quark flavour indices, and we have chosen to focus on those operators that couple the quark fields exclusively to the second lepton family.

The operators highlighted in Eq. (4.2.11) have received a lot of attention recently in the context of the LHCb anomalies reported in rare B-meson decays [109, 225, 108, 226]. The reason is that the CKM flavour structure relates the $b \to s\mu^+\mu^-$ decays to the neutral-current Drell-Yan process at the LHC $pp \to \mu^+\mu^-$ [83]. The explicit models which successfully describe the LHCb anomalies, based on the $U(2)$ flavour symmetry and dominant dynamics with the third generation fermions [227–230], predict that the flavour channel dominating EFT effects in the Drell-Yan production is $b\bar{b} \to \mu^+\mu^-$ (see [231] for an explicit model example). The direct $b\bar{s}$ production channel is suppressed by V_{ts} and is therefore irrelevant. If the observed deviations in $R(K^{(*)})$ are due to new physics, in this class of models we generically expect $|\mathbf{C}_{33}^{D\mu}| \gtrsim 0.001$.

The ATLAS dimuon search reported in Ref. [232] is recast in Ref. [83] to set the limit on this scenario. In particular, the reported 95% confidence level interval is

$$\mathbf{C}_{33}^{D\mu} \in [-0.026, 0.021] \ \ \text{(ATLAS 13 TeV, 36.1 fb}^{-1}\ \text{[232])}. \tag{4.2.12}$$

For this second benchmark scenario we will assume that, out of the operators listed in Eq. (4.2.11), only a single Wilson coefficient is allowed to be non-zero. Specifically, we allow $\mathbf{C}_{33}^{D\mu} \neq 0$, while setting to zero all the coefficients of the other four-fermion operators. This assumption implies that in this scenario the SMEFT Lagrangian is

reduced to

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \frac{\mathbf{C}_{33}^{Du}}{v^2}(\bar{d}_L^3\gamma_\mu d_L^3)(\bar{\mu}_L\gamma^\mu\mu_L) \ . \tag{4.2.13}$$

In contrast to the previous benchmark, now the electron channel is SM-like. This feature provides a useful handle to separate PDF and EFT effects in the Drell-Yan process, by using electron data to determine the former and muon data to constrain both. Another difference with respect to the first benchmark is that here the leading new physics effects arise at the dimension-6 squared level, since the interference of the operator in Eq. (4.2.13) with the SM is subleading [83]. (Dijet production is another prominent process relevant for flavour physics [233] that enters into PDF fits. A detail study of the interplay is left for future work.)

To summarise, in this second benchmark scenario there is a single non-zero Wilson coefficient, $\mathbf{C}_{33}^{Du}$. Therefore, the determination of the SMEFT PDFs from Drell-Yan data requires a simultaneous fit of the PDFs together with the $\mathbf{C}_{33}^{Du}$ parameter. One should also note that this operator enters in the description of the DIS neutral-current structure functions via $\mu\, b$ scattering, though this contribution is highly suppressed due to the smallness of the bottom PDFs and the low energy scale probed by DIS data.

4.3 Experimental data, theory predictions, and fit settings

In Sect. 4.3.1 we present the LHC experimental data that will be used in the present analysis for the simultaneous determination of the PDFs and the EFT coefficients from high-mass Drell-Yan cross sections. We then describe in Sect. 4.3.2 the corresponding theoretical calculations, both in the SM and in the two SMEFT benchmark scenarios described in Sect. 4.2. In Sect. 4.3.3 we discuss the settings of the baseline SM PDF fit and assess the specific impact of the Run I and Run II high-mass Drell-Yan data on PDFs. Finally, in Sect. 4.3.3 we outline the fitting methodology adopted for the determination of the PDFs in the SMEFT, along with their simultaneous determination with the EFT Wilson coefficients.

4.3.1 Experimental data

The present analysis is based on the DIS and DY measurements which were part of the strangeness study of [12], which in turn was a variant of the NNPDF3.1 global PDF determination [215], extended with additional high-mass DY cross sections. The DIS

structure functions include the same legacy HERA inclusive combination [234] used in the DIS-only joint fit of PDF and EFT effects of [93].

No other datasets beyond DIS and DY are considered. In particular, the inclusive jet and top quark production measurements used in [12] are excluded from the present analysis. The rationale behind this choice is the following. SMEFT at dimension-6 level introduces 2499 independent parameters, many of which contribute to the processes used to extract the parton distribution functions. The full PDF fit in the SMEFT (with the consistent power counting in the inverse powers of the new physics scale) is the ultimate future goal of this line of research. Before that, we are forced to make assumptions about the subset of operators and processes involved. The restricted choice of DIS and DY is motivated by the idea that other datasets, such as inclusive jet, could potentially receive corrections from other SMEFT operators, e.g. four-quark operators while being insensitive to the semi-leptonic operators. Including all datasets to effectively determine PDF, while considering one or two operators able to impact a subset of processes, would misrepresent the realistic case.

For the purposes of our study, the DY data can be classified into low-mass, on-shell, and high-mass datasets. Table 4.3.1 summarises the low-mass and on-shell datasets, where in each case we indicate the experiment, the centre-of-mass energy $\sqrt{s}$, the publication reference, the physical observable, and the number of data points. The only difference as compared to [12] is the removal of the $W \to e\nu$ asymmetry measurements from D0 [235], which were found to be inconsistent with the rest of the Drell-Yan data.

In Table 4.3.2 we provide the same information as in Table 4.3.1 but for the neutral-current high-mass Drell-Yan datasets. In Table 4.3.2 we also indicate the final state, whether the distribution is 1D or 2D (thus differential only in the lepton invariant mass or differential in the lepton invariant mass and rapidity), the integrated luminosity $\mathcal{L}$, and the values of the dilepton invariant mass $m_{\ell\ell}$ for the most energetic bin. We note that while the ATLAS and CMS measurements at $\sqrt{s} = 7$ TeV [258, 259] were already part of the strangeness study of [12], the corresponding 8 TeV and 13 TeV measurements from [223, 224, 260] were not and are being considered for the first time in this analysis. For those datasets where data are available in terms of both Born and dressed leptons, the ATLAS 7 TeV analysis being an example thereof, we use the Born data so that it is not necessary to supplement our fixed-order predictions with final-state QED radiation corrections. The CMS 13 TeV data on the other hand are only provided in terms of dressed leptons. In total, there are either 270 or 313 data points in this high-mass category, depending on whether the 13 TeV CMS data are included in the combined channel or in the separate electron and muon channels.

Exp.	$\sqrt{s}$ (TeV)	Ref.	Observable	$n_{\rm dat}$
E886	0.8	[236]	$d\sigma^d_{\rm DY}/d\sigma^p_{\rm DY}$	15
E886	0.8	[237, 238]	$d\sigma^p_{\rm DY}/(dy\, dm_{\ell\ell})$	89
E605	0.04	[239]	$\sigma^p_{\rm DY}/(dx_F\, dm_{\ell\ell})$	85
CDF	1.96	[240]	$d\sigma_Z/dy_Z$	29
D0	1.96	[241]	$d\sigma_Z/dy_Z$	28
D0	1.96	[242]	$d\sigma_{W\to\mu\nu}/d\eta_\mu$ asy.	9
ATLAS	7	[243]	$d\sigma_W/d\eta_l, d\sigma_Z/dy_z$	30
ATLAS	7	[244]	$d\sigma_{Z\to e^+e^-}/dm_{e^+e^-}$	6
ATLAS	7	[245]	$d\sigma_W/d\eta_l, d\sigma_Z/dy_z$	61
ATLAS	7	[246]	$d\sigma_{W+c}/dy_c$	22
ATLAS	8	[247]	$d\sigma_Z/dp_T$	82
ATLAS	8	[248]	$d\sigma_{W+j}/dp_T$	32
CMS	7	[249]	$d\sigma_{W\to l\nu}/d\eta_\ell$ asy.	22
CMS	7	[250]	$d\sigma_{W+c}/dy_c$	5
CMS	7	[250]	$d\sigma_{W^++c}/d\sigma_{W^-+c}$	5
CMS	8	[251]	$d\sigma_Z/dp_T$	28
CMS	8	[252]	$d\sigma_{W\to\mu\nu}/d\eta_\mu$	22
CMS	13	[253]	$d\sigma_{W+c}/dy_c$	5
LHCb	7	[254]	$d\sigma_{Z\to\mu^+\mu^-}/dy_{\mu^+\mu^-}$	9
LHCb	7	[255]	$d\sigma_{W,Z}/d\eta$	29
LHCb	8	[256]	$d\sigma_{Z\to e^+e^-}/dy_{e^+e^-}$	17
LHCb	8	[257]	$d\sigma_{W,Z}/d\eta$	30
Total				**659**

Table 4.3.1. The low-mass and on-shell Drell-Yan datasets used in the present study. For each dataset we indicate the experiment, the centre-of-mass energy $\sqrt{s}$, the publication reference, the physical observable, and the number of data points

From Table 4.3.2 one can observe that, with the exception of the CMS 13 TeV data, only one specific leptonic final state is available to be used in the fit. For the CMS 13 TeV measurement instead, one can select between the combined channel or the individual electron and muon final states, which are statistically independent. The separate use of the electron and muon channels is potentially beneficial when considering BSM effects that are not lepton-flavour universal. For example, in benchmark scenario

Exp.	$\sqrt{s}$ (TeV)	Ref.	$\mathcal{L}$ (fb^{-1})	Channel	1D/2D	$n_{\rm dat}$	$m_{\ell\ell}^{\rm max}$ (TeV)
ATLAS	7	[258]	4.9	e^-e^+	1D	13	[1.0, 1.5]
ATLAS (*)	8	[223]	20.3	$\ell^-\ell^+$	2D	46	[0.5, 1.5]
CMS	7	[259]	9.3	$\mu^-\mu^+$	2D	127	[0.2, 1.5]
CMS (*)	8	[224]	19.7	$\ell^-\ell^+$	1D	41	[1.5, 2.0]
CMS (*)	13	[260]	5.1	e^-e^+, $\mu^-\mu^+$ $\ell^-\ell^+$	1D	43, 43 43	[1.5, 3.0]
Total						**270 (313)**	

Table 4.3.2. Same as Table 4.3.1 for the neutral-current high-mass Drell-Yan datasets considered in this work. We also indicate the final-state, whether the distribution is 1D (which are differential in the invariant mass, $m_{\ell\ell}$, of the final-state leptons) or 2D (which are differential in both the invariant mass of the leptons, $m_{\ell\ell}$, and in their rapidity, $y_{\ell\ell}$), and the values of $m_{\ell\ell}$ for the most energetic bin. Datasets indicated with **(*)** are used for the first time in this analysis in comparison with [12].

II described in Sect. 4.2, the theoretical predictions for the DY electron data would be those of the SM while those of the muon data should include EFT corrections. On the other hand in the (flavour-universal) $\hat{W}$ and $\hat{Y}$ scenario, it is more convenient to include the data from the combined channel, which displays reduced systematic uncertainties.

4.3.2 Theoretical predictions

We now discuss the settings of the theoretical calculations, both in the SM and in the SMEFT. Appendix F contains further information regarding the computation and benchmarking of the SMEFT corrections for both the DIS structure functions, for which the effect in both scenarios is negligible, and the DY cross sections, for which the impact of SMEFT corrections is much more sizeable.

SM cross sections. The SM cross sections are computed at next-to-next-to-leading order (NNLO) in QCD and include next-to-leading order (NLO) EW corrections, the latter being especially significant in the high-mass region relevant for this study. In particular, the DIS reduced cross sections (combinations of structure functions) are evaluated at NNLO in the FONLL-C general-mass variable flavour number scheme [261] with `APFEL` [262] interfaced to `APFELgrid` [263]. The Drell-Yan differential distributions are computed using `MCFM` [264] and MadGraph5_aMC@NLO [265] interfaced to

`APPLgrid` [266] and `APFELgrid` to generate fast NLO interpolation tables which are then supplemented by bin-by-bin K-factors to account for the NNLO QCD and NLO EW corrections. These K-factors are defined as

$$d\sigma_{pp} = \left(d\hat{\sigma}_{ij}\Big|_{\text{NLO QCD}} \otimes \mathcal{L}_{ij}^{\text{NNLO}} \right) \times K_{\text{QCD}} \times K_{\text{EW}} \,, \tag{4.3.1}$$

where $\otimes$ represents the standard convolution product, $d\sigma_{pp}$ ($d\hat{\sigma}_{ij}$) is the short-hand notation for the bin-by-bin hadronic cross section (partonic cross section for partons i, j) differential in $m_{\ell\ell}$ (in case of neutral-current (NC) Drell-Yan) or m_T (in case of charged-current (CC) Drell-Yan) and the partonic luminosities $\mathcal{L}_{ij}$ are defined as

$$\mathcal{L}_{ij}(\tau, m) = \int_\tau^1 \frac{dx}{x} \, f_i(x, m) f_j(\tau/x, m) \,, \tag{4.3.2}$$

where $m = m_{\ell\ell}$ in the NC case and $m = m_T$ in the CC case and are evaluated at NNLO. The QCD and EW K-factors are defined as

$$K_{\text{QCD}} = \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij}\Big|_{\text{NNLO QCD}} \right) \Big/ \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij}\Big|_{\text{NLO QCD}} \right) \,, \tag{4.3.3}$$

$$K_{\text{EW}} = \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij}\Big|_{\text{NLO QCD+EW}} \right) \Big/ \left(\mathcal{L}_{ij}^{\text{NNLO}} \otimes d\hat{\sigma}_{ij}\Big|_{\text{NLO QCD}} \right) \,, \tag{4.3.4}$$

The NNLO QCD K-factors have been computed using either `MATRIX` [267] or `FEWZ` [268] and cross-checked with the analytic computations of [269, 270]. The NLO EW K-factors have been evaluated with MadGraph5_aMC@NLO [265]. Eq. (4.3.4) accounts also for photon-initiated contributions (using the `NNPDF3.1QED` PDF set [271]) and final-state radiation effects, except when the latter has already been subtracted in the corresponding experimental analysis.

Fig. 4.3.1 displays a comparison between the CMS Drell-Yan distributions at 13 TeV and the corresponding theoretical predictions as a function of the dilepton invariant mass $m_{\ell\ell}$, separately for the dielectron and dimuon final states. The theory calculations are presented at NLO QCD, NNLO QCD, and NNLO QCD combined with NLO EW corrections, in all cases with `NNPDF3.1QED_nnlo_as_0118` as input PDF set, to illustrate the effect of the K-factors of Eq. (4.3.3) and (4.3.4). The CMS data are provided in terms of dressed leptons, and hence final state radiation (FSR) QED effects must be included in the electroweak corrections. Accounting for these effects is essential to improve the agreement between theory and data in the region below the Z-mass peak. NLO electroweak corrections are also important in the high-energy tail in $m_{\ell\ell}$,

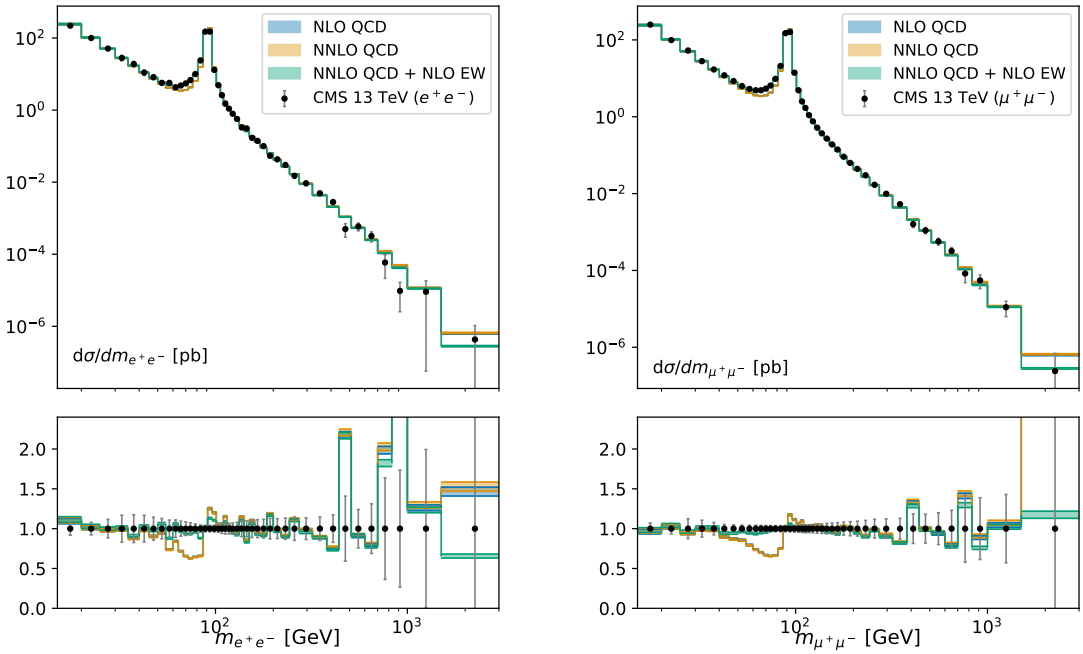


Fig. 4.3.1. Comparison of the CMS Drell-Yan 13 TeV data with the corresponding theoretical calculations at different perturbative orders as a function of the dilepton invariant mass $m_{\ell\ell}$ in the dielectron (left) and dimuon (right panel) final states. The bottom panels display the ratio of the theory calculations to the central value of the experimental data. We display the sum in quadrature of the experimental uncertainties, and the error band in the theory predictions correspond to the one-sigma PDF uncertainties.

where they are driven by the interplay between (negative) virtual EW effects and (positive) photon-initiated contributions.

A quantitative assessment of the agreement between theoretical predictions and experimental data for the high-mass DY datasets listed in Table 4.3.2 is presented in Table 4.3.3, which collects the values of the χ^2 per data point evaluated using the full information on correlated systematics provided by the experimental covariance matrix

$$\chi^2 = \frac{1}{n_{\rm dat}} \sum_{i,j=1}^{n_{\rm dat}} (D_i - T_i) \left(\text{cov}^{-1}\right)_{ij} (D_j - T_j), \tag{4.3.5}$$

where T_i are the theoretical predictions, D_i the central value of the experimental data and where the multiplicative uncertainties in the experimental covariance matrix (cov_{ij}) are treated as explained in [272, 273]. One can observe how in general the NNLO QCD

corrections are relatively small and that the NLO electroweak ones can be significant, especially for observables presented in terms of dressed leptons (such as the CMS 13 TeV ones) and are required to achieve a good description of the Drell-Yan data in the whole kinematical range available. Note that the input PDF sets used for these calculations include only a subset of these Drell-Yan measurements, in particular only the 7 TeV measurements, for which the data-theory agreement is comparable to the one observed in [215].

The data-theory agreement before including the 8 TeV and 13 TeV data in the PDF fit is generally good, once EW corrections are included, with the exception of the CMS 13 TeV data in the e^+e^- channel, for which the χ^2 per data point remains above 2. As can be observed in Fig. 4.3.1, the dielectron invariant mass distribution in this channel presents dips at about 500 GeV and 900 GeV which are not present in the $\mu^+\mu^-$ channel. These dips are the origin of this worse data-theory agreement, which is partially reduced once the dataset is included in the fit (see Sect. 4.3.3). We have verified that excluding this dataset from the fit does not change the results of the analysis, and therefore decided to keep it. Further experimental analysis based on the full Runs II and III datasets will tell whether the dips in the distributions in the electron invariant mass will stay.

SMEFT cross sections. In this work, we augment the SM calculations of the high-Q^2 DIS reduced cross sections discussed in Ref. [93] and the high-mass Drell-Yan cross sections listed in Table 4.3.2 with the effects of dimension-six SMEFT operators following the two benchmark scenarios presented in Sect. 4.2. These corrections are negligible for dilepton invariant masses of $m_{\ell\ell} \leq 200$ GeV and for DIS structure functions with $Q \leq (120)$ GeV, and hence there can safely adopt the SM calculations. In a similar manner as for higher-order QCD and EW corrections, we can define correction factors that encapsulate the linear and quadratic modifications induced by the dimension-six SMEFT operators. Adopting an operator normalisation such that

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_{n=1}^{n_{\rm op}} \frac{c_n}{v^2} \mathcal{O}_n \,, \tag{4.3.6}$$

with $n_{\rm op}$ indicating the number of operators that contribute to a given benchmark scenario and c_n being the (dimensionless) Wilson coefficient associated to $\mathcal{O}_n$, the linear EFT corrections can be parametrised as

$$R^{(n)}_{\rm SMEFT} \equiv \left(\mathcal{L}^{\rm NNLO}_{ij} \otimes d\widehat{\sigma}^{(n)}_{ij,\rm SMEFT}\right) \Big/ \left(\mathcal{L}^{\rm NNLO}_{ij} \otimes d\widehat{\sigma}_{ij,\rm SM}\right) \,, \quad n = 1 \ldots, n_{\rm op} \,, \tag{4.3.7}$$

Dataset	Final state	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$		
			NLO QCD	NNLO QCD	NNLO QCD + NLO EW
ATLAS 7 TeV	e^+e^-	13	1.45	1.77	1.73
ATLAS 8 TeV	$\ell^+\ell^-$	46	1.67	-	1.20
CMS 7 TeV	$\mu^+\mu^-$	127	3.40	1.27	1.54
CMS 8 TeV	$\ell^+\ell^-$	41	2.22	2.21	0.70
CMS 13 TeV	$\ell^+\ell^-$	43	18.7	19.7	1.91
CMS 13 TeV	e^+e^-	43	9.16	9.45	2.32
CMS 13 TeV	$\mu^+\mu^+$	43	15.7	15.8	0.81

Table 4.3.3. The values of the χ^2 per data point evaluated for the high-mass DY datasets listed in Table 4.3.2, using theoretical predictions computed at different perturbative accuracy. The PDF sets used here are `NNPDF31_nlo_as_0118`, `NNPDF31_nnlo_as_0118` and `NNPDF31_nnlo_as_0118_luxqed` for the NLO QCD, NNLO QCD and NNLO QCD + NLO EW predictions respectively. For CMS 13 TeV, where different final states are available, we indicate the χ^2 values for each of them. For the ATLAS 8 TeV data, we only evaluated the combined NNLO QCD + NLO EW correction, and hence the pure NNLO QCD result is not given.

with $\mathcal{L}_{ij}^{\rm NNLO}$ being the usual partonic luminosity evaluated at NNLO QCD, $d\widehat{\sigma}_{ij,\rm SM}$ the bin-by-bin partonic SM cross section, and $d\widehat{\sigma}^{(n)}_{ij,\rm SMEFT}$ the corresponding partonic cross section associated to the interference between $\mathcal{O}_n$ and the SM amplitude $\mathcal{A}_{\rm SM}$ when setting $c_n = 1$. Likewise, the ratio encapsulating the quadratic effects is defined as

$$R^{(n,m)}_{\rm SMEFT} \equiv \left(\mathcal{L}_{ij}^{\rm NNLO} \otimes d\widehat{\sigma}^{(n,m)}_{ij,\rm SMEFT}\right) \Big/ \left(\mathcal{L}_{ij}^{\rm NNLO} \otimes d\widehat{\sigma}_{ij,\rm SM}\right) , \quad n,m = 1\ldots, n_{\rm op} , \tag{4.3.8}$$

with the bin-by-bin partonic cross section $d\widehat{\sigma}^{(n,m)}_{ij,\rm SMEFT}$ now being evaluated from the squared amplitude $\mathcal{A}_n\mathcal{A}_m$ associated to the operators $\mathcal{O}_n$ and $\mathcal{O}_m$ when $c_n = c_m = 1$. The partonic cross sections in these ratios are computed at LO. In terms of Eqns. (4.3.7) and (4.3.8), we can define the EFT K-factors as

$$K_{\rm EFT} \;=\; 1 + \sum_{n=1}^{n_{\rm op}} c_n R^{(n)}_{\rm SMEFT} + \sum_{n,m=1}^{n_{\rm op}} c_n c_m R^{(n,m)}_{\rm SMEFT} , \tag{4.3.9}$$

which allow us to express a general Drell-Yan or DIS cross sections accounting for the dimension-six operators in Eq. (4.3.6) as

$$d\sigma_{\rm SMEFT} \;=\; d\sigma_{\rm SM} \times K_{\rm EFT} \tag{4.3.10}$$

where the $d\sigma_{\rm SM}$ is the state-of-the-art SM prediction including NNLO QCD and NLO EW corrections. In this approach, the SMEFT predictions inherit factorisable higher-order radiative correction [83, 9]. The SMEFT K-factors in Eq. (4.3.9) are precomputed before the fit using a reference SM PDF set and then kept fixed. The effect of varying the input NNLO PDF in Eqns. (4.3.7) and (4.3.8) is quantitatively assessed in Appendix F and it is found to be at the permil level in Scenario I and slightly more significant but still at most at the percent level in Scenario II. As a result, this effect will be neglected in the following. Further details about the implementation and validation of these K-factors can be found in Appendix F.

Fig. 4.3.2 illustrates the size of the EFT corrections in benchmark scenario I from Sect. 4.2 by comparing $(K_{\rm EFT}-1)$ with the relative experimental uncertainties for the ATLAS 7 TeV, CMS 8 TeV, and CMS 13 TeV Drell-Yan $m_{\ell\ell}$ distributions. We provide results for two representative points in the $(\hat{W}, \hat{Y})$ parameter space, namely $(\hat{W}, \hat{Y}) = (10^{-3}, 0)$ and $(0, 10^{-3})$. One can observe how for these values of $(\hat{W}, \hat{Y})$, and particularly for the ATLAS 8 TeV data, the SMEFT corrections to the Drell-Yan cross sections become comparable with the experimental uncertainties, increasing steadily with $m_{\ell\ell}$.

4.3.3 Methodology for the simultaneous PDF and EFT fits

Overview of the fitting methodology

Let us denote by $\mathbf{c} = (c_1, c_2, \ldots, c_{N_{\rm op}})$ the array containing the Wilson coefficients associated to the $N_{\rm op}$ dimension-six operators contributing to a given SMEFT scenario, where c_n are defined as in Eq. (4.3.6). We will perform a simultaneous EFT and PDF fit by scanning through an n-dimensional grid of values $\mathbf{c}_i$ in the EFT parameter space. For each point $\mathbf{c}_i$ in the scan, we evaluate the Drell-Yan and the DIS cross sections as described in Sect. 4.3.2. We then determine the $\chi^2(\mathbf{c})$ at each point in EFT space, where

$$\chi^2(\mathbf{c}) = \sum_{i,j=1}^{n_{dat}} (D_i - T_i(\mathbf{c}))(\mathrm{cov}^{-1})_{ij}(D_j - T_j(\mathbf{c})). \tag{4.3.11}$$

This procedure results in a sampling of the χ^2 values in the EFT parameter space, which we denote by $\chi^2(\mathbf{c}_i)$.

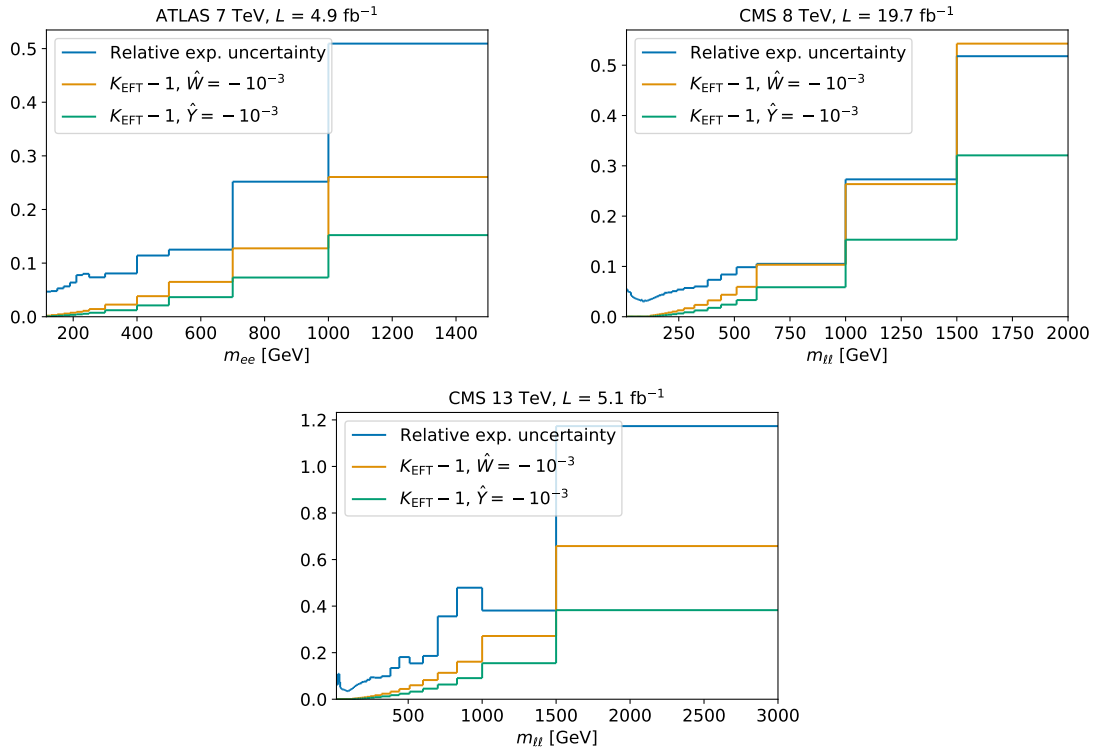


Fig. 4.3.2. Comparison between the (relative) experimental uncertainties and the corresponding EFT corrections, $K_{\rm EFT}(\hat{W},\hat{Y})-1$ in Eq. (4.3.9), for the ATLAS 7 TeV, CMS 8 TeV, and CMS 13 TeV Drell-Yan $m_{\ell\ell}$ distributions, for two representative values of $\hat{W}$ and $\hat{Y}$.

The calculation of the Drell-Yan and DIS cross sections depends on a choice of PDF set. Usually, SM PDFs are used for this purpose. These are PDFs which have been determined from a fit to data under the assumption of the SM. When sampling the χ^2, we use the same SM PDF at each value of $\mathbf{c}_i$, and we will denote the resulting χ^2 by $\chi^2_{\rm smp}(\mathbf{c}_i)$. In contrast, to perform a simultaneous fit of the EFT and PDF we make use of SMEFT PDFs. At each value of $\mathbf{c}_i$ in EFT space, we perform our theory calculations using a SMEFT PDF set which has been determined from a fit to data under the assumption of the SMEFT at that same value of $\mathbf{c}_i$. We will denote the resulting χ^2 by $\chi^2_{\rm eftp}(\mathbf{c}_i)$.

This implies that prior to obtaining the sampling $\chi^2_{\rm smp}(\mathbf{c}_i)$, we must first determine the SM PDFs using a single PDF fit under the assumption of the SM. To obtain $\chi^2_{\rm eftp}(\mathbf{c}_i)$ we must obtain a number of SMEFT PDFs, one for each value of $\mathbf{c}_i$ in the chosen grid, by performing a PDF fit at each point. We will begin by outlining the details and settings of these PDF fits, before elaborating on the details of the EFT fits.

PDF fit settings

Each PDF fit is performed using the standard `NNPDF3.1` [215] methodology, regardless of whether it is a SM PDF or a SMEFT PDF. This methodology evaluates the χ^2 given in Eq. 4.3.5 in the space of PDF parameters, determining the best-fit PDF from the minimum of the χ^2 (subject to cross-validation, to avoid overlearning). We note that this χ^2 keeps fully into account the experimental systematic correlations among all the measurements D_j included in the PDF analysis. We will begin by describing the fit settings used to assemble the PDF set that represents the baseline SM PDFs in this work. The same settings are then used to determine the SMEFT PDFs, with the caveat that we modify the theory predictions T_j in Eq. 4.3.5 at each point $\mathbf{c}_i$ for each SMEFT PDF produced.

The settings for this baseline SM PDF fit are the same as those used in the strangeness study of [12], itself a variant of NNPDF3.1 [215]. As described in Sect. 4.3.1, in this work we consider only DIS and Drell-Yan datasets, with the latter augmented as compared to [12] with the new high-mass measurements indicated in Table 4.3.2. A detailed comparison between this baseline SM PDF and the `NNPDF3.1_str` set from [12] is provided in Appendix D, while Table E.0.2 in App. E details the breakdown of the χ^2 for all the datasets that enter the fit. In general, the fit quality of the baseline SM PDF set is similar to that of the global fit of [12], although the description of the CMS 13 TeV invariant mass distribution in the combined electron and muon channels remains sub-optimal.

Having produced the SM PDFs, we may now study the impact of the high-mass DY datasets on the SM PDFs. Fig. 4.3.3 displays a comparison between the baseline SM PDF set, labelled "DIS+DY", with the same fit but without any datapoints from the high-mass DY datasets listed in Table 4.3.2, labelled "DIS+DY (no HM)". We show results at $Q = 100$ GeV both for the PDFs normalised to the central value of the baseline and for the relative PDF uncertainties. In the latter case we also display the PDF uncertainties from a corresponding DIS-only fit. The latter comparison shows that the DY cross sections significantly reduce the PDF uncertainties of the DIS-only fit. The addition of the high-mass DY data leads to a visible uncertainty reduction in the $0.005 \lesssim x \lesssim 0.3$ region as compared to the "DIS+DY(noHM)" reference as well an upwards shift of the up and down quarks and antiquark PDF. We therefore find that the available high-mass DY data can have an appreciable impact on the light quark and antiquark PDFs, despite the fact that in terms of Run II data our analysis is restricted to a single low-luminosity high-mass DY dataset. Yet more stringent constraints on the PDFs are expected from the measurements based on the full Runs II and III datasets,

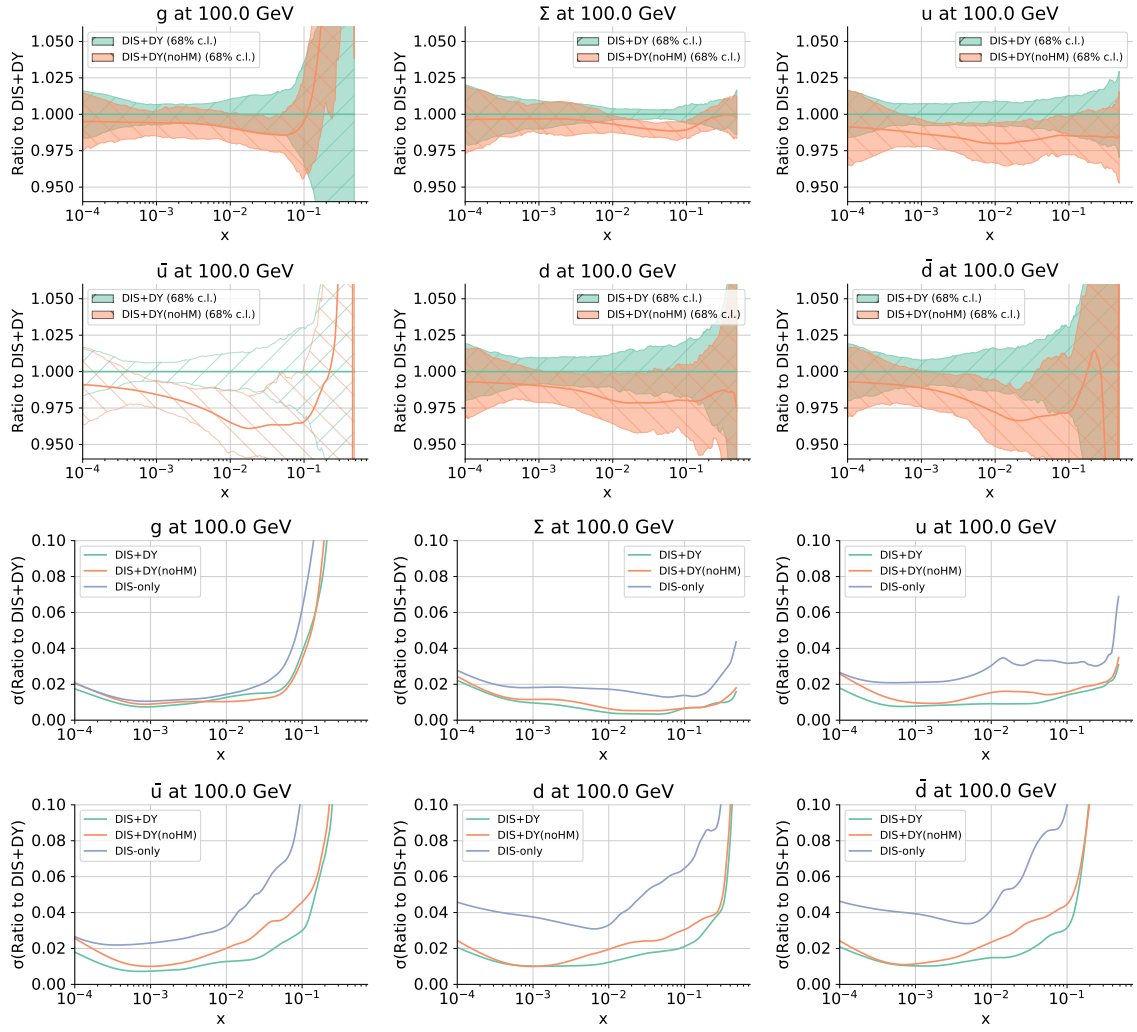


Fig. 4.3.3. Comparison between the baseline SM PDF set of this work, labelled "DIS+DY", with the corresponding fit without high-mass DY data. We show results at $Q = 100$ GeV for PDFs normalised to the central value of the baseline (upper) and for the relative PDF uncertainties (lower panels). In the latter case, we also display the PDF uncertainties from the DIS-only fit.

as well as from those to be provided by the HL-LHC [218]. We study the anticipated impact of the HL-LHC measurements in Sect. 4.5.

EFT fit settings

The strategy we follow to determine the constraints on the EFT parameters $\mathbf{c}$ follows the one adopted in our proof-of-concept DIS-only study [93], now extended to LHC processes. We produce two samplings of the χ^2 values in the EFT parameter space, one assuming SM PDFs ($\chi^2_{\rm smp}(\mathbf{c}_i)$) and one assmuming SMEFT PDFs ($\chi^2_{\rm eftp}(\mathbf{c}_i)$). The

comparison between the resulting bounds on the EFT coefficients obtained from $\chi^2_{\rm eftp}(\mathbf{c}_i)$ and from $\chi^2_{\rm smp}(\mathbf{c}_i)$ quantifies the relevance of producing consistent joint determinations of PDFs and Wilson coefficients when studying EFTs in high-energy tails.

Close enough to a local minimum $\chi^2_0 = \chi^2\left(\mathbf{c}^{(0)}\right)$ associated with best-fit values $\mathbf{c}^{(0)}$, the χ^2 as a function of the EFT coefficients can be approximated by a quadratic form

$$\chi^2_i \equiv \chi^2(\mathbf{c}_i) = \chi^2_0 + \sum_{n,m=1}^{N_{\rm op}} \left(c_{n,i} - c_n^{(0)}\right) H_{nm} \left(c_{m,i} - c_m^{(0)}\right) , \tag{4.3.12}$$

with H_{nm} being the usual Hessian matrix in the EFT parameter space. Restricting the EFT calculations to their linear, $\mathcal{O}\left(\Lambda^{-2}\right)$, contributions, Eq. (4.3.12) becomes exact in the case of $\chi^2_{\rm smp}(\mathbf{c}_i)$ (where cross sections are evaluated with SM PDFs). The reason is that in this case all dependence on the EFT coefficients is encoded in the partonic cross sections.

However, this is not true for $\chi^2_{\rm eftp}(\mathbf{c}_i)$, since now there will be a (non-linear) EFT back-reaction onto the PDFs and hence Eq. (4.3.12) is only valid up to higher orders in the EFT expansion, even if the EFT cross sections themselves are evaluated in the linear approximation. Eq. (4.3.12) can thus be only considered a reasonable approximation in the case that the SMEFT PDFs are not too different from their SM counterparts.

Hence, if we work with linear EFT calculations, provided the sampling in the EFT parameter space is sufficiently broad and fine-grained, and that the EFT-induced distortion on the PDFs is moderate, we can extract the parameters χ^2_0 and $\mathbf{c}^{(0)}$ and the Hessian matrix H using least-squares regression from Eq. (4.3.12), using $\chi^2_{\rm smp}$ for the SM PDFs and $\chi^2_{\rm eftp}$ for the SMEFT PDFs. The associated confidence level contours are determined by imposing

$$\Delta\chi^2(\mathbf{c}) \equiv \chi^2_i(\mathbf{c}) - \chi^2_0 = \sum_{n,m=1}^{N_{\rm op}} \left(c_n - c_n^{(0)}\right) H_{nm} \left(c_m - c_m^{(0)}\right) = \text{constant} , \tag{4.3.13}$$

where this constant depends on the number of degrees of freedom. For linear EFT two-parameter fits, such as those for benchmark scenario I in the context of HL-LHC projections, imposing Eq. (4.3.13) leads to elliptic contours in the $(\hat{W}, \hat{Y})$ plane.

In the case of fits to χ^2 profiles obtained from EFT calculations which include both linear, $\mathcal{O}\left(\Lambda^{-2}\right)$, and quadratic, $\mathcal{O}\left(\Lambda^{-4}\right)$, contributions, such as those arising in benchmark scenario II, rather than working in the Hessian approximation we instead

carry out a (one-dimensional) quartic fit of the form

$$\chi^2(c) = \sum_{k=0}^{k_{\max}=4} a_k (c)^k \tag{4.3.14}$$

with the χ^2 values being $\chi^2_{\rm smp}$ ($\chi^2_{\rm eftp}$) for the SM (SMEFT) PDFs, and then determine confidence level intervals by imposing $\Delta\chi^2(c) = \chi^2_i(c) - \chi^2_0 =$ constant. We determine this constant numerically by finding the likelihood contour, $\mathcal{L} \propto \exp(-\chi^2/2)$, containing 95% of the total probability (for the 95% CL intervals).

To conclude this section, we give details on how we account for sources of uncertainty in our EFT constraints. We account for PDF uncertainties and the statistical uncertainty associated to the finite replica sample of the NNPDF Monte Carlo sets that we use here.

PDF uncertainty. In Sects. 4.4.1, 4.4.2 and 4.5.3 we will present bounds on the EFT parameters using the SM PDFs with and without the PDF uncertainties being accounted for. In order to estimate these, we follow the procedure detailed above to determine the confidence level intervals for the EFT parameters but now using the kth Monte Carlo replica of the PDF set, rather than the central replica $k = 0$ as done when PDF uncertainties are neglected. One ends up with $N_{\rm rep}$ values of the upper and lower bounds:

$$\left[\mathbf{c}^{(k)}_{\rm min}, \mathbf{c}^{(k)}_{\rm max}\right] , \qquad k = 1, \ldots, N_{\rm rep} , \tag{4.3.15}$$

and then the outermost bounds in the 68% envelope are considered to be the bounds on the EFT parameters $\mathbf{c}$, now including the 1σ-PDF uncertainty. This is very important to account for, given that in the case of the bounds determined using $\chi^2_{\rm eftp}$, the PDF uncertainty is already included by construction, given that the Wilson coefficients are determined from the global set of PDFs, exactly as in the case of the α_s determination from a global set of PDFs of [274, 275]. A more sophisticated way to extract parameters such as α_s of the Wilson coefficients from a global fit of PDFs, that includes the correlations between these parameters and the PDFs, is given by the correlated replica method proposed in the more recent α_s determination in [276]. The latter would allow better accounting of the correlations between Wilson coefficients and PDFs. However we do not use it here due to the fact that these correlations of the PDFs with the Wilson coefficients are much smaller than those with the strong coupling constant and due to its large computational cost. We endeavour to address this issue in future work.

Methodological uncertainty. In a simultaneous fit of PDFs and EFT coefficients, for each set of Wilson coefficients $\mathbf{c}_i$ one has a PDF fit composed of $N_{\rm rep}$ Monte Carlo replicas. The major methodological uncertainty is associated to finite-$N_{\rm rep}$ effects can be estimated by bootstrapping across the replicas, as explained in the $\alpha_s(m_Z)$ extraction of [276]. Specifically, for each value of $\mathbf{c}_i$ we perform $N_{\rm res}$ re-samples of all $N_{\rm rep}$ replicas with replacement, and compute the theory predictions:

$$\mathbf{T}^{(\rm res)}_{i,lk} \,, \quad \begin{array}{l} l = 1, \ldots, N_{\rm res} \\ k = 1, \ldots, N_{\rm rep} \end{array} \,, \tag{4.3.16}$$

such that there are $N_{\rm res}$ re-samples each composed of an $N_{\rm rep}$-sized array of theory predictions. Since this re-sampling is done with replacement, it differs from the original sample in that it contains duplicates and missing values. The average theory prediction is then obtained for each of these bootstrapped sets:

$$\overline{\mathbf{T}}_{i,l} = \left\langle \mathbf{T}^{(\rm res)}_{i,lk} \right\rangle_{\rm rep} \,, \quad l = 1, \ldots, N_{\rm res} \,. \tag{4.3.17}$$

These bootstrapped theory predictions $\overline{\mathbf{T}}_{i,l}$ are used to evaluate the χ^2 to data, with the finite-size uncertainty given by the standard deviation across each bootstrap re-sample:

$$\sigma_{\chi^2_i} = \text{std}\left(\chi^2_{i,l}\right)\Big|_{\rm res} \,. \tag{4.3.18}$$

A value of $N_{\rm res} \simeq 10^4$ re-samples is found to be sufficient to achieve stable results for the estimate of the finite-size uncertainties defined by Eq. (4.3.18).

4.4 Results

In this section, we start by presenting results for the SMEFT PDFs extracted from DIS and Drell-Yan data in benchmark scenario I. We compare them with their SM counterparts at the level of partonic luminosities and assess how the bounds obtained on the $\hat{W}$ and $\hat{Y}$ parameters in this simultaneous SMEFT and PDF fit compare to those based on assuming SM PDFs. We then investigate the sensitivity of available high-mass Drell-Yan data to benchmark scenario II, where only the dimuon final state is modified by EFT effects. Finally, we quantify the impact that the consistent use of SMEFT PDFs has on the reinterpretation of high-mass dilepton BSM searches.

4.4.1 PDF and EFT interplay in current high-mass Drell-Yan data

By deploying the methodology described in §4.3.3, we have extended the PDF analysis based on SM predictions presented in §4.3.3 to account for the effects of non-zero EFT coefficients within benchmark scenario I defined in §4.2.1. Here, we present results for one-dimensional fits where only one of the $\hat{W}$ or the $\hat{Y}$ parameter is allowed to be non-zero. The reason for this choice is that, in a fit including only high-mass neutral-current Drell-Yan processes, there exists a flat direction when $\hat{W}$ and $\hat{Y}$ are varied simultaneously, since both operators scale as q^4 and thus cannot both be constrained by a single 1D distribution. This degeneracy can only be lifted once high-mass charged-current DY data is included in the fit. As we demonstrate in §4.5, thanks to the HL-LHC it will be possible to carry out a simultaneous fit of the PDFs and the two EFT parameters $(\hat{W}, \hat{Y})$.

