

NORTHWESTERN UNIVERSITY

PRECISION PHENOMENOLOGY AND NEW PHYSICS PROBES AT  
FUTURE COLLIDERS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Physics

By

KAAN ŞİMŞEK

EVANSTON, ILLINOIS

September 2025

© Copyright by KAAN ŞİMŞEK 2025

All Rights Reserved

# ABSTRACT

## PRECISION PHENOMENOLOGY AND NEW PHYSICS PROBES AT FUTURE COLLIDERS

KAAN ŞİMŞEK

This thesis explores how future colliders can be used to study physics beyond the Standard Model through precision observables. Most of the work is set within the Standard Model Effective Field Theory (SMEFT), but the broader focus is on collider-specific questions: what can realistically be measured, how far current designs can reach, and which theoretical parameters are actually testable in controlled environments. Four main directions are developed. First, we study neutral-current parity-violating observables at the Electron-Ion Collider and show that they can lift degeneracies introduced by Drell-Yan processes at the LHC in the semi-leptonic four-fermion operator subspace of the SMEFT framework. Second, we extend the analysis to include the complete set of SMEFT operators that modify the neutral-current deep inelastic scattering amplitude at leading order, next-to-leading order QCD corrections, and two additional machines, the Large Hadron-electron Collider and the Future Circular Collider. The results demonstrate how these colliders, with their complementary energy ranges, can resolve parameter degeneracies in global

fits and impose tighter bounds on new physics effects. Third, we shift focus to electron-positron collisions and introduce a novel observable at the Future Circular Collider that improves sensitivity to the electron Yukawa coupling, the smallest in the Standard Model, by a factor of five to six. Finally, we return to the SMEFT and examine Drell-Yan production with an associated jet, identifying observables that are sensitive to  $CP$ -violating gluonic operators at dimension eight. We focus on Collins-Soper angular moments that vanish in the Standard Model at leading order but become nonzero in the presence of the SMEFT operators of interest due to their  $CP$ -odd structure. All analyses are tied to realistic experimental conditions, including polarization effects, luminosity estimates, and systematic uncertainties. The methods developed here prioritize reproducibility, computational transparency, and applicability to actual collider programs. Together, these results show how precision and collider phenomenology can be used not only to refine the Standard Model but also to push its limits.



## Acknowledgements

No ocean is sailed alone, and I am particularly terrified of the sea when I can't see the bottom. I would like to take this opportunity to express my sincerest gratitude to the fantastic people who made this work possible.

I want to begin by expressing my deepest gratitude to my advisor, Professor Francis John Petriello. It has been nothing but a privilege to work with Frank. Imagine working in a field that *simply* makes sense to you, and doing so under the guidance of someone who not only leads the field but does so with quiet brilliance and precision. Frank is, without exaggeration, a rock star in this business. I am truly grateful for his leadership, support, and most importantly, for his god-tier patience and for tolerating all of my idiosyncrasies along the way. Throughout these years, I have had the experience of meeting people at conferences who, upon hearing his name, would immediately light up and say something like, "Oh, you're working with Frank! He eats two-loop calculations for breakfast." I still remember one such conversation at the DIS conference in Michigan, which somehow took place while we were jamming at the piano outside the conference hall, a fitting backdrop to the absurdity and charm of this field. Frank's name and reputation opened many doors for me, and I feel that privilege now more than ever. Looking back, I realize how fortunate I have been to work in a field that, as an old mentor of mine once said, is "not the trenches of hep-th," and to do so under the guidance of one of its finest practitioners.

Also, I am indebted to Dr. Radja Boughezal, my supervisor at Argonne National Laboratory. Collaborating with Radja, as well as her postdoctoral research, Dr. Chiara Bissolotti, was not only scientifically rewarding but also personally inspiring. I'm grateful for her guidance and the positive, energetic atmosphere she fostered during my time at Argonne.

In addition, I want to thank Dr. Daniel Wiegand, whose high-precision supervision during my on-boarding period helped me navigate the early stages of our line of work.

I also extend my gratitude to Professors André Luiz de Gouvêa and John Joseph M. Carrasco for their service in my candidacy and dissertation committee.

This research was supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology.

My sincere appreciation also goes out to my former mentor, Professor Takhnasib M. Aliev. He has always been there, offering support and guidance. A fond memory from my Master's days perfectly encapsulates his mentorship style. We were casually chatting on the fourth floor one spring afternoon when he off-handedly mentioned, "They're going to build this new machine, capable of colliding polarized electron beams with hadrons." It was, of course, the Electron-Ion Collider. Years later, while working on my first project here, I vividly recalled that moment and his persistent advice to study phenomenology and not get lost in the trenches of hep-th. That advice, as it turns out, shaped the entire direction of this thesis and continues to shape how I approach physics today.

The less I know about other people, the happier I am. I shared my office with numerous people for half a decade and never learned more than half of their names. Best colleagues I have ever had. We still never talk.

“Life’s good, brother”, said Nazım Hikmet, the greatest poet who ever lived. İ. Anıl Evren, my brother from a different father and a different mother, made sure to have constantly and consistently reminded me of this fact. Despite being separated by the Atlantic Ocean and an entire continent, we never stopped making music together, which *natürlich* helped me preserve my sanity in these trying times. He says “if it were really all about analytical thinking, the world would be a lot more epic place today”.

Lastly, and most importantly, I thank Melih, my brother, my absolute rock in this turbulent journey. He has been my beacon of light, my compass when I have felt completely lost amidst the enormous waves of life’s challenges. In my lowest moments, when it felt like no one else was there, his steady guidance, wisdom, and unconditional support kept me grounded and helped me regain perspective. Among the eight billion people inhabiting this chaotic planet, he is truly second to none.

## Preface

*Science is the truest guide for civilization, for life, for success in the world. Searching for a guide other than science is absurdity, ignorance, and heresy.*

Mustafa Kemal Atatürk

Since I started my program at Northwestern, the Earth has orbited the Sun five times. To the Earth, it is unworthy of notice, merely a microscopic moment in eternity. To me, five years of my life. The universe has no particular interest in what we do. Progress is slow, work is tedious, and meaning is not handed to anyone. This thesis is, at its core, my humble contribution to an otherwise indifferent cosmic harmony.

It is a collection of studies in precision phenomenology at future colliders, mostly framed within the Standard Model Effective Field Theory (SMEFT). Each chapter is the result of months well spent calculating, coding, debugging, analyzing, and occasionally questioning every decision that led me here. It is simply what I managed to produce in collaboration with some of the most brilliant particle physicists.

Chapter 1 introduces the broader context and motivation of this work. Chapter 2 outlines the theoretical framework and methodological tools used throughout the thesis, including collider kinematics, the SMEFT framework, and the computational infrastructure for amplitude generation, numerical routines, and statistical analysis workflows. Chapter

3 presents an analysis of neutral-current deep inelastic scattering (DIS) at the Electron-Ion Collider. Chapter 4 extends this study to higher-energy DIS processes at the Large Hadron-electron Collider and Future Circular Collider, incorporating next-to-leading order quantum chromodynamics corrections and performing multidimensional SMEFT parameter fits. Chapter 5 explores transverse spin asymmetries in electron-positron collisions at the Future Circular Collider as a probe of the electron Yukawa coupling, requiring the development of more advanced amplitude and phase-space tools to extract subtle interference effects. Chapter 6 investigates dimension-8 operators in Drell-Yan plus jet production at the High-Luminosity Large Hadron Collider, with a focus on angular observables sensitive to  $CP$ -odd structures. The backbone of the analysis is complete, yet the work is still in progress because we consider ambitious future directions to pursue, regarding UV matching, as well as the involvement of a spin-2 mediator.

This thesis is a log of work done under normal conditions, namely limited time, limited energy, the persistent noise of unanswered questions, and the stubborn passion that kept me wake at night for all the right reasons. I chose this path for a reason, and there is nothing else I would rather be doing. I enjoy the work—the physics, the coding, the thinking—especially when it gets dark in here.

## **Dedication**

To Melih, my brother, my hero.

## List of abbreviations

BSM: beyond the Standard Model

CEPC: Circular Electron-Positron Collider

CL: confidence level

c.m.: center of mass

CS: Collins-Soper

DIS: deep inelastic scattering

DOE: Department of Energy

DY: Drell-Yan

DY $j$ : Drell-Yan plus jet

EFT: effective field theory

EIC: Electron-Ion Collider

ERL: energy recovery line

EW: electroweak

FCC: Future Circular Collider

HERA: *Hadron-Electron-Ringanlage*

HL-LHC: high-luminosity Large Hadron Collider

LC: lepton-charge

LEP: Large Electron-Positron Collider

LHC: Large Hadron Collider

LHeC: Large Hadron-electron Collider

LIPS: Lorentz-invariant phase space

MWE: minimal working example

NC: neutral-current

NLO: next-to-leading order

PDF: parton distribution function

PV: parity-violating

QCD: quantum chromodynamics

QED: quantum electrodynamics

RHIC: Relativistic Heavy Ion Collider

SLC: SLAC Linear Collider

SLD: SLAC Large Detector

SM: Standard Model

SMEFT: Standard Model Effective Field Theory

U.S.: United States

UV: ultraviolet

VEV: vacuum expectation value



## Table of Contents

ABSTRACT	3
Acknowledgements	5
Preface	8
Dedication	10
List of abbreviations	11
Table of Contents	13
List of Tables	16
List of Figures	19
Chapter 1. Introduction	30
1.1. Position at the time	30
1.2. Future collider facilities for precision physics	34
1.3. SMEFT literature review	49
1.4. Outline of the thesis	56
Chapter 2. The Opening Repertoire	57
2.1. Collider phenomenology	58
2.2. The SMEFT formalism	77

	14
2.3. Particle momenta and many-particle Lorentz-invariant phase space	86
2.4. Statistical analysis	95
2.5. Computational frameworks	103
Chapter 3. EIC Wide Shut	118
3.1. Prelude	119
3.2. Neutral-current DIS physics at the EIC	121
3.3. Projections of PV and LC asymmetry data	133
3.4. Pseudodata generation, the uncertainty matrix, and the fitting	136
3.5. SMEFT fit results	142
3.6. Coda	153
Chapter 4. DISentangling SMEFT: A Few Colliders More	156
4.1. Prelude	156
4.2. Formalism	160
4.3. Observables	176
4.4. Analysis	178
4.5. SMEFT fit results	189
4.6. Coda	202
Chapter 5. Transcendental Étude in $e^-$ Minor	205
5.1. Prelude	206
5.2. Structure of the cross section	209
5.3. Analytical calculations	217
5.4. Dilution of the signal	222

	15
5.5. Sensitivity estimates	225
5.6. Coda	236
Chapter 6. $CP$ -odd to Joy	239
6.1. Prelude	240
6.2. Structure of the cross section	242
6.3. Collins-Soper or: how I learned to stop worrying and love the Frame	247
6.4. SMEFT formalism	256
6.5. Collins-Soper moments	259
6.6. Numerical analysis	260
6.7. HL-LHC simulation	265
6.8. Fits of Wilson coefficients on HL-LHC pseudodata	267
6.9. Coda	277
Chapter 7. Conclusion	286
7.1. Summary of work and findings	286
7.2. Contributions and perspective	289
7.3. Outlook and Future Directions	291
References	293

## List of Tables

3.1	The list of dimension-6 SMEFT operators relevant to DIS in the Warsaw basis before electroweak symmetry breaking and reexpressed in the vector and axial-vector current basis after electroweak symmetry breaking. The $c_{V/A_r}^f$ coefficients represent the chiral structure of each operator.	123
3.2	Expected runs at the EIC, in terms of beam energy, beam type, and nominal annual luminosity, together with our labeling scheme. P6 is the Yellow Report reference setting.	133
3.3	Anticipated uncertainty components for each observable at each family of runs at the EIC. NL means the case of nominal luminosity and HL indicates the case of a 10-fold higher luminosity.	136
4.1	Dimension-6 SMEFT operators in the Warsaw basis [85] that contribute to the NC DIS amplitudes at leading order. The 10 operators that modify $ffV$ vertices are shown on the left, and the 7 operators that introduce semi-leptonic four-fermion contact interactions are presented on the right.	162

4.2	The description of the EIC datasets used in our analysis. We indicate beam energies, polarizations, integrated luminosities, our labeling scheme, and the observable of interest.	179
4.3	The same as in Table 4.2 but for the LHeC.	180
4.4	The same as in Table 4.2 but for the FCC-eh.	180
4.5	Anticipated values or ranges of experimental uncertainties at the EIC for the <i>good</i> bins used in our analysis. All uncertainties are relative with respect to the observable.	182
4.6	The same as in Table 4.5 but for the LHeC.	182
4.7	The same as in Table 4.5 but for the FCC-eh.	183
4.8	Individual and combined 95% confidence level limits on semi-leptonic four-fermion Wilson coefficients at $\Lambda = 1$ TeV, using the EIC data sets P4 and $\Delta$ P4, the combined EIC fit with D4, $\Delta$ D4, P4, and $\Delta$ P4, the separate and combined LHeC runs, and the separate and combined FCC-eh runs. The related effective UV scales are also shown in TeV. “nonmarg.” means nonmarginalized and “marg.” means marginalized.	191
4.9	Marginalized 95% confidence level bounds on Wilson coefficients in the 17-dimensional fit assuming $\Lambda = 1$ TeV, along with the corresponding effective UV scales in TeV. The combined EIC fit of D4, $\Delta$ D4, P4, and $\Delta$ P4, the joint LHeC and FCC-eh fits, and the marginalized bounds and UV scales from the 34-dimensional fits of EW, diboson, Higgs, and top data [92] are shown.	199

- 5.1 The sensitivity estimates for the double-polarization and various single-polarization asymmetries, along with the reference observable, are obtained using the level-0 cut  $\mathcal{C}^0$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization values of  $P_{e-} = 80\%$  and  $P_{e+} = 30\%$  at the collider. 228
- 5.2 The sensitivity estimates for the double-polarization and various single-polarization asymmetries, as well as for the reference observable, are obtained using the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization reaches of  $P_{e-} = 80\%$  and  $P_{e+} = 30\%$  at the collider. 230
- 5.3 Sensitivity estimates for the double-polarization and various single-polarization asymmetries, along with the reference observable, are shown with the corresponding optimal polar-angle cuts. The results are based on the level-2 cuts  $\mathcal{C}^2$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization values of  $P_{e-} = 80\%$  and  $P_{e+} = 30\%$  at the collider. 234
- 6.1 Dimension-8 four-fermion operators with a gluon field that contribute to the Drell-Yan transverse momentum spectrum classified according to their  $CP$  signature. 257
- 6.2 Coarse bins used in our HL-LHC simulations. 268
- 6.3 Fine bins used in our HL-LHC simulations. 269

## List of Figures

- 1.1 An illustrative sketch of the event count corresponding to the total  $e^+e^- \rightarrow \text{hadrons}$  cross section as a function of the center-of-mass energy. Resonances corresponding to known particles, such as the  $\rho$ ,  $J/\psi$ ,  $\Upsilon$ ,  $Z$ , and Higgs bosons, appear as Breit-Wigner peaks due to poles in the scattering amplitude. While these are visible in clean initial states like  $e^+e^-$  collisions, not all particles necessarily appear as bumps. Broad widths, suppressed couplings, or incompatible quantum numbers may obscure resonances, motivating complementary search strategies beyond bump hunting. The dashed curve suggests potential future resonances at higher energy scales, such as those accessible at future colliders. 31
- 1.2 The Electron-Ion Collider (EIC) will be constructed at Brookhaven National Laboratory, utilizing the existing Relativistic Heavy Ion Collider (RHIC) infrastructure. The schematic shows the integration of the new EIC ring with the existing RHIC tunnel.  
*Image credit* Tiffany Bowman/Brookhaven National Laboratory, from <https://www.bnl.gov/newsroom/news.php?a=121805>. 36
- 1.3 Proposed layout of the Large Hadron-electron Collider (LHeC) in the Geneva basin, overlaid on the existing LHC (light blue) and SPS (dark

blue) tunnels [56]. The yellow racetrack indicates the baseline design for the LHeC, with two alternative configurations shown in orange. Also indicated is the location of the high-luminosity LHC (HL-LHC), which would share the LHC ring with the LHeC.

39

- 1.4 A schematic map showing a possible location for the Future Circular Collider (FCC) in the Geneva region, illustrating its relation to the existing LHC ring.

*Image credit* CERN, from <https://home.cern/science/accelerators/future-circular-collider>.

42

- 2.1 Schematic representation of a typical deep inelastic scattering event. An incoming lepton,  $\ell$ , scatters off a parton inside the hadron,  $H$ , via EW gauge-boson exchange, resulting in an outgoing lepton,  $\ell'$ , and an inclusive hadronic final state  $X$ . Throughout this thesis, blue lines indicate leptons, orange lines denote quarks, magenta lines represent EW gauge bosons, green lines represent gluons, and brown lines indicate scalar particles. This color scheme is used consistently throughout, and should be particularly helpful for the synesthetically inclined. Feynman diagrams are produced using Jaxodraw [38].

59

- 2.2 The Feynman diagram for the underlying partonic process for the DIS at leading order.

60

- 2.3 Schematic representation of Drell-Yan production in proton-proton collisions. A quark and an antiquark from the incoming protons



annihilate via an intermediate EW gauge boson,  $V = \gamma, Z, W$ , which decays into a lepton pair. Additional hadronic activity from the proton remnants is denoted by  $X$  and treated inclusively. This is a general schematic; in our analysis, we focus on neutral-current production and include a real jet in the final state.

- 2.4      The Feynman diagrams for the underlying partonic process for the Drell-Yan and jet production at leading order. 66
- 2.5      Representative diagrams for Higgs production (top) and its decay into EW bosons or fermions and gluons, alongside the dominant background processes (bottom). Adapted from Ref. [88]. 73
- 2.6      Leading-order Feynman diagrams for  $e^-e^+ \rightarrow b\bar{b}$  (top) and  $e^-e^+ \rightarrow W^-W^+ \rightarrow \ell\nu jj$  (bottom two). 74
- 2.7      The SMEFT Feynman diagram for the underlying partonic process of the DIS generated by the semi-leptonic four-fermion operators. 79
- 2.8      The SMEFT Feynman diagram for the underlying partonic process of the Drell-Yan production with a jet, generated by the semi-leptonic four-fermion operators coupled to a gluon field strength tensor. 85
- 2.9      The geometry of the 2-body decay in the rest frame of the mother particle. 87
- 2.10     The geometry of the 2-body decay in the center-of-mass frame of the incoming particles. 88

2.11	Comparison of nonmarginalized and marginalized constraints on the parameter pairs $(C_1, C_2)$ , $(C_1, C_3)$ , and $(C_2, C_3)$ .	115
3.1	The Feynman diagrams for $\ell + H \rightarrow \ell + X$ at the parton level from one-boson exchange (left) and SMEFT contact interactions (right).	122
3.2	The kinematic coverage at the EIC in terms of the Bjorken- $x$ variable and the momentum transfer for the lowest and highest collider energies. The shaded region indicates the <i>good</i> region used in our SMEFT analysis.	135
3.3	Error budget plots for representative datasets D4, $\Delta$ P4, and LP5.	138
3.4	The same as Figure 3.3 but for LP5.	139
3.5	95% CL bounds of $C_{\ell q}^{(1)}$ from single-parameters fits (darker) and from the $(1+1)$ -parameter fits with beam polarization as an additional fitting parameter (lighter) using the families of data sets D4, D5, P4, and P5 at $\Lambda = 1$ TeV.	142
3.6	The same as Figure 3.5 but for $C_{\ell q}^{(3)}$ .	143
3.7	The same as Figure 3.5 but for $C_{\ell u}$ .	143
3.8	The same as Figure 3.5 but for $C_{\ell d}$ .	144
3.9	The same as Figure 3.5 but for $C_{eu}$ .	144
3.10	The same as Figure 3.5 but for $C_{ed}$ .	145
3.11	The same as Figure 3.5 but for $C_{qe}$ .	145

- 3.12 Effective UV cut-off scales,  $\Lambda/\sqrt{C_{\ell q}^{(1)}}$ , defined in terms of the 95% CL bounds on the Wilson coefficient  $C_{\ell q}^{(1)}$  and with  $\Lambda = 1$  TeV. 147
- 3.13 The same as Figure 3.12 but for  $C_{\ell q}^{(3)}$ . 147
- 3.14 The same as Figure 3.12 but for  $C_{\ell u}$ . 148
- 3.15 The same as Figure 3.12 but for  $C_{\ell d}$ . 148
- 3.16 The same as Figure 3.12 but for  $C_{eu}$ . 149
- 3.17 The same as Figure 3.12 but for  $C_{ed}$ . 149
- 3.18 The same as Figure 3.12 but for  $C_{qe}$ . 150
- 3.19 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{qe}$  using the families of data sets D4 and P4 in the simultaneous  $(2 + 1)$ -parameter fits that includes the beam polarization as an additional fitting parameter. 151
- 3.20 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{qe}$  using the data sets D4 and P4 in the  $(2 + 1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [48]. 152
- 3.21 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{\ell q}^{(1)}$  using the nominal- and high-luminosity data set P4 in the  $(2 + 1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [47]. 153
- 3.22 95% CL ellipses for the Wilson coefficients  $C_{\ell q}^{(1)}$  and  $C_{\ell q}^{(3)}$  using the nominal-luminosity data set P4 in the  $(2 + 1)$ -parameter fit that includes

- the beam polarization as an additional fitting parameter, compared with the corresponding fit from the LHC data [47] and the combined fit of the two. 154
- 4.1 Feynman diagrams at tree level for the underlying partonic process of the scattering  $\ell + H \rightarrow \ell' + X$ . 163
- 4.2 Feynman diagrams describing NLO QCD corrections to the underlying partonic process of the scattering  $\ell + H \rightarrow \ell' + X$ . 163
- 4.3 NC DIS cross section with NLO QCD corrections for  $e^-p$  collisions at  $\sqrt{s} = 1.3$  TeV with  $E_e = 60$  GeV and  $P_\ell = +80\%$ . 176
- 4.4 Kinematic coverage of the EIC, LHeC, and FCC-eh data sets, indicating the the complementarity of these experiments to each other in terms of the *good* regions considered in our analysis. 181
- 4.5 The different sources of uncertainty that make up the diagonal elements of the uncertainty matrix are shown for the LHeC3 (left) and P4 (right) data sets. The red line shows statistical uncertainty, the blue line shows uncorrelated global efficiency uncertainty, the magenta line shows systematic uncertainty, and the orange line shows PDF uncertainty. For P4, the cyan line shows the uncertainty from beam polarization. 186
- 4.6 The same as in Figure 4.5 but for  $\Delta$ P4 (left) and for FCCeh1 (right). 186
- 4.7 Nonmarginalized (top) and marginalized (bottom) 95% confidence level ellipses for the parameter spaces formed by  $C_{\ell q}^{(1)}$  and  $C_{\ell u}$  (left) and  $C_{\ell q}^{(1)}$

- and  $C_{qe}$  (right) with  $\Lambda = 1$  TeV. The insets display a zoomed-in view of the combined LHeC and FCC-eh fits. 193
- 4.8 Marginalized 95% confidence level ellipses in the parameter spaces formed by  $C_{\ell q}^{(1)}$  and  $C_{\ell q}^{(3)}$  (left) and  $C_{\ell u}$  and  $C_{\ell d}$  (right) at  $\Lambda = 1$  TeV, comparing data sets with luminosities differing by a factor of 10: LHeC2 and LHeC5 (left), and LHeC3 and LHeC6 (right). 194
- 4.9 The same as in Figure 4.8 but for  $C_{\ell u}$  and  $C_{\ell d}$ , with data sets having different lepton species (LHeC3 and LHeC4). 195
- 4.10 Effective UV scales corresponding to marginalized 95% confidence level bounds on the Wilson coefficients  $C_{\ell q}^{(1)}$ ,  $C_{\ell q}^{(3)}$ , and  $C_{qe}$  in the analysis of flat directions for LHeC2, LHeC4, and LHeC5. 197
- 4.11 Correlation matrix of the 17d joint LHeC fit of Wilson coefficients. 200
- 4.12 The same as in Figure 4.11 but for FCC-eh. 201
- 4.13 Marginalized 95% confidence level ellipses in the two-parameter fits of  $C_{\varphi D}$  and  $C_{\varphi e}$  (left) and  $C_{\varphi \ell}^{(1)}$  and  $C_{\varphi e}$  (right) at  $\Lambda = 1$  TeV. The joint EIC, LHeC, and FCC-eh fits are shown, along with the EWPO fit adapted from [92]. 203
- 4.14 Marginalized 95% confidence level ellipses in the two-parameter fits of  $C_{\varphi D}$  and  $C_{\varphi e}$  (left) and  $C_{\varphi \ell}^{(1)}$  and  $C_{\varphi e}$  (right) at  $\Lambda = 1$  TeV. The joint EIC, LHeC, and FCC-eh fits are shown, along with the EWPO fit from [92]. 204

- 5.1 The collision of a transversely-polarized electron beam with a longitudinally-polarized positron beam for the  $b\bar{b}$  process in the center-of-mass frame of the incoming particles. Note that the vectors  $\mathbf{p}_i$  for  $i = 1, 2$  are not restricted to the  $xz$  plane. 212
- 5.2 The collision of a transversely-polarized electron beam with a longitudinally-polarized positron beam for the  $WW$  process in the center-of-mass frame of the incoming particles on the left. On the right, we draw the decays of the  $W$  bosons into a lepton-neutrino pair or two jets in their respective center-of-mass frames. Note that the vectors  $\mathbf{p}_i$  for  $i = 12, 34, 1, 2, 3, 4$  are not restricted to the  $xz$  plane. 213
- 5.3 The dependence of the sensitivity estimate of the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry on the invariant-mass and polar-angle cuts for the  $b\bar{b}$  ( $WW$ ) process is shown in the left (right) panel. The results are obtained using the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization reaches of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider. The green dot marks the highest sensitivity, achieved at  $m_{\text{inv,cut}} = m_h - \Gamma_h$ . 231
- 5.4 The left (right) panel shows how the sensitivity of the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry varies with the invariant-mass cut when the polar-angle cut is fixed at its optimal value,  $\theta_{\text{cut}} = 39\%$  ( $\theta_{12,\text{cut}} = 28\%$ ), for the  $b\bar{b}$  ( $WW$ ) process. The results are based on the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization levels of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider. 232

- 5.5 The left (right) panel shows the contributions of interference channels to the numerator (denominator) of the  $SP^+$  asymmetry in the  $WW$  process. Results are shown for the level-0 and level-1 cuts,  $\mathcal{C}^0$  and  $\mathcal{C}^1$ , represented by dark and light shading, respectively. Channels contributing less than 0.005 in the displayed units are omitted. 233
- 5.6 Sensitivity estimates for the double-polarization and various single-polarization asymmetries, as well as the reference observable, are shown using the specified invariant-mass cuts at all three cut levels. The results assume an integrated luminosity of  $L = 10 \text{ ab}^{-1}$  and beam polarization reaches of  $P_{e-} = 80\%$  and  $P_{e+} = 30\%$  at the collider. The phase-space cuts are  $5^\circ < \theta < 175^\circ$  for the  $b\bar{b}$  channel, and  $E_{j_1, j_2} < 52, 45 \text{ GeV}$ ,  $E_\ell > 10 \text{ GeV}$ ,  $E_{\text{miss}} > 20 \text{ GeV}$ , and  $m_{12} > 12 \text{ GeV}$  for the  $WW$  channels. The corresponding optimal polar-angle cuts are listed in Table 5.3. 235
- 5.7 The left (right) panel shows how the sensitivity of the  $SP^-$  ( $SP^+$ ) asymmetry depends on the center-of-mass energy spread and the integrated luminosity for the  $b\bar{b}$  ( $WW$ ) process. Results are obtained using the level-2 cuts  $\mathcal{C}^2$  and beam polarization values of  $P_{e-} = 80\%$  and  $P_{e+} = 30\%$ . The green dot marks the baseline choice of  $\delta = \Gamma_h$  and  $L = 10 \text{ ab}^{-1}$ . 236
- 6.1 The tree-level Feynman diagrams describing the underlying partonic processes. 243

6.2	SM cross section and SMEFT corrections characterized by Wilson coefficients of interest as functions of dilepton transverse momentum for $170 < m_{\ell\ell} < 350$ GeV, normalized with the SM cross section at the $Z$ -peak.	262
6.3	The same as in Figure 6.2 but for $350 < m_{\ell\ell} < 1000$ GeV.	263
6.4	SMEFT contributions to $A_6$ characterized by Wilson coefficients of interest as functions of dilepton transverse momentum at the dilepton invariant mass bins of $170 < m_{\ell\ell} < 350$ GeV and $350 < m_{\ell\ell} < 1000$ GeV.	265
6.5	The same as in Figure 6.4 but for $A_7$ .	266
6.6	The error budget plot for the observable $A_6$ . The SM value for this observable is zero. The black line is the statistical uncertainty. We also present the SMEFT corrections characterized by Wilson coefficients of interest to compare to the size of the uncertainty.	273
6.7	The same as in Figure 6.6 but for $A_7$ .	274
6.8	95% CL bounds of $CP$ -odd Wilson coefficients from single-parameter fits at $\Lambda = 1$ TeV for the observables $A_6$ and $A_7$ , as well as the combined fit assuming independence, with coarse (top) and fine (bottom) bins.	278
6.9	Effective UV cut-off scales, $\Lambda/C_w^{1/4}$ , corresponding to the nonmarginalized 95% CL bounds given in Figure 6.8.	279
6.10	95% CL bounds of $CP$ -odd Wilson coefficients from multi-parameter fits at $\Lambda = 1$ TeV for the observables $A_6$ and $A_7$ , as well as the combined fit assuming independence, with coarse (top) and fine (bottom) bins.	280



6.11	Effective UV cut-off scales, $\Lambda/C^{1/4}$ , corresponding to the marginalized 95% CL bounds given in Figure 6.10.	281
6.12	Nonmarginalized (top) and marginalized (bottom) 95% CL ellipses for the Wilson coefficients $C_{\ell^2 q^2 g}^{(1)}$ and $C_{\ell^2 u^2 g}$ at the HL-LHC for coarse (left) and fine (right) bins from the fits using the observable $A_6$ and $A_7$ , as well as the joint fit.	282
6.13	The same as in Figure 6.12 but for $C_{\ell^2 u^2 g}$ and $C_{e^2 u^2 g}$ .	283
6.14	Correlation matrix for the fits of $A_6$ with coarse (left) and fine (right) bins.	284
6.15	The same as in Figure 6.14 but for $A_7$ .	284
6.16	The same as in Figure 6.14 but for the joint fit of $A_6$ and $A_7$ .	285

## CHAPTER 1

### Introduction

*It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way—in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.*

Charles Dickens, *A Tale of Two Cities*

#### 1.1. Position at the time

Particle physics stands at a crossroads in the quest to uncover physics beyond the Standard Model (BSM). The Standard Model (SM) has been enormously successful in describing known particles and their interactions, crescendoing with the discovery of the Higgs boson in 2012 [5, 65]. As is the tragic fate of all that possess transcendent beauty, it suffers from certain shortcomings. It does not explain dark matter, neutrino masses, or the baryon asymmetry, and it leaves many theoretical questions open. No new particles have been conclusively observed at the Large Hadron Collider (LHC) beyond the Higgs particle, despite the LHC's impressive energy reach. A canonical depiction of resonant

particle production across collider energies is shown in Figure 1.1, highlighting both the discoveries of the past and the unknown terrain ahead.

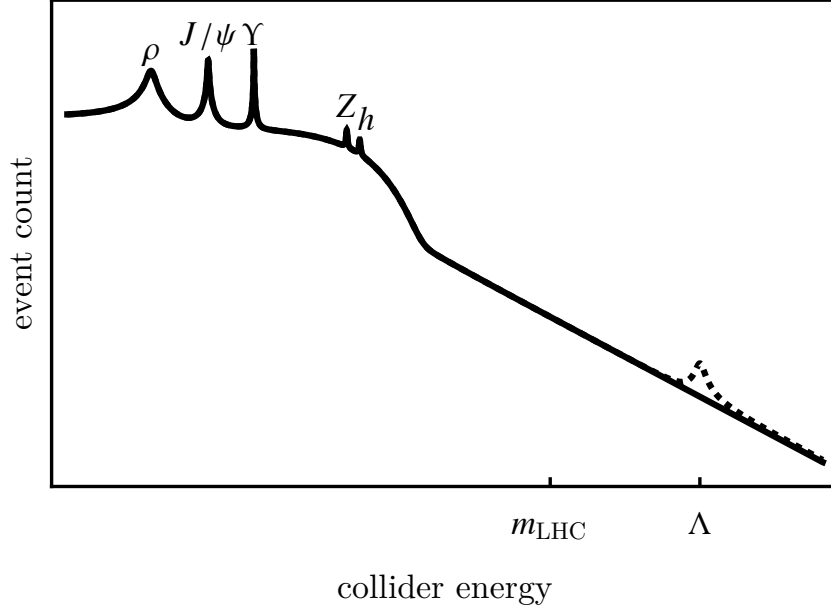


Figure 1.1. An illustrative sketch of the event count corresponding to the total  $e^+e^- \rightarrow \text{hadrons}$  cross section as a function of the center-of-mass energy. Resonances corresponding to known particles, such as the  $\rho$ ,  $J/\psi$ ,  $\Upsilon$ ,  $Z$ , and Higgs bosons, appear as Breit-Wigner peaks due to poles in the scattering amplitude. While these are visible in clean initial states like  $e^+e^-$  collisions, not all particles necessarily appear as bumps. Broad widths, suppressed couplings, or incompatible quantum numbers may obscure resonances, motivating complementary search strategies beyond bump hunting. The dashed curve suggests potential future resonances at higher energy scales, such as those accessible at future colliders.

The absence of clear new signals suggests that any new physics may either be very heavy, and we are talking about energy scales beyond the LHC’s direct reach, or very weakly coupled, which means it must be hiding subtly within precise measurements. In either case, precision phenomenology, namely making extremely accurate predictions and measurements of SM processes, becomes a powerful approach to indirectly probe new physics. By comparing high-precision experimental results with equally precise theoretical predictions, we can detect tiny deviations caused by heavy new particles, even if these particles cannot be produced on-shell. This strategy of indirect discovery through precision measurements has a long and successful history. For example, precision electroweak (EW) data predicted the top quark mass before its discovery and indicated a light Higgs boson. Today, with the energy frontier pushing against practical limits, the precision frontier is more important than ever for exploring BSM effects.

A robust theoretical framework supports this precision program. The Standard Model Effective Field Theory (SMEFT) has emerged as a universal language to describe potential new physics in a model-independent way. In the SMEFT, one upgrades the SM Lagrangian with higher-dimensional operators that might reflect the effects of heavy new particles as subtle shifts in SM interactions. Each operator comes with a Wilson coefficient capturing the strength of the new physics contribution. If new physics lies at an ultraviolet (UV) scale  $\Lambda$  well above observed particle masses, as well as current collider reaches, its low-energy effects can be organized on powers of  $1/\Lambda$ . The leading deviations appear as dimension-6 operators, which are suppressed by  $1/\Lambda^2$ , which can modify cross sections, decay rates, and asymmetry observables at colliders. Crucially, the SMEFT provides a systematic way to globally fit many different measurements for evidence of new

physics within a consistent theoretical framework, rather than invoking new particles. This has motivated a global effort to reanalyze collider data through the SMEFT lens. So far, global fits to precision observables, including LEP/SLD EW measurements and LHC Run 1/2 data, have found no significant deviations, but they have highlighted specific directions in the multidimensional parameter space that are weakly constrained due to overlapping effects of different operators. These flat directions, or degeneracies, in current data leave room for new physics at the TeV scale. Resolving them is a key motivation for both theoretical work and future experiments. Improvements in precision and variety of measurements, especially using new collider facilities, are hoped to tighten these constraints and provide clues of BSM physics. In this context, future colliders are being designed not only to extend the energy frontier, but also to deliver huge data samples and cleaner environments needed for ultraprecise tests of the SM.

Over the next decades, an array of new collider projects are planned or proposed to push the precision frontier. Each comes with unique strengths for exploring different aspects of high-energy physics. In this chapter, we survey the landscape of relevant future colliders and their physics goals, focusing on those facilities that play a role in later chapters of this thesis, namely the Electron-Ion Collider (EIC), the Large Hadron-electron Collider (LHeC), the Future Circular Collider (FCC) running in the electron-hadron mode or the electron-positron mode, and the high-luminosity LHC (HL-LHC) upgrade. We outline the motivations and design parameters of each of these machines, namely their intended physics outcomes, collision types, energies, luminosities, and timelines. We then review how precision measurements at these colliders can be used in SMEFT studies to probe new physics. In doing so, we summarize existing studies in the literature that project

the sensitivity of these experiments to BSM effects via SMEFT fits. This sets the stage for the detailed investigations presented in subsequent chapters. In short, this chapter aims to explain why precision collider phenomenology matters, what future facilities are on the horizon and what features they bring, and how these features and capabilities can be utilized, often in complementary ways, to search for new physics with unprecedented sensitivity.

## **1.2. Future collider facilities for precision physics**

### **1.2.1. Electron-Ion Collider**

The EIC will be a new facility at Brookhaven National Lab (BNL) in the United States (U.S.), designed primarily to explore the quark and gluon structure of nucleons and nuclei with unprecedented detail [154, 12, 10]. It will collide electron beams with beams of protons or heavier ions, such as nuclei of helium, gold, or uranium. A key motivation for the EIC is to create three-dimensional “snapshots” of nucleons, mapping how quarks and gluons carry momentum and spin inside the proton, thereby addressing fundamental questions of quantum chromodynamics (QCD), such as the origin of the proton spin and mass. In 2020, the U.S. Department of Energy approved the EIC project, with the construction beginning at the end of 2025, just when the Relativistic Heavy Ion Collider (RHIC) completes its scientific program, and operations starting in the early 2030s. This will be the first new high-energy collider built in the U.S. in decades, and it represents a cornerstone of the nuclear physics community’s long-term plan [1].

From an accelerator standpoint, the EIC involves a polarized electron beam, potentially 85% polarized at source and with nearly 70% polarization in the storage ring on

average, colliding with polarized proton or ion beams. For protons, a similar polarization reach of about 70% is planned. The electron beam energy will be tunable up to 18 GeV, and proton beam energies up to 275 GeV, allowing for a range of center-of-mass (c.m.) energies from roughly 20 to 140 GeV. These energies sit between fixed-target deep inelastic scattering (DIS) experiments of the past and the *Hadron-Elektron-Ringanlage* (HERA), which reached approximately 320 GeV with 27.6 GeV electrons and 920 GeV protons [127]. The EIC's design luminosity is extremely high for an electron-hadron collider, up to  $10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$  (or about  $300 \text{ fb}^{-1}$  per year), which is orders of magnitude beyond HERA's typical range of  $10^{31}$  to  $10^{32} \text{ cm}^{-2} \cdot \text{s}^{-1}$ . Such luminosity, combined with polarization of both beams, will enable precision EW measurements in addition to its QCD program. The collider will utilize the existing RHIC tunnel, with a circumference of 3.9 km (Figure 1.2), for the hadron ring and add a new electron storage ring of the same size, implementing advanced techniques like energy recovery, spin rotators, and strong focusing to achieve the desired performance.

While the EIC's primary mission is to study QCD phenomenology, such as parton distribution functions (PDFs) at small Bjorken- $x$ , the gluon saturation regime, and the spin structure of nucleons, it also provides an opportunity for BSM searches via precision EW observables. Because it will collide electrons with protons or ions at moderate energies, the EIC is ideal for measuring parity-violating (PV) asymmetries in DIS cross section. In neutral-current (NC) DIS, the photon- $Z$  boson interference causes a small difference in cross section when the electron's helicity is flipped. The EIC can measure this asymmetry, which in the SM is directly related to the running weak mixing angle,  $\sin^2(\theta_W)$ , and to the electron and quark NC couplings. A high-precision measurement of the weak mixing

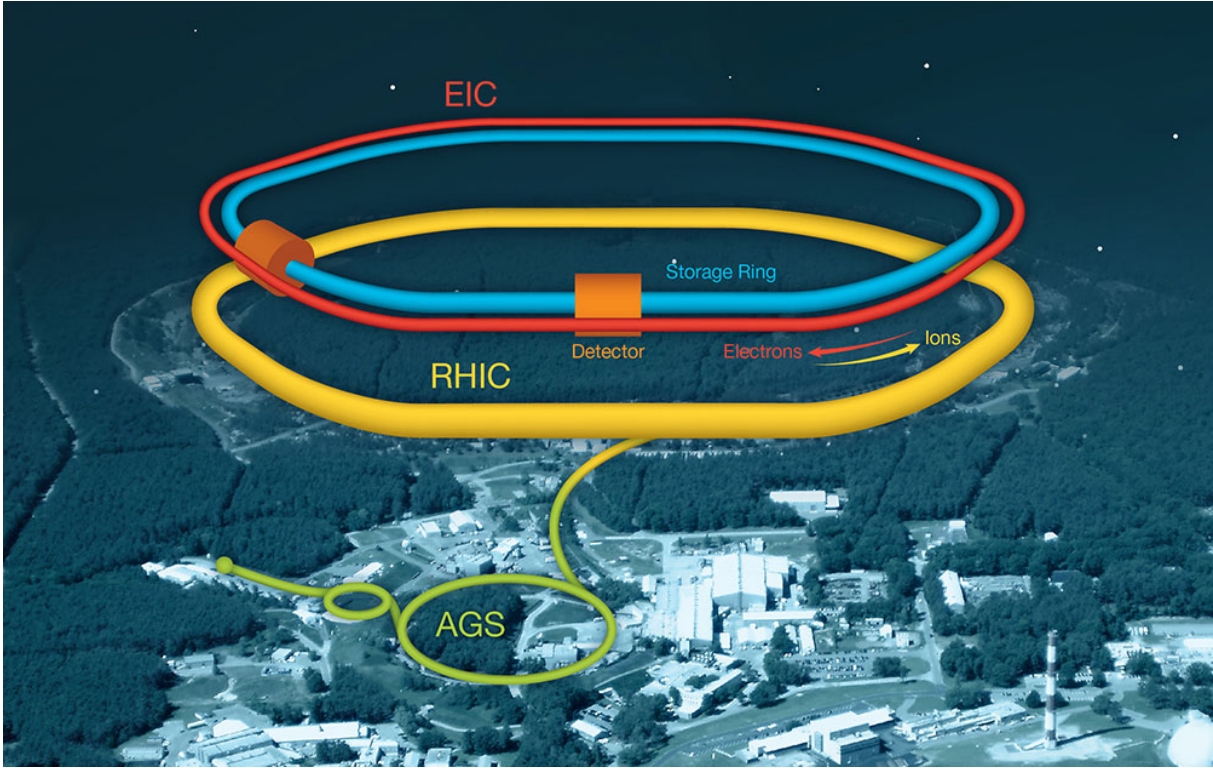


Figure 1.2. The Electron-Ion Collider (EIC) will be constructed at Brookhaven National Laboratory, utilizing the existing Relativistic Heavy Ion Collider (RHIC) infrastructure. The schematic shows the integration of the new EIC ring with the existing RHIC tunnel.

*Image credit* Tiffany Bowman/Brookhaven National Laboratory, from <https://www.bnl.gov/newsroom/news.php?a=121805>.

angle at low momentum transfer is a goal of the EIC's EW program. Moreover, these PV DIS measurements are sensitive to certain SMEFT operators, specifically semi-leptonic four-fermion operators that describe the contact interaction of lepton and quark currents.



The EIC’s true power, namely high luminosity, polarized beams, and a clean event environment with no color contamination and well-defined initial states, allows it to probe these effects ruthlessly. In addition, the EIC will measure electron-nucleus scattering, providing new tests of EW nuclear effects and potentially probing novel physics in nuclear matter. Overall, although the EIC is a machine designed to study hadron structure first and foremost, its high precision makes it a versatile facility where precision SM tests can reveal BSM signals. The knowledge of proton structure gained, for example in the form of improved PDFs, will also be invaluable for interpreting searches at the LHC and future colliders [56].

### 1.2.2. Large Hadron-electron Collider

The LHeC is a proposed upgrade to the LHC that will add a high-energy electron beam to enable electron-proton and electron-ion collisions using one of the LHC’s proton beams [56, 11]. The core idea is to maximize the reuse of existing infrastructure, namely the 27-km LHC tunnel and beams (Figure 1.3), by augmenting it with a new electron accelerator. The LHeC would open up to a TeV-scale DIS program. With a 60-GeV electron beam colliding with a 7-TeV proton beam, the c.m. energy would reach about 1.3 TeV [57], which is substantially higher than HERA’s 320 GeV. In addition, the LHeC aims for a luminosity around  $10^{33} \text{ cm}^{-2} \cdot \text{s}^{-1}$  in its baseline design, which is about 100 times HERA’s. This is expected to lead to enormous DIS data samples. The combined power of the LHeC’s energy and luminosity would push the kinematic region in Bjorken- $x$  and momentum transfer to new extremes, allowing exploration of proton’s structure down to  $x \sim 10^{-6}$  and up to momentum transfers  $Q \sim 10^3 \text{ GeV}$ , an order of magnitude beyond

what LHC measurements currently constrain. From a physics standpoint, the LHeC is motivated both as a precision adjunct to the LHC, providing better PDFs, especially at high  $x$  and high  $Q$ , and novel ways to study the Higgs, top quark, and EW bosons, and as a discovery machine in its own right, able to observe phenomena such as heavy flavor excitation in the proton, rare QCD parton dynamics, or even new particles produced in  $e^-p$  collisions.

The LHeC design has evolved to favor a *linac-ring* configuration, in which a new linear accelerator supplies an electron beam to collide with one of the proton beams circulating in the LHC. The baseline setup features an energy-recovery linac (ERL), which is essentially two superconducting linacs, each about 1 km long, arranged in a racetrack shape tangential to the LHC ring. Electrons would make three passes through each linac to reach 60 GeV before colliding with protons at one of the LHC interaction points [56]. After the collision, the electrons would be decelerated in the same linacs to recover energy. This ERL approach allows for a large number of electrons to be delivered steadily over time, achieving both high beam current and continuous operation without excessive power demands. The same electron accelerator could later be repurposed for other projects such as the Future Circular Collider running in the electron-hadron mode, as discussed below. The electron beam could be polarized, up to nearly 80% at source, although the baseline LHeC design does not require polarization for its core program. The 7-TeV protons in the LHC are not polarized, so LHeC would mainly measure unpolarized  $e^-p$  and rely on electron polarization for any asymmetry measurements. One nice feature of the linac-ring scheme is that it could operate simultaneously with the nominal LHC  $pp$  program by using one of the LHC's beam interaction points without disrupting the others. In

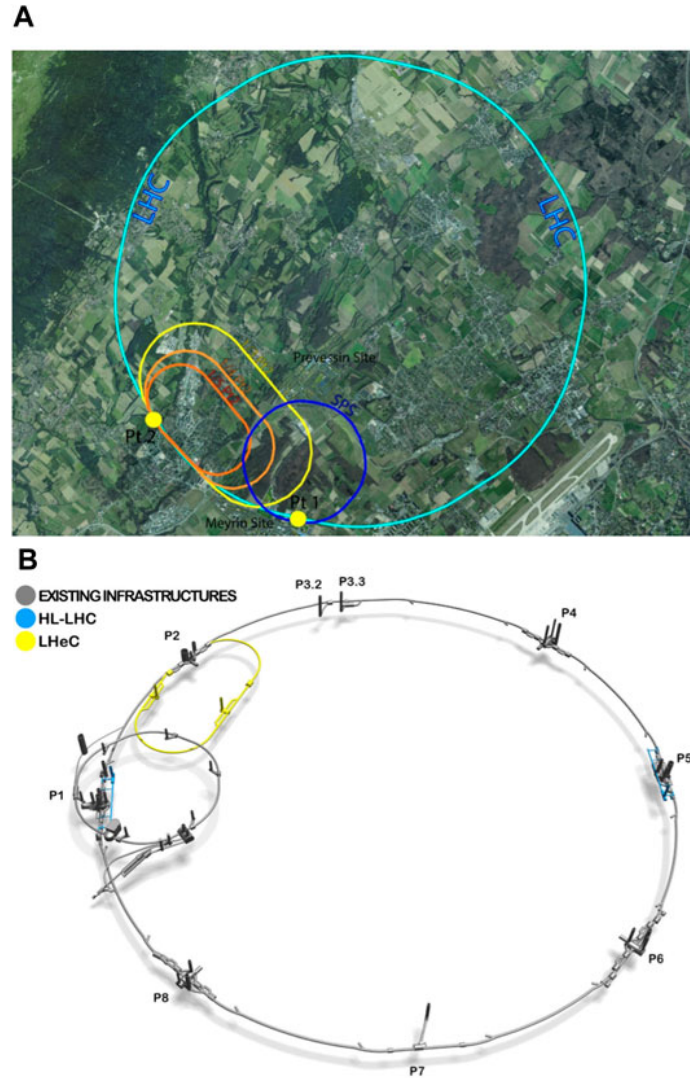


Figure 1.3. Proposed layout of the Large Hadron-electron Collider (LHeC) in the Geneva basin, overlaid on the existing LHC (light blue) and SPS (dark blue) tunnels [56]. The yellow racetrack indicates the baseline design for the LHeC, with two alternative configurations shown in orange. Also indicated is the location of the high-luminosity LHC (HL-LHC), which would share the LHC ring with the LHeC.

principle, the LHeC could quietly slip into the HL-LHC's schedule, offering  $e^-p$  collisions without interrupting  $pp$  operations. After the HL-LHC program is complete, the LHeC could even continue as a dedicated facility on its own. This staging flexibility is a selling point of the project.