Taking into account the existing bounds reported in §4.2.1, as well as the sensitivity of available high-mass Drell-Yan data to the EFT coefficients illustrated by Fig. 4.3.2, here we have adopted the following sampling ranges for the $\hat{W}$ and $\hat{Y}$ parameters:

$$\left(\hat{W} \times 10^4\right) \in [-22, 14] \ , \qquad \left(\hat{Y} \times 10^4\right) \in [-20, 20] \ . \tag{4.4.1}$$

We used 21 sampling values of $\hat{Y}_i$ equally spaced in this interval, hence in steps of $\Delta\hat{Y} = 2 \times 10^{-4}$. In the case of $\hat{W}_i$ it was found convenient to instead use 15 points equally spaced between -14×10^{-4} and 14×10^{-4} in steps of $\Delta\hat{W} = 2 \times 10^{-4}$, and then to add two more values at $\hat{W}_i = -18 \times 10^{-4}$ and -22×10^{-4}.

Fig. 4.4.1 displays the obtained values of $\Delta\chi^2$, Eq. (4.3.13), as a function of $\hat{W}_i$ and $\hat{Y}_i$ in the case of the SMEFT PDFs. That is, using the values of $\chi^2_{\rm eftp}(\mathbf{c}_i)$. These χ^2 values are evaluated as a sum over those datasets from Table 4.3.1 and 4.3.2 that receive non-zero EFT corrections, namely the DIS datasets that have a reach in Q^2 above $(120)^2$ GeV^2 (namely HERA and NMC), and the ATLAS and CMS high-mass Drell-Yan measurements in Table 4.3.2. The use of such a partial χ^2 rather than the global χ^2 is a necessary approximation due to the limitation of our current methodology. The statistical fluctuations of the global χ^2 are significantly larger than those of the partial χ^2 and can only be tamed by running a very large batch of replicas for each benchmark point in $\hat{W}$ and $\hat{Y}$ and by increasing the density of benchmark points in the region that is explored, as it was done for the scan of α_s in Ref. [276]. However, while for the scan of α_s all processes contribute to the parabolic behaviour of the $\Delta\chi^2$, in this case the dominant contributions to $\chi^2_{\rm eftp}$ come by far from the SMEFT corrections

to the hard cross section of these processes, and from the changes in the PDFs induced by non-zero Wilson coefficients. The latter changes in PDFs are confined to the large-x light quark and antiquark distributions, which affect the high-mass Drell-Yan data. The analysis of the $\chi^2_{\rm eftp}$ computed on the subset of data captures the dominant effects, while minimising the level of statistical fluctuations. A further approximation is given by the fact that only linear EFT effects are included in the calculation of the DIS and DY cross sections, while the (subleading) quadratic corrections are neglected in this scenario. The error bars in the $\Delta\chi^2_i$ points of Fig. 4.4.1 indicate the methodological finite-size uncertainties evaluated with the bootstrapping method described in §4.3.3 and the horizontal line corresponds to the $\Delta\chi^2 = 4$ condition associated to a 95% CL interval. We also show in Fig. 4.4.1 the results of the associated parabolic fits,

$$\Delta\chi^2(\hat{W}) = \left(\hat{W} - \hat{W}^{(0)}\right)^2 / \left(\delta\hat{W}\right)^2 , \tag{4.4.2}$$

and likewise for $\Delta\chi^2(\hat{Y})$. From the results in Fig. 4.4.1, one observes that both the $\hat{W}$ and $\hat{Y}$ parameters agree with the SM expectation within uncertainties.

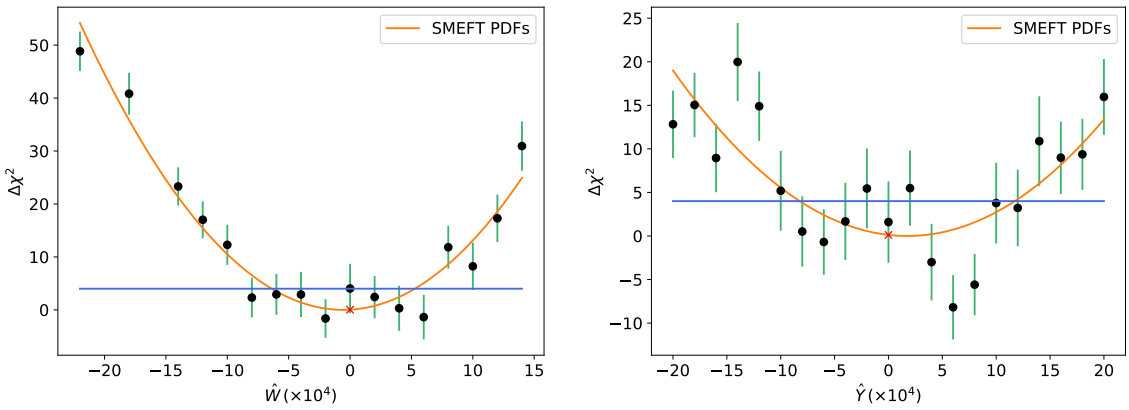


Fig. 4.4.1. The values of $\Delta\chi^2$, Eq. (4.3.13), obtained for the SMEFT PDFs (thus using the $\chi^2_{\rm eftp}(\mathbf{c}_i)$ values) as a function of $\hat{W}_i$ (left) and $\hat{Y}_i$ (right panel) in the sampling ranges of Eq. (4.4.1) together with the corresponding parabolic fits. The error bars indicate the finite-size uncertainties and the horizontal line corresponds to the $\Delta\chi^2 = 4$ condition defining the 95% CL intervals. The red cross indicates the SM expectation, $\hat{W} = \hat{Y} = 0$.

Fig. 4.4.2 then compares the results of the parabolic fits based on the SMEFT PDFs as displayed in Fig. 4.4.1 with their counterparts obtained in the case of the SM PDFs. That is, in the latter case one carries out parabolic fits to the $\chi^2_{\rm smp}$ values, as is customary in the literature for the EFT analyses. The insets highlight the region close to $\Delta\chi^2 \simeq 0$. For the $\hat{W}$ parameter, the consistent use of SMEFT PDFs leaves

the best-fit value essentially unchanged but increases the coefficient uncertainty $\delta\hat{W}$, leading to a broader parabola. Similar observations can be derived for the $\hat{Y}$ parameter, though here one also finds a upwards shift in the best-fit values by $\Delta\hat{Y} \simeq 2 \times 10^{-4}$ in addition to a parabola broadening, when SMEFT PDFs are consistently used. We note that the SM PDF parabolas in Fig. 4.4.2 are evaluated using the central PDF replica and hence do not account for PDF uncertainties.

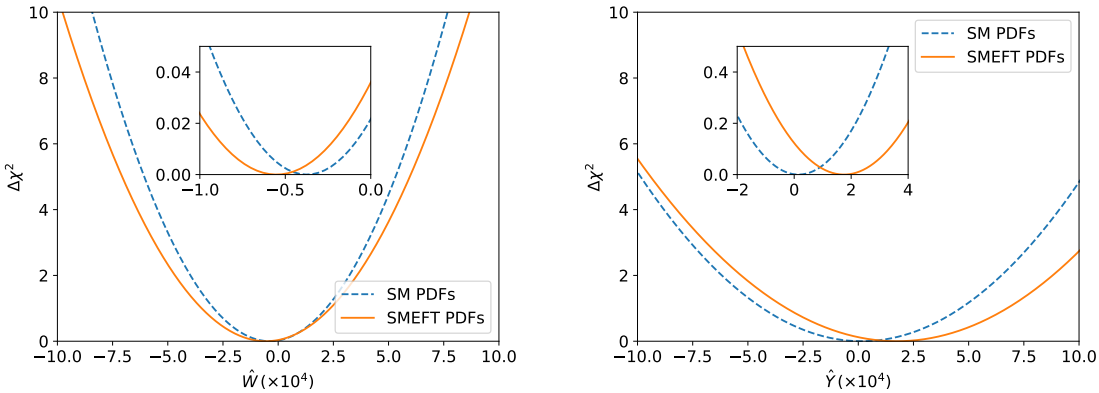


Fig. 4.4.2. Comparison between the results of the parabolic fits to $\Delta\chi^2$, Eq. (4.4.2), for the $\hat{W}$ (left) and $\hat{Y}$ (right panel) parameters for either the SMEFT PDFs ($\chi^2_{\rm eftp}$, already displayed in Fig. 4.4.1) or the SM PDFs (hence with $\chi^2_{\rm smp}$). The insets zoom on the region close to $\Delta\chi^2 \simeq 0$.

Table 4.4.1 summarises the 68% and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic $\Delta\chi^2$ fits using either the SM or the SMEFT PDFs shown in Fig. 4.4.2. The fourth and fifth column indicate the absolute shift in best-fit values and the percentage broadening of the fit parameter uncertainties when the SMEFT PDFs are consistently used instead of the SM PDFs (either without or with PDF uncertainties):

$$\text{best fit shift} \equiv \left(\hat{W}^{(0)}\Big|_{\rm SMEFT\,PDF} - \hat{W}^{(0)}\Big|_{\rm SM\,PDF}\right) , \tag{4.4.3}$$

$$\text{broadening} \equiv \left(\delta\hat{W}^{(0)}\Big|_{\rm SMEFT\,PDF} - \delta\hat{W}^{(0)}\Big|_{\rm SM\,PDF}\right) \Big/ \delta\hat{W}^{(0)}\Big|_{\rm SM\,PDF} , \tag{4.4.4}$$

and likewise for the $\hat{Y}$ parameter.

In the specific case of the SM PDF results, Table 4.4.1 indicates the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; recall that the SMEFT PDF bounds already include PDF uncertainties by construction (see §4.3.3).

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^4$ (68% CL)	$[-3.0, 2.2]$ $[-4.3, 3.8]$	$[-3.5, 2.4]$	-0.2 -0.3	$+13\%$ -27%
$\hat{W} \times 10^4$ (95% CL)	$[-5.5, 4.7]$ $[-6.8, 6.3]$	$[-6.4, 5.3]$	-0.2 -0.3	$+15\%$ -11%
$\hat{Y} \times 10^4$ (68% CL)	$[-4.4, 4.7]$ $[-6.7, 7.5]$	$[-3.4, 6.9]$	$+1.6$ $+1.4$	$+13\%$ -27%
$\hat{Y} \times 10^4$ (95% CL)	$[-8.8, 9.2]$ $[-11.1, 12.0]$	$[-8.3, 11.8]$	$+1.6$ $+1.3$	$+12\%$ -13%

Table 4.4.1. The 68% CL and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the $\Delta\chi^2$ values calculated from either the SM or the the SMEFT PDFs. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction, while the methodological (finite-size) uncertainty is included according to the approached described in §4.3.3. The fourth and fifth column indicate the absolute shift in best-fit values, Eq. (4.4.3) and the percentage broadening of the EFT parameter uncertainties, Eq. (4.4.4), when the SMEFT PDFs are consistently used instead of the SM PDFs.

The methodological (finite-size) uncertainty is included according to the approach described in §4.3.3 and it amounts to $4.7 \cdot 10^{-5}$ in the case of $\hat{W}$ and $1.0 \cdot 10^{-4}$ in the case of $\hat{Y}$, corresponding to 4% and 5% respectively of the 95% C.L. bounds for the $\hat{W}$ and $\hat{Y}$ coefficients.

By comparing the bounds obtained when PDF uncertainties are accounted for to those neglecting PDF uncertainty, one observes a systematic broadening of the bounds from both the lower and upper limits, as was also reported in [93].

When PDF uncertainties are neglected (accounted for) when using the SM PDFs to constrain the EFT parameters, the consistent use of the SMEFT PDFs leads to both a shift in the best-fit values of magnitude $\Delta\hat{W} = -2 \times 10^{-5}$ and $\Delta\hat{Y} = +1.6 \times 10^{-4}$ as well as to an increase (decrease) of the fit parameter uncertainties, with $\delta\hat{W}$ and $\delta\hat{Y}$ growing by 15% and 12% (decreasing by 11% and 13%) respectively. This result shows that, given available Drell-Yan data and once PDF uncertainties are accounted for, the bounds on the EFT parameters are actually *improved* once SMEFT PDFs are adopted.

All in all, the effect of the consistent treatment of the SMEFT PDFs in the interpretation of high-mass DY cross sections is moderate but not negligible, either loosening or tightening up the obtained bounds on the EFT parameters (depending on whether or not PDF uncertainties are accounted for to begin with) by up to 15% and, in the case of $\hat{Y}$ parameter, shifting its central value by one-third of the 68% CL parameter uncertainty. Such a relatively moderate effect can be partly understood from the limited availability of high-mass DY measurements for EFT interpretations, with a single dataset at 13 TeV, and even in this case, with it being restricted to a small fraction of the Run II luminosity. As we will demonstrate in §4.5, the impact of SMEFT PDFs becomes much more significant once higher-statistics measurements of the NC and CC Drell-Yan tails become available at the HL-LHC, loosening the bounds on $\hat{W}$ and $\hat{Y}$ by up to a factor 5.

Comparing the limits on the $\hat{W}$ and $\hat{Y}$ parameters from Table 4.4.1 with those of Ref. [8] and reported in §4.2, we observe that our bounds are more stringent. There are two main reasons that could explain this difference. On the one hand, on top of the ATLAS and CMS high-mass DY cross sections at 8 TeV, we also include the corresponding 7 and 13 TeV data that provide additional weight to the high invariant mass region of the spectrum in the fit. On the other hand, in our analysis we fit the whole invariant mass spectrum and do not cut away the low $m_{\ell\ell}$ region below 120 GeV, thus we do not ignore the correlations between the low and high ends of the spectrum which are important even if the former is not affected by SMEFT corrections.

We now move to assess how the SMEFT PDFs relate to their SM counterparts, and determine the extent to which it is possible to reabsorb EFT effects into the PDFs. Fig. 4.4.3 displays a comparison between the SM and the SMEFT PDF luminosities for representative values of the $\hat{W}$ (upper) and $\hat{Y}$ (lower panel) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in Table 4.4.1. The error band in the SM PDFs corresponds to the 68% CL PDF uncertainty, while for the SMEFT PDFs only the central values are shown.

In all cases, one finds that the EFT-induced shifts on the luminosities are smaller than their standard deviation. The biggest differences, relative to uncertainties, are observed in the quark-antiquark luminosities for $m_X \gtrsim 500$ GeV. This finding can be understood from the fact that the NC Drell-Yan cross section is proportional to the $u\bar{u}$ and $d\bar{d}$ combinations at leading order, but the up and down quark PDFs are already well constrained by lower-energy DIS measurements. Furthermore, we have verified that the size PDF uncertainties is unchanged in the SMEFT fits. The results of Fig. 4.4.3 are consistent with those of Table 4.4.1 and demonstrate that, with current

data, the interplay between EFT effects and PDFs in the high-mass Drell-Yan tails is appreciable but remains subdominant as compared to other sources of uncertainty.

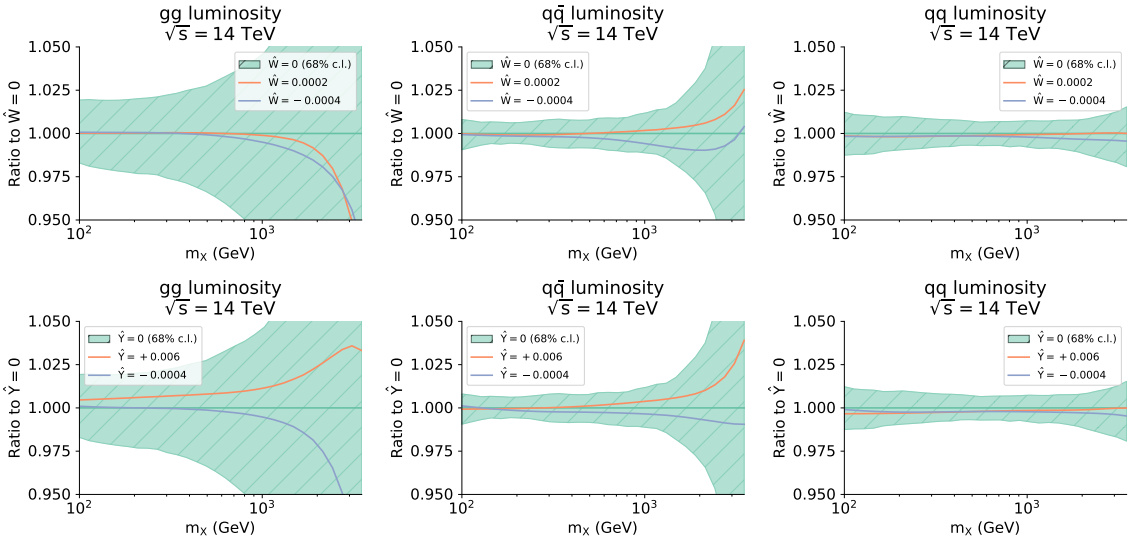


Fig. 4.4.3. Comparison between the SM PDF luminosities with their SMEFT counterparts, displayed as ratios to the central value of the SM luminosities, for representative values of the $\hat{W}$ (upper) and $\hat{Y}$ (lower panel) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 68% CL intervals reported in Table 4.4.1.

One important question in this context concerns how one could disentangle the EFT-induced shifts in the PDF luminosities displayed in Fig. 4.4.3 (see also the corresponding PDF comparisons in Fig. G.0.1) from other possible sources of deviations, such as internal inconsistencies in some datasets or missing higher orders in the SM calculations. An attractive strategy in this respect is based on exploiting the energy-growing effects associated to the higher-dimensional EFT operators, which translate into an enhanced sensitivity to the $\hat{W}$ and $\hat{Y}$ parameters for large values of the dilepton invariant mass $m_{\ell\ell}$. To this purpose, it is useful to define the following ratio:

$$R_{\chi^2}\left(m_{\ell\ell}^{(\max)}, \hat{W}, \hat{Y}\right) \equiv \frac{\chi^2\left(m_{\ell\ell}^{(\max)}, \hat{W}, \hat{Y}\right)}{\chi^2\left(m_{\ell\ell}^{(\max)} = 120 \text{ GeV}, \hat{W}, \hat{Y}\right)}, \tag{4.4.5}$$

where $m_{\ell\ell}^{(\max)}$ is the upper bound on the value of the dilepton invariant mass bins that enter the χ^2 calculation. In Eq. (4.4.5), both the numerator and the denominator are evaluated using either $\chi^2_{\rm smp}$ (for the SM PDFs) or $\chi^2_{\rm eftp}$ (for the SMEFT PDFs), and

the denominator corresponds to the χ^2 value (per data point) in the kinematic region for which EFT effects are negligible.[1]

The R_{χ^2} estimator defined in Eq. (4.4.5) allows for the isolation of the contribution to the total χ^2 that arises from the high-$m_{\ell\ell}$ bins that dominate the overall sensitivity to the $\hat{W}$ and $\hat{Y}$ parameters. For small values of $m_{\ell\ell}^{(\max)}$, say 200 GeV, one is cutting away all $m_{\ell\ell}$ bins with EFT sensitivity and hence one expects $R_{\chi^2} \simeq 1$. As $m_{\ell\ell}^{(\max)}$ is increased, the χ^2 will include the contributions from the $m_{\ell\ell}$ bins more sensitive to EFT effects, and thus one expects to find a large deviation with respect to the $R_{\chi^2} \simeq 1$ reference value. Furthermore, EFT effects should induce an approximately monotonic growth of R_{χ^2} with $m_{\ell\ell}^{(\max)}$, which would instead be absent from other possible sources of PDF distortion and thus represent a smoking gun for BSM physics in the high-energy DY tails.

These expectations are verified in Fig. 4.4.4, which displays the R_{χ^2} estimator (normalised to its SM value) as a function for $m_{\ell\ell}^{(\max)}$ for representative values of the $\hat{W}$ and $\hat{Y}$ parameters both for the SM and the SMEFT PDFs, where the horizontal line indicates its reference SM value. Indeed we observe an approximately monotonic growth of R_{χ^2} arising from the energy-growing effects in the EFT. Due to the limited experimental information the binning in $m_{\ell\ell}$ is rather coarse, explaining the observed fluctuations. In the specific case of the $\hat{W}$ parameter, the SMEFT PDF curve lies slightly below the SM PDF one, highlighting how EFT effects are being partially (but not completely) reabsorbed into the PDFs.

4.4.2 EFT constraints on scenario II from current high-mass Drell-Yan data

In contrast to benchmark scenario I, which is flavour universal, the second SMEFT scenario to be explored in this work and described in §4.2.2 contains a four-fermion interaction involving muons but not electrons, which therefore modifies the rates of the dilepton process $pp \to \mu^+\mu^-$ but not those of $pp \to e^+e^-$. This property implies that, without introducing further assumptions, the Wilson coefficient $\mathbf{C}_{33}^{Du}$ can be only constrained from DY measurements carried out in the dimuon (rather than in the dielectron or in the combined) final state. As indicated in Table 4.3.2, only the CMS data at 7 TeV and 13 TeV include DY distributions in the dimuon final state.

[1] Note that Eq. (4.4.5) is computed *a posteriori* using existing fits, and that the kinematical cut in $m_{\ell\ell}^{(\max)}$ is absent from the actual fits and it is only evaluated as a diagnosis tool.

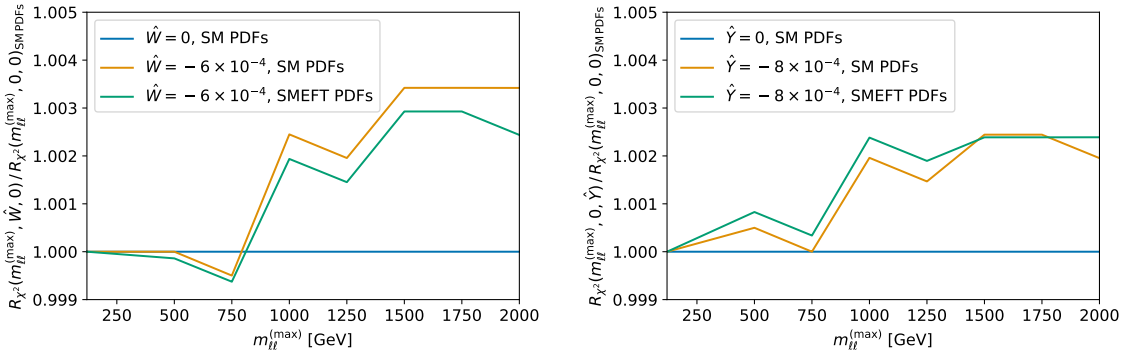


Fig. 4.4.4. The R_{χ^2} estimator, Eq. (4.4.5), normalised to its SM value, as a function for $m_{\ell\ell}^{(\max)}$ for representative values of $\hat{W}$ (left) and $\hat{Y}$ (right panel). We display the results obtained both with SM and SMEFT PDFs, with the horizontal line indicating the reference SM value of R_{χ^2}.

Due to these restrictions in the input dataset, the interplay between PDFs and SMEFT effects is expected to be milder as compared to the results presented in §4.4.1. For this reason, here we do not attempt to perform a joint determination of the PDFs and the $\mathbf{C}_{33}^{Dμ}$ coefficient, but rather restrict ourselves to quantifying the information that available DY data in the dimuon final state provide on this operator. We instead present a simultaneous determination including projections for the HL-LHC in §4.5.

Fig. 4.4.5 displays the results of three quartic fits to the $\chi^2\left(\mathbf{C}_{33}^{Dμ}\right)$ profile in benchmark scenario II, based on Eq. (4.3.14), where here $\chi^2_{\rm smp}$ includes only the contributions from the two available DY measurements in the dimuon final state. We present fits based on cross sections that account only for the linear, only for the quadratic, and for both the linear and quadratic terms in the EFT expansion. In all cases, these cross sections are computed using the baseline SM PDF set. The inset displays the outcome of the linear EFT fit with an enlarged x-axis range.

The results of Fig. 4.4.5 indicate that $\mathbf{C}_{33}^{Dμ}$ is essentially unconstrained at the linear EFT level, and only once quadratic corrections $\mathcal{O}(\Lambda^{-4})$ are accounted for is one able to obtain reasonable bounds on this coefficient. The reason for this behaviour is that for this operator the interference with the SM amplitude is suppressed, and hence the leading EFT effects arise at the quadratic level from the square of the EFT amplitude, thus being proportional to $\left(\mathbf{C}_{33}^{Dμ}\right)^2$ [83]. In the case of the polynomial fit to the χ^2 profile evaluated on the full quadratic EFT cross sections, we find the following 95%

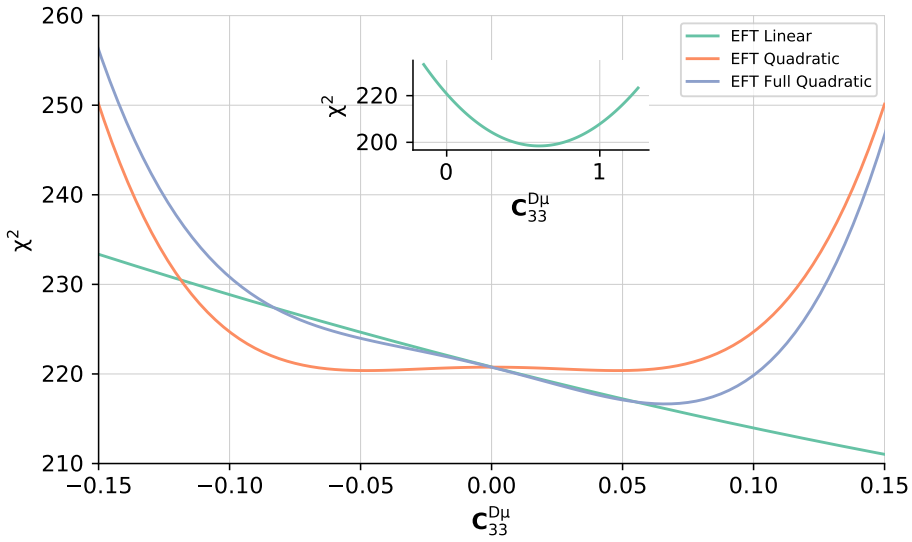


Fig. 4.4.5. The results of polynomial fits to $\chi^2\left(\mathbf{C}_{33}^{D\mu}\right)$, Eq. (4.4.2), in scenario II. This χ^2 includes only the contributions from the two DY measurements in the dimuon final state. We display results for fits based on cross sections that account only for the linear, only for the quadratic, and for both linear and quadratic terms in the EFT expansion, in all cases using the baseline SM PDF set. The inset displays the fit to the linear EFT values with an enlarged x-axis range.

CL limits on this Wilson coefficient:

$$\left(\mathbf{C}_{33}^{D\mu} \times 10^2\right) \in [-1.2, 10.7] \ , \tag{4.4.6}$$

which can be compared with the bounds on the same operator obtained in [83] from recasting the ATLAS dilepton search data of [232], given by Eq. (4.2.12). The fact that our bound in Eq. (4.4.6) is around a factor three looser than in Eq. (4.2.12) is explained because the dilepton search data from [232] benefits from an extended coverage in $m_{\ell\ell}$ as compared to the available unfolded DY cross sections. The same result, this time for the $\hat{W}$ and $\hat{Y}$ parameters, will be obtained in the next section where we assess the impact of the SMEFT PDFs in the EFT interpretation of the ATLAS dilepton search dataset.

4.4.3 On the EFT interpretation of high-mass dilepton searches

As mentioned above, a single high-mass DY cross section measurement is available at 13 TeV, and even in this case it is only based on a small subset of the Run II luminosity. As a consequence, the highest energy bin of this dataset is rather wide, $m_{\ell\ell} \in [1.5, 3.0]$

TeV. This implies limited sensitivity to deviations in the tails of DY distributions, for which using a large number of narrow bins is most beneficial to constrain heavy resonances, for instance. Here we would like to quantify the interplay between PDF and EFT effects at the level of a recent ATLAS 13 TeV search for Z' bosons in the dilepton channel [7] based on the complete Run II luminosity of $\mathcal{L} = 139$ fb^{-1}. Since these are detector-level measurements, which cannot therefore be included in a PDF analysis, our aim is to use the SMEFT PDFs to investigate how the bounds on BSM physics are modified as compared to the standard approach based on computing theory predictions using SM PDFs.

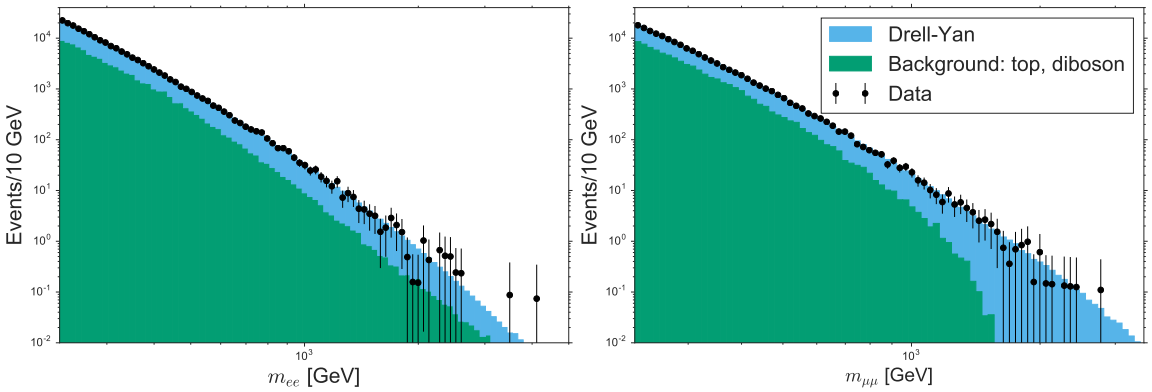


Fig. 4.4.6. The data (number of events per 10 GeV bin) from the ATLAS Z' search from [7] in the di-electron (left) and di-muon (right) channels. We also display the theoretical predictions associated to the contributions from Drell-Yan and from the rest of the backgrounds, taken from the ATLAS publication.

The ATLAS data, displayed in Fig. 4.4.6, consist of event counts in 100 dilepton invariant mass, $m_{\ell\ell}$, bins in both the dimuon and dielectron channels in the range $m_{\ell\ell} \in (225, 6000)$ GeV. We take this data from `HEPdata` [277] and denote the event count in the i^{th} bin by n_i. The narrow binning and broad $m_{\ell\ell}$ coverage allowed ATLAS to constrain Z' masses to $M_{Z'} \gtrsim 4$ TeV. This is a much higher reach than the DY cross section measurements used for the SMEFT PDF fits in the previous subsections, and should therefore provide stronger constraints on the EFT benchmark scenarios described in §4.2. By including this search data in our study, we can investigate whether such strong constraints are sensitive to the EFT-induced modifications in the PDF luminosities highlighted in Fig. 4.4.3.

In order to constrain the $\hat{W}$ and $\hat{Y}$ parameters in benchmark scenario I from the ATLAS dilepton search data, for each bin we compute a theory prediction $y_i = y_i(\hat{W}, \hat{Y})$ given by the sum of background b_i (top, diboson) and signal $s_i(\hat{W}, \hat{Y})$ (Drell-Yan) components. The ATLAS search provides an estimate of the total SM contribution

(sum of top, diboson, and DY) without a breakdown into components. This estimate is provided as a continuous function of $m_{\ell\ell}$. We thus estimate our background (top and diboson) by subtracting our own DY simulation from the estimated total SM event counts found by evaluating this function at each bin centre. We compute the DY signal in each bin as

$$s_i(\hat{W},\hat{Y}) = s_{i,\mathrm{SM}} \times K(\hat{W},\hat{Y})\,, \tag{4.4.7}$$

where $s_{i,\mathrm{SM}}$ indicates the detector-level prediction for the i^{th} bin of the $m_{\ell\ell}$ distribution evaluated at NLO QCD using MadGraph5_aMC@NLO [278], `Pythia` [279] and `Delphes` [280] and using the `NNPDF31_nnlo_as_0118` set as PDF input set. In Eq. (4.4.7), $K(\hat{W},\hat{Y})$ is the K-factor calculated as the ratio of cross sections in each bin, accounting for the impact of non-zero EFT corrections $\hat{W},\hat{Y} \neq 0$ both in the partonic cross section and in the PDFs

$$K_{\mathrm{eftp}}(\hat{W},\hat{Y}) \equiv \frac{\sum_q \int d\tau \mathcal{L}^{\mathrm{SMEFT}}_{q\bar{q}}(\tau,\mu_F,\hat{W},\hat{Y})\hat{\sigma}(\tau s_0,\hat{W},\hat{Y})}{\sum_q \int d\tau \mathcal{L}^{\mathrm{SM}}_{q\bar{q}}(\tau,\mu_F)\hat{\sigma}(\tau s_0,0,0)}, \tag{4.4.8}$$

where the integration in τ goes from τ_{min} to τ_{max} in each bin. For comparison, we will also present results where the K-factor is instead evaluated as usual in terms of the SM PDFs,

$$K_{\mathrm{smp}}(\hat{W},\hat{Y}) \equiv \frac{\sum_q \int d\tau \mathcal{L}^{\mathrm{SM}}_{q\bar{q}}(\tau,\mu_F)\hat{\sigma}(\tau s_0,\hat{W},\hat{Y})}{\sum_q \int d\tau \mathcal{L}^{\mathrm{SM}}_{q\bar{q}}(\tau,\mu_F)\hat{\sigma}(\tau s_0,0,0)}\,. \tag{4.4.9}$$

The likelihood $\mathcal{L}$ is defined as the product of Poisson probabilities in each bin,

$$\mathcal{L}(n|\hat{W},\hat{Y}) = \prod_{i=1}^{m} \frac{y_i(\hat{W},\hat{Y})^{n_i}}{n_i!} e^{-y_i(\hat{W},\hat{Y})}\,, \tag{4.4.10}$$

and the best-fit values of $\hat{W}$, $\hat{Y}$ are determined as the maximum-likelihood estimates. The test statistics in the individual fits of $\hat{W}$ and $\hat{Y}$ are the profile likelihood ratios defined as

$$\begin{aligned} \lambda(\hat{W}) &= -2\ln\left(\frac{\mathcal{L}(n|\hat{W},0)}{\mathcal{L}(n|\hat{W},0)_{\mathrm{max}}}\right) \\ \lambda(\hat{Y}) &= -2\ln\left(\frac{\mathcal{L}(n|0,\hat{Y})}{\mathcal{L}(n|0,\hat{Y})_{\mathrm{max}}}\right) \end{aligned} \tag{4.4.11}$$

and follow a χ^2 distribution with $n_{\rm dof} = 1$. The 1σ and 2σ bounds are found by solving the implicit equations $\lambda(\hat{W}) = 1, 4$ respectively.

PDF uncertainties can be included by taking the confidence level intervals on the bounds given by the NNPDF Monte Carlo replicas, as it is explained in §4.3.3.

The results for the fits obtained by using SM PDFs in the K-factors, Eq. (4.4.9), compared to those obtained by using SMEFT PDFs, Eq. (4.4.8), are displayed in Fig. 4.4.7, with the corresponding bounds being provided in Table 4.4.2. We find that inclusion of PDF uncertainties has a much smaller impact on the parabolic fit than in the previous analysis and thus only the parabola including PDF uncertainties is displayed. Secondly, we observe that the shift in the bounds that one has using SM versus SMEFT PDFs is not entirely negligible. This finding indicates that it is important to use consistent PDFs determined with the same settings as the theoretical predictions in the partonic cross section. We can similarly recast the ATLAS search data to constrain the scenario II Wilson coefficient $\mathbf{C}_{33}^{Dμ}$, finding the following constraints at 95% CL:

$$\left(\mathbf{C}_{33}^{Dμ} \times 10^2\right) \in [-1.6, 2.4] \ , \tag{4.4.12}$$

in good agreement with the results reported in Eq. (4.4.6).

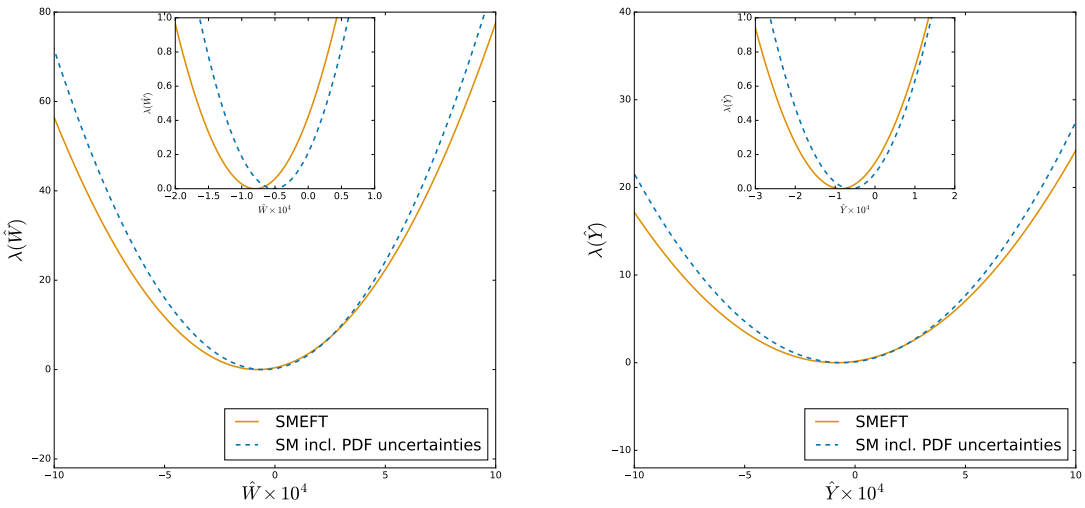


Fig. 4.4.7. Comparison between the results of the parabolic fits to the ATLAS search data [7] for $\hat{W}$ (left) and $\hat{Y}$ (right panel) when using the SMEFT PDFs ($\chi^2_{\rm eftp}$) as compared to the SM PDF baseline ($\chi^2_{\rm smp}$). The insets zoom on the region close to $\lambda(\hat{W}), \lambda(\hat{Y}) \simeq 0$.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^4$ (68% CL)	$[-1.6, 0.6]$ $[-1.6, 0.6]$	$[-1.9, 0.5]$	-0.2 -0.2	+9% +9%
$\hat{W} \times 10^4$ (95% CL)	$[-2.7, 1.7]$ $[-2.7, 1.7]$	$[-3.1, 1.6]$	-0.25 -0.25	+7% +7%
$\hat{Y} \times 10^4$ (68% CL)	$[-2.6, 1.5]$ $[-2.6, 1.6]$	$[-3.1, 1.4]$	-0.3 -0.35	+10% +7%
$\hat{Y} \times 10^4$ (95% CL)	$[-4.7, 3.5]$ $[-4.7, 3.6]$	$[-5.3, 3.6]$	-0.25 -0.30	+9% +7%

Table 4.4.2. The 68% and 95% CL bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the corresponding parabolic fits to the ATLAS search data of [7] when using either the SMEFT PDFs or their SM counterparts. For the SM PDF results, we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for; the SMEFT PDF bounds already include PDF uncertainties by construction. The fourth and fifth column indicate the absolute shift in best-fit values and the percentage broadening of the EFT parameter uncertainties when the SMEFT PDFs are consistently used instead of the SM PDFs.

4.4.4 Overview of current constraints

In order to summarise the results obtained in this section, Fig. 4.4.8 displays the 95% CL bounds derived on the EFT parameters $\hat{W}$ and $\hat{Y}$ (in scenario I) and on $\mathbf{C}_{33}^{Du}$ (in scenario II), both from the high-mass DY cross section measurements (Table 4.4.1) and from the ATLAS Z' search data (Table 4.4.2). These bounds are shown in the case of theoretical calculations evaluated either with SM PDFs or with SMEFT PDFs, and in the former case we indicate the results that account for PDF uncertainties (these are included by construction for the SMEFT PDFs). To compare with previous works, we also display the bounds derived in [8] for the $\hat{W}$ and $\hat{Y}$ parameters from the ATLAS 8 TeV data and in [83] for the $\mathbf{C}_{33}^{Du}$ coefficient from the same ATLAS Z' search data.

As discussed above, our main findings are that the consistent simultaneous determination of the PDFs together with the EFT parameters leads to a moderate increase in the uncertainties (in this case, up to 10%) as well as to a small shift in their central values. As we demonstrate in the next section, the interplay between PDFs and EFT coefficients becomes much more marked in the case of the high-mass DY measurements that will become available at the HL-LHC.

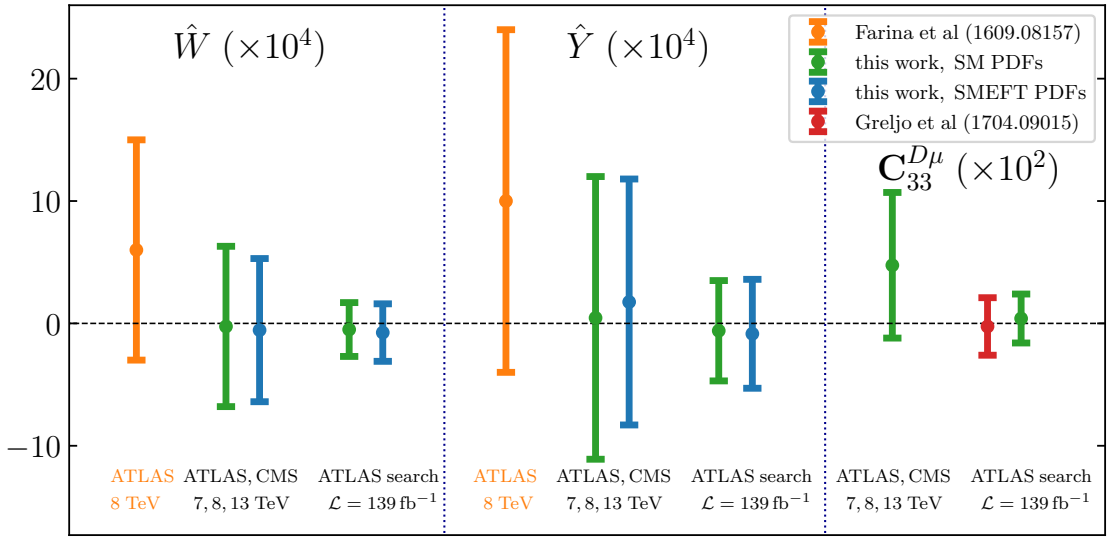


Fig. 4.4.8. Overview of the results obtained in this section concerning the EFT parameters $\hat{W}$ and $\hat{Y}$ (in scenario I) and $\mathbf{C}_{33}^{Du}$ (in scenario II). We compare the 95% CL bounds derived in [8] with those obtained in this work from the high-mass DY cross section measurements (Table 4.4.1) and from the ATLAS Z' search data (Table 4.4.2), in both cases displaying the results obtained with either the SM or the SMEFT PDFs. In the former case, we indicate the results that account for PDF uncertainties; these are included by construction for the SMEFT PDFs.

4.5 Projections for the High-Luminosity LHC

The results presented in the previous section indicate that, given the available unfolded Drell-Yan measurements, the impact of a simultaneous determination of the PDFs together with the EFT parameters remains moderate. However, it is conceivable that this interplay between PDFs and BSM effects in the high-energy tails of Drell-Yan cross sections will become more significant once more data are accumulated. With this motivation, we revisit the analysis of §4.4 now accounting for the impact of projected High-Luminosity LHC pseudo-data generated for the present study. We demonstrate that in the scenario under consideration, in which no other data apart from the high-mass Drell-Yan constrain the large-x quark and antiquark distributions, a consistent joint determination of PDFs is crucial for EFT studies at the HL-LHC. We will also discuss how the inclusion of further LHC data, which can constrain the large-x region without being affected by potential energy-growing new physics effects, can soften the interplay observed in this study and disentangle new-physics effects.

4.5.1 Generation of HL-LHC pseudo-data

Following the strategy adopted in [218] to estimate the ultimate PDF reach of the HL-LHC measurements (see also [281, 282]), here we generate HL-LHC pseudo-data for NC and CC high-mass Drell-Yan cross sections at $\sqrt{s} = 14$ TeV and for a total integrated luminosity of $\mathcal{L} = 6$ ab^{-1} (from the combination of ATLAS and CMS, which provide $\mathcal{L} = 3$ ab^{-1} each). For these projections, theoretical predictions are evaluated at NNLO in QCD including NLO EW corrections, as is explained in detail in §4.3.2. The PDF set used as an input to generate the theoretical prediction is the DIS+DY baseline that was presented in §4.3.3.

For the generation of the NC pseudo-data, we adopt as reference the CMS measurement at 13 TeV [260] based on $\mathcal{L} = 2.8$ fb^{-1}. The dilepton invariant mass distribution $m_{\ell\ell}$ is evaluated using the same selection and acceptance cuts of [260] but now with an extended binning in the $m_{\ell\ell}$ to account for the increase in luminosity. We assume equal cuts for electrons and muons and impose $|\eta_\ell| \leq 2.4$, $p_T^{\rm lead} \geq 20$ GeV, and $p_T^{\rm sublead} \geq 15$ GeV for the two leading charged leptons of the event. In the case of the CC pseudo-data, the lack of unfolded measurements of the m_T distribution at 13 TeV to be used as reference forces us to base our projections on the ATLAS search for W' bosons in the dilepton channel [283]. As in the case of the NC projections, theory predictions for the m_T distribution at high-mass are generated using the same selection and acceptance cuts as in [283] but now using an extended coverage in m_T.

Both in the case of NC and CC Drell-Yan cross sections, we restrict ourselves to events with either $m_{\ell\ell}$ or m_T greater than 500 GeV. Otherwise, the total experimental uncertainty would be limited by our modelling of the expected systematic errors and thus our projections could become unreliable. Furthermore, we require that the expected number of events per bin is bigger than 30 to ensure the applicability of Gaussian statistics. Taking into account these considerations, our choice of binning for the $m_{\ell\ell}$ (m_T) distribution at the HL-LHC is displayed in Fig. 4.5.1 (Fig. 4.5.2), with the highest energy bins reaching $m_{\ell\ell} \simeq 4$ TeV ($m_T \simeq 3.5$ TeV) for neutral-current (charged-current) scattering.

The percentage statistical and systematic uncertainties associated to the HL-LHC pseudo-data are displayed in the lower panels of Figs. 4.5.1 and 4.5.2 and have been estimated as follows. Let us denote by $\sigma_i^{\rm th}$ the theoretical prediction for the DY cross section, including all relevant selection cuts as well as the leptonic branching fractions. The expected number of events in this bin and the associated (relative) statistical

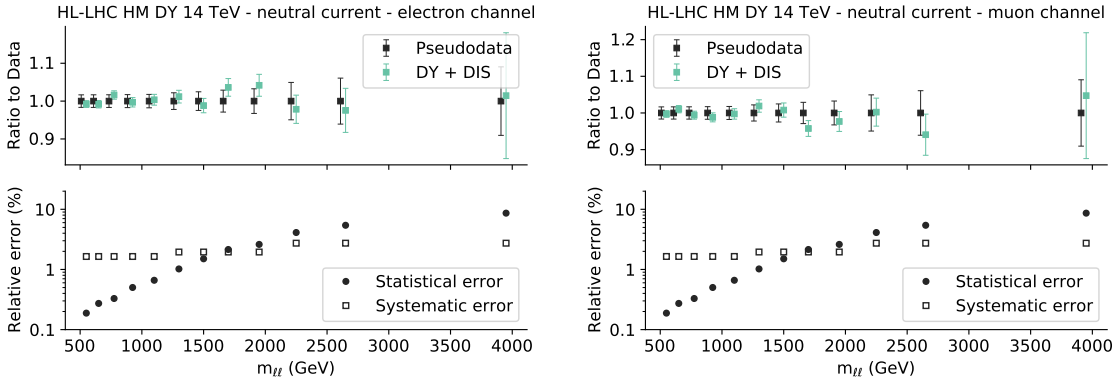


Fig. 4.5.1. Top panels: comparison of the projected HL-LHC pseudo-data for high-mass neutral-current Drell-Yan in the dielectron (left) and dimuon (right) final states as a function of $m_{\ell\ell}$ with the corresponding theory predictions obtained from the SM PDF baseline. The theoretical predictions, generated according to Eq. (4.5.3), are accompanied by their corresponding PDF uncertainties (green bars). Lower panels: the percentage statistical and systematic uncertainty in each $m_{\ell\ell}$ bin of the HL-LHC pseudo-data.