From a physics perspective, the LHeC would serve multiple roles. First, it would dramatically improve the knowledge of proton structures. With its high reach in momentum transfer and wide  $x$  range, it would pin down PDFs, especially the gluon and quark distributions at small  $x$ , with high precision. This has direct benefits for interpreting LHC results, for instance, reducing PDF uncertainties in precision measurements of the  $W$  mass and Higgs production. Second, the LHeC offers a unique environment to study the Higgs boson. Higgs production in  $ep$  primarily occurs via vector-boson fusion, e.g.  $W$  exchange from the electron scattering on the proton, and can produce a Higgs boson plus a final-state neutron or proton leftovers. Studies indicate that the LHeC could accumulate on the order of  $10^4$  to  $10^5$  Higgs events, enabling measurements of the  $h \rightarrow b\bar{b}$  decay in a cleaner environment than  $pp$  due to the absence of huge QCD multijet background. In fact, with sufficient luminosity ( $> 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ ), the LHeC could act as a Higgs factory in its own right, measuring couplings like the  $b$  quark Yukawa with percent-level precision complementary to LHC or future  $e^-e^+$  colliders. Third, the LHeC can probe the top quark via single top production, e.g.  $e^-p \rightarrow \nu_e t \bar{b}$ , and even top pair production at the highest energies. This provides an independent way of measuring the  $V_{tb}$  element of the Cabibbo-Kobayashi-Maskawa matrix and the top EW couplings. Finally, and most importantly for this thesis, the LHeC would enable new physics searches through precision EW measurements. By studying processes like NC DIS at high momentum transfer,

the LHeC would be sensitive to contact interactions or exchange of new heavy particles, such as leptoquarks or  $Z'$  bosons, that would manifest as deviations from SM predictions. Semi-leptonic four-fermion SMEFT operators would also yield distinctive effects in  $e^-p$  scattering observables. Thanks to its high energy, the LHeC could improve upon existing bounds from the LEP and even the LHC in many cases. For example, it could cleanly measure  $Z$  boson couplings to quarks by observing the interference pattern of photon and  $Z$  exchange in DIS, which helps break degeneracies that are difficult to resolve with  $Z$ -pole data alone. In conclusion, although the LHeC has not yet been approved and faces technical challenges, its potential impact on both SM measurements and indirect new physics searches is extensive. It stands as a natural next step in the evolution of DIS experiments, bringing the electron-proton probe to the energy scale of the TeV era.

### 1.2.3. Future Circular Collider in electron-hadron mode

The Future Circular Collider in the electron-hadron mode (FCC-eh) refers to an electron-hadron collider utilizing the planned Future Circular Collider infrastructure. In the long-term plan at CERN, the FCC would begin with an  $e^-e^+$  collider (FCC-ee) and later be followed by a 100-TeV proton-proton collider (FCC-hh) in the same 91-km tunnel [56, 2] (Figure 1.4). Once the 100-TeV hadron machine is running, one can imagine adding an electron beam similar to the LHeC's to collide with the 50-TeV proton beam of the Future Circular Collider in the hadron-hadron mode (FCC-hh). This is the FCC-eh concept, essentially the big brother of the LHeC, which would extend  $e^-p$  collisions to unprecedented energies. With a 60-GeV electron beam on a 50-TeV proton beam, the c.m. energy would reach about 3.5 TeV. This is almost three times higher than the LHeC and

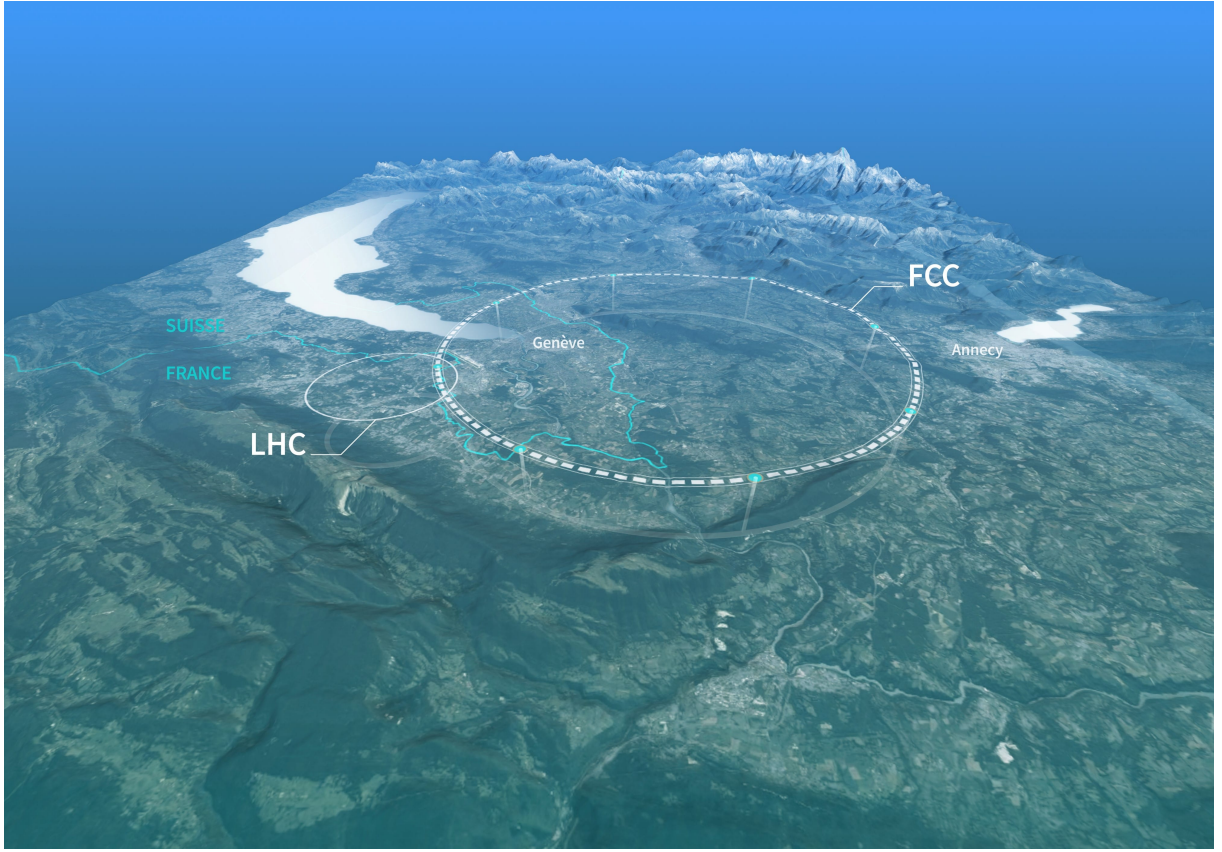


Figure 1.4. A schematic map showing a possible location for the Future Circular Collider (FCC) in the Geneva region, illustrating its relation to the existing LHC ring.

*Image credit* CERN, from <https://home.cern/science/accelerators/future-circular-collider>.

an order of magnitude beyond HERA. The kinematic reach in DIS would correspondingly extend to probes of quark structure with a resolution of  $10^{-5}$  fm. Such collisions could explore extremely small  $x$  values down to  $10^{-7}$  or lower, and momentum transfers near  $10^4$  GeV.

The design of FCC-eh would likely deploy the same ERL used for the LHeC. Indeed, one proposal is a staged approach: Build the LHeC ERL during LHC operations, use it as a full-energy injector for FCC-ee in the 2040s so as to avoid requiring a 100-km booster ring, and then configure it as the  $e^-p$  collider for FCC-hh in the 2070s. In this scenario, the hardware from the LHeC basically gets recycled, namely the racetrack linacs would first help FCC-ee and later serve FCC-eh. The conceptual similarity means that FCC-eh would inherit the same general features as LHeC, namely a polarized electron beam (80% or higher) in an ERL, multipass acceleration, and the possibility of concurrent running with the  $pp$  program, so FCC-hh and FCC-eh data could be collected simultaneously. The luminosity goal for FCC-eh is of order  $10^{33}$  to  $10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ . Achieving high luminosity with a 50-TeV proton beam would require significant technical innovation but studies indicate that a few  $10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$  is feasible in a dedicated  $e^-p$  mode.

In terms of physics reach, FCC-eh would combine the virtues of LHeC, namely high precision and a clean  $e^-p$  environment, with an extended energy range that even exceed the HL-LHC's Drell-Yan (DY) reach for certain interactions. It would further sharpen our picture of the proton's inner structure. When FCC-hh starts running, proton PDFs will be a limiting systematic for many measurements and FCC-eh data would be crucial to push those uncertainties well below the percent level. For Higgs physics, FCC-eh could produce Higgs bosons via vector-boson fusion at several times the rate of LHeC, enabling precision measurements of Higgs properties complementary to FCC-ee and FCC-hh. For BSM searches, the FCC-eh could indirectly sense new physics scales of a few 10 TeV. The FCC-eh would be particularly great at probing any lepton-quark contact interactions or new resonance in the lepton-quark channel, since it extends the search for effects in

$e^-q \rightarrow e^-q$  scattering well beyond the LHC’s kinematic range. It is important to note that FCC-eh is a very forward-looking project. In the current timeline, it would likely operate in the latter half of the 21st century, possibly after 2070, during FCC-hh’s run. By then, the chance of seeing the first data may depend less on luminosity and more on cardio, clean lungs, dietary discipline, good genetics, a bit of luck, and will to live. Nonetheless, the electron-hadron mode features in strategic plans as part of the complete physics program of the FCC.

#### 1.2.4. Future Circular Collider in $e^-e^+$ mode

The Future Circular Collider running in the electron-positron mode (FCC-ee) is a proposed electron-positron collider that would serve as the first phase of CERN’s Future Circular Collider project. It is conceived as a Higgs factory and a precision EW machine, exploiting the relatively low mass of the Higgs, which makes an  $e^-e^+$  collider viable at those energies. The FCC-ee would be placed in a new circular tunnel about 91 km in circumference, and it would operate at several c.m. energy values spanning the  $Z$ -boson peak up to the  $t\bar{t}$  threshold [116]. Specifically, the baseline includes runs at 91.2 GeV (the  $Z$  pole), to produce an enormous sample of  $Z$  bosons, which is often called the Tera- $Z$  run for aiming at more than  $10^{12}$   $Z$  decays, 160 GeV (the  $WW$  threshold) to measure the  $W$  mass and width precisely with threshold scans and to collect  $10^8$   $W$  pairs, 240 to 250 GeV (around the Higgsstrahlung maximum) to produce Higgs bosons in association with  $Z$  bosons, maximizing the rate of Higgs production, and 340 to 365 GeV (around the  $t\bar{t}$  threshold) to study top quark pairs and measure the top mass and its EW couplings. In practice, the machine could run at intermediate to extended energies up to around 365



GeV; however, the four aforementioned ones are in focus. Due to the virtue of a circular collider, having many bunches and continuous collisions, the luminosity goal is extremely high, at a level  $0.5$  to  $1 \times 10^{36} \text{ cm}^{-2} \cdot \text{s}^{-1}$  at the  $Z$  pole. We expect roughly  $5 \times 10^{12}$   $Z$  bosons,  $10^8$   $WW$  pairs,  $10^6$  Higgs bosons, and about  $10^6$  top quark pairs to be collected in total. These numbers are many orders of magnitude beyond what was achieved at LEP, and even beyond what the HL-LHC will produce for some of these particles. To illustrate,  $10^6$  Higgs events is about 30 times the Higgs yield of the HL-LHC (and even in a much cleaner environment), and  $10^{12}$   $Z$  bosons is six orders of magnitude beyond LEP's  $Z$  samples. This statistical power alone renders FCC-ee so appealing because we are talking about statistical uncertainties at the per-mille to  $10^{-5}$  level in some cases.

While the FCC-ee offers an exceptionally clean collision environment ideal for precision measurements, achieving longitudinal polarization of the incident beams remains a challenge. Unlike linear colliders, the FCC-ee cannot sustain longitudinally polarized beams in collisions due to depolarizing effects like beamstrahlung. Instead, polarization is primarily used at lower energies for precise energy calibration, particularly near the  $Z$  pole, via techniques developed at LEP. However, for most runs, the beams are effectively unpolarized, limiting direct access to observables such as left-right asymmetries. Nonetheless, indirect probes remain viable through final-state polarization and angular distributions across various fermionic channels.

The physics program of FCC-ee is rich and directly targets precision tests of the SM at the loop level, making it a potent tool to search for virtual effects of new physics. At the  $Z$  pole, by analyzing the shapes of angular distributions and asymmetries of  $e^+e^- \rightarrow f\bar{f}$  for a fermion  $f$ , FCC-ee will measure the weak mixing angle  $\sin^2(\theta_W)$  to  $10^{-5}$  precision, the  $Z$

boson width and couplings to quarks and leptons with improvements of one to two orders of magnitude over LEP. These measurements constrain EW SMEFT operators, such as ones modifying  $Z$  couplings, at the per-mille level, corresponding to probing new physics up to multi-TeV scales. Running at the  $W$  threshold, FCC-ee can determine the  $W$  mass to an accuracy within 1 to 2 MeV, an order of magnitude better than today's uncertainty, by analyzing the threshold cross section shape. At 240 GeV, the Higgs factory run will nail down the Higgs boson's couplings to gauge bosons and fermions with sub-percent uncertainties in many cases. For example, the  $hZZ$  coupling strength can be measured to 0.2%,  $hWW$  to 0.5%, the Higgs total width to a few percent, and rare decays like  $h \rightarrow \mu^- \mu^+$  observed with meaningful statistics. These Higgs measurements are of great interest for SMEFT since any deviation from SM coupling values signals new physics. One example is the  $\kappa$ -framework or a global effective field theory (EFT) fit, which allows coupling shifts to be translated into bounds on higher-dimensional operators. Finally, at near 350 GeV, FCC-ee will study the top quark with great precision. By performing a threshold scan of  $e^- e^+ \rightarrow t \bar{t}$ , it can extract the top quark mass to about 10 MeV and measure the top's NC couplings much more cleanly than hadron colliders can. All these high-precision observables provide indirect sensitivity to heavy new physics. In fact, it's estimated that FCC-ee data could detect the effects of new particles up to masses of order  $\Lambda \sim 10$  TeV or even higher, depending on the couplings [108]. For instance, deviations in Higgs couplings at the 0.1% level would point to BSM states in the 1 to 10 TeV range in many scenarios, and FCC-ee is exactly aimed at that level of precision. In summary, FCC-ee would set a new benchmark for precision tests of the SM. If new physics is hiding just beyond current reach, the FCC-ee's ultra-high statistics and clean environment offer one

of the best bets to find its footprints. Its operation, which is planned for about 15 years starting in the late 2040s if approved, would also lay the groundwork for its successor, namely the 100-TeV FCC-hh, by providing essential measurements and perhaps clues for what that higher-energy machine should target.

### 1.2.5. High-luminosity upgrade of the Large Hadron Collider

The HL-LHC is the nearest-term major project and will be an integral part of the precision frontier. The HL-LHC is not a new collider but an upgrade of the existing LHC at CERN to significantly increase its luminosity. After around 2025, the LHC will undergo its Long Shutdown 3, during which new high-field focusing magnets, crab cavities, and other upgrades will be installed. The goal is to reach a peak luminosity about 5 to  $7 \times 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ , which is roughly an order of magnitude beyond the LHC's initial design of  $1 \times 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ . By operating through the 2030s, the HL-LHC aims to deliver an integrated luminosity of about 3 to 4  $\text{ab}^{-1}$  to each of the major experiments [3, 129]. This is a factor of 10 more data than the entire first 14 years, namely Runs 1 to 3, of the LHC running will ever have produced; for comparison, by the end of Run 3 in 2025, the LHC expects 300 to 400  $\text{fb}^{-1}$  per experiment. We expect HL-LHC collisions to begin around 2029 and extending to 2040 [4].

The HL-LHC's primary physics motivation is to significantly improve measurements of processes that were either discovery channels or loosely measured in the first LHC runs. For example, with 3  $\text{ab}^{-1}$  of data, the properties of the Higgs boson can be determined much more precisely. The ATLAS and CMS experiments project uncertainties of order 2 to 5% on most Higgs couplings, to  $W$ ,  $Z$ ,  $b$ , and  $\tau$  to name a few, and around 10%

for the top coupling and the Higgs self-coupling. These precision levels, though still an order of magnitude above FCC-ee’s targets, will test the SM Higgs sector at the loop level. In the top quark sector, rare processes like  $h t \bar{t}$  and  $Z t \bar{t}$  will be accessible in large numbers, allowing detailed tests of the Yukawa and EW couplings of the top quark. EW gauge-boson processes, such as diboson production and vector-boson scattering, with high statistics will probe the self-interactions of  $W$  and  $Z$  bosons, looking for deviations that signal the beginning of new dynamics, such as the effects of a heavy resonance or a nonlinear Higgs sector. The HL-LHC will also push further the direct search reach for new particles, as well. Because cross sections fall steeply with mass, a tenfold increase in luminosity typically improves the sensitivity to heavy resonances by up to 30% in mass reach. Equally importantly, it enables searches for extremely rare processes, like flavor-changing NC decays of top quarks or boosted Higgs decays into invisible particles, which may have tiny branching fractions if induced by new physics.

One critical aspect of the HL-LHC is its challenging experimental environment. Higher luminosity comes at the cost of much higher pile-up. On average, about 140 to 200 overlapping  $pp$  interactions will occur every beam crossing at design luminosity. This means each recorded event is buried in debris from dozens of other proton collisions happening simultaneously. The detectors are being upgraded with new trackers, high-granularity calorimeters, and fast timing layers to cope with this by improving vertex separation and timing to distinguish pile-up tracks. From a theoretical standpoint, predictions must also be improved both in accuracy by means of higher-order QCD or EW calculations, as well as in incorporating effects like multi-parton interactions, to match the reduced

experimental uncertainties. The HL-LHC data will demand percent-level theoretical predictions for many key processes, which in turn justifies the use of frameworks like SMEFT to consistently include potential new physics effects in those predictions.

In summary, the HL-LHC will extend the LHC’s precision reach in measuring the familiar observables within the SM and searching for deviations. By the end of its run, if the SM still holds with no significant deviations, we will have substantially tightened the limits on many possible BSM scenarios. For instance, fits to the Higgs and EW data from HL-LHC are expected to constrain certain dimension-6 SMEFT operator scales to several TeV. However, there will remain blind spots and flat directions that only different collision systems, like  $e^-e^+$  or  $e^-p$  colliders, can address, due to the HL-LHC’s inherent hadronic initial state and limited energy for some observables. This is why the previously mentioned future colliders are seen as complementary. The HL-LHC sets the stage in the 2020s and 30s with huge data on the Higgs, top, and EW bosons and its results will inform and tighten the designs of those future machines.

### 1.3. SMEFT literature review

The SMEFT extends the SM Lagrangian with higher-dimensional operators built using the existing SM fields without introducing any new particles, suppressed by powers of a new-physics scale,  $\Lambda$ , which is assumed to be heavier than all SM particles and beyond accessible collider energy. Coupling strength of these new interactions are referred to as Wilson coefficients. In this section, we present an overview of the foundations and the current state of the SMEFT.

### 1.3.1. Early effective field theory foundations

The idea of parametrizing new heavy physics through an EFT dates back decades, having its foundation in the principle that heavy degrees of freedom can be integrated out of the low-energy theory [71, 26, 152]. The first systematic enumeration of higher-dimensional operators was given by Buchmüller and Wyler, who catalogued all dimension-6 terms allowed by SM gauge symmetries [58]. This pioneering list contained redundancies due to relations like equations of motion. Subsequent work clarified these issues and established more convenient operator bases [27]. Prior to the LHC era, EFT techniques were already employed to describe potential new physics in precision measurements. For instance, four-fermion contact interactions and anomalous couplings were constrained in LEP, HERA, and Tevatron data using effective operator frameworks [42, 41, 61]. These earlier works laid the groundwork for the modern SMEFT program.

### 1.3.2. SMEFT operator bases and theoretical advances

With the advent of the LHC, the SMEFT has been developed into a systematic framework. A complete, nonredundant operator basis up to dimension-6, the so-called Warsaw basis, was established in 2010 [110], building upon the earlier classifications. Alternative formulations or basis choices, for example the Higgs-centric “Higgs basis” and the Strongly Interacting Light Higgs basis, have also been proposed, which are physically equivalent to the Warsaw basis under proper field redefinitions [143, 99]. In recent years, the operator catalog has been extended to even higher orders. Independent groups have constructed the full set of dimension-8 SMEFT operators [140, 134], and efforts have progressed toward enumerating operators at dimension 12 and beyond [113]. Systematic

methods, such as Hilbert-series techniques, have been introduced to ensure all invariants are accounted for at a given order [117, 118, 104, 105]. Together, these works provide a structured foundation for the SMEFT as a general extension of the Standard Model.

Alongside the operator bases, significant work has gone into developing the theoretical toolkit for SMEFT. Complete Feynman rules for the SMEFT have been derived and implemented in public codes [85, 86, 87, 54, 52], enabling automated calculations of processes with dimension-6 effects. The consistency of the SMEFT framework has also been investigated extensively. This includes studies of renormalization and operator mixing, as well as constraints on the EFT parameter space coming from fundamental principles. For example, theoretical bounds such as positivity constraints have been derived, which restrict certain Wilson coefficients based on the requirement of a unitary, causal UV completion [31, 156, 84]. These constraints provide an important complementary guide to the allowed SMEFT parameter space, beyond direct experimental limits. For a comprehensive review of the SMEFT formalism and its theoretical underpinnings, see [55].

### 1.3.3. Phenomenology with SMEFT

The SMEFT framework has been extensively used to interpret and constrain new physics from experimental data. In the early applications, which took place around the LEP and Tevatron era and the start of the LHC program, studies typically focused on one sector or a limited set of processes at a time. For example, four-fermion interactions and EW precision observables were studied with dimension-6 operators to set bounds on contact interactions and anomalous couplings [112, 68]. Similarly, Higgs and gauge-boson processes have been analyzed in an EFT context. The effects of higher-dimensional

operators on Higgs boson decays, EW boson pair production, and triple gauge couplings were examined by many groups [66, 94, 153, 115]. Low-energy measurements and flavor physics provided additional complementary constraints on SMEFT coefficients in semi-leptonic and meson decay processes [67, 100]. These targeted studies yielded the first bounds on various Wilson coefficients, often under simplifying assumptions, such as one operator switched on at a time, and demonstrated the feasibility of probing heavy new physics indirectly.

As more data accumulated, especially after the Higgs discovery, the emphasis shifted to global fits that combine information from multiple channels. By the end of LHC Run 1, the first global SMEFT analyses appeared, incorporating Higgs, EW precision, and top-quark measurements into a simultaneous fit [79, 114]. This effort intensified with Run 2. Several comprehensive fits were performed, progressively including tens of independent Wilson coefficients in the analysis [37, 109, 29]. Such studies account for correlations between observables and provide a consistent overall interpretation of new physics scales. In recent years, global fits have reached even greater scope and sophistication [47, 44, 97, 62]. Some analyses incorporate heavy flavor and low-energy data under combined frameworks or impose flavor symmetry assumptions to manage the large operator set [93, 77]. Others have integrated PDF uncertainties into the EFT fit, reflecting the interplay between new physics and proton structure. These up-to-date fits constrain many dimension-6 operator coefficients at the percent level or better, corresponding to new physics scales on the order of multi-TeV, and represent a milestone in the SMEFT program’s maturity.

Recently, attention has also turned to effects beyond the leading dimension-6 order in the SMEFT. In particular, studies have begun exploring the phenomenological impact



of dimension-8 operators and higher-order  $1/\Lambda^4$  contributions. Several works investigate how dimension-8 terms, or equivalently, quadratic combinations of dimension-6 terms, could affect LHC observables such as diboson productions, DY distributions, and Higgs processes [15, 46, 70, 135, 75]. These analyses aim to discern whether subtle deviations could arise from neglected higher-dimensional effects, and to what extent current and future experiments might be sensitive to them. While so far no significant deviations attributable to dimension-8 operators have been detected, the inclusion of such terms is crucial for consistent interpretation as experimental precision increases. The SMEFT approach has now been adopted by the experimental community as well. To illustrate, the CMS collaboration has released results of fits interpreted in the SMEFT framework [149].

#### 1.3.4. Projections for future colliders

Boughezal, Petriello, and Wiegand showed that the planned EIC can play an important role in constraining SMEFT operators. In [47], they studied NC DIS at the EIC with polarized beams. They found that certain four-fermion operator combinations are accessible at the EIC but practically invisible in LHC Drell-Yan measurements. Thanks to polarization asymmetries and the clean leptonic initial state, the EIC can break degeneracies among SMEFT Wilson coefficients that persist even after LEP and LHC data are combined. Indeed, their fit projections showed the EIC probing new contact interactions at scales comparable to or beyond the reach of high-invariant-mass LHC dilepton searches, without suffering flat directions in parameter space. This demonstrates that precision

DIS measurements, even at relatively low energy scales of just a few TeV, complement high-energy colliders in constraining EFT parameters.

A Snowmass 2021 whitepaper by de Blas *et al.* (2022) performed global fits for future collider scenarios [80]. They considered several run plans: combinations of future  $e^+e^-$  Higgs factories (FCC-ee and the proposed Circular Electron-Positron Collider (CEPC) in China), a high-energy muon collider, and the HL-LHC. For each scenario, they included the relevant projected measurements and fit the expected precision on Wilson coefficients. For example, their Higgs+EW fit (updating the European Strategy study) showed percent-level or better measurements of Higgs couplings would dramatically tighten the bounds on certain coefficients. They also did a dedicated fit to four-fermion operators, which is challenging at the LHC but would be probed cleanly by lepton colliders running at the  $Z$  pole and above. Separately, they fit the top-quark sector using projected HL-LHC and future lepton collider top data. The outcome was a broad view that future colliders have complementary strengths: FCC-ee/CEPC excels in high-precision EW and Higgs measurements, Muon Collider (with multi-TeV energy) could probe contact interactions far beyond the LHC reach, and HL-LHC still plays a role in QCD-dominated processes. In all cases, the fits were reported in terms of Wilson coefficient sensitivities or equivalent coupling deviations, enabling comparison between scenarios.

A particularly comprehensive projection is the SMEFiT3.0 study by Celada *et al.* [63]. They updated the global data fit to include the latest Run 2 results for Higgs, top, and diboson channels, then extrapolated to the HL-LHC at 14 TeV with  $3 \text{ ab}^{-1}$  by scaling uncertainties and central values appropriately. The addition of HL-LHC data is expected

to significantly strengthen constraints on many coefficients, especially those affecting high- $p_T$  tails. Furthermore, Celada *et al.* incorporated detailed FCC-ee projections at multiple energies: the  $Z$ -pole, the  $WW$  threshold, the *Higgs factory* (240 GeV), and even  $t\bar{t}$  threshold and 365 GeV runs. Using optimal observable techniques for some channels, they showed that FCC-ee data would push SMEFT sensitivity to the  $10^{-3}$  level in many cases, effectively eliminating vast areas of parameter space or probing multi-TeV scales for new physics. For example, FCC-ee's precise measurements of the  $Z$  lineshape, asymmetries, and  $W$  mass can improve constraints on four-fermion and EW operators by one to two orders of magnitude beyond the capabilities of the HL-LHC. The study also examined how these improvements could enhance the ability to distinguish between different UV models, based on the characteristic patterns they induce in SMEFT coefficients. The clear message is that future high-energy and high-precision colliders will sharpen SMEFT tests of the SM dramatically, making full use of the EFT framework to search for new physics.

\*\*\*

In summary, a vast array of phenomenological studies [112, 103, 68, 66, 122, 123, 20, 91, 99, 95, 94, 153, 101, 67, 79, 115, 100, 16, 18, 17, 37, 109, 78, 29, 47, 44, 147, 119, 48, 97, 137, 143, 90, 74, 93, 114, 53, 150, 25, 92, 78, 29, 107, 98, 62, 15, 46, 97, 80, 125, 30, 45, 33, 39, 63] have firmly established the SMEFT as a powerful and unifying framework to describe potential new physics in a model-independent way. The continued refinement of this framework, both in theoretical consistency and in confrontation with data, is a central component of modern particle physics research.

### 1.4. Outline of the thesis

This thesis investigates how future colliders can constrain new physics through precision observables, primarily within the SMEFT framework. Each chapter addresses a different angle of this broader goal, with a focus on collider-specific reach, theoretical clarity, and realistic observables.

- **Chapter 2. The Opening Repertoire** introduces collider kinematics, the SMEFT formalism, and statistical tools used throughout the thesis.
- **Chapter 3. EIC Wide Shut** studies PV DIS observables at the EIC and their role in resolving parameter degeneracies in SMEFT fits.
- **Chapter 4. DISentangling SMEFT: A Few Colliders More** combines SMEFT projections at the EIC, LHeC, and FCC-eh to test complementarity, include next-to-leading-order QCD corrections, and resolve flat directions in global fits using EW observables.
- **Chapter 5. Transcendental Étude in  $e^-$  Minor** introduces a novel transverse spin asymmetry at the FCC-ee to significantly improve sensitivity to the electron Yukawa coupling.
- **Chapter 6.  $CP$ -odd to Joy** explores  $CP$ -violating dimension-8 gluonic operators via Collins-Soper angular moments in the DY production with an associated jet.
- **Chapter 7. Conclusion** summarizes the findings, contributions, and future directions in precision collider phenomenology.

## CHAPTER 2

**The Opening Repertoire**

*The only difference between screwing around and science is writing it down.*

Adam Savage

This chapter outlines the theoretical foundations and methodological tools used throughout this thesis. We begin with collider phenomenology, focusing on three main classes of collisions: deep inelastic scattering, Drell-Yan production, and electron-positron annihilation. These processes are central to precision measurements and new physics searches, providing complementary access to the quantum chromodynamics and electroweak structure of the Standard Model. The discussion then moves on to an overview of the Standard Model Effective Field Theory, which provides a systematic framework to parametrize deviations from Standard Model predictions. We then examine the kinematics of many-particle final states and the construction of Lorentz-invariant phase space. Afterwards, we summarize the statistical methods used to obtain bounds on new physics parameters. Finally, we describe the computational frameworks used throughout this work, including symbolic amplitude calculations, numerical integration routines, and statistical analysis workflows.

## 2.1. Collider phenomenology

We have three main classes of collisions that serve as the core tools of collider phenomenology: deep inelastic scattering (DIS), Drell-Yan (DY) production, and electron-positron ( $e^-e^+$ ) annihilation. They offer complementary views into the structure of matter and the quantum chromodynamics (QCD) and electroweak (EW) interactions. Their full theoretical treatment fills textbooks, and here we focus on the basics relevant to our works.

### 2.1.1. Deep inelastic scattering

DIS involves the collision of a lepton  $\ell$  with a hadron  $H$ , typically a proton or a light nucleus, illustrated in Figure 2.1. At leading order, the lepton scatters off a single parton inside the hadron via neutral or charged gauge-boson exchange,  $V = \gamma, Z, W$ . The scattered lepton  $\ell'$  is detected, while the hadronic final state  $X$  is treated inclusively. In this thesis, we focus on neutral-current (NC) DIS processes mediated by a photon or  $Z$ -boson exchange.

A typical DIS process can be expressed as

$$(2.1) \quad \ell(k) + H(P) \rightarrow \ell'(k') + X.$$

Here,  $k$ ,  $k'$ , and  $P$  are the momentum of the incoming lepton, the scattered lepton, and the incoming hadron, respectively. The underlying partonic process at leading order is

$$(2.2) \quad \ell(k) + q(p) \rightarrow \ell'(k') + q_f(p'),$$

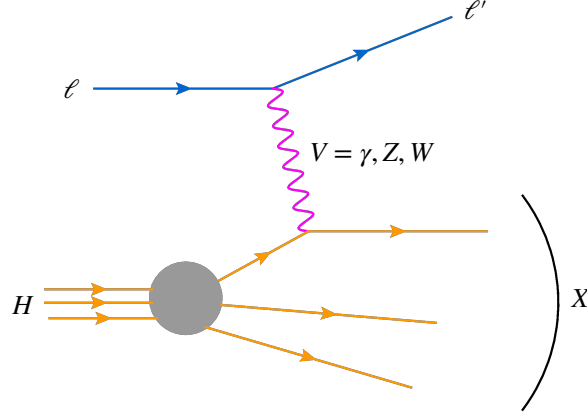


Figure 2.1. Schematic representation of a typical deep inelastic scattering event. An incoming lepton,  $\ell$ , scatters off a parton inside the hadron,  $H$ , via EW gauge-boson exchange, resulting in an outgoing lepton,  $\ell'$ , and an inclusive hadronic final state  $X$ . Throughout this thesis, blue lines indicate leptons, orange lines denote quarks, magenta lines represent EW gauge bosons, green lines represent gluons, and brown lines indicate scalar particles. This color scheme is used consistently throughout, and should be particularly helpful for the synesthetically inclined. Feynman diagrams are produced using Jaxodraw [38].

with  $q$  a quark or an antiquark inside the hadron. The Feynman diagram that represents this process is pictured in Figure 2.2.

Partonic and hadronic momenta are related by the Bjorken- $x$  parameter,  $p = xP$ . The standard notation for the momentum transfer is  $q = k - k'$ , with  $q^2 = -Q^2$ . Some of the fundamental kinematic relations relevant for DIS are given as follows:

$$(2.3) \quad p \cdot k = p_f \cdot k' = \frac{xs}{2}, \quad k \cdot k' = p \cdot q = \frac{Q^2}{2}, \quad P \cdot q = \frac{Q^2}{2x}, \quad \frac{P \cdot q}{P \cdot k} = \frac{p \cdot q}{p \cdot k} = y.$$

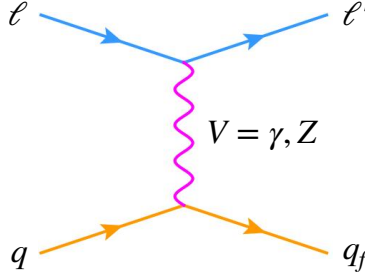


Figure 2.2. The Feynman diagram for the underlying partonic process for the DIS at leading order.

At leading order in massless approximation, these relations give us  $Q^2 = xys$ .

For the NC DIS, the matrix element has contributions from photon and  $Z$ -boson exchanges,

$$(2.4) \quad \mathcal{A} = \mathcal{A}_\gamma + \mathcal{A}_Z.$$

Let us write down the amplitudes in the parametric form that we call *the sleight of hand*. The sleight-of-hand parametrization tells us that the vertex factors are of the form  $V_{ffV}^\mu = i\gamma^\mu (C_{ffV}^L P_L + C_{ffV}^R P_R)$  and the propagators of the form  $\Delta_V^{\mu\nu}(k) = iN_V^{\mu\nu}(k)D_V(k)$ ; furthermore, when the propagator has the momentum that corresponds to one of the  $d = s, t, u$  channels, we use the shorthand notation  $\Delta_{Vd}^{\mu\nu}$ . Here, we call the  $C$  factors the *coupling strengths* and the  $D$  terms the *propagator denominators*, which we keep in the closed form until we actually plug in the numbers, which helps us keeping the input scheme general, as well as keeping track of any mistakes in writing down the amplitude. To illustrate the latter, if we ever see the factors  $C_{eeZ}^L{}^2$  and  $C_{qqZ}^L{}^2$  together with polarization factors  $(1 - \lambda_e)$  and  $(1 + \lambda_q)$  in an electron-quark (specifically not an antiquark)



scattering, we understand that the quark- $Z$  coupling must be actually  $C_{qqZ}^R$ . The propagator numerators, namely the  $N$  factors, are supposed to be evaluated immediately because they contain Lorentz structures. For a massless gauge boson in general, we have

$$(2.5) \quad N_V^{\mu\nu}(k) = -g^{\mu\nu}, \quad V = \gamma, g,$$

and for a massive gauge boson, we have

$$(2.6) \quad N_V^{\mu\nu}(k) = -g^{\mu\nu} + \frac{k^\mu k^\nu}{m_V^2}, \quad V = Z, W,$$

We assume unitary gauge because we work at tree level. It is beneficial to remark that if the massive gauge boson ever couples to external, massless fermions, the momentum-dependent term drops (because these terms give zero if we use the Dirac equation of motion). For the propagator denominators, we may use the full Breit-Wigner form only when we are certain that we assume a collider energy or an invariant-mass variable around a specific resonance point, and this can take place only at the numerical step, so we avoid having to carry around so many terms in our analytical expressions.

We take this opportunity also to introduce our convention for labeling external particles. Instead of deploying an index scheme such as  $1 + 2 \rightarrow 3 + 4 + \dots$ , we call the incoming particles  $a$  and  $b$ , and then we call the leading particle 1, subleading particle 2, subsubleading particle 3, and so on, namely for a collision, we have  $a + b \rightarrow 1 + 2 + 3 + \dots$ . For a decay, we drop  $b$ . Consequently, for a partonic process, we use the momenta  $p_a$ ,  $p_b$ ,  $p_1$ ,  $p_2$ ,  $p_3$ , and so on, and in fact, we go on to denote any given process in terms of momenta as  $p_a + p_b \rightarrow p_1 + p_2 + p_3 + \dots$ . We elaborate more on the momenta later in this chapter. Now, we write the underlying partonic processes for the DIS under consideration

in a manner consistent with our convention for particle labeling as

$$(2.7) \quad q(p_a) + e^-(p_b) \rightarrow q(p_1) + e^-(p_2),$$

$$(2.8) \quad q(p_a) + e^+(p_b) \rightarrow q(p_1) + e^+(p_2),$$

$$(2.9) \quad \bar{q}(p_a) + e^-(p_b) \rightarrow \bar{q}(p_1) + e^-(p_2),$$

$$(2.10) \quad \bar{q}(p_a) + e^+(p_b) \rightarrow \bar{q}(p_1) + e^+(p_2).$$

With our sleight-of-hand parametrization, we write the amplitudes as

$$(2.11) \quad \mathcal{A}_{ie^-} = \sum_{V=\gamma,Z} i[\bar{u}_1 V_{qqV}^\mu u_a][\bar{u}_2 V_{eeV}^\nu u_b] \Delta_{Vt}^{\mu\nu},$$

$$(2.12) \quad \mathcal{A}_{ie^+} = \sum_{V=\gamma,Z} i[\bar{u}_1 V_{qqV}^\mu u_a][\bar{v}_b V_{eeV}^\nu v_2] \Delta_{Vt}^{\mu\nu},$$

$$(2.13) \quad \mathcal{A}_{-ie^-} = \sum_{V=\gamma,Z} i[\bar{v}_a V_{qqV}^\mu v_1][\bar{u}_2 V_{eeV}^\nu u_b] \Delta_{Vt}^{\mu\nu},$$

$$(2.14) \quad \mathcal{A}_{-ie^+} = \sum_{V=\gamma,Z} i[\bar{v}_a V_{qqV}^\mu v_1][\bar{v}_b V_{eeV}^\nu v_2] \Delta_{Vt}^{\mu\nu},$$

where  $i = 1, 2, 3, 4, 5$ . Here, we assume the familiar Particle Data Group (PDG) particle identification numbers, namely the quarks are numbers  $(d, u, s, c, b) = (1, 2, 3, 4, 5)$ , anti-quarks picks up a minus sign, and the gluon is 0. For the sake of completeness, we write  $\mathcal{A}_{0e^\pm} = 0$ .

All Lorentz indices are written in the upstairs position. Since this work is purely phenomenological, we never deal with Lorentz indices showing up more than twice in any term, and raising/lowering would be superfluous. Aside from occasional SU(2) and SU(3) indices on the respective generator and the gauge bosons, which are always explicit, we

make no use of spinor, flavor, or other group-theoretical index structures. As a result, all index contractions are unambiguous.

We discuss the computation of squared amplitudes in the last section in this chapter, when we describe our computational infrastructure. Now, we write the partonic cross section as

$$(2.15) \quad \sigma_{ie^\pm}(\hat{s}) = F \int |\mathcal{A}_{ie^\pm}|^2 \text{dLIPS}, \quad i = q, \bar{q},$$

and the hadronic cross section is given by

$$(2.16) \quad \sigma_{He^\pm}(s) = \sum_{i=-N_f}^{N_f} \int \text{d}x_a \, f_i(x_a, \mu_F) \sigma_{ie^\pm}(\hat{s}),$$

where  $N_f$  is the number of active flavors,  $f_i(x_a, \mu_F)$  is the parton distribution function (PDF) evaluated at the Bjorken parameter  $x_a$  and the factorization scale  $\mu_F$ ,  $F$  is the flux factor, and dLIPS is the differential Lorentz-invariant phase space (LIPS), which we detail later in this chapter. We note that if any or both of the incoming beams,  $H$  or  $e^\pm$ , are polarized, then we can write a polarized cross section, or yet an observable based on various polarization states of the two beams that eventually depends on various linear combinations of polarized cross sections.

In processes involving hadrons in the initial state, such as the DIS process or the DY production discussed in the next section, the QCD factorization theorem allows the hadronic cross section to be expressed as a convolution of PDFs and perturbatively calculable partonic cross sections [96, 72]. PDFs encapsulate nonperturbative information about the momentum distribution of partons inside the hadron, while the partonic cross

section describes the short-distance interaction and is computed using standard techniques in perturbative QCD. This separation of scales is valid up to power-suppressed corrections and forms the theoretical foundation for the expressions used throughout this thesis.

### 2.1.2. Drell-Yan production

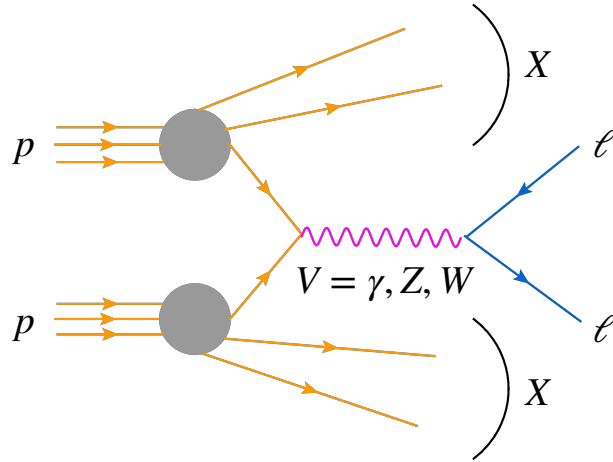


Figure 2.3. Schematic representation of Drell-Yan production in proton-proton collisions. A quark and an antiquark from the incoming protons annihilate via an intermediate EW gauge boson,  $V = \gamma, Z, W$ , which decays into a lepton pair. Additional hadronic activity from the proton remnants is denoted by  $X$  and treated inclusively. This is a general schematic; in our analysis, we focus on neutral-current production and include a real jet in the final state.

The DY process describes the production of an EW gauge boson in quark-antiquark annihilation, followed by its decay into a lepton pair; in this thesis, we focus exclusively on NC DY processes. At leading order, the partonic process is  $q\bar{q} \rightarrow \gamma^*/Z \rightarrow \ell^-\ell^+$ . The

lepton pair kinematics offer precision access to EW parameters and are highly sensitive to potential new physics contributions.

We study the DY production in association with a real jet (DY $j$ ) and in which the final leptonic product is an electron-positron pair. The process can be expressed as

$$(2.17) \quad p(P_a) + p(P_b) \rightarrow j(p_1) + V(p_{23}) \rightarrow j(p_1) + e^-(p_2) + e^+(p_3),$$

where  $ps$  are the incoming protons,  $V$  is the intermediate EW gauge boson that later decays into the lepton pair. We take this opportunity to emphasize our notation for larger processes that have more than four external states and that can be expressed as chain events. The entire process can be considered as a  $jV$  production, followed by the decay  $V \rightarrow e^-e^+$ . Our labeling scheme is  $a + b \rightarrow 1 + 2 + 3$  for the process in the big picture, but since  $V$  is the mother for particles 2 and 3, we find it appropriate to denote it 23. The underlying partonic process at leading order is

$$(2.18) \quad p_i(p_a) + p_j(p_b) \rightarrow j(p_1) + V(p_{23}) \rightarrow j(p_1) + e^-(p_2) + e^+(p_3),$$

where  $p_i$  and  $p_j$  could be a quark, an antiquark, or a gluon inside the protons. Since the initial states of the hadronic process are identical, there is no way of knowing which parton comes from which proton, which we take care of carefully when we write down the hadronic cross section. With the real jet accompanying the electron-positron pair, the Feynman diagrams depicted in Figure 2.4 look more populated compared to a usual DY production.

Partonic and hadronic momenta are related by the Bjorken- $x$  parameters,  $p_a = x_a P_a$  and  $p_b = x_b P_b$ . The standard quantities relevant to this process are the Mandelstam

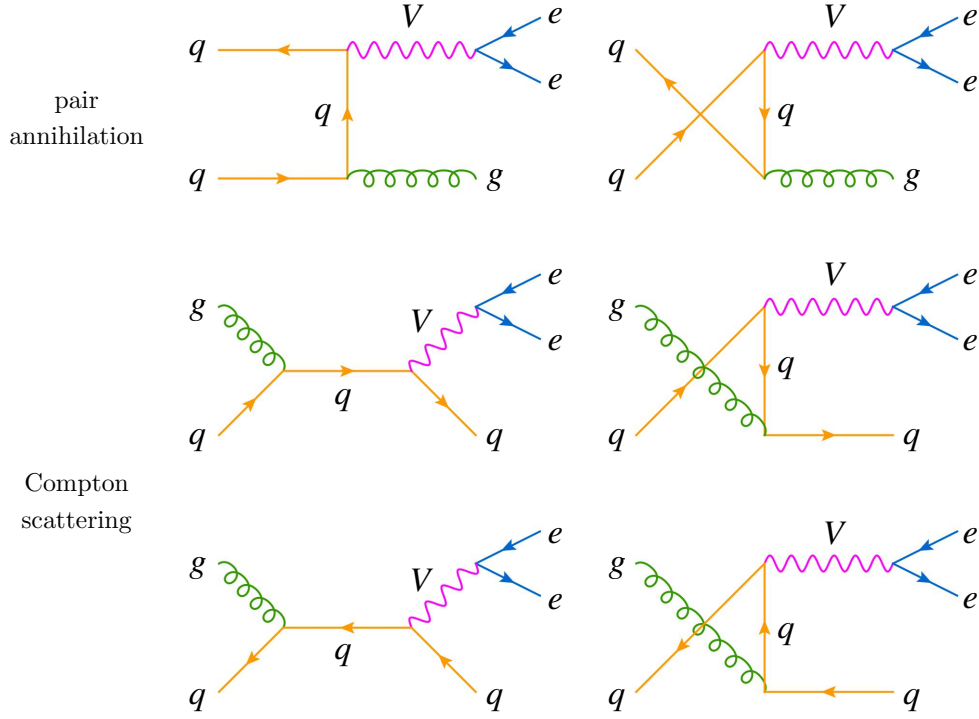


Figure 2.4. The Feynman diagrams for the underlying partonic process for the Drell-Yan and jet production at leading order.

invariants of the  $2 \rightarrow 2$  subprocess  $p_i(p_a) + p_j(p_b) \rightarrow j(p_1) + V(p_{23})$ , namely  $\hat{s} = (p_a + p_b)^2$ ,  $\hat{t} = (p_a - p_1)^2$ , and  $\hat{u} = (p_a - p_{23})^2$ . We denote the dilepton invariant mass by  $m_{\ell\ell}$  such that  $m_{\ell\ell}^2 = (p_2 + p_3)^2$ . From momentum conservation, we notice that this quantity is also the invariant mass of the EW gauge boson  $V$ , corresponding to the momentum  $p_{23}$ , namely  $m_{\ell\ell}^2 = p_{23}^2$ . Speaking of invariant mass, we have one more, namely the invariant mass of the  $jV$  system, denoted  $m_{jV}$ . From momentum conservation, we note that this is nothing but the Mandelstam- $s$  parameter for the partonic process, namely  $\hat{s} = (p_a + p_b)^2 = (p_1 + p_{23})^2 = m_{jV}^2$ . At the Large Hadron Collider, where the collider energy is at the TeV scale, we assume massless fermions. In conjunction with this, we note that the familiar relation involving the Mandelstam parameters holds true for the

$2 \rightarrow 2$  subprocess  $p_i(p_a) + p_j(p_b) \rightarrow j(p_1) + V(p_{23})$ ; to wit,

$$(2.19) \quad \hat{s} + \hat{t} + \hat{u} = m_a^2 + m_b^2 + m_1^2 + m_{23}^2 = 0 + 0 + 0 + m_{\ell\ell}^2,$$

so we can express  $\hat{u} = m_{\ell\ell}^2 - m_{jV}^2 - \hat{t}$  to eliminate one Mandelstam parameter. Other quantities relevant to the DY production are the dilepton transverse momentum, rapidity, pseudorapidity, azimuthal angle, and beam separation, denoted  $p_\top$ ,  $y$ ,  $\eta$ ,  $\phi$ , and  $\Delta R$ , respectively, which are all measured in the *lab* frame, namely the hadronic center-of-mass (c.m.) frame. For particle momenta  $k$  and  $k'$ , the standard definitions are

$$(2.20) \quad k_T = \sqrt{k_x^2 + k_y^2}, \quad y_k = \frac{1}{2} \log \left( \frac{E_k + k_z}{E_k - k_z} \right), \quad \eta_k = \frac{1}{2} \log \left( \frac{|\mathbf{k}| + k_z}{|\mathbf{k}| - k_z} \right),$$

$$(2.21) \quad \phi_k = \arctan \left( \frac{k_y}{k_x} \right), \quad \Delta R_{kk'} = \sqrt{\Delta\eta_{kk'} + \Delta\phi_{kk'}},$$

where  $\Delta\eta_{kk'} = \eta_k - \eta_{k'}$  and  $\Delta\phi_{kk'} = \phi_k - \phi_{k'}$ .