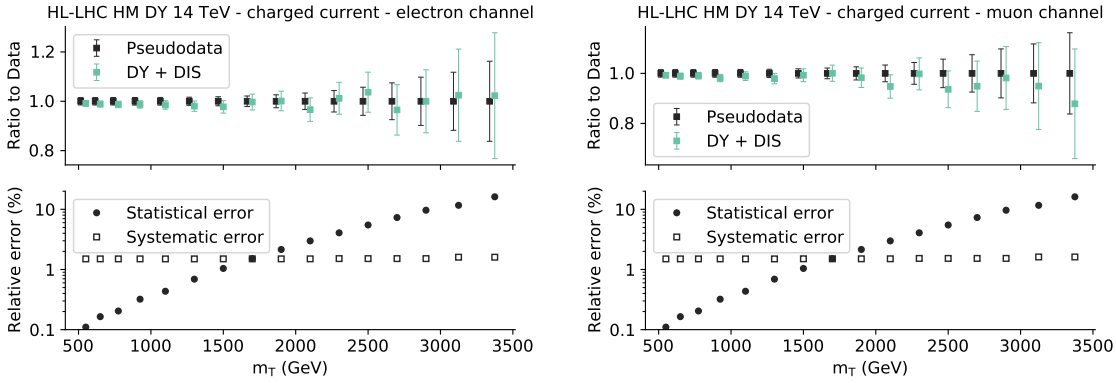


Fig. 4.5.2. Same as Fig. 4.5.1 for charged-current Drell-Yan in bins of the transverse mass m_T.

uncertainty $\delta_i^{\rm stat}$ are given by

$$N_i^{\rm th} = \sigma_i^{\rm th} \times \mathcal{L}\,, \quad \delta_i^{\rm stat} \equiv \frac{(\delta N_i)_{\rm stat}}{N_i^{\rm th}} = \frac{1}{\sqrt{N_i^{\rm th}}}\,. \tag{4.5.1}$$

Note that this bin-by-bin relative statistical uncertainty is the same both at the level of number of events and at the level of fiducial cross sections.

The HL-LHC systematic uncertainties are also estimated from the same reference measurements. If $\delta_{i,j}^{\rm sys}$ denotes the $j^{\rm th}$ relative systematic uncertainty associated to the $i^{\rm th}$ bin of the reference measurement, and if this bin contains $N_i^{\rm th}$ events, then for our projections we assume that the same systematic error associated to a bin with a similar number of expected events will be given by $f_{{\rm red},j}\delta_{i,j}^{\rm sys}$, where $f_{{\rm red},j}$ is the expected reduction in systematic errors foreseen at the HL-LHC.[2] This assumption is justified since most systematic errors improve with the sample size thanks to *e.g.* better calibration.

Adding in quadrature systematic uncertainties with the statistical error, the total relative uncertainty for the ith bin of our HL-LHC projections is

$$\delta_{{\rm tot},i}^{\rm exp} = \left(\left(\delta_i^{\rm stat}\right)^2 + \sum_{j=1}^{n_{\rm sys}} \left(f_{{\rm red},j}\delta_{i,j}^{\rm sys}\right)^2 \right)^{1/2} , \tag{4.5.2}$$

where $n_{\rm sys}$ indicates the number of systematic error sources. The central values for the HL-LHC pseudo-data is then generated by fluctuating the reference theory prediction by the expected total experimental uncertainty, namely

$$\sigma_i^{\rm hllhc} \equiv \sigma_i^{\rm th}\left(1+\lambda\delta_{\mathcal{L}}^{\rm exp} + r_i\delta_{{\rm tot},i}^{\rm exp}\right) , \qquad i=1,\ldots,n_{\rm bin} , \tag{4.5.3}$$

where λ, r_i are univariate Gaussian random numbers, $\delta_{{\rm tot},i}^{\rm exp}$ is the total (relative) experimental uncertainty corresponding to this specific bin (excluding the luminosity and normalisation uncertainties), and $\delta_{\mathcal{L}}^{\rm exp}$ is the luminosity uncertainty, which is fully correlated amongst all the pseudo-data bins of the same experiment. We take this luminosity uncertainty to be $\delta_{\mathcal{L}}^{\rm exp} = 1.5\%$ for both ATLAS and CMS, as done in Ref. [218].

Here we adopt the baseline SM PDF set described in §4.4, which is denoted as "DIS+DY", to evaluate the $\sigma_i^{\rm th}$ cross sections entering Eq. (4.5.3). We have verified that, both at the pre- and post-fit levels, the fit quality to the HL-LHC pseudo-data satisfies $\chi^2/n_{\rm bin} \simeq 1$ in the case of the SM PDFs as expected. Furthermore, we assume $f_{{\rm red},j} = 0.2$ for all systematic sources, as done in the optimistic scenario of Ref. [218]. We note that more conservative values for the reduction of systematic errors, such as $f_{{\rm red},j} = 0.5$, are not expected to qualitatively modify our results. The reason is that, as indicated by the bottom panels of Figs. 4.5.1 and 4.5.2, for the highest energy bins

[2] The binning of the CC reference measurement, the ATLAS W' search, is much finer than for our HL-LHC projections and hence we first match them by means of a weighted average.

(which dominate the EFT sensitivity), specifically above $m_{\ell\ell} \approx 1.7$ TeV and $m_T \approx 1.5$ TeV, the measurement will be limited by statistical uncertainties.

4.5.2 Impact on PDF uncertainties

From Figs. 4.5.1 and 4.5.2, one can observe that the PDF uncertainties in the SM PDF baseline used to generate the pseudo-data are either comparable or larger than the corresponding projected experimental uncertainties at the HL-LHC. Specifically, for the highest $m_{\ell\ell}$ bin of the NC distribution the PDF errors are twice the experimental ones, while in the CC case the associated PDF errors become clearly larger than the experimental ones starting at $m_T \simeq 2$ TeV. This comparison suggests that one should expect a significant uncertainty reduction once the HL-LHC pseudo-data is included in the PDF fit.

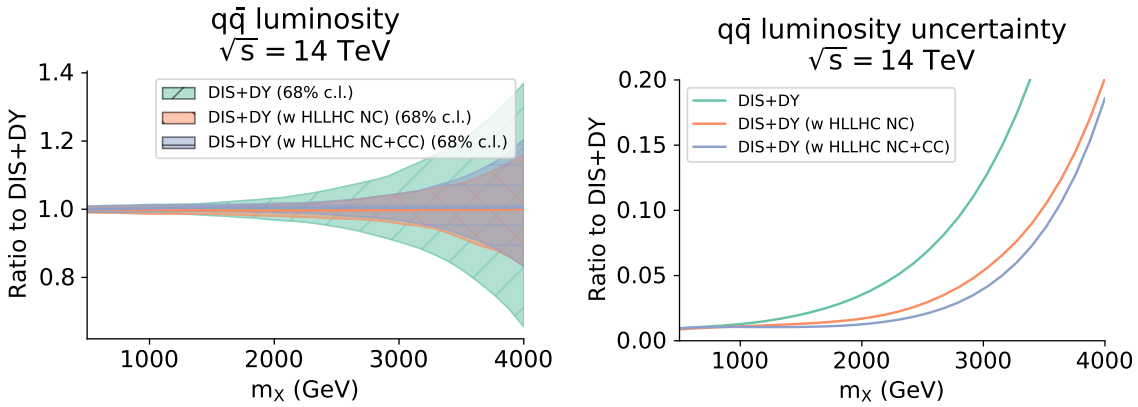


Fig. 4.5.3. Impact of the HL-LHC pseudo-data on the quark-antiquark luminosity $\mathcal{L}_{q\bar{q}}$ of the SM PDF baseline fit as a function of m_X. Left: the luminosities $\mathcal{L}_{q\bar{q}}$ for the DIS+DY baseline and the corresponding fits including the HL-LHC pseudo-data, either only NC or also with CC cross sections, presented as a ratio to the central value of the former. Right: the relative PDF uncertainty in $\mathcal{L}_{q\bar{q}}$ (with the central value of the DIS+DY baseline as reference) for the same fits.

To validate this expectation, Fig. 4.5.3 displays the impact of the HL-LHC pseudo-data on the quark-antiquark luminosity $\mathcal{L}_{q\bar{q}}$ as a function of the final state invariant mass m_X at $\sqrt{s} = 14$ TeV. We compare $\mathcal{L}_{q\bar{q}}$ for the SM PDF baseline fit (DIS+DY) with the same quantity from the corresponding fits including the HL-LHC pseudo-data, either only NC or also with CC cross sections. The right panel displays the associated relative PDF uncertainties. We find a significant reduction of the PDF uncertainties affecting the quark-antiquark luminosity (and hence the Drell-Yan cross sections) in

the high mass ($m_X \gtrsim 1$ TeV) region once the HL-LHC pseudo-data constraints are accounted for. For instance, at $m_X \gtrsim 2$ TeV, PDF uncertainties on $\mathcal{L}_{q\bar{q}}$ decrease from $\simeq 5\%$ in the baseline down to $\simeq 2.5\%$ ($\simeq 1.5\%$) once the NC (NC+CC) HL-LHC pseudo-data is included in the fit. The effect of the inclusion of HL-LHC projections becomes more dramatic as m_X increases. On the other hand, other partonic luminosities such as the quark-quark and gluon-gluon ones are essentially unaffected by the HL-LHC constraints. In terms of fit quality, the only noticeable effect is a mild improvement in the χ^2 of the high-mass DY datasets listed in Table 4.3.2.

4.5.3 PDF and EFT interplay at the HL-LHC

The finding that the projected HL-LHC pseudo-data has a significant impact on the quark-antiquark PDF luminosity, summarised in Fig. 4.5.3, suggests that the interplay between PDFs and EFT effects in the high-energy DY tails should become enhanced as compared to the results reported in the previous section. With this motivation, we first of all repeat the joint determination of PDFs and the $\hat{W}, \hat{Y}$ coefficients from EFT scenario I presented in §4.4.1 now accounting for the constraints of the HL-LHC pseudo-data. An important difference in this case is that the inclusion of CC data lifts the flat direction in the $(\hat{W}, \hat{Y})$ plane, making a full two-dimensional fit possible. Secondly, the availability of the HL-LHC pseudo-data allows us to assess the interplay between the PDFs and the EFT coefficient $\mathbf{C}_{33}^{Dμ}$ from benchmark scenario II, whose analysis in §4.4.2 was restricted to fixed SM PDFs.

Scenario I. For the simultaneous determination of PDFs and the $\hat{W}, \hat{Y}$ coefficients accounting for the constraints provided by the HL-LHC pseudo-data, we use 35 sampling values of $(\hat{W}_i, \hat{Y}_i)$, 25 of which are equally spaced in either $\hat{W} \in (-1.6, 1.6) \times 10^{-5}$ or $\hat{Y} \in (-8, +8) \times 10^{-5}$ (hence in steps of $\Delta\hat{W} = 0.8 \times 10^{-6}$ and $\Delta\hat{Y} = 4 \times 10^{-6}$ respectively), and then 10 additional points along the diagonals. In order to assess the robustness of the results, we added 12 more sampling values, 8 further away from the origin and 4 more along the $\hat{W} = 0$ and $\hat{Y} = 0$ axes, and verified that the confidence level countours are stable upon their addition.

We find that the constraints on the $(\hat{W}, \hat{Y})$ parameters are completely dominated by the HL-LHC projections and that current data exhibit a much smaller pull, consistent with the findings of previous studies [8, 9]. Also, the $\chi^2_{\rm eftp}$ contour is more stable and requires less replicas if only the HL-LHC projections are included in the computation of the χ^2. The corresponding marginalised bounds on $\hat{W}$ and $\hat{Y}$ are reported in Table 4.5.1 using the same format as in Table 4.4.1.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^5$ (68% CL)	$[-0.7, 0.5]$	$[-4.5, 6.9]$	1.3	850%
	$[-1.0, 0.9]$		1.3	500%
$\hat{W} \times 10^5$ (95% CL)	$[-1.0, 0.8]$	$[-8.1, 10.6]$	1.4	940%
	$[-1.4, 1.2]$		1.4	620%
$\hat{Y} \times 10^5$ (68% CL)	$[-1.8, 3.2]$	$[-6.4, 8.0]$	0.1	190%
	$[-3.7, 4.7]$		0.3	70%
$\hat{Y} \times 10^5$ (95% CL)	$[-3.4, 4.7]$	$[-11.1, 12.6]$	0.1	190%
	$[-5.3, 6.3]$		0.3	110%

Table 4.5.1. Same as Table 4.4.1 for the 68% CL and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional ($\hat{W}$,$\hat{Y}$) fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. As in Table 4.4.1, for the SM PDFs we indicate the bounds obtained without (upper) and with (lower entry) PDF uncertainties accounted for.

From Table 4.5.1, one can observe how including high-mass data at the LHC both in a fit of PDFs and in a fit of SMEFT coefficients and neglecting the interplay between them could result in a significant underestimate of the uncertainties associated to the EFT parameters. Indeed, the marginalised 95% CL bound on the $\hat{W}$ ($\hat{Y}$) parameter becomes looser once SMEFT PDFs are consistently used, with a broadening, defined in Eq. (4.4.4), of 500% (110%), even once PDF uncertainties are fully accounted for. This effect would have been even more marked if PDF uncertainties had not been accounted for in EFT fits based on SM PDFs, where the same broadening factors would be 940% and 190% respectively.

A further important question is whether the bounds obtained with SM PDFs appearing on the left column of Table 4.5.1 would become more comparable to those obtained from the simultaneous fit of PDFs and SMEFT coefficients, in case a conservative set of PDF was used in the analysis based on SM PDFs. To address this question, in Table 4.5.2 we display the bounds that are obtained using a PDF set that does not include any of the high-mass Drell-Yan sets (neither the HL-LHC projections nor the current datasets listed in Table. 4.3.2) and compare the bounds obtained using this set of PDFs to those obtained consistently using SMEFT PDFs. We observe that, once this set of conservative PDF is used as an input PDF set and the PDF uncertainty is included in the computation of the bounds, the latter increases as compared to the

	SM cons. PDFs	SMEFT PDFs	best-fit shift	broadening
$\hat{W} \times 10^5$ (68% CL)	$[-1.0, 0.0]$	$[-4.5, 6.9]$	1.7	1000%
	$[-4.0, 2.8]$		1.8	70%
$\hat{W} \times 10^5$ (95% CL)	$[-1.4, 0.4]$	$[-8.1, 10.6]$	1.8	940%
	$[-4.3, 3.1]$		1.9	150%
$\hat{Y} \times 10^5$ (68% CL)	$[2.1, 7.0]$	$[-6.4, 8.0]$	-3.7	190%
	$[-3.4, 11.2]$		-3.6	-1%
$\hat{Y} \times 10^5$ (95% CL)	$[0.5, 8.5]$	$[-11.1, 12.6]$	-3.7	200%
	$[-5.0, 13.7]$		-3.6	30%

Table 4.5.2. Same as Table 4.5.1 for the 68% and 95% CL marginalised bounds on the $\hat{W}$ and $\hat{Y}$ parameters obtained from the two-dimensional ($\hat{W}$,$\hat{Y}$) fits that include the HL-LHC pseudo-data for NC and CC Drell-Yan distributions. The input PDF set for the analysis done using fixed SM PDFs (corresponding to the results displayed in the column "SM cons. PDFs") is a conservative PDF set that does not include any of the high-mass distributions or the HL-LHC projections nor the Run I and Run II high-mass dataset listed in Table 4.3.2. The limits obtained from the simultaneous fit of PDFs and Wilson coefficients (corresponding to the results displayed on the column "SMEFT PDFs") are the same as those in Table 4.5.1.

bounds in Table 4.5.1. As a result, the size of the bounds obtained by keeping fixed SM PDFs is closer to the size obtained from the simultaneous fits, although still slightly underestimated. At the same time, the shift in the best-fit becomes more marked.

Results are graphically displayed in Fig. 4.5.4, where the 95% confidence level contours in the ($\hat{W}$,$\hat{Y}$) plane obtained from the DIS+DY fits that include the high-mass Drell-Yan HL-LHC pseudo-data when using either SM PDFs, SM conservative PDFs or SMEFT PDFs are compared. All solid countours include PDF uncertainties, while the dashed contours that do not include PDF uncertainties are also indicated to visualise the impact of the inclusion of the PDF uncertainties.

To conclude, we should also emphasise that, while in this work we use pseudo-data and hence the best-fit values are by construction unchanged, this would not necessarily be the case in the analysis of real data, where improper treatment of PDFs could result in a spurious EFT 'signal', or even missing a signal which is indeed present in the data. A detailed study aimed at a precise definition of 'conservative' PDFs in a more general scenario is beyond the scope of this paper and will be the topic of future work; a thorough comparison of the consistent simultaneous approach, versus the use

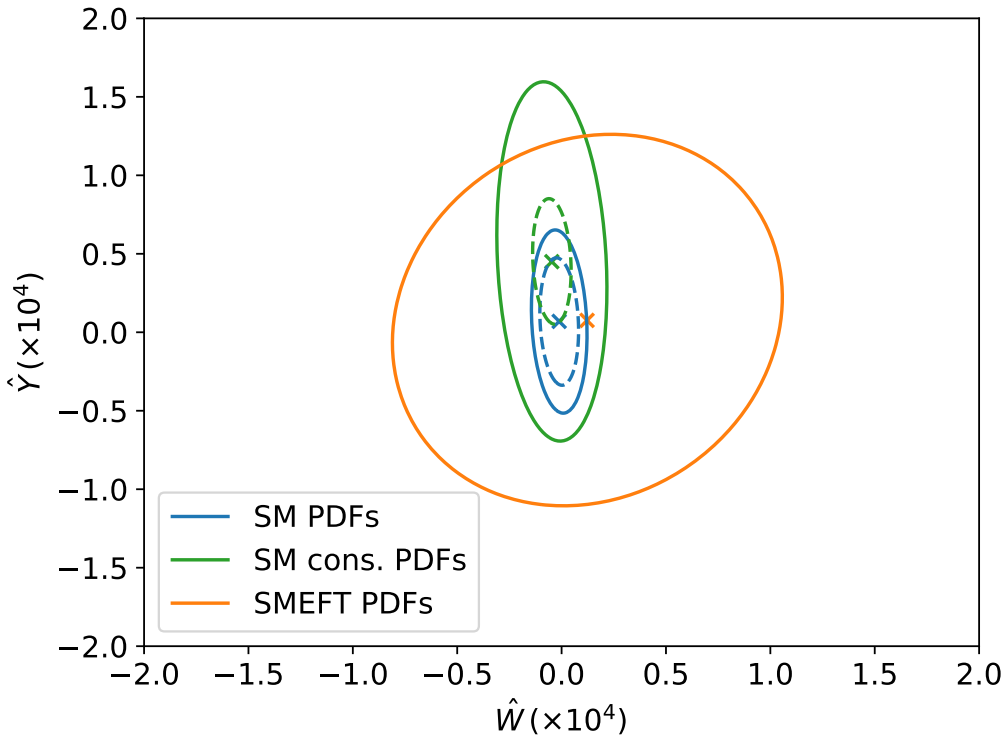


Fig. 4.5.4. The 95% confidence level contours in the $(\hat{W},\hat{Y})$ plane obtained from the DIS+DY fits that include the high-mass Drell-Yan HL-LHC pseudo-data (both in the NC and CC channels) when using either SM PDFs (blue) or conservative SM PDFs (green). In both cases the ellipses are obtained by performing a parabolic fit to $\chi^2_{\rm smp}$ with fixed PDFs. PDF uncertainties are included in the solid lines and not included in the dashed lines. The results are compared to those obtained in a simultaneous fit, namely with SMEFT PDFs (orange). In this case, the parabolic fit is performed to $\chi^2_{\rm eftp}$ by varying simultaneously the Wilson Coefficients and the PDFs. The crosses indicate the best fits in the three cases discussed in the text.

of conservative PDF sets, will be of particular interest in cases of EFT manifestations of new physics.

The increased role that the interplay between PDFs and EFT coefficients will play at the HL-LHC can also be illustrated by comparing the expected behaviour of the quark-antiquark luminosity, displayed in Fig. 4.5.5, for the SMEFT PDFs corresponding to representative values of the $\hat{W}$ and $\hat{Y}$ parameters of benchmark scenario I as compared to the SM PDFs. Note that the corresponding comparison for $\mathcal{L}_{q\bar{q}}$ in the fits to available Drell-Yan data was displayed in Fig. 4.4.3. Indeed, the central value of the quark-antiquark luminosity for SMEFT PDFs corresponding to values of $(\hat{W}_i, \hat{Y}_i)$ selected along the grid used to derive Fig. 4.5.5 changes greatly,

well outside the one-sigma error band of the SM PDFs, while the PDF uncertainties themselves are unchanged. This change in central value of the large-x PDFs partially reabsorbs the effects in the partonic cross section induced by the SMEFT operators and leads to better χ^2 values as compared to those obtained with the SM PDFs.

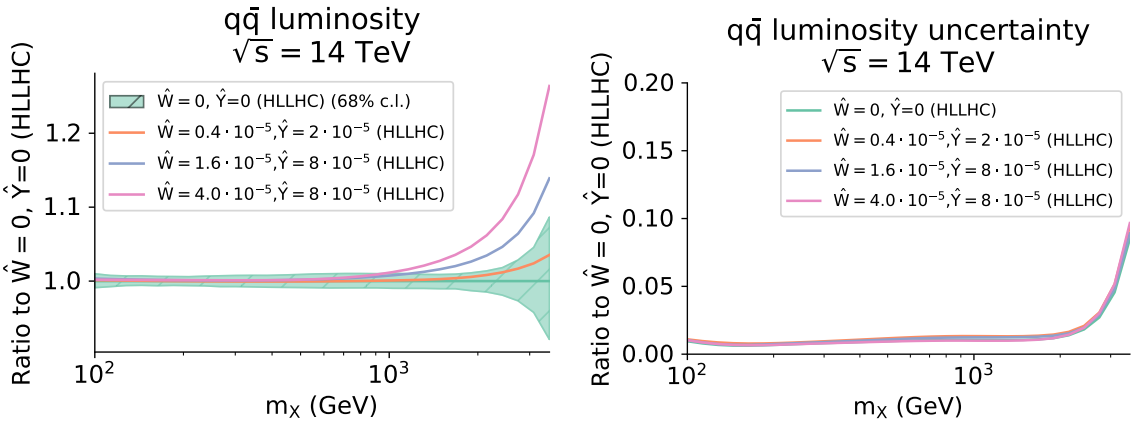


Fig. 4.5.5. Same as Fig. 4.5.3, now comparing the quark-antiquark SM PDF luminosity in the fits including the HL-LHC pseudo-data with those obtained in the SMEFT PDF fits for representative values of the $\hat{W}$ and $\hat{Y}$ parameters. The corresponding comparison in the case of fits to available Drell-Yan data was shown in Fig. 4.4.3.

Even neglecting SMEFT PDF effects, we note that our marginalised bounds on the $\hat{W}$ and $\hat{Y}$ coefficients from HL-LHC pseudo-data using SM PDFs turn out to be more stringent than those reported in [9] by around a factor of 4 for $\hat{W}$ and a factor 2 for $(\hat{Y})$. This is due to a combination of factors. First of all we use the 13 TeV measurements as reference to produce the HL projections. Furthermore we assume a total integrated luminosity of $\mathcal{L} = 6$ fb^{-1} (from the combination of ATLAS and CMS) rather than 3 fb^{-1} as well as a more optimistic scenario concerning the reduction of the experimental systematic uncertainties.

Fig. 4.5.6 then displays the R_{χ^2} estimator, defined in Eq. (4.4.5) and shown in Fig. 4.4.4 for the case of available LHC data, now evaluated from the fits including the HL-LHC pseudo-data. In this case, the $m_{\ell\ell}^{(\max)}$ cut applies to $m_{\ell\ell}$ for the neutral-current distributions and to the transverse mass m_T for the charged-current ones. As in the case of Fig. 4.4.4, we observe an approximately monotonic growth of R_{χ^2} for the SM PDFs arising from the energy-growing EFT corrections that dominate the high-energy DY tails, an effect which is now rather larger thanks to the presence of the HL-LHC pseudo-data. The most striking difference as compared to Fig. 4.4.4 is that now the

SMEFT PDF curve is much flatter, indicating that EFT effects are being almost totally reabsorbed into the PDFs.

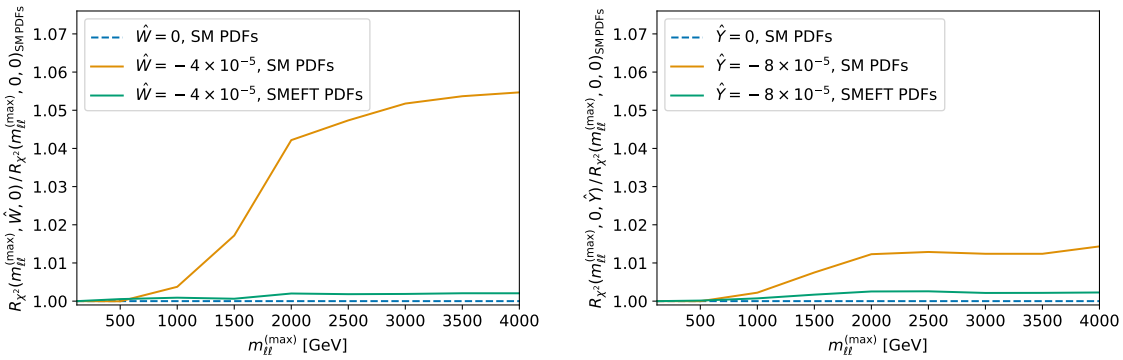


Fig. 4.5.6. Same as Fig. 4.4.4 now for the fits including the HL-LHC pseudo-data. Note that in this case the $m_{\ell\ell}^{(\max)}$ cut refers to the transverse mass m_T for the charged-current distributions.

The findings summarised by Fig. 4.5.6 demonstrate that, at the HL-LHC, EFT-induced deviations could be indeed inadvertently "fitted away" into a PDF redefinition, explaining the large broadenings reported in Fig. 4.5.5, and highlight the need to devise novel strategies to disentangle the effects of PDFs and EFT contributions from the high-energy tails of LHC cross-sections. Such strategies could exploit, for instance, the availability of measurements sensitive to large-x PDFs but not to high scales, such as forward electroweak gauge boson production by LHCb [218].

Scenario II. We now turn to present the corresponding results of the simultaneous fits of the PDFs and the EFT coefficients including the HL-LHC pseudo-data for the case of benchmark scenario II. As motivated in §4.2.2, a non-zero value of the $\mathbf{C}_{33}^{D\mu}$ coefficient affects only the NC and CC muon final states, while the electron ones remain described by the SM calculations. This property implies that, in fits presented below, the EFT corrections modify only the shapes of the HL-LHC distributions in the muon channel, the right panels in Figs. 4.5.1 and 4.5.2, but not those of the electron pseudo-data.

Fig. 4.5.7 displays the values of $\Delta\chi^2$ obtained for the SMEFT PDFs as a function of the EFT parameter $\mathbf{C}_{33}^{D\mu}$ from the joint fits that include the HL-LHC pseudo-data. The sampling is constituted by 21 points uniformly distributed in $\mathbf{C}_{33}^{D\mu} \in [-0.02, 0.02]$. As in Fig. 4.4.1, the error bars indicate the uncertainties associated to the finite number of Monte Carlo replicas used for each value of $\mathbf{C}_{33}^{D\mu}$. The profile in $\Delta\chi^2$ exhibits a

double minimum structure (bimodal distribution), explained by the fact that in this scenario it is the quadratic rather than the linear terms in the EFT expansion that dominate. The corresponding quartic polynomial fit using Eq. (4.4.2) can be seen to successfully reproduce the $\Delta\chi^2$ values obtained in this joint analysis. The right panel of Fig. 4.5.7 then compares the polynomial fit obtained with the SMEFT PDFs with the corresponding one when using instead fixed SM PDFs to determine the $\Delta\chi^2$ values, with the inset focusing on the region close to $\Delta\chi^2 \simeq 0$. The associated 68% and 95% CL bounds are then reported in Table 4.5.3, where we note that since the 68% CL interval is disjoint we evaluate the shift and broadening only for the 95% CL bounds.

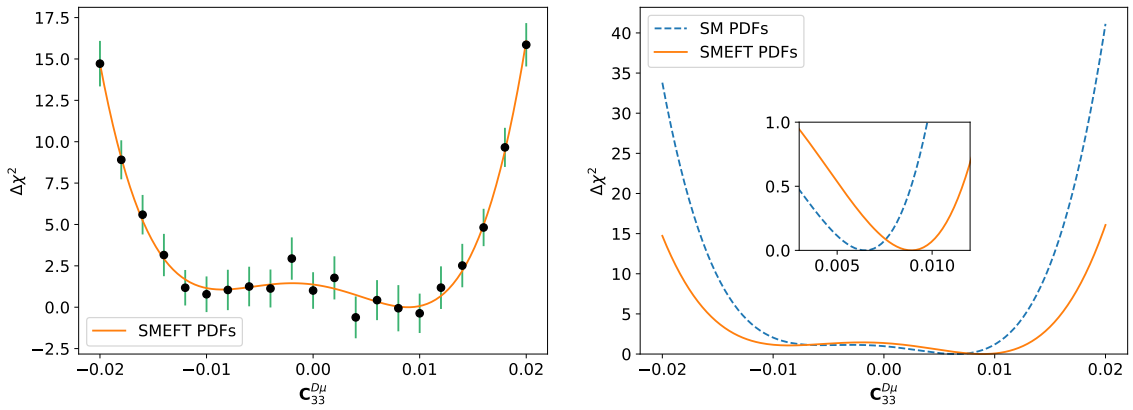


Fig. 4.5.7. Left: the values of $\Delta\chi^2$ obtained for the SMEFT PDFs as a function of $\mathbf{C}_{33}^{D\mu}$ from the fits including the HL-LHC pseudo-data, together with the corresponding quartic polynomial fit. Right: comparison of the polynomial fit obtained with SMEFT PDFs and displayed in the left panel with its counterpart based on SM PDFs.

	SM PDFs	SMEFT PDFs	best-fit shift	broadening
$\mathbf{C}_{33}^{D\mu} \times 10^2$ (68% CL)	$[-0.1, 1.1]$	$[-0.3, 1.2]$	0.06	25%
$\mathbf{C}_{33}^{D\mu} \times 10^2$ (95% CL)	$[-1.0, 1.2]$	$[-1.2, 1.4]$	0.06	18%

Table 4.5.3. Same as Table 4.5.1, now for the $\mathbf{C}_{33}^{D\mu}$ parameter from EFT benchmark scenario II.

Inspection of Fig. 4.5.7 and Table 4.5.3 indicates that, even at the HL-LHC, the interplay between PDFs and EFT coefficients remains moderate in this particular scenario. Indeed, in contrast with the marked effects in scenario I (Fig. 4.5.4), where the bounds on the $\hat{W}$ and $\hat{Y}$ worsened by up to an order of magnitude when the SMEFT PDFs were consistently used, in scenario II the obtained bounds on $\mathbf{C}_{33}^{D\mu}$

would only loosen by around 30%. The origin of this rather different behaviour can be traced back to the fact that in scenario II the electron channel data do not receive EFT corrections, and hence all the information that they provide makes it possible to exclusively constrain the PDFs. The muon channel distributions then determine the allowed range for $\mathbf{C}_{33}^{D\mu}$, restricted by the well-constrained large-x quarks and antiquark PDFs from the electron data. This finding demonstrates how the availability of measurements in separate leptonic final states is of utmost importance to test BSM scenarios that account for violations of Lepton Flavour Universality.

In the same manner as in Fig. 4.5.5, Fig. 4.5.8 displays the comparison of the quark-antiquark luminosities at $\sqrt{s} = 14$ TeV in the fits with HL-LHC pseudo-data in the case of the SM PDFs and for the SMEFT PDFs for representative values of $\mathbf{C}_{33}^{D\mu}$. Specifically, we show $\mathbf{C}_{33}^{D\mu} = -0.004$ and 0.012, chosen to lie at the boundary of the 68% CL interval reported in Table 4.5.3. The result that the two values lead to the same effect on $\mathcal{L}_{q\bar{q}}$ follows from the dominance of the quadratic EFT terms in this scenario. One finds that the shift in the central values of the quark-antiquark luminosity induced by a non-zero value of $\mathbf{C}_{33}^{D\mu}$ is well within PDF uncertainties. This is consistent with the result of Fig. 4.5.7 indicating that bounds on $\mathbf{C}_{33}^{D\mu}$ obtained with SM and with SMEFT PDFs are relatively similar in this scenario even after accounting for the HL-LHC constraints.

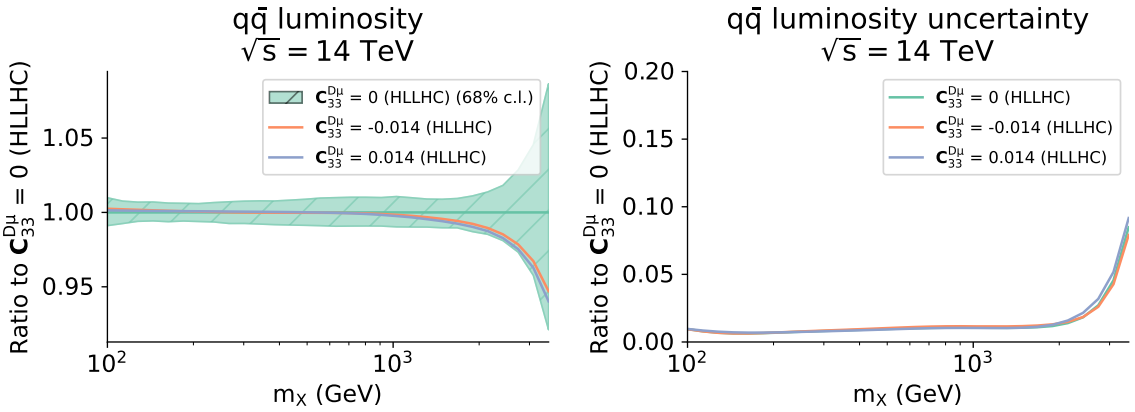


Fig. 4.5.8. Same as Fig. 4.5.5 now in the case of the simultaneous fits of the PDFs and the $\mathbf{C}_{33}^{D\mu}$ EFT parameter taking into account HL-LHC pseudo-data.

4.6 Concluding discussion

In this Chapter we have presented a first simultaneous determination of PDFs and EFT coefficients from high-energy LHC data, specifically from high-mass Drell-Yan cross sections. Our analysis has considered available unfolded measurements, detector-level searches based on the full Run II luminosity, and tailored HL-LHC projections. The EFT interpretation of the Drell-Yan data is formulated in terms of two benchmark scenarios, first a flavour universal one leading to modifications of the $\hat{W}$ and $\hat{Y}$ electroweak parameters [8], and second a flavour-specific scenario motivated by the recent evidence for lepton flavour universality violation in B-meson decays [83].

The main findings of this work are summarised in Fig. 4.6.1. We demonstrate how, for the analysis of all available unfolded Drell-Yan data, the consistent simultaneous extraction of the PDFs together with the EFT parameters leads to a modest increase in the uncertainties of the latter (up to 15%, in the case of the $\hat{W}$ and $\hat{Y}$ parameters), as well as to a shift in their central values by up to a third of a sigma. Furthermore, while our results indicate that for current data the interplay between PDF and EFT effects remains moderate, the impact of their cross-talk will become much larger at the HL-LHC: using SM rather than SMEFT PDFs would lead to artificially precise bounds, even mimicking new physics effects. This result indicates that including high-energy data in PDF fits should be done with care, as PDFs can actually absorb the effects of new physics.

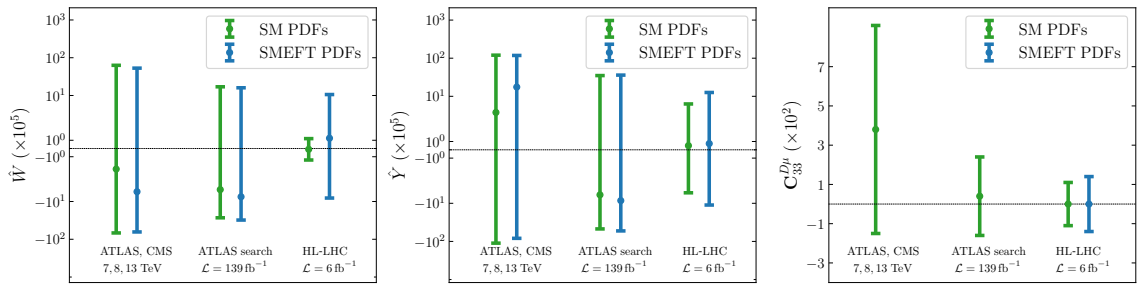


Fig. 4.6.1. Overview of the 95% CL bounds obtained in this work on the EFT parameters $\hat{W}$ (left), $\hat{Y}$ (middle), and $\mathbf{C}_{33}^{Dμ}$ (right panel) based on either the SM PDFs or the SMEFT PDFs. Both PDF and methodological uncertainties are accounted for in the bounds, when available.

We have discussed estimators, such as shown in Fig. 4.4.4, which provide information on the kinematic dependence of any possible deviation between the data and the SM predictions, and the extent to which this can be reabsorbed into the PDFs. These

represent a powerful diagnostic tool to separate QCD effects from genuine BSM deviations. A complementary strategy to disentangle QCD effects from BSM effects would be to account for the constraints on the large-x PDFs arising from other processes for which EFT corrections can be neglected, such as forward W, Z production at LHCb. This way the uncertainties associated to the PDFs at large-x would be reduced and the indirect signal for new physics could be more easily disentangled. A detailed study aimed at a definition of conservative PDFs in a more general scenario is beyond the scope of this work, and will be the topic of future studies.

Exploiting the full potential of future precision measurements at the HL-LHC for indirect BSM searches requires that we no longer neglect the interplay between PDFs and EFT effects. The development of novel data interpretation frameworks capable of accounting for this interplay is crucial as we move towards the HL-LHC. In Chapter 5 we will continue to study the potential of the HL-LHC to discover BSM physics. Leaving behing the subtlety of indirect searches, we will investigate the potential of the HL-LHC and other future colliders to directly detect new physics. In this context we neglect the interplay with PDFs, this time searching for a clear 'bump' in the data resulting from the direct on-shell production of a new particle.

Chapter 5

Leptoquarks at Future Hadron Colliders

5.1 Introduction

As we saw in Chapter 2, working in the framework of the weak effective theory allows us to interpret signs of new physics in rare B meson decays. Although individually each measurement is discrepant from the SM at the level of only 2-3 σ, the addition of new physics in the form of a scalar leptoquark (LQ) may improve the collective description of these neutral current B anomalies (NCBAs) by a pull of 5.9 σ. The goal of this Chapter is to investigate the prospects for detecting such a scalar LQ. Based on Ref. [3], we will determine the sensitivity of future hadron colliders to LQ pair production in the di-muon di-jets ($\mu^+\mu^- jj$) channel.

We first recall from Chapter 2 that in order to produce a new physics contribution to the $\mathcal{O}_{LL}$ operator of the weak effective theory, we may extend the SM by the scalar S_3 which couples to the SM quarks and leptons as follows:

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} = &- \sqrt{2}\,\overline{\mathbf{d}_L^C} Y_{de} \mathbf{e}_L S^{+4/3} - \overline{\mathbf{u}_L^C} Y_{ue} \mathbf{e}_L S^{+1/3} \\ &- \overline{\mathbf{d}_L^C} Y_{d\nu} \nu_L S^{+1/3} + \sqrt{2}\,\overline{\mathbf{u}_\mathbf{L}^\mathbf{C}} Y_{u\nu} \nu_L S^{-2/3} + h.c.\,.\end{aligned} \tag{5.1.1}$$

Furthermore, we will specialise to LQs which solve the NCBAs, taking only nonzero values of $(Y_{de})_{32}$ and $(Y_{de})_{22}$ as determined from the matching of $\mathcal{L}_{\text{Yukawa}}$ onto the weak effective theory Lagrangian:

$$\frac{(Y_{de})_{32}(Y_{de}^*)_{22}}{m_{LQ}^2} = C_{LL} \frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \frac{e^2}{16\pi^2}. \tag{5.1.2}$$

Our LQs contribute to $B_s - \bar{B}_s$ mixing at one-loop order, as shown in Fig. 5.1.1. A recent determination of the SM prediction for $B_s - \bar{B}_s$ mixing is broadly in agreement with the experimental measurement, and so upper bounds can be placed on the LQ contribution. However, a recent determination [284] finds that this is not very constraining for S_3 leptoquarks that fit the NCBAs. Perturbative unitarity provides the stronger constraint that $m_{LQ} < 68$ TeV.

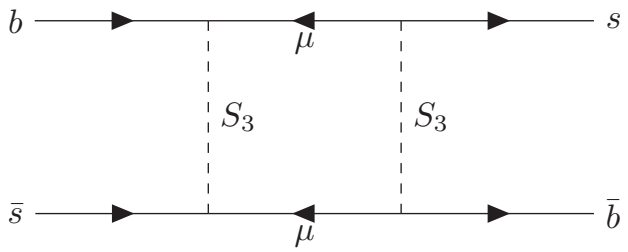


Fig. 5.1.1. A scalar leptoquark will contribute to $B_s - \bar{B}_s$ mixing at one-loop order.

A previous estimate of future collider sensitivity to S_3 LQs consistent with the NCBAs was made in Ref. [285], which projected current sensitivity to higher centre of mass energies and luminosities. However, the sensitivity estimate had two important approximations. The first was that experimental efficiency and acceptance did not change with centre of mass energy. In fact, at large m_{LQ} and at high energies (particularly at the FCC-hh), the decay products from LQs will be highly boosted. This has two effects: the muons will be pushed closer to the jets, meaning that more of them will fail isolation criteria. Also, at higher energies, the muon momentum resolution is likely to be very poor, since such hard muons can only be bent to a limited extent by the magnets. This will also affect the signal efficiency from peak broadening. The second approximation was that the LQs are produced exactly at threshold. This is likely to introduce large uncertainties. We shall rectify these approximations in this Chapter by performing a fast simulation of the signal and SM backgrounds, including the effects of the detector response. The first of these approximations has already been found to have non-trivial effects upon the predicted future hadron collider sensitivity of Z' explanations of the NCBAs [286, 78]. We expect, therefore, that the estimate in this Chapter should be much more accurate than the previous pioneering determination in Ref. [285].

Searches for LQ pair production with subsequent decays of each into a muon and a jet have already been performed at the 13 TeV LHC. The ATLAS Collaboration set a 95% confidence level lower limit on m_{LQ} of 1.05 TeV from 3.2 fb^{-1} of pp collisions [287]. This is a simple cut-based analysis, which we adopt in this work for estimating

future hadron collider sensitivity. More recent experimental analyses were made more sophisticated in order to squeeze more sensitivity out of them. The CMS Collaboration maximise their sensitivity using a multi-dimensional optimisation of the final selection for each m_{LQ} in 36 fb^{-1} of delivered beam at the LHC [288], finding a 95% CL lower bound of $m_{LQ} > 1.28$ TeV. The ATLAS collaboration has also performed a search in 36 fb^{-1} of 13 TeV pp collisions for LQs decaying to muons and jets. They utilise differential cross section measurements and boosted decision trees to obtain a lower bound of $m_{LQ} > 1.23$ TeV. However, such a level of sophistication is unnecessary for our purposes, where the uncertainties involved in estimating future collider sensitivities (for example because we do not yet know the experimental design) are much larger than the gain in sensitivity. Thus, following the much simpler methodology in Ref. [287] is sufficient for our purposes. We also note that at the time of writing Ref. [3] (on which this Chapter is based) ATLAS had not published their analysis of the full LHC Run II data, using 139 fb^{-1} data to place limits of $m_{\mathrm{LQ}} > 1.7$ TeV on scalar LQ masses [133]. As a result, we include here projections for the Run II LHC in our future collider studies, and will comment on the success of our projections with a comparision in §5.5.

Collider sensitivity to LQ pair production is limited by SM background rates. Therefore, the estimation of such background rates is of vital importance to the estimate of the sensitivity to LQ pair production. We begin in §5.2 by describing in detail the methodology used to simulate these SM backgrounds. In §5.3 we discuss the LQ signal used in our simulations and introduce the statistical methods we use to quantify sensitivity. Finally, we present our projections in §5.4 before concluding in §5.5.

5.2 Standard Model Backgrounds

5.2.1 Methodology

Consider the pair production of LQs and their decay to a $\mu^+\mu^- jj$ final state. Following previous searches for LQ pair production, we define the parameter $m_{\mathrm{min}}(\mu, j)$ from the kinematics of these four final state particles by finding the configuration of muon-jet pairings which minimises the difference in invariant masses $|m(\mu_1, j_1) - m(\mu_2, j_2)|$ and choosing $m_{\mathrm{min}}(\mu, j) = \min[m(\mu_1, j_1), m(\mu_2, j_2)]$, where j_1 and j_2 are the hardest two jets in an event. In an on-shell LQ pair production event this parameter will approximate the LQ mass m_{LQ}. We will simulate the distribution of the SM background

in $m_{\min}(\mu, j)$, selecting events containing *exactly two* muons with no charge requirement and *at least two* jets with no flavour requirement.

We generate the SM background events at parton level in `Madgraph5` [289]. These events are then passed to `Pythia8` [279] for the simulation of initial state radiation, parton showering and hadronisation. Finally, the hadron-level events are passed to `Delphes3` [280] for detector simulation. We use the 5-flavour `NNPDF2.3LO` [165] parton distribution function via `LHAPDF6` [290] for all background simulations except for di-boson production, for which the 4-flavour `NNPDF2.3LO` parton distribution function is used. This choice is made to remove interference in di-boson production, as outlined in more detail later in this Section.