For the NC DY production, the amplitude has contributions from photon and  $Z$ -boson channels,

$$(2.22) \quad \mathcal{A} = \mathcal{A}_\gamma + \mathcal{A}_Z,$$

and the interference between these contributions plays a significant role in observables like  $A_{\text{FB}}$ , especially in the vicinity of the  $Z$  pole. Let us write down the amplitudes,  $\mathcal{A}_{ij}$ , with the sleight-of-hand parametrization. Here,  $i$  and  $j$  are parton flavors. With  $N_f = 5$ , in principle we have 121 amplitudes. Since we have massless fermions and since we focus exclusively on neutral current (so the flavor change is forbidden, up and charm amplitudes are equal, and down, strange, and bottom amplitudes are identical), this

number becomes 25. Furthermore, not all processes exist. The only nontrivial processes are  $ij = 20, 2 - 2, 10, 1 - 1, 02, 01, 0 - 1, 0 - 2, -11, -10, -22, -20$ . The sleight-of-hand amplitudes are

$$(2.23) \quad \mathcal{A}_{i0} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} i[\bar{u}_1 V_{qqV}^\nu \Delta_{qs} V_{qqg}^\rho u_a][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_b^\rho \\ + i[\bar{u}_1 V_{qqg}^\rho \Delta_{qu} V_{qqV}^\nu u_a][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_b^\rho \end{array} \right], \quad i = 1, 2$$

$$(2.24) \quad \mathcal{A}_{i-i} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} [\bar{v}_b V_{qqV}^\nu \Delta_{qs} V_{qqg}^\rho u_a][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_1^{*\rho} \\ + i[\bar{v}_b V_{qqg}^\rho \Delta_{qu} V_{qqV}^\nu u_a][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_1^{*\rho} \end{array} \right], \quad i = 1, 2$$

$$(2.25) \quad \mathcal{A}_{0i} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} i[\bar{u}_1 V_{qqV}^\nu \Delta_{qs} u_b][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_a^\rho \\ + i[\bar{u}_1 V_{qqg}^\rho \Delta_{qt} V_{qqV}^\nu u_b][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_a^\rho \end{array} \right], \quad i = 1, 2$$

$$(2.26) \quad \mathcal{A}_{0-i} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} i[\bar{v}_b V_{qqg}^\rho \Delta_{qs} V_{qqV}^\nu v_1][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_a^\rho \\ + i[\bar{v}_b V_{qqV}^\nu \Delta_{qt} V_{qqg}^\rho v_1][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_a^\rho \end{array} \right], \quad i = 1, 2$$

$$(2.27) \quad \mathcal{A}_{-ii} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} i[\bar{v}_a V_{qqg}^\rho \Delta_{qt} V_{qqV}^\nu u_b][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_1^{*\rho} \\ + i[\bar{v}_a V_{qqV}^\nu \Delta_{qu} V_{qqg}^\rho u_b][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_1^{*\rho} \end{array} \right], \quad i = 1, 2$$

$$(2.28) \quad \mathcal{A}_{-i0} = \sum_{V=\gamma,Z} \left[ \begin{array}{l} i[\bar{v}_a V_{qqg}^\rho \Delta_{qs} V_{qqV}^\nu v_1][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_b^\rho \\ + i[\bar{v}_a V_{qqV}^\nu \Delta_{qu} V_{qqg}^\rho v_1][\bar{u}_2 V_{eeV}^\mu v_3] \Delta_{V23}^{\mu\nu} \varepsilon_b^\rho \end{array} \right], \quad i = 1, 2.$$



Here,  $\varepsilon_k^{(*)}$  is the gluon polarization vector corresponding to the momentum  $p_k$  when the gluon is incoming (outgoing), and we have a new vertex factor,  $V_{qqg}^\mu = iC_{qqg}T_{ab}^A\gamma^\mu$ , and multiple new propagators,  $\Delta_{qd}$  and  $\Delta_{V23}^{\mu\nu}$ . The former is the quark propagator in the indicated channel,  $d = s, t, u$ , namely  $\Delta_{qd} = i\hat{k}/\hat{d}$ , where  $k$  is such that  $k^2 = \hat{d}$ . The latter is the good old EW gauge-boson propagator but now having the momentum  $p_{23}$ .

Now, the partonic cross section is given by

$$(2.29) \quad \sigma_{ij}(\hat{s}) = F \int |\mathcal{A}_{ij}|^2 \text{dLIPS},$$

and the hadronic cross section is obtained as

$$(2.30) \quad \sigma(s) = \sum_{i=-N_f}^{N_f} \sum_{j=-N_f}^{N_f} \int \text{d}x_a \text{d}x_b f_i(x_a, \mu_F) f_j(x_b, \mu_F) \sigma_{ij}(\hat{s}).$$

Adding the partonic cross section this way, weighted by the corresponding factors of PDFs, makes sure that we remain blind as to which parton comes from which proton.

The angular analysis of the DY production typically takes place by expressing the differential hadronic cross section in terms of spherical harmonics:

$$(2.31) \quad \frac{\text{d}\sigma}{\text{d}\Omega^\star} = \frac{3\sigma}{16\pi} \left[ 1 + c_{\theta^\star}^2 + \sum_{m=0}^7 A_m Y_m(\Omega^\star) \right],$$

where  $\Omega^\star = (\theta^\star, \varphi^\star)$  are the Collins-Soper (CS) angles [73], the  $Y_m$  are just some linear combinations of the spherical harmonics  $Y_1^0$ ,  $Y_1^1 \pm Y_1^{-1}$ ,  $Y_2^0$ ,  $Y_2^1 \pm Y_2^{-1}$ , and  $Y_2^2 \pm Y_2^{-2}$ , namely

$$(2.32) \quad Y_0 = \frac{1}{2}(1 - 3c_{\theta^\star}^2), \quad Y_1 = s_{2\theta^\star}c_{\varphi^\star}, \quad Y_2 = \frac{1}{2}s_{\theta^\star}^2c_{2\varphi^\star}, \quad Y_3 = s_{\theta^\star}c_{\varphi^\star}, \quad Y_4 = c_{\theta^\star},$$

$$(2.33) \quad Y_5 = s_{\theta^*}^2 s_{2\varphi^*}, \quad Y_6 = s_{2\theta^*} s_{\varphi^*}, \quad Y_7 = s_{\theta^*} s_{\varphi^*},$$

with

$$(2.34) \quad \int d\Omega^* Y_m Y_n \propto \delta_{mn},$$

and the  $A_m$  are the CS moments. Here,  $s$  and  $c$  are shorthand notations for the sine and cosine of the angle given in the subscript, respectively. We note that the CS angles are directly related to the lab momenta of the electron and the positron:

$$(2.35) \quad c_{\theta^*} = \frac{2(p_z^{e^-} E^{e^+} - p_z^{e^+} E^{e^-})}{m_{\ell\ell} \sqrt{m_{\ell\ell}^2 + p_{\top}^2}}, \quad \varphi^* = \arctan \left[ \frac{2p_y^{e^-} \sqrt{m_{\ell\ell}^2 + p_{\top}^2}}{m_{\ell\ell}(p_x^{e^-} - p_x^{e^+})} \right],$$

where  $p_{\top}$  is the transverse momentum of the dilepton system, and  $E^{e^{\pm}}$  and  $p_{x,y,z}^{e^{\pm}}$  are the energy and momentum components of the electron and positron, provided that the  $jV$  system is produced in the  $xz$  plane in the lab frame. Using the orthogonality of the spherical harmonics, we obtain

$$(2.36) \quad A_0 = \frac{20}{3} \langle Y_0 \rangle + \frac{2}{3}, \quad A_1 = 5 \langle Y_1 \rangle, \quad A_2 = 20 \langle Y_2 \rangle, \quad A_3 = 4 \langle Y_3 \rangle, \quad A_4 = 4 \langle Y_4 \rangle,$$

$$(2.37) \quad A_5 = 5 \langle Y_5 \rangle, \quad A_6 = 5 \langle Y_6 \rangle, \quad A_7 = 4 \langle Y_7 \rangle,$$

with

$$(2.38) \quad \langle Y_m \rangle = \frac{\int Y_m d\sigma}{\sigma}.$$

In a proton-proton collision, the incoming quark typically carries a larger momentum fraction than the antiquark, imparting a net boost to the intermediate gauge boson  $V$  along

the beam axis. Consequently, the CS polar angle  $\theta^*$  must be reconstructed experimentally as

$$(2.39) \quad c_{\theta_R^*} = \text{sign}(p_z^V) c_{\theta^*}$$

to assign the correct direction [6, 148], where  $p_z^V$  is the  $z$  component of the lab momentum of the EW gauge boson  $V$ . This ensures that the angular distributions reflect the true kinematics of the process. Notably, the azimuthal angle  $\varphi^*$  does not require such an adjustment. We use the reconstructed expression in our calculations of the expectation values of the angular structures.

These angular coefficients, or the CS moments, encode the full angular structure of the dilepton final state in the DY production. The moments  $A_0$  through  $A_7$  correspond to spherical harmonics up to  $\ell = 2$ , reflecting the spin-1 nature of the intermediate EW boson in the Standard Model (SM) at leading order. In this picture, the lepton pair is produced from an  $s$ -channel vector current, and the angular distribution reflects the interference patterns of different helicity amplitudes. Specifically,  $A_0$  and  $A_2$  are sensitive to longitudinal and transverse polarizations,  $A_1$  and  $A_3$  probe spin correlations and parity-violating effects, and  $A_4$  captures the forward-backward asymmetry. The remaining coefficients  $A_5$  through  $A_7$  are zero at leading order in the SM but can be activated at higher QCD orders, by  $CP$ -odd and higher-dimensional SMEFT operators, or simply by including additional jets in the final state at tree level (so the azimuthal symmetry of the intermediate states  $jj \dots jV$  is broken). Notably, angular structures beyond  $\ell = 2$ , such as those corresponding to the  $B_m$  harmonics [15], arise only through spin-2 interference patterns and are a hallmark of dimension-8 SMEFT contributions, as these operators

induce  $\ell = 2$  partial waves at the amplitude level, allowing for  $\ell = 3$  harmonics in the squared amplitude.

### 2.1.3. Electron-positron annihilation

Electron-positron ( $e^-e^+$ ) annihilation offers a clean theoretical and experimental environment, free from hadronic uncertainties such as PDFs. In this thesis, we study the  $b\bar{b}$  and semi-leptonically decaying  $W^-W^+$  channels, namely  $e^-e^+ \rightarrow b\bar{b}$  and  $e^-e^+ \rightarrow W^-W^+ \rightarrow \ell\nu jj$ . With these processes, we can probe EW interactions and quantum interference near the Higgs resonance, providing an opportunity to access the electron Yukawa coupling.

The use of transversely polarized electron beams and longitudinally polarized positron beams allows the construction of single and double spin asymmetries that are linearly sensitive to the electron Yukawa coupling. These asymmetries emerge from the interference between the Higgs-mediated signal amplitude and the SM background. In the  $b\bar{b}$  final state, the Higgs contribution interferes with the dominant EW production channels; in the semi-leptonic  $WW$  final state, the interference arises between the Higgs and the continuum  $WW$  production processes. In both cases, the signal is enhanced through beam polarization and the use of angular observables that isolate the relevant interference terms.

Among all possible Higgs decay modes and relevant background processes illustrated in Figure 2.5, the  $b\bar{b}$  and semi-leptonic  $WW$  channels are chosen due to their large branching ratios. These channels offer the best statistical reach while maintaining manageable experimental complexity. Ref. [88] established the baseline sensitivity using unweighted cross section measurements in these final states and noted that the use of polarized beams

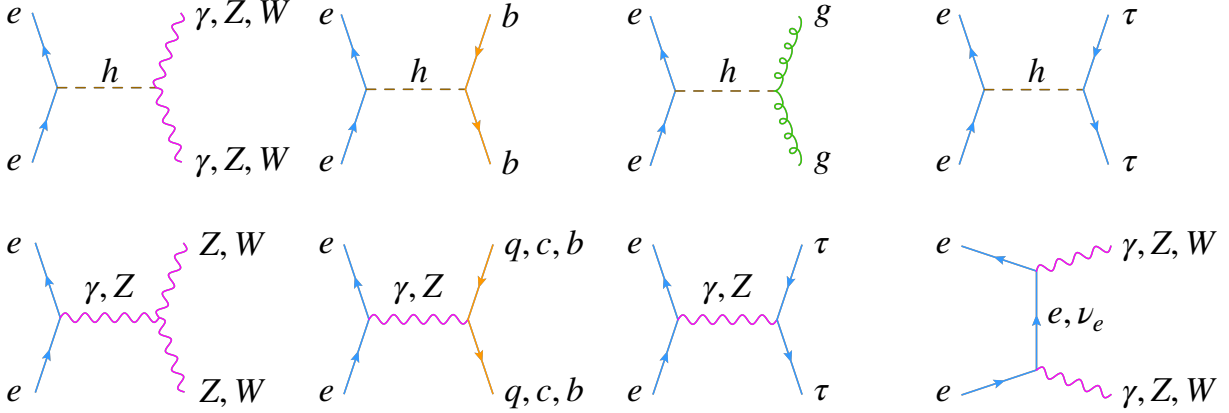


Figure 2.5. Representative diagrams for Higgs production (top) and its decay into EW bosons or fermions and gluons, alongside the dominant background processes (bottom). Adapted from Ref. [88].

could enhance the reach, particularly through interference effects involving the Higgs-mediated amplitude. While that study did not employ polarization-sensitive observables or optimized event weighting, it pointed out the opportunity for further improvement via tailored asymmetries. This motivates our focus on constructing spin asymmetries that are linearly sensitive to the electron Yukawa coupling.

These annihilation processes, due to their cleanliness and sensitivity to small couplings, complement the hadronic probes discussed earlier and play a critical role in testing the SM at sub-percent precision.

The processes of interest are

$$(2.40) \quad e^-(p_a) + e^+(p_b) \rightarrow b(p_1) + \bar{b}(p_2),$$

$$(2.41) \quad e^-(p_a) + e^+(p_b) \rightarrow W^-(p_{12}) + W^+(p_{34}) \rightarrow \ell^-(p_1) + \bar{\nu}_\ell(p_2) + u_f(p_3) + \bar{d}_f(p_4),$$

$$(2.42) \quad e^-(p_a) + e^+(p_b) \rightarrow W^+(p_{12}) + W^-(p_{34}) \rightarrow \ell^+(p_1) + \nu_\ell(p_2) + \bar{u}_f(p_3) + d_f(p_4).$$

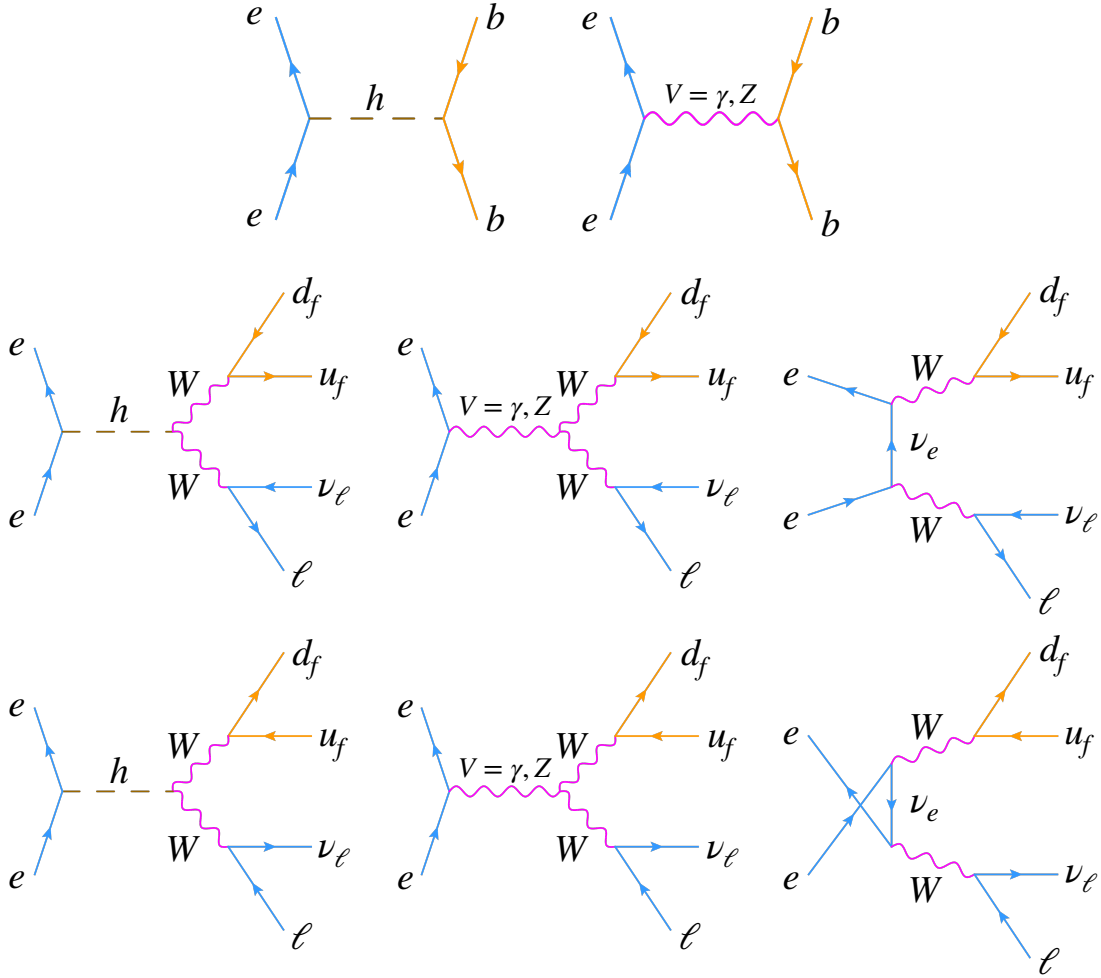


Figure 2.6. Leading-order Feynman diagrams for  $e^-e^+ \rightarrow b\bar{b}$  (top) and  $e^-e^+ \rightarrow W^-W^+ \rightarrow \ell\nu jj$  (bottom two).

The leading-order Feynman diagrams within the SM are presented in Figure 2.6. Here,  $\ell$  could be an electron, muon, or tau, and  $u_f$  ( $d_f$ ) represents an up (down) or charm (strange) quark. In our work, we set the Cabibbo-Kobayashi-Maskawa matrix equal to

identity. The sleight-of-hand amplitudes are

$$(2.43) \quad \mathcal{A} = \left\{ \begin{aligned} & i[\bar{v}_b V_{eeh} u_a][\bar{u}_1 V_{bbh} v_2] \Delta_{hs} \\ & + \sum_{V=\gamma, Z} i[\bar{v}_b V_{eeV}^\mu u_a][\bar{u}_1 V_{bbV}^\nu v_2] \Delta_{Vs}^{\mu\nu} \end{aligned} \right\}$$

for the  $b\bar{b}$  process,

$$(2.44) \quad \mathcal{A} = \left\{ \begin{aligned} & i[\bar{v}_b V_{eeh} u_a][\bar{u}_1 V_{ff'W}^\mu v_2][\bar{u}_3 V_{ff'W}^\rho v_4] V_{hWW}^{\nu\sigma} \Delta_{hs} \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \\ & + \sum_{V=\gamma, Z} [\bar{v}_b V_{eeV}^\alpha u_a][\bar{u}_1 V_{ff'W}^\mu v_2][\bar{u}_3 V_{ff'W}^\rho v_4] V_{VWW}^{\beta\nu\sigma} \Delta_{Vs}^{\alpha\beta} \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \\ & + i[\bar{v}_b V_{ff'W}^\sigma \Delta_{\nu t} V_{ff'W}^\nu u_a][\bar{u}_1 V_{ff'W}^\mu v_2][\bar{u}_3 V_{ff'W}^\rho v_4] \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \end{aligned} \right\}$$

for the  $WW$  process if the lepton flying off has negative charge, and

$$(2.45) \quad \mathcal{A} = \left\{ \begin{aligned} & i[\bar{v}_b V_{eeh} u_a][\bar{u}_2 V_{ff'W}^\mu v_1][\bar{u}_4 V_{ff'W}^\rho v_3] V_{hWW}^{\nu\sigma} \Delta_{hs} \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \\ & + \sum_{V=\gamma, Z} [\bar{v}_b V_{eeV}^\alpha u_a][\bar{u}_2 V_{ff'W}^\mu v_1][\bar{u}_4 V_{ff'W}^\rho v_3] V_{VWW}^{\beta\nu\sigma} \Delta_{Vs}^{\alpha\beta} \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \\ & + i[\bar{v}_b V_{ff'W}^\nu \Delta_{\nu u} V_{ff'W}^\sigma u_a][\bar{u}_2 V_{ff'W}^\mu v_1][\bar{u}_4 V_{ff'W}^\rho v_3] \Delta_{W12}^{\mu\nu} \Delta_{W34}^{\rho\sigma} \end{aligned} \right\}$$

for the  $WW$  process if the lepton flying off has positive charge. Here, we have defined new quantities. For the vertex factors, we have  $V_{ff'W} = iC_{ff'W}\gamma^\mu P_L$  as the universal  $ff'W$  coupling for leptons and quarks,  $V_{ffh} = iC_{ffh}$  the fermion-Higgs vertex, which means  $C_{ffh}$  is essentially the Yukawa coupling of fermion  $f$ ,  $V_{hWW}^{\mu\nu} = iC_{hWW}g^{\mu\nu}$  the coupling of the Higgs particle to two  $W$  bosons, and  $V_{VWW}^{\mu\nu\lambda} = iC_{VWW}S_{p_V p_W^- p_W^+}^{\mu\nu\lambda}$  the triple EW

gauge-boson interaction, with  $S_{qpk}^{\mu\nu\lambda} = g^{\mu\nu}(q-p)^\lambda + g^{\nu\lambda}(p-k)^\mu + g^{\lambda\mu}(k-q)^\nu$ . As for the propagators,  $\Delta_{\nu d} = i\hat{k}/\hat{d}$  is the neutrino propagator for the  $d$  channel, namely  $k^2 = \hat{d}$ ,  $\Delta_{hs}$  is the Higgs propagator in the  $s$  channel,  $\Delta_{V_s}^{\mu\nu}$  is the propagator for  $V = \gamma, Z$  in the  $s$  channel, and  $\Delta_{W12(34)}^{\mu\nu}$  is the  $W$  propagator with momentum  $p_{12(34)}$ . We note that at the SM, the Yukawa couplings are proportional to the mass of the fermion; however, in our work, we keep it as  $C_{ffh}$  to distinguish it from the mass term deriving from the equation of motion. We are interested in the squared amplitudes to the leading order in electron mass. We expand squared amplitudes to leading order in bottom mass for the  $b\bar{b}$  process and we assume massless final states for the  $WW$  process.

We detail squared amplitude computations later in this chapter, and we emphasize here an important aspect. We are interested in transversely polarized electrons and longitudinally polarized positrons. Since we are assuming massive initial states, this requires careful use of generalized projectors [43]. In the computation of the squared amplitude, we make use of the relations

$$(2.46) \quad u_\lambda(p)\bar{u}_\lambda(p) = (\not{p} + m)\mathbb{P}_\lambda^+(S),$$

$$(2.47) \quad v_\lambda(p)\bar{v}_\lambda(p) = (\not{p} - m)\mathbb{P}_\lambda^-(S).$$

Here,  $\mathbb{P}_\lambda^\pm(S)$  is the generalized projector for the particle/antiparticle spinors, given by

$$(2.48) \quad \mathbb{P}_\lambda^\pm(S) = \frac{1 \pm \lambda\gamma_5 \not{S}}{2},$$

where  $S^\mu$  is the polarization 4-vector. We detail our analysis in the relevant chapter later in the thesis.



Now, the cross section is given by

$$(2.49) \quad \sigma(s) = F \int |\mathcal{A}|^2 \text{dLIPS},$$

for both of the  $b\bar{b}$  and  $WW$  processes. We note that since we specify the polarization states of the incoming beams explicitly by introducing the helicity signs  $\lambda_{a/b}$  for the incoming electron/positron beam, the cross section is technically a function of  $\lambda_a$  and  $\lambda_b$  through the squared amplitude, a fact that we later exploit to define asymmetries.

## 2.2. The SMEFT formalism

The Standard Model Effective Field Theory (SMEFT) is a model-independent extension of the SM. In this framework, one constructs operators of mass dimension  $n > 4$ , denoted  $O_k^{(n)}$ , deploying the existing SM particle spectrum. Each operator is introduced with an effective coupling strength,  $C_k^{(n)}$ , which are called Wilson coefficients. These effective couplings are defined at an ultraviolet cutoff scale,  $\Lambda$ . We assume that  $\Lambda$  is heavier than all SM fields and beyond accessible collider energy. The SMEFT Lagrangian is given schematically by

$$(2.50) \quad \mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_{n>4} \frac{1}{\Lambda^{n-4}} \sum_k C_k^{(n)} O_k^{(n)}.$$

We note that the SMEFT operators modify the SM vertices in a gauge-invariant manner and generate gauge-invariant amplitudes if it is a new interactions absent in the SM. Gauge invariance is a fundamental symmetry of the SM, ensuring the consistency of the theory and the preservation of key principles such as unitarity and renormalizability. In the SMEFT framework, maintaining gauge invariance is essential because it guarantees

that the extended theory respects the same underlying symmetry structure as the SM. This constrains the possible forms of higher-dimensional operators. The construction of gauge-invariant operators prevents the introduction of unphysical degrees of freedom and maintains the predictive power of the SMEFT extension.

Let's consider the operators that are relevant to our studies and that we use in our works presented later in the thesis. We start with the DIS. At the parton level, the DIS is represented by the interaction of a leptonic current with a quark current by exchanging an EW gauge boson, which we assume to be neutral. Thus, there are operators in the form  $(\bar{\ell}\Gamma\ell)(\bar{q}\Gamma'q)$ , where  $\ell$  and  $q$  are Dirac fermions and  $\Gamma$  and  $\Gamma'$  are some Dirac matrices, that could contribute to the amplitude. At high energies, we assume massless fermions, which are then chiral eigenstates, and since the SM amplitude has definite chiral structure, we can have only operators of the form  $(\bar{\ell}\gamma^\mu P_X\ell)(\bar{q}\gamma_\mu P_Yq)$  that conserve helicity, where  $P_X$  and  $P_Y$  are the chiral projections operators,  $P_L$  or  $P_R$ . A fermionic field has mass dimension 3/2, so such an operator would have dimension 6, so this is the lowest dimension we can consider to extend our amplitudes to a new physics scenario. Thus, our dimension-6 operators are of the form

$$(2.51) \quad O_{XY} = (\bar{\ell}\gamma^\mu P_X\ell)(\bar{q}\gamma_\mu P_Yq).$$

We call operators of this form *semi-leptonic four-fermion operators*. Now let  $\ell$  and  $q$  denote SU(2) left-handed doublets and  $e$ ,  $u$ , and  $d$  denote SU(2) right-handed singlets. In the Warsaw basis [110], there are seven such operators:

$$(2.52) \quad O_{\ell q}^{(1)} = (\bar{\ell}\gamma^\mu\ell)(\bar{q}\gamma_\mu q), \quad O_{\ell q}^{(3)} = (\bar{\ell}\gamma^\mu\tau^I\ell)(\bar{q}\gamma_\mu\tau^Iq),$$

$$(2.53) \quad O_{\ell u} = (\bar{\ell}\gamma^\mu\ell)(\bar{u}\gamma_\mu u), \quad O_{\ell d} = (\bar{\ell}\gamma^\mu\ell)(\bar{d}\gamma_\mu d),$$

$$(2.54) \quad O_{eu} = (\bar{e}\gamma^\mu e)(\bar{u}\gamma_\mu u), \quad O_{ed} = (\bar{e}\gamma^\mu e)(\bar{d}\gamma_\mu d),$$

$$(2.55) \quad O_{qe} = (\bar{e}\gamma^\mu e)(\bar{q}\gamma_\mu q),$$

where the  $\tau^I$  are the Pauli matrices. The corresponding vertex factors are of the form

$$(2.56) \quad V_{XY} = iC_{XY}[\gamma^\mu P_X]_\ell[\gamma_\mu P_Y]_q,$$

where  $\ell$  and  $q$  indicate the leptonic and quark currents, and

$$(2.57) \quad C_{LL}^{u/d} = C_{\ell q}^{(1)} \mp C_{\ell q}^{(3)}, \quad C_{LR}^{u/d} = C_{\ell u/d}, \quad C_{RL}^{u/d} = C_{qe}, \quad C_{RR}^{u/d} = C_{eu/d}$$

for the up or down-like quarks. The SMEFT Feynman diagram for the underlying partonic process of the DIS is painted in Figure 2.7.

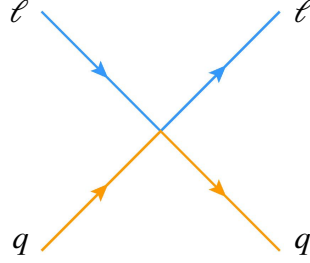


Figure 2.7. The SMEFT Feynman diagram for the underlying partonic process of the DIS generated by the semi-leptonic four-fermion operators.

The corresponding amplitudes are

$$(2.58) \quad \mathcal{A}_{ie^-} = \sum_{XY} C_{XY}^i [\bar{u}_2 \gamma^\mu P_X u_b] [\bar{u}_1 \gamma^\mu P_Y u_a],$$

$$(2.59) \quad \mathcal{A}_{-ie^-} = \sum_{XY} C_{XY}^i [\bar{u}_2 \gamma^\mu P_X u_b] [\bar{v}_a \gamma^\mu P_Y v_1],$$

$$(2.60) \quad \mathcal{A}_{ie^+} = \sum_{XY} C_{XY}^i [\bar{v}_b \gamma^\mu P_X v_2] [\bar{u}_1 \gamma^\mu P_Y u_a],$$

$$(2.61) \quad \mathcal{A}_{-ie^+} = \sum_{XY} C_{XY}^i [\bar{v}_b \gamma^\mu P_X v_2] [\bar{v}_a \gamma^\mu P_Y v_1],$$

where  $i = 1, 2, 3, 4, 5$  represent quark flavors  $d, u, s, c, b$ . We add these amplitudes to the corresponding SM amplitudes given in Eqs. (2.11)–(2.14). At this point, we emphasize that we assume flavor universality of the SMEFT operators. To illustrate, the operator  $O_{\ell q}^{(1)}$  could describe the electron-up quark coupling, as well as muon-strange quark coupling. Furthermore, since we have NC interaction and since we are in the massless limit, up and charm quarks, as well as down, strange, and bottom quarks are identical at the amplitude level. This is the basis of our calculations for Chapter 3.

The next we can do is to consider the SMEFT corrections to the  $ffV$  vertices, where  $V$  is a neutral EW gauge boson. We expect these corrections to be of the form

$$(2.62) \quad V_{ffV}^\mu = V_{ffV,\text{SM}}^\mu \left( 1 + \sum_k C_k v_k \right),$$

where  $v_k$  is some SMEFT shift to the SM vertex factor characterized by the Wilson coefficient  $C_k$ . We note that these modifications respect gauge invariance and all the underlying symmetries of the SM Lagrangian. In the Warsaw basis, there are 10 dimension-6 operators that cause this kind of a modification:

$$(2.63) \quad O_{\varphi WB} = (\varphi^\dagger \tau^I \varphi) W_{\mu\nu}^I B^{\mu\nu}, \quad O_{\varphi D} = (\varphi^\dagger D_\mu \varphi)^* (\varphi^\dagger D^\mu \varphi),$$

$$(2.64) \quad O_{\varphi \ell}^{(1)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi) (\bar{\ell} \gamma^\mu \ell), \quad O_{\varphi \ell}^{(3)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \tau^I \varphi) (\bar{\ell} \gamma^\mu \tau^I \ell),$$

$$(2.65) \quad O_{\varphi e} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{e}\gamma^\mu e),$$

$$(2.66) \quad O_{\varphi q}^{(1)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{q}\gamma^\mu q), \quad O_{\varphi q}^{(3)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \tau^I \varphi)(\bar{q}\gamma^\mu \tau^I q),$$

$$(2.67) \quad O_{\varphi u} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{u}\gamma^\mu u), \quad O_{\varphi d} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{d}\gamma^\mu d),$$

$$(2.68) \quad O_{\ell\ell} = (\bar{\ell}\gamma^\mu \ell)(\bar{\ell}\gamma^\mu \ell),$$

where  $\varphi$  is the SU(2) scalar doublet,  $W_{\mu\nu}^I$  and  $B_{\mu\nu}$  are SU(2) and U(1) gauge boson field strength tensors,  $D_\mu$  is the covariant derivative, and as is standard in this basis,  $\ell$  and  $q$  are SU(2) left-handed doublets and  $e$ ,  $u$ , and  $d$  are right-handed singlets. The left-right covariant derivative is such that

$$(2.69) \quad \varphi^\dagger i \overleftrightarrow{D}^\mu \varphi = \varphi^\dagger i D^\mu \phi + \text{h.c.},$$

where h.c. denotes the Hermitian conjugate. We are only interested in the Higgs vacuum expectation value,  $v$ , from the scalar doublet and the gauge-boson coupling terms in the covariant derivative. It might not be obvious at first glance how the operators without two fermionic fields and a covariant derivative could possibly modify the  $ffV$  vertices, and the answer is via the input scheme. In the  $G_\mu$  input scheme, also known as the  $\{G_F, \alpha, m_Z\}$  input scheme, we perform the transformation from the bare parameters  $\{g_2, g_1, v\}$ , where  $g_2$  is the SU(2) coupling and  $g_1$  is the U(1) coupling, and the operators  $O_{\ell\ell}$ ,  $O_{\varphi\ell}^{(3)}$ ,  $O_{\varphi D}$ , and  $O_{\varphi WB}$  naturally contributes to said transformation, namely the first two to the Fermi constant via the muon decay and the last three via the spontaneous symmetry breaking, or the Higgs mechanism. We note that the SMEFT shifts to the  $ffV$  vertices scale as  $v^2/\Lambda^2$ , whereas the semi-leptonic four-fermion operators scale as  $s/\Lambda^2$ . Ref. [20] has the

expressions  $G_F$ ,  $m_Z$ , and  $\alpha$  (or actually  $e = \sqrt{4\pi\alpha}$ ) in terms of bare parameters  $g_2$ ,  $g_1$ , and  $v$ , as well as the relevant Wilson coefficients:

$$(2.70) \quad -\frac{4G_F}{\sqrt{2}} = -\frac{2}{v^2} + \frac{2C_{\ell\ell}}{\Lambda^2} - \frac{4C_{\varphi\ell}^{(3)}}{\Lambda^2},$$

$$(2.71) \quad m_Z^2 = \frac{v^2}{4}(g_1^2 + g_2^2) + \frac{1}{8}\frac{v^4}{\Lambda^2}C_{\varphi D}(g_1^2 + g_2^2) + \frac{1}{2}\frac{v^4}{\Lambda^2}g_1g_2C_{\varphi WB},$$

$$(2.72) \quad e = \frac{g_1g_2}{\sqrt{g_1^2 + g_2^2}} \left( 1 - \frac{v^2}{\Lambda^2} \frac{g_1g_2}{g_1^2 + g_2^2} C_{\varphi WB} \right).$$

We can then solve these three equations for the bare parameters  $g_2$ ,  $g_1$ , and  $v$  perturbatively, namely by assuming  $q = q^{(0)} + \epsilon q^{(1)}$ , where  $q$  is any of the bare parameters and  $\epsilon = 1/\Lambda^2$  and then borrow the  $ffV$  vertices in terms of the bare parameters from [85]:

$$(2.73) \quad V_{ff\gamma}^\mu = V_{ee\gamma, \text{SM}}^\mu \left\{ 1 - \frac{g_1g_2v^2C_{\varphi WB}}{(g_1^2 + g_2^2)\Lambda^2} \right\},$$

$$(2.74) \quad \begin{aligned} V_{eeZ}^\mu = & V_{eeZ, \text{SM}}^L \left\{ 1 - \frac{v^2 \left( (g_1^2 + g_2^2)^2 C_{\varphi\ell}^{(1)} + (g_1^2 + g_2^2)^2 C_{\varphi\ell}^{(3)} + g_1(g_1 - g_2)g_2(g_1 + g_2)C_{\varphi WB} \right)}{(g_1^4 - g_2^4)\Lambda^2} \right\} \\ & + V_{eeZ, \text{SM}}^R \left\{ 1 - \frac{v^2 \left( (g_1^2 + g_2^2)^2 C_{\varphi e} - 2g_1g_2^3C_{\varphi WB} \right)}{2g_1^2(g_1^2 + g_2^2)\Lambda^2} \right\}, \end{aligned}$$

$$(2.75) \quad \begin{aligned} V_{uuZ}^\mu = & V_{uuZ, \text{SM}}^L \left\{ 1 + \frac{v^2 \left( 3(g_1^2 + g_2^2)^2 C_{\varphi q}^{(1)} - 3(g_1^2 + g_2^2)^2 C_{\varphi q}^{(3)} + g_1g_2(g_2^2 - 3g_1^2)C_{\varphi WB} \right)}{(g_1^2 - 3g_2^2)(g_1^2 + g_2^2)\Lambda^2} \right\} \\ & + V_{uuZ, \text{SM}}^R \left\{ 1 + \frac{v^2 (4g_1g_2^3C_{\varphi WB} + 3(g_1^2 + g_2^2)^2 C_{\varphi u})}{4g_1^2(g_1^2 + g_2^2)\Lambda^2} \right\}, \end{aligned}$$

$$\begin{aligned}
V_{ddZ}^\mu = & \\
(2.76) \quad & V_{ddZ,\text{SM}}^L \left\{ 1 + \frac{v^2 \left( 3(g_1^2 + g_2^2)^2 C_{\varphi q}^{(1)} + 3(g_1^2 + g_2^2)^2 C_{\varphi q}^{(3)} + g_1 g_2 (3g_1^2 + g_2^2) C_{\varphi WB} \right)}{(g_1^2 + g_2^2)(g_1^2 + 3g_2^2) \Lambda^2} \right\} \\
& + V_{ddZ,\text{SM}}^R \left\{ 1 - \frac{v^2 (3(g_1^2 + g_2^2)^2 C_{\varphi d} - 2g_1 g_2^3 C_{\varphi WB})}{2g_1^2 (g_1^2 + g_2^2) \Lambda^2} \right\},
\end{aligned}$$

where  $V_{ffV,\text{SM}}^\mu$  are the corresponding SM couplings, for which we have considered left and right-handed  $Z$  couplings explicitly separately. Now armed with these modification, we can go back to the SM amplitudes and introduce these SMEFT contributions, which is what we do in Chapter 4, on top of introducing the more familiar semi-leptonic four-fermion operators.

For the DY process accompanied by a single jet, we focus on one particular set of operators, namely ones that can generate  $CP$ -odd observables. The unpolarized SM cross section is  $CP$ -even, and the LIPS is  $CP$ -even. If we consider our favorite class of operators, namely the semi-leptonic four-fermion operators, coupled to the gluonic field strength tensor, we can generate  $CP$ -odd observables via SM-SMEFT interference. One immediate example is the CS moments  $A_5$ ,  $A_6$ , and  $A_7$ . A four-fermion operator coupled to a gluonic field strength tensor means a dimension-8 operator. Emphasizing the helicity conservation in SM-SMEFT interference, we need operators of the form

$$(2.77) \quad O_{XYg} = (\bar{\ell} \gamma^\mu P_X \ell) (\bar{q} \gamma^\nu P_Y T^A q) G_{\mu\nu}^A.$$

Such operators have nonzero contributions to the DY $j$  amplitudes. In the Murphy basis [140], there are seven such operators:

$$(2.78) \quad O_{\ell^2 q^2 g}^{(1)} = (\bar{\ell}\gamma^\mu \ell)(\bar{q}\gamma^\nu T^A q)G_{\mu\nu}^A, \quad O_{\ell^2 q^2 g}^{(3)} = (\bar{\ell}\gamma^\mu \tau^I \ell)(\bar{q}\gamma^\nu T^A \tau^I q)G_{\mu\nu}^A,$$

$$(2.79) \quad O_{\ell^2 u^2 g} = (\bar{\ell}\gamma^\mu \ell)(\bar{u}\gamma^\nu T^A u)G_{\mu\nu}^A, \quad O_{\ell^2 d^2 g} = (\bar{\ell}\gamma^\mu \ell)(\bar{d}\gamma^\nu T^A d)G_{\mu\nu}^A,$$

$$(2.80) \quad O_{e^2 u^2 g} = (\bar{e}\gamma^\mu e)(\bar{u}\gamma^\nu T^A u)G_{\mu\nu}^A, \quad O_{e^2 d^2 g} = (\bar{e}\gamma^\mu e)(\bar{d}\gamma^\nu T^A d)G_{\mu\nu}^A,$$

$$(2.81) \quad O_{q^2 e^2 g} = (\bar{e}\gamma^\mu e)(\bar{q}\gamma^\nu T^A q)G_{\mu\nu}^A,$$

where  $\ell$  and  $q$  are SU(2) left-handed doublets,  $e$ ,  $u$ , and  $d$  are right-handed singlets,  $G_{\mu\nu}^A$  is the gluon field strength tensor, the  $T^A$  are the SU(3) generators, and the  $\tau^I$  are the Pauli matrices. Since we want specifically one jet, we discard the nonabelian term in the gluon field strength. Thus, the vertex factors are of the form

$$(2.82) \quad V_{XYg}^{\mu\nu\rho} = iC_{XY}[\gamma^\mu P_X]_\ell[\gamma^\nu P_Y]_q T_{ab}^A [(-i)(p_g^\mu g^{\nu\rho} - p_g^\nu g^{\mu\rho})],$$

where  $p_g$  is the gluon momentum, assumed outgoing from the vertex, and

$$(2.83) \quad C_{LL}^{u/d} = C_{\ell^2 q^2 g}^{(1)} \mp C_{\ell^2 q^2 g}^{(3)}, \quad C_{LR}^{u/d} = C_{\ell^2 u^2/d^2 g}, \quad C_{RL}^{u/d} = C_{q^2 e^2 g}, \quad C_{RR}^{u/d} = C_{e^2 u^2/d^2 g}.$$

The SMEFT Feynman diagram for the underlying partonic process of the DY $j$  is depicted in Figure 2.8.

The corresponding amplitudes are

$$(2.84) \quad \mathcal{A}_{ij} = \sum_{XY} C_{XY}[\bar{u}_2 \gamma^\mu P_X v_3][\bar{q}_j \gamma^\nu P_Y q_i] T_{ab}^A (p_g^\mu g^{\nu\rho} - p_g^\nu g^{\mu\rho}) \varepsilon_g^\rho.$$



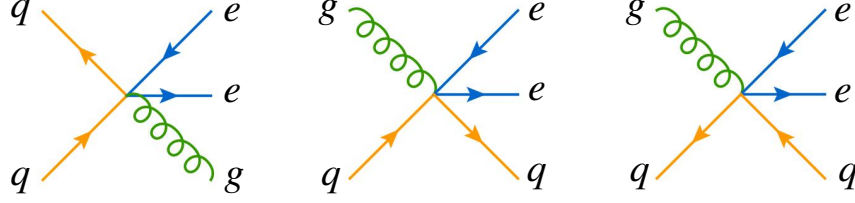


Figure 2.8. The SMEFT Feynman diagram for the underlying partonic process of the Drell-Yan production with a jet, generated by the semi-leptonic four-fermion operators coupled to a gluon field strength tensor.

Here,  $i, j = -2, -1, 0, 1, 2$  are the parton flavors  $\bar{u}, \bar{d}, g, d$ , and  $u$ . Technically, we go from  $-N_f$  to  $N_f$  but even if we consider multiple generations, going from  $-2$  to  $2$  is sufficient because we consider only neutral current interactions so flavor is conserved, and we assume massless quarks, so up and charm amplitudes, as well as down, strange, and bottom amplitudes are equivalent. Meanwhile,  $q_i$  and  $\bar{q}_j$  are quark spinors depending on the process,  $p_g$  is the gluon momentum, and  $\varepsilon_g^\rho$  is the gluon polarization. We have

$$(2.85) \quad ij = 2 - 2, 1 - 1 : q_i = u_a, \bar{q}_j = \bar{v}_b, p_g = p_1, \varepsilon_g^\rho = \varepsilon_1^{*\rho},$$

$$(2.86) \quad ij = 20, 10 : q_i = u_a, \bar{q}_j = \bar{u}_1, p_g = -p_b, \varepsilon_g^\rho = \varepsilon_b^\rho,$$

$$(2.87) \quad ij = -20, -10 : q_i = v_1, \bar{q}_j = \bar{v}_a, p_g = -p_b, \varepsilon_g^\rho = \varepsilon_b^\rho,$$

$$(2.88) \quad ij = -22, -11 : q_i = u_b, \bar{q}_j = \bar{v}_a, p_g = p_1, \varepsilon_g^\rho = \varepsilon_1^{*\rho},$$

$$(2.89) \quad ij = 02, 01 : q_i = u_b, \bar{q}_j = \bar{u}_1, p_g = -p_a, \varepsilon_g^\rho = \varepsilon_a^\rho,$$

$$(2.90) \quad ij = 0 - 2, 0 - 1 : q_i = v_1, \bar{q}_j = \bar{v}_b, p_g = -p_a, \varepsilon_g^\rho = \varepsilon_a^\rho.$$

We add these amplitudes to the corresponding SM amplitudes given in Eqs. (2.23)–(2.28).

### 2.3. Particle momenta and many-particle Lorentz-invariant phase space

Consider a two-body decay in the form  $a \rightarrow 1 + 2$ . In the rest frame of the decaying particle, we know the expressions for the particle 4-momenta, which derives from a cute exercise of taking dot products in conjunction with the conservation of momentum,  $p_a = p_1 + p_2$ , and the invariant masses,  $p_a^2 = m_a^2$ ,  $p_1^2 = m_1^2$ , and  $p_2^2 = m_2^2$ . If we write  $p = (E, |\mathbf{p}| \hat{\mathbf{p}})$ , where we use bold letters to denote the usual 3-momentum, we have

$$(2.91) \quad E_a^\bullet = m_a, \quad \mathbf{p}_a^\bullet = (0, 0, 0),$$

$$(2.92) \quad E_{1/2}^\bullet = \frac{m_a^2 + m_{1/2}^2 - m_{2/1}^2}{2m_a}, \quad |\mathbf{p}_{1/2}^\bullet| = \frac{\nabla(m_a^2, m_1^2, m_2^2)}{2m_a},$$

$$(2.93) \quad \hat{\mathbf{p}}_{1/2}^\bullet = \pm \hat{\mathbf{n}}_1^\bullet.$$

Here, the superscript bullet indicates that this is the rest frame of the mother particle, and we have defined

$$(2.94) \quad \nabla(x, y, z) = \sqrt{\lambda(x, y, z)}$$

as the square root of the Källén function,

$$(2.95) \quad \lambda(x, y, z) = x^2 + y^2 + z^2 - 2(xy + yz + zx),$$

and  $\hat{\mathbf{n}}_1$  is the spherical radial unit vector (with an arbitrary choice for the  $z^\bullet$  direction) defined by the spherical angles of the leading particle 1,

$$(2.96) \quad \hat{\mathbf{n}}_1 = (s_{\theta_1} c_{\varphi_1}, s_{\theta_1} s_{\varphi_1}, c_{\theta_1}),$$

where  $s$  and  $c$  are the sine and cosine functions of the angles indicated in the subscript, respectively. The quantity  $m_a$ , which is the mass of particle  $a$ , is the available energy for this decay. The geometry of the decay is illustrated in Figure 2.9.

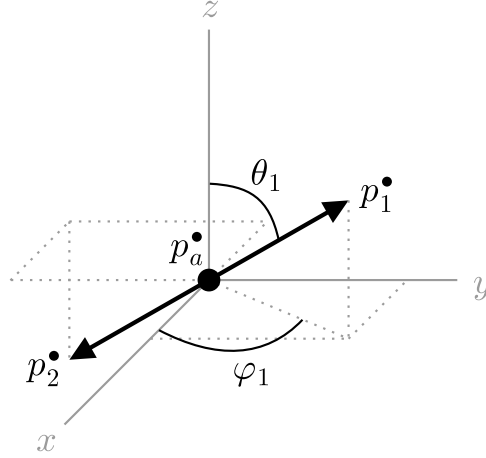


Figure 2.9. The geometry of the 2-body decay in the rest frame of the mother particle.

Next, consider a  $2 \rightarrow 2$  scatter in the form  $a + b \rightarrow 1 + 2$ . In the c.m. frame of the  $ab$  system, we also know the expressions for the particle 4-momenta, which satisfy  $p_a + p_b = p_1 + p_2$  and  $p_i^2 = m_i^2$  for  $i = a, b, 1, 2$ . The momentum components are given by

$$(2.97) \quad E_{a/b}^\bullet = \frac{\mathcal{E}^2 + m_{a/b}^2 - m_{b/a}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{a/b}^\bullet| = \frac{\nabla(\mathcal{E}^2, m_a^2, m_b^2)}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{a/b} = \pm(0, 0, 1),$$

$$(2.98) \quad E_{1/2}^\bullet = \frac{\mathcal{E}^2 + m_{1/2}^2 - m_{2/1}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{1/2}^\bullet| = \frac{\nabla(\mathcal{E}^2, m_1^2, m_2^2)}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{1/2} = \pm\hat{\mathbf{n}}_1^\bullet.$$

Here, the superscript bullet now indicates the c.m. frame of the incoming particle, and we note that by assuming the incoming particles to move along the  $z$  axis, we explicitly assume that this is the first of a chain of processes, upon which we elaborate shortly in this section. The quantity  $\mathcal{E}$ , which is the invariant mass of the  $ab$  system, is the available

energy for the final states, namely  $(p_a + p_b)^2 = \mathcal{E}^2$ . Clearly, it is the Mandelstam- $s$  parameter of this  $2 \rightarrow 2$  process. The geometry of the scattering is depicted in Figure 2.10.

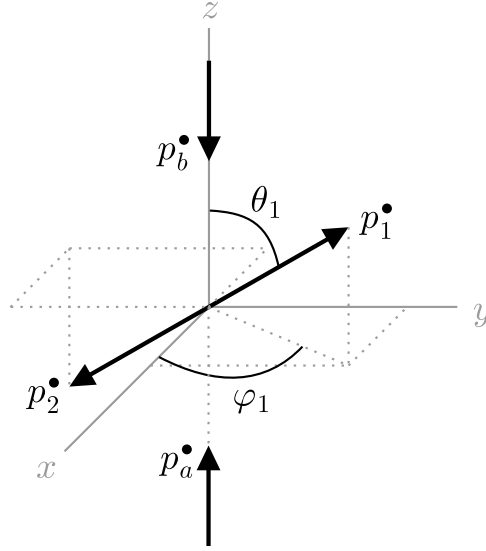


Figure 2.10. The geometry of the 2-body decay in the center-of-mass frame of the incoming particles.