Process definitions

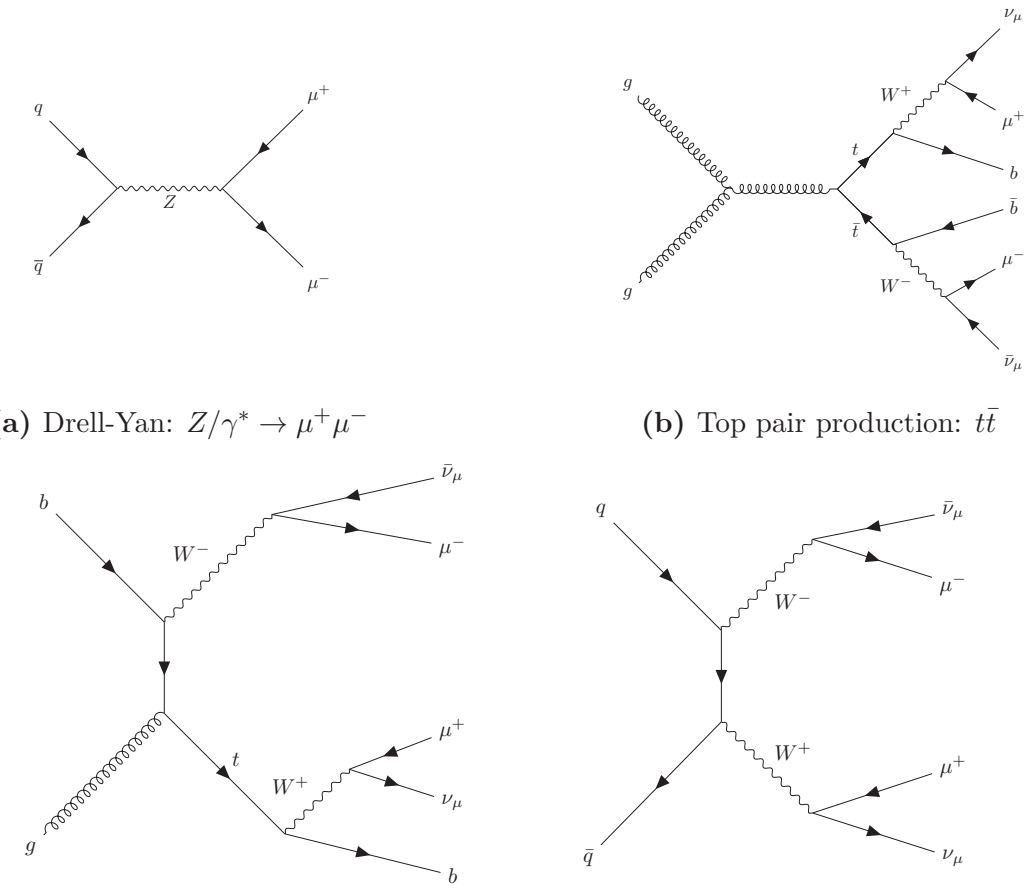
There are four significant contributions to the SM background in the $\mu^+\mu^- jj$ channel. These are Drell-Yan ($Z/\gamma^* \to \mu^+\mu^-$), top pair production ($t\bar{t}$), single top production in association with a W boson (Wt) and di-boson production (W^+W^-), where top quarks decay leptonically to muons. An example of the production of each component of the background is shown in Fig. 5.2.1. Other sources of background include misidentified muons from W+jets, single top production in the s-channel and t-channel, or multi-jet events. These form a negligible component of the background in comparison and therefore we treat Drell-Yan, top pair production, single top and di-boson production as the only sources of background.

To contribute to the $\mu^+\mu^- jj$ signature, Drell-Yan and di-boson production require the addition of at least two jets from initial and final state QCD radiation. Similarly at least one extra jet must be added to single top production. To account for this we generate events from processes of a range of different jet multiplicities according to the following definitions:

$$\begin{aligned} &\text{DY} + \text{0,1,2,3 jets}, \quad && t\bar{t} + \text{ 0,1 jets}, \\ &Wt + \text{0,1 jets}, && W^+W^- + \text{ 0,1,2 jets}. \end{aligned} \tag{5.2.1}$$

We include processes with less than two final state jets at parton level to account for the possibility that sufficiently hard jets may be produced by the parton shower algorithm. We use MLM matching [291] to match the final state partons generated from matrix elements in `Madgraph5` to those produced by parton showering in `Pythia8`. This removes overcounting between multi-jet final states and accounts for the fact that while soft and collinear jets are well described by the parton shower, the matrix elements

are more suited to simulating hard and well-separated partons. MLM matching is implemented in `Madgraph5` by specifying a nonzero value of the jet cut-off *xqcut* to be approximately 1/3 times a hard scale in the process for each component of the SM background and for each collider. We confirm our choice of *xqcut* value in each case by checking that the differential jet rate distributions are smooth and that observables such as the total cross section are insensitive to changes in *xqcut* about the chosen value.



(a) Drell-Yan: $Z/\gamma^* \to \mu^+\mu^-$

(b) Top pair production: $t\bar{t}$

(c) Single top production: Wt

(d) Diboson production: W^+W^-

Fig. 5.2.1. There are four contributions to the SM background of the $\mu^+\mu^- jj$ signal. Here we show a representative Feynman diagram for each process.

Interference

Interference arises between WW + 2 jets, Wt + 1 jet and $t\bar{t}$ production: all three processes may produce a $WWb\bar{b}$ final state via $t\bar{t}$ production. Ideally we would simulate

all contributions to the $WWb\bar{b}$ final state simultaneously, but this was found to be very computationally difficult and impractical. Instead, we generate events from each process separately and then combine them to produce the SM background. This means we must define each process in our simulations in such a way that any overcounting is removed.

A number of methods have been suggested to define Wt+1 jet production such that large contributions from $t\bar{t}$ diagrams are removed [292]. We use the diagram removal method as it is the most straightforward to implement in `Madgraph5` [293]. Let us denote the amplitude for Wt+1 jet by $\mathcal{A}_{Wt}$. We can write this as $\mathcal{A}_{Wt} = \mathcal{A}_1 + \mathcal{A}_2$ where $\mathcal{A}_2$ is the amplitude for all diagrams containing $t\bar{t}$ production. Double counting arises from the appearance of $|\mathcal{A}_2|^2$ in $|A_{Wt}|^2 = |\mathcal{A}_1|^2 + |\mathcal{A}_2|^2 + 2\mathrm{Re}(A_1^\dagger A_2)$. Diagram removal is implemented by setting $\mathcal{A}_2 = 0$ in our definition of the Wt+1 jet process, which removes the double counting. Although this method also neglects the interference term $2\mathrm{Re}(\mathcal{A}_1^\dagger \mathcal{A}_2)$, it has been shown that the effect of this on observables is moderate and that this method approximates Wt production well. We will validate this choice by comparing our simulations to data in §5.2.2. The violation of gauge invariance in the diagram removal method is found to have no observable effect [292].

The production of a W^+W^-jj final state in the di-boson channel features overcounting as a result of interference with both $t\bar{t}$ and Wt+1 jet production. This happens only when the two jets originate from b quarks. In our simulations we remove this interference by treating the b quarks as massive and neglecting them from the definitions of the proton and jets i.e. by working in a 4-flavour scheme. This is the method used by ATLAS in their simulations at 13 TeV [294]. 4-flavour parton distribution functions are used. This removes all $t\bar{t}$ and Wt+1 jet production from the di-boson channel, but also neglects processes with initial and final state b quarks which do contribute to di-boson production. A study of how well this 4-flavour scheme approximates the full di-boson production cross section at centre of mass energies $\sqrt{s} = 14, 100$ TeV was undertaken in Ref. [295] by comparing the leading order cross sections of di-boson production in the 4 and 5-flavour schemes, where it was found that the difference is negligible at 14 TeV and $\sim 5\%$ at 100 TeV.

Detector simulations

To produce an accurate simulation of the SM background at each future collider, we use `Delphes3` to simulate the response of the detectors and the decay of short-lived particles. Jets are clustered using the anti-k_T clustering algorithm [296] with jet radius $R = 0.4$. This value is chosen from the ATLAS analysis at 13 TeV [287] to reproduce

their analysis as closely as possible. To mimic the response of different detectors at each future collider we specify detector configurations as follows. The ATLAS configuration is used in all simulations at 13 TeV. At 14 TeV and 27 TeV we use the `Delphes3` HL-LHC configuration designed to reproduce the average response of the ATLAS and CMS detectors at high energies and luminosities. Similarly in our simulations of the 100 TeV FCC-hh we use the FCC-hh configuration provided by `Delphes3`.

We maintain the default settings in our simulations except in the case of muon isolation. Muon isolation is defined by finding the sum of the transverse momentum p_T of all objects within a cone of radius $R^{\max}$ around a muon, excluding the p_T of the muon itself. If the sum satisfies $p_T^{sum} < p_T^{\max}$ for fixed $p_T^{\max}$, the muon is considered isolated. At 13 TeV and 14 TeV we select only isolated muons with $p_T^{\max} = 0.2$ GeV and $R^{\max} = 0.2$, choosing these parameters to reproduce the 13 TeV ATLAS analysis. At 27 TeV and 100 TeV we make no selection on the muon isolation criteria, following the same reasoning as in [78]. This choice is made because the overall normalisation of the SM background is found to be very dependent on the muon isolation criteria and the specific selection made will likely vary in different future analyses. Relative to our simulations, any selection on muon isolation at future experiments will only reduce the SM background producing a better sensitivity to the LQ signal.

Event reweighting

We are interested in the search for TeV-scale LQs, which are expected to manifest as a resonance at high $m_{\min}(\mu, j)$. Producing a large number of events in the tail of the $m_{\min}(\mu, j)$ distribution is therefore necessary to achieve good statistics in this region. We find that binning the generation of events in $m_{\min}(\mu, j)$ at parton-level or in parameters such as the dimuon invariant mass $M_{\mu\mu}$ and $H_T = p_T^{j_1} + p_T^{j_2}$ is inefficient for producing a sufficient number of tail events. Instead we reweight the generation of each event x by applying a bias $b(x) \propto s(x)^5$. For each SM background process $s(x)$ is defined at parton-level as the invariant mass of the final state muons and jets, where we only include the minimum number of jets in the multi-jet process definitions of Equation 5.2.1, accounting for jets originating from top quarks. For example, for Drell-Yan we define $s(x)$ as the invariant mass of the di-muon final state. All physical observables and distributions shown in this Chapter have been obtained by unweighting the events after parton showering and detector simulation, in order to remove the effect of this bias.

5.2.2 Validation of Standard Model background simulations

We first validate our methods by simulating the SM background at $\sqrt{s} = 13$ TeV for an integrated luminosity $\mathcal{L} = 3.2$ fb^{-1} and comparing with the ATLAS search for second generation LQs at the same centre of mass energy and integrated luminosity [287]. We compare our simulations to the ATLAS data in two regions of phase space: the preselection region and the signal region. Both are defined by cuts on p_T, $|\eta|$ and ΔR designed to increase the significance of a LQ signal above the SM background and are summarised in Table 5.2.1. All jet cuts are placed on the two hardest jets in the event, denoted by j_1, j_2. The signal region is subject to further cuts on $S_T = p_T^{\mu_1} + p_T^{\mu_2} + p_T^{j_1} + p_T^{j_2}$ and $M_{\mu\mu}$. These significantly reduce features of the SM background due to soft jets and W and Z boson resonances. In both the signal and preselection regions we also reject muons falling in the range $1.01 < |\eta| < 1.1$ as specified by the ATLAS analysis to avoid potential p_T mismeasurement in this range. A preliminary selection on muon isolation is made at the level of detector simulation as outlined in §5.2.1.

Region	p_T^j (GeV)	p_T^μ (GeV)	$\lvert\eta_\mu\rvert$	$\lvert\eta_j\rvert$	$\Delta R_{\mu j}$	$\Delta R_{\mu\mu}$	$M_{\mu\mu}$ (GeV)	S_T (GeV)
Preselection	> 50	> 40	< 2.5	< 2.8	> 0.4	> 0.3		
Signal	> 50	> 40	< 2.5	< 2.8	> 0.4	> 0.3	> 130	> 600

Table 5.2.1. Phase space cuts defining the preselection and signal regions at $\sqrt{s} = 13$ TeV. All cuts are applied in the analysis after parton showering and detector simulation.

To efficiently simulate events in these regions of phase space, we generate events subject to a subset of the phase space cuts. These are applied at parton-level in the `Madgraph5` run card and summarised in Table 5.2.2. We will refer to this subset as generator cuts. The jet cut off *xqcut* required for MLM matching is found for each process in the presence of the generator cuts, as outlined in §5.2.1. Note however that we set the parameter `auto_ptj_mjj = True` for DY and di-boson production, allowing the jet matching procedure to automatically set the cuts on p_T^j and $M_{j_1 j_2}$ equal to the chosen value of *xqcut*. We use $xqcut = 30, 60, 60, 30$ GeV for DY, top pair, single top and di-boson production, respectively.

$p_T^{j_1}$ (GeV)	p_T^μ (GeV)	$\lvert\eta_\mu\rvert$	$\lvert\eta_j\rvert$	$\Delta R_{\mu j}$	$\Delta R_{\mu\mu}$	$M_{\mu\mu}$ (GeV)
> 35	> 30	< 2.5	< 2.8	> 0.4	> 0.3	> 20 (preselection) > 120 (signal region)

Table 5.2.2. Cuts applied at parton-level to efficiently simulate events at $\sqrt{s} = 13$ TeV.

Fig. 5.2.2 shows the distribution of preselection events in the parameter $M_{\mu\mu}$. The Monte Carlo (MC) error on each bin is shown in grey and is computed from the MC event weights w_i by $\text{Err}_i = \sqrt{\sum w_i^2}$. Systematic uncertainties are not included. Our simulations are not in perfect agreement with the ATLAS data (simulations) shown in black (green). This is expected because we generate all events at leading order, and the dominant process in this region is $t\bar{t}$ production which has large NLO corrections. However, our simulations provide a good estimate of the order of magnitude of the SM background in each bin.

Our methods are further validated in Fig. 5.2.3 which shows the distribution of signal region events in the parameter $m_{\text{min}}(\mu, j)$. As in the preselection region, we underestimate the SM background slightly by working only at leading order. However, compared to the preselection region, this provides a less fair comparison as normalisation factors have been applied to rescale to the ATLAS simulations in the signal region. Overall we take this comparison as a validation of our methods for simulating the SM background.

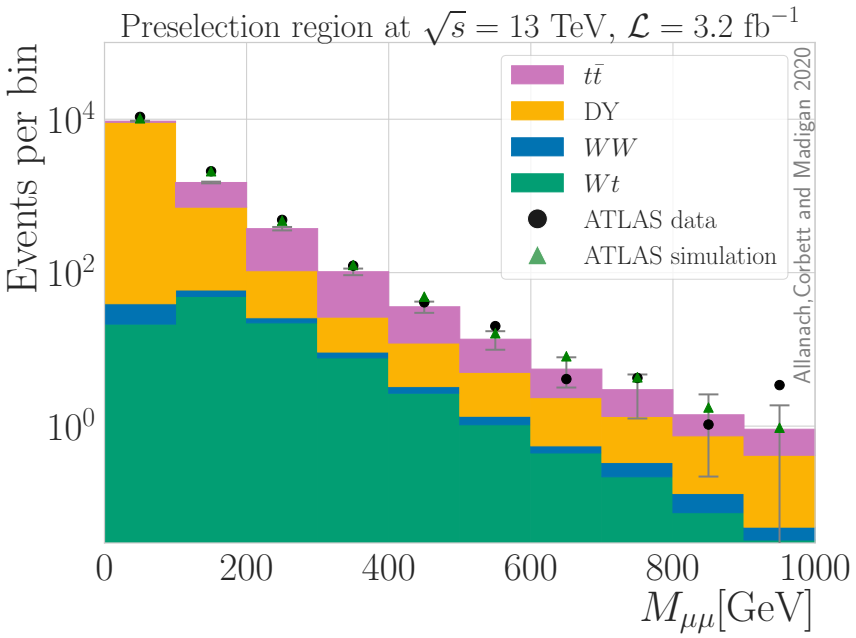


Fig. 5.2.2. Validation plot showing our simulations of the $M_{\mu\mu}$ distribution of the SM backgrounds in the search for the pair production of second generation LQs in the $\mu^+\mu^- jj$ channel at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} in the preselection region. We compare our simulations to the ATLAS simulations and data for validation.

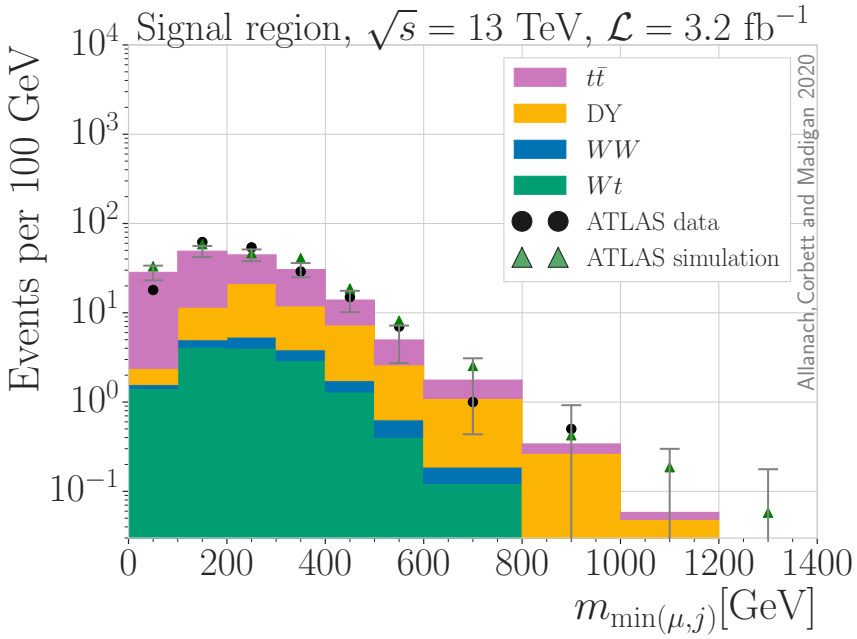


Fig. 5.2.3. Validation plot showing our simulations of the $m_{\min}(\mu, j)$ distribution of the SM backgrounds in the search for the pair production of second generation LQs in the $\mu^+\mu^- jj$ channel at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} in the signal region. We compare our simulations to the ATLAS simulations and data for validation.

5.2.3 Future collider Standard Model backgrounds

Collider	p_T^j (GeV)	p_T^μ (GeV)	$\lvert\eta_\mu\rvert$	$\lvert\eta_j\rvert$	$M_{\mu\mu}$ (GeV)	S_T (GeV)
LHC	> 50	> 40	< 2.5	< 2.8	> 130	> 600
HL-LHC	> 50	> 40	< 2.5	< 2.8	> 130	> 600
HE-LHC	> 100	> 80	< 4.0	< 4.0	> 260	> 1200
FCC-hh	> 400	> 320	< 4.0	< 4.0	> 1000	> 4000

Table 5.2.3. Phase space cuts defining the signal regions in simulations of the 13 TeV LHC and future colliders. All cuts are applied in the analysis after parton showering and detector simulation.

Having validated our simulation methodology, we now turn to future colliders. We generate the SM background at the LHC with the full Run II integrated luminosity (13 TeV, 140 fb^{-1}) and at the three future colliders introduced in Chapter 2, with details specified in Table 2.4.1. We define the signal region at each collider by a set of phase space cuts based on the 13 TeV ATLAS analysis as follows. All angular separations ΔR

are kept unchanged. For the LHC and HL-LHC we keep the same cuts on $|\eta|$, while at the HE-LHC and FCC-hh these are increased to $|\eta| < 4$ to allow for a difference in detector topologies at the future colliders. Cuts with dimensions of energy (p_T, S_T, $M_{\mu\mu}$) are kept constant for the HL-LHC and scaled up by approximately the ratio of centre of mass energies $\sqrt{s}/(13\text{ TeV})$ for the HE-LHC and FCC-hh simulations. These signal region cuts are summarised in Table 5.2.3. As in our 13 TeV simulations, we generate events subject to generator cuts applied at parton-level in the `Madgraph5` run card, summarised in Table 5.2.4. Table 5.2.5 specifies the values of *xqcut* used in MLM matching for each component of the SM background.

Collider	$p_T^{j_1}$ (GeV)	p_T^{μ} (GeV)	$\lvert\eta_\mu\rvert$	$\lvert\eta_j\rvert$	$M_{\mu\mu}$ (GeV)
LHC	> 35	> 30	< 2.5	< 2.8	> 130
HL-LHC	> 35	> 30	< 2.5	< 2.8	> 130
HE-LHC	> 85	> 60	< 4.0	< 4.0	> 200
FCC-hh	> 380	> 300	< 4.0	< 4.0	> 900

Table 5.2.4. Phase space cuts applied at parton-level in `Madgraph5` to efficiently simulate events in the signal region for the 13 TeV LHC and future colliders.

Figs. 5.2.4, 5.2.5, 5.2.6 and 5.2.7 show the resulting distributions of the SM background at the LHC, HL-LHC, HE-LHC and FCC-hh respectively. As before the MC error is shown in grey and systematic uncertainties are not included.

Collider	DY	$t\bar{t}$	Wt	W^+W^-
LHC	30	60	30	30
HL-LHC	30	60	30	30
HE-LHC	45	120	120	60
FCC-hh	90	300	120	120

Table 5.2.5. Values of the jet cut-off parameter *xqcut* in GeV used in MLM matching of multi-jet events at the 13 TeV LHC and future colliders. All jet matching parameters are found in the presence of the generator cuts summarised in Table 5.2.4.

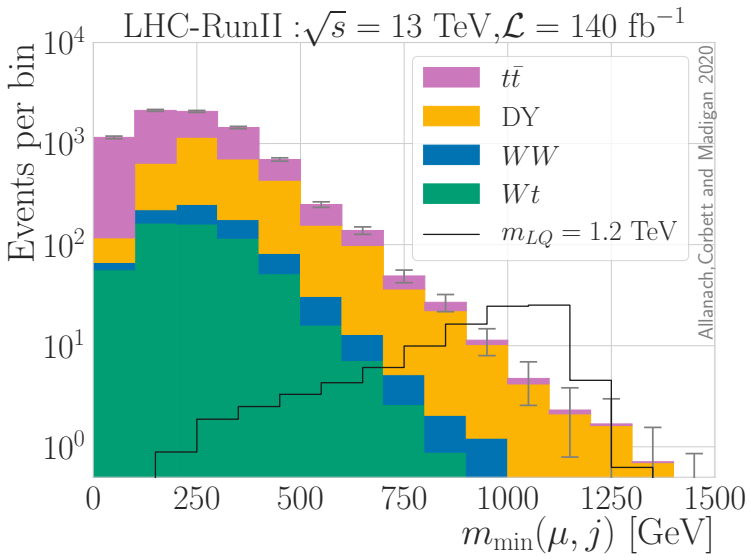


Fig. 5.2.4. Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 13 TeV LHC with full Run II integrated luminosity of $\mathcal{L} = 140$ fb^{-1}. The LQ signal corresponds to $m_{\rm LQ} = 1.2$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 5σ relative to the SM background.

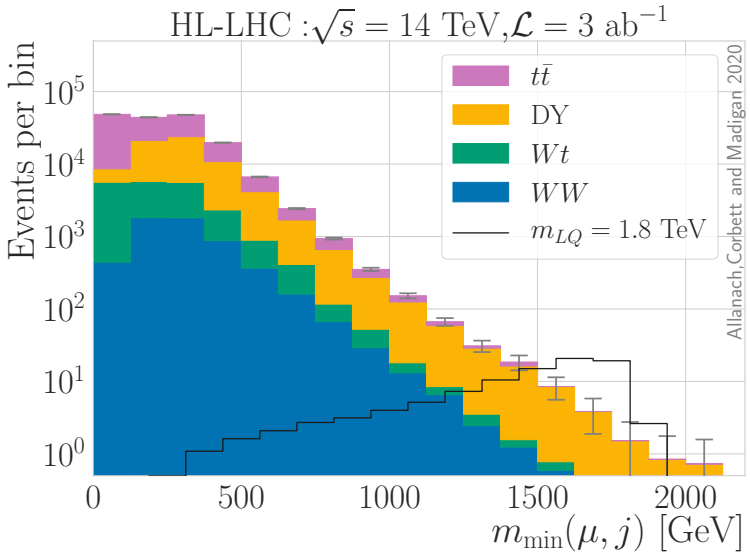


Fig. 5.2.5. Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 14 TeV HL-LHC with $\mathcal{L} = 3$ ab^{-1}. The LQ signal corresponds to $m_{\rm LQ} = 1.8$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 7σ relative to the SM background.

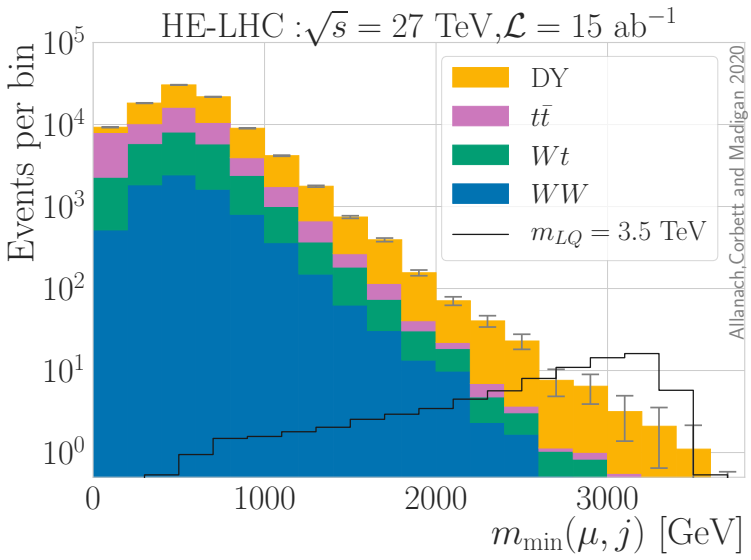


Fig. 5.2.6. Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 27 TeV HE-LHC with $\mathcal{L} = 15$ ab^{-1}. The LQ signal corresponds to $m_{\mathrm{LQ}} = 3.5$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 6σ relative to the SM background.

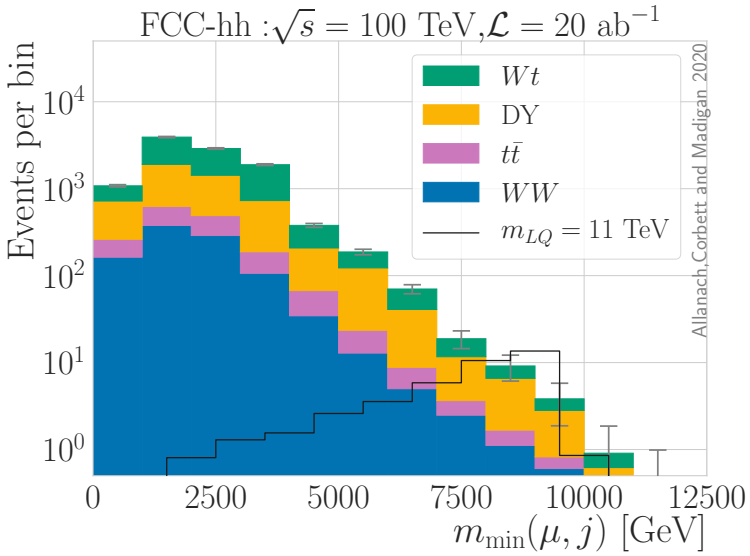


Fig. 5.2.7. Predicted $m_{\min}(\mu, j)$ distribution of the SM background and an example of a LQ signal at the 100 TeV FCC-hh with $\mathcal{L} = 20$ ab^{-1}. The LQ signal corresponds to $m_{\mathrm{LQ}} = 11$ TeV and couplings chosen to fit the NCBAs as outlined in §5.3.1, and has a statistical significance of 3σ relative to the SM background.

5.3 Projections of collider sensitivity

5.3.1 Leptoquark signal simulations

To find the significance of a LQ at mass m_{LQ} relative to the simulated SM backgrounds, we simulate the distribution of a LQ resonance in $m_{\mathrm{min}}(\mu, j)$. We simulate LQ pair production and decay into a $\mu^+\mu^- jj$ final state at leading order[1] in `Madgraph5`. We work in a 5-flavour scheme using the 5-flavour `NNPDF2.3LO` [165] parton distribution function, and correct for parton showering and detector effects using `Pythia8` and `Delphes3` as before. We simulate LQs from the S_3 LQ model provided by Ref.[2] [302].

We specify LQ couplings as follows. For each m_{LQ} the product of couplings $|(Y_{de})_{32}(Y^*_{de})_{22}|$ is fixed by fits to the NCBAs as given in Equation 5.1.2. We choose $(Y_{de})_{22} = (Y_{de})_{32}$ and set all other $(Y_{de})_{ij}$ to zero. This couples the LQ to $b_L\mu_L$ and $s_L\mu_L$ pairs as required by the NCBAs. Following the conventions of Ref. [126], all CKM and PMNS mixing occurs within the up and neutrino sector respectively i.e. we set $V_{d_L} = V_{e_L} = I$ in Equation 2.5.7. Our choice of couplings then corresponds to setting $(Y_l)_{22} = (Y_l)_{32} \neq 0$, inducing further couplings of the LQ to $\mathbf{u}_L\mu_L$, $b_L\boldsymbol{\nu}_L$, $s_L\boldsymbol{\nu}_L$ and $\mathbf{u}_L\boldsymbol{\nu}_L$ pairs where $\mathbf{u}_L$ and $\boldsymbol{\nu}_L$ denote the vectors of left-handed up-type quarks and neutrinos respectively. These include CKM suppressed couplings to $u_L\mu_L$ and $c_L\mu_L$ which will contribute to the muon-jet decay channel of the LQ, increasing the number of events in the $\mu^+\mu^- jj$ signal.

We account for these additional CKM and PMNS suppressed couplings when calculating the theory predictions for $\sigma \times \mathrm{BR}$. The partial width of the decay $\mathrm{LQ} \to lq$ parametrised by the coupling y_{lq} is given to leading order in α_s by [303]

$$\Gamma = \frac{|y_{lq}|^2 m_{\mathrm{LQ}}}{16\pi}. \tag{5.3.1}$$

The branching ratios of the decays of each S_3 component $S_3^{\frac{1}{3}}$, $S_3^{-\frac{2}{3}}$ and $S_3^{\frac{4}{3}}$ into μj final states (where j denotes all u, d, s, c, b quarks and antiquarks) are then given by

[1] The LQ pair production signal at hadron colliders was calculated some time ago at leading order [297, 298]. Next-to-leading order effects [299, 300] and parton shower effects [301] on the signal have also been studied.

[2] A link to the `UFO` model files can be found within this reference.

$$\begin{aligned}
\mathrm{BR}(\mathrm{S}_3^{\frac{1}{3}} \to \mu^+\mathrm{j}) &= \frac{|V_{21}+V_{31}|^2+|V_{22}+V_{32}|^2}{|V_{21}+V_{31}|^2+|V_{22}+V_{32}|^2+|V_{23}+V_{33}|^2+2|U_{21}|^2+2|U_{22}|^2+2|U_{23}|^2} \\
&= \mathrm{BR}(\bar{\mathrm{S}}_3^{\frac{1}{3}} \to \mu^-\mathrm{j}), \\
\mathrm{BR}(\mathrm{S}_3^{-\frac{2}{3}} \to \mu\mathrm{j}) &= 0, \\
\mathrm{BR}(\mathrm{S}_3^{\frac{4}{3}} \to \mu^+\mathrm{j}) &= \mathrm{BR}(\bar{\mathrm{S}}_3^{\frac{4}{3}} \to \mu^-\mathrm{j}) = 1.
\end{aligned} \tag{5.3.2}$$

Note that the branching ratios are independent of $(Y_{de})_{32}$, $(Y_{de})_{22}$, and the LQ mass m_{LQ}. This feature is model dependent, relying on our choice of $(Y_{de})_{32} = (Y_{de})_{22}$ and our choice to set all other $(Y_{ed})_{ij} = 0$. We calculate the value of $\mathrm{BR}(\mathrm{S}_3^{\frac{1}{3}} \to \mu^+\mathrm{j})$ from the central values of V_{CKM} and U_{PMNS} [304] assuming normal ordering of neutrino masses. The pair production of scalar leptoquarks is dominated by $gg \to S_3\bar{S}_3$, as discussed in Section 2.5.3. This interaction is independent of any leptoquark couplings y_{lq}. As a result, the pair production cross section σ is independent of $(Y_{de})_{22}$, $(Y_{de})_{32}$.

As outlined in §5.2.2 and §5.2.3 we generate events subject to the generator cuts in Table 5.2.4 at parton-level, applying the full set of signal region phase space cuts from Table 5.2.3 in the analyses. Examples of the predicted distributions of LQ events at each future collider are included in Figs. 5.2.4, 5.2.5, 5.2.6 and 5.2.7, superimposed over the SM backgrounds.

5.3.2 Statistical analysis

Each experiment consists of measurements of events in N bins of a histogram, denoted by n_i where $i = 1, \ldots, N$. We find the expected number of background events b_i and signal events s_i in bin i from our MC simulations, and parametrise the signal present in our data sample by the signal strength $\mu \in [0, 1]$. The likelihood is defined by taking the product of Poisson probabilities in each bin

$$L(\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}, \tag{5.3.3}$$

where θ denotes all nuisance parameters. The profile likelihood ratio is defined as $\lambda(\mu) = L(\mu, \hat{\hat{\theta}})/L(\hat{\mu}, \hat{\theta})$ where $\hat{\mu}$ and $\hat{\theta}$ are the maximum likelihood estimates of μ and θ, and $\hat{\hat{\theta}}$ is found by maximising the likelihood with fixed μ.

Mass exclusion limits

Suppose the measured data n_i show no fluctuations above the SM background b_i. To set exclusion limits on $\sigma \times \mathrm{BR}$ we test the $b + \mu s$ hypothesis and find the maximum value of μ compatible with the data. We quantify compatibility by computing the p-value from the modified frequentist CL_s method [305] and the test statistic q_μ defined by

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{5.3.4}$$

The upper limit at 95% CL on μ, denoted by μ_{lim}, is then given by the value of μ at which $\mathrm{CL}_s = 0.05$. We compute the CL_s values using `pyhf` [306], a Python implementation of `HistFactory` [307].

Note that, as discussed in §5.3.1, both σ and the branching ratio are independent of LQ couplings y_{lq} given our choice of model. As a result, $\sigma \times \mathrm{BR}$ depends on the LQ mass m_{LQ} but not the LQ couplings y_{lq}. We translate μ_{lim} into a limit on $\sigma \times \mathrm{BR}$ by defining $(\sigma \times \mathrm{BR})_{lim} = \mu_{lim}(\sigma \times \mathrm{BR})$, and calculating $(\sigma \times \mathrm{BR})_{lim}$ as a function of m_{LQ}. We then compare this quantity with the theoretical predictions $(\sigma \times \mathrm{BR})_{\mathrm{theory}}$ as a function of m_{LQ}. Any value of m_{LQ} for which $(\sigma \times \mathrm{BR})_{\mathrm{theory}} > (\sigma \times \mathrm{BR})_{\mathrm{lim}}$ can then be excluded: at these values of m_{LQ}, our predictions show that $\sigma \times \mathrm{BR}$ is expected to be larger than the upper limit we expect to observe. The point of intersection between the curves $(\sigma \times \mathrm{BR})_{lim}$ and $(\sigma \times \mathrm{BR})_{theory}$ determines the mass sensitivity i.e. the maximum m_{LQ} that could be excluded at 95% CL at each future collider.

Discovery reach

Alternatively, suppose an excess of events is seen in the data n_i. To find the significance of such an observation we test the compatibility of the background-only hypothesis $\mu = 0$ with the data. The test statistic q_0 is defined by

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \geq 0 \\ 0 & \hat{\mu} < 0. \end{cases} \tag{5.3.5}$$

The discovery reach of each future collider is found by determining, for each m_{LQ} of interest, the integrated luminosity $\mathcal{L}$ required for a p-value of $\mathrm{CL}_s = 2.9 \times 10^{-7}$ or equivalently a statistical significance of 5σ.

In determining the discovery and exclusion sensitivities, we work in the large sample approximation and use the Asimov data set to calculate the median CL_s [73].

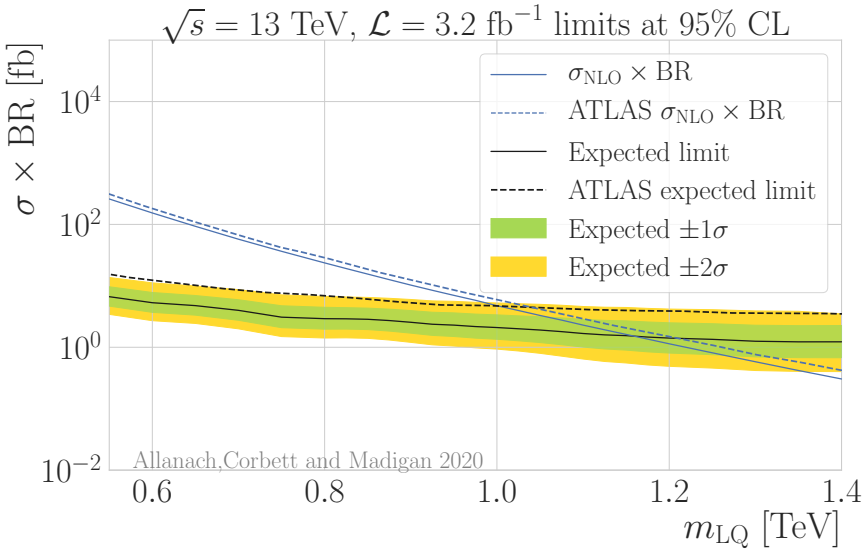


Fig. 5.3.1. Validation plot comparing our expected limits at 95% CL on $\sigma \times$ BR for LQ pair production and decay into a $\mu^+\mu^- jj$ final state at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} to the expected limits obtained by ATLAS.

5.3.3 Validation

We validate the method of calculating exclusion limits by using the signal region data generated at $\sqrt{s} = 13$ TeV, $\mathcal{L} = 3.2$ fb^{-1} in Fig. 5.2.3 to place limits on LQs in the range $m_{\mathrm{LQ}} \in [500, 1400]$ GeV. The resulting limits on $\sigma \times$ BR as a function of m_{LQ} are shown in Fig. 5.3.1, excluding LQ masses up to approximately 1.15 TeV. This limit is compared to the exclusion limits found by ATLAS, shown by the black dashed curve, indicating sensitivity to LQ masses up to 1.05 TeV. Note that for the purposes of this comparison only we generate events and compute the $\sigma \times$ BR from a model of second generation LQs decaying into a $\mu^-\mu^+ c\bar{c}$ final state with coupling $y_{\mu c} = \sqrt{0.01 \times 4\pi\alpha_{em}}$ from the minimal Buchmüller-Rückl-Wyler model [130], following the ATLAS 13 TeV analysis. All other LQ events and values of $\sigma \times$ BR in this Chapter are found as outlined in §5.3.1 and §5.4.1.

This shows that our methods have slightly overestimated the sensitivity to LQ. This is to be expected from the fact that we have underestimated the SM background and do not include systematic uncertainties in setting limits. However, as an estimate

of the sensitivity this is a good approximation, and so we take this comparison as a validation of our methods and proceed by using the same methods for future colliders.

5.4 Results

The resulting limits on $\sigma \times \mathrm{BR}$ as a function of m_{LQ} are shown in Figs. 5.4.1 and 5.4.2 for the LHC, HL-LHC, HE-LHC and FCC-hh at design integrated luminosities. We compare our limits with theory predictions for $\sigma \times \mathrm{BR}$, shown by the blue curves, and determine the mass to which each collider is sensitive from the point of intersection. We see that while the sensitivity will be increased up to $m_{\mathrm{LQ}} = 1.75$ TeV and eventually $m_{\mathrm{LQ}} = 2.5$ TeV by the LHC Run II and HL-LHC respectively, the HE-LHC and FCC-hh have the potential to explore a much larger range of LQ parameter space, excluding masses up to $m_{\mathrm{LQ}} = 4.8$ TeV and 13.5 TeV respectively.

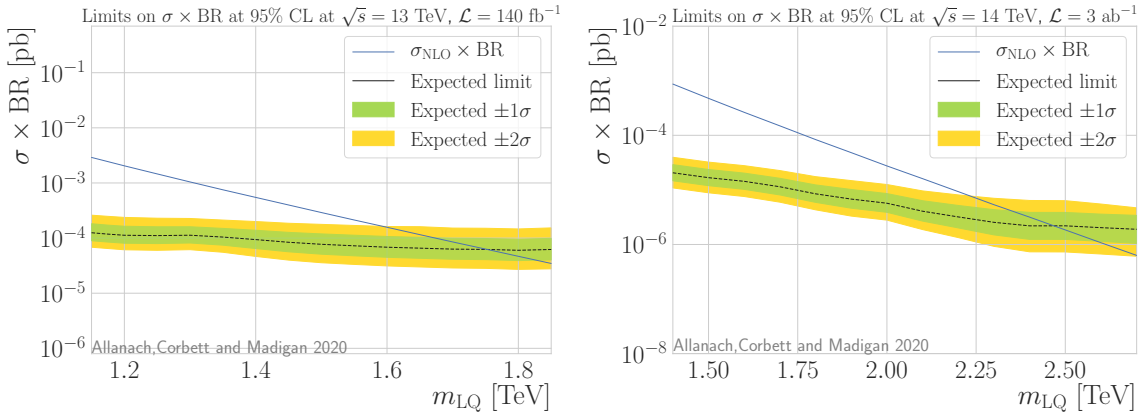


Fig. 5.4.1. Expected limits at 95% CL on $\sigma \times \mathrm{BR}$ for the pair production of LQs and decay into a $\mu^+\mu^- jj$ final state at the LHC with full Run II integrated luminosity (left) and the HL-LHC (right). Theory curves $\sigma_{\mathrm{NLO}} \times \mathrm{BR}$ are calculated for narrow width LQs with couplings chosen to fit the NCBAs.

To further investigate the potential of future colliders to exclude high-mass LQs, we scan over a range of integrated luminosities up to $\mathcal{L} = 3, 15$ and 20 ab^{-1} for the HL-LHC, HE-LHC and FCC-hh respectively and determine the mass sensitivity at 95% CL for each. Similarly, we perform a scan over integrated luminosities and determine the discovery reach of each future collider. These results are shown in Fig. 5.4.3. In both plots the filled circles correspond to the design integrated luminosities of $\mathcal{L} = 3, 15$ and 20 ab^{-1}. The highest m_{LQ} that can be observed with a 5σ significance is $m_{\mathrm{LQ}} = 9.5$ TeV: we predict that narrow width scalar LQs could be discovered at this

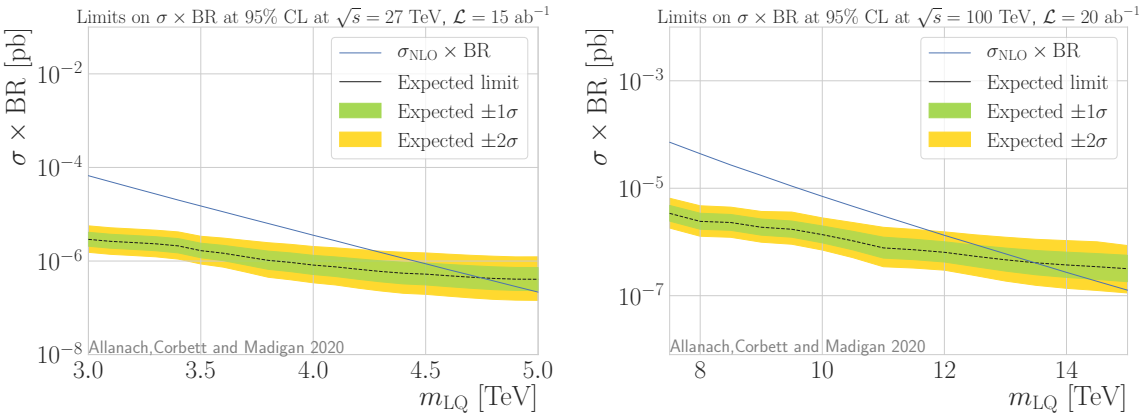


Fig. 5.4.2. Expected limits at 95% CL on $\sigma \times \mathrm{BR}$ for the pair production of LQs and decay into a $\mu^+\mu^- jj$ final state at the HE-LHC (left) and the FCC-hh (right). Theory curves $\sigma_{\mathrm{NLO}} \times \mathrm{BR}$ are calculated for narrow width LQs with couplings chosen to fit the NCBAs.

mass assuming the FCC-hh operates at the full $\mathcal{L} = 20$ ab^{-1}. Similarly, the HE-LHC and HL-LHC have the potential to observe narrow width scalar LQs of masses up to $m_{\mathrm{LQ}} = 3.6$ TeV and 1.9 TeV respectively. Finally, we compute the discovery reach of the LHC Run II with $\sqrt{s} = 13$ TeV, $\mathcal{L} = 140$ fb^{-1} to be $m_{\mathrm{LQ}} = 1.2$ TeV, right on the edge of the 95% exclusion limits already found by the 13 TeV LHC as discussed in §5.1.

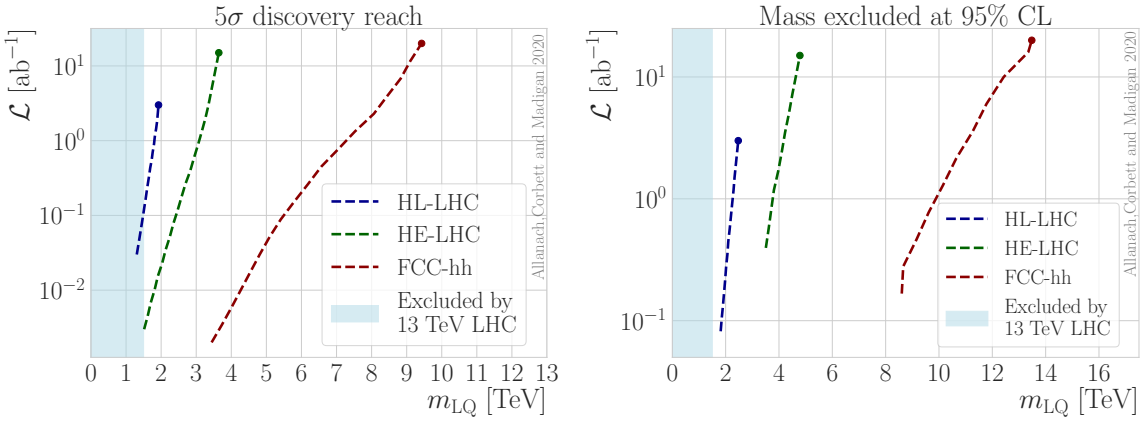


Fig. 5.4.3. Predicted 5σ discovery reach (left) and mass exclusion at 95% CL (right) of the HL-LHC, HE-LHC and FCC-hh. Points correspond to the design integrated luminosities of each future collider of $\mathcal{L} = 3, 15, 20$ ab^{-1} respectively.

5.4.1 Wide resonances

Before concluding, we note that the $m_{\rm LQ}$ exclusion limits of Figs. 5.4.1 and 5.4.2 may be applied to scalar LQ models which are not solutions to the NCBAs, as long as the LQ width is sufficiently narrow. In general, if we depart from the NCBA limit the LQs may acquire an appreciable and non-negligible width. We wish to estimate how big the effect might be on the resulting sensitivity. Our SM background simulations and statistical methods can be applied to determine the approximate change in sensitivity to wider LQ resonances, where some couplings $|Y_{ql}|$ may be large,[3] with one important caveat: we do not include interference between signal and background. Thus, our final estimate of sensitivity will be an over-estimate because we may expect the signal to be broadened further by signal-background interference effects. Our purpose however, is just to see the approximate shift in sensitivity rather than to provide a true and accurate calculation of the sensitivity itself. We shall see that the sensitivity is not drastically changed by including large width effects, and we expect that this qualitative conclusion holds once signal-background interference effects have been included.