Let's discuss what happens if we have a process of the form  $a + b \rightarrow 1 + 2 + 3$  or  $a + b \rightarrow 1 + 2 + 3 + 4$ . To illustrate the notation, we can give the processes  $p + p \rightarrow j + e^- + e^+$  and  $e^- + e^+ \rightarrow \ell^- + \bar{\nu} + j + j$  as examples. The former is a DY production with a real jet emission accompanying the EW gauge boson  $V$ , which later decays into the electron-positron pair. The latter is more intricate in the sense that we now have a lot more possibilities; thus, we simplify the picture by focusing on an electron-positron pair colliding to produce two  $W$  bosons, one of which decays into the lepton pair and the other into two jets, which is one of the processes we study later in this thesis. A fine trick to study these many-particle processes is to break them into known pieces, e.g. a  $2 \rightarrow 2$

scattering or a two-body decay, in a chain of subprocesses. Thus, the first example can actually be rewritten as  $p + p \rightarrow j + V$  followed by  $V \rightarrow e^- + e^+$ , and the second example as  $e^- + e^+ \rightarrow W^- + W^+$  followed by  $W^- \rightarrow \ell^- + \bar{\nu}$  and  $W^+ \rightarrow j + j$ , which we precisely know how to tackle.

This brings us to the next topic, namely the many-particle LIPS. For a process with two particles in the final state, e.g.  $a \rightarrow 1 + 2$  and  $a + b \rightarrow 1 + 2$ , the formula is well known:

$$(2.99) \quad \int d\text{LIPS}_2 = \frac{\nabla(P^2, m_1^2, m_2^2)}{32\pi^2 P^2} \int d\Omega_1^\bullet,$$

where  $P$  is the total incoming momentum in the relevant c.m. frame, i.e.  $P = p_a^\bullet$  for the decay (so  $P^2 = m_a^2$ ) or  $P = p_a^\bullet + p_b^\bullet$  for the scattering (so  $P^2 = \mathcal{E}^2$ ), and  $\Omega_1^\bullet = (\theta_1, \varphi_1)$  are the spherical angles of the leading particle in the said c.m. frame. Now we want a formula for the many-particle LIPS, which we build up from scratch. To this end, we can consider either a many-particle decay or a scattering with many particles in the final states, which are equivalent for our purposes if we simply express the processes in terms of particle momenta as  $P \rightarrow p_1 + \dots + p_n$ . Each outgoing particle is massive with mass  $m_i$ . We have

$$(2.100) \quad \int d\text{LIPS}_n = \int \left\{ \prod_{i=1}^n \frac{d^4 p_i}{(2\pi)^4} (2\pi) \delta(p_i^2 - m_i^2) \theta(p_i^0) \right\} (2\pi)^4 \delta^4 \left( P - \sum_{i=1}^n p_i \right),$$

where  $\delta$  is the Dirac delta function and  $\theta$  is the Heaviside step function. Now consider the process  $P \rightarrow p_1 + p_2 + p_3$ , so  $n = 3$  explicitly, and break it into  $P \rightarrow p_1 + p_{23}$  followed by  $p_{23} \rightarrow p_2 + p_3$ . Let  $m_{23}$  be the invariant mass corresponding to the momentum  $p_{23}$ , i.e.

$p_{23}^2 = m_{23}^2$ . We have

$$(2.101) \quad \int d\text{LIPS}_3 = \frac{1}{(2\pi)^5} \int d^4 p_1 d^4 p_2 d^4 p_3 \delta(p_1^2 - m_1^2) \theta(p_1^0) \\ \times \delta(p_2^2 - m_2^2) \theta(p_2^0) \delta(p_3^2 - m_1^2) \theta(p_3^0) \delta^4(P - p_1 - p_2 - p_3).$$

Resolve an identity of the form

$$(2.102) \quad 1 = \int d^4 Q dM^2 \delta(Q^2 - M^2) \delta^4\left(Q - \sum_k p_k\right),$$

where  $Q$  is a momentum equal to the sum of some other momenta,  $\sum_k p_k$ , and  $M$  is the invariant mass corresponding to  $Q$ . With  $Q = p_{23}$  and  $M = m_{23}$ , we obtain

$$(2.103) \quad \int d\text{LIPS}_3 = \frac{1}{(2\pi)^5} \int d^4 p_1 d^4 p_2 d^4 p_3 d^4 p_{23} dm_{23}^2 \delta(p_1^2 - m_1^2) \theta(p_1^0) \\ \times \delta(p_2^2 - m_2^2) \theta(p_2^0) \delta(p_3^2 - m_1^2) \theta(p_3^0) \delta^4(P - p_1 - p_2 - p_3) \\ \times \delta(p_{23}^2 - m_{23}^2) \delta^4(p_{23} - p_2 - p_3).$$

If we rearrange the terms, we get

$$(2.104) \quad \int d\text{LIPS}_3 = \frac{1}{(2\pi)^5} \int dm_{23}^2 \\ \times \int d^4 p_1 d^4 p_{23} \delta(p_1^2 - m_1^2) \theta(p_1^0) \delta(p_{23}^2 - m_{23}^2) \delta^4(P - p_1 - p_{23}) \\ \times \int d^4 p_2 d^4 p_3 \delta(p_2^2 - m_2^2) \theta(p_2^0) \delta(p_3^2 - m_3^2) \theta(p_3^0) \delta^4(p_{23} - p_2 - p_3),$$

which is nothing but the product of two 2-particle LIPSs convoluted over the invariant mass of the particle that lives in the intermediate state at which we split the process,

$$\begin{aligned}
 \int d\text{LIPS}_3[P \rightarrow p_1 + p_2 + p_3] &= \int dm_{23}^2 \\
 &\times \int d\text{LIPS}_2[P \rightarrow p_1 + p_{23}] \\
 &\times \int d\text{LIPS}_2[p_{23} \rightarrow p_2 + p_3].
 \end{aligned}
 \tag{2.105}$$

If we repeat the calculation for the 4-particle final state,  $P \rightarrow p_1 + p_2 + p_3 + p_4$ , by first breaking it into  $P \rightarrow p_{12} + p_{34}$  followed by  $p_{12} \rightarrow p_1 + p_2$  and  $p_{34} \rightarrow p_3 + p_4$ , we find that we need two invariant-mass integrals, corresponding to the two split points of the process:

$$\begin{aligned}
 \int d\text{LIPS}_4[P \rightarrow p_1 + p_2 + p_3 + p_4] &= \int dm_{12}^2 \int dm_{34}^2 \\
 &\times \int d\text{LIPS}_2[P \rightarrow p_{12} + p_{34}] \\
 &\times \int d\text{LIPS}_2[p_{12} \rightarrow p_1 + p_2] \\
 &\times \int d\text{LIPS}_2[p_{34} \rightarrow p_3 + p_4].
 \end{aligned}
 \tag{2.106}$$

This is a robust method of building the  $n$ -particle LIPS with a sleight of hand. Furthermore, since each 2-particle LIPS is Lorentz-invariant on its own, we are at liberty to evaluate each in any frame for which our hearts seek. As it turns out, the most convenient frame is the c.m. frame of the total incoming momentum that provides the available energy for the process. To illustrate, in (2.106), for the first subprocess, we go to a frame where  $\mathbf{P} = 0$ , and  $\mathbf{p}_{12} = \mathbf{0}$  for the next one, and  $\mathbf{p}_{34} = \mathbf{0}$  for the last one. We already

know the expression for the 2-particle LIPS. Each such LIPS goes like

$$(2.107) \quad \int d\text{LIPS}_2[p \rightarrow k + q] = \frac{\nabla(p^2, k^2, q^2)}{8p^2} \int d\Omega_k^\bullet,$$

where the factor 8 in the denominator is merely for convenience, and the rest is just figuring out the factors of  $2\pi$  in the total phase space. After careful power counting, we find this factor to be equal to  $1/(2\pi)^{3n-4}$  for the  $n$ -particle LIPS. Namely, for the 3-particle LIPS for the process  $P \rightarrow p_1 + p_{23} \rightarrow p_1 + p_2 + p_3$ , we have

$$(2.108) \quad \begin{aligned} \int d\text{LIPS}_3 = & \left[ \frac{1}{(2\pi)^{3n-4}} \right]_{n=3} \int dm_{23}^2 \\ & \times \left[ \frac{\nabla(P^2, m_1^2, m_{23}^2)}{8P^2} \int d\Omega_1^\bullet \right] \left[ \frac{\nabla(m_{23}^2, m_2^2, m_3^2)}{8m_{23}^2} \int d\Omega_2^\circ \right], \end{aligned}$$

where the solid angles are the spherical angles of the leading particles for each subprocess and the superscripts bullet and circle denote the c.m. frame where  $\mathbf{P} = \mathbf{0}$  and  $\mathbf{p}_{23} = \mathbf{0}$ , respectively. As for the 4-particle LIPS for the process  $P \rightarrow p_{12} + p_{34} \rightarrow p_1 + p_2 + p_3 + p_4$ , we have

$$(2.109) \quad \begin{aligned} \int d\text{LIPS}_4 = & \left[ \frac{1}{(2\pi)^{3n-4}} \right]_{n=4} \int dm_{12}^2 dm_{34}^2 \left[ \frac{\nabla(P^2, m_{12}^2, m_{34}^2)}{8P^2} \int d\Omega_{12}^\bullet \right] \\ & \times \left[ \frac{\nabla(m_{12}^2, m_1^2, m_2^2)}{8m_{12}^2} \int d\Omega_1^\circ \right] \left[ \frac{\nabla(m_{34}^2, m_3^2, m_4^2)}{8m_{34}^2} \int d\Omega_3^* \right], \end{aligned}$$

where this time the superscripts bullet, circle, and asterisk indicate the c.m. frame where  $\mathbf{P} = \mathbf{0}$ ,  $\mathbf{p}_{12} = \mathbf{0}$ , and  $\mathbf{p}_{34} = \mathbf{0}$ , respectively. The integration limits for the solid angles are as usual, and the bounds of the invariant-mass integrals are obtained by requiring the



nabla functions to be real or equivalently the Källén functions to be positive. This can be done for instance by using the `Reduce` command on Mathematica.

We have seen the power of splitting the process at convenient and perhaps physically meaningful points to facilitate the phase-space calculations. One important take-home message is that we can work in any frame for each subprocess to write down the momentum. This frame is usually the rest frame of the mother if it is a decay or the c.m. frame of the initial states if it is a  $2 \rightarrow 2$  scattering. Once we start splitting our process, we actually generate subprocesses that are nothing but 2-particle decays. For the final-state particles of these processes, the explicit momenta are always given as discussed earlier: the mother is at rest, the leading particle flies off in a direction defined by the familiar spherical angles, and the subleading particles takes off in the opposite direction. We note that since we are breaking our processes at intermediate states, the mother is always massive with the corresponding invariant mass, not the on-shell mass. We also note that whilst we can make simplifying assumptions such as letting the jet and the gauge boson be emitted in the  $xz$  plane of the lab frame for the aforementioned  $DYj$  process so as to eliminate one azimuthal angle (so its integration just contributes a factor of  $2\pi$ ); however, both states are intermediate, as the two  $W$  particles in the above-mentioned example of the electron-positron collision, we cannot make this simplification.

Now we know how to write down the LIPS for any process and the particle momenta explicitly. However, we have mentioned only the relevant c.m. frame for each subprocess. We still need to bring all the momenta into one common frame, which is usually the lab frame, i.e. the c.m. frame of the two protons in the said  $DYj$  process or the c.m. frame of the electron-positron pair in the aforementioned example. The transformation takes

place using a Lorentz matrix,  $\Lambda$ . Given the form of the momenta in the c.m. frame of any given subprocess, we perform the Lorentz transformation as a rotation around some  $z$  axis (to eliminate the azimuthal angle) followed by a rotation around the  $y$  axis of the frame under consideration (to get rid of the polar angle so now we have a one-dimensional motion) followed by a boost. This transformation is carried out with respect to the intermediate particle at which we split the process. We illustrate this yet another robust method of handling momenta later in the thesis, specifically designed for the process under consideration.

Once we have the all momenta boosted into a single frame, there are certain sanity tests that we can perform to verify our results. For instance, for the process  $a + b \rightarrow 1 + \dots + n$ , we should be able to verify

$$(2.110) \quad \int d\text{LIPS}_n (p_1 + p_2)^2 = \int d\text{LIPS}_n (p_1 + p_3)^2 = \dots = \int d\text{LIPS}_n (p_{n-1} + p_n)^2,$$

and

$$(2.111) \quad \begin{aligned} \int d\text{LIPS}_n p_a \cdot p_1 &= \dots = \int d\text{LIPS}_n p_a \cdot p_n \\ &= \int d\text{LIPS}_n p_b \cdot p_1 = \dots = \int d\text{LIPS}_n p_b \cdot p_n, \end{aligned}$$

when the initial-state particles have the same mass and the final-state particles have the same mass. We call these the *symmetry integrals*, where the symmetry refers to the permutation symmetry of the final states. The LIPS becomes essentially particle-blind if all the final state particles have the same mass. If the initial-state particles have different

masses, then the second relation breaks into two, namely the  $p_a$  integrals are equal to one number and the  $p_b$  integrals are equal to another.

## 2.4. Statistical analysis

Statistical analysis is, without exaggeration, half of the game we play here, and a basic  $\chi^2$  test statistics constitutes the backbone of our calculations. Let's take our favorite observable,  $\mathcal{Q}$ , and consider a set of data points with  $N_B$  bins,  $\{\mathcal{Q}_b\}_{b=1}^{N_B}$ , each with a set of uncertainties,  $\{\delta\mathcal{Q}_{1b}, \delta\mathcal{Q}_{2b}, \dots\}_{b=1}^{N_B}$ . These may be statistical uncertainties, systematical uncertainties, uncertainties deriving from higher-order corrections, PDF uncertainties, and renormalization and factorization scale uncertainties. Some days, all or only a subset of these are correlated across bins. We discuss how to build the uncertainty matrix in each study in the subsequent chapters of the thesis, whenever relevant, in great detail. This section aims to guide the user what to do next once they have in their hands the uncertainty matrix, denoted  $\mathcal{E}$ .

We need a fit or model function to make predictions. At this point, we use the SMEFT version of our observable, which looks like

$$(2.112) \quad \hat{\mathcal{Q}}_b = \mathcal{Q}_{\text{SM},b} + \sum_{w=1}^{N_W} C_w \mathcal{Q}_{wb}$$

after linearization, where  $\mathcal{Q}_{\text{SM},b}$  is the SM value at the  $b^{\text{th}}$  bin,  $\mathcal{Q}_{wb}$  is the SMEFT correction characterized by the Wilson coefficient  $C_w$ , and we assume  $N_W$  Wilson coefficients activated. This is actually the case in our works involving SMEFT because we activate either dimension-6 or dimension-8 operators alone and in the squared amplitude, we take into account only the SM squared and SM-SMEFT interference terms. Thus, we have a

linear fit model. With this, we can write the  $\chi^2$  test statistic as

$$(2.113) \quad \chi^2 = \sum_{b=1}^{N_B} \sum_{b'=1}^{N_B} (\hat{\mathcal{Q}}_b - \mathcal{Q}_b) \mathcal{H}_{bb'} (\hat{\mathcal{Q}}_{b'} - \mathcal{Q}_{b'}),$$

where  $\mathcal{H} = \mathcal{E}^{-1}$  is the inverse of the uncertainty matrix. At this point, one has every right to question what the data points we use are since the colliders of interest live in the future. We use made-up data, which we technically refer to as pseudodata (but we continue calling them data because that's all we've got). We generate pseudodata by smearing the SM predictions with the experimental uncertainties predicted by people who know better:

$$(2.114) \quad \mathcal{Q}_b = \mathcal{Q}_{\text{SM},b} + r_b \delta \mathcal{Q}_b^{\text{uncorr}} + \sum_j r'_j \delta \mathcal{Q}_b^{\text{corr},j},$$

where  $\delta \mathcal{Q}_b^{\text{uncorr}}$  is the total uncorrelated experimental uncertainty at the  $b^{\text{th}}$  bin added in quadrature,  $\delta \mathcal{Q}_b^{\text{corr},j}$  is the  $j^{\text{th}}$  correlated uncertainty, and  $r_b$  and  $r'_j$  are unit normal variates,  $r_b, r'_j \sim \mathcal{N}(0, 1)$ . We introduce the correlated uncertainties with a different unit normal variate, which is fixed across the bins, to ensure that all the bins feel the shift uniformly.

Since our fit model is a linear function of Wilson coefficients, the  $\chi^2$  function is a quadratic function of Wilson coefficients. We may as well write

$$(2.115) \quad \chi^2 = k_0 + \sum_{w=1}^{N_W} k_{1w} C_w + \sum_{w=1}^{N_W} \sum_{w'=1}^{N_W} k_{2ww'} C_w C_{w'} = k_0 + \mathbf{k}_1 \cdot \mathbf{C} + \mathbf{C} \cdot k_2 \mathbf{C}.$$

We now minimize this expression with respect to the Wilson coefficients. The vanishing gradient of the expression gives us the values of the Wilson coefficients that minimize the

$\chi^2$  function,

$$(2.116) \quad \nabla \chi^2(\overline{\mathbf{C}}) = \mathbf{0},$$

and the Hessian evaluated at these values give us the Fisher information matrix [136, 13, 92],

$$(2.117) \quad \mathcal{F} = \frac{1}{2} \nabla \nabla \chi^2(\overline{\mathbf{C}}).$$

Here, all the derivatives are with respect to the variables of the parameter space, namely the Wilson coefficients. With the quadratic form of the  $\chi^2$  function presented above, we have

$$(2.118) \quad \overline{\mathbf{C}} = \frac{1}{2} k_2^{-1} \mathbf{k}_1,$$

and

$$(2.119) \quad \mathcal{F} = k_2.$$

Once we have the Fisher information matrix, we can do anything. We can obtain non-marginalized bounds for Wilson coefficients around some central value with a desired confidence level (CL), which is equivalent to the best-fit analysis of the model function as if we turn on one Wilson coefficient at a time, or confidence ellipses at a desired CL, which is equivalent to the best-fit analysis with just two Wilson coefficients activated at a time, or marginalized bounds and marginalized confidence ellipses when a larger set of or all the Wilson coefficients of interest activated at a time.

Before we proceed, there is one more concern we need to address. Our statistical analysis strongly relies on pseudodata. If we generate one set of pseudodata, namely if we perform just one pseudoexperiment, we might as well obtain best-fitted  $\overline{\mathcal{C}}$  values away from the origin. However, we know that the best-fit values of the Wilson coefficients must be centered at zero. We are sure because our pseudodata generation procedure deploys unit normal variates, which are just random numbers picked from a Gaussian distribution centered at the origin. Therefore, in order to ensure that our fits are reliable, we need to create statistics. This goes on to say that we repeat a finite number of pseudoexperiments, denoted  $N_E$ . Our best-fitted values for Wilson coefficients and our Fisher information matrix are meaningful only when we average over the pseudoexperiments. Let's put a subscript  $e$  to indicate the  $e^{\text{th}}$  pseudoexperiment. Then, our previous notation for the best-fit values and the Fisher information matrix evolve into  $\overline{\mathcal{C}}_e$  and  $\mathcal{F}_e$ , respectively. The best-fit values of the Wilson coefficients averaged over pseudoexperiments is given by

$$(2.120) \quad \overline{\mathcal{C}} = \left( \sum_{e=1}^{N_E} \mathcal{F}_e \right)^{-1} \sum_{e=1}^{N_E} \mathcal{F}_e \overline{\mathcal{C}}_e,$$

and the Fisher information matrix averaged over the pseudoexperiments is given by

$$(2.121) \quad \mathcal{F} = \frac{1}{N_E} \sum_{e=1}^{N_E} \mathcal{F}_e.$$

Now we claim that we do not even need to run pseudoexperiments and we prove it. The disclaimer is that what follows works only when the model function is a multilinear function of Wilson coefficients so the  $\chi^2$  function is quadratic in Wilson coefficients; for instance, it works when we have only dimension-6 SMEFT operators or only dimension-8

SMEFT operators and when their effects are included in the squared amplitude only via a SM-SMEFT interference term. Let's show how. In the  $\chi^2$  function, we have the difference between the observed value, which is the SM entry plus uncertainties added after scaled by random numbers pick from unit normal distribution, and the predicted value, which is just the SM value plus the SMEFT corrections each characterized by a particular Wilson coefficient. The SM parts cancel out and we are left with

$$(2.122) \quad \chi^2 = (C_w \mathcal{Q}_{wb} - r_b \delta \mathcal{Q}_b^{\text{uncorr}} - r' \delta \mathcal{Q}_b^{\text{corr}}) \mathcal{H}_{bb'} (C_{w'} \mathcal{Q}_{w'b'} - r_{b'} \delta \mathcal{Q}_{b'}^{\text{uncorr}} - r' \delta \mathcal{Q}_{b'}^{\text{corr}}),$$

where we assume summation over  $b$ ,  $b'$ ,  $w$ , and  $w'$ . If we rewrite this as

$$(2.123) \quad \chi^2 = k_0 + \mathbf{k}_1 \cdot \mathbf{C} + \mathbf{C} \cdot k_2 \mathbf{C},$$

then we see that

$$(2.124) \quad k_{1w} = -\mathcal{H}_{bb'} (r' \mathcal{Q}_{wb} \delta \mathcal{Q}_{b'}^{\text{corr}} + r' \mathcal{Q}_{wb'} \delta \mathcal{Q}_b^{\text{corr}} + r_b \mathcal{Q}_{wb'} \delta \mathcal{Q}_b^{\text{uncorr}} + r_{b'} \mathcal{Q}_{wb} \delta \mathcal{Q}_{b'}^{\text{uncorr}}),$$

and

$$(2.125) \quad k_{2ww'} = \mathcal{H}_{bb'} \mathcal{Q}_{wb} \mathcal{Q}_{w'b'}.$$

Thus, whilst the best-fit values of Wilson coefficients certainly depend on pseudoexperiments (because they depend on  $\mathbf{k}_1$  and  $\mathbf{k}_1$  depends on the random variates), the Fisher information matrix depends only on the uncertainty matrix and the SMEFT corrections,

which is independent of random variates and hence of pseudoexperiments. With  $\mathcal{F}$  being constant across pseudoexperiments, the formula for the averaged best-fitted values of Wilson coefficients simplifies to

$$(2.126) \quad \overline{\mathcal{C}} = \frac{1}{N_E} \sum_{e=1}^{N_E} \overline{\mathcal{C}}_e.$$

Now we can hypothetically carry out infinitely many pseudoexperiments to ensure that the Wilson coefficients are best-fitted around zero.

The take-home message is that if the fit model is linear in Wilson coefficients, then we don't really need to run pseudoexperiments to create statistics. This may seem like a trivial statement but it is *au contraire*. Experience shows that for a single Wilson coefficient, running 1k pseudoexperiments guarantees that  $\{\overline{\mathcal{C}}_e\}_{e=1}^{N_{\text{exp}}}$  is distributed normally around the origin beautifully by visual inspection. When we activate two Wilson coefficients, this number easily becomes 50k, and for six Wilson coefficients, we quickly reach 10M pseudoexperiments to make sure the results are sensible. This is nothing but an unfortunate waste of computational resources and time. Occasionally, we introduce other parameters in an attempt to improve the bounds on SMEFT parameters, and we might have to do it in a nonlinear manner. In this unfortunate occasion, one has to run a couple pseudoexperiments for statistically sensible results.

We are now at a point where we have  $\mathcal{F}$  in our hands and we want to obtain bounds and draw confidence ellipses. First, we emphasize that the Fisher matrix is always symmetric and positive definite. Furthermore, the inverse of the Fisher matrix gives the symmetric



covariance matrix,  $\mathcal{V}$ , which looks like

$$(2.127) \quad \mathcal{V} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \cdots & \rho_{1N_W}\sigma_1\sigma_{N_W} \\ & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \cdots & \rho_{2N_W}\sigma_2\sigma_{N_W} \\ & & \ddots & \vdots & \vdots \\ & & & \ddots & \vdots \\ & & & & \sigma_{N_W}^2 \end{pmatrix}_{\text{sym}},$$

where  $\sigma_w$  is correlated (or marginalized) 1-sigma, or about 68% CL, uncertainty of the Wilson coefficient  $C_w$  and  $\rho_{ww'}$  is the correlation of  $C_w$  and  $C_{w'}$ . The correlation matrix is obtained by

$$(2.128) \quad \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1N_W} \\ & 1 & \rho_{23} & \cdots & \rho_{2N_W} \\ & & 1 & \cdots & \vdots \\ & & & \ddots & \vdots \\ & & & & 1 \end{pmatrix}_{\text{sym}} = \text{diag} \left( \frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_{N_W}} \right) \mathcal{V} \text{diag} \left( \frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_{N_W}} \right)$$

Furthermore, if we are combining more than one independent experiments or data sets, we add up the individual Fisher matrices and then take the inverse to obtain the correlations and the uncertainties. This is to say, if we combine data from distinct data sets or independent experiments, the factor  $1/N_E$  in Eq. (2.121) drops. The reason is that pseudoexperiments correspond to one set of run parameter or configuration so they are technically not independent and that pseudoexperiments are required only to generate statistics (so they should not accidentally improve the allowed bounds). Now, we discuss

how to actually obtain bounds and confidence ellipses. The nonmarginalized interval of Wilson coefficient  $C_w$  is given by  $[-x, x]$ , where  $x$  is the positive solution of the equation

$$(2.129) \quad (x)\mathcal{F}_{ww}(x) = \Delta\chi^2(1, c),$$

where  $\Delta\chi^2(p, c)$  is the scaling factor of the bounds depending on the number of fitted parameters  $p$  and the CL  $c$ , which can be numerically computed using MATHEMATICA by `Quantile[ChiSquareDistribution[p], c]`, where `p` is an integer and `c` is between 0 and 1, which we usually like to take 95% or 0.95. The reason why we write this equation in this form shall become apparent shortly. On the other hand, the marginalized interval is, in general, larger because it includes the correlation effects among all activated Wilson coefficients (or essentially, if the data size is kept fixed but the set of parameters is extended, then we lose information about each parameter). The marginalized interval is given by  $[-x, x]$ , where  $x$  is the positive solution of the equation

$$(2.130) \quad (x)(\mathcal{V}_{ww})^{-1}(x) = \Delta\chi^2(1, c).$$

The fundamental difference between the two intervals is that in the first, we take the  $(w, w)$  entry of the Fisher matrix before inversion, which immediately eliminates all other variables as if they were not activated in the first place, which in turn avoids the correlation effects. The situation is similar in confidence ellipses, the only difference being that now we have one more variable. Suppose we just active two Wilson coefficients at a time,  $C_w$

and  $C_{w'}$ . Then, the confidence ellipse at a CL  $c$  is given by equation

$$(2.131) \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \mathcal{F}_{ww} & \mathcal{F}_{ww'} \\ \mathcal{F}_{w'w} & \mathcal{F}_{w'w'} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \Delta\chi^2(2, c).$$

This is the non-marginalized ellipse, which completely ignores the correlation effects from other Wilson coefficients. The marginalized ellipse when a larger subset of or all the Wilson coefficients are activated is given by

$$(2.132) \quad \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \mathcal{V}_{ww} & \mathcal{V}_{ww'} \\ \mathcal{V}_{w'w} & \mathcal{V}_{w'w'} \end{pmatrix}^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \Delta\chi^2(2, c).$$

We present a minimal working example (MWE), or rather a hypothetical case study at the end of the next section.

## 2.5. Computational frameworks

We outline the analytical and numerical tools used in this thesis. Our main tool for analytical and numerical computations, statistical analysis, and data visualization is Mathematica, as long as the observable does not require full phase-space integration. When the observable necessitates inclusive phase-space integration especially for more crowded final states as in  $2 \rightarrow 3$  and  $2 \rightarrow 4$  topologies, we deploy Monte-Carlo methods designed for this purpose on Python, or some days Fortran if we are in the mood or believe it might run faster. We use FeynArts [111] and FeynCalc [146, 145, 144, 138] to compute traditional amplitudes and Vegas [133, 132] for numerical integration. For hadronic processes, we use ManeParse [69] on Mathematica or LHAPDF [59] on Python and Fortran for PDFs.

### 2.5.1. Amplitudes with FeynArts and FeynCalc

We obtain the SM amplitudes using FeynCalc in conjunction with FeynArts because it is already a closed ecosystem that works just fine, especially when we have a significant amount of amplitudes that we want to generate systematically with the least possible number of keystrokes.

Some of the key ideas are as follows. In the standard model files (pun intended), all the couplings are given in terms of electric charge, particle masses, and weak mixing angle explicitly, which is not immediately helpful. We parametrize the vertex factors as  $V = iCTS$ , where  $C$  is the coupling strength,  $\Gamma$  is some Dirac matrix, and  $S$  is a momentum structure tensor. We let  $\Gamma$  and  $S$  ride along because we want them explicitly to get traced or to join Lorentz contractions. We do not substitute expressions for  $C$ s because it creates larger files already and because we want to keep track of helicity structures for diagnosis, for instance using the  $ffZ$  couplings, defined as  $C_{ffZL}$  and  $C_{ffZR}$ , where  $f$  is some fermion. To illustrate, in the array `M$Couplings` in our custom `.mod` file, we define the interactions  $ee\gamma$ ,  $uuZ$ ,  $udW$ ,  $eeh$ ,  $hWW$ , and  $\gamma WW$  as follows:

```
(* eeA *)
C[-F[2,{j1}],F[2,{j2}],V[1]] ==
  I CeeA IndexDelta[j1,j2] {{1, 0}, {1, 0}},
(* uuZ *)
C[-F[3,{j1,o1}],F[3,{j2,o2}],V[2]] ==
  I IndexDelta[j1,j2] IndexDelta[o1,o2] {{CuuZL,0},{CuuZR,0}},
(* udW, CKM = Identity *)
```

```

C[-F[3,{j1,o1}],F[4,{j2,o2}],-V[3]] ==
  I CffW IndexDelta[j1,j2] IndexDelta[o1,o2] {{1,0},{0,0}},
C[F[3,{j1,o1}],-F[4,{j2,o2}],V[3]] ==
  I CffW IndexDelta[j1,j2] IndexDelta[o1,o2] {{1,0},{0,0}},
(* eeh *)
C[-F[2,{j1}],F[2,{j2}],S[1]] ==
  I Ceeh IndexDelta[j1,j2] {{1,0},{1,0}},
(* hWW *)
C[S[1], -V[3], V[3]] ==
  I ChWW {{1,0}},
(* AWW *)
C[V[1],-V[3],V[3]] ==
  I CAWW {{1,0}}

```

Once we modify the model file following this vision, we generate the amplitudes and then perform another set of simplifying substitutions for the propagator denominators. FeynArts introduces `FeynAmpDenominator` and we simply replace it by `Dp`, where `p` is some particle identifier. For instance, if we have a single photon or a  $Z$  boson in a given amplitude, which is the case for our studies, then we replace these lengthy objects by `DA` or `DZ` with the appropriate momenta. Another example would be a quark living in different channel propagators, for example `Dqs` or `Dqt`. In the former, the quark is the  $s$ -channel resonant particle and in the latter, it is the  $t$ -channel exchanged particle, with  $Dqd = D_{qd} = i\hat{k}/\hat{d}$ , with  $k^2 = \hat{d}$ . Also, if we have successive decays as in the case of  $e^-e^+$  annihilation, we introduce `DW12` and `DW34`. Here, the  $W$  bosons have the momentum

$p_{12}$  and  $p_{34}$ , respectively, so they have the invariant masses  $m_{12}$  and  $m_{34}$ . This is also illustrative in the sense that the first  $W$  decays into a pair of particles with momenta  $p_1$  and  $p_2$ , and the second one into a pair with momenta  $p_3$  and  $p_4$ . Whenever relevant, all  $D$  factors are assumed complex so they have the full Breit-Wigner form, which is also defined in numerical calculations later. All this leads to simple-looking traditional amplitudes even for larger topologies such as  $2 \rightarrow 3$  and  $2 \rightarrow 4$ .

As far as we know, FeynArts do not immediately support four-fermion interactions yet. Thus, we define the SMEFT amplitudes manually. Doing so, we follow the sign convention of FeynArts and FeynCalc, namely derivatives are replaced by  $-ip$ , assuming outgoing momenta, and we append a factor of  $i$  to obtain the vertex factor after functional derivatives of the interaction Lagrangian. Then, we introduce one more factor of  $i$  to form the amplitude after writing down all the currents, vertex factors, and perhaps gauge-boson polarization vectors.

In order to obtain the squared amplitude, there is a certain order of operations we follow. Once we have the SM and SMEFT amplitudes, first we add them with trackers,

```
amp["SMEFT"] =  
  track["SM"] amp["SM"] + track["X"] amp["X"]
```

where we use X to denote the pure SMEFT amplitudes because the corresponding Feynman diagrams look like  $\times$ . Next, we compute the squared amplitudes as

```
amp["SMEFT"] //  
  
csq //  
  
FermionSpinSum[#, ExtraFactor -> extraFactor]& //  
  
DoPolarizationsSum[#, pV]& //
```

```

projections //
DiracSimplify //
colorSimplify //
momentumSimplify

```

where `csq` is the complex square defined as

```

csq[expr_] :=
  expr ComplexConjugate[expr, Conjugate -> {(* BW propagators, any other
    ↪ complex param *)}]

```

`FermionSpinSum` is the command to convert expressions of the form  $[\bar{u}_2 \Gamma u_1][\bar{u}_1 \Gamma' u_2]$  into  $\text{tr}((\not{p}_2 + m_2)\Gamma(\not{p}_1 + m_1)\Gamma')$ , the `extraFactor` is the factors for initial spin and color averaging, whenever relevant, `DoPolarizationSums` is the polarization sum for the gauge boson  $V$  with momentum  $p_V$  whenever relevant, `projections` is the command where we replace  $\not{p}$  by  $\mathbb{P}_\lambda^\pm \not{p}$  for incoming massless fermions or  $\not{p} \pm m$  by  $(\not{p} \pm m)\mathbb{P}_\lambda^\pm(S)$  for incoming massive fermions when we collide polarized beams, defined for example as

```

projections[expr_] :=
  expr /. {
    DiracGamma[Momentum[k1]] -> proj[sgn1, lam1].GS[k1],
    DiracGamma[Momentum[k2]] + m2 -> (GS[k2]+m2).projGen[+1, lam2, S2]
  } /. {
    proj[sgn_, lam_] :> (1+sgn lam GA5)/2,
    projGen[sgn_, lam_, S_] :> (1+sgn lam GA5.GS[S])/2
  }

```

where the projection operators are defined as  $\mathbb{P}_\lambda^\pm = (1 \pm \lambda \gamma_5)/2$  and  $\mathbb{P}_\lambda^\pm(S) = (1 \pm \lambda \gamma_5 \not{S})/2$ ,  $\mathbf{k1} = k_1$  is the momentum of some massless incoming particle longitudinally polarized with helicity sign  $\mathbf{lam1} = \lambda_1$ ,  $\mathbf{k2} = k_2$  is the momentum of some massive particle having polarization vector  $\mathbf{S} = S^\mu$  with helicity sign  $\mathbf{lam2} = \lambda_2$ ,  $\mathbf{sgn}$  is the particle sign, namely +1 for particles and -1 for antiparticles. Continuing with squaring the amplitude, `colorSimplify` takes care of SU(3) algebra, `momentumSimplify` replaces various momenta in terms of others as much as possible so as to have a squared amplitude that depends on the least amount of independent momenta. For a  $2 \rightarrow 2$  process, this is *natürlich* irrelevant because all dot products are already accounted for by either invariant masses or Mandelstam parameters, but for  $2 \rightarrow 3$  or  $2 \rightarrow 4$  processes, this is especially useful if we can write the process at hand in terms of successive subprocesses so as to have multiple equations for momentum conservation. Finally, since we assume a linear SMEFT observable in Wilson coefficient, or to be more precise at leading order in power of  $1/\Lambda$ , throughout this thesis, we kill the terms proportional to `Track["X"]`<sup>2</sup>. With that, we are ready for numerical calculations.

### 2.5.2. Numerical calculations

For our numerical routines, depending on the complexity of the problem, we either stick to Mathematica if the observable does not require the full phase-space integration, or else we use purely numerical integration tools on Python or Fortran. On Mathematica, we use `ManeParse` to call and use PDFs. We do this in Chapters 3 and 4, where we evaluate our observable at the central points bin by bin, so no integration is necessary. For these particular studies, it is actually a more tedious exercise to keep track of data sets, generate our



own data tables filled with SM and SMEFT predictions and generated uncertainties. For the successive studies, which are slightly different in nature and philosophy, we have more populated final states so the inclusive phase-space integrals require more sophisticated tools than Mathematica's naive `NIntegrate`. We use the Vegas routine for this purpose on Python or Fortran. For the PDFs, we use the LHAPDF library. The involvement of PDFs is more straightforward than preparing the integrand for the Vegas routine.

Modern Vegas integrators allow integration variables to vary between general limits. However, some days we have dependent variables so we need to pay more attention. Furthermore, we still opt to scale our integrals so they run from 0 to 1, which is preferred to increase sampling efficiency. We refer to this process as *unitizing variables*. On Python, with batching and multiprocessing, we can reach lightning speeds on our personal computers compared to widely accepted open-source packages such as MadGraph and MCFM. Of course, this is because our integration routines are written solely for the process under consideration, and because of our manipulation of the integration order so as to further maximize the sampling efficiency. Thus, we include an MWE for Vegas on Python.

Consider the following  $3D$  integral:

$$(2.133) \quad f(x_1, x_2, x_3) = \int_0^1 dx_1 \int_0^{1-x_1} dx_2 \int_0^{x_1^2-x_2^2} dx_3 \ x_1 x_2 x_3.$$

The analytical result of this integral is  $1/480$ , or approximately 0.00208333. Let's write a Vegas routine, pretending that this integration represents our observable, so all the constants, PDFs, and the good stuff are defined or called beforehand. Below is a batched, multiprocessed vegas routine:

```

import vegas

nitn = 10
ncall = int(1e7)
nproc = 10
ntrain = 10

@vegas.lbatchintegrand
def f(var):
    x1min = 0
    x1max = 1
    x1jac = x1max-x1min
    x1 = x1jac*var[:, 0]+x1min

    x2min = 0
    x2max = 1-x1
    x2jac = x2max-x2min
    x2 = x2jac*var[:, 1]+x2min

    x3min = 0
    x3max = x1**2-x2**2
    x3jac = x3max-x3min
    x3 = x3jac*var[:, 2]+x3min

```

```

J = x1jac*x2jac*x3jac

integ = x1*x2*x3

return J*integ

def main():

    integ = vegas.Integrator(3*[[0, 1]], nproc = nproc, nitn = nitn)

    for _ in range(ntrain):

        integ(f, neval = ncall/10)

    result = integ(f, neval = ncall, adapt = False)

    mean = result.mean

    error = result.sdev

    perror = abs(error/mean*100)

    chi2dof = result.chi2/result.dof

    print(

        f"result = "

        f"{round(mean, 8)}"

        f" +- {error:.1e}"

        f" [{perror:.1e}%],"

        f" chi2/dof = {round(chi2dof, 1)}")

```

```
if __name__ == '__main__':
    main()
```

This prints

```
result = 0.00208333 +- 1.1e-08 [5.2e-04%], chi2/dof = 1.0
```

The rest is just to apply this to the observables.

### 2.5.3. Statistical analysis with Mathematica

We conclude this chapter with MWEs of cases that one might encounter in our line of work. Suppose that we activate only three Wilson coefficients,  $C_1$ ,  $C_2$ , and  $C_3$ , to illustrate marginalization and we consider the cases with weak correlation between  $C_1$  and  $C_2$ , moderate correlation between  $C_1$  and  $C_3$ , and strong correlation between  $C_2$  and  $C_3$  to illustrate various shapes of ellipses. We demonstrate how to obtain the nonmarginalized bounds, the marginalized bounds, and the equation for the confidence ellipse at 95% CL in a systematic way that can be easily generalized to a larger number of parameters. We use a Mathematica code snippet accompanying the explanations.

Suppose that we obtain a Fisher matrix

$$(2.134) \quad \mathcal{F} = \begin{pmatrix} 329.825 & 380.117 & -333.333 \\ 380.117 & 487.329 & -416.667 \\ -333.333 & -416.667 & 364.583 \end{pmatrix}.$$

We define this as [F](#). The nonmarginalized 1-sigma (or 68% CL) bounds, [sigmaNM68](#), are the inverse square roots of the diagonal entries of the Fisher matrix. If we multiply these

uncertainties by  $\Delta\chi^2(1, 0.95) = 3.841$ , we obtain the nonmarginalized 2-sigma (or 95% CL) bounds, `sigmaNM95`.

```
In[1]:= sigmaNM68 = 1/Sqrt[Diagonal[F]];
dchisq1 = Quantile[ChiSquareDistribution[1], 0.95];
sigmaNM95 = Sqrt[dchisq1] sigmaNM68;
```

The marginalized 1-sigma bounds, `sigmaM68`, are the square roots of the diagonal entries of the inverse of the Fisher matrix. The marginalized 2-sigma bounds, `sigmaM95`, are obtained with the same factor of  $\Delta\chi^2(1, 0.95)$ .

```
In[4]:= V = Inverse[F];
sigmaM68 = Sqrt[Diagonal[V]];
dchisq1 = Quantile[ChiSquareDistribution[1], 0.95];
sigmaM95 = Sqrt[dchisq1] sigmaM68;
```

The correlation matrix is given by Eq. (2.128).

```
In[8]:= R = DiagonalMatrix[1/sigmaM68].V.DiagonalMatrix[1/sigmaM68];
```

The nonmarginalized confidence ellipses, namely the confidence ellipses when only two Wilson coefficients are activated at a time, at 95% CL are given by Eq. (2.131).

```
In[9]:= NW = Length[F];
Wpairs = Sort /@ Permutations[Range[NW], {2}] //
DeleteDuplicates;
dchisq2 = Quantile[ChiSquareDistribution[2], 0.95];
ellipsesNM = Table[
  {x, y}.{
```

```

{F[[w[[1]], w[[1]]], F[[w[[1]], w[[2]]]]},
{F[[w[[2]], w[[1]]], F[[w[[2]], w[[2]]]]}
}.{x, y} == dchisq2, {w, Wpairs}];

```

Here, the confidence ellipse of the pairs  $(C_1, C_2)$ ,  $(C_1, C_3)$ , and  $(C_2, C_3)$  is called with the indices 1, 2, and 3, respectively. If we now replace the Fisher matrix here by the covariance matrix and take the inverse of the created  $2 \times 2$  matrix, we obtain the equations the 95% CL confidence ellipses described by Eq. (2.132).

```

In[13]:= ellipsesM = Table[
  {x, y}.Inverse[{
    {V[[w[[1]], w[[1]]], V[[w[[1]], w[[2]]]]},
    {V[[w[[2]], w[[1]]], V[[w[[2]], w[[2]]]]}
  }].{x, y} == dchisq2, {w, Wpairs}];

```

The rest is now just to paint these ellipses, like Bob Ross painting on *The Joy of Painting*. The correlation matrix can be plotted with a simple `ArrayPlot` and the ellipses with `ContourPlot`. In Figure 2.11, we present the nonmarginalized bounds (the red dashed rectangles), the marginalized bounds (the red solid rectangles), and the nonmarginalized ellipses (black dashed) and the marginalized ellipses (black) at 95% CL. The correlation values indicated in the plot labels are the ones obtained with the three-dimensional fit. These ellipses beautifully demonstrate the boot camp basics.

To summarize this section, let's note the following.

- Nonmarginalized bounds are obtained by fixing all other parameters to zero.

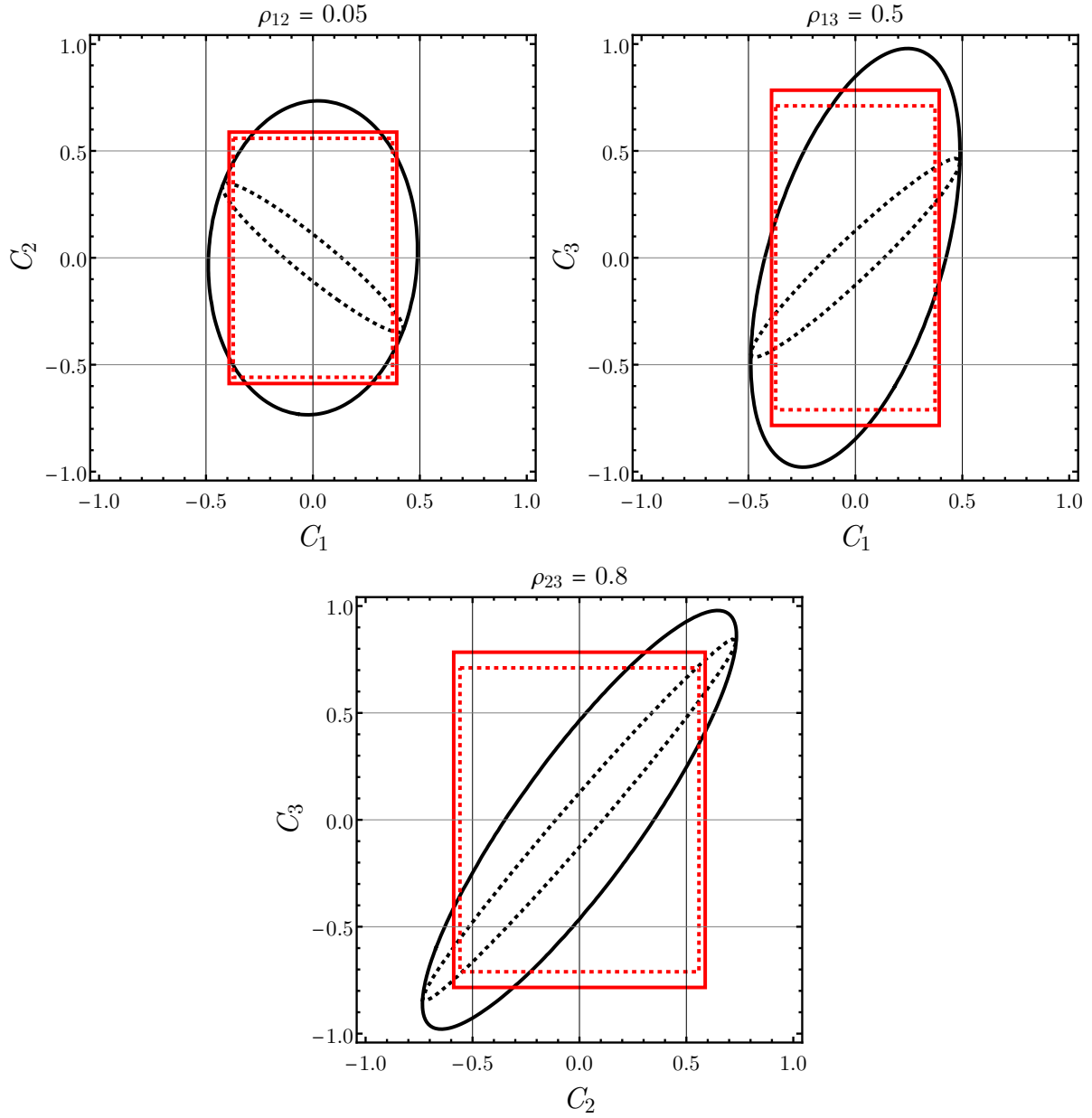


Figure 2.11. Comparison of nonmarginalized and marginalized constraints on the parameter pairs  $(C_1, C_2)$ ,  $(C_1, C_3)$ , and  $(C_2, C_3)$ .

- Marginalized bounds are obtained by allowing the other parameters to vary freely in the fit. They account for degeneracies and correlations, and therefore are generally wider or less stringent.

- Marginalization leads to an effective loss of information due to projection onto a lower-dimensional subspace, which naturally increases the allowed region for each parameter.
- Nonmarginalized ellipses are typically tighter and reflect the direct correlation between two parameters, assuming all others are fixed.
- Marginalized ellipses incorporate correlations with the remaining parameters. This can change the orientation, size, and even the qualitative shape of the allowed region.
- The correlation coefficient between two parameters can differ between nonmarginalized and marginalized cases. Marginalizing over a third parameter can weaken or even flip the sign of the apparent correlation.
- When the correlation between two parameters is small and their uncertainties are comparable, the resulting confidence ellipse is approximately circular in shape.
- As the correlation strength increases, the ellipse becomes increasingly elongated, with its major axis oriented along the principal direction of degeneracy.
- In the limiting case of  $\rho = \pm 1$ , the ellipse becomes a narrow band or even collapses into a line, corresponding to a *flat direction* in the parameter space.
- Such flat directions typically signal *degeneracies* at the observable level, where certain combinations of Wilson coefficients appear together in analytical expressions and cannot be disentangled by the data.
- Identifying these directions is important for both interpreting the fit and designing future measurements that can break the degeneracy and constrain the orthogonal directions.



- In linear models such as SMEFT with interference-only terms, these features are directly encoded in the structure of the Fisher matrix and its inverse, allowing for analytical control and efficient estimation of confidence regions.

## CHAPTER 3

**EIC Wide Shut**

*I didn't need an explanation. Making an effort he took the plunge: "It's all about trying to create ties, you see." Well, sure, I understood. In this life that sometimes seems to be a vast, ill-defined landscape without signposts, amid all of the vanishing lines and the lost horizons, we hope to find reference points, to draw up some sort of land registry so as to shake the impression that we are navigating by chance. So we forge ties, we try to find stability in chance encounters.*

Patrick Modiano, *In the Café of Lost Youth*

In this chapter, we study the potential of the Electron-Ion Collider (EIC) to search for physics beyond the Standard Model using cross section asymmetries in neutral-current deep inelastic scattering. We include a complete accounting of anticipated experimental and theoretical uncertainties. The analysis covers both proton and deuteron beams, with a wide range of beam energies and integrated luminosities. We also consider possible extensions such as a positron beam and a 10-fold luminosity upgrade. The Standard Model Effective Field Theory framework is used to parametrize new physics effects, focusing on semi-leptonic four-fermion operators. A simultaneous fit of the beam polarization and luminosity difference parameters with the Wilson coefficients is also performed. The results show that the EIC can constrain the Wilson coefficients competitively with, and in some cases better than, neutral-current Drell-Yan measurements at the Large Hadron

Collider. Unlike the latter, EIC pseudodata does not yield strong degeneracies in the parameter space.

### 3.1. Prelude

The Standard Model (SM) describes all laboratory phenomena to date. With the discovery of the Higgs boson, the predicted particle spectrum has now been complete. Yet no new particles have been discovered, and no remarkable deviation from the SM has emerged. Despite its success, the theory has several shortcomings. It offers no explanation for dark matter, the baryon-antibaryon asymmetry of the universe, or the origin of neutrino masses. The hierarchy problem remains unresolved. Even within the known parts, there are features that resist interpretation, such as the proton spin decomposition in terms of the spin and orbital angular momentum of its constituent quarks and gluons.

Many experimental programs are running or under design to address these open questions. Our focus in this work is on the Electron-Ion Collider (EIC), currently under construction at Brookhaven National Laboratory. The EIC will collide electrons with protons and nuclei in an intermediate-energy regime, connecting the realms of fixed-target experiments and high-energy colliders. Its luminosity is expected to exceed that of *Hadron-Electron-Ring Anlage* (HERA) by several orders of magnitude, making it the highest-luminosity electron-proton collider to date. It will also be the first lepton-ion collider with both beams polarized, with a fast spin-flip capacity. In addition, uncertainties are expected to be significantly reduced due to improved luminosity measurements and detector acceptance and efficiency. These features together allow for a clean extraction

of parity-violating (PV) asymmetries in neutral-current (NC) deep inelastic scattering (DIS), associated with spin flips of either the electron or the hadron beam.

The EIC is primarily designed as a machine of quantum chromodynamics (QCD), no one promised that it would be a new physics machine, yet its potential reach in probing physics beyond the Standard Model (BSM) is strong. The accessible range of momentum transfer complements both  $Z$ -pole measurements and low-energy precision experiments. The availability of polarized beams offers unique handles on potential new physics effects.

Our goal in this chapter is to provide a detailed assessment of the EIC's sensitivity to new physics, incorporating a complete accounting of the expected experimental and theoretical uncertainties. The primary observables of interest are PV asymmetries. We also consider lepton-charge (LC) asymmetries, assuming that a positron beam will become available.

Given the absence of new particle discoveries, we employ the Standard Model Effective Field Theory (SMEFT) framework to parametrize potential deviations from SM predictions. SMEFT introduces operators of mass dimension greater than four, constructed using SM fields, and suppressed by powers of an ultraviolet (UV) scale  $\Lambda$  higher than SM particle masses and beyond collider reach. We neglect the dimension-5 operator that violates lepton number, and work instead with the leading dimension-6 operators, following the Warsaw basis [58, 28, 110].

We find that the EIC can probe SMEFT scales of a few TeV. The combination of different asymmetry observables allows us to constrain the relevant Wilson coefficients without leaving flat directions, unlike what is observed in NC Drell-Yan (DY) at the Large Hadron Collider (LHC) [23, 47, 142].

This chapter is organized as follows. Section 2 introduces the formalism of NC DIS at the EIC, including SMEFT contributions. Section 3 presents the projected asymmetry datasets and associated uncertainties. Section 4 details the generation of pseudodata, the construction of the uncertainty matrix, and the fitting procedure. Section 5 shows the SMEFT fit results for both single and two-parameter cases. We conclude with a summary in Section 6.

### 3.2. Neutral-current DIS physics at the EIC

#### 3.2.1. SMEFT meets DIS

We generalize the SM DIS cross section and asymmetries to include SMEFT effects. The process of interest is

$$(3.1) \quad \ell(k) + H(P) \rightarrow \ell'(k') + X,$$

where the incoming lepton  $\ell$  is either an electron or a positron, the hadron  $H$  is either a proton or a deuteron, and  $X$  is the final-state hadron. The kinematic variables are

$$(3.2) \quad s = (P + k)^2, \quad Q^2 = -(k - k')^2,$$

$$(3.3) \quad x = \frac{Q^2}{2P \cdot (k - k')}, \quad y = \frac{P \cdot (k - k')}{P \cdot k},$$

$$(3.4) \quad W^2 = (P + k - k')^2,$$

where  $s$  is the collider energy,  $Q$  is the momentum transfer,  $x$  is the momentum fraction carried by the parton,  $y$  is the fractional energy loss of the lepton in the hadron rest frame, and  $W$  is the invariant mass of the hadronic system. These variables are related

via  $Q^2 = xy(s - M^2)$ , where  $M$  is the mass of the incoming hadron. The Feynman diagrams at tree level are shown in Figure 3.1

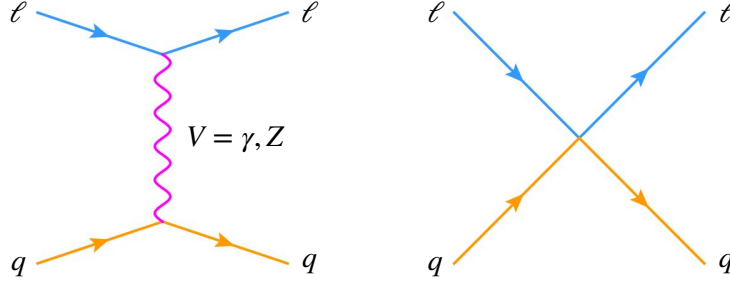


Figure 3.1. The Feynman diagrams for  $\ell + H \rightarrow \ell + X$  at the parton level from one-boson exchange (left) and SMEFT contact interactions (right).

In SMEFT, one constructs operators  $\mathcal{O}_r^{(n)}$  of mass dimension  $n > 4$  at a UV scale  $\Lambda$ , and introduces Wilson coefficients  $C_r^{(n)}$  as effective couplings. The SMEFT Lagrangian looks like

$$(3.5) \quad \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{n>4} \frac{1}{\Lambda^{n-4}} \sum_r C_r^{(n)} \mathcal{O}_r^{(n)}$$

We focus on  $n = 6$  and keep only the SMEFT effects at leading order in the Wilson coefficients, or more precisely, at order  $E^2/\Lambda^2$ , where  $E$  is a typical energy scale, which is the momentum transfer in our case. Higher-order terms may become important for the DY process at the LHC [15, 46], but the EIC runs at low enough momentum transfer to justify neglecting them. We therefore keep only the SM-SMEFT interference and ignore the squared SMEFT amplitudes. Consequently, all observables are linearized with respect to the SMEFT parameters.

The dimension-6 operators of interest are listed in Table 3.1. Following a common phenomenological approach, we write these operators in the vector-axial vector basis

Table 3.1. The list of dimension-6 SMEFT operators relevant to DIS in the Warsaw basis before electroweak symmetry breaking and reexpressed in the vector and axial-vector current basis after electroweak symmetry breaking. The  $c_{V/A_r}^f$  coefficients represent the chiral structure of each operator.

$C_r$	$O_r$	$\tilde{C}_r$	$c_{V_r}^e$	$c_{A_r}^e$	$c_{V_r}^u$	$c_{A_r}^u$	$c_{V_r}^d$	$c_{A_r}^d$
$C_{\ell q}^{(1)}$	$O_{\ell q}^{(1)} = (\bar{\ell}\gamma^\mu\ell)(\bar{q}\gamma_\mu q)$	$C_{\ell q}^{(1)}/4$	1	1	1	1	1	1
$C_{\ell q}^{(3)}$	$O_{\ell q}^{(3)} = (\bar{\ell}\gamma^\mu\tau^I\ell)(\bar{q}\gamma_\mu\tau^I q)$	$C_{\ell q}^{(3)}/4$	1	1	-1	-1	1	1
$C_{eu}$	$O_{eu} = (\bar{e}\gamma^\mu e)(\bar{u}\gamma_\mu u)$	$C_{eu}/4$	1	-1	1	-1	0	0
$C_{ed}$	$O_{ed} = (\bar{e}\gamma^\mu e)(\bar{d}\gamma_\mu d)$	$C_{ed}/4$	1	-1	0	0	1	-1
$C_{\ell u}$	$O_{\ell u} = (\bar{\ell}\gamma^\mu\ell)(\bar{u}\gamma_\mu u)$	$C_{\ell u}/4$	1	1	1	-1	0	0
$C_{\ell d}$	$O_{\ell d} = (\bar{\ell}\gamma^\mu\ell)(\bar{d}\gamma_\mu d)$	$C_{\ell d}/4$	1	1	0	0	1	-1
$C_{qe}$	$O_{qe} = (\bar{e}\gamma^\mu e)(\bar{q}\gamma_\mu q)$	$C_{qe}/4$	1	-1	1	1	1	1

using massive Dirac fields [157]:

$$(3.6) \quad \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \frac{1}{\Lambda^2} \sum_r \tilde{C}_r \left\{ \sum_f [\bar{e}\gamma^\mu (c_{V_r}^e - c_{A_r}^e \gamma_5) e] [\bar{q}_f \gamma_\mu (c_{V_r}^f - c_{A_r}^f \gamma_5) q_f] \right\}.$$

Table 3.1 also shows the electroweak couplings and the rescaled SMEFT parameters.

The differential hadronic cross section is written in terms of the usual SM DIS structure functions and additional SMEFT-induced structure functions:

$$(3.7) \quad \frac{d^2\sigma}{dx dy} = \frac{2\pi y \alpha^2}{Q^4} \left\{ \eta^\gamma L_{\mu\nu}^\gamma W_\gamma^{\mu\nu} + \eta^{\gamma Z} L_{\mu\nu}^{\gamma Z} W_{\gamma Z}^{\mu\nu} + \eta^Z L_{\mu\nu}^Z W_Z^{\mu\nu} + \sum_r \xi^{\gamma r} L_{\mu\nu}^{\gamma r} W_{\gamma r}^{\mu\nu} + \sum_r \xi^{Zr} L_{\mu\nu}^{Zr} W_{Zr}^{\mu\nu} \right\},$$

where  $\alpha$  is the electromagnetic fine structure constant, and  $L_{\mu\nu}^j$  and  $W_{\mu\nu}^j$  with  $j = \gamma, Z, \gamma Z, \gamma r, Zr$  represent the leptonic and hadronic tensors, respectively. These include contributions from the photon and  $Z$  channels, as well as the photon- $Z$ , photon-SMEFT, and  $Z$ -SMEFT interference terms. The  $\eta^j$  and  $\xi^j$  coefficients are

$$(3.8) \quad \eta^\gamma = 1, \quad \eta^{\gamma Z} = \frac{G_F m_Z^2}{2\sqrt{2}\pi\alpha} \frac{Q^2}{Q^2 + m_Z^2}, \quad \eta^Z = (\eta^{\gamma Z})^2,$$

$$(3.9) \quad \xi^{\gamma r} = \frac{\tilde{C}_r}{4\pi\alpha} \frac{Q^2}{\Lambda^2}, \quad \xi^{Zr} = \eta^{\gamma Z} \frac{\tilde{C}_r}{4\pi\alpha} \frac{Q^2}{\Lambda^2},$$

where  $G_F$  is the Fermi constant and  $m_Z$  is the  $Z$  mass. The lepton tensors are

$$(3.10) \quad L_{\mu\nu}^\gamma = 2[k_\mu k'_\nu + k'_\mu k_\nu - k \cdot k' g_{\mu\nu} - i\lambda_e \epsilon_{\mu\nu\alpha\beta} k^\alpha k'^\beta],$$

$$(3.11) \quad L_{\mu\nu}^{\gamma Z} = -(g_V^e - \lambda_e g_A^e) L_{\mu\nu}^\gamma,$$

$$(3.12) \quad L_{\mu\nu}^Z = (g_V^e - \lambda_e g_A^e)^2 L_{\mu\nu}^\gamma,$$

$$(3.13) \quad L_{\mu\nu}^{\gamma r} = (c_{V_r}^e - \lambda_e c_{A_r}^e) L_{\mu\nu}^\gamma,$$

$$(3.14) \quad L_{\mu\nu}^{Zr} = -(c_{V_r}^e - \lambda_e c_{A_r}^e)(g_V^e - \lambda_e g_A^e) L_{\mu\nu}^\gamma,$$

where  $\lambda_e = \pm 1$  is the lepton helicity. For positron beams, one flips the signs of all  $g$  and  $c$  couplings, and also the overall signs of  $L_{\mu\nu}^{\gamma Z}$  and  $L_{\mu\nu}^{\gamma r}$ .

The cross section becomes more explicit when these lepton tensors are inserted:

$$(3.15) \quad \frac{d^2\sigma}{dx dy} = \frac{2\pi y \alpha^2}{Q^4} L_{\mu\nu}^\gamma \left\{ \eta^\gamma W_\gamma^{\mu\nu} - \eta^{\gamma Z} (g_V^e - \lambda_e g_A^e) W_{\gamma Z}^{\mu\nu} + \eta^Z (g_V^e - \lambda_e g_A^e)^2 W_Z^{\mu\nu} \right. \\ \left. + \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) W_{\gamma r}^{\mu\nu} - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) W_{Zr}^{\mu\nu} \right\}.$$



The hadronic structure tensor is parametrized using structure functions as

$$\begin{aligned}
 W_{\mu\nu}^j = & \left( -g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2} \right) F_1^j + \frac{\hat{P}_\mu \hat{P}_\nu}{P \cdot q} F_2^j + \frac{i\epsilon_{\mu\nu\alpha\beta}}{2P \cdot q} (P^\alpha q^\beta F_3^j + 2q^\alpha S^\beta g_1^j) \\
 (3.16) \quad & - \frac{S \cdot q}{P \cdot q} \left[ \frac{\hat{P}_\mu \hat{P}_\nu}{P \cdot q} g_4^j + \left( g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) g_5^j \right].
 \end{aligned}$$

This form follows from general Lorentz structures, available four-momenta, and the spin vector  $S^\mu$  of the nucleus. We define  $\hat{P}^\mu = P^\mu - q^\mu(P \cdot q)/q^2$ . The functions  $F_i^j$  and  $g_i^j$  represent the unpolarized and polarized structure functions. We omit  $g_2^j$  and  $g_3^j$  because they are suppressed by  $M^2/Q^2$ . The spin vector satisfies  $S^2 = -M^2$  and  $S \cdot P = 0$ . For longitudinally polarized nuclei, we use  $S^\mu = \lambda_H(|\mathbf{p}|, E\hat{\mathbf{p}})$ , where  $\lambda_H = \pm 1$  is the nucleon helicity and  $P^\mu = (E, \mathbf{p})$  is its four-momentum.

The structure functions  $F_i$  and  $g_i$  each include the SM piece and a SMEFT contribution:

$$(3.17) \quad F_i = F_i^{\text{SM,NC}} + F_i^{\text{SMEFT}},$$

$$(3.18) \quad g_i = g_i^{\text{SM,NC}} + g_i^{\text{SMEFT}}.$$

The SM terms are given in terms of the familiar NC structure functions as

$$(3.19) \quad F_i^{\text{SM,NC}} = F_i^\gamma - \eta^{\gamma Z} (g_V^e - \lambda_e g_A^e) F_i^{\gamma Z} + \eta^Z (g_V^e - \lambda_e g_A^e)^2 F_i^Z,$$

$$(3.20) \quad g_i^{\text{SM,NC}} = g_i^\gamma - \eta^{\gamma Z} (g_V^e - \lambda_e g_A^e) g_i^{\gamma Z} + \eta^Z (g_V^e - \lambda_e g_A^e)^2 g_i^Z$$

The SMEFT corrections are

$$(3.21) \quad F_i^{\text{SMEFT}} = \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) F_i^{\gamma r} - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) F_i^{Zr},$$

$$(3.22) \quad g_i^{\text{SMEFT}} = \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) g_i^{\gamma r} - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) g_i^{Zr}.$$

The expressions for these structure functions in the parton model are

$$(3.23) \quad \begin{aligned} & \left[ F_2^\gamma, F_2^{\gamma Z}, F_2^Z, F_2^{\gamma r}, F_2^{Zr} \right] = \\ & x \sum_f \left[ Q_f^2, 2Q_f g_V^f, g_V^{f2} + g_A^{f2}, 2Q_f c_{V_r}^f, 2(g_V^f c_{V_r}^f + g_A^f c_{A_r}^f) \right] (q_f + \bar{q}_f), \end{aligned}$$

$$(3.24) \quad \begin{aligned} & \left[ F_3^\gamma, F_3^{\gamma Z}, F_3^Z, F_3^{\gamma r}, F_3^{Zr} \right] = \\ & \sum_f \left[ 0, 2Q_f g_A^f, 2g_V^f g_A^f, 2Q_f c_{A_r}^f, 2(g_V^f c_{A_r}^f + g_A^f c_{V_r}^f) \right] (q_f - \bar{q}_f), \end{aligned}$$

$$(3.25) \quad \begin{aligned} & \left[ g_1^\gamma, g_1^{\gamma Z}, g_1^Z, g_1^{\gamma r}, g_1^{Zr} \right] = \\ & \frac{1}{2} \sum_f \left[ Q_f^2, 2Q_f g_V^f, g_V^{f2} + g_A^{f2}, 2Q_f c_{V_r}^f, 2(g_V^f c_{V_r}^f + g_A^f c_{A_r}^f) \right] (\Delta q_f + \Delta \bar{q}_f), \end{aligned}$$

$$(3.26) \quad \begin{aligned} & \left[ g_5^\gamma, g_5^{\gamma Z}, g_5^Z, g_5^{\gamma r}, g_5^{Zr} \right] = \\ & \sum_f \left[ 0, Q_f g_A^f, g_V^f g_A^f, Q_f c_{A_r}^f, g_V^f c_{A_r}^f + g_A^f c_{V_r}^f \right] (\Delta q_f - \Delta \bar{q}_f). \end{aligned}$$

Here,  $(\Delta)q_f$  refers to the un(polarized) parton distribution functions (PDFs) and  $Q_f$  is the electric charge of parton flavor  $f$ . In the parton model at leading order, the Callan–Gross relations hold, namely  $F_2^j = 2xF_1^j$  and  $g_4^j = 2xg_5^j$ , for all channels  $j$ .

If the incoming hadron is an ion or a nucleus, we also need neutron PDFs. From isospin symmetry, neutron PDFs are given by

$$(3.27) \quad q_{u/n}(x, Q^2) = q_{d/p}(x, Q^2),$$

$$(3.28) \quad q_{d/n}(x, Q^2) = q_{u/p}(x, Q^2),$$

$$(3.29) \quad q_{s/n}(x, Q^2) = q_{s/p}(x, Q^2),$$

$$(3.30) \quad q_{c/n}(x, Q^2) = q_{c/p}(x, Q^2),$$

$$(3.31) \quad \Delta q_{u/n}(x, Q^2) = \Delta q_{d/p}(x, Q^2),$$

$$(3.32) \quad \Delta q_{d/n}(x, Q^2) = \Delta q_{u/p}(x, Q^2),$$

$$(3.33) \quad \Delta q_{s/n}(x, Q^2) = \Delta q_{s/p}(x, Q^2),$$

$$(3.34) \quad \Delta q_{c/n}(x, Q^2) = \Delta q_{c/p}(x, Q^2).$$

For the deuteron, which is an isoscalar combination of a proton and a neutron, the PDFs are given by

$$(3.35) \quad q_{f/D}(x, Q^2) = \frac{1}{2}[q_{f/p}(x, Q^2) + q_{f/n}(x, Q^2)],$$

$$(3.36) \quad \Delta q_{f/D}(x, Q^2) = \frac{1}{2}[\Delta q_{f/p}(x, Q^2) + \Delta q_{f/n}(x, Q^2)].$$

The hadronic cross section, now written in terms of the structure functions  $F_i$  and  $g_i$ , is

$$(3.37) \quad \frac{d^2\sigma^{\lambda_e\lambda_H}}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 F_1 + (1-y)F_2 - \lambda_e \frac{y}{2}(y-2)x F_3 \right. \\ \left. + \lambda_e \lambda_H (2-y)xy g_1 - \lambda_H (1-y)g_4 - \lambda_H xy^2 g_5 \right\},$$

for specific helicities of the electron and nucleon. We neglect the electron mass and corrections of order  $M^2/Q^2$ . The unpolarized and polarized cross sections are built by summing combinations of this expression evaluated at all helicity configurations:

$$(3.38) \quad d\sigma_0 = \frac{1}{4} (d\sigma^{++} + d\sigma^{+-} + d\sigma^{-+} + d\sigma^{--}) ,$$

$$(3.39) \quad d\sigma_e = \frac{1}{4} (d\sigma^{++} + d\sigma^{+-} - d\sigma^{-+} - d\sigma^{--}) ,$$

$$(3.40) \quad d\sigma_H = \frac{1}{4} (d\sigma^{++} - d\sigma^{+-} + d\sigma^{-+} - d\sigma^{--}) ,$$

$$(3.41) \quad d\sigma_{eH} = \frac{1}{4} (d\sigma^{++} - d\sigma^{+-} - d\sigma^{-+} + d\sigma^{--}) .$$

Here,  $d\sigma_0$  is the unpolarized cross section,  $d\sigma_e$  refers to a polarized electron and unpolarized hadron,  $d\sigma_H$  corresponds to a polarized hadron and unpolarized electron, the fully polarized case is  $d\sigma_{eH}$ , and we have suppressed the differentials  $dx$  and  $dy$ . The SM parts of these cross sections are

$$(3.42) \quad \begin{aligned} \frac{d^2\sigma_0}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 \left[ F_1^\gamma - g_V^e \eta_{\gamma Z} F_1^{\gamma Z} + (g_V^e)^2 + g_A^e{}^2 \right] \eta_Z F_1^Z \right. \\ &\quad + (1-y) \left[ F_2^\gamma - g_V^e \eta_{\gamma Z} F_2^{\gamma Z} + (g_V^e)^2 + g_A^e{}^2 \right] \eta_Z F_2^Z \\ &\quad \left. - \frac{xy}{2} (2-y) \left[ g_A^e \eta_{\gamma Z} F_3^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_3^Z \right] \right\} , \end{aligned}$$

$$(3.43) \quad \begin{aligned} \frac{d^2\sigma_e}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 \left[ g_A^e \eta_{\gamma Z} F_1^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_1^Z \right] \right. \\ &\quad + (1-y) \left[ g_A^e \eta_{\gamma Z} F_2^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_2^Z \right] \\ &\quad \left. + \frac{xy}{2} (2-y) \left[ g_V^e \eta_{\gamma Z} F_3^{\gamma Z} - (g_V^e)^2 + g_A^e{}^2 \right] \eta_Z F_3^Z \right\} , \end{aligned}$$

$$\frac{d^2\sigma_H}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \left\{ (2-y) xy \left[ g_A^e \eta_{\gamma Z} g_1^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_1^Z \right] \right.$$

$$(3.44) \quad \begin{aligned} & + (1-y) \left[ g_V^e \eta_{\gamma Z} g_4^{\gamma Z} - (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_4^Z \right] \\ & - xy^2 \left[ g_V^e \eta_{\gamma Z} g_5^{\gamma Z} - (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_5^Z \right] \Big\}, \end{aligned}$$

$$(3.45) \quad \begin{aligned} \frac{d^2 \sigma_{eH}}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \Big\{ & (2-y) xy \left[ g_1^\gamma - g_V^e \eta_{\gamma Z} g_1^{\gamma Z} + (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_1^Z \right] \\ & - (1-y) \left[ g_A^e \eta_{\gamma Z} g_4^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_4^Z \right] \\ & - xy^2 \left[ g_A^e \eta_{\gamma Z} g_5^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_5^Z \right] \Big\}. \end{aligned}$$

The SMEFT contributions are

$$(3.46) \quad \begin{aligned} \frac{d^2 \sigma_0^{\text{SMEFT}}}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \sum_r \Big[ & xy^2 (c_{V_r}^e \xi_{\gamma r} F_1^{\gamma r} - (c_{V_r}^e g_V^e + c_{A_r}^e g_A^e) \xi_{Zr} F_1^{Zr}) \\ & + (1-y) (c_{V_r}^e \xi_{\gamma r} F_2^{\gamma r} - (c_{V_r}^e g_V^e + c_{A_r}^e g_A^e) \xi_{Zr} F_2^{Zr}) \\ & + \frac{xy}{2} (2-y) (c_{A_r}^e \xi_{\gamma r} F_3^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_3^{Zr}) \Big], \end{aligned}$$

$$(3.47) \quad \begin{aligned} \frac{d^2 \sigma_e^{\text{SMEFT}}}{dx dy} = -\frac{4\pi\alpha^2}{xyQ^2} \sum_r \Big[ & xy^2 (c_{A_r}^e \xi_{\gamma r} F_1^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_1^{Zr}) \\ & + (1-y) (c_{A_r}^e \xi_{\gamma r} F_2^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_2^{Zr}) \\ & + \frac{xy}{2} (2-y) (c_{V_r}^e \xi_{\gamma r} F_3^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} F_3^{Zr}) \Big], \end{aligned}$$

$$(3.48) \quad \begin{aligned} \frac{d^2 \sigma_H^{\text{SMEFT}}}{dx dy} = -\frac{4\pi\alpha^2}{xyQ^2} \sum_r \Big[ & xy(2-y) (c_{A_r}^e \xi_{\gamma r} g_1^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_1^{Zr}) \\ & + (1-y) (c_{V_r}^e \xi_{\gamma r} g_4^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_4^{Zr}) \\ & + xy^2 (c_{V_r}^e \xi_{\gamma r} g_5^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_5^{Zr}) \Big], \end{aligned}$$

$$\frac{d^2 \sigma_{eH}^{\text{SMEFT}}}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \sum_r \Big[ xy(2-y) (c_{V_r}^e \xi_{\gamma r} g_1^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_1^{Zr})$$

$$\begin{aligned}
& + (1 - y)(c_{A_r}^e \xi_{\gamma r} g_4^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_4^{Zr}) \\
(3.49) \quad & + xy^2(c_{A_r}^e \xi_{\gamma r} g_5^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_5^{Zr}) \Big].
\end{aligned}$$

The observables of interest are PV and LC asymmetries. To be more precise, we consider the left-right asymmetry of electrons with unpolarized hadrons,

$$(3.50) \quad A_{\text{PV}}^e = \frac{d\sigma_e}{d\sigma_0},$$

left-right asymmetry of hadrons with unpolarized electrons,

$$(3.51) \quad A_{\text{PV}}^H = \frac{d\sigma_H}{d\sigma_0},$$

and unpolarized electron-positron asymmetry with unpolarized hadrons,

$$(3.52) \quad A_{\text{LC}} = \frac{d\sigma_0^{e^+} - d\sigma_0^{e^-}}{d\sigma_0^{e^+} + d\sigma_0^{e^-}}.$$

### 3.2.2. Measurement of PV asymmetries at the EIC

In DIS experiments with polarized electron and hadron beams, the measured differential cross section with the beam polarizations  $P_e$  and  $P_H$  is given by

$$(3.53) \quad d\sigma = d\sigma_0 + P_e d\sigma_e + P_H d\sigma_H + P_e P_H d\sigma_{eH}.$$

Here,  $P_e$  and  $P_H$  play a role similar to helicity signs, but in practice they are not exactly  $\pm 100\%$ . Instead, they take values somewhere in between. PV asymmetries correspond to flipping the spin direction of either the electron or the hadron.

At the EIC, electron and hadron beams with opposite polarization signs will be entered alternately into the storage rings. This means both beam polarizations will flip periodically over short time scales. This is in contrast to HERA, where data were collected with one polarization at a time. The intervals between different polarization configurations were so long that each setting effectively formed an independent experiment. The EIC setup avoids that limitation.

The DIS event counts during a given beam helicity configuration are expressed as

$$(3.54) \quad N^{++} = a_{\text{det}} L^{++} \left( d\sigma_0 + |P_e^{++}| d\sigma_e + |P_H^{++}| d\sigma_H + |P_e^{++}| |P_H^{++}| d\sigma_{eH} \right),$$

$$(3.55) \quad N^{+-} = a_{\text{det}} L^{+-} \left( d\sigma_0 + |P_e^{+-}| d\sigma_e - |P_H^{+-}| d\sigma_H - |P_e^{+-}| |P_H^{+-}| d\sigma_{eH} \right),$$

$$(3.56) \quad N^{-+} = a_{\text{det}} L^{-+} \left( d\sigma_0 - |P_e^{-+}| d\sigma_e + |P_H^{-+}| d\sigma_H - |P_e^{-+}| |P_H^{-+}| d\sigma_{eH} \right),$$

$$(3.57) \quad N^{--} = a_{\text{det}} L^{--} \left( d\sigma_0 - |P_e^{--}| d\sigma_e - |P_H^{--}| d\sigma_H + |P_e^{--}| |P_H^{--}| d\sigma_{eH} \right),$$

where  $L^{XY}$  is the integrated luminosity for configuration  $XY$ , and  $P_e^{XY}$  and  $P_H^{XY}$  are the electron and hadron beam polarizations for that helicity bunch. The label  $XY = ++, +-, -+, --$  corresponds to the electron and hadron helicity signs, respectively. The factor  $a_{\text{det}}$  accounts for detector acceptance, efficiency, and phase space coverage. Assuming beam polarization, luminosity, and detector effects to be constant over the measurement period, we get

$$(3.58) \quad d\sigma_0 = \frac{1}{4} \left( d\sigma^{++} + d\sigma^{+-} + d\sigma^{-+} + d\sigma^{--} \right),$$

$$(3.59) \quad d\sigma_e = \frac{1}{4|P_e|} \left( d\sigma^{++} + d\sigma^{+-} - d\sigma^{-+} - d\sigma^{--} \right),$$

$$(3.60) \quad d\sigma_H = \frac{1}{4|P_H|} (d\sigma^{++} - d\sigma^{+-} + d\sigma^{-+} - d\sigma^{--}),$$

$$(3.61) \quad d\sigma_{eH} = \frac{1}{4|P_e||P_H|} (d\sigma^{++} - d\sigma^{+-} - d\sigma^{-+} + d\sigma^{--}),$$

where the experimentally measured cross section is defined by  $d\sigma^{XY} = N^{XY}/L^{XY}/a_{\text{det}}$ .

The PV asymmetries are then obtained as ratios of these measured cross sections. Since both beam helicities flip frequently on a short time scale, it is safe to treat  $a_{\text{det}}$  as constant.

We extract the asymmetries from the measured yields  $Y^{XY} = N^{XY}/L^{XY}$  using

$$(3.62) \quad A_{\text{PV}}^e = \frac{d\sigma_e}{d\sigma_0} = \frac{1}{|P_e|} \frac{Y^{++} + Y^{+-} - Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}},$$

$$(3.63) \quad A_{\text{PV}}^H = \frac{d\sigma_H}{d\sigma_0} = \frac{1}{|P_H|} \frac{Y^{++} - Y^{+-} + Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}}.$$

The EIC has been designed to keep point-to-point luminosity uncertainties at the level of  $\mathcal{O}(10^{-4})$ . As a result, the dominant experimental uncertainty is expected to come from polarimetry.

### 3.2.3. Measurement of LC asymmetries at the EIC

PV asymmetries can be extracted from the yields within a single run. In contrast, forming the LC asymmetry requires two separate runs: one with electrons, the other with positrons. To reduce possible systematic effects coming from differences in electron and positron detection, the magnet polarity can be flipped between the runs. In this case, the dominant uncertainty is expected to be from luminosity. For the LC asymmetry, we assume a 2% relative uncertainty in luminosity, treated as an absolute uncertainty on the asymmetry itself.



### 3.3. Projections of PV and LC asymmetry data

The anticipated runs and corresponding pseudodatasets, including beam energies, integrated luminosities, and the labeling convention used throughout this study, are listed in Table 3.2.

Table 3.2. Expected runs at the EIC, in terms of beam energy, beam type, and nominal annual luminosity, together with our labeling scheme. P6 is the Yellow Report reference setting.

D1	5 GeV $\times$ 41 GeV $e^-D$ , 4.4 fb $^{-1}$	P1	5 GeV $\times$ 41 GeV $e^-p$ , 4.4 fb $^{-1}$
D2	5 GeV $\times$ 100 GeV $e^-D$ , 36.8 fb $^{-1}$	P2	5 GeV $\times$ 100 GeV $e^-p$ , 36.8 fb $^{-1}$
D3	10 GeV $\times$ 100 GeV $e^-D$ , 44.8 fb $^{-1}$	P3	10 GeV $\times$ 100 GeV $e^-p$ , 44.8 fb $^{-1}$
D4	10 GeV $\times$ 137 GeV $e^-D$ , 100 fb $^{-1}$	P4	10 GeV $\times$ 275 GeV $e^-p$ , 100 fb $^{-1}$
D5	18 GeV $\times$ 137 GeV $e^-D$ , 15.4 fb $^{-1}$	P5	18 GeV $\times$ 275 GeV $e^-p$ , 15.4 fb $^{-1}$
		P6	18 GeV $\times$ 275 GeV $e^-p$ , 100 fb $^{-1}$

We refer to these configurations as datasets for simplicity. D# and P# labels denote the unpolarized PV asymmetry datasets in  $e^-D$  and  $e^-p$  collisions, while  $\Delta$ D# and  $\Delta$ P# indicate the polarized ones. LD# and LC# labels are used for the LC asymmetry datasets. P6 represents the unrealistic  $e^-p$  collision setting with the highest energy and luminosity, and corresponds to the reference configuration used in the Yellow Report [9]. We also include a high-luminosity scenario, HL-EIC, in which the integrated luminosity of each run is increased by a factor of 10.

The projected data are subject to a series of baseline cuts. We require  $Q > 1$  GeV to suppress nonperturbative QCD effects. To reduce uncertainties from bin migration

and unfolding, we apply a cut  $y > 0.1$ . To suppress photoproduction backgrounds from the final-state hadron, we apply  $y < 0.9$ . For the scattered lepton, we require  $|\eta'| < 3.5$  to restrict events to the main detector region and  $E' > 2$  GeV to ensure high sample purity. In the SMEFT analysis, we apply additional cuts of  $x < 0.5$  and  $Q > 10$  GeV. These remove regions where uncertainties from nonperturbative QCD and nuclear effects become large. The kinematic region covered by the datasets spans  $\sqrt{s} = 70$  to 140 GeV, with  $0.1 \leq y \leq 0.9$ . This coverage is shown in Figure 3.2.

Now let's talk about the money. The statistical uncertainty on the unpolarized PV asymmetry is determined by the event count and corrected by the electron beam polarization. It is given by  $\delta A_{\text{PV, stat}}^{(e)} = 1/P_e/\sqrt{N}$ . For the HL-EIC scenario, where the integrated luminosity is increased by a factor of 10, this improves to  $\delta A_{\text{PV, stat, HL}}^{(e)} = \delta A_{\text{PV, stat}}^{(e)}/\sqrt{10}$ . For the polarized PV asymmetry, the hadron beam polarization should be included, and the uncertainty becomes  $\delta A_{\text{PV, stat}}^{(H)} = (P_e/P_H) \delta A_{\text{PV, stat}}^{(e)}$ . In the case of the LC asymmetry, both beams are unpolarized, so the polarization factors are removed. We also introduce a  $\sqrt{10}$  penalty to account for the expected 10-fold reduction in positron luminosity. The resulting uncertainty is  $\delta A_{\text{LC, stat}} = \sqrt{10} P_e \delta A_{\text{PV, stat}}^{(e)}$ . We consider a HL-EIC upgrade also for the polarized PV asymmetry, but not for the LC case. The remaining sources of uncertainty are treated as follows. A 1% uncorrelated systematic uncertainty is assigned to all asymmetries. Beam polarization uncertainties are taken as fully correlated, and 1% for the unpolarized PV asymmetry and 2% for the polarized case, since the latter is smaller in magnitude. The LC asymmetry has no associated polarization uncertainty. We include an absolute luminosity uncertainty of 2% for the LC asymmetry, treated as fully

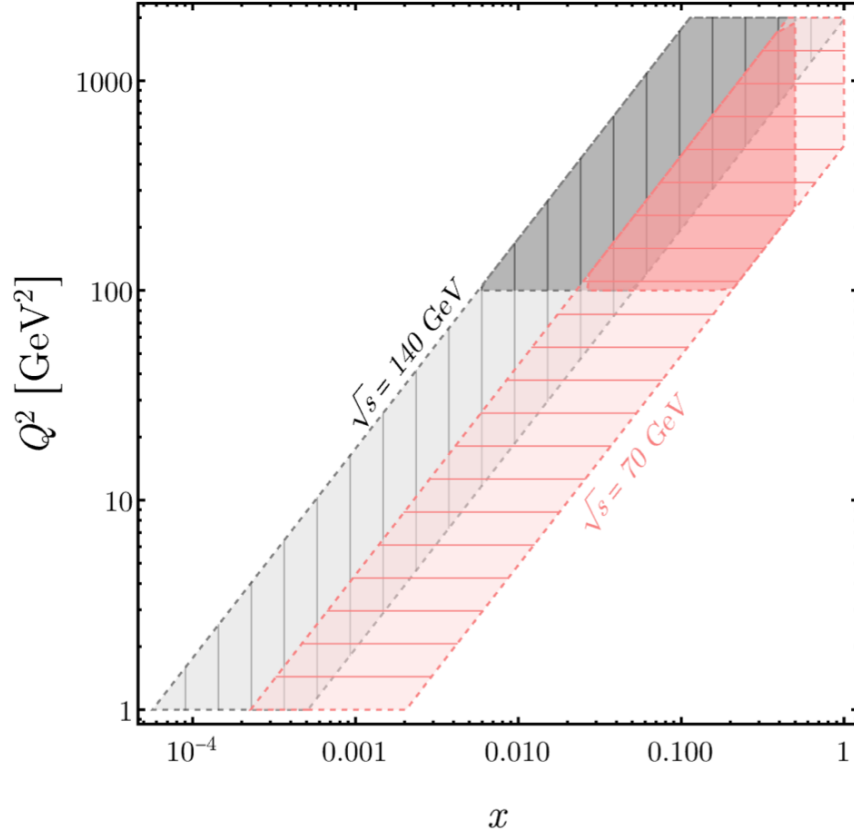


Figure 3.2. The kinematic coverage at the EIC in terms of the Bjorken- $x$  variable and the momentum transfer for the lowest and highest collider energies. The shaded region indicates the *good* region used in our SMEFT analysis.

correlated. For the LC observables, we also include higher-order quantum electrodynamics (QED) effects by assigning the 5% of the difference between the next-to-leading-order (NLO) and Born-level values as uncorrelated uncertainty. Finally, PDF variations are included as fully correlated theoretical uncertainties. All experimental and theoretical uncertainties are summarized in Table 3.3.

Table 3.3. Anticipated uncertainty components for each observable at each family of runs at the EIC. NL means the case of nominal luminosity and HL indicates the case of a 10-fold higher luminosity.

Uncertainty	$A_{\text{PV}}^e$ (D, P)	$A_{\text{PV}}^H$ ( $\Delta D$ , $\Delta P$ )	$A_{\text{LC}}$ (LD, LC)
statistical (NL)	$\delta A_{\text{PV,stat}}^e = \frac{1}{P_e \sqrt{N}}$	$\frac{P_e}{P_H} \delta A_{\text{PV,stat}}^e$	$\sqrt{10} P_e \delta A_{\text{PV,stat}}^e$
statistical (HL)	$\frac{1}{\sqrt{10}} \delta A_{\text{PV,stat}}^e$	$\frac{1}{\sqrt{10}} \frac{P_e}{P_H} \delta A_{\text{PV,stat}}^e$	—
uncorrelated systematics	1% rel.	1% rel.	1% rel.
fully correlated beam polarization	1% rel.	2% rel.	—
fully correlated luminosity	—	—	2% abs.
uncorrelated NLO QED	—	—	$5\% \times (A_{\text{LC}}^{\text{NLO}} - A_{\text{LC}}^{\text{Born}})$
fully correlated PDF	yes	yes	yes

### 3.4. Pseudodata generation, the uncertainty matrix, and the fitting

For the PV asymmetries, both unpolarized and polarized, three sources of uncertainty are considered: statistical, uncorrelated systematics, and correlated beam polarization uncertainty. For the  $b^{\text{th}}$  bin, the pseudodata are generated as

$$(3.64) \quad A_{\text{PV},b}^{\text{pseudo}} = A_{\text{PV,SM},b} + r_b(\delta A_{\text{PV,stat},b} \oplus \delta A_{\text{PV,sys},b}) + r' \delta A_{\text{PV,pol},b},$$

where  $\oplus$  means sum in quadrature, and  $r_b, r' \sim \mathcal{N}(0, 1)$  are independent unit normal variates. For the LC asymmetries, four types of uncertainty are included: statistical, uncorrelated systematics, uncorrelated NLO QED effects, and correlated luminosity uncertainty. The pseudodata in the  $b^{\text{th}}$  bin are generated as

$$(3.65) \quad A_{\text{LC},b}^{\text{pseudo}} = A_{\text{LC,SM},b} + r_b(\delta A_{\text{LC,stat},b} \oplus \delta A_{\text{LC,sys},b} \oplus \delta A_{\text{LC,NLO QED},b}) + r' \delta A_{\text{LC,lum},b},$$

The total uncertainty matrix has two parts, experimental and theoretical,

$$(3.66) \quad \mathcal{E} = \mathcal{E}_{\text{exp}} + \mathcal{E}_{\text{theo}}$$

The only theoretical uncertainty we consider is from PDF variations. The experimental part is defined as

$$(3.67) \quad \mathcal{E}_{\text{exp},bb'} = \begin{cases} (\delta A_b^{\text{uncorr}} \oplus \delta A_b^{\text{corr}})^2, & b = b', \\ \rho_{bb'} \delta A_b^{\text{corr}} \delta A_{b'}^{\text{corr}}, & b \neq b', \end{cases}$$

where  $\delta A_b^{\text{un(corr)}}$  denotes the total un(correlated) uncertainty and we assume full correlation  $\rho_{bb'} = 1$ . The PDF uncertainty matrix is given by

$$(3.68) \quad \mathcal{E}_{\text{pdf},bb'} = \frac{1}{N_P} \sum_{m=1}^{N_P} (A_{\text{SM},m,b} - A_{\text{SM},0,b})(A_{\text{SM},m,b'} - A_{\text{SM},0,b'}),$$

where  $A_{\text{SM},0(m),b}$  is the observable evaluated at the central ( $m^{\text{th}}$ ) member of the PDF set. We use NNPDF3.1 NLO [32] for the unpolarized cross section and NNPDFPOL1.1 [141] for the polarized one.

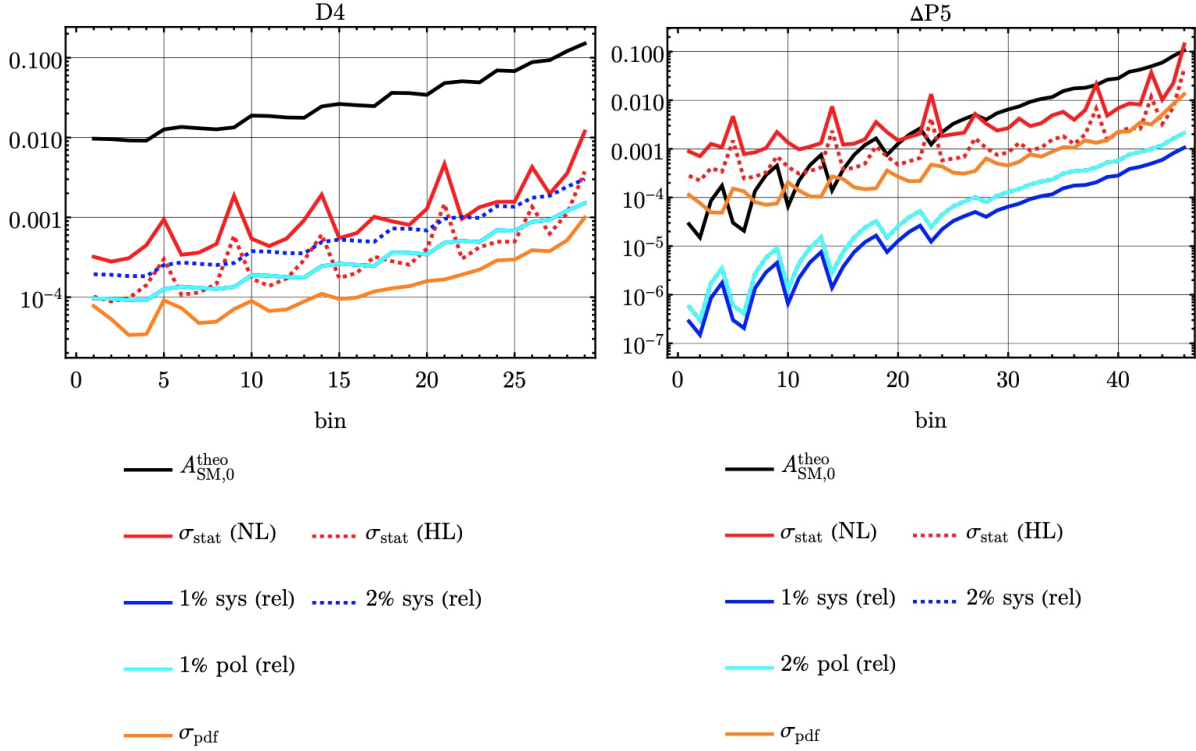


Figure 3.3. Error budget plots for representative datasets D4,  $\Delta P4$ , and LP5.

In Figures 3.3 and 3.4, we present the error budget plots for representative datasets D4,  $\Delta P5$ , and LC5. These plots show the contribution of the uncertainty contributions to the diagonal entries of the uncertainty matrix. The black line corresponds to the central value of the asymmetry,  $A_{\text{PV}}^{(e)}$ ,  $A_{\text{PV}}^{(H)}$ , or  $A_{\text{LC}}$ , depending on the dataset. The solid red line shows the statistical uncertainty for the nominal luminosity case, while the dashed red line shows the HL-EIC scenario. No high-luminosity configuration is assumed for  $A_{\text{LC}}$ . The solid blue line corresponds to the default uncorrelated systematic uncertainty of 1% relative, and the dashed blue line represents a hypothetical 2% case, which is shown for illustration but not used in the analysis. The cyan lines indicate the correlated beam polarization uncertainty, 1% for  $A_{\text{PV}}^{(e)}$  and 2% for  $A_{\text{PV}}^{(H)}$ , and 2% absolute luminosity uncertainty for

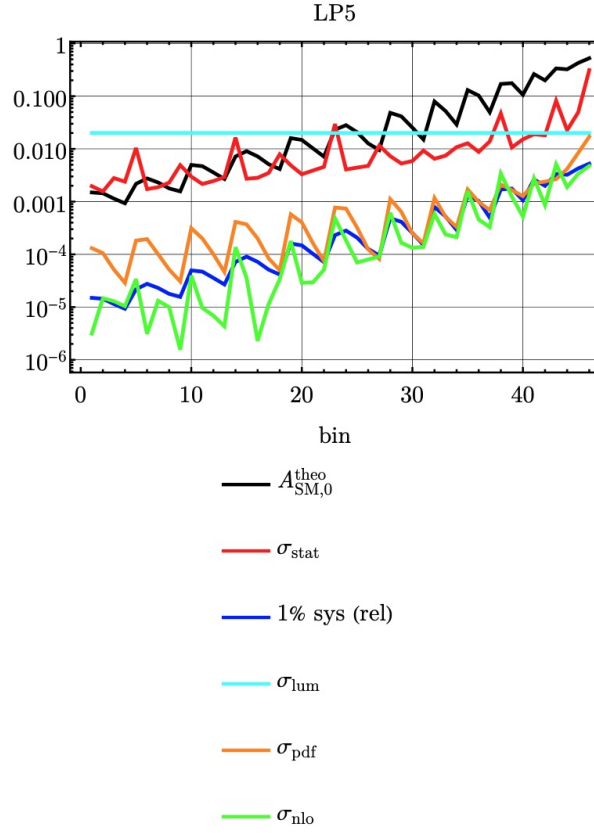


Figure 3.4. The same as Figure 3.3 but for LP5.

$A_{LC}$ . The orange line shows the uncertainty from PDF variations, and the green line, included only for  $A_{LC}$ , corresponds to NLO QED uncertainty introduced as higher-order effects. The horizontal axes show only those bins that pass both detector-level cuts and the additional cuts used in the SMEFT analysis. The bins are sorted by momentum transfer first and then by Bjorken- $x$ , which explains the wave pattern. In the nominal luminosity case, statistical uncertainty dominates the PV asymmetries. In the high-luminosity case, systematic and beam polarization uncertainties are comparable. The LC asymmetry is dominated by luminosity uncertainty. PDF uncertainties are subdominant for  $A_{PV}^{(e)}$  but become more important in the polarized case.