The partial decay width of a LQ into a lepton l and quark q is related to the mass $m_{\rm LQ}$ and coupling Y_{ql} by Eq. 5.3.1. Given the choice of couplings in our signal simulations as outlined in §5.3.1, we have so far only considered narrow LQ resonances satisfying $\Gamma/m_{\rm LQ} < 0.01$. However, any narrow width LQ will still produce a wide resonance in the distribution of $m_{\rm min}(\mu, j)$ as shown in Figs. 5.2.4, 5.2.5, 5.2.6 and 5.2.7. This is an effect of changes in the kinematics of the final state particles due to parton showering and detector resolution, as well as the ambiguity in defining $m_{\rm min}(\mu, j)$. For narrow width LQs, the width of this resonance is determined only by the experimental resolution $\Gamma_{\rm res}$. By fitting a normal distribution to the resonances of Figs. 5.2.4 - 5.2.7 and approximating $\Gamma_{\rm res}$ by twice the standard deviation, we estimate the $\Gamma_{\rm res}$ to be $\Gamma_{\rm res} = 0.1\ m_{\rm LQ}$. To investigate the effects of wide resonances we then simulate LQ events with decay width $\Gamma \geq \Gamma_{\rm res}$. We do this by switching on the same couplings $(Y_{de})_{22} = (Y_{de})_{32} \neq 0$ as before, determining their values from Equation 5.3.1 for benchmark values of $\Gamma = 0.1, 0.2$ and 0.5.

Fig. 5.4.4 (left) compares our simulations of the $m_{\rm min}(\mu, j)$ distributions of large width LQs at $m_{\rm LQ} = 3.2$ TeV and $\Gamma/m_{\rm LQ} = 0.1, 0.2$ and 0.5. These simulations are performed assuming the HE-LHC collider *i.e.* assuming $\sqrt{s} = 27$ TeV, $\mathcal{L} = 15$ ab^{-1}. Fig. 5.4.4 (right) shows the corresponding expected limits on $\sigma \times$ BR at 95% CL.

[3] If any of the $|Y_{ql}|$ involving quarks from the first two families are large, single LQ production, which is beyond the scope of the present Chapter, may also prove a profitable search channel [285, 308].

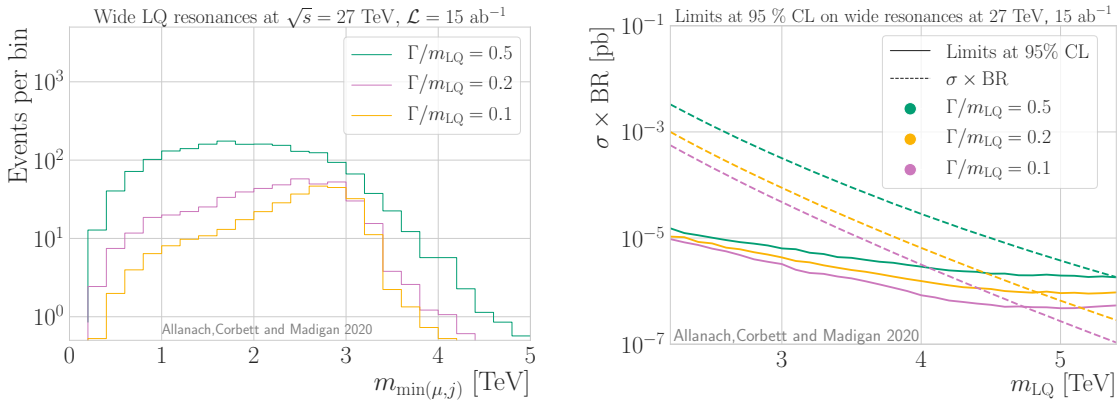


Fig. 5.4.4. Comparison of the predicted $m_{\min}(\mu, j)$ distribution of LQ signal events for a LQ with large decay width Γ (left), and expected limits at 95% CL on $\sigma \times$ BR for the pair production of wide LQs decaying into a $\mu^+\mu^- jj$ final state at $\sqrt{s} = 27$ TeV, $\mathcal{L} = 15$ ab^{-1} (right).

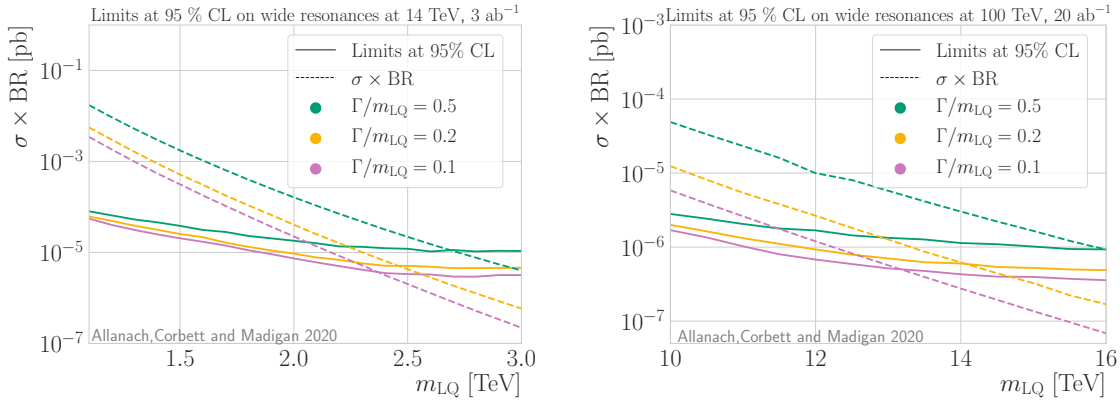


Fig. 5.4.5. Expected limits at 95% CL on $\sigma \times$ BR for the pair production of LQs with large decay width Γ, decaying into a $\mu+\mu-jj$ final state at $\sqrt{s} = 14$ TeV, $\mathcal{L} = 3$ ab^{-1} (left) and $\sqrt{s} = 100$ TeV, $\mathcal{L} = 20$ ab^{-1} (right).

To provide a sample estimate of sensitivity we compare our limits to the values of $\sigma_{\text{LO}} \times$ BR calculated for a wide LQ signal, as shown by the dashed curves.

Fig. 5.4.4 (left) shows that the distribution of signal events spreads out in $m_{\min}(\mu, j)$ with increasing Γ. We expect that the sensitivity to LQs is decreased as a result of the signal events spreading out in this way rather than being peaked around a few bins. This effect is seen in the increase in the upper limits on $\sigma \times$ BR with increasing Γ in Fig. 5.4.4 (right). We can see from the intersection of the $\Gamma/m_{\text{LQ}} = 0.1$ theory curve (purple, dashed) with each set of expected limits (solid curves) that if the theory

predictions for $\sigma \times \mathrm{BR}$ were independent of LQ couplings $(Y_{de})_{22}$ and $(Y_{de})_{32}$, an increase in LQ width from $\Gamma/m_{\mathrm{LQ}} = 0.1$ to $\Gamma/m_{\mathrm{LQ}} = 0.5$ would result in a loss of sensitivity from approximately $m_{\mathrm{LQ}} = 4.8$ TeV to $m_{\mathrm{LQ}} = 4$ TeV. However, this effect is mitigated by the fact that at such large couplings, pair production is no longer dominated by gluon-gluon interactions. Instead, pair production via quark-lepton interactions has a significant contribution to the total cross section. As a result, the theory prediction for $\sigma \times \mathrm{BR}$ depends strongly on the choice of couplings $(Y_{de})_{22}$ and $(Y_{de})_{32}$. This can be seen by the overall increase in the number of signal events with Γ in Fig. 5.4.4 (left), and by the large increase in the values of $\sigma \times \mathrm{BR}$ with Γ in Fig. 5.4.4 (right). Overall this leads to an increase in sensitivity to LQs with increasing Γ. A similar effect is seen in our predictions for wide LQs at the HL-LHC and FCC-hh, as shown in Fig. 5.4.5. Were we to include signal-background interference effects in the calculation, we expect the sensitivity would be degraded. This leads us to conclude that the overall change in sensitivity from the larger widths is not dramatic and would remain small were we to include signal-background interference effects.

5.5 Concluding discussion

In this Chapter we have estimated the exclusion and discovery sensitivities of future hadron colliders to LQ pair production for the case that each LQ decays to a muon and a jet. Such a decay channel is motivated in part by the LQ solution to the NCBAs. It is also motivated by the fact (regardless of the NCBAs) that muons are empirically robust objects, which are good for tagging and beating down irreducible backgrounds. By concentrating on LQ pair production (rather than single LQ production, for example) we cover a large volume of model parameter space where LQs, being perturbatively coupled, are narrow and the pair production cross section varies only with the LQ mass m_{LQ}. For such LQs, their production is dominated by production from glue-glue interactions, their interactions with initial state quarks being negligible. This is typically true for LQs that have a coupling-mass relation consistent with the NCBAs, but we emphasise that our sensitivities extend beyond this coupling-mass relation more generally, as discussed below.

The previous estimate of the exclusion sensitivity in Ref. [285] extrapolated LHC search limits using two highly dubious approximations. The first approximation is that experimental efficiency and acceptance would not change with centre of mass energy, and the second that LQs are produced exactly at threshold. With respect to the first point, at large m_{LQ} and at high energies (particularly at FCC-hh), the decay

products from LQs will be highly boosted resulting in muons collinear to the jets, and therefore more muons failing isolation criteria. The muon momentum resolution is also likely to be very poor at higher energies, since such hard muons will only be bent to a limited extent by the magnets. This also affects signal efficiency due to peak broadening. Secondly, the assumption that LQs are produced exactly at threshold is likely to introduce large uncertainties.

Collider	$\sqrt{s}$ [TeV]	$\mathcal{L}$ [ab^{-1}]	5σ discovery reach [TeV]	Mass excluded at 95% CL [TeV]
LHC Run II	13	0.14	1.2	1.8
HL-LHC	14	3	1.9	2.5
HE-LHC	27	15	3.6	4.8
FCC-hh	100	20	9.5	13.5

Table 5.5.1. Summary of the expected 5σ discovery sensitivity and expected 95% CL exclusion sensitivity to S_3 for hadron colliders, from LQ pair production.

We rectify these two bad approximations in this Chapter by performing a fast simulation of the signal and detector response. We summarise our expected discovery and exclusion sensitivities in Table 5.5.1. Ref. [285] estimated that the HL-LHC could exclude 2.2 TeV at 95% CL, to be compared with 1.8 TeV. The HE-LHC was estimated to cover up to 4.1 TeV, but this was for a higher centre of mass energy (33 TeV) and a different luminosity (15 ab^{-1}), precluding a direct comparison. The FCC-hh exclusion sensitivity was calculated at an integrated luminosity of 10 ab^{-1} to be 12.0 TeV, to be compared with 12.5 TeV from our estimate (see Fig. 5.4.3). It is somewhat surprising that the comparable estimates are so similar, since as we have argued, the old ones were based on simple approximations. The results in Table 5.5.1 are on a much firmer footing.

As mentioned in §5.1, at the time of writing Ref. [3] on which this Chapter is based, the ATLAS search for scalar leptoquarks using 139 fb^{-1} of Run II data had not yet been published [133]. By searching in the $\mu^+\mu^- jj$ channel they exclude second generation scalar LQs with masses up to $m_{\mathrm{LQ}} = 1.7$ TeV. This constraint can be compared with our limit of 1.8 TeV shown in Table 5.5.1. Accounting for small differences arising from the model dependence of the limits, these constraints are in agreement, providing further validation of the methodology used in projections of the HL-LHC, HE-LHC and FCC-hh.

Fig. 5.5.1 displays a summary of our projections for the mass exclusion limits of the HL-LHC, HE-LHC and FCC-hh. The projections are shown as dashed vertical lines, labelled by their respective future colliders. The fact that these projections are

vertical, with no dependence on the value of $|(Y_{de})_{32}(Y_{de})_{22}|$, is a result of the fact that the mass exclusion limits are independent of LQ couplings given our choice of LQ model, as described in §5.3.2. If we had not assumed $(Y_{de})_{32} = (Y_{de})_{22}$, or if we allowed for nonzero $(Y_{de})_{ij}$ in general, the mass exclusion limits may have some dependence on the LQ couplings. This is also true of the constraints from searches for the pair production of scalar LQs by ATLAS and CMS shown in green. We compare these constraints to the constraints coming from searches for the single production of scalar LQs by CMS and a constraint from perturbative unitarity. We take our constraint from single LQ production from the recast of a CMS search at 8 TeV, 20 fb^{-1} [131] provided by Ref. [132]. Finally, in dark blue we indicate the values of couplings and LQ mass favoured by fits to the NCBAs, taking the $\pm 1\sigma$ range allowed by Eq. 2.5.9. We see that although our projections indicate that a reasonable range of LQ parameter space may be excluded by the FCC-hh, there remains a wide region at high LQ mass which will not be explored. If a LQ exists at this high scale, we may have to rely on indirect search methods for its detection.

Before concluding this Chapter, we note that the sensitivities phrased in terms of LQ mass have a dependence on the the model, since different choices of LQ model couplings may affect the BR of the muon-jet decay rate. However, all limits on $\sigma \times \mathrm{BR}$, such as the dashed curves in Fig. 5.4.1, also apply to narrow width LQ models with different (but still small) LQ couplings to fermions. Only when one or more of the LQ couplings approaches the non-perturbative régime does the LQ width become comparable to the experimental resolution, potentially affecting sensitivity. To cover this case, we considered a wider LQ: see §5.4.1. Of and by itself, the width does not change the sensitivity much. Increasing the width divided by mass of the LQ from 0.1 to 0.5 but keeping the cross section times branching ratio constant only results in a 10% degradation or so in FCC-hh mass reach, as the right-hand panel of Fig. 5.4.5 shows.

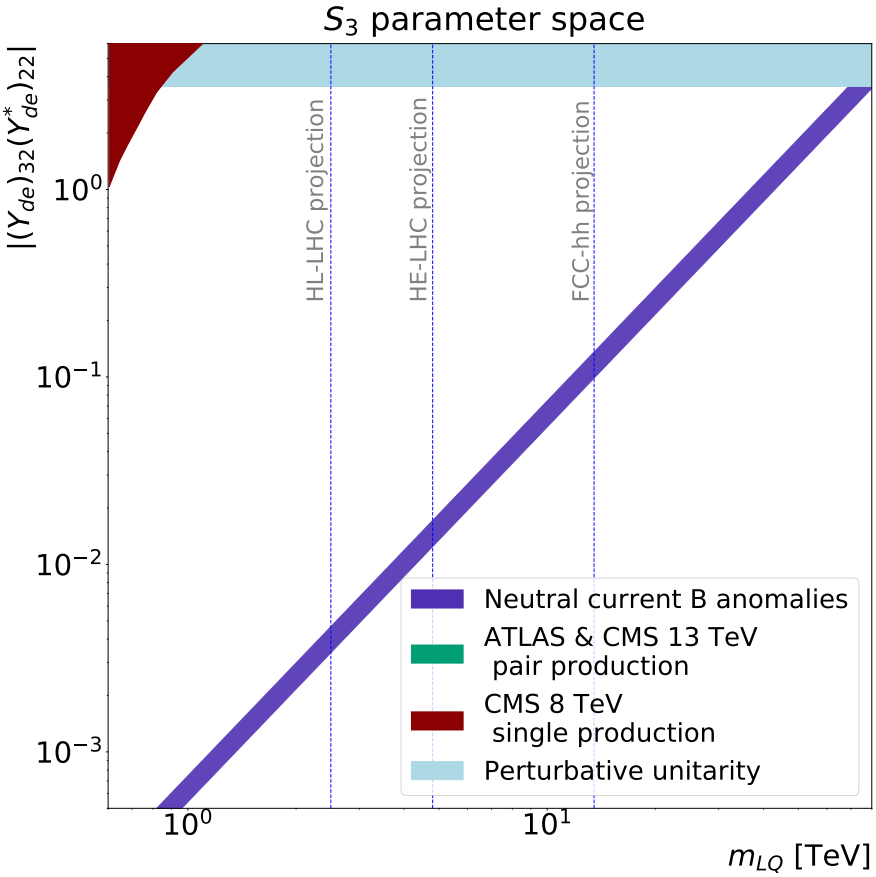


Fig. 5.5.1. We summarise constraints on the LQ mass and the product of couplings $|(Y_{de})_{32}(Y_{de})_{22}|$ from searches for LQ single and pair production by ATLAS and CMS, as well as constraints from perturbative unitarity. We overlay a curve indicating the LQ parameters which best-fit the NCBAs, taking the $\pm 1\sigma$ range of Eq. 2.5.9. These constraints are compared to our projections for the mass sensitivity of the HL-LHC, HE-LHC and FCC-hh.

Chapter 6

Conclusions and future directions

We began this thesis by motivating the use of effective field theories in the search for new physics beyond the Standard Model. Having observed no significant evidence for new physics at the LHC so far, we are presented with a challenge: if new physics must lie at high mass scales, how can we best interpret its effect on low energy observables? Throughout this thesis we have demonstrated the use of effective field theories in addressing this challenge.

The benefit of the model independence of an EFT comes at the price of a vast number of Wilson coefficients which must be determined from measurements. This task is undertaken in Chapter 3, in which we perform a global fit of 34 dimension-6 operators of the SMEFT to data involving the top, Higgs, diboson and electroweak sectors. Although we detect no signs of new physics beyond the Standard Model, our global analysis combines the top sector with measurements from the Higgs, diboson and electroweak sectors for the first time, shedding light on their interplay. An analysis of the correlation matrix indicates that between these sectors, there are a total of 22 correlations of magnitude greater than or equal to 0.22. This indicates that there is a mild but non-negligible interplay between these sectors. Our ability to detect this interplay comes as a result of the increase in precision in measurements of top and Higgs processes from LHC Run II.

Since the publication of our analysis, a comparable global fit to the top, diboson and Higgs sectors has been published by the `SMEFiT` collaboration [67]. A comparison between these fits has been made in Ref. [309]. However, the differences we see in this comparison may not be due to issues in the fit: as discussed in Chapter 2, different choices such as the electroweak input scheme impact our results. An important and interesting direction of future work is to provide tuned comparisons of these fits,

benchmarking different fitting frameworks against one another. Doing so would provide confidence in our fitting methodology, especially if we are to continue to rely on effective field theories in the search for new physics at the HL-LHC.

We have demonstrated the use of a new public code, `Fitmaker`. Further work on the `Fitmaker` code is in progress: we aim to make the code user friendly, expanding the database regularly and updating the publicly available theory predictions to include UV models matched to the SMEFT. The χ^2 fitting methodology of `Fitmaker` is capable of producing fast fits at linear order in the SMEFT coefficients. We benefit from this speed by being able to redo each fit in interesting subspaces of the SMEFT, for example in analysing subsets of operators relevant to specific UV models. However, although the speed of the χ^2 fitter is advantageous, this fitter is limited. Using the nested sampling fitter allows us to relax many approximations: we can extend our observables to depend on the Wilson coefficients at quadratic order, and the Gaussian approximation of the likelihood is no longer required. We have demonstrated the use of this nested sampler in a fit to Higgs data in Appendix B, and made use of it in producing constraints on UV models in §3.7. A full fit to the 34 dimension-6 operators of the SMEFT considered in our analysis is yet to be completed, requiring further developments and tests of our implementation of the nested sampling code. We leave this to future work.

Our ability to constrain the SMEFT in Chapter 3 is partially governed by the measurements available to us. The Higgs sector provides an excellent example of this: we see a marked increase in sensitivity resulting from the inclusion of STXS measurements of Higgs production and decays from LHC Run II. This binned, differential information has far more sensitivity to EFT effects compared to measurements of total cross sections. Sensitivity to EFT coefficients could be further maximised by the use of observables produced using unbinned multivariate analysis techniques, such as those first presented in Ref. [310]. Here, 'optimal observables' are constructed specifically for EFT fits by training a neural network to reconstruct a likelihood ratio. Making use of the likelihood ratio as a test statistic is well-motivated by the Neyman Pearson Lemma [311]. These neural networks are then used to analyse a dataset on an event-by-event basis. An interesting line of future work would be to develop this methodology further such that these observables could be incorporated into a global fit along with existing measurements of cross sections and differential distributions.

In Chapter 4 we delved into the elusive nature of EFT effects, quantifying their

interplay with the parton distribution functions in measurements of high-mass Drell-Yan tails. By performing a simultaneous fit to the PDF and SMEFT coefficients, we found a subtle interplay. The impact of this interplay on the constraints on Wilson coefficients is moderate and within the PDF uncertainties, indicating that our work in constraining the SMEFT thus far is safe: we have not invalidated any previously-obtained constraints.

However, our work serves as a warning for the future: if we include high-mass measurements in PDF fits, particularly those from the HL-LHC, we may amplify this PDF and EFT interplay. Our projections show that by ignoring this effect, we risk overconstraining the SMEFT. This work has important implications for global EFT fits. As global fits incorporate increasing amounts of data, an overlap with the data in PDF fits is inevitable. In this context, a simultaneous determination of the EFT and PDFs is necessary.

An exciting and challenging direction for future work is to perform a study of the PDF and EFT interplay in new data sectors, developing the tools for a truly simultaneous global fit in the process. Jet data would provide a complementary dataset to our Drell-Yan study. As we discussed in Chapter 3, dijet observables produce strong constraints on the $\mathcal{O}_G$ operator of the SMEFT. Jet observables are also important in PDF fits, for example in constraining the gluon PDF. This line of future work does not come without difficulties, however: the modifications to the PDF from the SMEFT are more complex than in the case of Drell-Yan, with SMEFT operators producing corrections to the DGLAP evolution equations.

It is important to acknowledge that although the work of Chapter 4 recommends the use of simultaneous PDF and EFT fits in the future, these will not always be computationally feasible. Alternative methods of accounting for the effects of the PDF and EFT interplay are needed. In Chapter 4 we consider 'conservative PDFs', incorporating only low-mass and on-shell DY observables as well as DIS observables into the PDF fit. These are defined such that the assumption of the SM in the PDF fit may be made less problematic, by excluding the high-mass data most sensitive to the SMEFT. However, our projections show that using this definition of conservative PDFs is not sufficient: our constraints on the SMEFT remain over-precise. Further work is needed on the definition of conservative PDF sets.

Chapter 5 serves as a case study of the use of EFTs in the interpretation of signs of new physics. Once a deviation has been observed, how do we proceed? Interpreting the deviations observed in rare B meson decays in the weak effective field theory

framework is just the first step. Recall the example of muon decay in the Fermi theory of low-energy weak interactions in §2.1.2. We saw that by matching the low energy EFT to the SM, the Wilson coefficient G_F was related to the parameters of the SM by $2\sqrt{2}G_F = \frac{g_L^2}{2m_W^2}$. This indicates that measurements at low energy and the resulting constraints on the Wilson coefficients can only provide information on *ratios* of the new physics couplings and masses. Although a measurement of a nonzero Wilson coefficient provides evidence for new physics, it does not fully characterise this new physics: the direct measurement of the BSM particle is required to disentangle its parameters. This is how we proceed in Chapter 5, studying the direct pair production of scalar leptoquarks at future hadron colliders, as motivated by the neutral current B anomalies. We estimate the sensitivity of the HL-LHC, HE-LHC and FCC-hh to leptoquarks in the $\mu^+\mu^- jj$ channel, finding that the FCC-hh could discover leptoquarks up to masses of 9.5 TeV.

This work, and similar estimates of the future hadron collider sensitivity to Z' production [286, 78], are based on simple cut-based analyses searching for excesses in distributions of kinematic parameters, such as $m_{\text{min}}(\mu, j)$. Many of the recent searches for new physics by ATLAS and CMS use much more sophisticated analysis techniques. As we discussed in Chapter 5, a recent search for scalar leptoquarks by ATLAS uses boosted decision trees [312], and a search for Z' particles by ATLAS uses a data-driven method to estimate the SM background [313]. Including these analysis techniques is beyond the scope of this work, especially as our estimates are much more dependent on unknowns related to detectors at the FCC-hh which have not yet been built. However, when creating projections for the near future such as the HL-LHC it would be interesting to incorporate these techniques into our phenomenological analyses, particularly if they are capable of maximising sensitivity to new physics: this can only strengthen the case for future colliders.

Finally, we note that Chapter 5 is based on Ref. [3], published in 2019. Since then, evidence for new physics in the neutral current B anomalies has grown: the most recent measurement of R_K showed a deviation from the SM with a significance of 3.1σ, an increase from the previous 2.5σ. This is an intriguing development, generating cautious excitement in the search for new physics beyond the Standard Model. We will conclude this thesis on this positive note, in the hope that the methods and insights presented here will contribute to the detection and interpretation of many more signs of new physics yet to be observed.

Appendix A

Datasets

The following tables summarise the observables that have been encoded into the `Fitmaker` database, some of which are used to produce the global fit discussed in Chapter 3. Those that are not included in the final fit are shaded, as they are not statistically independent from other data that we include.

Diboson LEP & LHC	$n_{\mathbf{obs}}$	**Ref.**
W^+W^- angular distribution measurements at LEP II.	8	[314]
W^+W^- total cross section measurements at L3 in the $\ell\nu\ell\nu$, $\ell\nu qq$ & $qqqq$ final states for 8 energies	24	[315]
W^+W^- total cross section measurements at OPAL in the $\ell\nu\ell\nu$, $\ell\nu qq$ & $qqqq$ final states for 7 energies	21	[316]
W^+W^- total cross section measurements at ALEPH in the $\ell\nu\ell\nu$, $\ell\nu qq$ & $qqqq$ final states for 8 energies	21	[317]
ATLAS W^+W^- differential cross section in the $e\nu\mu\nu$ channel, $\frac{d\sigma}{dp^T_{\ell_1}}$, $p_T > 120$ GeV overflow bin	1	[318]
ATLAS W^+W^- fiducial differential cross section in the $e\nu\mu\nu$ channel, $\frac{d\sigma}{dp^T_{\ell_1}}$	14	[184]
ATLAS $W^\pm Z$ fiducial differential cross section in the $\ell^+\ell^-\ell^\pm\nu$ channel, $\frac{d\sigma}{dp^T_Z}$	7	[186]
CMS $W^\pm Z$ normalised fiducial differential cross section in the $\ell^+\ell^-\ell^\pm\nu$ channel, $\frac{1}{\sigma}\frac{d\sigma}{dp^T_Z}$	11	[185]
ATLAS Zjj fiducial differential cross section in the $\ell^+\ell^-$ channel, $\frac{d\sigma}{d\Delta\phi_{jj}}$	12	[187]

Table A.0.1. Diboson measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3.

EW precision observables	$n_{\mathbf{obs}}$	**Ref.**
Precision electroweak measurements on the Z resonance. Γ_Z, $\sigma^0_{\text{had.}}$, R^0_ℓ, A^ℓ_{FB}, A_ℓ(SLD), A_ℓ(Pt), R^0_b, R^0_c A^b_{FB}, A^c_{FB}, A_b & A_c	12	[319]
Combination of CDF and D0 W-Boson Mass Measurements	1	[320]
LHC run 1 W boson mass measurement by ATLAS	1	[14]

Table A.0.2. Electroweak precision measurements included in the Fitmaker database.

LHC Run 1 Higgs	$n_{\mathbf{obs}}$	**Ref.**
ATLAS and CMS LHC Run 1 combination of Higgs signal strengths. Production: ggF, VBF, ZH, WH & ttH Decay: $\gamma\gamma$, ZZ, W^+W^-, $\tau^+\tau^-$ & $b\bar{b}$	21	[321]
ATLAS inclusive $Z\gamma$ signal strength measurement	1	[322]

Table A.0.3. Higgs Run 1 measurements included in the Fitmaker database.

LHC Run 2 Higgs (new)	$n_{\mathbf{obs}}$	**Ref.**
ATLAS combination of signal strengths and stage 1.0 STXS in $H \to 4\ell$ including ratios of branching fractions to $\gamma\gamma$, WW^*, $\tau^+\tau^-$ & $b\bar{b}$ Signal strengths|coarse STXS bins| fine STXS bins	16|19|25	[323]
CMS LHC combination of Higgs signal strengths. Production: ggF, VBF, ZH, WH & ttH Decay: $\gamma\gamma$, ZZ, W^+W^-, $\tau^+\tau^-$, $b\bar{b}$ & $\mu^+\mu^-$	23	[59]
CMS stage 1.0 STXS measurements for $H \to \gamma\gamma$. 13 parameter fit | 7 parameter fit	13|7	[324]
CMS stage 1.0 STXS measurements for $H \to \tau^+\tau^-$	9	[325]
CMS stage 1.1 STXS measurements for $H \to 4\ell$	19	[326]
CMS differential cross section measurements of inclusive Higgs production in the $WW^* \to \ell\nu\ell\nu$ final state. $\frac{d\sigma}{dn_{\text{jet}}}$ | $\frac{d\sigma}{dp_H^T}$	5|6	[327]
ATLAS $H \to Z\gamma$ signal strength.	1	[328]
ATLAS $H \to \mu^+\mu^-$ signal strength.	1	[329]

Table A.0.4. Higgs Run 2 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3.

Tevatron & Run 1 top	$n_{\mathbf{obs}}$	**Ref.**
Tevatron combination of differential $t\bar{t}$ forward-backward asymmetry, $A_{FB}(m_{t\bar{t}})$.	4	[330]
ATLAS $t\bar{t}$ differential distributions in the dilepton channel. $\frac{d\sigma}{dm_{t\bar{t}}}$	6	[331]
ATLAS $t\bar{t}$ differential distributions in the ℓ+jets channel. $\frac{d\sigma}{dm_{t\bar{t}}} \mid \frac{d\sigma}{d\lvert y_{t\bar{t}}\rvert} \mid \frac{d\sigma}{dp_t^T} \mid \frac{d\sigma}{d\lvert y_t\rvert}$.	7|5|8|5	[332]
CMS $t\bar{t}$ differential distributions in the ℓ+jets channel. $\frac{d\sigma}{dm_{t\bar{t}}} \mid \frac{d\sigma}{dy_{t\bar{t}}} \mid \frac{d\sigma}{dp_t^T} \mid \frac{d\sigma}{dy_t}$.	7|10|8 |10	[333, 334]
CMS measurement of differential $t\bar{t}$ charge asymmetry, $A_C(m_{t\bar{t}})$ in the dilepton channel.	3	[335]
ATLAS inclusive measurement $t\bar{t}$ charge asymmetry, $A_C(m_{t\bar{t}})$ in the dilepton channel.	1	[336]
ATLAS & CMS combination of differential $t\bar{t}$ charge asymmetry, $A_C(m_{t\bar{t}})$, in the ℓ+jets channel.	6	[337]
CMS $t\bar{t}$ double differential distributions in the dilepton channel. $\frac{d\sigma}{dm_{t\bar{t}}dy_t} \mid \frac{d\sigma}{dm_{t\bar{t}}dy_{t\bar{t}}} \mid \frac{d\sigma}{dm_{t\bar{t}}dp_{t\bar{t}}^T} \mid \frac{d\sigma}{dy_t dp_t^T}$.	16|16 |16|16	[338, 339]
ATLAS & CMS Run 1 combination of W-boson helicity fractions in top decay. f_0, f_L & f_R	3	[340]
ATLAS measurement of W-boson helicity fractions in top decay. f_0, f_L & f_R	3	[341]
CMS measurement of W-boson helicity fractions in top decay. f_0, f_L & f_R	3	[342]
ATLAS $t\bar{t}W$ & $t\bar{t}Z$ cross section measurements. $\sigma_{t\bar{t}W}\lvert\sigma_{t\bar{t}Z}$	2	[196]
CMS $t\bar{t}W$ & $t\bar{t}Z$ cross section measurements. $\sigma_{t\bar{t}W}\lvert\sigma_{t\bar{t}Z}$	2	[195]
ATLAS $t\bar{t}\gamma$ cross section measurement in the ℓ+ jets channel.	1	[343]
CMS $t\bar{t}\gamma$ cross section measurement in the ℓ+ jets channel.	1	[344]
ATLAS t-channel single-top differential distributions. $\frac{d\sigma}{dp_t^T} \mid \frac{d\sigma}{dp_{\bar{t}}^T} \mid \frac{d\sigma}{d\lvert y_t\rvert} \mid \frac{d\sigma}{d\lvert y_{\bar{t}}\rvert}$	4|4|4|5	[345]
CMS s-channel single-top cross section measurement.	1	[346]
CMS t-channel single-top differential distributions. $\frac{d\sigma}{dp_{t+\bar{t}}^T} \mid \frac{d\sigma}{d\lvert y_{t+\bar{t}}\rvert}$	6 |6	[347]
CMS measurement of the t-channel single-top and anti-top cross sections. $\sigma_t \mid \sigma_{\bar{t}} \mid \sigma_{t+\bar{t}} \mid R_t$.	1|1|1|1	[348]
ATLAS s-channel single-top cross section measurement.	1	[349]
CMS tW cross section measurement.	1	[350]
ATLAS tW cross section measurement in the single lepton channel.	1	[351]
ATLAS tW cross section measurement in the dilepton channel.	1	[352]

Table A.0.5. Top Run 1 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3.

Run 2 top	$n_{\mathbf{obs}}$	**Ref.**
CMS $t\bar{t}$ differential distributions in the dilepton channel. $\frac{d\sigma}{dm_{t\bar{t}}}$	6	[353, 354]
CMS $t\bar{t}$ differential distributions in the ℓ+jets channel. $\frac{d\sigma}{dm_{t\bar{t}}}$	10	[5]
ATLAS measurement of differential $t\bar{t}$ charge asymmetry, $A_C(m_{t\bar{t}})$.	5	[355]
ATLAS $t\bar{t}W$ & $t\bar{t}Z$ cross section measurements. $\sigma_{t\bar{t}W}\vert\sigma_{t\bar{t}Z}$	2	[197]
CMS $t\bar{t}W$ & $t\bar{t}Z$ cross section measurements. $\sigma_{t\bar{t}W}\vert\sigma_{t\bar{t}Z}$	1|1	[356]
CMS $t\bar{t}Z$ differential distributions. $\frac{d\sigma}{dp_Z^T}$ | $\frac{d\sigma}{d\cos\theta^*}$	4|4	[357]
ATLAS $t\bar{t}\gamma$ differential distribution. $\frac{d\sigma}{dp_\gamma^T}$	11	[358]
CMS measurement of differential cross sections and charge ratios for t-channel single-top quark production. $\frac{d\sigma}{dp_{t+\bar{t}}^T}$ | $R_t\left(p_{t+\bar{t}}^T\right)$	5|5	[194]
CMS measurement of t-channel single-top and anti-top cross sections. σ_t, $\sigma_{\bar{t}}$, $\sigma_{t+\bar{t}}$ & R_t.	4	[359]
CMS measurement of the t-channel single-top and anti-top cross sections. $\sigma_t \vert \sigma_{\bar{t}} \vert \sigma_{t+\bar{t}} \vert R_t$.	1|1|1|1	[360]
CMS t-channel single-top differential distributions. $\frac{d\sigma}{dp_{t+\bar{t}}^T}$ | $\frac{d\sigma}{d\vert y_{t+\bar{t}}\vert}$	4|4	[361]
ATLAS tW cross section measurement.	1	[362]
CMS tZ cross section measurement.	1	[363]
CMS tW cross section measurement.	1	[364]
ATLAS tZ cross section measurement.	1	[365]
CMS $tZ\,(Z \to \ell^+\ell^-)$ cross section measurement	1	[366]

Table A.0.6. Top Run 2 measurements included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3.

Run 2 top (continued)	$n_{\mathbf{obs}}$	**Ref.**
ATLAS four-top search in the multi-lepton and same-sign dilepton channels.	1	[367]
ATLAS four-top search in the single-lepton and opposite-sign dilepton channels.	1	[368]
CMS four-top search in the multi-lepton and same-sign dilepton channels.	1	[369]
CMS four-top search in the single-lepton and opposite-sign dilepton channels.	1	[370]
CMS $t\bar{t}b\bar{b}$ cross section measurement in the all-jet channel.	1	[371]
CMS $t\bar{t}b\bar{b}$ cross section measurement in the dilepton channel.	1	[372]

Table A.0.7. Top Run 2 measurements (continued) included in the Fitmaker database. Shaded datasets are not included in the global fit of Chapter 3.

Appendix B

Numerical fits with nested sampling

Throughout the global SMEFT analysis presented in Chapter 3, we make use of the fitting procedure outlined in §3.2, in which we minimise a χ^2 function and determine the least-squares estimator for each coefficient $\hat{\vec{C}}$. This procedure has the advantage of being analytic, as it is linearised in the coefficients $\vec{C}$, and therefore fast to implement. However, it relies on a number of key assumptions: the linear approximation to the SMEFT predictions $\mu(\vec{C})$ must be accurate, and the experimental covariance matrix V must be symmetrised. Even if these assumptions are satisfied and we can write down a Gaussian likelihood $\mathcal{L}(\vec{C}|D)$ where D denotes the dataset, the analytic fitting procedure excludes the possibility that a nontrivial prior $\pi(\vec{C})$ may lead to a non-Gaussian posterior $p(\vec{C}|D)$ through Bayes' theorem: $p(\vec{C}|D) \propto \mathcal{L}(\vec{C}|D)\pi(\vec{C})$.

With this motivation, we implement in `Fitmaker` an option to constrain parameters numerically using the method of nested sampling. This is provided through `MultiNest` [138], incorporated into our code using `Pymultinest` [373].

As in traditional Markov Chain Monte Carlo (MCMC) parameter estimation, nested sampling is a method of sampling from the posterior distribution. `MultiNest` uses the ellipsoidal nested sampling algorithm, described in more detail in [138], in which the samples are drawn from ellipsoids in parameter space. As the algorithm progresses the ellipsoids close in on the regions of high likelihood. Overlapping and distinct ellipsoids allow for the possibility of degeneracies and multiple modes in the posterior respectively. This is an advantage over traditional MCMC methods, in which the Markov chain may get stuck and fail to explore more than one mode of the posterior. The importance of this feature in the context of quadratic SMEFT contributions will be seen below.

We apply nested sampling in §3.7 to find constraints on the 1-parameter UV models in Table 3.7.1. In many of these models we are constraining a positive quantity $|\lambda|^2$, where λ denotes a coupling of the new field to the SM. Nested sampling allows us to produce the constraints due to positivity bounds on $|\lambda|^2$ using a Heaviside prior: $\pi(|\lambda|^2 < 0) = 0$. The constraints in Figure 3.7.1 are found in this way.

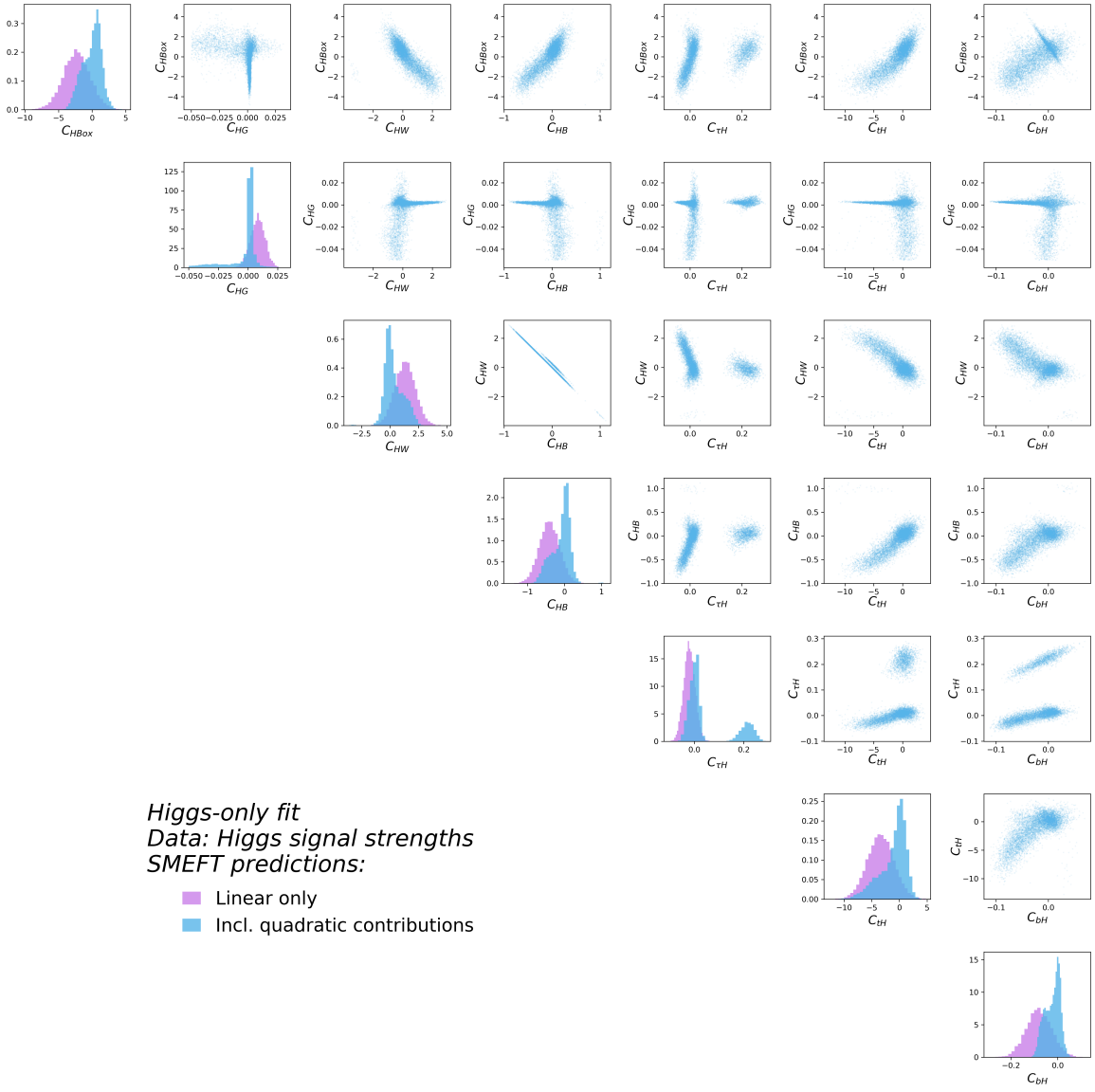


Fig. B.0.1. Samples produced by nested sampling, projected onto 2- and 1-dimensional subspaces of the 7-dimensional parameter space. Along the diagonal we compare the distributions found with and without quadratic SMEFT contributions in blue and purple, respectively.

As a proof-of-concept of the capabilities of nested sampling, we investigate in this Section the effects of including quadratic contributions from dimension-6 operators in SMEFT predictions from Higgs data. We perform a Higgs-only fit using just the Run 2 signal strength measurements from ATLAS [323] and CMS [59], and constrain 7 operators: $C_{H\Box}$, C_{HG}, C_{HW}, C_{HB}, $C_{\tau H}$, C_{tH} and C_{bH}. For the purpose of this proof-of-concept fit, we take our SMEFT predictions from [374], rotating the SILH basis operators into the Warsaw basis using the `Rosetta` code [40]. This fit differs from the results in the main text: as well as the differences in the SMEFT predictions, we use just a subset of the full dataset and use only signal strengths, not STXS measurements.

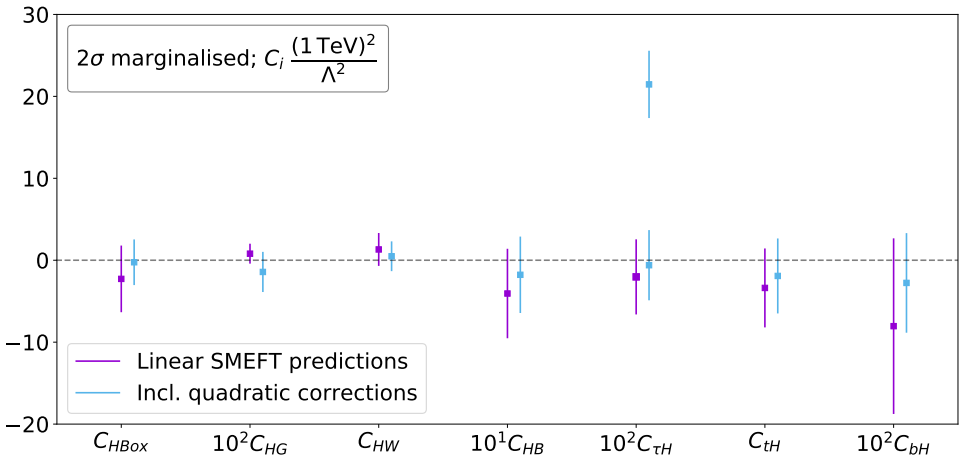


Fig. B.0.2. Marginalised 95 % credible intervals for each of the indicated operator coefficients resulting from a fit to Higgs signal strength data using nested sampling. We compare the effects of including quadratic contributions from dimension-6 operators (blue) to the case of linearised SMEFT predictions (purple).

Figure B.0.1 shows the distribution of 30,000 samples produced by nested sampling (with a sampling efficiency of 0.8 and an evidence tolerance of 0.5, taking approximately 1 hour). Each 2-dimensional distribution is a projection of the full 7-dimensional posterior distribution onto a 2-parameter subspace. These plots highlight the non-Gaussianity of the posterior distributions when quadratic contributions are included. In particular, we see multiple modes in the distribution of $C_{\tau H}$ as well as highly skewed distributions in C_{tH} and C_{bH}.

The histograms along the diagonal in Fig. B.0.1 show the distributions of samples in each of the 7 parameters. Here we compare the results of nested sampling with and without quadratic contributions in blue and purple respectively. We see that although

the distributions are generally peaked close to the same value, the shapes of the distributions differ, with the quadratic contributions leading to more asymmetric and multimodal distributions. For example, the distribution of C_{HG} is more skewed towards the negative region when quadratic corrections are included, and C_{bH} is much more narrowly constrained than in the linear-only case. These differences in distributions translate into differences in the marginalised 95 % credible intervals shown in Fig. B.0.2, computed as highest posterior density intervals.

Although there are visible differences between the credible regions found in the linear and quadratic SMEFT fits, they are sufficiently similar that one may consider the linear approximation to be usefully robust. The most notable difference is in $C_{\tau H}$, in which a distinct second mode is found in the quadratic SMEFT fit, while the first mode is in good agreement with the linear SMEFT fit. There is only one other instance, namely C_{HG}, where the mode in the quadratic fit lies outside the 95% CL range found in the linear approximation, and only one instance, namely C_{bH}, where the size of the quadratic 95% credible interval is much smaller than the linear 95% CL range. Apart from these exceptions, the ranges estimated in the linear fit are encouraging approximations to the results from the quadratic fit. We note that a global quadratic fit would require calculations of many currently unknown quadratic operator contributions and, for consistency, a full treatment of the linear contributions of dimension-8 operators as discussed in the Higgs sector in [170].

Appendix C

Tables of numerical fit results

In this Appendix we share the numerical results of the global analysis of Chapter 3.

Table C.0.1 provides the numerical values of the individual and marginalised constraints of Fig. 3.6.4 presented in terms of SMEFT coefficients (for $\Lambda = 1$ TeV) and scale $\Lambda/\sqrt{C}$.

Table C.0.2 uses the Fisher Information matrix to provide the relative constraining power of the datasets on each SMEFT coefficient taken individually, *i.e.* setting all other coefficients to zero. For each SMEFT coefficient this provides an estimate of the dataset which will dominate the constraint on this coefficient.

Finally, Table C.0.3 provides the coefficients of the eigenvectors found in the principal component analysis of the global fit displayed in Fig. 3.6.11.

	Individual			Marginalised		
SMEFT Coeff.	Best fit [$\Lambda = 1$ TeV]	95% CL range	Scale $\frac{\Lambda}{\sqrt{C}}$ [TeV]	Best fit [$\Lambda = 1$ TeV]	95% CL range	Scale $\frac{\Lambda}{\sqrt{C}}$ [TeV]
C_{HWB}	0.00	[-0.0043, +0.0026]	17.0	0.18	[-0.36, +0.73]	1.4
C_{HD}	-0.01	[-0.023, +0.0027]	8.8	-0.39	[-1.6, +0.81]	0.91
C_{ll}	0.01	[-0.005, +0.019]	9.2	-0.03	[-0.084, +0.02]	4.4
$C_{Hl}^{(3)}$	0.00	[-0.01, +0.003]	12.0	-0.03	[-0.13, +0.055]	3.3
$C_{Hl}^{(1)}$	0.00	[-0.0044, +0.013]	11.0	0.11	[-0.19, +0.41]	1.8
C_{He}	0.00	[-0.015, +0.0071]	9.6	0.19	[-0.41, +0.79]	1.3
$C_{Hq}^{(3)}$	0.00	[-0.017, +0.012]	8.3	-0.05	[-0.11, +0.012]	4.1
$C_{Hq}^{(1)}$	0.02	[-0.1, +0.14]	2.9	-0.04	[-0.27, +0.18]	2.1
C_{Hd}	-0.03	[-0.13, +0.071]	3.1	-0.39	[-0.91, +0.13]	1.4
C_{Hu}	0.00	[-0.075, +0.073]	3.7	-0.19	[-0.63, +0.25]	1.5
$C_{H\Box}$	-0.27	[-1, +0.47]	1.2	-0.9	[-3, +1.2]	0.69
C_{HG}	0.00	[-0.0034, +0.0032]	17.0	0.00	[-0.014, +0.0086]	9.4
C_{HW}	0.00	[-0.012, +0.006]	11.0	0.12	[-0.38, +0.62]	1.4
C_{HB}	0.00	[-0.0034, +0.002]	19.0	0.07	[-0.09, +0.22]	2.5
C_{W}	0.18	[-0.071, +0.42]	2.0	0.15	[-0.11, +0.4]	2.0
C_{G}	-0.46	[-0.77, -0.14]	1.8	-1.4	[-2.2, -0.72]	1.2
$C_{\tau H}$	0.01	[-0.015, +0.025]	7.1	0.01	[-0.016, +0.028]	6.7
$C_{\mu H}$	0.00	[-0.0057, +0.005]	14.0	0.00	[-0.0058, +0.005]	14.0
C_{bH}	0.00	[-0.016, +0.024]	7.1	0.01	[-0.034, +0.052]	4.8
C_{tH}	-0.09	[-1, +0.84]	1.0	1.5	[-2.8, +5.7]	0.48
$C_{HQ}^{(3)}$	0.01	[-0.032, +0.048]	5.0	-0.1	[-0.67, +0.46]	1.3
$C_{HQ}^{(1)}$	0.01	[-0.031, +0.049]	5.0	-0.01	[-0.59, +0.58]	1.3
C_{Ht}	0.87	[-1.2, +2.9]	0.7	6.6	[+2, +11]	0.47
C_{tG}	-0.01	[-0.1, +0.086]	3.2	0.36	[+0.12, +0.6]	2.0
C_{tW}	0.19	[-0.12, +0.51]	1.8	0.23	[-0.088, +0.55]	1.8
C_{tB}	-1.6	[-4.5, +1.2]	0.59	-1.4	[-5.2, +2.5]	0.51
$C_{Qq}^{3,1}$	0.06	[-0.043, +0.16]	3.2	0.05	[-0.071, +0.17]	2.9
$C_{Qq}^{3,8}$	-1.2	[-2.4, +0.036]	0.91	-6.8	[-18, +4.5]	0.3
$C_{Qq}^{1,8}$	-0.12	[-0.56, +0.31]	1.5	-0.65	[-4.9, +3.6]	0.48
C_{Qu}^{8}	-0.6	[-1.3, +0.06]	1.2	6.3	[-2.5, +15]	0.34
C_{Qd}^{8}	-1.4	[-2.9, +0.07]	0.83	1.8	[-9.5, +13]	0.3
C_{tq}^{8}	-0.4	[-0.85, +0.059]	1.5	-5.6	[-13, +2.2]	0.36
C_{tu}^{8}	-0.45	[-1.1, +0.23]	1.2	4.0	[-11, +19]	0.26
C_{td}^{8}	-1.0	[-2.5, +0.38]	0.83	-0.42	[-12, +11]	0.29

Table C.0.1. Table of numerical results in Fig. 3.6.4 from the global fit to the electroweak, diboson, Higgs and top data in the top-specific $SU(2)^2 \times SU(3)^3$ scenario.

C_i	EWPO	LEP WW	Run 1 SS	Run 2 SS	STXS	LHC WW	WZ	Zjj	$t\bar{t}$	$W_{\text{hel.}}$	tX	$t\bar{t}V$
C_{HWB}	51	–	7	14	28	–	–	–	–	–	–	–
C_{HD}	100	–	–	–	–	–	–	–	–	–	–	–
C_{ll}	99	–	–	–	–	–	–	–	–	–	–	–
$C_{Hl}^{(3)}$	99	–	–	–	–	–	–	–	–	–	–	–
$C_{Hl}^{(1)}$	100	–	–	–	–	–	–	–	–	–	–	–
C_{He}	100	–	–	–	–	–	–	–	–	–	–	–
$C_{Hq}^{(3)}$	89	1	–	–	2	–	6	–	–	–	–	–
$C_{Hq}^{(1)}$	99	–	–	–	–	–	–	–	–	–	–	–
C_{Hd}	99	–	–	–	–	–	–	–	–	–	–	–
C_{Hu}	98	–	–	–	1	–	–	–	–	–	–	–
$C_{H\Box}$	–	–	22	46	32	–	–	–	–	–	–	–
C_{HG}	–	–	22	42	36	–	–	–	–	–	–	–
C_{HW}	–	–	14	29	56	–	–	–	–	–	–	–
C_{HB}	–	–	14	29	57	–	–	–	–	–	–	–
C_W	–	3	–	–	–	–	13	84	–	–	–	–
C_G	–	–	–	–	–	–	–	–	43	–	–	56
$C_{\tau H}$	–	–	22	45	34	–	–	–	–	–	–	–
$C_{\mu H}$	–	–	5	95	–	–	–	–	–	–	–	–
C_{bH}	–	–	19	35	47	–	–	–	–	–	–	–
C_{tH}	–	–	21	45	34	–	–	–	–	–	–	–
$C_{HQ}^{(3)}$	99	–	–	–	–	–	–	–	–	–	–	–
$C_{HQ}^{(1)}$	100	–	–	–	–	–	–	–	–	–	–	–
C_{Ht}	–	–	–	–	–	–	–	–	–	–	–	100
C_{tG}	–	–	13	29	24	–	–	–	24	–	–	9
C_{tW}	–	–	–	–	–	–	–	–	–	84	15	–
C_{tB}	–	–	–	–	–	–	–	–	–	–	–	100
$C_{Qq}^{3,1}$	–	–	–	–	–	–	–	–	–	–	100	–
$C_{Qq}^{3,8}$	–	–	–	–	–	–	–	–	87	–	–	13
$C_{Qq}^{1,8}$	–	–	–	–	–	–	–	–	82	–	–	17
C_{Qu}^{8}	–	–	–	–	–	–	–	–	91	–	–	7
C_{Qd}^{8}	–	–	–	2	–	–	–	–	92	–	–	6
C_{tq}^{8}	–	–	–	1	–	–	–	–	89	–	–	10
C_{tu}^{8}	–	–	–	–	–	–	–	–	96	–	–	3
C_{td}^{8}	–	–	–	2	–	–	–	–	92	–	–	5

Table C.0.2. Relative constraining power in percent of different datasets on each coefficient of the global fit individually. Entries below 1% are not displayed. 'SS', $W_{\text{hel.}}$ and tX refer to Higgs signal strength, W-helicity fraction and single top data, respectively.

$2\sigma[\Lambda = 1\text{TeV}]$	Λ[TeV]	Eigenvector
0.0021	22	$+0.73C_{HB} - 0.51C_{HWB} - 0.37C_{HG} + 0.22C_{HW} - 0.07C_{Hl}^{(3)} - 0.06C_{bH}$
0.0036	17	$-0.79C_{HG} + 0.45C_{HWB} + 0.30C_{Hl}^{(3)} - 0.17C_{ll} + 0.15C_{HD} - 0.10C_{bH} - 0.09C_{He} + 0.09C_{\mu H}$
0.0042	15	$+0.51C_{HB} + 0.46C_{Hl}^{(3)} + 0.46C_{HG} + 0.35C_{HWB} - 0.25C_{ll} + 0.22C_{HD} - 0.16C_{He} + 0.15C_{HW} - 0.10C_{\mu H} - 0.09C_{Hq}^{(3)}$
0.0054	14	$+0.99C_{\mu H} + 0.12C_{HG}$
0.0066	12	$-0.75C_{Hl}^{(1)} + 0.42C_{He} + 0.31C_{Hq}^{(3)} - 0.23C_{Hl}^{(3)} + 0.21C_{HWB} + 0.14C_{HB} - 0.12C_{ll} + 0.11C_{HD} + 0.08C_{HQ}^{(1)} + 0.08C_{HQ}^{(3)} + 0.06C_{Hu}$
0.015	8.3	$-0.56C_{Hq}^{(3)} + 0.49C_{He} + 0.37C_{Hl}^{(3)} - 0.35C_{HWB} - 0.26C_{ll} - 0.19C_{Hl}^{(1)} - 0.18C_{HB} - 0.10C_{Hu} - 0.10C_{HQ}^{(3)} - 0.10C_{HQ}^{(1)} + 0.07C_{Hd} + 0.06C_{HD} - 0.05C_{HW} - 0.05C_{Hq}^{(1)}$
0.019	7.3	$-0.62C_{He} - 0.51C_{Hl}^{(1)} - 0.50C_{HD} - 0.23C_{ll} - 0.14C_{Hq}^{(3)} - 0.13C_{HWB} - 0.07C_{\tau H} - 0.07C_{HB} + 0.06C_{Hl}^{(3)}$
0.019	7.2	$-0.96C_{\tau H} + 0.27C_{bH}$
0.03	5.8	$-0.52C_{HQ}^{(3)} - 0.52C_{HQ}^{(1)} + 0.48C_{Hq}^{(3)} - 0.34C_{bH} + 0.17C_{Hl}^{(3)} - 0.16C_{ll} - 0.14C_{HWB} - 0.13C_{\tau H} - 0.09C_{HB} + 0.06C_{Hd}$
0.035	5.3	$+0.88C_{bH} - 0.27C_{HQ}^{(3)} - 0.27C_{HQ}^{(1)} + 0.24C_{\tau H} + 0.12C_{Hq}^{(3)} - 0.10C_{HG}$
0.057	4.2	$-0.85C_{ll} + 0.29C_{Hl}^{(1)} - 0.26C_{Hl}^{(3)} + 0.18C_{Hq}^{(3)} + 0.17C_{HQ}^{(3)} + 0.17C_{HQ}^{(1)} - 0.11C_{HWB} + 0.07C_{bH} - 0.07C_{HB}$
0.086	3.4	$-0.60C_{HW} - 0.43C_{Hl}^{(3)} - 0.37C_{Hq}^{(3)} + 0.31C_{HB} - 0.24C_{HQ}^{(3)} + 0.23C_{HWB} - 0.22C_{HQ}^{(1)} - 0.13C_{bH} - 0.11C_{ll} + 0.09C_{Hu} + 0.09C_{Hq}^{(1)} + 0.07C_{Qq}^{3,1} - 0.07C_{HD} + 0.06C_{Hl}^{(1)}$
0.1	3.2	$-0.98C_{Qq}^{3,1} - 0.17C_{HW} + 0.08C_{HQ}^{(3)}$
0.11	3	$+0.66C_{HW} - 0.39C_{Hl}^{(3)} - 0.37C_{Hq}^{(3)} - 0.21C_{HQ}^{(1)} + 0.20C_{HD} - 0.20C_{HQ}^{(3)} + 0.19C_{Hu} - 0.18C_{Qq}^{3,1} + 0.15C_{Hq}^{(1)} - 0.14C_{HB} + 0.12C_{HWB} - 0.10C_{He} - 0.07C_{bH} + 0.07C_{W} - 0.07C_{Hd}$
0.14	2.7	$+0.93C_{tG} + 0.30C_{G} + 0.12C_{Qq}^{1,8} + 0.11C_{tq}^{8} + 0.07C_{Qu}^{8} + 0.06C_{tu}^{8}$
0.2	2.2	$+0.97C_{Hq}^{(1)} + 0.13C_{Hl}^{(3)} - 0.08C_{HD} - 0.08C_{Hd} + 0.08C_{HQ}^{(3)} + 0.07C_{HQ}^{(1)}$
0.25	2	$-0.99C_{W} + 0.07C_{Hu}$
0.28	1.9	$-0.92C_{Hu} + 0.24C_{HD} - 0.19C_{Hl}^{(3)} + 0.13C_{Hd} - 0.12C_{He} - 0.08C_{HQ}^{(3)} - 0.08C_{Hl}^{(1)} - 0.07C_{HQ}^{(1)} - 0.06C_{W} + 0.05C_{Hq}^{(1)}$
0.31	1.8	$+0.57C_{Qq}^{1,8} - 0.53C_{tq}^{8} + 0.39C_{tu}^{8} - 0.37C_{Qu}^{8} + 0.21C_{Qq}^{3,8} + 0.17C_{td}^{8} - 0.16C_{Qd}^{8}$
0.32	1.8	$+1.00C_{tW}$
0.38	1.6	$+0.82C_{G} - 0.35C_{tG} + 0.27C_{tq}^{8} + 0.24C_{Qq}^{1,8} + 0.15C_{Qu}^{8} + 0.10C_{Ht} + 0.09C_{tu}^{8} + 0.08C_{Qd}^{8} - 0.06C_{HBox} + 0.06C_{tH} + 0.05C_{td}^{8}$
0.51	1.4	$+0.97C_{Hd} + 0.17C_{Hu} + 0.10C_{HQ}^{(3)} + 0.09C_{Hq}^{(1)} - 0.06C_{Hl}^{(3)} + 0.05C_{HD}$
0.59	1.3	$-0.49C_{tq}^{8} - 0.47C_{Qq}^{1,8} + 0.43C_{G} - 0.39C_{Qu}^{8} - 0.31C_{tu}^{8} - 0.17C_{Qd}^{8} - 0.15C_{td}^{8} - 0.14C_{Qq}^{3,8} + 0.07C_{Ht} - 0.07C_{HBox} + 0.06C_{tH} + 0.06C_{HQ}^{(3)} - 0.05C_{HQ}^{(1)}$
0.77	1.1	$+0.70C_{HQ}^{(1)} - 0.69C_{HQ}^{(3)} + 0.11C_{G} - 0.09C_{Qq}^{3,1} + 0.06C_{Hd} - 0.06C_{Qq}^{1,8} - 0.05C_{Ht}$
1.1	0.96	$+0.59C_{HBox} - 0.58C_{HD} + 0.29C_{He} + 0.27C_{HWB} + 0.23C_{HW} - 0.19C_{Hu} + 0.14C_{Hl}^{(1)} - 0.10C_{tH} + 0.09C_{Ht} + 0.08C_{Hd} + 0.08C_{HB} - 0.07C_{Qq}^{3,8} - 0.06C_{HQ}^{(1)}$
1.7	0.78	$-0.64C_{Qq}^{3,8} + 0.51C_{Qq}^{1,8} - 0.40C_{tu}^{8} + 0.29C_{Ht} - 0.16C_{td}^{8} - 0.12C_{Qu}^{8} - 0.12C_{HBox} - 0.12C_{G} + 0.08C_{HQ}^{(1)} - 0.07C_{HQ}^{(3)} - 0.06C_{tB} + 0.05C_{HD}$
2.1	0.7	$+0.73C_{HBox} + 0.44C_{HD} - 0.31C_{tH} - 0.22C_{He} - 0.20C_{HWB} - 0.19C_{HW} + 0.12C_{Hu} - 0.11C_{Hl}^{(1)} + 0.10C_{G} - 0.06C_{tB} - 0.06C_{Qu}^{8} - 0.06C_{HB}$
2.8	0.6	$+0.85C_{tB} - 0.31C_{Ht} - 0.20C_{Qd}^{8} + 0.19C_{Qu}^{8} + 0.17C_{tH} - 0.16C_{Qq}^{3,8} + 0.13C_{HBox} - 0.11C_{tq}^{8} - 0.10C_{tu}^{8} + 0.09C_{Qq}^{1,8}$
3.4	0.54	$-0.71C_{Ht} - 0.45C_{tB} + 0.40C_{tH} + 0.16C_{Qq}^{1,8} - 0.14C_{tu}^{8} + 0.13C_{HBox} + 0.13C_{Qu}^{8} - 0.13C_{Qq}^{3,8} - 0.12C_{tq}^{8} - 0.10C_{td}^{8} + 0.06C_{Qd}^{8} + 0.06C_{G}$
4.4	0.48	$+0.82C_{tH} + 0.46C_{Ht} + 0.25C_{HBox} + 0.10C_{HD} + 0.09C_{tu}^{8} - 0.09C_{G} + 0.07C_{Qq}^{3,8} + 0.07C_{tq}^{8} - 0.06C_{Qd}^{8}$
9.0	0.33	$+0.55C_{Qu}^{8} - 0.46C_{td}^{8} - 0.40C_{Qd}^{8} + 0.36C_{Qq}^{3,8} - 0.26C_{tq}^{8} + 0.19C_{Ht} - 0.19C_{tu}^{8} + 0.15C_{Qq}^{1,8} - 0.13C_{tH} - 0.12C_{tB}$
9.6	0.32	$+0.70C_{td}^{8} - 0.40C_{Qd}^{8} + 0.39C_{Qu}^{8} - 0.31C_{Qq}^{3,8} - 0.21C_{tB} - 0.17C_{tq}^{8} - 0.13C_{Qq}^{1,8} + 0.11C_{Ht}$
13.0	0.28	$+0.75C_{Qd}^{8} - 0.48C_{tq}^{8} + 0.33C_{Qu}^{8} + 0.18C_{Ht} + 0.16C_{td}^{8} - 0.16C_{tu}^{8} + 0.09C_{tB}$
21.0	0.22	$+0.69C_{tu}^{8} - 0.50C_{Qq}^{3,8} - 0.41C_{td}^{8} + 0.20C_{Qu}^{8} - 0.18C_{Qq}^{1,8} - 0.17C_{tq}^{8} + 0.06C_{Qd}^{8}$

Table C.0.3. Components of the eigenvectors found in the principal component analysis of the global fit displayed in Fig. 3.6.11. Components with coefficients of magnitude less than 0.05 are omitted.

Appendix D

Detailed SM PDF comparisons

In this appendix we present detailed comparisons between different sets of SM PDFs to complement the discussions in §4.3 of Chapter 4. To begin with, we compare the baseline SM PDF of this work, based on DIS+DY data, with the recent global `NNPDF3.1_str` fit obtained in the context of the proton strangeness study of [12]. Fig. D.0.1 is the counterpart of Fig. 4.3.3, now displaying the gluon, singlet, up, anti-up, down, and anti-down quark PDFs at $Q = 100$ GeV both for the baseline SM PDF (labelled "DIS+DY") and for the global `NNPDF3.1_str` determination.

We observe an overall good compatibility between our DIS+DY baseline and the `NNPDF3.1_str` global fit, with PDFs in agreement at the one-sigma level in all cases except for the quark singlet Σ in the region $0.01 \lesssim x \lesssim 0.1$. As discussed in §4.3.3, the new high-mass DY data included in this analysis as compared to [12] are responsible for the bulk of the differences observed in Fig. D.0.1, both in terms of central values and uncertainties, for the quark and anti-quark PDFs. Specifically, the upwards shift in the central values of the quark and anti-quark PDFs in this x-region for the DIS+DY baseline as compared to the `NNPDF3.1_str` determination is consistent with the comparisons in Fig. 4.3.3 illustrating the impact of the high-mass Drell-Yan data in the fit, and the same applies for the associated reduction of the quark and anti-quark PDF uncertainties.

As is well known, the PDF uncertainties on the gluon become rather enlarged in the DIS+DY baseline due to the lack of information from the top and jet cross sections. However, this does not impact the results of the present joint PDF and EFT interpretation, given that gluon-induced contributions to inclusive Drell-Yan processes enter only starting at NLO. Furthermore, we also find somewhat larger uncertainties in the strangeness of the DIS+DY baseline as compared to `NNPDF3.1_str` due to the missing constraints from the NOMAD neutrino dimuon cross sections. All in all, with

the exception of gluon-initiated processes, we can conclude that the DIS+DY baseline to be used in this work is competitive with a full-fledged global PDF determination.

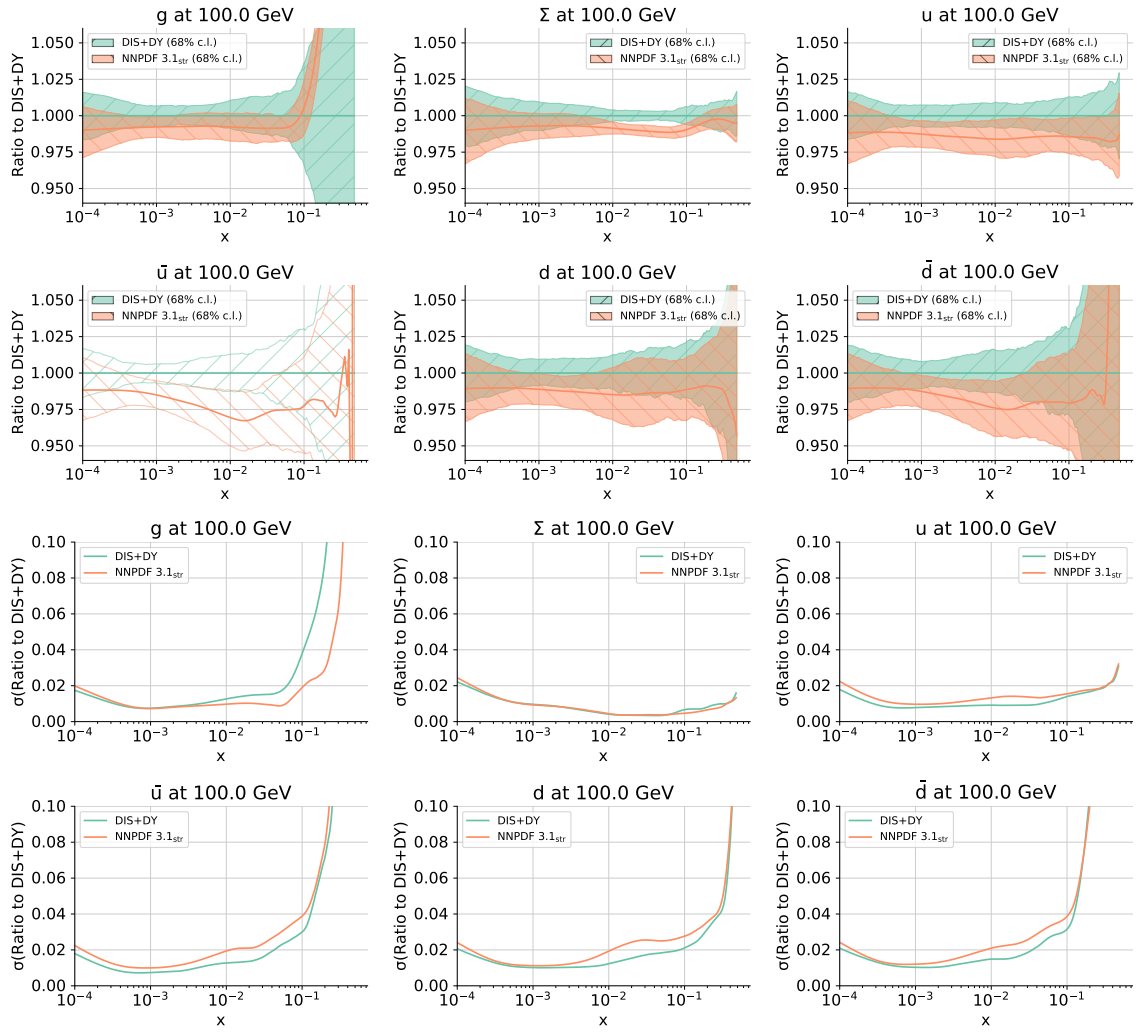


Fig. D.0.1. Same as Fig. 4.3.3 comparing the baseline SM PDF set used in this work (labelled "DIS+DY") with the global `NNPDF3.1_str` determination.

Next, we display in Fig. D.0.2 the corresponding comparison between the baseline SM PDF set based on DIS and DY data (dubbed "DIS+DY") with the same fit but only including DIS structure functions. Note that the comparison between the PDF uncertainties in these two fits was already displayed in the lower panels of Fig. 4.3.3. One can observe how in general there is excellent consistency between the two fits. Indeed, PDFs are in agreement at the one-sigma level except for very specific cases, such as the up quark PDF at $x \simeq 0.05$, but even there the differences are at most at the 1.5σ level. The very marked reduction of PDF errors is also appreciable in the DIS+DY fit as compared to the DIS-only fit, highlighting the importance of the DY

data in the global PDF fit to constrain the light quark and antiquark PDFs in a broad range of x.

Finally, in Fig. D.0.3 we compare the PDF luminosities in the DIS+DY baseline with those from the same fit excluding all the data of the high-mass DY datasets listed in Table 4.3.2. The corresponding comparisons at the PDF level was shown in Fig. 4.3.3

We focus on the gluon-gluon, quark-antiquark, and quark-quark luminosities at $\sqrt{s} = 14$ TeV as a function of the invariant mass m_X of the produced final state, and display both the luminosity ratio to the reference as well as the relative PDF uncertainties in each case.

Again, one finds that the high-mass DY measurements constrain the luminosities in the range 100 GeV $\lesssim m_X \lesssim$ 2 TeV, consistent with the kinematic coverage in $m_{\ell\ell}$ of the data used in the fit. Their main effects are a reduction of the $q\bar{q}$ uncertainty for m_X between 500 GeV and 2 TeV and an upwards (downwards) shift in the central values of the $q\bar{q}$ (gg) luminosities within this m_X region. The uncertainty of $\mathcal{L}_{qq}$ is barely changed and its central value is shifted within 1σ PDF uncertainties once the high-mass Drell-Yan datasets of Table. 4.3.2 are included in the fit. This comparison further highlights how the high-mass DY data provide useful information for constraining the PDF luminosities and in turn the high-p_T processes relevant for both direct and indirect BSM searches at the LHC.

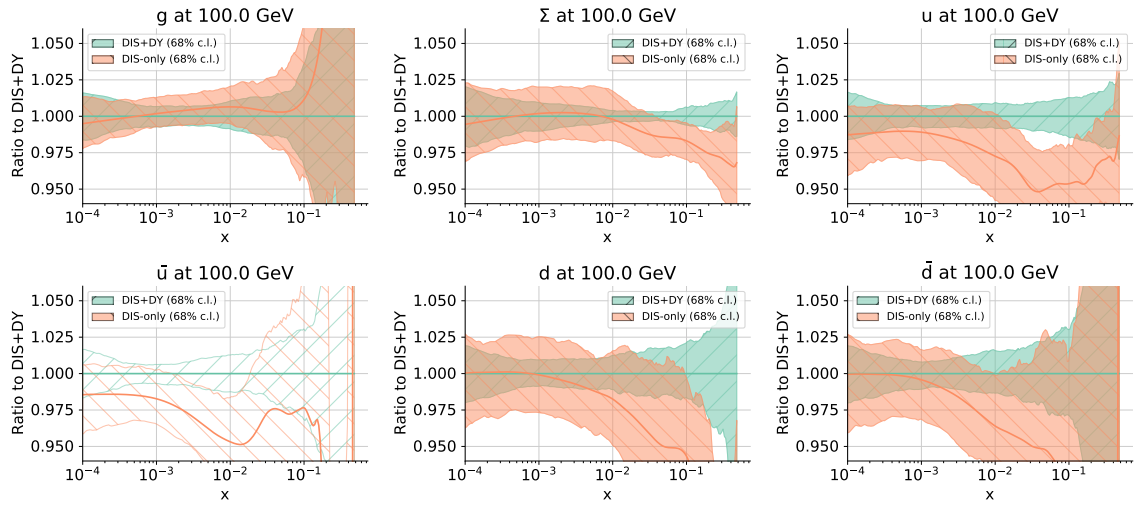


Fig. D.0.2. Same as the upper panels of Fig. 4.3.3 comparing now the baseline SM PDF set used in this work (labelled "DIS+DY") with the corresponding DIS-only fit. Note that the comparison between the PDF uncertainties in these two fits was already displayed in the lower panels of Fig. 4.3.3.

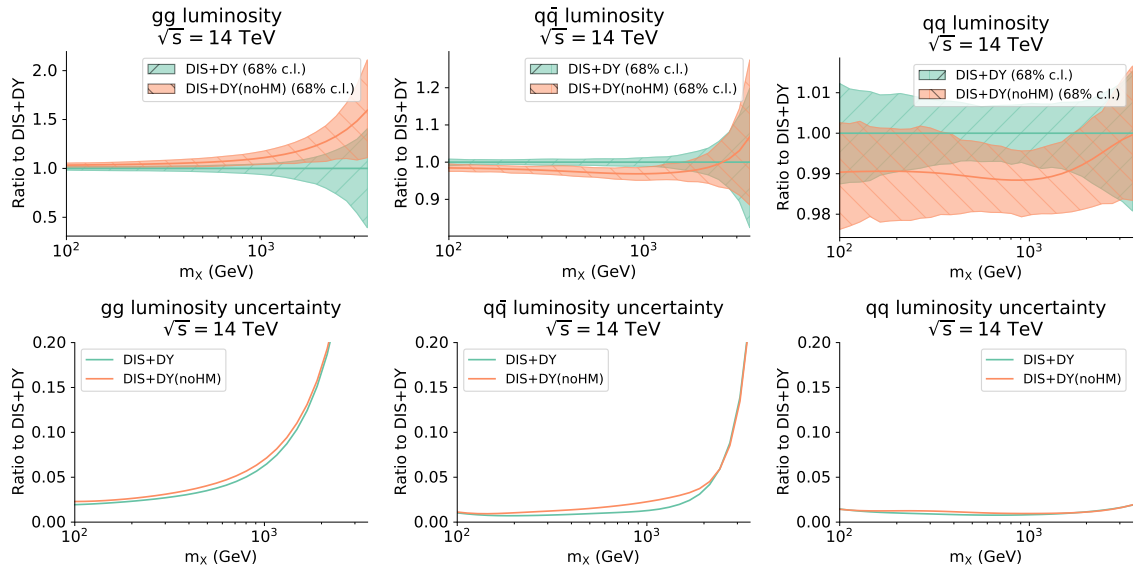


Fig. D.0.3. The gluon-gluon, quark-antiquark, and quark-quark luminosities at $\sqrt{s} = 14$ TeV as a function of m_X for the DIS+DY baseline and for the same fit excluding the all the datapoints in the high-mass DY experimental sets listed in Table. 4.3.2. The top panels display the ratio of luminosities to the central value of the DIS+DY baseline, while the bottom panels compare the relative PDF uncertainties in each case.

Appendix E

Fit quality for SM and SMEFT PDFs

In this appendix, we provide detailed information about the PDF fit quality of Chapter 4, quantifying fit quality by the figure of merit used in the fits: the χ^2 per data point, defined in Eq. (4.3.5) and evaluated with the t_0 prescription described in Ref. [272]. We will do this both for the SM PDFs based on different datasets and for the SMEFT PDFs from the fits with the baseline dataset and for different values of the EFT parameters $\hat{W}$ and $\hat{Y}$.

Dataset	NNPDF3.1	This work
ATLAS 7 TeV	$m_{ee} \leq 200$ GeV	no cuts
ATLAS 8 TeV	not included	$y_{\ell\ell} < 1.68$ (for $m_{\ell\ell} \geq 1000$ GeV)
CMS 7 TeV	$m_{\mu\mu} \leq 200$ GeV $y_{\mu\mu} \leq 2.2$	$y_{\mu\mu} \leq 2.2$ (for all $m_{\mu\mu}$) $y_{\mu\mu} < 1.5$ (for $m_{\mu\mu} \geq 850$ GeV)
CMS 8 TeV	not included	no cuts
CMS 13 TeV	not included	no cuts

Table E.0.1. The kinematic cuts applied to the high-mass Drell-Yan datasets listed in Table 4.3.2, compared to those used in NNPDF3.1.

For completeness, we also provide here the χ^2 values obtained when using `NNPDF3.1_str` as the input PDF set, with all other settings such as the partonic matrix elements unchanged. Note that here the kinematical cuts are slightly different as compared to [12], the differences being summarised in Table E.0.1. The rationale behind having different cuts is that in this work we include electroweak corrections to the high-mass

DY cross sections, thus the $m_{\ell\ell} \leq 200$ GeV restriction applied in NNPDF3.1 is not necessary anymore. With the current cuts, essentially all high-mass DY data points can be included in the fits. The exception is a subset of points from the CMS 7 TeV and ATLAS 8 TeV datasets, where we restrict ourselves to the tree-level kinematic condition $|y_{\ell\ell}| \leq \ln(\sqrt{s}/m_{\ell\ell})$. The reason is that our calculation of the EFT corrections is based on tree-level SM cross sections which must satisfy this requirement. Furthermore, for the CMS 7 TeV dataset the last rapidity bin is excluded for all $m_{\mu\mu}$ bins, since it is found to deviate from the SM predictions by a large amount suggesting the need to account for threshold resummation effects [375]. Henceforth, here we evaluate the predictions based on `NNPDF3.1_str` for the same set of kinematical cuts as in this work.

Table E.0.2 summarises the values of the χ^2 for the baseline SM PDF fit labelled "DIS+DY" compared to the most recent NNPDF global fit `NNPDF3.1_str`, as well as for the corresponding fits based on reduced datasets, namely the DIS-only fit and the fit excluding the high-mass DY data. The entries in italic indicate the datasets that do not enter the corresponding fit. We observe that the quality of the description of the DIS data is similar across all fits considered. As far as hadronic data are concerned, we observe that the fit quality of the LHCb data slightly deteriorates when the high-mass Drell-Yan data are included. Also, the description of the CMS 13 TeV invariant mass distribution in the combined electron and muon channels is not optimal, even after including the data in the fit. However, the overall χ^2 is statistically equivalent to the most recent NNPDF3.1 set.

Then in Table E.0.3 we list again the χ^2 values in the SM PDF fit (same as the "DIS+DY" column of Table E.0.2) and compare them with those obtained with the SMEFT PDFs for the same representative values of $\hat{W}$ and $\hat{Y}$ parameters as used in §4.4.2, see also the PDF-level comparisons in Fig. G.0.1. Clearly the theoretical predictions are computed consistently, namely the partonic cross sections of the SM baseline "DIS+DY" are computed in the SM, while the partonic cross section of the other columns are augmented by the SMEFT contributions of the corresponding operators.

First of all, we observe that as expected the addition of the EFT corrections does not affect the description of the DIS structure functions. Differences are also small for the low-mass and on-shell DY data, and slightly larger for the HM measurements. For instance, the χ^2 to the high-mass Drell-Yan datasets is 1.471 for $\hat{W} = 0.0006$ to be compared with 1.478 for the SM PDFs. In any case, the differences at the level of χ^2

Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$ (SM PDFs) This work DIS-only	DIS+DY(noHM)	DIS+DY	Reference `NNPDF3.1_str`
SLAC	67	1.032	0.807	0.780	0.772
BCDMS	581	1.150	1.222	1.230	1.229
NMC	325	1.320	1.347	1.378	1.346
CHORUS	832	1.058	1.188	1.228	1.191
NuTeV	76	0.796	0.642	0.684	0.703
HERA inclusive	1145	1.238	1.250	1.242	1.264
HERA charm	37	1.654	1.433	1.445	1.424
HERA bottom	29	1.304	1.328	1.326	1.343
Total DIS	**3092**	**1.172**	**1.217**	**1.230**	**1.225**
E886 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	15	*49.94*	0.484	0.484	0.509
E886 $\sigma^p_{\rm DY}$	89	*1.306*	1.061	1.094	1.064
E605 $\sigma^p_{\rm DY}$	85	*2.682*	0.972	0.982	1.006
CDF $d\sigma_Z/dy_Z$	29	*1.796*	1.443	1.460	1.459
D0 $d\sigma_Z/dy_Z$	28	*0.650*	0.595	0.602	0.594
D0 $W \to \mu\nu$ asy.	9	*6.729*	1.411	1.488	1.582
ATLAS W, Z 2010	30	*1.353*	0.817	0.866	0.846
ATLAS low-mass $Z \to ee$	6	*1.038*	0.985	0.949	0.995
ATLAS W, Z 2011	61	*6.077*	1.704	1.681	1.760
ATLAS $W + c$ rapidity	22	*0.497*	0.469	0.468	0.487
ATLAS $Z\, p_T$	92	*1.110*	0.989	0.942	1.029
ATLAS $p_{T,W}$ in W+jets	32	*2.074*	1.574	1.690	1.567
CMS W asy.	22	*5.362*	1.291	1.287	1.292
CMS σ_{W+c} 7 TeV	5	*0.555*	0.495	0.478	0.505
CMS $\sigma_{W^++c}/\sigma_{W^-+c}$ 7 TeV	5	*2.526*	1.826	1.687	1.710
CMS $Z\, p_T$	28	*1.289*	1.336	1.296	1.354
CMS $W \to \mu\nu$ rapidity	22	*5.022*	1.006	1.070	1.077
CMS $W + c$ rapidity 13 TeV	5	*0.638*	0.661	0.658	0.671
LHCb $Z \to \mu\mu$	9	*2.440*	1.630	1.652	1.676
LHCb $W, Z \to \mu$ 7 TeV	29	*13.62*	2.032	2.209	2.136
LHCb $Z \to ee$	17	*1.273*	1.118	1.124	1.114
LHCb $W, Z \to \mu$ 8 TeV	30	*8.835*	1.496	1.769	1.475
Total DY (excl. HM)	**670**	***4.185***	**1.166**	**1.191**	**1.193**
ATLAS DY high-mass 7 TeV	13	*2.261*	*2.014*	1.885	1.945
ATLAS DY high-mass 8 TeV	46	*1.393*	*1.227*	1.181	*1.215*
CMS DY high-mass 7 TeV	117	*1.603*	*1.617*	1.589	1.584
CMS DY high-mass 8 TeV	41	*0.796*	*0.891*	0.805	*0.838*
CMS DY high-mass 13 TeV	43	*1.837*	*1.981*	2.013	*1.952*
Total DY (HM-only)	**260**	***1.510***	***1.514***	**1.478**	***1.480***
Total	**4022**	**1.733**	**1.258**	**1.243**	**1.266**

Table E.0.2. The values of the χ^2 per data point for the baseline SM PDF fit, labelled "DIS+DY", and for the corresponding fits based on reduced datasets. Here Eq. (4.3.5) is evaluated using the t_0 prescription. We also include the results obtained using `NNPDF3.1_str` with the kinematic cuts used in this work and summarised in Table E.0.1. Values in italics indicate datasets that do not enter the corresponding fit.

between the SM and SMEFT PDFs are reasonably small, consistent with the finding that the best-fit values of the $\hat{W}$ and $\hat{Y}$ parameters are close to the SM expectation.

Dataset	$n_{\rm dat}$	$\chi^2/n_{\rm dat}$ SM	$\hat{W}=0.0006$	$\hat{W}=-0.0008$	$\hat{Y}=0.0012$	$\hat{Y}=-0.0006$
SLAC	67	0.780	0.806	0.798	0.763	0.833
BCDMS	581	1.230	1.223	1.224	1.228	1.227
NMC	325	1.378	1.349	1.346	1.377	1.364
CHORUS	832	1.228	1.233	1.223	1.230	1.237
NuTeV	76	0.684	0.698	0.651	0.706	0.655
HERA inclusive	1145	1.242	1.247	1.244	1.243	1.245
HERA charm	37	1.445	1.440	1.431	1.433	1.416
HERA bottom	29	1.326	1.324	1.314	1.330	1.323
Total DIS	**3092**	**1.230**	**1.229**	**1.224**	**1.230**	**1.231**
E886 $\sigma^d_{\rm DY}/\sigma^p_{\rm DY}$	15	0.484	0.454	0.475	0.513	0.458
E886 $\sigma^p_{\rm DY}$	89	1.094	1.067	1.060	1.075	1.049
E605 $\sigma^p_{\rm DY}$	85	0.982	1.025	1.013	1.005	1.032
CDF $d\sigma_Z/dy_Z$	29	1.460	1.446	1.452	1.459	1.463
D0 $d\sigma_Z/dy_Z$	28	0.602	0.604	0.600	0.602	0.603
D0 $W\to\mu\nu$ asy.	9	1.488	1.464	1.517	1.507	1.536
ATLAS W,Z 2010	30	0.866	0.875	0.869	0.868	0.863
ATLAS low-mass $Z\to ee$	6	0.949	0.960	0.964	0.949	0.952
ATLAS W,Z 2011 CC	46	1.837	1.921	1.899	1.893	1.864
ATLAS W,Z 2011 CF	15	1.254	1.264	1.269	1.260	1.253
ATLAS $W+c$ rapidity	22	0.468	0.470	0.445	0.478	0.460
ATLAS $Z\,p_T$	92	0.942	0.973	0.967	0.951	0.978
ATLAS $p_{T,W}$ in W+jets	32	1.690	1.675	1.737	1.640	1.697
CMS W e asy.	11	0.838	0.829	0.827	0.856	0.825
CMS W μ asy.	11	1.736	1.747	1.746	1.760	1.735
CMS σ_{W+c} 7 TeV	5	0.478	0.483	0.487	0.471	0.501
CMS $\sigma_{W^++c}/\sigma_{W^-+c}$ 7 TeV	5	1.687	1.686	1.728	1.678	1.735
CMS $Z\,p_T$	28	1.296	1.280	1.276	1.293	1.273
CMS $W\to\mu\nu$ rapidity	22	1.070	1.054	1.088	1.031	1.048
CMS $W+c$ rapidity 13 TeV	5	0.658	0.671	0.680	0.654	0.695
LHCb $Z\to\mu\mu$	9	1.652	1.648	1.662	1.651	1.645
LHCb $W,Z\to\mu$	29	2.209	2.165	2.176	2.161	2.257
LHCb $Z\to ee$	17	1.124	1.139	1.129	1.130	1.143
LHCb $W,Z\to\mu$	30	1.769	1.773	1.748	1.741	1.823
Total DY (excl. HM)	**670**	**1.191**	**1.199**	**1.198**	**1.192**	**1.202**
ATLAS DY high-mass 7 TeV	13	1.885	2.090	1.643	2.127	1.781
ATLAS DY high-mass 8 TeV	46	1.181	1.230	1.243	1.247	1.181
CMS DY high-mass 7 TeV	117	1.589	1.585	1.553	1.598	1.577
CMS DY high-mass 8 TeV	41	0.805	0.793	0.833	0.820	0.787
CMS DY high-mass 13 TeV	43	2.013	1.876	2.203	1.881	2.064
Total DY (HM-only)	**260**	**1.478**	**1.471**	**1.497**	**1.486**	**1.473**
Total	**4022**	**1.243**	**1.244**	**1.242**	**1.245**	**1.247**

Table E.0.3. Same as Table E.0.2, now comparing the χ^2 values (computed using the t_0 prescription, as above) of the SM PDFs with those of the SMEFT PDFs for different values of the $\hat{W}$ and $\hat{Y}$ parameters in benchmark scenario I, specifically those displayed in Figs. 4.4.3 and G.0.1.

Appendix F

Validation of the SMEFT K-factors

As described in Sect. 4.3.2, in this work the effect of the dimension-six SMEFT operators considered in the two benchmark scenarios is accounted for at the level of cross sections via the K-factor approach, Eq. (4.3.10). In this appendix, we provide further details about the calculation and validation of these EFT K-factors. Specifically, we compare the numerical values for these K-factors, which have been obtained using `SMEFTsim` [164, 376] interfaced with MadGraph5_aMC@NLO, with the analytic calculation presented in [83].

Drell-Yan distributions. As demonstrated in Ref. [83], in the case of the dilepton $m_{\ell\ell}$ distribution in neutral-current Drell-Yan, one can derive the following analytic expression

$$\left(\frac{d\sigma}{dm_{\ell\ell}}\right)_{\text{SMEFT}} = \left(\frac{d\sigma}{dm_{\ell\ell}}\right)_{\text{SM}} \times \frac{\sum_{q,\ell}\mathcal{L}_{q\bar{q}}(m^2_{\ell\ell}/s, m_{\ell\ell})|F_{q\ell}(m_{\ell\ell}, \epsilon^{q\ell})|^2}{\sum_{q,\ell}\mathcal{L}_{q\bar{q}}(m^2_{\ell\ell}/s, m_{\ell\ell})|F_{q\ell}(m_{\ell\ell}, 0)|^2}\,, \quad \text{(F.0.1)}$$

where $F_{ql}(m_{ll}, \epsilon^{q\ell})$ represents a form factor that depends on the values of the SMEFT coefficients

$$F_{q\ell}(m_{\ell\ell}, \epsilon^{q\ell}) = \delta^{ij}\frac{e^2 Q_q Q_\ell}{m^2_{\ell\ell}} + \delta^{ij}\frac{g^q_Z g^\ell_Z}{m^2_{\ell\ell} - m^2_Z + i m_Z \Gamma_Z} + \delta^{ij}\frac{\epsilon^{q\ell}}{v^2}\,. \quad \text{(F.0.2)}$$

One can then match the Warsaw-basis parametrisation of Eq. (4.2.3) to the contact terms $\epsilon^{q\ell}_{ijkl} \equiv \epsilon^{q\ell}\delta_{ij}\delta_{kl}$ in the above equation, where i, j (k, l) are the quark (lepton) flavour indices. There are four combinations for quarks $q = u_L, u_R, d_L, d_R$ and two combinations for charged leptons $\ell = e_L, e_R$ relevant for the description of the neutral-

current Drell-Yan process. Specifically, the matching is

$$\begin{aligned}
&\epsilon^{u_L e_L} = \hat{W} + \frac{1}{3}t_\theta^2\hat{Y}\,, \quad \epsilon^{u_R e_L} = \frac{4}{3}t_\theta^2\hat{Y}\,, \quad \epsilon^{d_L e_L} = -\hat{W} + \frac{1}{3}t_\theta^2\hat{Y}\,, \quad \epsilon^{d_R e_L} = -\frac{2}{3}t_\theta^2\hat{Y}\,,\\
&\epsilon^{u_L e_R} = \frac{2}{3}t_\theta^2\hat{Y}\,, \qquad \epsilon^{u_R e_R} = \frac{8}{3}t_\theta^2\hat{Y}\,, \quad \epsilon^{d_L e_R} = \frac{2}{3}t_\theta^2\hat{Y}\,, \qquad \epsilon^{d_R e_R} = -\frac{4}{3}t_\theta^2\hat{Y}\,,
\end{aligned} \tag{F.0.3}$$

where t_θ is the tangent of the Weinberg angle. We confirm these results using the modified form of the Feynman propagator for W^3 and B fields following directly from Eq. (4.2.1). In the unbroken phase, where $\hat{s} \gg m_W^2$, this simply amounts to the replacement $1/p^2 \to 1/p^2 - \hat{Y}/m_W^2$ for B and $1/p^2 \to 1/p^2 - \hat{W}/m_W^2$ for W^3 in the Standard Model calculation of $pp \to \ell^+\ell^-$. Note that there are two diagrams at the leading order with W^3 and B vector bosons propagating in the s-channel.