We perform fits using both  $e^-p$  and  $e^-D$  data, focusing on datasets #4 and #5, where the center-of-mass (c.m.) energy is higher and SMEFT effects become more visible. The fit is based on a standard  $\chi^2$  function:

$$(3.69) \quad \chi^2 = \sum_{b=1}^{N_B} \sum_{b'=1}^{N_B} (A_{\text{SMEFT}} - A^{\text{pseudo}})_b \mathcal{H}_{bb'} (A_{\text{SMEFT}} - A^{\text{pseudo}})_{b'},$$

where  $\mathcal{H} = \mathcal{E}^{-1}$  is the inverse of the uncertainty matrix. The error budget plots suggest that the uncertainties from polarimetry and luminosity differences could become limiting factors. To assess this, we also perform simultaneous fits of SMEFT parameters along with either the beam polarization parameter  $P$  or the luminosity difference parameter  $A_{\text{lum}}$ , in order to potentially improve the bounds on the Wilson coefficients. In the fits where  $P$  is included as a free parameter, the  $\chi^2$  function is defined as

$$(3.70) \quad \chi^2 = \sum_{b=1}^{N_B} \sum_{b'=1}^{N_B} (PA_{\text{SMEFT}} - A^{\text{pseudo}})_b \mathcal{H}_{bb'} (PA_{\text{SMEFT}} - A^{\text{pseudo}})_{b'} + \frac{(P - \bar{P})^2}{\delta P^2}.$$

In this case, we omit the beam polarization uncertainty from the error matrix, since  $P$  is treated as a fit parameter. We set the reference value  $\bar{P} = 1$ . The idea is that polarimetry gives an external estimate of  $P$ , but a better value may be inferred directly from the data, constrained within the uncertainty provided by polarimetry. In the fits where  $A_{\text{lum}}$  is included as a free parameter, the  $\chi^2$  function is defined as

$$(3.71) \quad \chi^2 = \sum_{b=1}^{N_B} \sum_{b'=1}^{N_B} (A_{\text{SMEFT}} - A^{\text{pseudo}} - A_{\text{lum}})_b \mathcal{H}_{bb'} (A_{\text{SMEFT}} - A^{\text{pseudo}} - A_{\text{lum}})_{b'}.$$

Here, we omit the luminosity uncertainty from the uncertainty matrix.



Let  $\mathbf{W}$  denote the full set of fit parameters, including the Wilson coefficients and, where applicable, the additional nuisance parameter,  $P$  or  $A_{\text{lum}}$ . The best-fit values are obtained by solving the condition

$$(3.72) \quad \nabla \chi^2(\overline{\mathbf{W}}) = 0.$$

The Fisher information matrix is constructed from the Hessian of the  $\chi^2$  function evaluated at the best-fit point:

$$(3.73) \quad \mathcal{F} = \frac{1}{2} \nabla \nabla \chi^2(\overline{\mathbf{W}}).$$

The inverse of the Fisher matrix gives the covariance matrix  $\mathcal{V}$ .

Since a single pseudodataset reflects just one outcome, we repeat the procedure over  $N_E = 1000$  pseudoexperiments to generate statistics. The final best-fit values and corresponding covariance matrix are then obtained by averaging over this ensemble:

$$(3.74) \quad \overline{\mathbf{W}} = \left[ \sum_{e=1}^{N_E} \mathcal{F}_e \right]^{-1} \sum_{e=1}^{N_E} \mathcal{F}_e \overline{\mathbf{W}}_e,$$

$$(3.75) \quad \mathcal{F} = \frac{1}{N_E} \sum_{e=1}^{N_E} \mathcal{F}_e.$$

The fit results are shown in the next section.

### 3.5. SMEFT fit results

#### 3.5.1. Single Wilson coefficients

We now present the results of the single-parameter fits for the Wilson coefficients. These are quoted as 95% confidence level (CL) intervals, averaged over 1000 pseudoexperiments.

The bounds on the Wilson coefficients are shown in Figures 3.5–3.11.

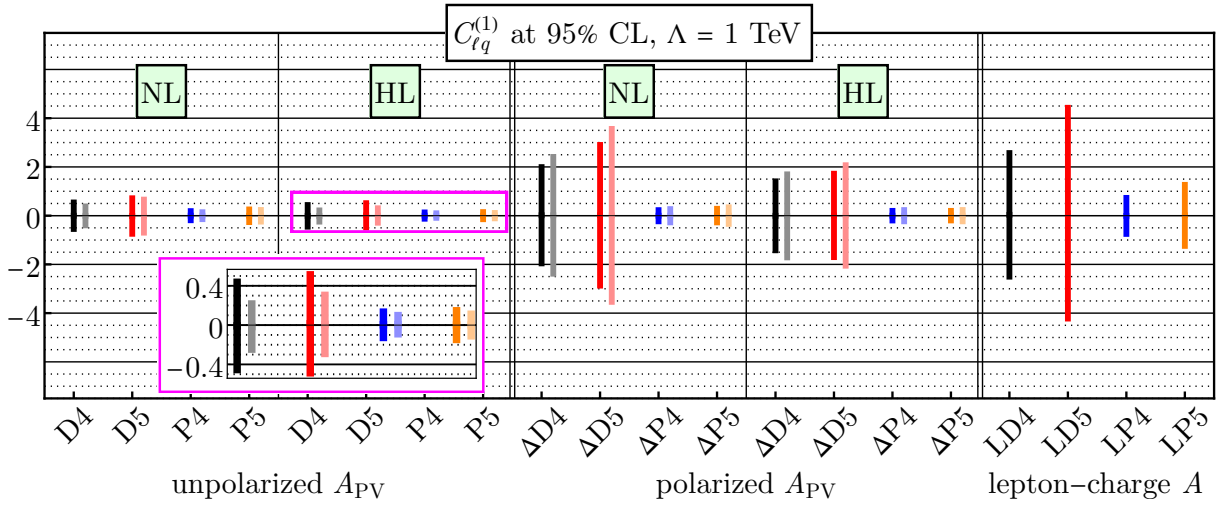


Figure 3.5. 95% CL bounds of  $C_{\ell q}^{(1)}$  from single-parameters fits (darker) and from the  $(1+1)$ -parameter fits with beam polarization as an additional fitting parameter (lighter) using the families of data sets D4, D5, P4, and P5 at  $\Lambda = 1$  TeV.

The plots are grouped into three categories based on the observable: unpolarized PV asymmetry, polarized PV asymmetry, and the LC asymmetry. For each PV asymmetry type, we show results for both the nominal and high-luminosity scenarios. The fits are restricted to datasets #4 and #5, where the c.m. energy is highest and sensitivity to

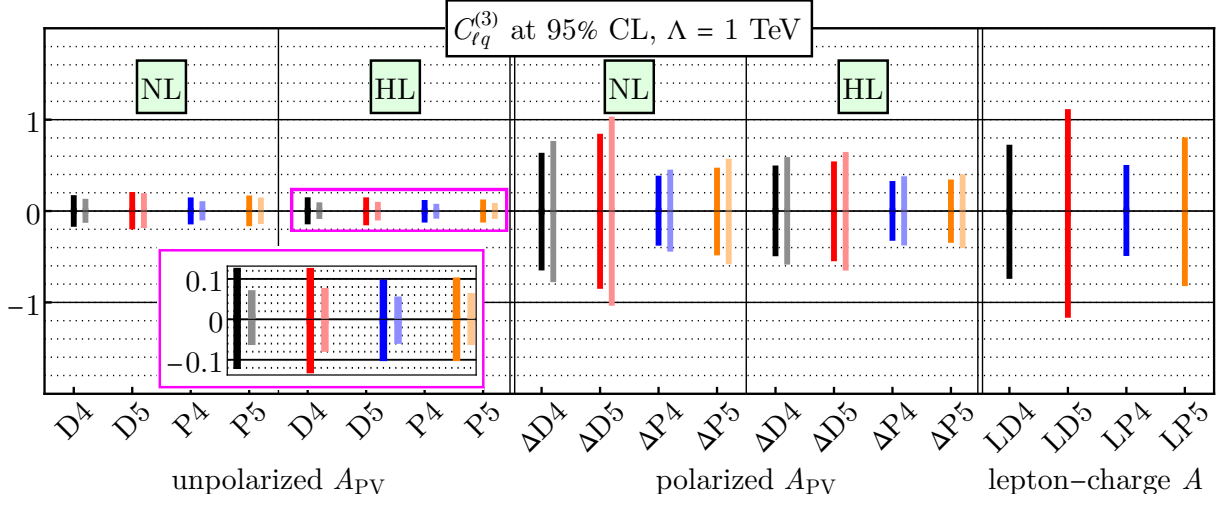


Figure 3.6. The same as Figure 3.5 but for  $C_{\ell q}^{(3)}$ .

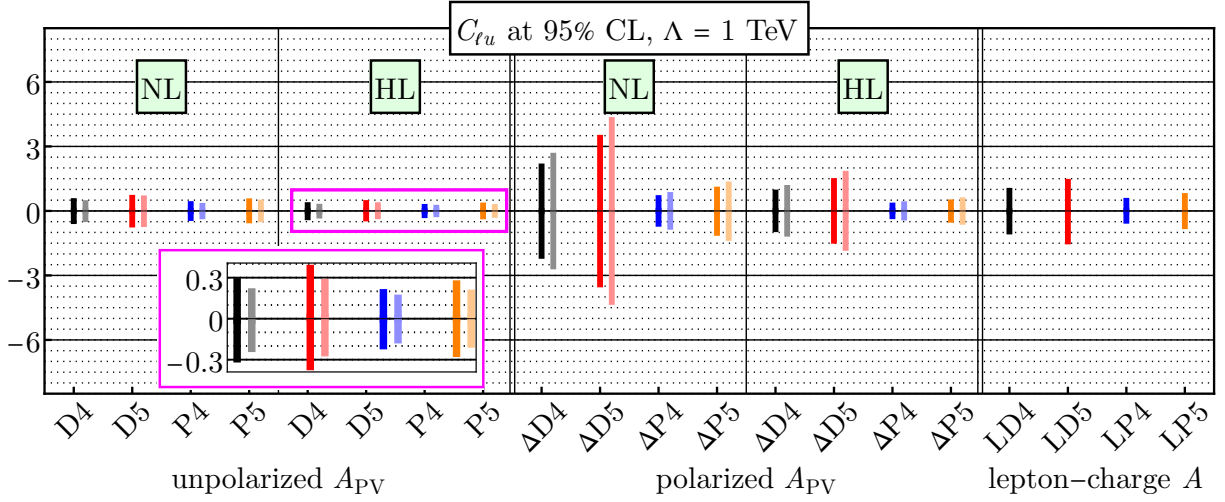


Figure 3.7. The same as Figure 3.5 but for  $C_{\ell u}$ .

SMEFT effects is enhanced. In each plot, two lines are shown per parameter. The darker line corresponds to the fit where only the Wilson coefficient is fitted, and the lighter line

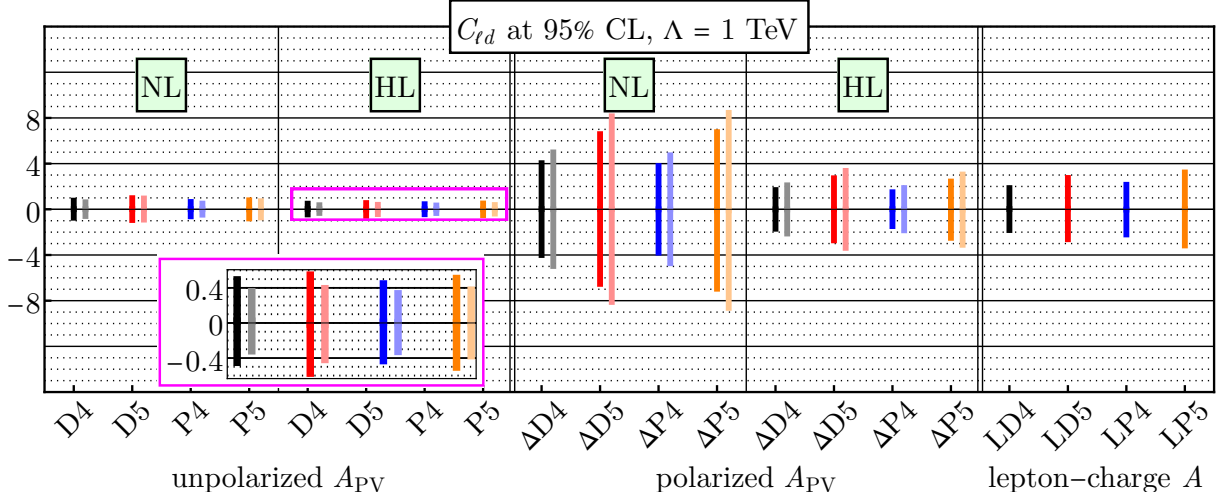


Figure 3.8. The same as Figure 3.5 but for  $C_{ld}$ .

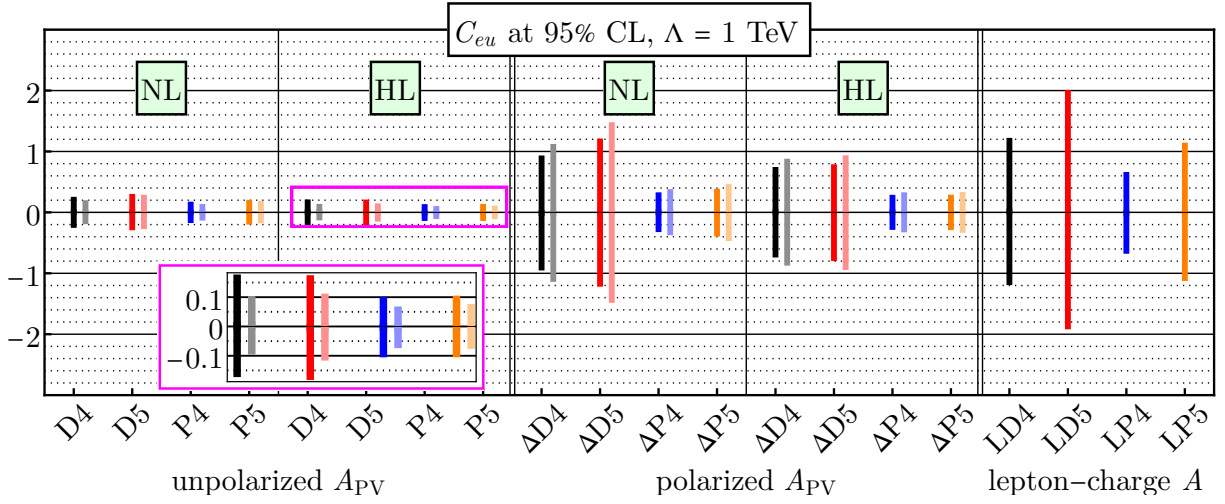
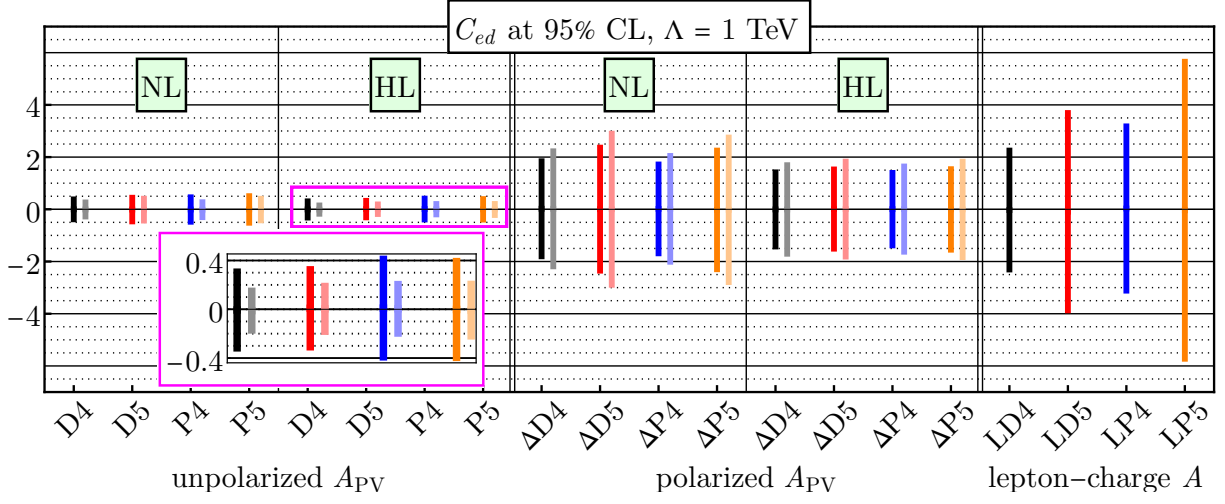
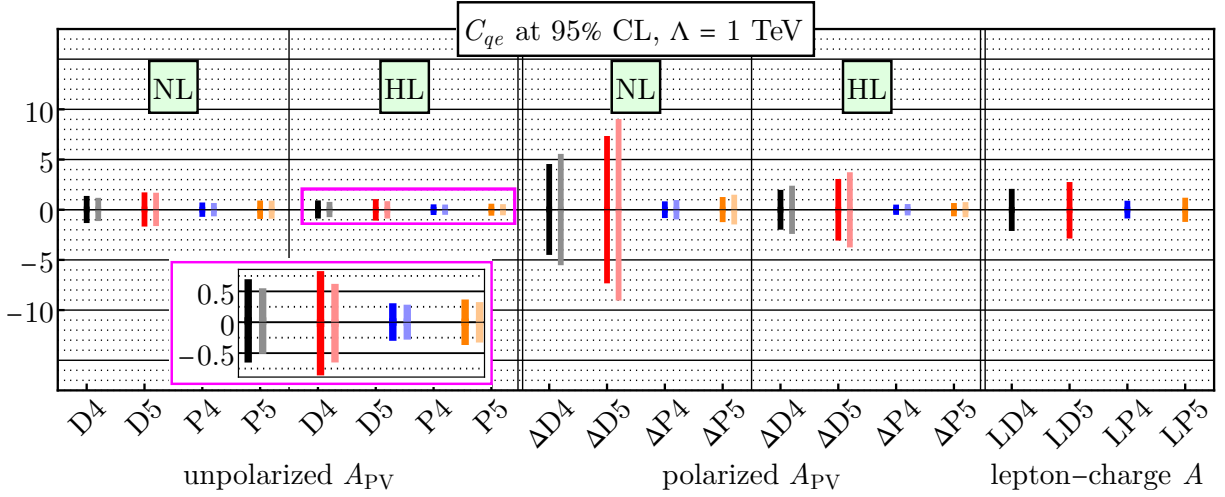


Figure 3.9. The same as Figure 3.5 but for  $C_{eu}$ .

corresponds to the fit where the beam polarization parameter  $P$  is also included. The

Figure 3.10. The same as Figure 3.5 but for  $C_{ed}$ .Figure 3.11. The same as Figure 3.5 but for  $C_{ge}$ .

color scheme is as follows: black for the D4-family (D4,  $\Delta$ D4, LD4), red for D5-family, blue for P4-family, and orange for P5-family.

From these figures, we note the common patterns. Across all observables, proton targets lead to stronger bounds than deuteron targets. The higher-energy but lower-luminosity datasets, D5 and P5, yield weaker constraints than the less-energetic but higher-luminosity D4 and P4 datasets. Among the observables, unpolarized PV asymmetries yield the strongest bounds, followed by polarized PV, and then the LC asymmetries. The HL-EIC scenario results in a noticeable improvement in sensitivity, particularly for the deuteron datasets.

For the unpolarized PV asymmetries, we find that the SMEFT parameters and the beam polarization parameter are strongly correlated, with correlation coefficients typically greater than 0.7. Including  $P$  in the fit leads to a 30 to 50% improvement in the bounds on the Wilson coefficients. In contrast, for the polarized PV asymmetries, the correlation between the SMEFT parameters and  $P$  is weaker, generally below 0.2, and the bounds weaken by 15 to 20%. Since the potential gain outweighs the loss, including  $P$  in the fit can be justified. For fits involving  $A_{\text{lum}}$ , the correlation with the SMEFT parameters is moderate, around 0.4, and the resulting bounds become 15 to 20% weaker. This suggests that treating  $A_{\text{lum}}$  as a fit parameter does not offer a meaningful improvement.

Figures 3.12–3.18 show the effective ultraviolet scales corresponding to these 95% confidence intervals. The same grouping and color scheme are used. Darker bars correspond to  $C$ -only fits, while lighter bars represent  $C + P$  fits.

With nominal annual luminosity, the EIC can probe scales up to 3 TeV in some channels. With the 10-fold HL-EIC upgrade, the reach extends up to 4 TeV.

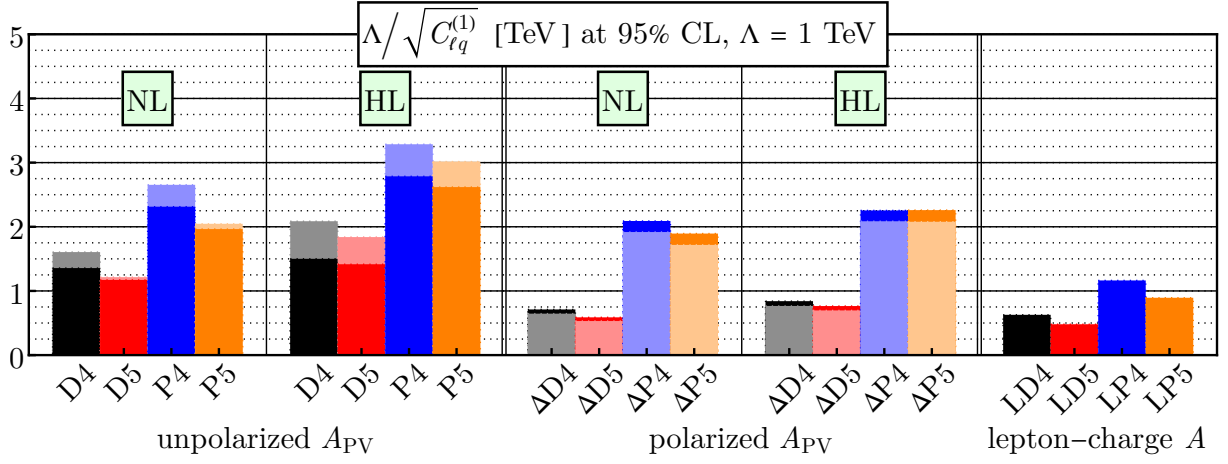


Figure 3.12. Effective UV cut-off scales,  $\Lambda/\sqrt{C_{\ell q}^{(1)}}$ , defined in terms of the 95% CL bounds on the Wilson coefficient  $C_{\ell q}^{(1)}$  and with  $\Lambda = 1$  TeV.

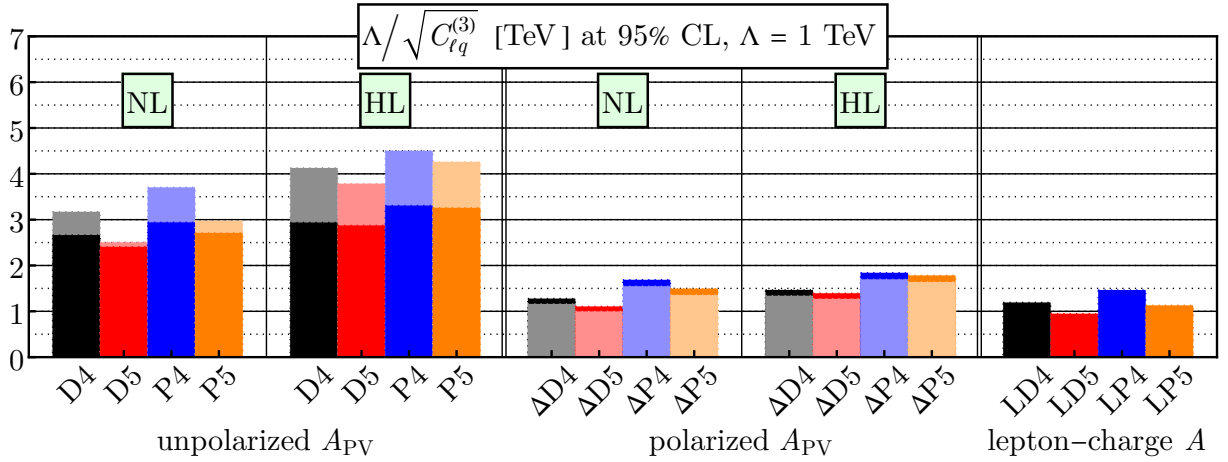


Figure 3.13. The same as Figure 3.12 but for  $C_{\ell q}^{(3)}$ .

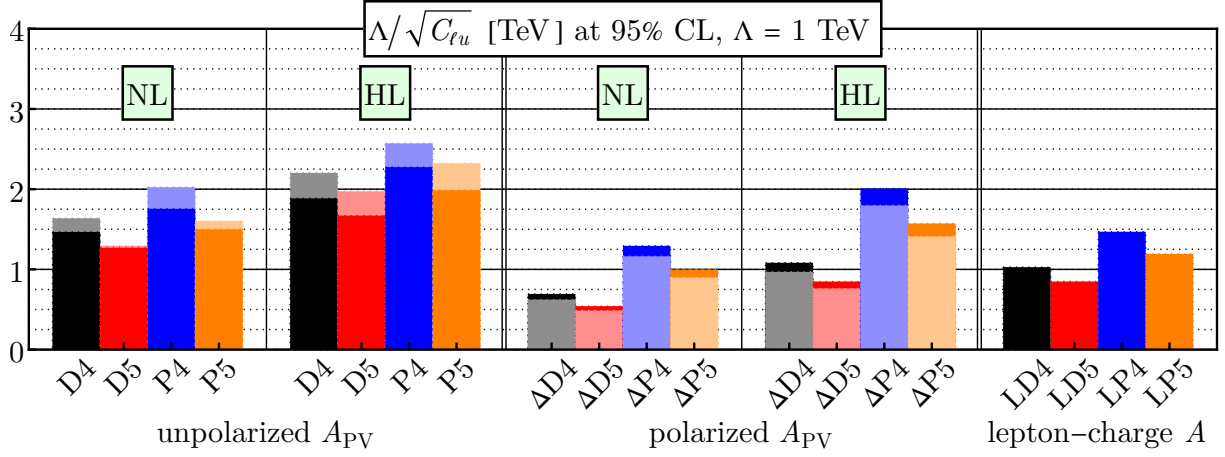


Figure 3.14. The same as Figure 3.12 but for  $C_{\ell u}$ .

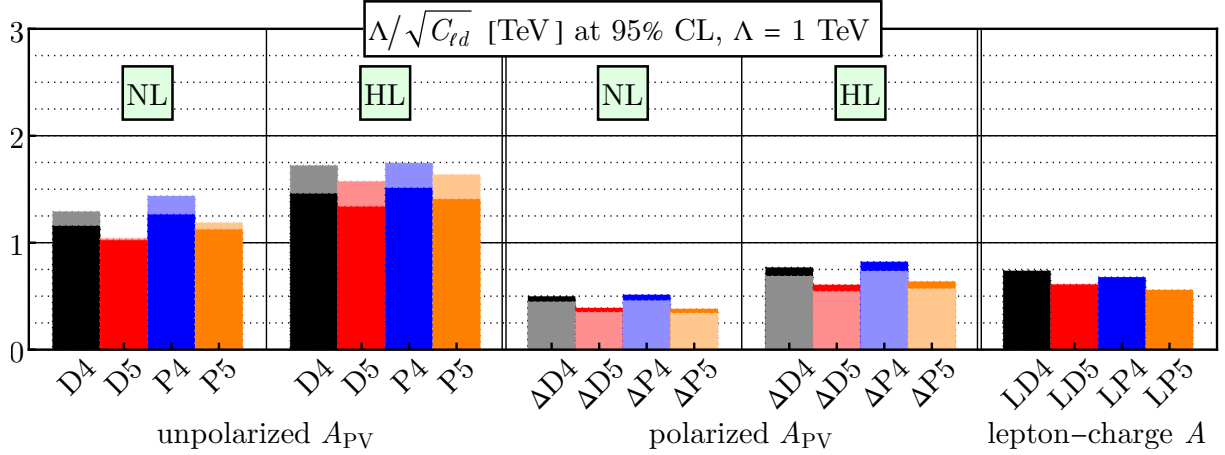


Figure 3.15. The same as Figure 3.12 but for  $C_{\ell d}$ .

### 3.5.2. Double Wilson coefficients

We now turn to the results of the two-parameter fits, where pairs of Wilson coefficients are varied simultaneously. These fits include beam polarization as a nuisance parameter.



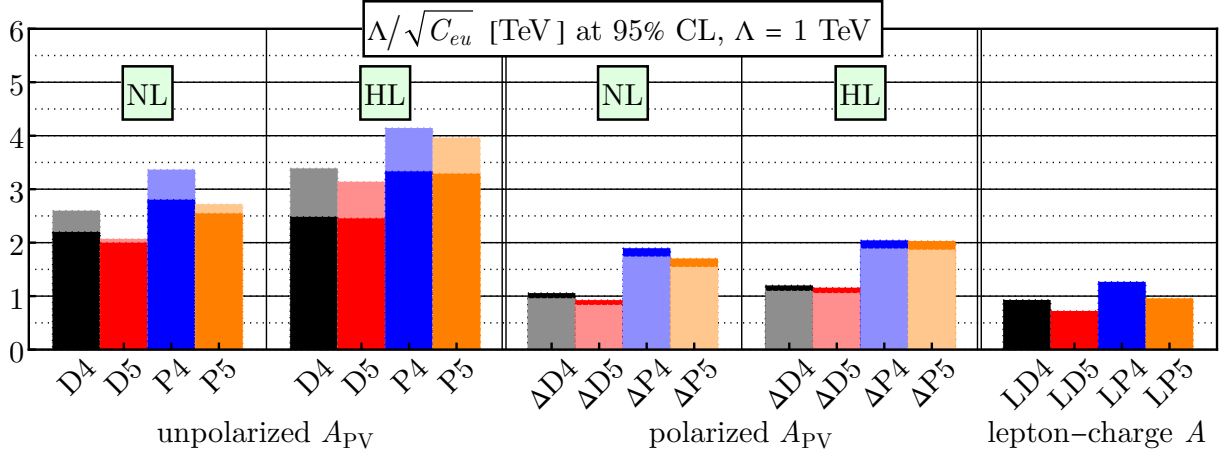


Figure 3.16. The same as Figure 3.12 but for  $C_{eu}$ .

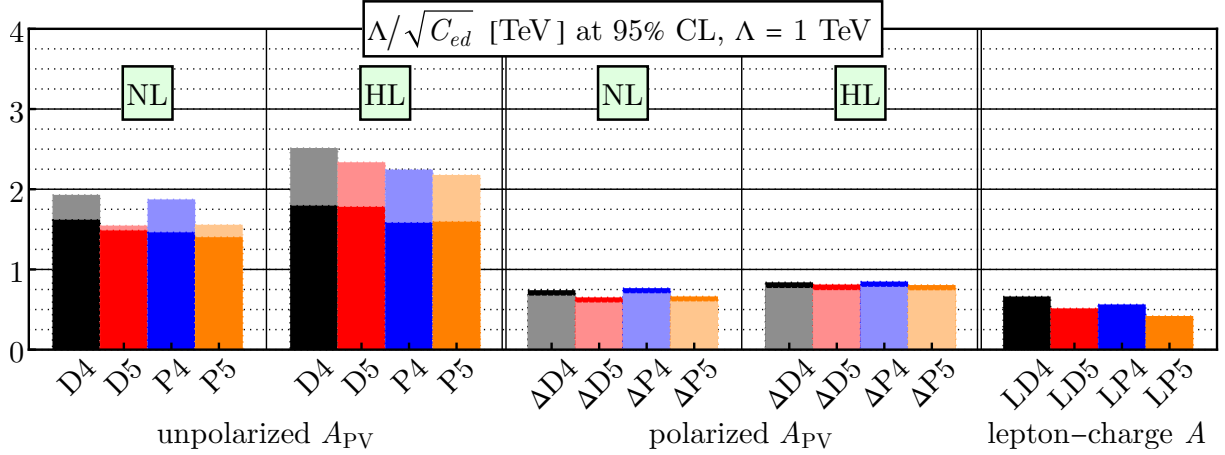


Figure 3.17. The same as Figure 3.12 but for  $C_{ed}$ .

A known limitation of NC DY measurements at the LHC is that they suffer from parameter degeneracies in this sector [47, 48]. One of the aims of this study is to test whether the EIC can lift those degeneracies.

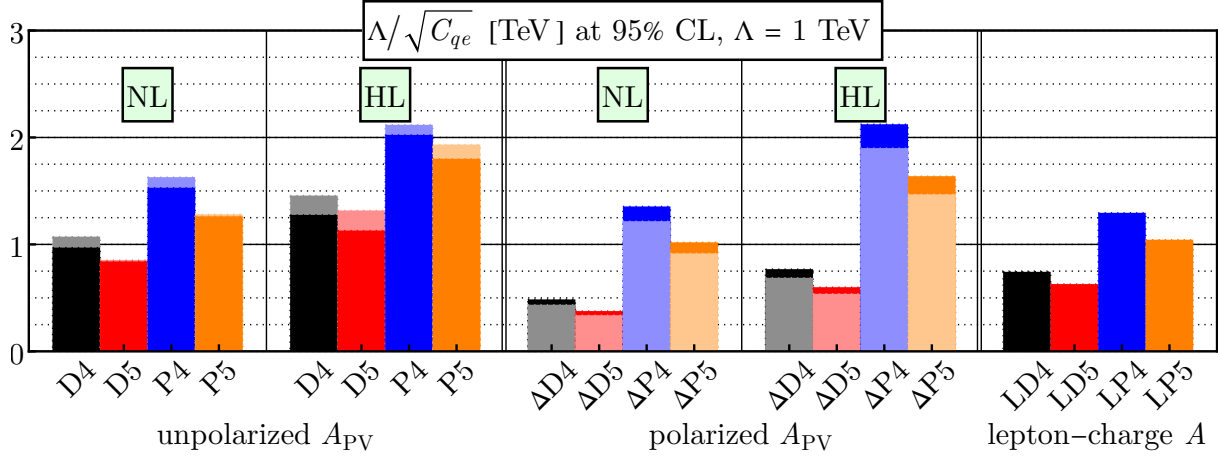


Figure 3.18. The same as Figure 3.12 but for  $C_{qe}$ .

Figure 3.19 shows the joint constraints on  $(C_{eu}, C_{qe})$  using datasets from the D4 and P4 families. Each observable leads to a distinct correlation pattern, and the resulting constraints are complementary. The LC asymmetry produces elongated ellipses, while the unpolarized PV asymmetry gives the tightest bounds. As in the single-parameter case, proton data are significantly more constraining than deuteron data.

Figure 3.20 compares the EIC reach on  $(C_{eu}, C_{lu})$  using D4 and P4 datasets against the LHC bound adapted from [48]. The LHC result is based on 8 TeV, 20 fb<sup>-1</sup> NC DY data. That measurement exhibits a clear flat direction in this parameter subspace, namely this specific parameter combination cannot be resolved fully using DY observables alone.

Figure 3.21 presents a similar comparison for  $(C_{eu}, C_{\ell q}^{(1)})$ , using both nominal and high-luminosity P4 datasets at the EIC. In this case, the EIC not only resolves the flat direction but also provides stronger bounds with visibly distinct correlation contours.

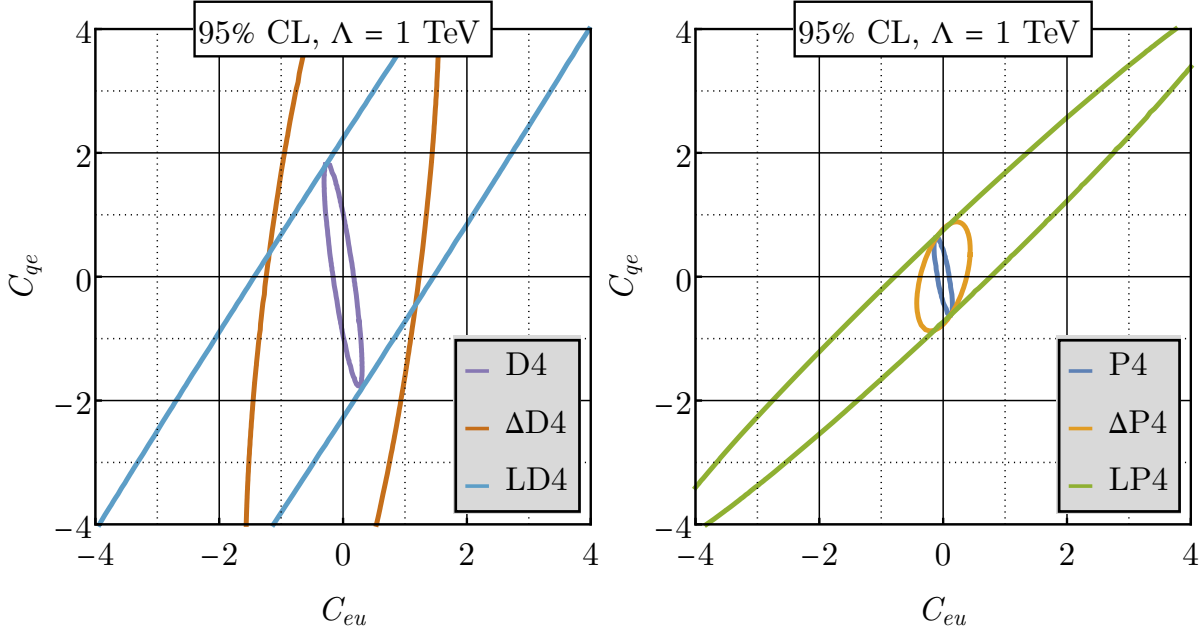


Figure 3.19. 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{qe}$  using the families of data sets D4 and P4 in the simultaneous  $(2 + 1)$ -parameter fits that includes the beam polarization as an additional fitting parameter.

In Figure 3.22, we show the results for  $(C_{\ell q}^{(1)}, C_{\ell q}^{(3)})$  using P4 at the EIC, compared to LHC data from [47]. This particular subspace is already tightly constrained by the DY data at the LHC, but we find that the EIC can lead to even stronger bounds. We also include the result of a combined fit, which highlights the complementarity of the two datasets.

The deuteron data at the EIC also exhibits flat directions for certain pairs of SMEFT parameters, such as  $(C_{eu}, C_{ed})$  and  $(C_{lu}, C_{ld})$ . This behavior can be understood analytically. In the analytical expression for asymmetries, these coefficients appear in linear combinations, for example,  $2C_{eu} - C_{ed}$ . Such combinations prevent the individual coefficients from being disentangled using deuteron data alone. However, the degeneracy is

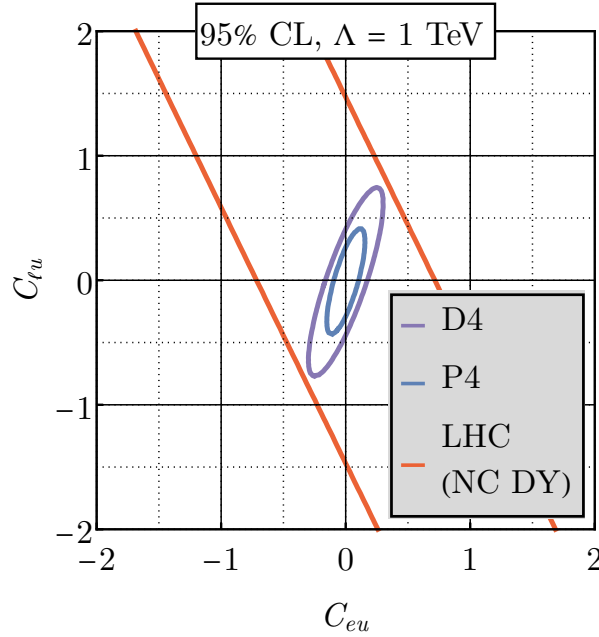


Figure 3.20. 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{qe}$  using the data sets D4 and P4 in the  $(2 + 1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [48].

specific to the deuteron target and does not appear in the corresponding proton datasets. As a result, the flat direction is lifted when data from different hadron beams are combined. This highlights the importance of a physics program that includes multiple hadron species at various energies at the EIC.

Let's summarize our findings. Proton data consistently yield tighter bounds than deuteron data. Among observables, unpolarized PV asymmetries provide the strongest constraints. When taken together, the three types of observables, namely unpolarized and polarized PV, and LC asymmetries, from both deuteron and proton collisions, as well as the NC DY data at the LHC, form a complementary set. Each contributes a distinct

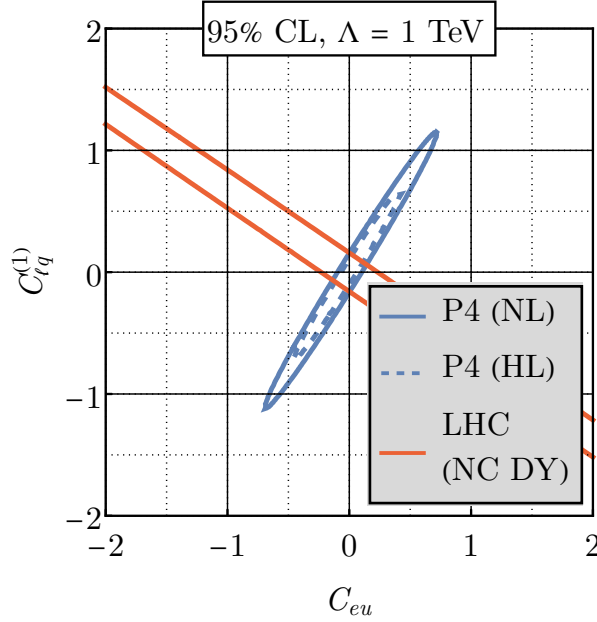


Figure 3.21. 95% CL ellipses for the Wilson coefficients  $C_{eu}$  and  $C_{\ell q}^{(1)}$  using the nominal- and high-luminosity data set P4 in the  $(2 + 1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [47].

correlation pattern. The EIC is capable of resolving all the flat directions that persist in LHC DY data. In several subspaces, the EIC even outperforms the LHC, highlighting its important role in future SMEFT analyses.

### 3.6. Coda

We conclude with a brief summary of the methodology and findings of this study. Our goal was to assess the BSM sensitivity of the EIC. We adopted the model-independent

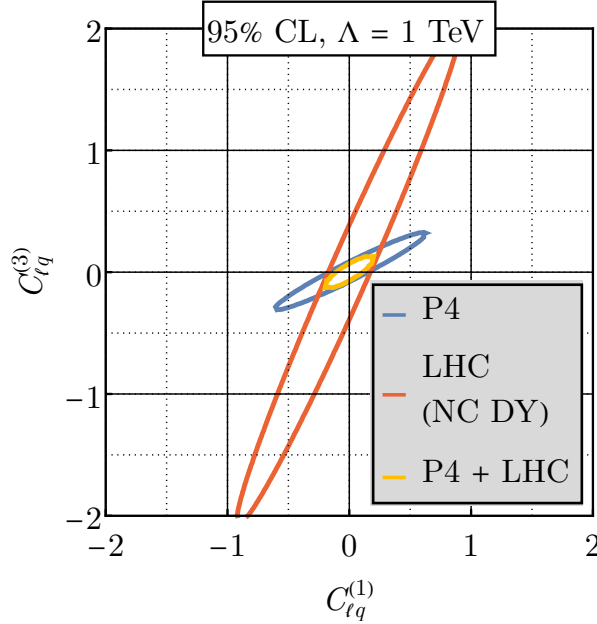


Figure 3.22. 95% CL ellipses for the Wilson coefficients  $C_{\ell q}^{(1)}$  and  $C_{\ell q}^{(3)}$  using the nominal-luminosity data set P4 in the (2+1)-parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding fit from the LHC data [47] and the combined fit of the two.

SMEFT approach, focusing on semi-leptonic four-fermion operators. The analysis incorporated a detailed treatment of anticipated experimental and theoretical uncertainties, and explored the potential improvement in sensitivity through simultaneous fits of Wilson coefficients with beam polarization and luminosity normalization parameters.

Our results show that the EIC can probe ultraviolet scales above 3 TeV using nominal integrated luminosity per year. With a 10-fold luminosity upgrade, the reach extends beyond 4 TeV. The most stringent constraints are obtained from polarized electron beams scattering off unpolarized protons. These results are complementary to, and in some cases competitive with, existing LHC bounds. In particular, the EIC's clean environment and

different initial states allow it to disentangle operator degeneracies that remain unresolved in hadron collider data.

Although the EIC was primarily designed as a precision QCD facility, this study demonstrates that it can also serve as a powerful probe of new physics.

## CHAPTER 4

**DISentangling SMEFT: A Few Colliders More**

*Blondie: Two hundred thousand dollars is a lot of money. We're gonna have to earn it.*

Sergio Leone, *The Good, the Bad, and the Ugly*

In this chapter, we extend our previous study of parity-violating deep inelastic scattering asymmetries at the Electron-Ion Collider to include additional simulated data from the Large Hadron-electron Collider and the Future Circular Collider. We upgrade the leading-order analysis of Chapter 3 with the framework of Standard Model Effective Field Theory by incorporating the complete set of dimension-6 operators that affect the amplitude, namely the shifts to the fermion couplings to neutral gauge bosons, and include next-to-leading order corrections from quantum chromodynamics to the structure functions. This allows us to directly compare the new physics sensitivity of all three future machines under consistent assumptions.

**4.1. Prelude**

The Standard Model (SM) is a remarkably successful theory. It accurately describes all known particles and interactions, and with the discovery of the Higgs boson, the SM particle spectrum is now complete. But its success is limited to what it includes. There is no explanation for the existence of dark matter, no mechanism to account for the matter-antimatter asymmetry, and no built-in origin for neutrino masses. There are also structural issues, such as the hierarchy problem and the vast spread of fermion



Yukawa couplings. All this is just Nature’s way of saying “in your face”. Meanwhile, in humankind’s long history of extensive and expensive experimental efforts, no new particles have been found. If new physics exists, it is likely to appear as subtle deviations in precision measurements before any new particle states are observed directly.

Electron-hadron colliders provide an ideal platform for these precision measurements. They constitute the ultimate tools for high-precision quantum chromodynamics (QCD) studies and microscopes for probing internal structures of hadrons. Electron stands out as a desirable probe to look into the proton because it doesn’t get involved in color interactions, so all the interactions are strictly electroweak (EW), which already has a solid footing with the unmatched precision of quantum electrodynamics. Furthermore, kinematics are uniquely determined by the incoming electron beam, the scattered lepton, or the hadronic final state, which all can be measured with great accuracy. These machines serve both as microscopes for QCD and as sensitive instruments for EW and BSM physics.

Historically, the only electron-hadron collider ever operated was *Hadron-Electron-Ringanlage* (HERA), which ran at DESY in Germany between 1991 and 2007. Since then, three next-generation facilities have been proposed or are under development. The Electron-Ion Collider (EIC) [12] is a United States (U.S.) Department of Energy (DOE) project now under construction at Brookhaven National Laboratory. It will be first high-energy deep inelastic scattering (DIS) machine to collide polarized electrons with polarized protons and ions at center-of-mass (c.m.) energies between 70 and 140 GeV, which is a range between fixed-target-scattering and high-energy collider experiments. It is anticipated to start operating within a decade. Electron beams will have energies 5 to 18 GeV, proton beams 41 to 275 GeV, light ions up to 166 GeV, and heavy ions up to 110 GeV.

With yearly integrated luminosities up to  $100 \text{ fb}^{-1}$  and high detector acceptance and efficiency rates, it will help improve extraction of parity-violating (PV) DIS asymmetries in EW neutral-current (NC) DIS cross section with reduced uncertainties. Just to give a timeline of the history of the EIC, an electron-ion collider of this caliber was first offered in 2012. In 2015, the U.S. DOE officially named the machine the EIC and released the public announcement of construction at Brookhaven in 2020. The construction will start at the end of 2025 and is planned to be completed by 2040.

The Large Hadron-electron Collider (LHeC) [11] is a proposed upgrade of the Large Hadron Collider (LHC) at CERN in Switzerland, which is awaiting approval at the time of writing. It would run simultaneously with the LHC, using its proton and ion beams, and a new, dedicated electron beamline, reaching c.m. energies of 1.5 TeV. The planned integrated luminosity is  $100 \text{ fb}^{-1}$ . Its primary design purpose is novel measurements in QCD, DIS physics at low Bjorken- $x$ , EW precision studies, and BSM physics. Historically, the idea of an electron-proton collider in the LEP-LHC tunnel was discussed for the first time in 1984, which is also the year HERA was approved. In 2005, it was found feasible to simultaneously run  $pp$  collisions in the LHC and the  $e^-p$  collisions in the new machine named LHeC. The first complete draft of the conceptual design was published in 2011. Its earliest operational period is estimated to be around 2032, which would coincide with the LHC Run 5.

The Future Circular Collider (FCC) [8, 34, 35, 36] would be a brand new collider to be built at CERN; however, it will take at least 30 years for design and construction. The designed collider energy is 3.5 TeV, and the planned integrated luminosity is at the order of  $1 \text{ ab}^{-1}$ . It would have a broad physics program similar to the LHeC's, namely QCD and

EW precision studies and new physics searches. As the history unfolds, European studies of post-LHC circular accelerators that are at energy frontiers were published between 2010 and 2013. In 2014, these efforts were combined into the FCC study. The first complete draft of the conceptual design was released in 2019. The feasibility studies were very recently published in May, 2025. The design and construction are expected to be completed by 2050.

In this work, we explore the potential of these future machines to probe physics beyond the Standard Model (BSM) by presenting the complete set of uncertainties projected by people who know better. Our observable is the PV asymmetries at the EIC based on our previous study in Chapter 3, and the NC DIS cross section at the LHeC and FCC-eh. Since there has been no definitive sign of new particles beyond the current spectrum of the SM, we use the Standard Model Effective Field Theory as our BSM framework. We consider the full spectrum of SMEFT operators that contribute to the NC DIS amplitude at leading order in SMEFT couplings at dimension 6, namely the good old semi-leptonic four-fermion operators, as well as operators that modify the fermion couplings to NC EW gauge bosons.

The best method to measure and constrain  $ffV$  vertex corrections is via  $Z$ -pole EW precision observables (EWPO) at LEP and SLC; however, data being limited yields degeneracies among SMEFT parameters. This is illustrated in [92]. In the current study, we demonstrate that DIS measurements at the future colliders can not only resolve these degeneracies but also impose more stringent constraints.

We find that for semi-leptonic four-fermion operators, the LHeC and FCC-eh can probe effective ultraviolet (UV) scales exceeding 10 TeV, whilst the reach of the EIC remains at a

few TeV. We include all possibilities in our analysis: beam energy, polarization, luminosity, and lepton species. No single choice of a configuration is sufficient to probe the parameter space fully. Furthermore, a positron beam can significantly extend the UV reach for certain Wilson coefficients due to the structure of the underlying amplitude. In addition, the EWPO global fits can give the most stringent bounds; however, that is true for merely single-parameter fits. Multi-parameter fits paints a picture, namely the allowed bounds of the SMEFT parameter get weakened by an order due to strong correlations, which signals flat directions. At this very point, the LHeC and FCC-eh become important.

This chapter is organized as follows. In Section 4.2, we review the SMEFT framework, define the operator basis, and summarize the relevant kinematic structure of DIS. Section 4.3 describes the observables used and the simulated pseudodata sets in terms of their run parameters, anticipated uncertainties, and error budgets for the SMEFT fits, as well as our fitting procedure. In Section 4.4, we present first our fit results for the semi-leptonic four-fermion operators, discussing the impact of different runs, luminosities, and lepton species. In Section 4.5, we activate all 17 operators and shift the focus to the  $ffV$  vertex corrections. We compare our results to those from global EWPO fits and highlight where the LHeC and FCC-eh improve upon them. Section 4.6 concludes with a summary of the main findings and their implications for the SMEFT program.

## 4.2. Formalism

### 4.2.1. The SMEFT Lagrangian

The SMEFT provides a systematic expansion of the SM Lagrangian in inverse powers of a heavy new physics scale  $\Lambda$ , which is assumed to be above the SM particle masses and

beyond collider reach. One builds operators,  $O_k^{(n)}$ , of mass dimension higher than  $n > 4$  using the existing SM spectrum:

$$(4.1) \quad \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{n>4} \frac{1}{\Lambda^{n-4}} \sum_k C_k^{(n)} O_k^{(n)},$$

where Wilson coefficients  $C_k^{(n)}$  are introduced as effective coupling strengths. In this study, we restrict ourselves to the case of dimension 6. We investigate only the leading-order effects of the SMEFT operators, so we retain only the SM-SMEFT interference amplitudes and discard the squared SMEFT contributions, and therefore all observables are linearized in the Wilson coefficients.

There are 17 dimension-6 operators in the Warsaw basis [85] that contribute to NC DIS. These include 7 semi-leptonic four-fermion operators, and 10 operators that shift the fermion couplings to the neutral EW gauge bosons. The operators are listed in Table 4.1. Here,  $\varphi$  is the SU(2) Higgs doublet,  $\ell$  and  $q$  are left-handed lepton and quark doublets,  $e$ ,  $u$ , and  $d$  are the right-handed electron, up quark, and down quark singlets, respectively, the  $\tau^I$  are the Pauli matrices, and the double-arrow covariant derivative is defined such that

$$(4.2) \quad \varphi^\dagger i \overleftrightarrow{D}^\mu (\tau^I) \varphi = \varphi^\dagger i D^\mu (\tau^I) \varphi + \text{h.c.}.$$

Operators involving dipoles or scalar bilinears are neglected because the corresponding vertex factors are proportional to fermion masses, which we assume zero. We also assume flavor universality and hence suppress flavor indices.

Table 4.1. Dimension-6 SMEFT operators in the Warsaw basis [85] that contribute to the NC DIS amplitudes at leading order. The 10 operators that modify  $ffV$  vertices are shown on the left, and the 7 operators that introduce semi-leptonic four-fermion contact interactions are presented on the right.

$ffV$	semi-leptonic four-fermion
$O_{\varphi WB} = (\varphi^\dagger \tau^I \varphi) W_{\mu\nu}^I B^{\mu\nu}$	$O_{\ell q}^{(1)} = (\bar{\ell} \gamma^\mu \ell)(\bar{q} \gamma_\mu q)$
$O_{\varphi D} = (\varphi^\dagger D^\mu \varphi)^* (\varphi^\dagger D_\mu \varphi)$	$O_{\ell q}^{(1)} = (\bar{\ell} \gamma^\mu \tau^I \ell)(\bar{q} \gamma_\mu \tau^I q)$
$O_{\varphi \ell}^{(1)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{\ell} \gamma_\mu \ell)$	$O_{eu} = (\bar{e} \gamma^\mu e)(\bar{u} \gamma_\mu u)$
$O_{\varphi \ell}^{(3)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \tau^I \varphi)(\bar{\ell} \gamma_\mu \tau^I \ell)$	$O_{ed} = (\bar{e} \gamma^\mu e)(\bar{d} \gamma_\mu d)$
$O_{\varphi e} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{e} \gamma_\mu e)$	$O_{\ell u} = (\bar{\ell} \gamma^\mu \ell)(\bar{u} \gamma_\mu u)$
$O_{\varphi q}^{(1)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{q} \gamma_\mu q)$	$O_{\ell d} = (\bar{\ell} \gamma^\mu \ell)(\bar{d} \gamma_\mu d)$
$O_{\varphi q}^{(3)} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \tau^I \varphi)(\bar{q} \gamma_\mu \tau^I q)$	$O_{qe} = (\bar{e} \gamma^\mu e)(\bar{q} \gamma_\mu q)$
$O_{\varphi u} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{u} \gamma_\mu u)$	
$O_{\varphi d} = (\varphi^\dagger i \overleftrightarrow{D}^\mu \varphi)(\bar{d} \gamma_\mu d)$	
$O_{\ell \ell} = (\bar{\ell} \gamma^\mu \ell)(\bar{\ell} \gamma_\mu \ell)$	

#### 4.2.2. DIS and Structure Functions

We study the NC DIS in the process  $\ell + H \rightarrow \ell' + X$ , where  $\ell$  is an electron or a positron,  $H$  can be proton or deuteron, and  $\ell'$  and  $X$  are the final-state lepton and hadron, respectively, within the framework of the SMEFT including next-to-leading-order (NLO) QCD corrections. One has to worry about the missing energy and the reconstruction of hadronic final states to determine kinematic variables in the CC DIS. This typically

brings in larger systematic uncertainties, so we find it feasible not to dedicate time to the CC DIS studies.

At parton level to leading order, said process can be mediated by single photon or Z-boson exchange or by the SMEFT contact interaction of leptons with partons. The tree-level Feynman diagrams are presented in Figure 4.1.

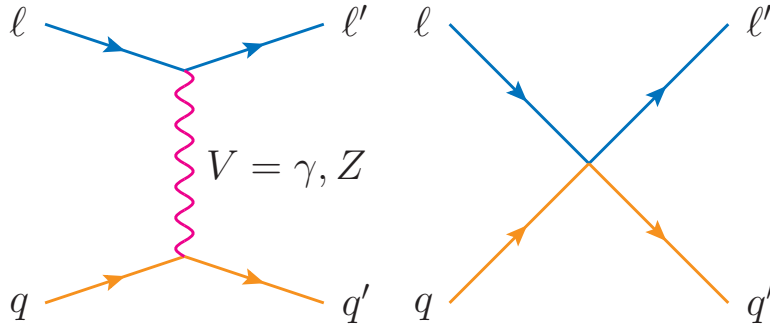


Figure 4.1. Feynman diagrams at tree level for the underlying partonic process of the scattering  $\ell + H \rightarrow \ell' + X$ .

NLO QCD corrections to the SM process are well known [82, 22, 151, 21, 83]. These corrections modify only the quark lines, as illustrated in Figure 4.2; therefore, said corrections are identical for both SM and SMEFT cross sections.

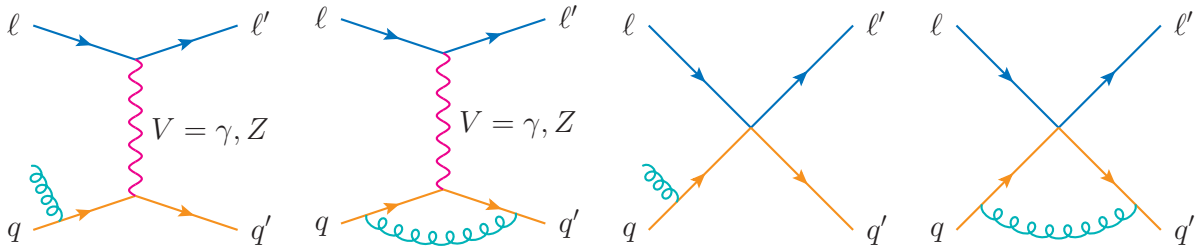


Figure 4.2. Feynman diagrams describing NLO QCD corrections to the underlying partonic process of the scattering  $\ell + H \rightarrow \ell' + X$ .

The NC DIS cross-section expressions for collisions of a lepton  $\ell$  with an unpolarized or polarized hadron are given in terms of the NC structure functions  $F_{1,3,L}^{\text{NC}}$  and  $g_{1,5,L}^{\text{NC}}$  by

$$(4.3) \quad \frac{d^2 \sigma_{\text{NC}}^\ell}{dx dQ^2} = \frac{2\pi\alpha^2}{xQ^4} \left\{ [1 + (1-y)^2] 2xF_1^{\text{NC}} + \text{sgn}(\ell)[1 - (1-y)^2] xF_3^{\text{NC}} + (1-y) 2F_L^{\text{NC}} \right\},$$

and

$$(4.4) \quad \frac{d^2 \Delta \sigma_{\text{NC}}^\ell}{dx dQ^2} = \frac{8\pi\alpha^2}{xQ^4} \left\{ [1 + (1-y)^2] xg_5^{\text{NC}} - \text{sgn}(\ell)[1 - (1-y)^2] xg_1^{\text{NC}} + (1-y) g_L^{\text{NC}} \right\},$$

respectively, where  $\text{sgn}$  is the *particle signum* function that returns +1 for particles and -1 for antiparticles. The dimension-6 SMEFT NC DIS structure functions  $F_{1,3,L}^{\text{NC},e^\pm}$  and  $g_{1,5,L}^{\text{NC},e^\pm}$  are given in terms of the structure functions  $F_{1,3,L}^{\tilde{V}}$  and  $g_{1,5,L}^{\tilde{V}}$  by

$$(4.5) \quad (\Delta) \mathcal{S}_i^{\text{NC},e^\pm} = \left\{ \begin{array}{l} (\Delta) \Phi_i^{\text{NC},e^\pm,\gamma} \mathcal{S}_i^\gamma \\ + (\Delta) \Phi_i^{\text{NC},e^\pm,Z} \mathcal{S}_i^Z \\ + (\Delta) \Phi_i^{\text{NC},e^\pm,\gamma Z} \mathcal{S}_i^{\gamma Z} \\ \mp (\lambda_e - 1) (\Delta) \Phi_i^{\text{NC},e^\pm,\gamma \times \pm} \mathcal{S}_i^{\gamma \times \pm} \\ \pm (\lambda_e + 1) (\Delta) \Phi_i^{\text{NC},e^\pm,\gamma \times \mp} \mathcal{S}_i^{\gamma \times \mp} \\ \mp (\lambda_e - 1) (\Delta) \Phi_i^{\text{NC},e^\pm,Z \times \pm} \mathcal{S}_i^{Z \times \pm} \\ \pm (\lambda_e + 1) (\Delta) \Phi_i^{\text{NC},e^\pm,Z \times \mp} \mathcal{S}_i^{Z \times \mp} \end{array} \right\},$$

where  $\mathcal{S}_i^{\text{NC},e^\pm} = F_i^{\text{NC},e^\pm}$  and  $\Delta \mathcal{S}_i^{\text{NC},e^\pm} = g_i^{\text{NC},e^\pm}$ , and  $\lambda_e = \pm 1$  is the incoming lepton helicity sign. The *structure factors*  $(\Delta) \Phi_i^{\text{NC},e^\pm,\tilde{V}}$  are given by

$$(4.6) \quad \Phi_{1,L}^{\text{NC},e^\pm,\gamma} = Q_{e\gamma}^2,$$



$$(4.7) \quad \Phi_{1,L}^{\text{NC},e^\pm,Z} = \eta_{\gamma Z}^2 \left( (\lambda_e + 1) \left( Q_{eZ}^{L/R} \right)^2 - (\lambda_e - 1) \left( Q_{eZ}^{R/L} \right)^2 \right),$$

$$(4.8) \quad \Phi_{1,L}^{\text{NC},e^\pm,\gamma Z} = \pm Q_{e\gamma} \eta_{\gamma Z} \left( (\lambda_e \pm 1) Q_{eZ}^L - (\lambda_e \mp 1) Q_{eZ}^R \right),$$

$$(4.9) \quad \Phi_{1,L}^{\text{NC},e^\pm,\gamma \times +} = \mp Q_{e\gamma},$$

$$(4.10) \quad \Phi_{1,L}^{\text{NC},e^\pm,\gamma \times -} = \mp Q_{e\gamma},$$

$$(4.11) \quad \Phi_{1,L}^{\text{NC},e^\pm,Z \times +} = \mp \eta_{\gamma Z} Q_{eZ}^R,$$

$$(4.12) \quad \Phi_{1,L}^{\text{NC},e^\pm,Z \times -} = \mp \eta_{\gamma Z} Q_{eZ}^L,$$

$$(4.13) \quad \Phi_3^{\text{NC},e^\pm,\gamma} = 0,$$

$$(4.14) \quad \Phi_3^{\text{NC},e^\pm,Z} = \pm \eta_{\gamma Z}^2 \left( (\lambda_e \pm 1) \left( Q_{eZ}^L \right)^2 + (\lambda_e \mp 1) \left( Q_{eZ}^R \right)^2 \right),$$

$$(4.15) \quad \Phi_3^{\text{NC},e^\pm,\gamma Z} = \pm Q_{e\gamma} \eta_{\gamma Z} \left( (\lambda_e \pm 1) Q_{eZ}^L + (\lambda_e \mp 1) Q_{eZ}^R \right),$$

$$(4.16) \quad \Phi_3^{\text{NC},e^\pm,\gamma \times +} = \mp Q_{e\gamma},$$

$$(4.17) \quad \Phi_3^{\text{NC},e^\pm,\gamma \times -} = \mp Q_{e\gamma},$$

$$(4.18) \quad \Phi_3^{\text{NC},e^\pm,Z \times +} = \mp \eta_{\gamma Z} Q_{eZ}^R,$$

$$(4.19) \quad \Phi_3^{\text{NC},e^\pm,Z \times -} = \mp \eta_{\gamma Z} Q_{eZ}^L,$$

and

$$(4.20) \quad \Delta \Phi_{5,L}^{\text{NC},e^\pm,\gamma} = 0,$$

$$(4.21) \quad \Delta \Phi_{5,L}^{\text{NC},e^\pm,Z} = \lambda_q \left( \pm \eta_{\gamma Z}^2 \right) \left( (\lambda_e \pm 1) \left( Q_{eZ}^L \right)^2 - (\lambda_e \mp 1) \left( Q_{eZ}^R \right)^2 \right),$$

$$(4.22) \quad \Delta \Phi_{5,L}^{\text{NC},e^\pm,\gamma Z} = \lambda_q \left( \pm Q_{e\gamma} \right) \eta_{\gamma Z} \left( (\lambda_e \pm 1) Q_{eZ}^L - (\lambda_e \mp 1) Q_{eZ}^R \right),$$

$$(4.23) \quad \Delta \Phi_{5,L}^{\text{NC},e^\pm,\gamma \times +} = \lambda_q \left( \pm Q_{e\gamma} \right),$$

$$(4.24) \quad \Delta\Phi_{5,L}^{\text{NC},e^\pm,\gamma\times-} = \mp\lambda_q Q_{e\gamma},$$

$$(4.25) \quad \Delta\Phi_{5,L}^{\text{NC},e^\pm,Z\times+} = \lambda_q (\pm\eta_{\gamma Z}) Q_{eZ}^R,$$

$$(4.26) \quad \Delta\Phi_{5,L}^{\text{NC},e^\pm,Z\times-} = \mp\lambda_q \eta_{\gamma Z} Q_{eZ}^L,$$

$$(4.27) \quad \Delta\Phi_1^{\text{NC},e^\pm,\gamma} = \lambda_e \lambda_q (\pm Q_{e\gamma}^2),$$

$$(4.28) \quad \Delta\Phi_1^{\text{NC},e^\pm,Z} = \lambda_q (\pm\eta_{\gamma Z}^2) ((\lambda_e \pm 1) (Q_{eZ}^L)^2 + (\lambda_e \mp 1) (Q_{eZ}^R)^2),$$

$$(4.29) \quad \Delta\Phi_1^{\text{NC},e^\pm,\gamma Z} = \lambda_q (\pm Q_{e\gamma}) \eta_{\gamma Z} ((\lambda_e \pm 1) Q_{eZ}^L + (\lambda_e \mp 1) Q_{eZ}^R),$$

$$(4.30) \quad \Delta\Phi_1^{\text{NC},e^\pm,\gamma\times+} = \lambda_q (\pm Q_{e\gamma}),$$

$$(4.31) \quad \Delta\Phi_1^{\text{NC},e^\pm,\gamma\times-} = \mp\lambda_q Q_{e\gamma},$$

$$(4.32) \quad \Delta\Phi_1^{\text{NC},e^\pm,Z\times+} = \lambda_q (\pm\eta_{\gamma Z}) Q_{eZ}^R,$$

$$(4.33) \quad \Delta\Phi_1^{\text{NC},e^\pm,Z\times-} = \mp\lambda_q \eta_{\gamma Z} Q_{eZ}^L.$$

The structure functions are given by

$$(4.34) \quad F_1^{\tilde{V}}(x, Q) = \frac{1}{2} \sum_q \lambda_V^{\tilde{V}} q_1(x, Q),$$

$$(4.35) \quad F_3^{\tilde{V}}(x, Q) = \sum_q \text{sgn}(q) (-\lambda_A^{\tilde{V}}) q_3(x, Q),$$

$$(4.36) \quad F_L^{\tilde{V}}(x, Q) = x \sum_q \text{sgn}(q) \lambda_V^{\tilde{V}} q_L(x, Q),$$

$$(4.37) \quad g_1^{\tilde{V}}(x, Q) = \frac{1}{2} \sum_q \lambda_V^{\tilde{V}} \Delta q_1(x, Q),$$

$$(4.38) \quad g_5^{\tilde{V}}(x, Q) = \frac{1}{2} \sum_q \text{sgn}(q) \lambda_A^{\tilde{V}} \Delta q_5(x, Q),$$

$$(4.39) \quad g_L^{\tilde{V}}(x, Q) = \frac{1}{2} \sum_q \text{sgn}(q) \lambda_A^{\tilde{V}} \Delta q_L(x, Q),$$

where  $\tilde{V} = \gamma, Z, \gamma Z, \gamma \times +, \gamma \times -, Z \times +, Z \times -$  is the channel label and

$$(4.40) \quad q_1(x, Q) = q(x, Q) + \frac{\alpha_s(Q)}{2\pi} \left[ \tilde{C}_{q_1} \otimes q(x, Q) + \tilde{C}_{g_1} \otimes g(x, Q) \right],$$

$$(4.41) \quad q_3(x, Q) = q(x, Q) + \frac{\alpha_s(Q)}{2\pi} \left[ \tilde{C}_{q_3} \otimes q(x, Q) + \tilde{C}_{g_3} \otimes g(x, Q) \right],$$

$$(4.42) \quad q_L(x, Q) = \frac{\alpha_s(Q)}{2\pi} \left[ \tilde{C}_{q_L} \otimes q(x, Q) + \tilde{C}_{g_L} \otimes g(x, Q) \right],$$

$$(4.43) \quad \Delta q_1(x, Q) = \Delta q(x, Q) + \frac{\alpha_s(Q)}{2\pi} \left[ \Delta \tilde{C}_{q_1} \otimes \Delta q(x, Q) + \Delta \tilde{C}_{g_1} \otimes \Delta g(x, Q) \right],$$

$$(4.44) \quad \Delta q_3(x, Q) = \Delta q(x, Q) + \frac{\alpha_s(Q)}{2\pi} \left[ \Delta \tilde{C}_{q_3} \otimes \Delta q(x, Q) + \Delta \tilde{C}_{g_3} \otimes \Delta g(x, Q) \right],$$

$$(4.45) \quad \Delta q_L(x, Q) = \frac{\alpha_s(Q)}{2\pi} \left[ \Delta \tilde{C}_{q_L} \otimes \Delta q(x, Q) + \Delta \tilde{C}_{g_L} \otimes \Delta g(x, Q) \right],$$

with  $(\Delta)q$  and  $(\Delta)g$  being the un(polarized) quark and gluon parton distribution functions (PDFs), respectively. The quark couplings  $\lambda_V^{\tilde{V}}$  and  $\lambda_A^{\tilde{V}}$  are given by

$$(4.46) \quad \lambda_V^\gamma = Q_{q\gamma}^2,$$

$$(4.47) \quad \lambda_V^Z = \frac{1}{4} \left( (Q_{qZ}^L)^2 + (Q_{qZ}^R)^2 \right),$$

$$(4.48) \quad \lambda_V^{\gamma Z} = \frac{1}{2} Q_{q\gamma} (Q_{qZ}^L + Q_{qZ}^R),$$

$$(4.49) \quad \lambda_V^{\gamma \times +} = \frac{Q_{q\gamma} (C_q^{RL} + C_q^{RR})}{2C_\gamma^2 P_\gamma},$$

$$(4.50) \quad \lambda_V^{\gamma \times -} = \frac{Q_{q\gamma} (C_q^{LR} + C_q^{LL})}{2C_\gamma^2 P_\gamma},$$

$$(4.51) \quad \lambda_V^{Z \times +} = \frac{C_q^{RL} Q_{qZ}^L + C_q^{RR} Q_{qZ}^R}{2C_\gamma^2 P_\gamma},$$

$$(4.52) \quad \lambda_V^{Z \times -} = \frac{C_q^{LR} Q_{qZ}^R + C_q^{LL} Q_{qZ}^L}{2C_\gamma^2 P_\gamma},$$

and

$$(4.53) \quad \lambda_A^\gamma = 0,$$

$$(4.54) \quad \lambda_A^Z = \frac{1}{4} \left( (Q_{qZ}^R)^2 - (Q_{qZ}^L)^2 \right),$$

$$(4.55) \quad \lambda_A^{\gamma Z} = \frac{1}{2} Q_{q\gamma} (Q_{qZ}^R - Q_{qZ}^L),$$

$$(4.56) \quad \lambda_A^{\gamma \times +} = \frac{Q_{q\gamma} (C_q^{RL} - C_q^{RR})}{2C_\gamma^2 P_\gamma},$$

$$(4.57) \quad \lambda_A^{\gamma \times -} = \frac{Q_{q\gamma} (C_q^{LR} - C_q^{LL})}{2C_\gamma^2 P_\gamma},$$

$$(4.58) \quad \lambda_A^{Z \times +} = \frac{C_q^{RL} Q_{qZ}^L - C_q^{RR} Q_{qZ}^R}{2C_\gamma^2 P_\gamma},$$

$$(4.59) \quad \lambda_A^{Z \times -} = \frac{C_q^{LR} Q_{qZ}^R - C_q^{LL} Q_{qZ}^L}{2C_\gamma^2 P_\gamma}.$$

The parametrization of the  $ffV$  and  $\ell\ell qq$  vertex factors in the  $L/R$  basis in our analysis is given by

$$(4.60) \quad V_{fV}^{L/R} = iC_V Q_{fV}^{L/R},$$

$$(4.61) \quad V_q^{PP'} = iC_q^{PP'},$$

where we denote the single-photon and single- $Z$  exchange and the semi-leptonic four-fermion interaction channels by  $V = \gamma, Z, \times$ , respectively, in the subscripts. The  $C_V$  and  $C_q^{PP'}$  are the coupling strenghts and  $P, P' = L, R$  indicate the structure of the lepton and

quark currents in the interaction term. To illustrate,  $C_q^{LL}$  denotes the coupling of  $\bar{\ell}\gamma_\mu P_L \ell$  and  $\bar{q}\gamma^\mu P_L q$ . The  $Q_{fV}^{L/R}$  are the SMEFT quantum numbers.

The gauge-boson propagators are defined as

$$(4.62) \quad P_V = \frac{1}{Q^2 + M_V^2}.$$

The coupling strengths are given by

$$(4.63) \quad C_\gamma = -\sqrt{4\pi\alpha},$$

$$(4.64) \quad C_Z = -2\sqrt{\sqrt{2}G_F M_Z^2},$$

$$(4.65) \quad C_{u/d}^{LL} = \frac{C_{\ell q}^{(1)} \mp C_{\ell q}^{(3)}}{\Lambda^2},$$

$$(4.66) \quad C_{u/d}^{LR} = \frac{C_{\ell u/d}}{\Lambda^2},$$

$$(4.67) \quad C_{u/d}^{RL} = \frac{C_{qe}}{\Lambda^2},$$

$$(4.68) \quad C_{u/d}^{RR} = \frac{C_{e/u}}{\Lambda^2},$$

$$(4.69)$$

where the upper (lower) signs and indices are for the up (down) quark. The quantum number of fermions in the photon interactions,  $Q_{f\gamma}$ , is still given by the electric charge,  $Q_f$ , at dimension 6:

$$(4.70) \quad Q_{f\gamma}^L = Q_{f\gamma}^R = Q_{f\gamma} = Q_f.$$

The quantum numbers of the  $ffZ$  interactions are given by

$$\begin{aligned}
Q_{eZ}^L = & \frac{1}{4\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z \left(g_{1(0)}^2 + g_{2(0)}^2\right)^{3/2}} \left\{ g_{1(0)}^4 \left( \Lambda^2 - v_{(0)}^2 \left( C_{\varphi\ell}^{(1)} + C_{\varphi\ell}^{(3)} \right) \right) \right. \\
& - g_{2(0)}g_{1(0)}^2 \left( 2v_{(0)}^2g_{2(0)} \left( C_{\varphi\ell}^{(1)} + C_{\varphi\ell}^{(3)} \right) + 3\Lambda^2g_{2(1)} \right) \\
& - g_{2(0)}^3 \left( g_{2(0)} \left( v_{(0)}^2 \left( C_{\varphi\ell}^{(1)} + C_{\varphi\ell}^{(3)} \right) + \Lambda^2 \right) + \Lambda^2g_{2(1)} \right) \\
& + g_{1(0)}^3 \left( \Lambda^2g_{1(1)} - v_{(0)}^2g_{2(0)}C_{\varphi WB} \right) \\
& \left. + g_{2(0)}^2g_{1(0)} \left( v_{(0)}^2g_{2(0)}C_{\varphi WB} + 3\Lambda^2g_{1(1)} \right) \right\},
\end{aligned}
\tag{4.71}$$

$$\begin{aligned}
Q_{eZ}^R = & \frac{1}{4\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z \left(g_{1(0)}^2 + g_{2(0)}^2\right)^{3/2}} \left\{ 2g_{2(0)}^2g_{1(0)} \left( v_{(0)}^2g_{2(0)}C_{\varphi WB} + 2\Lambda^2g_{1(1)} \right) \right. \\
& + 2g_{2(0)}g_{1(0)}^2 \left( g_{2(0)} \left( \Lambda^2 - v_{(0)}^2C_{\varphi e} \right) - \Lambda^2g_{2(1)} \right) + 2\Lambda^2g_{1(0)}^3g_{1(1)} \\
& \left. + g_{1(0)}^4 \left( 2\Lambda^2 - v_{(0)}^2C_{\varphi e} \right) - v_{(0)}^2g_{2(0)}^4C_{\varphi e} \right\},
\end{aligned}
\tag{4.72}$$

$$\begin{aligned}
Q_{uZ}^L = & -\frac{1}{12\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z \left(g_{1(0)}^2 + g_{2(0)}^2\right)^{3/2}} \left\{ g_{1(0)}^4 \left( 3v_{(0)}^2 \left( C_{\varphi q}^{(1)} - C_{\varphi q}^{(3)} \right) + \Lambda^2 \right) \right. \\
& - g_{2(0)}g_{1(0)}^2 \left( 2g_{2(0)} \left( 3v_{(0)}^2 \left( C_{\varphi q}^{(3)} - C_{\varphi q}^{(1)} \right) + \Lambda^2 \right) + 7\Lambda^2g_{2(1)} \right) \\
& - 3g_{2(0)}^3 \left( g_{2(0)} \left( v_{(0)}^2 \left( C_{\varphi q}^{(3)} - C_{\varphi q}^{(1)} \right) + \Lambda^2 \right) + \Lambda^2g_{2(1)} \right) \\
& + g_{1(0)}^3 \left( \Lambda^2g_{1(1)} - 3v_{(0)}^2g_{2(0)}C_{\varphi WB} \right) \\
& \left. + g_{2(0)}^2g_{1(0)} \left( v_{(0)}^2g_{2(0)}C_{\varphi WB} + 5\Lambda^2g_{1(1)} \right) \right\},
\end{aligned}
\tag{4.73}$$

$$\begin{aligned}
Q_{uZ}^R = & -\frac{1}{12\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z\left(g_{1(0)}^2+g_{2(0)}^2\right)^{3/2}}\left\{4g_{2(0)}^2g_{1(0)}\left(v_{(0)}^2g_{2(0)}C_{\varphi WB}+2\Lambda^2g_{1(1)}\right)\right. \\
& +2g_{2(0)}g_{1(0)}^2\left(3v_{(0)}^2g_{2(0)}C_{\varphi u}+2\Lambda^2g_{2(0)}-2\Lambda^2g_{2(1)}\right)+4\Lambda^2g_{1(0)}^3g_{1(1)} \\
& \left.+g_{1(0)}^4\left(3v_{(0)}^2C_{\varphi u}+4\Lambda^2\right)+3v_{(0)}^2g_{2(0)}^4C_{\varphi u}\right\},
\end{aligned}
\tag{4.74}$$

$$\begin{aligned}
Q_{dZ}^L = & -\frac{1}{12\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z\left(g_{1(0)}^2+g_{2(0)}^2\right)^{3/2}}\left\{g_{1(0)}^4\left(3v_{(0)}^2\left(C_{\varphi q}^{(1)}+C_{\varphi q}^{(3)}\right)+\Lambda^2\right)\right. \\
& +g_{2(0)}g_{1(0)}^2\left(6v_{(0)}^2g_{2(0)}\left(C_{\varphi q}^{(1)}+C_{\varphi q}^{(3)}\right)+4\Lambda^2g_{2(0)}+5\Lambda^2g_{2(1)}\right) \\
& +3g_{2(0)}^3\left(g_{2(0)}\left(v_{(0)}^2\left(C_{\varphi q}^{(1)}+C_{\varphi q}^{(3)}\right)+\Lambda^2\right)+\Lambda^2g_{2(1)}\right) \\
& +g_{1(0)}^3\left(3v_{(0)}^2g_{2(0)}C_{\varphi WB}+\Lambda^2g_{1(1)}\right) \\
& \left.+g_{2(0)}^2g_{1(0)}\left(v_{(0)}^2g_{2(0)}C_{\varphi WB}-\Lambda^2g_{1(1)}\right)\right\},
\end{aligned}
\tag{4.75}$$

$$\begin{aligned}
Q_{dZ}^R = & \frac{1}{12\sqrt[4]{2}\Lambda^2\sqrt{G_F}M_Z\left(g_{1(0)}^2+g_{2(0)}^2\right)^{3/2}}\left\{2g_{2(0)}^2g_{1(0)}\left(v_{(0)}^2g_{2(0)}C_{\varphi WB}+2\Lambda^2g_{1(1)}\right)\right. \\
& +2g_{2(0)}g_{1(0)}^2\left(g_{2(0)}\left(\Lambda^2-3v_{(0)}^2C_{\varphi d}\right)-\Lambda^2g_{2(1)}\right)+2\Lambda^2g_{1(0)}^3g_{1(1)} \\
& \left.+g_{1(0)}^4\left(2\Lambda^2-3v_{(0)}^2C_{\varphi d}\right)-3v_{(0)}^2g_{2(0)}^4C_{\varphi d}\right\},
\end{aligned}
\tag{4.76}$$

with

$$(4.77) \quad g_{1(0)} = \frac{g_{SW}}{c_W},$$

$$(4.78) \quad g_{1(1)} = -\frac{c_WM_Z^2s_W^2\left(s_W\left(4C_{\varphi\ell}^{(3)}+C_{\varphi D}-2C_{\ell\ell}\right)+4c_WC_{\varphi WB}\right)}{g\Lambda^2(2s_W^2-1)},$$

$$(4.79) \quad g_{2(0)} = g,$$

$$(4.80) \quad g_{2(1)} = \frac{c_W^2 M_Z^2 \left( -s_W^2 \left( 4C_{\varphi\ell}^{(3)} + C_{\varphi D} \right) + 4C_{\varphi\ell}^{(3)} + 4c_W s_W C_{\varphi WB} - 2c_W^2 C_{\ell\ell} + C_{\varphi D} \right)}{g\Lambda^2 (2s_W^2 - 1)},$$

$$(4.81) \quad v_{(0)} = \frac{1}{\sqrt[4]{2}\sqrt{G_F}},$$

$$(4.82) \quad v_{(1)} = -\frac{C_{\ell\ell} - 2C_{\varphi\ell}^{(3)}}{2 \cdot 2^{3/4} \Lambda^2 G_F^{3/2}},$$

where we have defined

$$(4.83) \quad g = \frac{\sqrt{4\pi\alpha}}{s_W},$$

$$(4.84) \quad c_W = \sqrt{1 - s_W^2},$$

$$(4.85) \quad s_W = \sqrt{\frac{1}{2} \left( 1 - \frac{\sqrt{G_F M_Z^2 (G_F M_Z^2 - 2\sqrt{2}\pi\alpha)}}{G_F M_Z^2} \right)},$$

as a shorthand notation. Numerically, in the input basis  $(G_F, \alpha, M_Z)$  and with  $\Lambda = 1$  TeV, we have

$$(4.86) \quad \begin{aligned} Q_{eZ}^L = & -0.0303121C_{\varphi\ell}^{(1)} + 0.0223408C_{\varphi\ell}^{(3)} + 0.0430523C_{\varphi WB} + 0.0131632C_{\varphi D} \\ & - 0.0263264C_{\ell\ell} - 0.287848, \end{aligned}$$

$$(4.87) \quad \begin{aligned} Q_{eZ}^R = & 0.0223408C_{\varphi\ell}^{(3)} + 0.0430523C_{\varphi WB} + 0.00558519C_{\varphi D} - 0.0303121C_{\varphi e} \\ & - 0.0111704C_{\ell\ell} + 0.212152, \end{aligned}$$



$$\begin{aligned}
Q_{uZ}^L &= -0.0303121C_{\varphi q}^{(1)} + 0.0303121C_{\varphi q}^{(3)} - 0.0452059C_{\varphi \ell}^{(3)} - 0.0287015C_{\varphi WB} \\
(4.88) \quad &- 0.0113015C_{\varphi D} + 0.022603C_{\ell \ell} + 0.358566,
\end{aligned}$$

$$\begin{aligned}
Q_{uZ}^R &= -0.0148938C_{\varphi \ell}^{(3)} - 0.0287015C_{\varphi WB} - 0.00372346C_{\varphi D} - 0.0303121C_{\varphi u} \\
(4.89) \quad &+ 0.00744692C_{\ell \ell} - 0.141434,
\end{aligned}$$

$$\begin{aligned}
Q_{dZ}^L &= -0.0303121C_{\varphi q}^{(1)} - 0.0303121C_{\varphi q}^{(3)} + 0.037759C_{\varphi \ell}^{(3)} + 0.0143508C_{\varphi WB} \\
(4.90) \quad &+ 0.00943975C_{\varphi D} - 0.0188795C_{\ell \ell} - 0.429283,
\end{aligned}$$

$$\begin{aligned}
Q_{dZ}^R &= 0.00744692C_{\varphi \ell}^{(3)} + 0.0143508C_{\varphi WB} - 0.0303121C_{\varphi d} + 0.00186173C_{\varphi D} \\
(4.91) \quad &- 0.00372346C_{\ell \ell} + 0.0707172.
\end{aligned}$$

The energy-dependent eta factors are defined by

$$(4.92) \quad \eta_{\gamma Z} = \frac{C_Z^2 P_Z}{C_\gamma^2 P_\gamma}, \quad \eta_Z = \eta_{\gamma Z}^2.$$

The convolution operator that appears in Equations (4.40)–(4.45) is defined as

$$(4.93) \quad C \otimes f(x) = \int_x^1 \frac{dz}{z} C(z) f\left(\frac{x}{z}\right).$$

The *normalized* unpolarized and polarized coefficient functions,  $\tilde{C}_{q,g_i}$  and  $\Delta\tilde{C}_{q,g_i}$ , are borrowed from [82]. Before setting the factorization scale equal to the transfer momentum,

we have

$$(4.94) \quad \begin{aligned} \tilde{C}_{q_1}(z) = C_F \Big\{ \ln \left( \frac{Q^2}{\mu_F^2} \right) \tilde{P}_{qq}(z) + \delta(1-z) \left( -\frac{9}{2} - \frac{\pi^2}{3} \right) \\ + L_1(z) - L_2(z) - \frac{3}{2} \left[ \frac{1}{1-z} \right]_+ + 3 \Big\}, \end{aligned}$$

$$(4.95) \quad \tilde{C}_{q_3}(z) = \tilde{C}_{q_1}(z) - C_F(1-z),$$

$$(4.96) \quad \tilde{C}_{q_L}(z) = C_F 2z,$$

$$(4.97) \quad \tilde{C}_{g_1}(z) = \frac{1}{2} \left\{ \tilde{P}_{qg}(z) \left[ \ln \left( \frac{Q^2}{\mu_F^2} \frac{1-z}{z} \right) - 1 \right] + 2z(1-z) \right\},$$

$$(4.98) \quad \tilde{C}_{g_3}(z) = 0,$$

$$(4.99) \quad \tilde{C}_{g_L}(z) = 2z(1-z),$$

$$(4.100) \quad \Delta \tilde{C}_{q_1}(z) = \tilde{C}_{q_1}(z) - C_F(1-z),$$

$$(4.101) \quad \Delta \tilde{C}_{q_5}(z) = \tilde{C}_{q_1}(z),$$

$$(4.102) \quad \Delta \tilde{C}_{q_L}(z) = \tilde{C}_{q_L}(z),$$

$$(4.103) \quad \Delta \tilde{C}_{g_1}(z) = \frac{1}{2} \left\{ \Delta \tilde{P}_{qg}(z) \left[ \ln \left( \frac{Q^2}{\mu_F^2} \frac{1-z}{z} \right) - 1 \right] + 2(1-z) \right\},$$

$$(4.104) \quad \Delta \tilde{C}_{g_5}(z) = 0,$$

$$(4.105) \quad \Delta \tilde{C}_{g_L}(z) = 0,$$

where the *auxiliary* functions are defined by

$$(4.106) \quad \tilde{P}_{qq}(z) = \frac{1}{C_F} P_{qq}(z), \quad \tilde{P}_{qg}(z) = 2P_{qg}(z), \quad \Delta \tilde{P}_{qg}(z) = 2z - 1,$$

$$(4.107) \quad L_1(z) = (1+z^2) \left[ \frac{\ln(1-z)}{1-z} \right]_+, \quad L_2(z) = \frac{1+z^2}{1-z} \ln(z),$$

with  $C_F = 4/3$  and  $P_{qq,qq}$  being the leading-order Altarelli-Parisi splitting functions (see e.g. [155]), and the *plus prescription* is defined via

$$(4.108) \quad \int_x^1 dz [f(z)]_+ g(z) = \int_x^1 dz f(z)[g(z) - g(1)] - g(1) \int_0^x dz f(z).$$

In our case, the plus-prescription functions are of the form

$$(4.109) \quad f(z) = \left[ \frac{\ln(1-z)^r}{1-z} \right]_+,$$

where  $r$  is a natural number. With this, Equation (4.108) becomes

$$(4.110) \quad \int_x^1 dz \left[ \frac{\ln(1-z)^r}{1-z} \right]_+ g(z) = g(1) \frac{\ln(1-x)^{r+1}}{r+1} + \int_x^1 dz \frac{\ln(1-z)^r}{1-z} [g(z) - g(1)].$$

We define the reduced cross sections as usual by

$$(4.111) \quad \frac{d^2 \sigma_{r,NC}^\ell}{dx dQ^2} = \left\{ \frac{2\pi\alpha^2}{xQ^4} [1 + (1-y)^2] \right\}^{-1} \frac{d^2 \sigma_{NC}^\ell}{dx dQ^2},$$

$$(4.112) \quad \frac{d^2 \Delta \sigma_{r,NC}^\ell}{dx dQ^2} = \left\{ \frac{4\pi\alpha^2}{xQ^4} [1 + (1-y)^2] \right\}^{-1} \frac{d^2 \Delta \sigma_{NC}^\ell}{dx dQ^2}.$$

From this point onward, when we mention cross sections, we mean the reduced ones and denote them simply by  $(\Delta)\sigma_{NC}$ .

In Figure 4.3, we show the NC DIS cross section with NLO QCD corrections for  $e^-p$  collisions at  $\sqrt{s} = 1.3$  TeV with RH electrons of beam energy 60 GeV and polarization  $P_\ell = +80\%$  and the corresponding  $k$  factors as a function of  $Q$  for various  $x$  values.

From Figure 4.3, we observe that NLO QCD corrections to the NC DIS cross section are 30% at most in either direction and exhibit high sensitivity to  $Q$  and low sensitivity

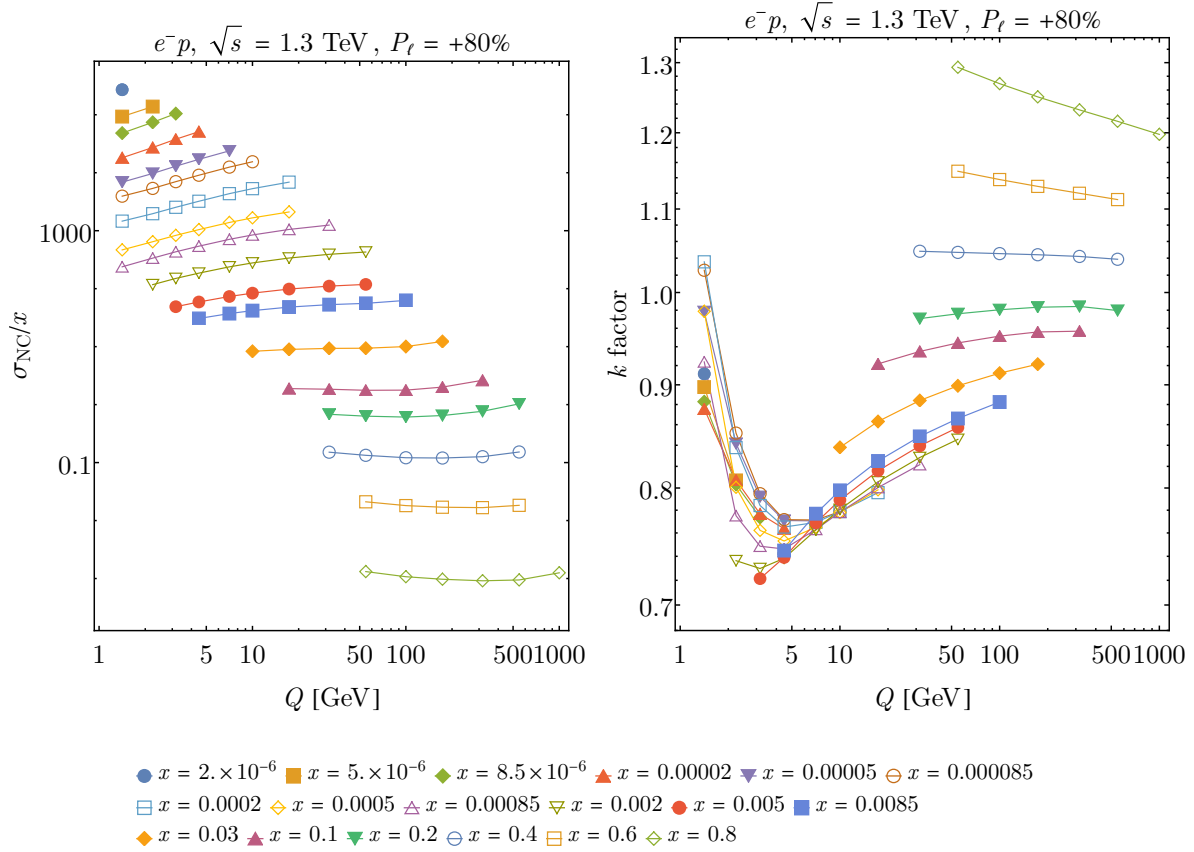


Figure 4.3. NC DIS cross section with NLO QCD corrections for  $e^-p$  collisions at  $\sqrt{s} = 1.3$  TeV with  $E_e = 60$  GeV and  $P_\ell = +80\%$ .

to  $x$  for  $Q \lesssim 30$  GeV and low sensitivity to  $Q$  and high sensitivity to  $x$  for higher values of  $Q$ .

### 4.3. Observables

The EIC observable is the PV asymmetries based on the polarized DIS cross sections following our previous study. The unpolarized PV asymmetry is defined as

$$(4.113) \quad A_{\text{PV}} = \frac{\sigma_{\text{NC}}^+ - \sigma_{\text{NC}}^-}{\sigma_{\text{NC}}^+ + \sigma_{\text{NC}}^-},$$

and the polarized PV asymmetry as

$$(4.114) \quad \Delta A_{\text{PV}} = \frac{\Delta \sigma_{\text{NC}}^0}{\sigma_{\text{NC}}^0},$$

where  $\sigma_{\text{NC}}^\pm$  is the NC DIS  $e^-H$  ( $H = p, D$ ) cross section with unpolarized hadron and polarized lepton beams with  $\lambda_\ell = \pm P_\ell$ ,  $\sigma_{\text{NC}}^0$  is the all-unpolarized cross section, and  $\Delta \sigma_{\text{NC}}^0$  is the cross section with polarized hadron and unpolarized lepton beams. Here,  $P_\ell$  denotes the lepton-beam polarization reach at the EIC. The observable of interest at the LHeC and FCC-eh is the NC DIS cross section with unpolarized hadron and polarized lepton beams. This observable is chosen so we can compare our simulated pseudodata with previous studies in the literature [50, 51].

Since we consider only the SM-SMEFT interference at the amplitude level, all our observables are linearized in SMEFT parameters, or to be more precise, are kept to leading order in the SMEFT expansion,  $E^2/\Lambda^2$ , where  $E$  is a relevant variable with energy dimensions, which is momentum transfer in our case. Thus, our observables have the generic SMEFT form

$$(4.115) \quad \mathcal{O} = \mathcal{O}_{\text{SM}} + \sum_k C_k \mathcal{O}_k,$$

where  $k$  runs over the active Wilson coefficients,  $\mathcal{O} = (\Delta)\sigma_{\text{NC}}$  or  $(\Delta)A_{\text{PV}}$  is the observable,  $\mathcal{O}_{\text{SM}}$  is the SM prediction, and  $\mathcal{O}_k$  is the SMEFT contribution to the observable characterized by the Wilson coefficient  $C_k$ .

#### 4.4. Analysis

In our analysis, we use the EIC pseudodata from the previous work presented in Chapter 3, the most recent public LHeC pseudodata available at

<http://hep.ph.liv.ac.uk/~mklein/lhecddata/> [126],

and generate our own pseudodata for the FCC-eh following the procedure explained in Chapter 3 using the run parameters found in [51]. To simplify the language, from now on, we simply refer to these pseudodatasets as datasets. The configurations of these datasets regarding beam energies, polarizations, and integrated luminosities, as well as our labeling scheme, are summarized in Tables 4.2–4.4. We note that the case of a 10-fold luminosity upgrade at the EIC (so as to have a high-luminosity EIC), which is considered in the previous chapter, is not included here.

On top of the detector cuts, we introduce the kinematic cuts

$$(4.116) \quad x \leq 0.5, \quad Q \geq 10 \text{ GeV}, \quad 0.1 \leq y \leq 0.9,$$

and we refer to the bins that satisfy these conditions as the *good* bins. These cuts help us avoid large uncertainties from nonperturbative QCD and nuclear dynamics. We expect SMEFT effects to be minimal in the leftout regions anyway. The kinematic coverage of the datasets is illustrated in Figure 4.4.

Table 4.2. The description of the EIC datasets used in our analysis. We indicate beam energies, polarizations, integrated luminosities, our labeling scheme, and the observable of interest.

Label	Configuration	Observable
D4	10 GeV $\times$ 137 GeV $e^-D$ , $P_\ell = 80\%$ , $\mathcal{L} = 100 \text{ fb}^{-1}$	$A_{\text{PV}}$
D5	18 GeV $\times$ 137 GeV $e^-D$ , $P_\ell = 80\%$ , $\mathcal{L} = 15.4 \text{ fb}^{-1}$	
P4	10 GeV $\times$ 275 GeV $e^-p$ , $P_\ell = 80\%$ , $\mathcal{L} = 100 \text{ fb}^{-1}$	
P5	18 GeV $\times$ 275 GeV $e^-p$ , $P_\ell = 80\%$ , $\mathcal{L} = 15.4 \text{ fb}^{-1}$	
$\Delta D4$	The same as D4 but with $P_\ell = 0$ and $P_H = 70\%$	$\Delta A_{\text{PV}}$
$\Delta D5$	The same as D5 but with $P_\ell = 0$ and $P_H = 70\%$	
$\Delta P4$	The same as P4 but with $P_\ell = 0$ and $P_H = 70\%$	
$\Delta P5$	The same as P5 but with $P_\ell = 0$ and $P_H = 70\%$	
LD4	The same as D4 but with $P_\ell = 0$	$A_{\text{LC}}$
LD5	The same as D5 but with $P_\ell = 0$	
LP4	The same as P4 but with $P_\ell = 0$	
LP5	The same as P5 but with $P_\ell = 0$	

Next, we discuss the anticipated uncertainties. For the EIC, we have previously assumed uncertainties, which are statistical uncertainties,

$$(4.117) \quad \delta A_{\text{PV}}^{\text{stat}} = \frac{1}{P_\ell \sqrt{N}}, \quad \delta \Delta A_{\text{PV}}^{\text{stat}} = \frac{P_\ell}{P_H} \delta A_{\text{PV}}^{\text{stat}},$$

Table 4.3. The same as in Table 4.2 but for the LHeC.