An analogous expression involving the transverse mass m_T can be derived in the CC case. These K-factors can be matched to the experimental data kinematics by integrating over the suitable ranges in $m_{\ell\ell}$ (m_T) and rapidity $y_{\ell\ell}$ (y_ℓ).

1D distributions. For the benchmarking between the analytical and numerical calculations, we use the relation between the $\hat{W}$ coefficient and the $c_{lq}^{(3)}$ coefficient in the Warsaw basis, $c_{lq}^{(3)} = -\hat{W}\Lambda^2/v^2$, where Λ is the SMEFT cut-off and v is the Higgs vev. We set $\Lambda = 1$ TeV and determine v using the (α_{EW}, m_Z, G_F) input scheme. We set m_Z and G_F using their PDG values: $m_Z = 91.1876$ GeV and $G_F = 1.1663787 \times 10^{-5}$ GeV^{-2}, and set $\alpha_{EW} = 1/127.951$. These values yield $v = 246.22$ GeV for the Higgs vev.

Fig. F.0.1 displays the comparison of the SMEFT K-factors, Eq. (4.3.10), linearised in the EFT parameters, between the numerical and the analytical approaches for a representative value of $\hat{W} = -10^{-3}$ ($c_{lq}^{(3)} = 1.65 \times 10^{-2}$) and the kinematics of the ATLAS 7 TeV DY data. The label "cuts" indicates that we impose acceptance requirements of $p_T^\ell \geq 25$ GeV and $|\eta_\ell| \leq 2.5$ in the numerical (`SMEFTsim`) calculation on the final-state leptons; these cuts cannot be applied in the analytical calculation. The right panel shows the relative difference in these K-factors, with the analytical calculation as a reference.

From this comparison we observe, first of all, the perfect agreement between the analytical and numerical K-factors in the case of no acceptance cuts, and second, that the acceptance cuts on the leptonic variances leave the K-factor value essentially unchanged. Furthermore, we have verified that the same level of agreement in the calculation of the EFT K-factors between the numerical and analytical approaches is obtained in the case of the $\hat{Y}$ parameter, as well as once the quadratic EFT corrections are accounted for. In addition, the calculation of the EFT K-factors for the neutral

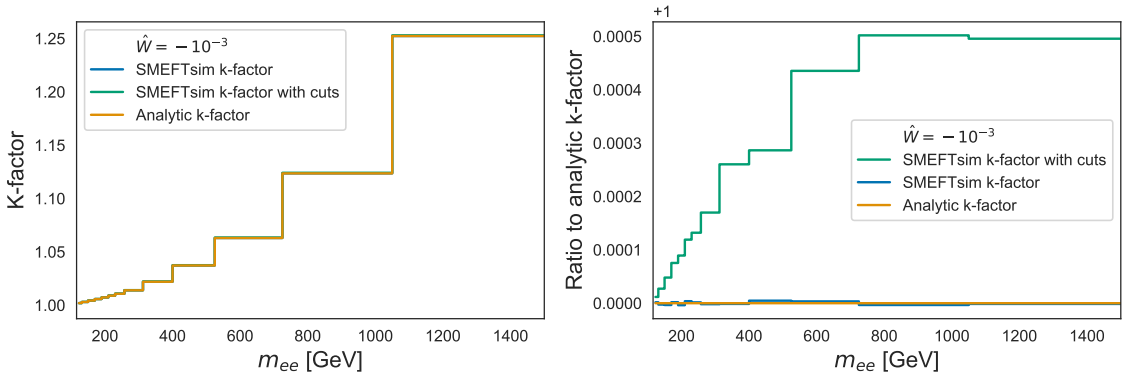


Fig. F.0.1. Left: comparison of the SMEFT K-factors, Eq. (4.3.10), linearised in the EFT parameters, between the numerical and the analytical approaches for $\hat{W} = -10^{-3}$ and the kinematics of the ATLAS 7 TeV DY data. The label "cuts" indicates that we impose acceptance requirements of $p_T^\ell \geq 25$ GeV and $|\eta_\ell| \leq 2.5$ in the numerical (`SMEFTsim`) calculation. Right: relative difference in the K-factors shown in the left panel taking the analytical calculation as a reference.

current DY pseudo-data used in the HL-LHC projections of Sect. 4.5 has been validated by ensuring that the analytical and `SMEFTsim` calculations are in perfect agreement.

As discussed in [83], far above the Z peak it is sufficient to include higher-order corrections to the SM cross section in Eq. (F.0.1) to achieve good theoretical accuracy in the SMEFT, due to the (approximate) factorisation of higher-order QCD corrections in this region. We also note that renormalisation group evolution effects [29, 30] are not required in this calculation, since for the operators considered in our benchmark scenarios the corresponding anomalous dimensions are either Yukawa-suppressed or suppressed by NLO electroweak contributions.

2D distributions. In Drell-Yan datasets such as the CMS 7 TeV data of [259], the measurement is presented as a double-differential distribution in the dilepton invariant mass $m_{\ell\ell}$ (or equivalently $\tau = m_{\ell\ell}^2/s$) and rapidity $|y_{\ell\ell}|$. Also in this case, bin-by-bin EFT K-factors can be computed using the prescription of Eq. (4.3.10). Fig. F.0.2 displays the same comparison as in Fig. F.0.1 (also for $\hat{W} = -10^{-3}$) now for the EFT K-factors for the double-differential Drell-Yan cross sections in $m_{\ell\ell}$ and $|y_{\ell\ell}|$, for the highest $m_{\ell\ell}$ bin of the ATLAS 7 TeV DY measurement. Note that we only compute these K-factors for the $|y_{\ell\ell}|$ that satisfy the LO kinematics. As in the case of the 1D distributions, we find good agreement between the analytical and numerical

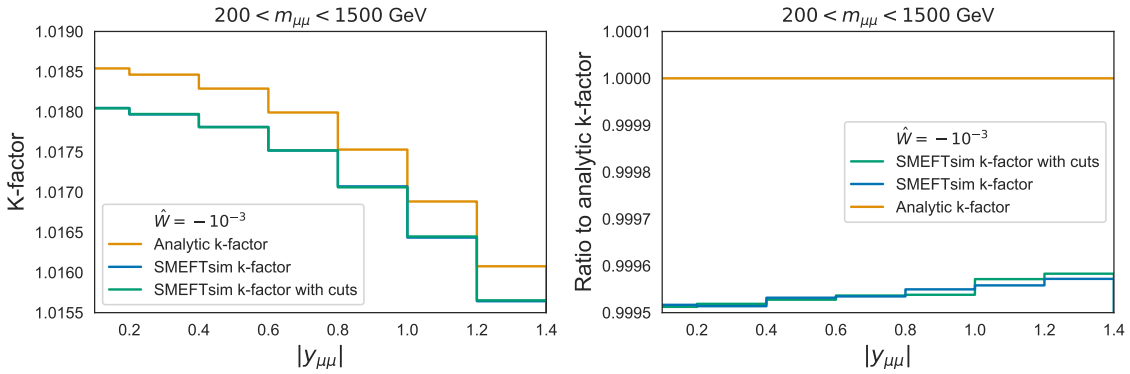


Fig. F.0.2. Same as Fig. F.0.1 (also for $\hat{W} = 10^{-3}$) for the EFT K-factors for the double-differential Drell-Yan cross sections in $m_{\ell\ell}$ and $|y_{\ell\ell}|$, for the highest $m_{\ell\ell}$ bin of the CMS 7 TeV DY measurement. We only compute K-factors for $|y_{\ell\ell}|$ satisfying LO kinematics.

calculations, with differences well below the absolute magnitude of the K-factors, and also that the impact of the acceptance cuts in the leptonic variables is negligible.

EFT K-factors for CC Drell-Yan at the HL-LHC. The left panel of Fig. F.0.3 displays the EFT K-factor for high-mass charged-current DY production at the HL-LHC as a function of m_T, the transverse mass of the neutrino-lepton pair, for a parameter value of $\hat{W} = -10^{-3}$. We find rather large EFT corrections, with K-factors as large as $K_{\rm EFT} \simeq 5$ for the highest m_T bin. To validate this `SMEFTsim`-based calculation, we show a comparison between k_{EFT} and the linear k-factor $k_{Ricci\ et.\ al}$ provided by the authors of [9]. In the right panel we plot the relative k-factor $\frac{k_{EFT}}{k_{Ricci\ et.et}} - 1$, finding good agreement between the two calculations.

Effect of varying PDFs in the computation of the SMEFT K-factors. The SMEFT K-factors in Eq. (4.3.9) are precomputed before the fit using a reference SM PDF set and then kept fixed. Here, we quantitatively assess the effect of varying the input NNLO PDF in Eqns. (4.3.7) and (4.3.8).

We will first spell out the approximation made when using SMEFT K-factors, and then assess its impact. For definiteness, consider corrections from a non-zero $\hat{W}$ coefficient. Denote the partonic cross section by $\hat{\sigma}_{\rm SMEFT} = \hat{\sigma}_{SM} + \hat{W}\hat{\sigma}_{\hat{W}}$ and let $\mathcal{L}_{\rm SMEFT}$ denote the luminosity calculated with SMEFT PDFs. The total cross section for $\hat{W} \neq 0$ is given by

$$\begin{aligned}\sigma(\hat{W}) &= (\hat{\sigma}_{\rm SM} + \hat{W}\hat{\sigma}_{\hat{W}})_{ij} \otimes \mathcal{L}_{\rm SMEFT,ij} \\ &= \hat{\sigma}_{\rm SM,ij} \otimes \mathcal{L}_{\rm SMEFT,ij} + \hat{W}\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\rm SMEFT,ij},\end{aligned} \tag{F.0.4}$$

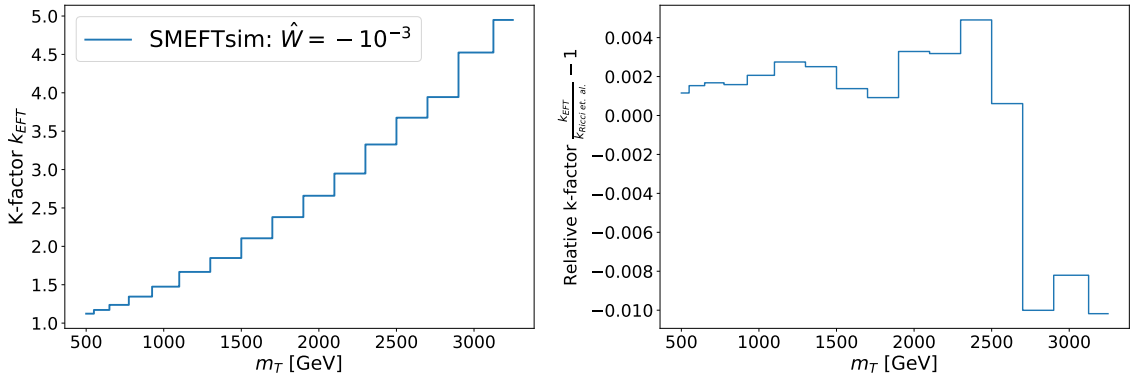


Fig. F.0.3. Left: the EFT K-factor for high-mass charged-current DY production at the HL-LHC as a function of m_T, the transverse mass of the neutrino-lepton pair, for $\hat{W} = -10^{-3}$. Right: the same quantity, now compared to the corresponding k-factors provided by the authors of [9].

where we sum over the partons i, j. By using the SMEFT K-factor approach we approximate Eq. (F.0.4) as

$$\begin{aligned}\sigma(\hat{W}) &\approx \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} \left(1 + \hat{W} \frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}\right) \\ &= \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} + W \hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}} \left(\frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}\right).\end{aligned} \tag{F.0.5}$$

Note that the first term of Eqs. (F.0.4) and (F.0.5) are equal, thus we can express the approximation as

$$\frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SMEFT,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SMEFT,ij}}} \approx \frac{\hat{\sigma}_{\hat{W},ij} \otimes \mathcal{L}_{\text{SM,ij}}}{\hat{\sigma}_{\text{SM,ij}} \otimes \mathcal{L}_{\text{SM,ij}}}, \tag{F.0.6}$$

or, equivalently, as

$$R_{\hat{W}}(\text{SMEFT}) \approx R_{\hat{W}}(\text{SM}) \tag{F.0.7}$$

where R is defined in Eqns. (4.3.7) and (4.3.8) and is computed either with SMEFT PDFs or fixed SM PDFs. In what follows we test whether R(SM) is a good approximation for R(SMEFT), taking each of the coefficients of Scenario I and II at a time.

We will first consider Scenario I and the $\hat{W}$ parameter. We use the PDFs including HL-LHC pseudodata and calculate $R_{\hat{W}}$ for the HL-LHC NC Drell-Yan bins outlined in Section 4.5. In Fig. F.0.4 we observe that, using fixed SM PDFs in the computation of $R_{\hat{W}}$ yields a 2% deviation in the highest invariant mass bin. In the same figure we assess

the impact of these differences on the SMEFT K-factors themselves, $K_{\hat{W}} = 1 + \hat{W} R_{\hat{W}}$, calculated at each of the benchmark points $\hat{W} = \pm 4 \cdot 10^{-5}$ and we observe that the difference at the level of the observable is completely negligible, at the permil level.

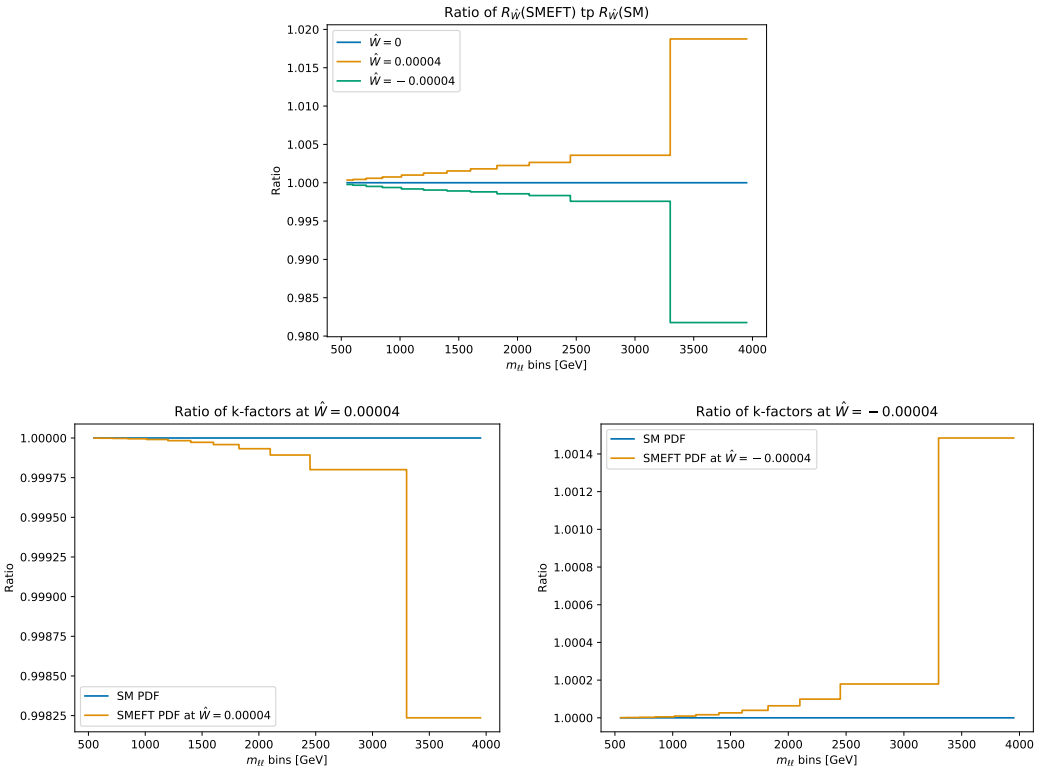


Fig. F.0.4. The impact of the SMEFT PDFs on R_W (above) and the SMEFT K-factor $K = 1 + \hat{W} R_W$ (below) calculated at the Scenario I benchmark points $\hat{W} = \pm 4 \cdot 10^{-5}$.

In the case of the $\hat{Y}$ parameter we observe a rather larger deviation in the highest invariant mass bin, which reaches to 4%, as shown in Fig F.0.5. However, as in the case of $\hat{W}$, the impact of these discrepancies on the SMEFT K-factors, calculated at each of the benchmark points $\hat{Y} = \pm 1.2 \cdot 10^{-4}$, thus on the observable is below the percent level, which is still negligible compared to the experimental and theoretical uncertainties associated to the last bin if the invariant mass distribution.

Finally, we turn to Scenario II, in which we expect to observe the largest deviation between our approximation, based on using fixed SM PDFs in the computation of the SMEFT K-factors, and the full calculation. This is expected because we include both the linear and quadratic terms in the EFT expansion, and because of the flavour non-universal structure of Scenario II. Here we will denote the K-factor by $K = 1 + C_{33}^{D\mu} R^{\text{lin}} + (C_{33}^{D\mu})^2 R^{\text{quad}}$ and calculate the dependence of each R^{lin}, R^{quad} on the SMEFT PDFs. We observe a 10% deviation in the computation of R in the highest

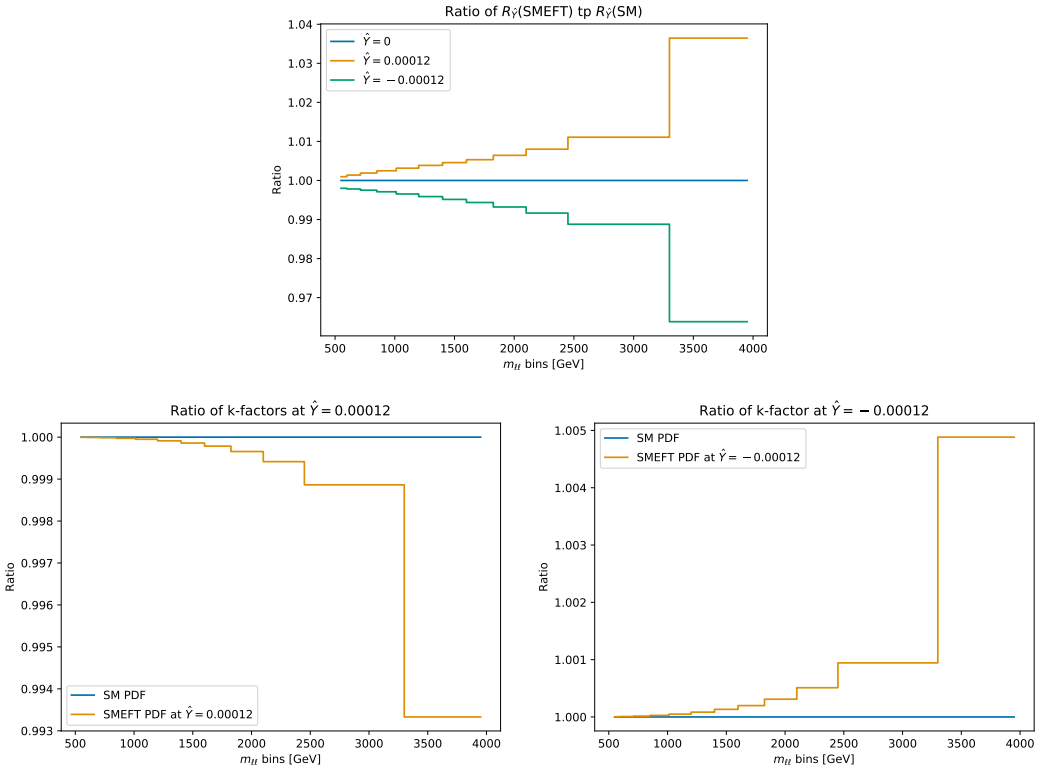


Fig. F.0.5. Same as Fig. F.0.4, in the case of $\hat{Y}$ parameter.

invariant mass bin, as shown in Fig F.0.6. However, the impact of such differences on the actual SMEFT K-factors, thus on the observables, calculated at each of the benchmark points $C_{33}^{D\mu} = \pm 0.014$, is still at the percent level, which is acceptable compared to the experimental and theoretical uncertainties.

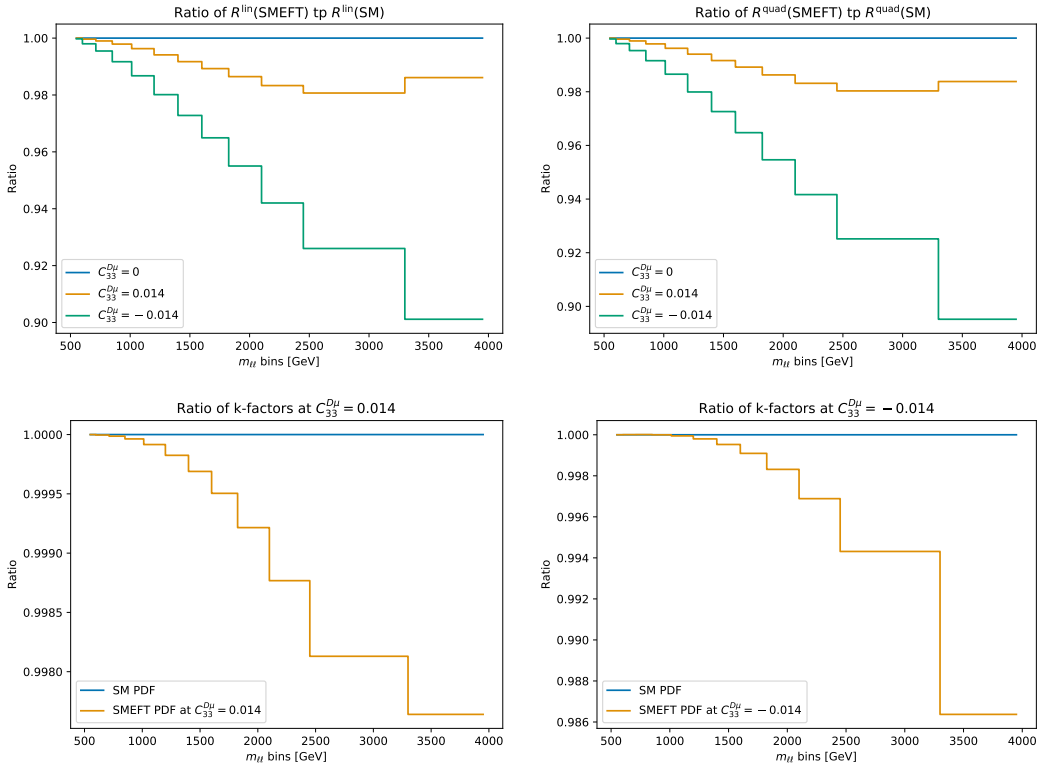


Fig. F.0.6. The impact of the SMEFT PDFs on R^{lin}, R^{quad} (above) and the SMEFT K-factor $K = 1 + C_{33}^{D\mu} R^{\text{lin}} + (C_{33}^{D\mu})^2 R^{\text{quad}}$ (below) calculated at the Scenario II benchmark points $C_{33}^{D\mu} = \pm 0.014$.

Appendix G

Flavour dependence of the SMEFT PDFs

In §4.4 of Chapter 4, when discussing the results of the joint PDF and EFT fits to available Drell-Yan cross-section data, we presented the comparison between the SM and SMEFT PDFs in benchmark scenario I for different values of the $\hat{W}$ and $\hat{Y}$ parameters in terms of the partonic luminosities, Fig. 4.4.3. In this appendix we present the corresponding comparisons between SM and SMEFT PDFs at the level of individual PDF flavours. Fig. G.0.1 displays a comparison between the SM and the SMEFT PDFs at $Q = 100$ GeV for representative values of the $\hat{W}$ (upper) and of $\hat{Y}$ (lower panels) parameters. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 95% CL intervals reported in Table 4.4.1. The error band in the SM PDFs corresponds to the 68% CL PDF uncertainty, while for the SMEFT PDFs only the central values are shown.

In all cases, one finds that the EFT-induced shifts on the PDFs are smaller than their uncertainties, though in some cases these shifts can represent up to one-third of a standard deviation. In particular, the up and down antiquarks in the region $x \gtrsim 10^{-2}$ are the PDF flavours most affected by the EFT effects. This finding can be understood from the fact that the NC Drell-Yan cross section is proportional to the $u\bar{u}$ and $d\bar{d}$ combinations at leading order, but the up and down quark PDF are already well constrained by lower-energy DIS measurements. Furthermore, we have verified that the PDF uncertainties themselves are unchanged in the SMEFT fits. The results of Fig. G.0.1 are consistent with those of Table 4.4.1 and demonstrate that, with current data, the interplay between EFT effects and PDFs in the high-mass Drell-Yan tails is appreciable but remains subdominant as compared to other sources of uncertainty.

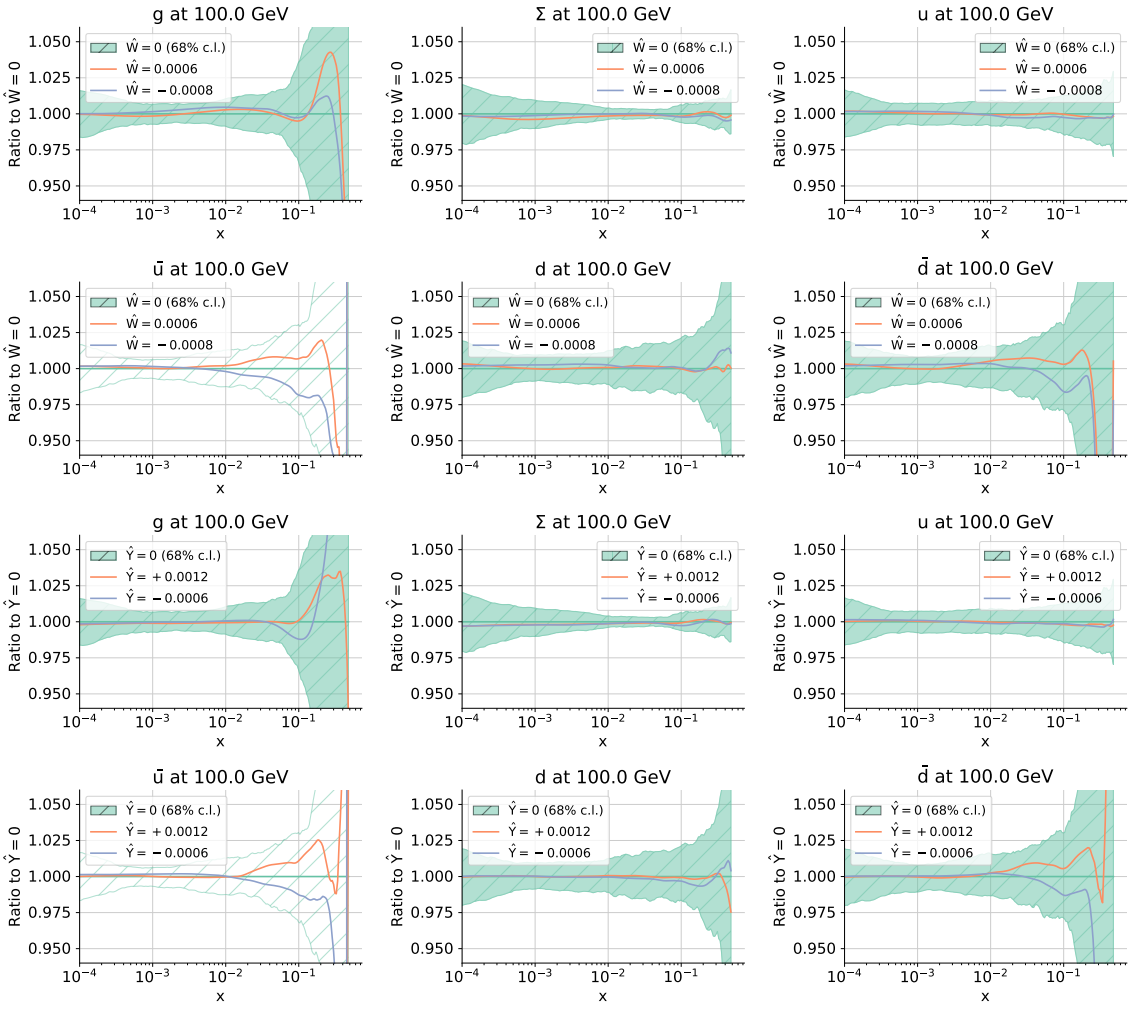


Fig. G.0.1. Comparison between the SM and the SMEFT PDFs at $Q = 100$ GeV, displayed as ratios to the central value of the SM PDFs, for representative values of the $\hat{W}$ (upper) and of $\hat{Y}$ (lower panels) parameters. We show the gluon, the total quark singlet, the up quark and antiquark, and the down quark and antiquark PDFs. The values of $\hat{W}$ and $\hat{Y}$ are chosen to be close to the upper and lower limits of the 95% CL intervals reported in Table 4.4.1.

References

[1] J. Ellis, M. Madigan, K. Mimasu, V. Sanz, and T. You, *Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory*, *JHEP* **04** (2021) 279, [arXiv:2012.02779].

[2] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *Parton distributions in the SMEFT from high-energy Drell-Yan tails*, *JHEP* **07** (2021) 122, [arXiv:2104.02723].

[3] B. C. Allanach, T. Corbett, and M. Madigan, *Sensitivity of Future Hadron Colliders to Leptoquark Pair Production in the Di-Muon Di-Jets Channel*, *Eur. Phys. J. C* **80** (2020), no. 2 170, [arXiv:1911.04455].

[4] I. Brivio, T. Corbett, and M. Trott, *The Higgs width in the SMEFT*, *JHEP* **10** (2019) 056, [arXiv:1906.06949].

[5] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Rev. D* **97** (2018), no. 11 112003, [arXiv:1803.08856].

[6] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, *Top-pair production at the LHC through NNLO QCD and NLO EW*, *JHEP* **10** (2017) 186, [arXiv:1705.04105].

[7] **ATLAS** Collaboration, G. Aad et al., *Search for high-mass dilepton resonances using 139* fb^{-1} *of pp collision data collected at* $\sqrt{s}$ *=13 TeV with the ATLAS detector*, *Phys. Lett. B* **796** (2019) 68–87, [arXiv:1903.06248].

[8] M. Farina, G. Panico, D. Pappadopulo, J. T. Ruderman, R. Torre, and A. Wulzer, *Energy helps accuracy: electroweak precision tests at hadron colliders*, *Phys. Lett. B* **772** (2017) 210–215, [arXiv:1609.08157].

[9] L. Ricci, R. Torre, and A. Wulzer, *On the W&Y interpretation of high-energy Drell-Yan measurements*, arXiv:2008.12978.

[10] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [arXiv:1008.4884].

[11] D. Barducci et al., *Interpreting top-quark LHC measurements in the standard-model effective field theory*, arXiv:1802.07237.

[12] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali, *The Strangest Proton?*, *Eur. Phys. J. C* **80** (2020), no. 12 1168, [arXiv:2009.00014].

[13] **CMS** Collaboration, A. M. Sirunyan et al., *Observation of* $\mathrm{t\bar{t}}H$ *production*, *Phys. Rev. Lett.* **120** (2018), no. 23 231801, [arXiv:1804.02610].

[14] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the W-boson mass in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, *Eur. Phys. J. C* **78** (2018), no. 2 110, [arXiv:1701.07240]. [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[15] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769.

[16] **Muon g-2** Collaboration, B. Abi et al., *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (2021), no. 14 141801, [arXiv:2104.03281].

[17] **ALEPH** Collaboration, R. Barate et al., *Observation of an excess in the search for the standard model Higgs boson at ALEPH*, *Phys. Lett. B* **495** (2000) 1–17, [hep-ex/0011045].

[18] **CDF, D0** Collaboration, *Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6* fb^{-1} *of Data*, 7, 2011. arXiv:1107.5518.

[19] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson*, arXiv:1201.6045.

[20] A. V. Manohar, *Introduction to Effective Field Theories*, arXiv:1804.05863.

[21] A. A. Petrov and A. E. Blechman, *Effective Field Theories*. WSP, 2016.

[22] C. P. Burgess, *Introduction to Effective Field Theory*, *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 329–362, [hep-th/0701053].

[23] B. Gripaios, *Lectures on Effective Field Theory*, arXiv:1506.05039.

[24] D. B. Kaplan, *Five lectures on effective field theory*, 10, 2005. nucl-th/0510023.

[25] R. Penco, *An Introduction to Effective Field Theories*, arXiv:2006.16285.

[26] S. Willenbrock and C. Zhang, *Effective Field Theory Beyond the Standard Model*, *Ann. Rev. Nucl. Part. Sci.* **64** (2014) 83–100, [arXiv:1401.0470].

[27] A. Falkowski, *Lectures on Effective Field Theory*, .

[28] E. E. Jenkins, A. V. Manohar, and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence*, *JHEP* **10** (2013) 087, [arXiv:1308.2627].

[29] E. E. Jenkins, A. V. Manohar, and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence*, *JHEP* **01** (2014) 035, [arXiv:1310.4838].

[30] R. Alonso, E. E. Jenkins, A. V. Manohar, and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology*, *JHEP* **04** (2014) 159, [arXiv:1312.2014].

[31] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1–98, [arXiv:1706.08945].

[32] B. Henning, X. Lu, and H. Murayama, *How to use the Standard Model effective field theory*, *JHEP* **01** (2016) 023, [arXiv:1412.1837].

[33] A. Falkowski, *Effective field theory approach to LHC Higgs data*, *Pramana* **87** (2016), no. 3 39, [arXiv:1505.00046].

[34] T. Cohen, N. Craig, X. Lu, and D. Sutherland, *Is SMEFT Enough?*, arXiv:2008.08597.

[35] M. B. Einhorn and J. Wudka, *The Bases of Effective Field Theories*, *Nucl. Phys. B* **876** (2013) 556–574, [arXiv:1307.0478].

[36] B. Gripaios and D. Sutherland, *DEFT: A program for operators in EFT*, *JHEP* **01** (2019) 128, [arXiv:1807.07546].

[37] G. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, *The Strongly-Interacting Light Higgs*, *JHEP* **06** (2007) 045, [hep-ph/0703164].

[38] R. Contino, M. Ghezzi, C. Grojean, M. Muhlleitner, and M. Spira, *Effective Lagrangian for a light Higgs-like scalar*, *JHEP* **07** (2013) 035, [arXiv:1303.3876].

[39] K. Hagiwara, S. Ishihara, R. Szalapski, and D. Zeppenfeld, *Low energy effects of new interactions in the electroweak boson sector*, *Phys. Rev. D* **48** (Sep, 1993) 2182–2203.

[40] A. Falkowski, B. Fuks, K. Mawatari, K. Mimasu, F. Riva, and V. Sanz, *Rosetta: an operator basis translator for Standard Model effective field theory*, *Eur. Phys. J. C* **75** (2015), no. 12 583, [arXiv:1508.05895].

[41] B. Henning, X. Lu, T. Melia, and H. Murayama, *2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT*, *JHEP* **08** (2017) 016, [arXiv:1512.03433]. [Erratum: JHEP 09, 019 (2019)].

[42] L. Lehman and A. Martin, *Hilbert Series for Constructing Lagrangians: expanding the phenomenologist's toolbox*, *Phys. Rev. D* **91** (2015) 105014, [arXiv:1503.07537].

[43] L. Lehman and A. Martin, *Low-derivative operators of the Standard Model effective field theory via Hilbert series methods*, *JHEP* **02** (2016) 081, [arXiv:1510.00372].

[44] B. Henning, X. Lu, T. Melia, and H. Murayama, *Hilbert series and operator bases with derivatives in effective field theories*, *Commun. Math. Phys.* **347** (2016), no. 2 363–388, [arXiv:1507.07240].

[45] L. Lehman, *Extending the Standard Model Effective Field Theory with the Complete Set of Dimension-7 Operators*, *Phys. Rev. D* **90** (2014), no. 12 125023, [arXiv:1410.4193].

[46] Y. Liao and X.-D. Ma, *Renormalization Group Evolution of Dimension-seven Baryon- and Lepton-number-violating Operators*, *JHEP* **11** (2016) 043, [arXiv:1607.07309].

[47] C. W. Murphy, *Dimension-8 operators in the Standard Model Effective Field Theory*, *JHEP* **10** (2020) 174, [arXiv:2005.00059].

[48] H.-L. Li, Z. Ren, J. Shu, M.-L. Xiao, J.-H. Yu, and Y.-H. Zheng, *Complete Set of Dimension-8 Operators in the Standard Model Effective Field Theory*, arXiv:2005.00008.

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

[50] **Super-Kamiokande** Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567, [hep-ex/9807003].

[51] J. Gerard, *Fermion mass spectrum in* $SU(2)_L \times U(1)$, *Z. Phys. C* **18** (1983) 145.

[52] R. Chivukula and H. Georgi, *Composite Technicolor Standard Model*, *Phys. Lett. B* **188** (1987) 99–104.

[53] C. Cesarotti, Q. Lu, Y. Nakai, A. Parikh, and M. Reece, *Interpreting the Electron EDM Constraint*, *JHEP* **05** (2019) 059, [arXiv:1810.07736].

[54] J. Ellis, C. W. Murphy, V. Sanz, and T. You, *Updated Global SMEFT Fit to Higgs, Diboson and Electroweak Data*, *JHEP* **06** (2018) 146, [arXiv:1803.03252].

[55] E. da Silva Almeida, A. Alves, N. Rosa Agostinho, O. J. Éboli, and M. Gonzalez-Garcia, *Electroweak Sector Under Scrutiny: A Combined Analysis of LHC and Electroweak Precision Data*, *Phys. Rev. D* **99** (2019), no. 3 033001, [arXiv:1812.01009].

[56] A. Biekoetter, T. Corbett, and T. Plehn, *The Gauge-Higgs Legacy of the LHC Run II*, *SciPost Phys.* **6** (2019), no. 6 064, [arXiv:1812.07587].

[57] A. Falkowski and D. Straub, *Flavourful SMEFT likelihood for Higgs and electroweak data*, *JHEP* **04** (2020) 066, [arXiv:1911.07866].

[58] **ATLAS** Collaboration, *Interpretations of the combined measurement of Higgs boson production and decay*, Tech. Rep. ATLAS-CONF-2020-053, 2020.

[59] **CMS** Collaboration, *Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV*, Tech. Rep. CMS-PAS-HIG-19-005, 2020.

[60] A. Buckley, C. Englert, J. Ferrando, D. J. Miller, L. Moore, M. Russell, and C. D. White, *Constraining top quark effective theory in the LHC Run II era*, *JHEP* **04** (2016) 015, [arXiv:1512.03360].

[61] I. Brivio, S. Bruggisser, F. Maltoni, R. Moutafis, T. Plehn, E. Vryonidou, S. Westhoff, and C. Zhang, *O new physics, where art thou? A global search in the top sector*, *JHEP* **02** (2020) 131, [arXiv:1910.03606].

[62] S. Bißmann, J. Erdmann, C. Grunwald, G. Hiller, and K. Kröninger, *Constraining top-quark couplings combining top-quark and $\boldsymbol{B}$ decay observables*, *Eur. Phys. J. C* **80** (2020), no. 2 136, [arXiv:1909.13632].

[63] 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].

[64] G. Durieux, A. Irles, V. Miralles, A. Peñuelas, R. Pöschl, M. Perelló, and M. Vos, *The electro-weak couplings of the top and bottom quarks – global fit and future prospects*, *JHEP* **12** (2019) 098, [arXiv:1907.10619].

[65] S. van Beek, E. R. Nocera, J. Rojo, and E. Slade, *Constraining the SMEFT with Bayesian reweighting*, *SciPost Phys.* **7** (2019), no. 5 070, [arXiv:1906.05296].

[66] **CMS** Collaboration, *Using associated top quark production to probe for new physics within the framework of effective field theory*, tech. rep., 7, 2020.

[67] J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang, *Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC*, arXiv:2105.00006.

[68] R. Trotta, *Bayesian Methods in Cosmology*, 1, 2017. arXiv:1701.01467.

[69] **Particle Data Group** Collaboration, P. A. Zyla et al., *Review of Particle Physics*, *PTEP* **2020** (2020), no. 8 083C01.

[70] H. B. Prosper, *Practical Statistics for Particle Physicists*, *CERN Yellow Rep. School Proc.* **6** (2019) 261–292.

[71] R. J. Barlow, *Practical Statistics for Particle Physics*, *CERN Yellow Rep. School Proc.* **5** (May, 2019) 149–197. 49 p, [arXiv:1905.12362].

[72] E. Gross, *Practical Statistics for High Energy Physics*, *CERN Yellow Rep. School Proc.* **3** (2018) 199–221.

[73] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J.* **C71** (2011) 1554, [arXiv:1007.1727]. [Erratum: Eur. Phys. J.C73,2501(2013)].

[74] R. Barlow, *Asymmetric statistical errors*, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, pp. 56–59, 6, 2004. physics/0406120.

[75] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *The Annals of Mathematical Statistics* **9** (1938), no. 1 60–62.

[76] S. Kokoska and C. Nevison, *Critical Values For The Chi-Square Distribution*, pp. 58–59. Springer New York, New York, NY, 1989.

[77] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17** (2020) 261–272.

[78] C. Helsens, D. Jamin, M. L. Mangano, T. G. Rizzo, and M. Selvaggi, *Heavy resonances at energy-frontier hadron colliders*, arXiv:1902.11217.

[79] **ATLAS** Collaboration, G. Aad et al., *Search for new non-resonant phenomena in high-mass dilepton final states with the ATLAS detector*, *JHEP* **11** (2020) 005, [arXiv:2006.12946]. [Erratum: JHEP 04, 142 (2021)].

[80] **CMS** Collaboration, A. M. Sirunyan et al., *Search for resonant and nonresonant new phenomena in high-mass dilepton final states at* $\sqrt{s} = 13$ *TeV*, arXiv:2103.02708.

[81] A. Greljo, J. Martin Camalich, and J. D. Ruiz-Álvarez, *Mono-*τ *Signatures at the LHC Constrain Explanations of B-decay Anomalies*, *Phys. Rev. Lett.* **122** (2019), no. 13 131803, [arXiv:1811.07920].

[82] S. Alioli, M. Farina, D. Pappadopulo, and J. T. Ruderman, *Catching a New Force by the Tail*, *Phys. Rev. Lett.* **120** (2018), no. 10 101801, [arXiv:1712.02347].

[83] A. Greljo and D. Marzocca, *High-*p_T *dilepton tails and flavor physics*, *Eur. Phys. J.* **C77** (2017), no. 8 548, [arXiv:1704.09015].

[84] D. A. Faroughy, A. Greljo, and J. F. Kamenik, *Confronting lepton flavor universality violation in B decays with high-*p_T *tau lepton searches at LHC*, *Phys. Lett.* **B764** (2017) 126–134, [arXiv:1609.07138].

[85] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[86] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

[87] M. D. Schwartz, *Quantum Field Theory and the Standard Model*. Cambridge University Press, 3, 2014.

[88] J. C. Collins, D. E. Soper, and G. F. Sterman, *Soft Gluons and Factorization*, *Nucl. Phys. B* **308** (1988) 833–856.

[89] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017), no. 10 663, [arXiv:1706.00428].

[90] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Parton distributions in the LHC era: MMHT 2014 PDFs*, *Eur. Phys. J. C* **75** (2015), no. 5 204, [arXiv:1412.3989].

[91] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, *New parton distribution functions from a global analysis of quantum chromodynamics*, *Phys. Rev. D* **93** (2016), no. 3 033006, [arXiv:1506.07443].

[92] S. Alte, M. König, and W. Shepherd, *Consistent Searches for SMEFT Effects in Non-Resonant Dijet Events*, *JHEP* **01** (2018) 094, [arXiv:1711.07484].

[93] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Can New Physics hide inside the proton?*, *Phys. Rev. Lett.* **123** (2019), no. 13 132001, [arXiv:1905.05215].

[94] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, *Phys. Rev. D* **2** (Oct, 1970) 1285–1292.

[95] S. Gori, *TASI lectures on flavor physics*, *PoS* **TASI2018** (2019) 013.

[96] M. Blanke, *Introduction to Flavour Physics and CP Violation*, *CERN Yellow Rep. School Proc.* **1705** (2017) 71–100, [arXiv:1704.03753].

[97] **LHCb** Collaboration, R. Aaij et al., *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, *JHEP* **06** (2014) 133, [arXiv:1403.8044].

[98] **ATLAS** Collaboration, M. Aaboud et al., *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, *JHEP* **04** (2019) 098, [arXiv:1812.03017].

[99] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment*, *JHEP* **04** (2020) 188, [arXiv:1910.12127].

[100] **LHCb** Collaboration, R. Aaij et al., *Analysis of neutral B-meson decays into two muons*, arXiv:2108.09284.

[101] **LHCb** Collaboration, R. Aaij et al., *Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays*, *JHEP* **06** (2015) 115, [arXiv:1503.07138]. [Erratum: JHEP 09, 145 (2018)].

[102] **LHCb** Collaboration, R. Aaij et al., *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, *JHEP* **09** (2015) 179, [arXiv:1506.08777].

[103] **LHCb** Collaboration, R. Aaij et al., *Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction*, *JHEP* **11** (2016) 047, [arXiv:1606.04731]. [Erratum: JHEP 04, 142 (2017)].

[104] **LHCb** Collaboration, R. Aaij et al., *Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$*, *Phys. Rev. Lett.* **111** (2013) 191801, [arXiv:1308.1707].

[105] **LHCb** Collaboration, R. Aaij et al., *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity*, *JHEP* **02** (2016) 104, [arXiv:1512.04442].

[106] **ATLAS** Collaboration, I. Carli, *Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decay with the ATLAS detector*, *PoS* **FPCP2017** (2017) 043.

[107] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of angular parameters from the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett. B* **781** (2018) 517–541, [arXiv:1710.02846].

[108] **LHCb** Collaboration, R. Aaij et al., *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **122** (2019), no. 19 191801, [arXiv:1903.09252].

[109] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **113** (2014) 151601, [arXiv:1406.6482].

[110] **Belle** Collaboration, A. Abdesselam et al., *Test of Lepton-Flavor Universality in $B \to K^*\ell^+\ell^-$ Decays at Belle*, *Phys. Rev. Lett.* **126** (2021), no. 16 161801, [arXiv:1904.02440].

[111] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, *Eur. Phys. J. C* **77** (2017), no. 2 112, [arXiv:1607.00299].

[112] M. Beneke, C. Bobeth, and R. Szafron, *Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \to \mu^+\mu^-$*, *Phys. Rev. Lett.* **120** (2018), no. 1 011801, [arXiv:1708.09152].

[113] L.-S. Geng, B. Grinstein, S. Jäger, S.-Y. Li, J. Martin Camalich, and R.-X. Shi, *Implications of new evidence for lepton-universality violation in $b \to s\ell+\ell-$ decays*, *Phys. Rev. D* **104** (2021), no. 3 035029, [arXiv:2103.12738].

[114] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, and D. M. Straub, *B-decay discrepancies after Moriond 2019*, arXiv:1903.10434.

[115] W. Altmannshofer and P. Stangl, *New Physics in Rare B Decays after Moriond 2021*, arXiv:2103.13370.

[116] Q.-Y. Hu, X.-Q. Li, and Y.-D. Yang, *$b \to c\tau\nu$ transitions in the standard model effective field theory*, *Eur. Phys. J. C* **79** (2019), no. 3 264, [arXiv:1810.04939].

[117] C. Bobeth and U. Haisch, *New Physics in Γ_{12}^s: $(\bar{s}b)(\bar{\tau}\tau)$ Operators*, *Acta Phys. Polon. B* **44** (2013) 127–176, [arXiv:1109.1826].

[118] F. F. Deppisch, S. Kulkarni, H. Päs, and E. Schumacher, *Leptoquark patterns unifying neutrino masses, flavor anomalies, and the diphoton excess*, *Phys. Rev.* **D94** (2016), no. 1 013003, [arXiv:1603.07672].

[119] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias, and J. Virto, *Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data*, *JHEP* **01** (2018) 093, [arXiv:1704.05340].

[120] G. Hiller and I. Nisandzic, *R_K and R_{K^*} beyond the standard model*, *Phys. Rev.* **D96** (2017), no. 3 035003, [arXiv:1704.05444].

[121] G. D'Amico, M. Nardecchia, P. Panci, F. Sannino, A. Strumia, R. Torre, and A. Urbano, *Flavour anomalies after the R_{K^*} measurement*, *JHEP* **09** (2017) 010, [arXiv:1704.05438].

[122] J. Alda, J. Guasch, and S. Penaranda, *Some results on Lepton Flavour Universality Violation*, *Eur. Phys. J.* **C79** (2019), no. 7 588, [arXiv:1805.03636].

[123] J. Kumar, D. London, and R. Watanabe, *Combined Explanations of the $b \to s\mu^+\mu^-$ and $b \to c\tau^-\bar{\nu}$ Anomalies: a General Model Analysis*, *Phys. Rev.* **D99** (2019), no. 1 015007, [arXiv:1806.07403].

[124] E. Alvarez, L. Da Rold, A. Juste, M. Szewc, and T. Vazquez Schroeder, *A composite pNGB leptoquark at the LHC*, *JHEP* **12** (2018) 027, [arXiv:1808.02063].

[125] L. Da Rold and F. Lamagna, *A vector leptoquark for the B-physics anomalies from a composite GUT*, arXiv:1906.11666.

[126] I. Doršner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Košnik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1–68, [arXiv:1603.04993].

[127] I. de Medeiros Varzielas and J. Talbert, *Simplified Models of Flavourful Leptoquarks*, *Eur. Phys. J.* **C79** (2019), no. 6 536, [arXiv:1901.10484].

[128] B. Gripaios, M. Nardecchia, and S. A. Renner, *Composite leptoquarks and anomalies in B-meson decays*, *JHEP* **05** (2015) 006, [arXiv:1412.1791].

[129] B. C. Allanach and J. Davighi, *Third family hypercharge model for $R_{K^{(*)}}$ and aspects of the fermion mass problem*, *JHEP* **12** (2018) 075, [arXiv:1809.01158].

[130] W. Buchmuller, R. Ruckl, and D. Wyler, *Leptoquarks in Lepton - Quark Collisions*, *Phys. Lett.* **B191** (1987) 442–448. [Erratum: Phys. Lett.B448,320(1999)].

[131] **CMS** Collaboration, V. Khachatryan et al., *Search for single production of scalar leptoquarks in proton-proton collisions at* $\sqrt{s} = 8$ *TeV*, *Phys. Rev. D* **93** (2016), no. 3 032005, [arXiv:1509.03750]. [Erratum: Phys.Rev.D 95, 039906 (2017)].

[132] M. Schmaltz and Y.-M. Zhong, *The leptoquark Hunter's guide: large coupling*, *JHEP* **01** (2019) 132, [arXiv:1810.10017].

[133] **ATLAS** Collaboration, G. Aad et al., *Search for pairs of scalar leptoquarks decaying into quarks and electrons or muons in* $\sqrt{s} = 13$ *TeV pp collisions with the ATLAS detector*, *JHEP* **10** (2020) 112, [arXiv:2006.05872].

[134] A. Cerri et al., *Opportunities in Flavour Physics at the HL-LHC and HE-LHC*, arXiv:1812.07638.

[135] J. Albrecht, F. Bernlochner, M. Kenzie, S. Reichert, D. Straub, and A. Tully, *Future prospects for exploring present day anomalies in flavour physics measurements with Belle II and LHCb*, arXiv:1709.10308.

[136] E. Kou et al., *The Belle II Physics book*, arXiv:1808.10567.

[137] D. Liu, A. Pomarol, R. Rattazzi, and F. Riva, *Patterns of Strong Coupling for LHC Searches*, *JHEP* **11** (2016) 141, [arXiv:1603.03064].

[138] F. Feroz, M. Hobson, and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics*, *Mon. Not. Roy. Astron. Soc.* **398** (2009) 1601–1614, [arXiv:0809.3437].

[139] C. Degrande, G. Durieux, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Automated one-loop computations in the SMEFT*, arXiv:2008.11743.

[140] J. Ellis, V. Sanz, and T. You, *The Effective Standard Model after LHC Run I*, *JHEP* **03** (2015) 157, [arXiv:1410.7703].

[141] F. Krauss, S. Kuttimalai, and T. Plehn, *LHC multijet events as a probe for anomalous dimension-six gluon interactions*, *Phys. Rev. D* **95** (2017), no. 3 035024, [arXiv:1611.00767].

[142] V. Hirschi, F. Maltoni, I. Tsinikos, and E. Vryonidou, *Constraining anomalous gluon self-interactions at the LHC: a reappraisal*, *JHEP* **07** (2018) 093, [arXiv:1806.04696].

[143] R. Goldouzian and M. D. Hildreth, *LHC dijet angular distributions as a probe for the dimension-six triple gluon vertex*, *Phys. Lett. B* **811** (2020) 135889, [arXiv:2001.02736].

[144] **Particle Data Group** Collaboration, P. Zyla et al., *Review of Particle Physics*, *PTEP* **2020** (2020), no. 8 083C01.

[145] J. Ellis and T. You, *Sensitivities of Prospective Future e+e- Colliders to Decoupled New Physics*, *JHEP* **03** (2016) 089, [arXiv:1510.04561].

[146] J. De Blas, G. Durieux, C. Grojean, J. Gu, and A. Paul, *On the future of Higgs, electroweak and diboson measurements at lepton colliders*, *JHEP* **12** (2019) 117, [arXiv:1907.04311].

[147] **FCC** Collaboration, A. Abada et al., *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2*, *Eur. Phys. J. ST* **228** (2019), no. 2 261–623.

[148] Z. Zhang, *Time to Go Beyond Triple-Gauge-Boson-Coupling Interpretation of W Pair Production*, *Phys. Rev. Lett.* **118** (2017), no. 1 011803, [arXiv:1610.01618].

[149] C. Grojean, M. Montull, and M. Riembau, *Diboson at the LHC vs LEP*, *JHEP* **03** (2019) 020, [arXiv:1810.05149].

[150] S. Dawson, P. Giardino, and A. Ismail, *Standard model EFT and the Drell-Yan process at high energy*, *Phys. Rev. D* **99** (2019), no. 3 035044, [arXiv:1811.12260].

[151] S. Alte, M. König, and W. Shepherd, *Consistent Searches for SMEFT Effects in Non-Resonant Dilepton Events*, *JHEP* **07** (2019) 144, [arXiv:1812.07575].

[152] J. Fuentes-Martin, A. Greljo, J. Martin Camalich, and J. D. Ruiz-Alvarez, *Charm physics confronts high-p_T lepton tails*, *JHEP* **11** (2020) 080, [arXiv:2003.12421].

[153] B. Henning, D. Lombardo, M. Riembau, and F. Riva, *Measuring Higgs Couplings without Higgs Bosons*, *Phys. Rev. Lett.* **123** (2019), no. 18 181801, [arXiv:1812.09299].

[154] A. Falkowski, S. Ganguly, P. Gras, J. M. No, K. Tobioka, N. Vignaroli, and T. You, *Light quark Yukawas in triboson final states*, arXiv:2011.09551.

[155] C. Degrande, J.-M. Gerard, C. Grojean, F. Maltoni, and G. Servant, *Non-resonant New Physics in Top Pair Production at Hadron Colliders*, *JHEP* **03** (2011) 125, [arXiv:1010.6304].

[156] J. D'Hondt, A. Mariotti, K. Mimasu, S. Moortgat, and C. Zhang, *Learning to pinpoint effective operators at the LHC: a study of the $t\bar{t}b\bar{b}$ signature*, *JHEP* **11** (2018) 131, [arXiv:1807.02130].

[157] G. Banelli, E. Salvioni, J. Serra, T. Theil, and A. Weiler, *The Present and Future of Four Tops*, arXiv:2010.05915.

[158] S. Bißmann, J. Erdmann, C. Grunwald, G. Hiller, and K. Kröninger, *Correlating uncertainties in global analyses within SMEFT matters*, arXiv:1912.06090.

[159] M. Czakon, D. Heymes, and A. Mitov, *fastNLO tables for NNLO top-quark pair differential distributions*, arXiv:1704.08551.

[160] M. L. Czakon, C. Gütschow, J. M. Lindert, A. Mitov, D. Pagani, A. S. Papanastasiou, M. Schönherr, I. Tsinikos, and M. Zaro, *NNLO versus NLO multi-jet merging for top-pair production including electroweak corrections*, in *11th International Workshop on Top Quark Physics*, 1, 2019. arXiv:1901.04442.

[161] M. Czakon, S. Dulat, T.-J. Hou, J. Huston, A. Mitov, A. S. Papanastasiou, I. Sitiwaldi, Z. Yu, and C.-P. Yuan, *An exploratory study of the impact of CMS double-differential top distributions on the gluon parton distribution function*, *J. Phys. G* **48** (2021), no. 1 015003, [arXiv:1912.08801].

[162] A. Falkowski, M. González-Alonso, and K. Mimouni, *Compilation of low-energy constraints on 4-fermion operators in the SMEFT*, *JHEP* **08** (2017) 123, [arXiv:1706.03783].

[163] O. Domenech, A. Pomarol, and J. Serra, *Probing the SM with Dijets at the LHC*, *Phys. Rev. D* **85** (2012) 074030, [arXiv:1201.6510].

[164] I. Brivio, Y. Jiang, and M. Trott, *The SMEFTsim package, theory and tools*, *JHEP* **12** (2017) 070, [arXiv:1709.06492].

[165] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [arXiv:1207.1303].

[166] C. Englert and M. Spannowsky, *Effective Theories and Measurements at Colliders*, *Phys. Lett. B* **740** (2015) 8–15, [arXiv:1408.5147].

[167] C. Hays, A. Helset, A. Martin, and M. Trott, *Exact SMEFT formulation and expansion to $\mathcal{O}(v^4/\Lambda^4)$*, *JHEP* **11** (2020) 087, [arXiv:2007.00565].

[168] A. Horne, J. Pittman, M. Snedeker, W. Shepherd, and J. W. Walker, *Shift-Type SMEFT Effects in Dileptons at the LHC*, arXiv:2007.12698.

[169] E. Keilmann and W. Shepherd, *Dijets at Tevatron Cannot Constrain SMEFT Four-Quark Operators*, *JHEP* **09** (2019) 086, [arXiv:1907.13160].

[170] C. Hays, A. Martin, V. Sanz, and J. Setford, *On the impact of dimension-eight SMEFT operators on Higgs measurements*, *JHEP* **02** (2019) 123, [arXiv:1808.00442].

[171] J. Baglio, S. Dawson, S. Homiller, S. D. Lane, and I. M. Lewis, *Validity of standard model EFT studies of VH and VV production at NLO*, *Phys. Rev. D* **101** (2020), no. 11 115004, [arXiv:2003.07862].

[172] V. Cirigliano, W. Dekens, J. de Vries, and E. Mereghetti, *Constraining the top-Higgs sector of the Standard Model Effective Field Theory*, *Phys. Rev. D* **94** (2016), no. 3 034031, [arXiv:1605.04311].

[173] R. Aoude, T. Hurth, S. Renner, and W. Shepherd, *The impact of flavour data on global fits of the MFV SMEFT*, arXiv:2003.05432.

[174] L. Silvestrini and M. Valli, *Model-independent Bounds on the Standard Model Effective Theory from Flavour Physics*, *Phys. Lett. B* **799** (2019) 135062, [arXiv:1812.10913].

[175] **ATLAS** Collaboration, *Methodology for EFT interpretation of Higgs boson Simplified Template Cross-section results in ATLAS*, .

[176] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1*, arXiv:1906.02754.

[177] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, *Phys. Lett. B* **579** (2004) 149–156, [hep-ph/0307206].

[178] K. Mimasu, V. Sanz, and C. Williams, *Higher Order QCD predictions for Associated Higgs production with anomalous couplings to gauge bosons*, *JHEP* **08** (2016) 039, [arXiv:1512.02572].

[179] O. Bessidskaia Bylund, F. Maltoni, I. Tsinikos, E. Vryonidou, and C. Zhang, *Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD*, *JHEP* **05** (2016) 052, [arXiv:1601.08193].

[180] C. Englert, R. Rosenfeld, M. Spannowsky, and A. Tonero, *New physics and signal-background interference in associated $pp \to HZ$ production*, *EPL* **114** (2016), no. 3 31001, [arXiv:1603.05304].

[181] A. Azatov, C. Grojean, A. Paul, and E. Salvioni, *Resolving gluon fusion loops at current and future hadron colliders*, *JHEP* **09** (2016) 123, [arXiv:1608.00977].

[182] E. Vryonidou and C. Zhang, *Dimension-six electroweak top-loop effects in Higgs production and decay*, *JHEP* **08** (2018) 036, [arXiv:1804.09766].

[183] **ATLAS** Collaboration, G. Aad et al., *Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Eur. Phys. J. C* **80** (2020), no. 10 957, [arXiv:2004.03447].

[184] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of fiducial and differential W^+W^- production cross-sections at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019), no. 10 884, [arXiv:1905.04242].

[185] **CMS** Collaboration, A. M. Sirunyan et al., *Measurements of the $pp \to WZ$ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV*, *JHEP* **04** (2019) 122, [arXiv:1901.03428].

[186] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of $W^\pm Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019), no. 6 535, [arXiv:1902.05759].

[187] **ATLAS** Collaboration, G. Aad et al., *Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton-proton collisions at ATLAS*, arXiv:2006.15458.

[188] L. Berthier, M. Bjørn, and M. Trott, *Incorporating doubly resonant $W^{\pm}$ data in a global fit of SMEFT parameters to lift flat directions*, *JHEP* **09** (2016) 157, [arXiv:1606.06693].

[189] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, *Top-quark charge asymmetry at the LHC and Tevatron through NNLO QCD and NLO EW*, *Phys. Rev. D* **98** (2018), no. 1 014003, [arXiv:1711.03945].

[190] J. Ellis, V. Sanz, and T. You, *Complete Higgs Sector Constraints on Dimension-6 Operators*, *JHEP* **07** (2014) 036, [arXiv:1404.3667].

[191] A. Azatov, R. Contino, C. S. Machado, and F. Riva, *Helicity selection rules and noninterference for BSM amplitudes*, *Phys. Rev. D* **95** (2017), no. 6 065014, [arXiv:1607.05236].

[192] **CMS** Collaboration, A. M. Sirunyan et al., *Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair*, *Phys. Rev. Lett.* **120** (2018), no. 7 071802, [arXiv:1709.05543].

[193] R. Contino, A. Falkowski, F. Goertz, C. Grojean, and F. Riva, *On the Validity of the Effective Field Theory Approach to SM Precision Tests*, *JHEP* **07** (2016) 144, [arXiv:1604.06444].

[194] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020), no. 5 370, [arXiv:1907.08330].

[195] **CMS** Collaboration, V. Khachatryan et al., *Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **01** (2016) 096, [arXiv:1510.01131].

[196] **ATLAS** Collaboration, G. Aad et al., *Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **11** (2015) 172, [arXiv:1509.05276].

[197] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **99** (2019), no. 7 072009, [arXiv:1901.03584].

[198] C. Degrande, F. Maltoni, K. Mimasu, E. Vryonidou, and C. Zhang, *Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation*, *JHEP* **10** (2018) 005, [arXiv:1804.07773].

[199] J. A. Dror, M. Farina, E. Salvioni, and J. Serra, *Strong tW Scattering at the LHC*, *JHEP* **01** (2016) 071, [arXiv:1511.03674].

[200] F. Maltoni, L. Mantani, and K. Mimasu, *Top-quark electroweak interactions at high energy*, *JHEP* **10** (2019) 004, [arXiv:1904.05637].

[201] E. H. Simmons, *Higher dimension gluon operators and hadronic scattering*, *Phys. Lett. B* **246** (1990) 471–476.

[202] H. K. Dreiner, A. Duff, and D. Zeppenfeld, *How well do we know the three gluon vertex?*, *Phys. Lett. B* **282** (1992) 441–447.

[203] P. L. Cho and E. H. Simmons, *Searching for G3 in $t\bar{t}$ production*, *Phys. Rev. D* **51** (1995) 2360–2370, [hep-ph/9408206].

[204] D. Ghosh and M. Wiebusch, *Dimension-six triple gluon operator in Higgs+jet observables*, *Phys. Rev. D* **91** (2015), no. 3 031701, [arXiv:1411.2029].

[205] D. Bardhan, D. Ghosh, P. Jain, and A. M. Thalapillil, *Towards constraining triple gluon operators through tops*, arXiv:2010.13402.

[206] J. de Blas, J. Criado, M. Perez-Victoria, and J. Santiago, *Effective description of general extensions of the Standard Model: the complete tree-level dictionary*, *JHEP* **03** (2018) 109, [arXiv:1711.10391].

[207] D. Marzocca et al., *BSM Benchmarks for Effective Field Theories in Higgs and Electroweak Physics*, arXiv:2009.01249.

[208] S. Dawson, S. Homiller, and S. D. Lane, *Putting standard model EFT fits to work*, *Phys. Rev. D* **102** (2020), no. 5 055012, [arXiv:2007.01296].

[209] A. Crivellin, F. Kirk, C. A. Manzari, and M. Montull, *Global Electroweak Fit and Vector-Like Leptons in Light of the Cabibbo Angle Anomaly*, arXiv:2008.01113.

[210] A. Drozd, J. Ellis, J. Quevillon, and T. You, *Comparing EFT and Exact One-Loop Analyses of Non-Degenerate Stops*, *JHEP* **06** (2015) 028, [arXiv:1504.02409].

[211] B. Henning, X. Lu, and H. Murayama, *What do precision Higgs measurements buy us?*, arXiv:1404.1058.

[212] J. R. Espinosa, C. Grojean, V. Sanz, and M. Trott, *NSUSY fits*, *JHEP* **12** (2012) 077, [arXiv:1207.7355].

[213] A. Banfi, A. Bond, A. Martin, and V. Sanz, *Digging for Top Squarks from Higgs data: from signal strengths to differential distributions*, *JHEP* **11** (2018) 171, [arXiv:1806.05598].

[214] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [arXiv:1410.8849].

[215] **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].

[216] T.-J. Hou et al., *New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC*, *Phys. Rev. D* **103** (2021), no. 1 014013, [arXiv:1912.10053].

[217] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, *Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs*, arXiv:2012.04684.

[218] R. Abdul Khalek, S. Bailey, J. Gao, L. Harland-Lang, and J. Rojo, *Towards Ultimate Parton Distributions at the High-Luminosity LHC*, *Eur. Phys. J. C* **78** (2018), no. 11 962, [arXiv:1810.03639].

[219] M. E. Peskin and T. Takeuchi, *Estimation of oblique electroweak corrections*, *Phys. Rev.* **D46** (1992) 381–409.

[220] G. Altarelli, R. Barbieri, and S. Jadach, *Toward a model independent analysis of electroweak data*, *Nucl. Phys.* **B369** (1992) 3–32. [Erratum: Nucl. Phys.B376,444(1992)].

[221] R. Barbieri, A. Pomarol, R. Rattazzi, and A. Strumia, *Electroweak symmetry breaking after LEP-1 and LEP-2*, *Nucl. Phys.* **B703** (2004) 127–146, [hep-ph/0405040].

[222] J. D. Wells and Z. Zhang, *Effective theories of universal theories*, *JHEP* **01** (2016) 123, [arXiv:1510.08462].

[223] **ATLAS** Collaboration, G. Aad et al., *Measurement of the double-differential high-mass Drell-Yan cross section in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, *JHEP* **08** (2016) 009, [arXiv:1606.01736].

[224] **CMS** Collaboration, V. Khachatryan et al., *Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV*, *Eur. Phys. J.* **C75** (2015), no. 4 147, [arXiv:1412.1115].

[225] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality with* $B^0 \to K^{*0}\ell^+\ell^-$ *decays*, *JHEP* **08** (2017) 055, [arXiv:1705.05802].

[226] **LHCb** Collaboration, R. Aaij et al., *Test of lepton universality in beauty-quark decays*, arXiv:2103.11769.

[227] R. Barbieri, G. Isidori, J. Jones-Perez, P. Lodone, and D. M. Straub, $U(2)$ *and Minimal Flavour Violation in Supersymmetry*, *Eur. Phys. J. C* **71** (2011) 1725, [arXiv:1105.2296].

[228] A. L. Kagan, G. Perez, T. Volansky, and J. Zupan, *General Minimal Flavor Violation*, *Phys. Rev. D* **80** (2009) 076002, [arXiv:0903.1794].

[229] J. Fuentes-Martín, G. Isidori, J. Pagès, and K. Yamamoto, *With or without U(2)? Probing non-standard flavor and helicity structures in semileptonic B decays*, *Phys. Lett. B* **800** (2020) 135080, [arXiv:1909.02519].

[230] D. A. Faroughy, G. Isidori, F. Wilsch, and K. Yamamoto, *Flavour symmetries in the SMEFT*, *JHEP* **08** (2020) 166, [arXiv:2005.05366].

[231] A. Greljo, P. Stangl, and A. E. Thomsen, *A Model of Muon Anomalies*, arXiv:2103.13991.

[232] **ATLAS** Collaboration, *Search for new high-mass phenomena in the dilepton final state using 36.1 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, .

[233] M. Bordone, A. Greljo, and D. Marzocca, *Exploiting dijet resonance searches for flavor physics*, arXiv:2103.10332.

[234] **H1, ZEUS** Collaboration, H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, *Eur. Phys. J.* **C75** (2015), no. 12 580, [arXiv:1506.06042].

[235] **D0** Collaboration, V. M. Abazov et al., *Measurement of the electron charge asymmetry in $\boldsymbol{p\bar{p} \to W + X \to e\nu + X}$ decays in $\boldsymbol{p\bar{p}}$ collisions at $\boldsymbol{\sqrt{s} = 1.96}$ TeV*, *Phys. Rev. D* **91** (2015), no. 3 032007, [arXiv:1412.2862]. [Erratum: Phys.Rev.D 91, 079901 (2015)].

[236] **NuSea** Collaboration, R. S. Towell et al., *Improved measurement of the anti-d / anti-u asymmetry in the nucleon sea*, *Phys. Rev.* **D64** (2001) 052002, [hep-ex/0103030].

[237] **NuSea** Collaboration, J. C. Webb et al., *Absolute Drell-Yan dimuon cross-sections in 800 GeV / c pp and pd collisions*, hep-ex/0302019.

[238] J. C. Webb, *Measurement of continuum dimuon production in 800-GeV/C proton nucleon collisions*. PhD thesis, New Mexico State U., 2003. hep-ex/0301031.

[239] G. Moreno et al., *Dimuon production in proton - copper collisions at $\sqrt{s} = 38.8$-GeV*, *Phys. Rev. D* **43** (1991) 2815–2836.

[240] **CDF** Collaboration, T. A. Aaltonen et al., *Measurement of $d\sigma/dy$ of Drell-Yan e^+e^- pairs in the Z Mass Region from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Lett.* **B692** (2010) 232–239, [arXiv:0908.3914].

[241] **D0** Collaboration, V. M. Abazov et al., *Measurement of the Shape of the Boson Rapidity Distribution for $p\bar{p} \to Z/\gamma^* \to e^+e^- + X$ Events Produced at $\sqrt{s}$ of 1.96-TeV*, *Phys. Rev.* **D76** (2007) 012003, [hep-ex/0702025].

[242] **D0** Collaboration, V. M. Abazov et al., *Measurement of the Muon Charge Asymmetry in $p\bar{p} \to W+X \to \mu\nu + X$ Events at $\sqrt{s}$=1.96 TeV*, *Phys. Rev. D* **88** (2013) 091102, [arXiv:1309.2591].

[243] **ATLAS** Collaboration, G. Aad et al., *Measurement of the inclusive $W^{\pm}$ and Z/gamma cross sections in the electron and muon decay channels in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Rev.* **D85** (2012) 072004, [arXiv:1109.5141].

[244] **ATLAS** Collaboration, G. Aad et al., *Measurement of the low-mass Drell-Yan differential cross section at $\sqrt{s} = 7$ TeV using the ATLAS detector*, *JHEP* **06** (2014) 112, [arXiv:1404.1212].

[245] **ATLAS** Collaboration, M. Aaboud et al., *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, *Eur. Phys. J.* **C77** (2017), no. 6 367, [arXiv:1612.03016].

[246] **ATLAS** Collaboration, G. Aad et al., *Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *JHEP* **05** (2014) 068, [arXiv:1402.6263].

[247] **ATLAS** Collaboration, G. Aad et al., *Measurement of the transverse momentum and ϕ^*_η distributions of Drell–Yan lepton pairs in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C76** (2016), no. 5 291, [arXiv:1512.02192].

[248] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of differential cross sections and W^+/W^- cross-section ratios for W boson production in association with jets at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **05** (2018) 077, [arXiv:1711.03296]. [Erratum: JHEP 10, 048 (2020)].

[249] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the Electron Charge Asymmetry in Inclusive W Production in pp Collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **109** (2012) 111806, [arXiv:1206.2598].

[250] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of Associated W + Charm Production in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2014) 013, [arXiv:1310.1138].

[251] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the Z boson differential cross section in transverse momentum and rapidity in proton–proton collisions at 8 TeV*, *Phys. Lett.* **B749** (2015) 187–209, [arXiv:1504.03511].

[252] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the differential cross section and charge asymmetry for inclusive* $\mathrm{pp} \to \mathrm{W}^{\pm} + X$ *production at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **76** (2016), no. 8 469, [arXiv:1603.01803].

[253] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} =$ 13 TeV*, *Eur. Phys. J. C* **79** (2019), no. 3 269, [arXiv:1811.10021].

[254] **LHCb** Collaboration, R. Aaij et al., *Inclusive W and Z production in the forward region at $\sqrt{s} = 7$ TeV*, *JHEP* **06** (2012) 058, [arXiv:1204.1620].

[255] **LHCb** Collaboration, R. Aaij et al., *Measurement of the forward Z boson production cross-section in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **08** (2015) 039, [arXiv:1505.07024].

[256] **LHCb** Collaboration, R. Aaij et al., *Measurement of the cross-section for $Z \to e^+e^-$ production in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2013) 106, [arXiv:1212.4620].

[257] **LHCb** Collaboration, R. Aaij et al., *Measurement of forward W and Z boson production in pp collisions at* $\sqrt{s} = 8$ *TeV*, *JHEP* **01** (2016) 155, [arXiv:1511.08039].

[258] **ATLAS** Collaboration, G. Aad et al., *Measurement of the high-mass Drell–Yan differential cross-section in pp collisions at sqrt(s)=7 TeV with the ATLAS detector*, *Phys. Lett.* **B725** (2013) 223–242, [arXiv:1305.4192].

[259] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the Differential and Double-Differential Drell-Yan Cross Sections in Proton-Proton Collisions at* $\sqrt{s} = 7$ *TeV*, *JHEP* **12** (2013) 030, [arXiv:1310.7291].

[260] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the differential Drell-Yan cross section in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **12** (2019) 059, [arXiv:1812.10529].

[261] S. Forte, E. Laenen, P. Nason, and J. Rojo, *Heavy quarks in deep-inelastic scattering*, *Nuclear Physics B* **834** (Jul, 2010) 116–162.

[262] V. Bertone, S. Carrazza, and J. Rojo, *Apfel: A pdf evolution library with qed corrections*, *Computer Physics Communications* **185** (Jun, 2014) 1647–1668.

[263] V. Bertone, S. Carrazza, and N. P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, *Comput. Phys. Commun.* **212** (2017) 205–209, [arXiv:1605.02070].

[264] J. Campbell and T. Neumann, *Precision Phenomenology with MCFM*, *JHEP* **12** (2019) 034, [arXiv:1909.09117].

[265] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *The automation of next-to-leading order electroweak calculations*, *JHEP* **07** (2018) 185, [arXiv:1804.10017].

[266] T. Carli, D. Clements, A. Cooper-Sarkar, C. Gwenlan, G. P. Salam, F. Siegert, P. Starovoitov, and M. Sutton, *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, *Eur. Phys. J. C* **66** (2010) 503–524, [arXiv:0911.2985].

[267] M. Grazzini, S. Kallweit, and M. Wiesemann, *Fully differential nnlo computations with matrix*, *The European Physical Journal C* **78** (Jun, 2018).

[268] Y. Li and F. Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, *Phys. Rev. D* **86** (2012) 094034, [arXiv:1208.5967].

[269] C. Duhr, F. Dulat, and B. Mistlberger, *The Drell-Yan cross section to third order in the strong coupling constant*, *Phys. Rev. Lett.* **125** (2020), no. 17 172001, [arXiv:2001.07717].

[270] C. Duhr, F. Dulat, and B. Mistlberger, *Charged current Drell-Yan production at* N^3LO, *JHEP* **11** (2020) 143, [arXiv:2007.13313].

[271] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, and J. Rojo, *Illuminating the photon content of the proton within a global PDF analysis*, *SciPost Phys.* **5** (2018), no. 1 008, [arXiv:1712.07053].

[272] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Fitting Parton Distribution Data with Multiplicative Normalization Uncertainties*, *JHEP* **05** (2010) 075, [arXiv:0912.2276].

[273] R. D. Ball et al., *Parton Distribution Benchmarking with LHC Data*, *JHEP* **04** (2013) 125, [arXiv:1211.5142].

[274] S. Lionetti, R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *Precision determination of α_s using an unbiased global NLO parton set*, *Phys. Lett. B* **701** (2011) 346–352, [arXiv:1103.2369].

[275] R. D. Ball, V. Bertone, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, S. Lionetti, J. Rojo, and M. Ubiali, *Precision NNLO determination of $\alpha_s(M_Z)$ using an unbiased global parton set*, *Phys. Lett. B* **707** (2012) 66–71, [arXiv:1110.2483].

[276] **NNPDF** Collaboration, R. D. Ball, S. Carrazza, L. Del Debbio, S. Forte, Z. Kassabov, J. Rojo, E. Slade, and M. Ubiali, *Precision determination of the strong coupling constant within a global PDF analysis*, *Eur. Phys. J. C* **78** (2018), no. 5 408, [arXiv:1802.03398].

[277] E. Maguire, L. Heinrich, and G. Watt, *HEPData: a repository for high energy physics data*, *J. Phys. Conf. Ser.* **898** (2017), no. 10 102006, [arXiv:1704.05473].

[278] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [arXiv:1405.0301].

[279] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [arXiv:1410.3012].

[280] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [arXiv:1307.6346].

[281] P. Azzi et al., *Report from Working Group 1: Standard Model Physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 1–220, [arXiv:1902.04070].

[282] M. Cepeda et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC*, *CERN Yellow Rep. Monogr.* **7** (2019) 221–584, [arXiv:1902.00134].

[283] **ATLAS** Collaboration, G. Aad et al., *Search for a heavy charged boson in events with a charged lepton and missing transverse momentum from pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, *Phys. Rev. D* **100** (2019), no. 5 052013, [arXiv:1906.05609].

[284] D. King, A. Lenz, and T. Rauh, *B_s mixing observables and $|V_{td}/V_{ts}|$ from sum rules*, *JHEP* **05** (2019) 034, [arXiv:1904.00940].

[285] B. C. Allanach, B. Gripaios, and T. You, *The case for future hadron colliders from $B \to K^{(*)}\mu^+\mu^-$ decays*, *JHEP* **03** (2018) 021, [arXiv:1710.06363].

[286] B. C. Allanach, T. Corbett, M. J. Dolan, and T. You, *Hadron collider sensitivity to fat flavourful Z's for $R_{K^{(*)}}$*, *JHEP* **03** (2019) 137, [arXiv:1810.02166].

[287] **ATLAS** Collaboration, M. Aaboud et al., *Search for scalar leptoquarks in pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS experiment*, *New J. Phys.* **18** (2016), no. 9 093016, [arXiv:1605.06035].

[288] **CMS** Collaboration, A. M. Sirunyan et al., *Search for pair production of second-generation leptoquarks at* $\sqrt{s} = 13$ *TeV*, *Phys. Rev.* **D99** (2019), no. 3 032014, [arXiv:1808.05082].

[289] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **06** (2011) 128, [arXiv:1106.0522].

[290] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J.* **C75** (2015) 132, [arXiv:1412.7420].

[291] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, *JHEP* **01** (2007) 013, [hep-ph/0611129].

[292] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, and C. D. White, *Single-top hadroproduction in association with a W boson*, *JHEP* **07** (2008) 029, [arXiv:0805.3067].

[293] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, *tWH associated production at the LHC*, *Eur. Phys. J.* **C77** (2017), no. 1 34, [arXiv:1607.05862].

[294] **ATLAS Collaboration** Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, Tech. Rep. ATL-PHYS-PUB-2017-005, CERN, Geneva, May, 2017.

[295] M. Mangano, *Physics at the FCC-hh, a 100 TeV pp collider*, *CERN Yellow Rep. Monogr.* **3** (2017) [arXiv:1710.06353].

[296] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[297] N. Cabibbo and R. Gatto, *Electron-positron colliding beam experiments*, *Phys. Rev.* **124** (Dec, 1961) 1577–1595.

[298] J. Blumlein, E. Boos, and A. Kryukov, *Leptoquark pair production in hadronic interactions*, *Z. Phys.* **C76** (1997) 137–153, [hep-ph/9610408].

[299] M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas, *Pair production of scalar leptoquarks at the Tevatron*, *Phys. Rev. Lett.* **79** (1997) 341–344, [hep-ph/9704322].

[300] M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas, *Pair production of scalar leptoquarks at the CERN LHC*, *Phys. Rev.* **D71** (2005) 057503, [hep-ph/0411038].

[301] T. Mandal, S. Mitra, and S. Seth, *Pair Production of Scalar Leptoquarks at the LHC to NLO Parton Shower Accuracy*, *Phys. Rev.* **D93** (2016), no. 3 035018, [arXiv:1506.07369].

[302] I. Doršner and A. Greljo, *Leptoquark toolbox for precision collider studies*, *JHEP* **05** (2018) 126, [arXiv:1801.07641].

[303] T. Plehn, H. Spiesberger, M. Spira, and P. M. Zerwas, *Formation and decay of scalar leptoquarks / squarks in e p collisions*, *Z. Phys.* **C74** (1997) 611–614, [hep-ph/9703433].

[304] **Particle Data Group** Collaboration, M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev.* **D98** (2018), no. 3 030001.

[305] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A434** (1999) 435–443, [hep-ex/9902006].

[306] Lukas, M. Feickert, G. Stark, R. Turra, and J. Forde, *diana-hep/pyhf v0.1.0*, June, 2019.

[307] **ROOT** Collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, .

[308] G. Hiller, D. Loose, and I. Nišandžić, *Flavorful leptoquarks at hadron colliders*, *Phys. Rev.* **D97** (2018), no. 7 075004, [arXiv:1801.09399].

[309] J. Rojo, "https://juanrojocom.files.wordpress.com/2021/07/rojo-smefit-epshep21.pdf."

[310] S. Chen, A. Glioti, G. Panico, and A. Wulzer, *Parametrized classifiers for optimal EFT sensitivity*, *JHEP* **05** (2021) 247, [arXiv:2007.10356].

[311] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933), no. 694-706 289–337.

[312] **ATLAS** Collaboration, M. Aaboud et al., *Searches for scalar leptoquarks and differential cross-section measurements in dilepton-dijet events in proton-proton collisions at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV with the ATLAS experiment*, *Eur. Phys. J. C* **79** (2019), no. 9 733, [arXiv:1902.00377].

[313] **ATLAS** Collaboration, G. Aad et al., *Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Phys. Lett. B* **796** (2019) 68–87, [arXiv:1903.06248].

[314] **ALEPH, DELPHI, L3, OPAL, LEP Electroweak** Collaboration, S. Schael et al., *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, *Phys. Rept.* **532** (2013) 119–244, [arXiv:1302.3415].

[315] **L3** Collaboration, P. Achard et al., *Measurement of the cross section of W-boson pair production at LEP*, *Phys. Lett. B* **600** (2004) 22–40, [hep-ex/0409016].

[316] **OPAL** Collaboration, G. Abbiendi et al., *Measurement of the e+ e- —> W+ W- cross section and W decay branching fractions at LEP*, *Eur. Phys. J. C* **52** (2007) 767–785, [arXiv:0708.1311].

[317] **ALEPH** Collaboration, A. Heister et al., *Measurement of W-pair production in e+ e- collisions at centre-of-mass energies from 183-GeV to 209-GeV*, *Eur. Phys. J. C* **38** (2004) 147–160.

[318] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the W^+W^- production cross section in pp collisions at a centre-of-mass energy of $\sqrt{s}$ = 13 TeV with the ATLAS experiment*, *Phys. Lett. B* **773** (2017) 354–374, [arXiv:1702.04519].

[319] **ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group** Collaboration, S. Schael et al., *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257–454, [hep-ex/0509008].

[320] **CDF, D0** Collaboration, T. A. Aaltonen et al., *Combination of CDF and D0 W-Boson Mass Measurements*, *Phys. Rev. D* **88** (2013), no. 5 052018, [arXiv:1307.7627].

[321] **ATLAS, CMS** Collaboration, G. Aad et al., *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **08** (2016) 045, [arXiv:1606.02266].

[322] **ATLAS** Collaboration, G. Aad et al., *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment*, *Eur. Phys. J. C* **76** (2016), no. 1 6, [arXiv:1507.04548].

[323] **ATLAS** Collaboration, G. Aad et al., *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS experiment*, *Phys. Rev. D* **101** (2020), no. 1 012002, [arXiv:1909.02845].

[324] **CMS** Collaboration, *Measurements of Higgs boson production via gluon fusion and vector boson fusion in the diphoton decay channel at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-HIG-18-029, 2019.

[325] **CMS** Collaboration, *Measurement of Higgs boson production and decay to the $\tau\tau$ final state*, Tech. Rep. CMS-PAS-HIG-18-032, 2019.

[326] **CMS** Collaboration, *Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-HIG-19-001, 2019.

[327] **CMS** Collaboration, *Measurements of differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-HIG-19-002, 2019.

[328] **ATLAS** Collaboration, G. Aad et al., *A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Lett. B* **809** (2020) 135754, [arXiv:2005.05382].

[329] **ATLAS** Collaboration, G. Aad et al., *A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector*, arXiv:2007.07830.

[330] **CDF, D0** Collaboration, T. A. Aaltonen et al., *Combined Forward-Backward Asymmetry Measurements in Top-Antitop Quark Production at the Tevatron*, *Phys. Rev. Lett.* **120** (2018), no. 4 042001, [arXiv:1709.04894].

[331] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of top quark pair differential cross-sections in the dilepton channel in pp collisions at $\sqrt{s}$ = 7 and 8 TeV with ATLAS*, *Phys. Rev. D* **94** (2016), no. 9 092003, [arXiv:1607.07281]. [Addendum: Phys.Rev.D 101, 119901 (2020)].

[332] **ATLAS** Collaboration, G. Aad et al., *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016), no. 10 538, [arXiv:1511.04716].

[333] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **75** (2015), no. 11 542, [arXiv:1505.04480].

[334] **CMS** Collaboration, V. Khachatryan et al., *Measurements of the $\mathrm{t\bar{t}}$ production cross section in lepton+jets final states in pp collisions at 8 TeV and ratio of 8 to 7 TeV cross sections*, *Eur. Phys. J. C* **77** (2017), no. 1 15, [arXiv:1602.09024].

[335] **CMS** Collaboration, V. Khachatryan et al., *Measurements of $t\bar{t}$ charge asymmetry using dilepton final states in pp collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett. B* **760** (2016) 365–386, [arXiv:1603.06221].

[336] **ATLAS** Collaboration, G. Aad et al., *Measurements of the charge asymmetry in top-quark pair production in the dilepton final state at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Rev. D* **94** (2016), no. 3 032006, [arXiv:1604.05538].

[337] **ATLAS, CMS** Collaboration, M. Aaboud et al., *Combination of inclusive and differential $\mathrm{t\bar{t}}$ charge asymmetry measurements using ATLAS and CMS data at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **04** (2018) 033, [arXiv:1709.05327].

[338] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV and impact on parton distribution functions*, *Eur. Phys. J. C* **77** (2017), no. 7 459, [arXiv:1703.01630].

[339] **CMS** Collaboration, S. Chatrchyan et al., *Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **02** (2014) 024, [arXiv:1312.7582]. [Erratum: JHEP 02, 102 (2014)].

[340] **CMS, ATLAS** Collaboration, G. Aad et al., *Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8$ TeV*, *JHEP* **08** (2020), no. 08 051, [arXiv:2005.03799].

[341] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the W boson polarisation in $t\bar{t}$ events from pp collisions at $\sqrt{s} = 8$ TeV in the lepton + jets channel with ATLAS*, *Eur. Phys. J. C* **77** (2017), no. 4 264, [arXiv:1612.02577]. [Erratum: Eur.Phys.J.C 79, 19 (2019)].

[342] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the W boson helicity fractions in the decays of top quark pairs to lepton + jets final states produced in pp collisions at $\sqrt{s} = 8$TeV*, *Phys. Lett. B* **762** (2016) 512–534, [arXiv:1605.09047].

[343] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the $t\bar{t}\gamma$ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *JHEP* **11** (2017) 086, [arXiv:1706.03046].

[344] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the semileptonic $\mathrm{t\bar{t}}$ + γ production cross section in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **10** (2017) 006, [arXiv:1706.08128].

[345] **ATLAS** Collaboration, M. Aaboud et al., *Fiducial, total and differential cross-section measurements of t-channel single top-quark production in pp collisions at 8 TeV using data collected by the ATLAS detector*, *Eur. Phys. J. C* **77** (2017), no. 8 531, [arXiv:1702.02859].

[346] **CMS** Collaboration, V. Khachatryan et al., *Search for s channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV*, *JHEP* **09** (2016) 027, [arXiv:1603.02555].

[347] **CMS** Collaboration, *Single top t-channel differential cross section at 8 TeV*, Tech. Rep. CMS-PAS-TOP-14-004, 2014.

[348] **CMS** Collaboration, V. Khachatryan et al., *Measurement of the t-channel single-top-quark production cross section and of the* $|V_{tb}|$ *CKM matrix element in pp collisions at* $\sqrt{s}$*= 8 TeV*, *JHEP* **06** (2014) 090, [arXiv:1403.7366].

[349] **ATLAS** Collaboration, G. Aad et al., *Evidence for single top-quark production in the s-channel in proton-proton collisions at* $\sqrt{s}$ *=8 TeV with the ATLAS detector using the Matrix Element Method*, *Phys. Lett. B* **756** (2016) 228–246, [arXiv:1511.05980].

[350] **CMS** Collaboration, S. Chatrchyan et al., *Observation of the associated production of a single top quark and a W boson in pp collisions at* $\sqrt{s}$ *=8 TeV*, *Phys. Rev. Lett.* **112** (2014), no. 23 231802, [arXiv:1401.2942].

[351] **ATLAS** Collaboration, G. Aad et al., *Measurement of single top-quark production in association with a W boson in the single-lepton channel at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, arXiv:2007.01554.

[352] **ATLAS** Collaboration, G. Aad et al., *Measurement of the production cross-section of a single top quark in association with a W boson at 8 TeV with the ATLAS experiment*, *JHEP* **01** (2016) 064, [arXiv:1510.03752].

[353] **CMS** Collaboration, A. Sirunyan et al., *Measurement of normalized differential* $\mathrm{t\bar{t}}$ *cross sections in the dilepton channel from pp collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **04** (2018) 060, [arXiv:1708.07638].

[354] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the* $\mathrm{t\bar{t}}$ *production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Eur. Phys. J. C* **79** (2019), no. 5 368, [arXiv:1812.10505].

[355] **ATLAS** Collaboration, *Inclusive and differential measurement of the charge asymmetry in* $t\bar{t}$ *events at 13 TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2019-026, 7, 2019.

[356] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **08** (2018) 011, [arXiv:1711.02547].

[357] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of top quark pair production in association with a Z boson in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **03** (2020) 056, [arXiv:1907.11270].

[358] **ATLAS** Collaboration, G. Aad et al., *Measurements of inclusive and differential cross-sections of combined* $t\bar{t}\gamma$ *and* $tW\gamma$ *production in the* $e\mu$ *channel at 13 TeV with the ATLAS detector*, *JHEP* **09** (2020) 049, [arXiv:2007.06946].

[359] **CMS** Collaboration, A. M. Sirunyan et al., *Cross section measurement of t-channel single top quark production in pp collisions at* $\sqrt{s} = 13$ *TeV*, *Phys. Lett. B* **772** (2017) 752–776, [arXiv:1610.00678].

[360] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *JHEP* **04** (2017) 086, [arXiv:1609.03920].

[361] **CMS** Collaboration, *Measurement of the differential cross section for t-channel single-top-quark production at $\sqrt{s} = 13$ TeV*, Tech. Rep. CMS-PAS-TOP-16-004, 2016.

[362] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS*, *JHEP* **01** (2018) 063, [arXiv:1612.07231].

[363] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **779** (2018) 358–384, [arXiv:1712.02825].

[364] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2018) 117, [arXiv:1805.07399].

[365] **ATLAS** Collaboration, M. Aaboud et al., *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, *Phys. Lett. B* **780** (2018) 557–577, [arXiv:1710.03659].

[366] **CMS** Collaboration, A. M. Sirunyan et al., *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV*, *Phys. Rev. Lett.* **122** (2019), no. 13 132003, [arXiv:1812.05900].

[367] **ATLAS** Collaboration, *Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2020-013, 2020.

[368] **ATLAS** Collaboration, M. Aaboud et al., *Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, *Phys. Rev. D* **99** (2019), no. 5 052009, [arXiv:1811.02305].

[369] **CMS** Collaboration, A. M. Sirunyan et al., *Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **80** (2020), no. 2 75, [arXiv:1908.06463].

[370] **CMS** Collaboration, A. M. Sirunyan et al., *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **11** (2019) 082, [arXiv:1906.02805].

[371] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **803** (2020) 135285, [arXiv:1909.05306].

[372] **CMS** Collaboration, A. M. Sirunyan et al., *Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **07** (2020) 125, [arXiv:2003.06467].

[373] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, *X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue*, *Astron. Astrophys.* **564** (2014) A125, [arXiv:1402.0004].

[374] C. Hays, V. Sanz Gonzalez, and G. Zemaityte, *Constraining EFT parameters using simplified template cross sections*, Tech. Rep. LHCHXSWG-2019-004, 2019.

[375] M. Bonvini, S. Marzani, J. Rojo, L. Rottoli, M. Ubiali, R. D. Ball, V. Bertone, S. Carrazza, and N. P. Hartland, *Parton distributions with threshold resummation*, *JHEP* **09** (2015) 191, [arXiv:1507.01006].

[376] I. Brivio, *SMEFTsim 3.0 – a practical guide*, arXiv:2012.11343.