Label	Configuration	Observable
LHeC1	60 GeV $\times$ 1000 GeV $e^-p$ , $P_\ell = 0$ , $\mathcal{L} = 100 \text{ fb}^{-1}$	$\sigma_{\text{NC}}$
LHeC2	60 GeV $\times$ 7000 GeV $e^-p$ , $P_\ell = -80\%$ , $\mathcal{L} = 100 \text{ fb}^{-1}$	
LHeC3	60 GeV $\times$ 7000 GeV $e^-p$ , $P_\ell = +80\%$ , $\mathcal{L} = 30 \text{ fb}^{-1}$	
LHeC4	60 GeV $\times$ 7000 GeV $e^+p$ , $P_\ell = +80\%$ , $\mathcal{L} = 10 \text{ fb}^{-1}$	
LHeC5	60 GeV $\times$ 7000 GeV $e^-p$ , $P_\ell = -80\%$ , $\mathcal{L} = 1000 \text{ fb}^{-1}$	
LHeC6	60 GeV $\times$ 7000 GeV $e^-p$ , $P_\ell = +80\%$ , $\mathcal{L} = 300 \text{ fb}^{-1}$	
LHeC7	60 GeV $\times$ 7000 GeV $e^+p$ , $P_\ell = 0\%$ , $\mathcal{L} = 100 \text{ fb}^{-1}$	

Table 4.4. The same as in Table 4.2 but for the FCC-eh.

Label	Configuration	Observable
FCCeh1	60 GeV $\times$ 50000 GeV $e^-p$ , $P_\ell = -80\%$ , $\mathcal{L} = 2 \text{ ab}^{-1}$	$\sigma_{\text{NC}}$
FCCeh2	60 GeV $\times$ 50000 GeV $e^-p$ , $P_\ell = +80\%$ , $\mathcal{L} = 0.5 \text{ ab}^{-1}$	
FCCeh3	60 GeV $\times$ 50000 GeV $e^+p$ , $P_\ell = 0$ , $\mathcal{L} = 0.2 \text{ ab}^{-1}$	

where  $P_H$  is the assumed hadron polarization reach, uncorrelated systematic uncertainties,  $\delta(\Delta)A_{\text{PV}}^{\text{sys}}$ , 1% relative to the asymmetry due to particle background and other imperfections in measurements, and correlated lepton and hadron beam polarization uncertainties,  $\delta(\Delta)A_{\text{PV}}^{\text{pol}}$ , 1% and 2% relative to the asymmetry, respectively. For the LHeC and FCC-eh, we borrow the estimates from [50, 51]. These cover uncorrelated statistical,  $\delta\sigma_{\text{stat}}$ , uncorrelated efficiency,  $\delta\sigma_{\text{eff}}$ , and correlated systematic uncertainties,  $\delta\sigma_{\text{sys}}$ . Systematics include lepton energy scale and polar angle measurements,  $\delta\sigma_{\text{len}}$  and  $\delta\sigma_{\text{lpol}}$ , hadronic energy scale,  $\delta\sigma_{\text{hen}}$ , radiative corrections,  $\delta\sigma_{\text{rad}}$ , photoproduction background,  $\delta\sigma_{\text{gam}}$ , global



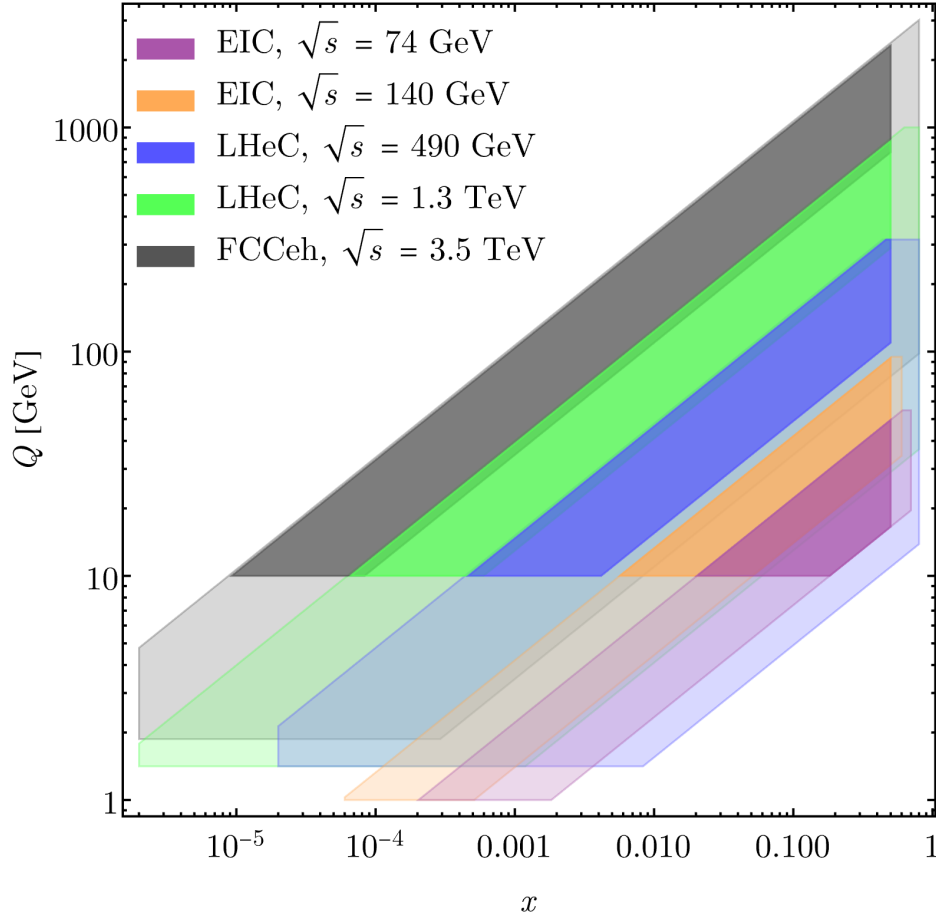


Figure 4.4. Kinematic coverage of the EIC, LHeC, and FCC-eh data sets, indicating the the complementarity of these experiments to each other in terms of the *good* regions considered in our analysis.

efficiency factor,  $\delta\sigma_{\text{geff}}$ , calorimetry noise,  $\delta\sigma_{\text{cal}}$ , and luminosity,  $\delta\sigma_{\text{lum}}$ . Luminosity uncertainties are 1% relative to the cross section. Systematics are assumed fully correlated. In addition to anticipated experimental uncertainties, we also include fully correlated PDF uncertainties,  $\delta(\Delta)A_{\text{PV}}^{\text{pdf}}$  and  $\delta\sigma_{\text{pdf}}$ . We summarize the anticipated values and ranges of experimental uncertainties for all the machines in Tables 4.5–4.7.

Table 4.5. Anticipated values or ranges of experimental uncertainties at the EIC for the *good* bins used in our analysis. All uncertainties are relative with respect to the observable.

Source of uncertainty	Value or range of uncertainty [%]	Observable
Statistical	1.53–65.87	$A_{\text{PV}}$
Systematical	1.00	
Lepton beam polarization	1.00	
Statistical	1.74–75.28	$\Delta A_{\text{PV}}$
Systematical	1.00	
Hadron beam polarization	2.00	

Table 4.6. The same as in Table 4.5 but for the LHeC.

Source of uncertainty	Value or range of uncertainty [%]	Observable
Statistical	0.10–6.83	$\sigma_{\text{NC}}$
Uncorrelated efficiency	0.50	
Lepton energy	0.11–0.49	
Lepton polar angle	0.00–0.13	
Hadron energy	0.00–1.81	
Radiative corrections	0.30	
Photoproduction background	0.00–1.00	
Global efficiency	0.50	
Calorimetry noise	0.00	
Luminosity	1.00	

Table 4.7. The same as in Table 4.5 but for the FCC-eh.

Source of uncertainty	Value or range of uncertainty [%]	Observable
Statistical	0.10–5.49	$\sigma_{\text{NC}}$
Lepton energy	0.90	
Lepton polar angle	0.40	
Hadron energy	2.00	
Radiative corrections	0.30	
Photoproduction background	0.00–1.00	
Global efficiency	0.50	
Luminosity	1.00	

With all the uncertainties accounted for, we can build the uncertainty matrices for all data sets of all machines. An uncertainty matrix has two parts, experimental and theoretical:

$$(4.118) \quad \mathcal{E} = \mathcal{E}_{\text{exp}} + \mathcal{E}_{\text{theo}}.$$

The experimental error matrix is defined by

$$(4.119) \quad \mathcal{E}_{\text{exp},bb'} = \begin{cases} (\delta\mathcal{O}_{\text{uncorr},b} \oplus \delta\mathcal{O}_{\text{corr},b})^2, & b = b', \\ \rho_{bb'} \delta\mathcal{O}_{\text{corr},b} \delta\mathcal{O}_{\text{corr},b'}, & b \neq b', \end{cases}$$

where  $b, b' = 1, 2, \dots, N_B$  are the bin indices,  $N_B$  is the number of good bins,  $\mathcal{O}$  is the observable,  $\delta\mathcal{O}_{(\text{un})\text{corr},b}$  are the (un)correlated uncertainties added in quadrature at the  $b^{\text{th}}$

bin, and we have defined

$$(4.120) \quad \delta\mathcal{O}_1 \oplus \delta\mathcal{O}_2 \oplus \cdots = \sqrt{\delta\mathcal{O}_1^2 + \delta\mathcal{O}_2^2 + \cdots}.$$

We assume full correlations among bins by taking  $\rho_{bb'} = 1$ . For the EIC data sets, we have

$$(4.121) \quad \delta(\Delta)A_{\text{PV,uncorr},b} = \delta(\Delta)A_{\text{PV,stat},b} \oplus \delta(\Delta)A_{\text{PV,sys},b},$$

$$(4.122) \quad \delta(\Delta)A_{\text{PV,corr},b} = \delta(\Delta)A_{\text{PV,pol},b},$$

and for the LHeC and FCC-eh data sets, we have

$$(4.123) \quad \delta\sigma_{\text{uncorr},b} = \delta\sigma_{\text{stat},b} \oplus \delta\sigma_{\text{ueff},b},$$

$$(4.124) \quad \delta\sigma_{\text{corr},b} = \delta\sigma_{\text{sys},b},$$

with

$$(4.125) \quad \delta\sigma_{\text{sys},b} = \delta\sigma_{\text{len},b} \oplus \delta\sigma_{\text{lpol},b} \oplus \delta\sigma_{\text{hen},b} \oplus \delta\sigma_{\text{rad},b} \oplus \delta\sigma_{\text{gam},b} \oplus \delta\sigma_{\text{geff},b} \oplus \delta\sigma_{\text{lum},b}.$$

On the theoretical side, the only source of uncertainty is the PDFs. The PDF error matrix is defined as

$$(4.126) \quad \mathcal{E}_{\text{pdf},bb'} = \frac{1}{N_P} \sum_{m=1}^{N_P} (\mathcal{O}_{m,b} - \mathcal{O}_{0,b})(\mathcal{O}_{m,b'} - \mathcal{O}_{0,b'}),$$

where  $N_P$  is the number of PDF set members and  $\mathcal{O}_{0(m),b}$  is the SM prediction for the observable  $\mathcal{O}$  at the  $b^{\text{th}}$  bin evaluated at the central ( $m^{\text{th}}$ ) member of the relevant PDF set.

After obtaining the uncertainty matrices, we can compare various uncertainty components to the observables at the EIC, LHeC, and FCC-eh. In Figures 4.5 and 4.6, we exhibit the error budget for the fits by comparing the uncertainty components to the observables for representative data sets. On the horizontal axis, we have the bin numbers, where the bins are sorted first by  $Q$  and then by  $x$  in increasing order. On the vertical axis, we have the central values of the observables and the uncertainty components that contribute to the diagonal entries of the error matrix.

In these error-budget plots, we see that systematics dominate at the LHeC and FCC-eh and that PDF uncertainties are non-negligible, which means at some point, one should consider a simultaneous fit of PDF in conjunction with SMEFT parameters, much like [62, 107, 102]. At the EIC, statistics dominate by an order for the unpolarized PV asymmetries, and for the polarized PV asymmetries, PDF uncertainties become more important than statistics.

Later in our study, we consider joint fits of certain datasets within a given machine. For the joint EIC fit, we assume beam polarization and PDF uncertainties to be correlated among the combined runs, and for the joint LHeC and FCC-eh fits, we assume all the uncertainties under systematics except for photoproduction background, as well as PDF uncertainties to be correlated. A joint uncertainty matrix is given by the uncertainty matrices of individual runs on the block-diagonal entries,  $\mathcal{E}$ , and we define uncertainty

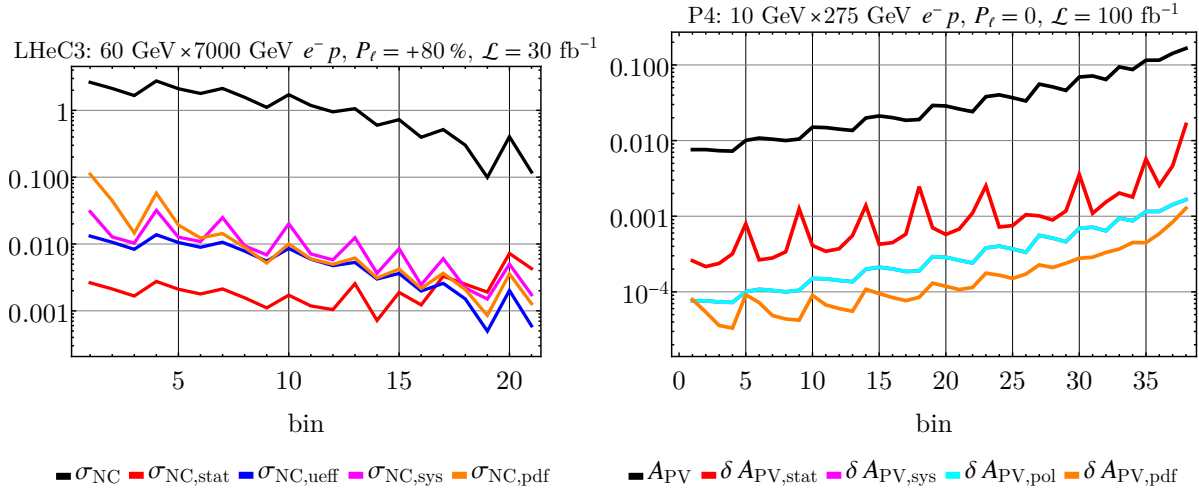


Figure 4.5. The different sources of uncertainty that make up the diagonal elements of the uncertainty matrix are shown for the LHeC3 (left) and P4 (right) data sets. The red line shows statistical uncertainty, the blue line shows uncorrelated global efficiency uncertainty, the magenta line shows systematic uncertainty, and the orange line shows PDF uncertainty. For P4, the cyan line shows the uncertainty from beam polarization.

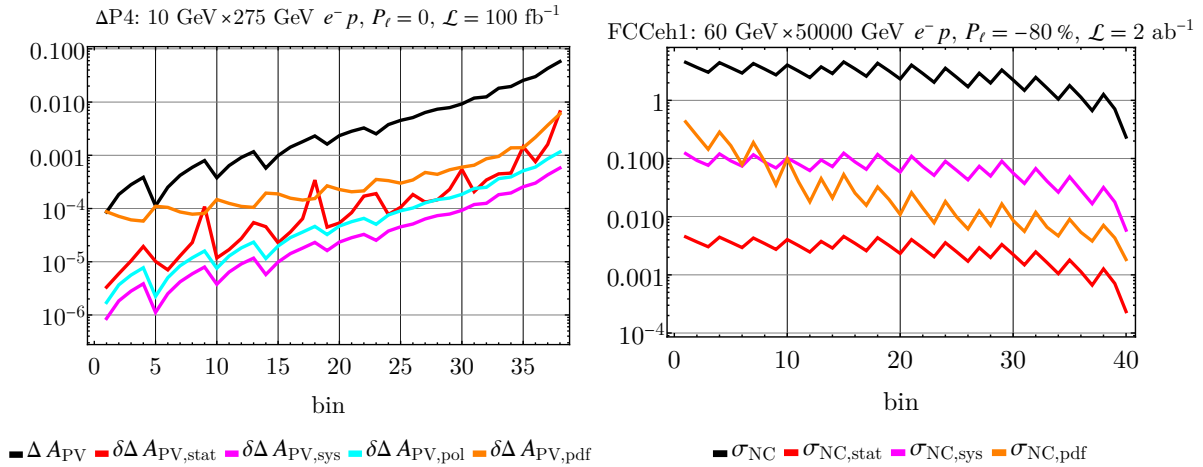


Figure 4.6. The same as in Figure 4.5 but for  $\Delta$ P4 (left) and for FCCeh1 (right).

matrices of correlation,  $\mathcal{J}$ , in the off-block-diagonal entries by

$$(4.127) \quad \mathcal{J}_{nn'} = \mathcal{J}_{\text{exp},nn'} + \mathcal{J}_{\text{pdf},nn'},$$

where  $n, n' = 1, 2, \dots, N_R$  are the run indices,  $N_R$  is the number of datasets or runs combined, and

$$(4.128) \quad \mathcal{J}_{\text{exp},nn',bb'} = \rho_{nn',bb'} \delta\mathcal{O}_{\text{corr},n,b} \delta\mathcal{O}_{\text{corr},n',b'},$$

$$(4.129) \quad \mathcal{J}_{\text{pdf},nn',bb'} = \frac{1}{N_P} \sum_{m=1}^{N_P} (\mathcal{O}_{n,m,b} - \mathcal{O}_{n,0,b})(\mathcal{O}_{n',m,b'} - \mathcal{O}_{n',0,b'}),$$

where  $b = 1, 2, \dots, N_{B,n}$ ,  $N_{B,n}$  is the number of good bins of the  $n^{\text{th}}$  run in the combined datasets,  $\delta\mathcal{O}_{\text{corr},n,b}$  is the relevant correlated uncertain described above, and  $\mathcal{O}_{n,0(m),b}$  is the observable evaluated at the central ( $m^{\text{th}}$ ) member of the PDF set at the  $b^{\text{th}}$  bin of the  $n^{\text{th}}$  run. At the end of the day, the joint uncertainty matrix looks like

$$(4.130) \quad \mathcal{E} = \begin{pmatrix} \mathcal{E}_1 & \mathcal{J}_{12} & \cdots & \mathcal{J}_{1N_R} \\ & \mathcal{E}_2 & \cdots & \mathcal{J}_{2N_R} \\ & & \ddots & \vdots \\ & & & \mathcal{E}_{N_R} \end{pmatrix}_{\text{sym}},$$

where  $\mathcal{E}_n$  is the error matrix of the  $n^{\text{th}}$  run of the datasets combined.

Now we have the observable, as well as the uncertainty matrix for a given individual or joint dataset. With that, we define a  $\chi^2$  test function as

$$(4.131) \quad \chi_e^2 = \sum_{b,b'=1}^{N_B} (\mathcal{O} - \mathcal{Q}_e)_b \mathcal{H}_{bb'} (\mathcal{O} - \mathcal{Q}_e)_{b'}$$

for the  $e^{\text{th}}$  pseudoexperiment, where  $\mathcal{O}_b$  is the SMEFT observable,  $\mathcal{O}_{e,b}$  is the pseudoexperimental value for the observable at the  $b^{\text{th}}$  bin, and  $\mathcal{H} = \mathcal{E}^{-1}$  is the inverse uncertainty matrix. Pseudoexperimental values are obtained by smearing the SM values with the uncertainties using random variates as

$$(4.132) \quad \mathcal{O}_{e,b} = \mathcal{O}_b^{\text{SM}} + r_{e,b} \delta\mathcal{O}_{\text{uncorr},b} + \sum_j r'_{j,e} \delta\mathcal{O}_{\text{corr}_j,b},$$

where  $\mathcal{O}_b^{\text{SM}}$  is the SM prediction for the observable,  $r_{e,b}, r'_{j,e} \sim \mathcal{N}(0,1)$  are unit normal random variates,  $\delta\mathcal{O}_{\text{uncorr},b}$  is the total uncorrelated uncertainty, and  $\delta\mathcal{O}_{\text{corr}_j,b}$  is the  $j^{\text{th}}$  correlated uncertainty. The best-fit values are obtained as the solutions of the equation

$$(4.133) \quad \nabla \chi_e^2(\overline{\mathcal{C}}_e) = \mathbf{0},$$

and the Fisher information matrix is simply the hessian evaluated at the best-fit values:

$$(4.134) \quad \mathcal{F} = \frac{1}{2} \nabla \nabla \chi_e^2(\overline{\mathcal{C}}_e),$$

where all the derivatives are with respect to the SMEFT parameters. For a linear model, the Fisher matrix is constant for all pseudoexperiments, hence we drop the subscript  $e$  now on. Using the machinery of statistical analysis introduced in Section 2.4, we present the SMEFT fit results in the next section.



#### 4.5. SMEFT fit results

For our numerical analysis, we use the input scheme  $\{G_F, \alpha, m_Z\}$  with numerical values

$$(4.135) \quad G_F = 1.1663787 \times 10^{-5} \text{ GeV}^{-2},$$

$$(4.136) \quad \alpha^{-1} = 137.036,$$

$$(4.137) \quad m_Z = 91.1876 \text{ GeV}.$$

We assume polarization reaches of  $P_\ell = 80\%$  and  $P_H = 70\%$  at the EIC, and use the polarization values given with the LHeC and FCC-eh datasets. We set  $\Lambda = 1 \text{ TeV}$ . We use NNPDF3.1 NLO [32] and NNPDF1.1 NLO [141] PDF sets with unpolarized and polarized cross sections, respectively. We compute  $\alpha_s$  from the renormalization group equation running at two loops

$$(4.138) \quad \mu_R^2 \frac{d\alpha_s}{d\mu_R^2} = \beta(\alpha_s) = -(b_0\alpha_s^2 + b_1\alpha_s^3),$$

where  $b_0 = (33 - 2N_f)/(12\pi)$  and  $b_1 = (153 - 19N_f)/(24\pi^2)$ , with the initial condition  $\alpha_s(m_Z) = 0.1185$ . We set  $N_f = 5$  and  $\mu_R = Q$ .

##### 4.5.1. Semi-leptonic four-fermion operators

First, we activate only the semi-leptonic four-fermion operators. The Drell-Yan (DY) process at the LHC has difficulty probing certain linear combinations of SMEFT parameters in this subspace [47, 23]. As we have shown in Chapter 3, future DIS experiments can

resolve these degeneracies (see also [47]). Thus, we want to compare the BSM potential of the EIC, LHeC, and FCC-eh to the DY at the LHC first.

In Table 4.8, we present the nonmarginalized bounds and the corresponding effective UV scales at 95% confidence level (CL) by activating one Wilson coefficient at a time, as well as the marginalized bounds and the marginalized bounds in the seven-parameter (7d) fit and the UV scales by activating them all. Our analysis focuses on the datasets P4,  $\Delta$ P4, the joint set of D4,  $\Delta$ D4, P4, and  $\Delta$ P4 at the EIC because #4 datasets are the strongest at the EIC (see Chapter 3), individual and the combined LHeC runs, and individual and the combined FCC-eh runs.

There are strong correlations among SMEFT parameters, which is why marginal and nonmarginal bounds often look wildly different. This tells us that there are degeneracies in the parameter space. But when we combine datasets, these degeneracies start to break, and we get better results. The marginal effective UV scales range from about 500 GeV to 1 TeV at the EIC, from 2.5 to 14 TeV at the LHeC, and from 2 to 18 TeV at the FCC-eh. The trend is clear: higher energy, stronger reach. No man can eat 50 eggs, but FCC-eh comes close. Polarized PV asymmetries at the EIC give weaker bounds on its own, but plays an important role in joint fits. No single LHeC or FCC-eh run is able to strongly constrain all semi-leptonic four-fermion operators. Each run uses different lepton types and helicities, which affect sensitivity. Still, the LHeC bounds are generally stronger than the EIC because of access to higher momentum transfers, where SMEFT effects become more important. For most operators, FCC-eh joint fits give the strongest bounds overall. Among the different configurations,  $e^-p$  with right-handed  $e^-$  beams are best for constraining  $C_{eu}$  and  $C_{ed}$ . High-luminosity  $e^-p$  runs with left-handed  $e^-$  beams

Table 4.8. Individual and combined 95% confidence level limits on semi-leptonic four-fermion Wilson coefficients at  $\Lambda = 1$  TeV, using the EIC data sets P4 and  $\Delta$ P4, the combined EIC fit with D4,  $\Delta$ D4, P4, and  $\Delta$ P4, the separate and combined LHeC runs, and the separate and combined FCC-eh runs. The related effective UV scales are also shown in TeV. “nonmarg.” means nonmarginalized and “marg.” means marginalized.

		$C_{eu}$	$\frac{\Lambda}{\sqrt{C_{eu}}}$	$C_{ed}$	$\frac{\Lambda}{\sqrt{C_{ed}}}$	$C_{tq}^{(1)}$	$\frac{\Lambda}{\sqrt{C_{tq}^{(1)}}}$	$C_{tq}^{(3)}$	$\frac{\Lambda}{\sqrt{C_{tq}^{(3)}}}$	$C_{tu}$	$\frac{\Lambda}{\sqrt{C_{tu}}}$	$C_{td}$	$\frac{\Lambda}{\sqrt{C_{td}}}$	$C_{qe}$	$\frac{\Lambda}{\sqrt{C_{qe}}}$
P4	nonmarg.	0.15	2.6	0.54	1.4	0.23	2.1	0.13	2.7	0.36	1.7	0.75	1.2	0.51	1.4
	marg.	19.	0.23	$2.5 \times 10^2$	0.063	$1.2 \times 10^2$	0.091	$1.3 \times 10^2$	0.089	43.	0.15	39.	0.16	46.	0.15
$\Delta$ P4	nonmarg.	0.34	1.7	2.2	0.68	0.28	1.9	0.44	1.5	0.81	1.1	5.3	0.43	0.74	1.2
	marg.	39.	0.16	92.	0.10	52.	0.14	32.	0.18	36.	0.17	$1.1 \times 10^2$	0.097	30.	0.18
Joint EIC	nonmarg.	0.12	2.9	0.34	1.7	0.17	2.4	0.10	3.2	0.28	1.9	0.57	1.3	0.39	1.6
	marg.	2.1	0.69	7.2	0.37	2.8	0.59	4.2	0.49	9.1	0.33	9.8	0.32	8.9	0.33
LHeC1	nonmarg.	0.054	4.3	0.29	1.9	0.048	4.6	0.028	6.0	0.27	1.9	0.94	1.0	0.37	1.7
	marg.	89.	0.11	$1.9 \times 10^2$	0.073	67.	0.12	8.0	0.35	23.	0.21	62.	0.13	52.	0.14
LHeC2	nonmarg.	0.080	3.5	0.35	1.7	0.0089	11.	0.0043	15.	0.059	4.1	0.24	2.1	0.75	1.2
	marg.	$7.8 \times 10^2$	0.036	$1.7 \times 10^3$	0.024	66.	0.12	9.4	0.33	28.	0.19	58.	0.13	$5.1 \times 10^2$	0.044
LHeC3	nonmarg.	0.0066	12.	0.026	6.2	0.064	4.0	0.028	6.0	0.36	1.7	1.0	0.99	0.050	4.5
	marg.	58.	0.13	$1.3 \times 10^2$	0.089	$4.0 \times 10^2$	0.050	55.	0.13	$1.7 \times 10^2$	0.077	$3.4 \times 10^2$	0.054	38.	0.16
LHeC4	nonmarg.	0.28	1.9	0.69	1.2	0.037	5.2	0.013	8.7	0.015	8.1	0.057	4.2	0.18	2.4
	marg.	$8.4 \times 10^2$	0.035	$1.9 \times 10^3$	0.023	72.	0.12	11.	0.30	33.	0.17	63.	0.13	$5.8 \times 10^2$	0.041
LHeC5	nonmarg.	0.053	4.3	0.30	1.8	0.0052	14.	0.0031	18.	0.037	5.2	0.18	2.4	0.49	1.4
	marg.	$3.0 \times 10^2$	0.058	$6.6 \times 10^2$	0.039	25.	0.20	3.6	0.53	11.	0.31	22.	0.21	$2.0 \times 10^2$	0.071
LHeC6	nonmarg.	0.0037	16.	0.019	7.3	0.030	5.8	0.017	7.7	0.20	2.2	0.77	1.1	0.032	5.6
	marg.	20.	0.22	45.	0.15	$1.4 \times 10^2$	0.084	20.	0.22	59.	0.13	$1.2 \times 10^2$	0.090	13.	0.27
LHeC7	nonmarg.	0.024	6.5	0.075	3.7	0.023	6.6	0.011	9.5	0.014	8.5	0.065	3.9	0.020	7.1
	marg.	41.	0.16	93.	0.10	32.	0.18	5.0	0.45	15.	0.26	28.	0.19	29.	0.19
Joint LHeC	nonmarg.	0.0022	21.	0.0097	10.	0.0031	18.	0.0017	24.	0.0084	11.	0.036	5.3	0.011	9.7
	marg.	0.0053	14.	0.026	6.2	0.020	7.1	0.011	9.5	0.032	5.6	0.16	2.5	0.018	7.4
FCCeh1	nonmarg.	0.015	8.3	0.043	4.8	0.0020	22.	0.00070	38.	0.0061	13.	0.016	8.0	0.065	3.9
	marg.	$3.6 \times 10^2$	0.053	$8.2 \times 10^2$	0.035	31.	0.18	4.9	0.45	14.	0.26	27.	0.19	$2.6 \times 10^2$	0.063
FCCeh2	nonmarg.	0.0013	28.	0.0031	18.	0.019	7.2	0.0046	15.	0.042	4.9	0.077	3.6	0.0046	15.
	marg.	20.	0.22	46.	0.15	$1.4 \times 10^2$	0.083	22.	0.21	66.	0.12	$1.2 \times 10^2$	0.090	14.	0.26
FCCeh3	nonmarg.	0.0059	13.	0.012	9.2	0.012	9.3	0.0021	22.	0.0074	12.	0.018	7.4	0.0083	11.
	marg.	69.	0.12	$1.5 \times 10^2$	0.081	53.	0.14	7.8	0.36	25.	0.20	44.	0.15	47.	0.15
Joint FCCeh	nonmarg.	0.00056	42.	0.0012	28.	0.0014	27.	0.00038	51.	0.0028	19.	0.0061	13.	0.0016	25.
	marg.	0.0031	18.	0.0070	12.	0.035	5.4	0.014	8.4	0.068	3.8	0.26	2.0	0.0092	10.

are best for  $C_{\ell q}^{(1)}$  and  $C_{\ell q}^{(3)}$ . The only polarized  $e^+p$  LHeC dataset is especially good for  $C_{lu}$  and  $C_{ld}$ , because of the kinematic structures in the amplitudes. Finally, unpolarized  $e^+p$  runs are best for probing  $C_{qe}$ .

The effective field theory is based on an expansion in powers of  $Q/\Lambda_{\text{eff}} = Q/(\Lambda/\sqrt{(C)})$ . In our case, this ratio stays small for all runs. This means the expansion works well. So, keeping only the dimension-6 terms and treating their effects to first order is a safe and good choice.

We show examples of  $2d$  nonmarginalized fits and marginalized results from the  $7d$  fit in Figure 4.7. The strongest individual EIC data set, the strongest LHeC and FCC-eh sets for these Wilson coefficients, and the combined EIC, FCC-eh, and LHeC fits are shown. In the  $2d$  case, we activate two Wilson coefficients at a time. These fits often look clean because the data can separate the effects of the chosen pair. But in the  $7d$  marginalized case, some information is lost, and flat directions appear. These are combinations that the data cannot fully resolve. The EIC is especially weak in this case. Even the joint EIC fit does not break these flat directions. So, one must consider running the machine under different setups to better cover the parameter space.

In both the LHeC and FCC-eh datasets, there are three parameters that vary: the integrated luminosity, the lepton beam polarization, and the lepton species. To keep things simple, we focus on the LHeC as a representative case. Figure 4.8 shows a comparison between two data sets that differ by a factor of 10 in total luminosity. The improvement in the fit is small. This tells us that just increasing the event count is not always the most effective way to gain sensitivity.

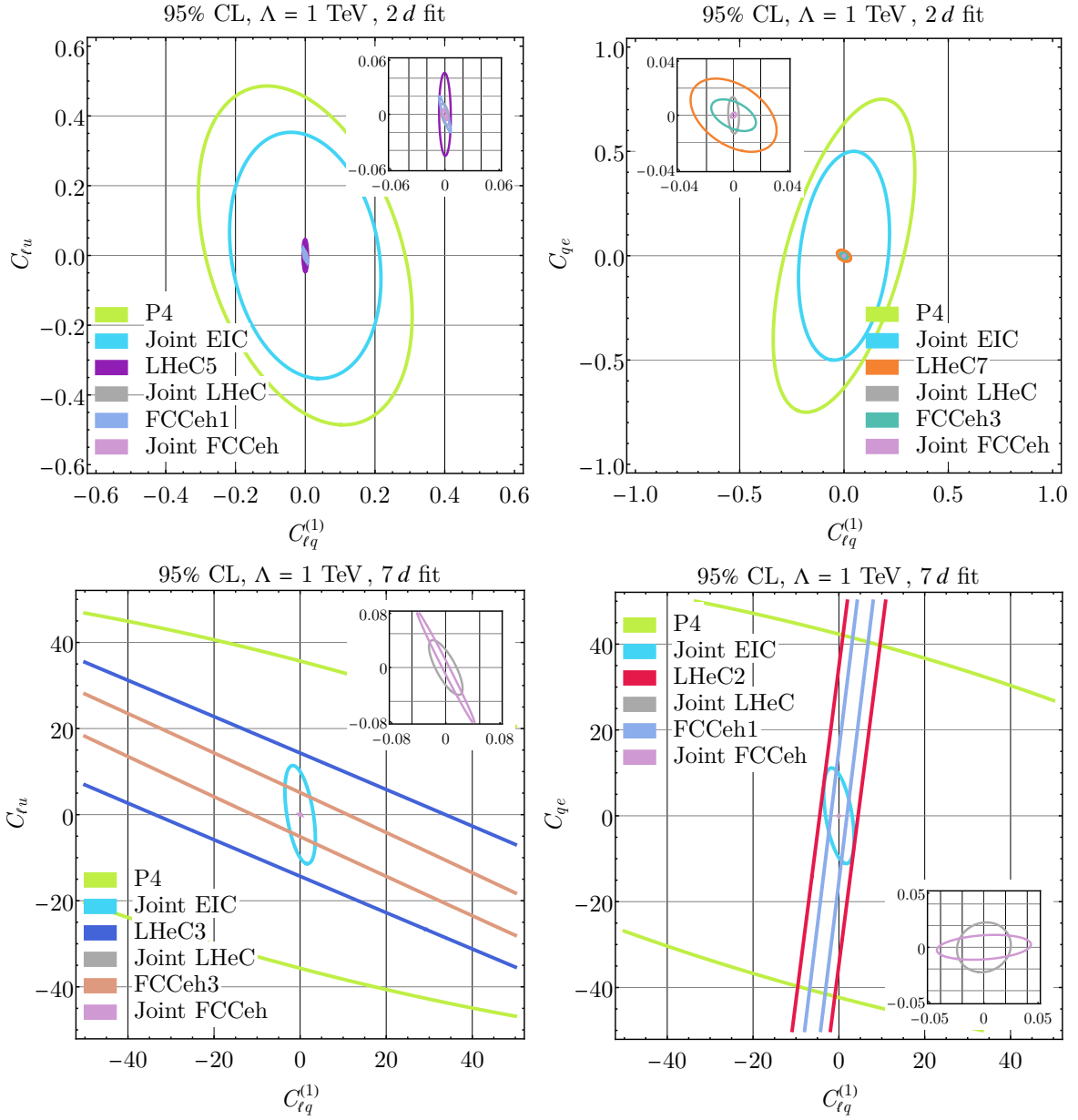


Figure 4.7. Nonmarginalized (top) and marginalized (bottom) 95% confidence level ellipses for the parameter spaces formed by  $C_{\ell q}^{(1)}$  and  $C_{\ell u}$  (left) and  $C_{\ell q}^{(1)}$  and  $C_{qe}$  (right) with  $\Lambda = 1$  TeV. The insets display a zoomed-in view of the combined LHeC and FCC-eh fits.

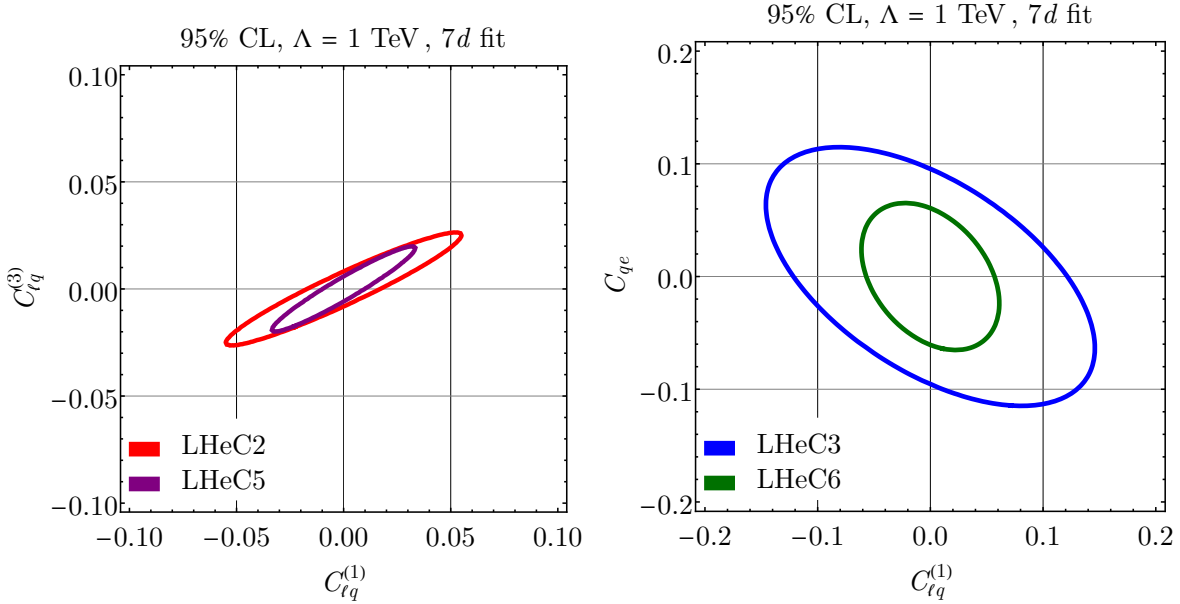


Figure 4.8. Marginalized 95% confidence level ellipses in the parameter spaces formed by  $C_{lq}^{(1)}$  and  $C_{lq}^{(3)}$  (left) and  $C_{lu}$  and  $C_{ld}$  (right) at  $\Lambda = 1$  TeV, comparing data sets with luminosities differing by a factor of 10: LHeC2 and LHeC5 (left), and LHeC3 and LHeC6 (right).

Figure 4.9 compares runs with electron and positron beams. Here, even though the positron run has lower luminosity, the bounds improve significantly. The reason is in the structure of the amplitudes. For electron and positron beams, different Wilson coefficients contribute. When we switch from electron to positron beams, we replace  $C_{eu}$  with  $C_{lu}$ ,  $C_{ed}$  with  $C_{ld}$ , and  $C_{lq}^{(1)} \pm C_{lq}^{(3)}$  with  $C_{qe}$ . These new combinations remove the factor of  $(1 - y)^2$  that suppresses the cross section in the electron case. On average, this factor is about 1/4. Removing it gives a much larger SMEFT contribution. This shows that changing the lepton species can improve the fit more than increasing luminosity. It also supports the case for including positron beams in future DIS programs.

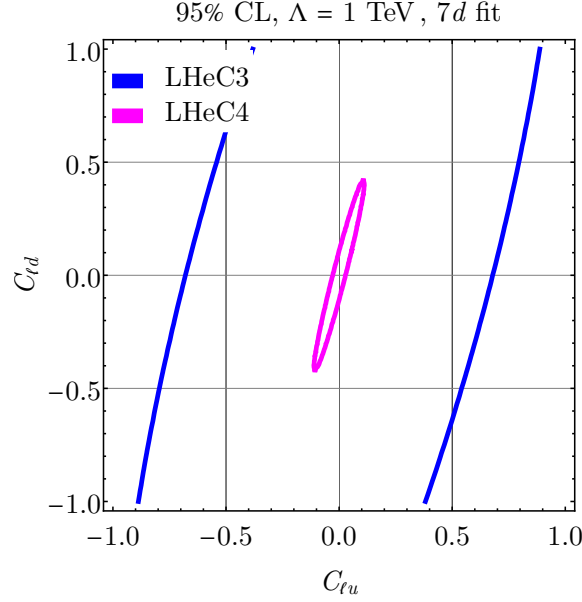


Figure 4.9. The same as in Figure 4.8 but for  $C_{\ell u}$  and  $C_{\ell d}$ , with data sets having different lepton species (LHeC3 and LHeC4).

The DY data from the LHC leaves blind spots in the SMEFT parameter space. These directions are poorly constrained and limit the reach of global fits. DIS measurements can remove them by probing different operator combinations. But DIS measurements can also have their own degeneracies, especially at high energies. To illustrate this, we examine one LHeC run configuration as an example.

The NC DIS amplitude contains two structures: one constant in  $y$ , and one proportional to  $(1 - y)^2$ . We set the coefficients of both structures to zero, for both up and down quark amplitudes. This gives four equations involving seven Wilson coefficients. So only three directions are independent. We choose  $C_{lq}^{(1)}$ ,  $C_{lq}^{(3)}$ , and  $C_{qe}$  as a basis. For the

SMEFT contribution to  $e^-p$  scattering to vanish, the solutions are given by

$$(4.139) \quad C_{eu} = \frac{P_\ell - 1}{P_\ell + 1} \frac{Q_u - g_+^e g_+^u \hat{\eta}_{\gamma Z}}{Q_u - g_-^e g_-^u \hat{\eta}_{\gamma Z}} (C_{\ell q}^{(1)} - C_{\ell q}^{(3)}),$$

$$(4.140) \quad C_{\ell u} = \frac{P_\ell + 1}{P_\ell - 1} \frac{Q_u - g_-^e g_+^u \hat{\eta}_{\gamma Z}}{Q_u - g_+^e g_-^u \hat{\eta}_{\gamma Z}} C_{qe},$$

$$(4.141) \quad C_{ed} = \frac{P_\ell - 1}{P_\ell + 1} \frac{Q_d - g_+^e g_+^d \hat{\eta}_{\gamma Z}}{Q_d - g_-^e g_-^d \hat{\eta}_{\gamma Z}} (C_{\ell q}^{(1)} + C_{\ell q}^{(3)}),$$

$$(4.142) \quad C_{\ell d} = \frac{P_\ell + 1}{P_\ell - 1} \frac{Q_d - g_-^e g_+^d \hat{\eta}_{\gamma Z}}{Q_d - g_+^e g_-^d \hat{\eta}_{\gamma Z}} C_{qe},$$

where  $Q_{u/d}$  is the up/down quark electric charge,  $g_\pm^f = g_V^f \pm g_A^f$ ,  $g_{V/A}^f$  are the usual SM vector/axial fermion couplings to the  $Z$  boson, and the energy-dependent  $\eta$  factor is defined by

$$(4.143) \quad \hat{\eta}_{\gamma Z} = \frac{G_F M_Z^2}{2\sqrt{2}\pi\alpha} \frac{Q^2}{Q^2 + M_Z^2}.$$

These same combinations also cancel the SMEFT contribution to  $e^+p$  scattering if we flip the lepton polarization.

Because of energy-dependent factors in the amplitude, these cancellations are only approximate. But they become more accurate at large  $Q$ . At the LHeC,  $Q$  can reach up to 1 TeV, which is much larger than  $m_Z$ . So the degeneracies become relevant. Since the cancellations depend on the lepton polarization, they can be broken by running with different polarization settings. This holds for both LHeC and FCC-eh.

This is a bottom-up construction of a flat direction. We do not try to connect it to any specific UV model. For  $P_\ell = -80\%$ , we study how this direction appears in the LHeC2,



LHeC4, and LHeC5 data sets. In the limit  $Q/m_Z \rightarrow \infty$ , the solutions are given by

$$(4.144) \quad C_{eu} \approx -13(C_{\ell q}^{(1)} - C_{\ell q}^{(3)}) \equiv C_{eu}^{(1)}$$

$$(4.145) \quad C_{\ell u} \approx -0.052 C_{qe} \equiv C_{\ell u}^{(1)}$$

$$(4.146) \quad C_{ed} \approx -22(C_{\ell q}^{(1)} + C_{\ell q}^{(3)}) \equiv C_{ed}^{(1)}$$

$$(4.147) \quad C_{\ell d} \approx 0.12 C_{qe} \equiv C_{\ell d}^{(1)}.$$

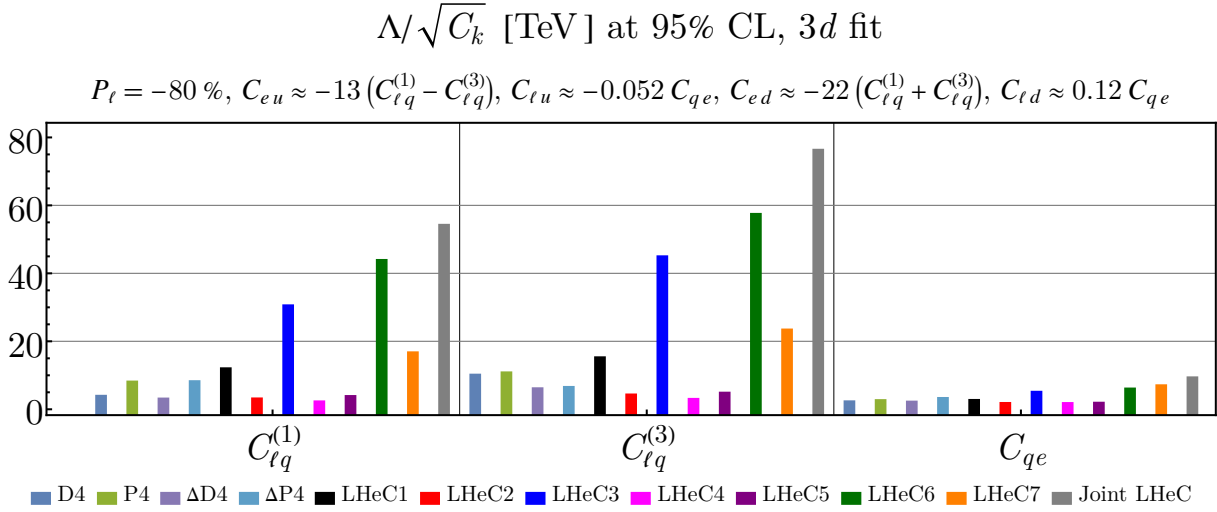


Figure 4.10. Effective UV scales corresponding to marginalized 95% confidence level bounds on the Wilson coefficients  $C_{\ell q}^{(1)}$ ,  $C_{\ell q}^{(3)}$ , and  $C_{qe}$  in the analysis of flat directions for LHeC2, LHeC4, and LHeC5.

We apply these equations to the amplitudes and re-fit the data in the three-dimensional parameter space spanned by  $C_{\ell q}^{(1)}$ ,  $C_{\ell q}^{(3)}$ , and  $C_{qe}$ . The resulting bounds and effective UV scales at 95% CL are shown in Figure 4.10. As expected, the individual runs give weak

bounds, since the SMEFT effects are tuned to cancel out. But the joint fit removes the flat direction and leads to strong bounds. This highlights the importance of running future DIS programs with a variety of beam configurations.

#### 4.5.2. $ffV$ vertex corrections

We now activate a 17-dimensional parameter set that includes modifications to the fermion couplings to neutral EW gauge bosons. In general, precision observables at the  $Z$  pole provide strong constraints on these parameters. In some cases, they push the effective UV scales up to 10 TeV, as shown in [76, 49]. But the number of independent measurements is limited, and once we move beyond single-parameter fits, the bounds weaken by about an order of magnitude. This is due to strong correlations among parameters, which create flat directions in the global fit. These effects are discussed in [92].

Table 4.9 shows our results from a 17-dimensional marginalized fit of the  $ffV$  operators using joint data sets from the EIC, LHeC, and FCC-eh. For comparison, we also show the corresponding bounds from the 34-dimensional EW, diboson, Higgs, and top-quark fits adapted from [92]. The correlation matrices from our joint LHeC and FCC-eh fits are given in Figures 4.11 and 4.12. It is important to note that this is not a one-to-one comparison with [92], since their fit includes 34 parameters while ours has only 17.

Several features stand out. First, the bounds from the LHeC are stronger than those from the joint EW fit. If added to a global analysis, the LHeC would be a major contributor. The FCC-eh bounds are even stronger than both the LHeC and the EIC in most

Table 4.9. Marginalized 95% confidence level bounds on Wilson coefficients in the 17-dimensional fit assuming  $\Lambda = 1$  TeV, along with the corresponding effective UV scales in TeV. The combined EIC fit of D4,  $\Delta$ D4, P4, and  $\Delta$ P4, the joint LHeC and FCC-eh fits, and the marginalized bounds and UV scales from the 34-dimensional fits of EW, diboson, Higgs, and top data [92] are shown.

	Joint EIC	Joint LHeC	Joint FCCeh	EW diboson, Higgs, and top data
$C_{\varphi D}$	[-3.8, 3.8]	[-0.019, 0.019]	[-0.013, 0.013]	[-1.6, 0.81]
$\frac{\Lambda}{\sqrt{C_{\varphi D}}}$	0.51	7.2	8.8	0.91
$C_{\varphi WB}$	[-9.9, 9.9]	[-0.098, 0.098]	[-0.034, 0.034]	[-0.36, 0.73]
$\frac{\Lambda}{\sqrt{C_{\varphi WB}}}$	0.32	3.2	5.4	1.4
$C_{\varphi q}^{(1)}$	[-38., 38.]	[-0.40, 0.40]	[-0.39, 0.39]	[-0.27, 0.18]
$\frac{\Lambda}{\sqrt{C_{\varphi q}^{(1)}}}$	0.16	1.6	1.6	2.1
$C_{\varphi q}^{(3)}$	[-4.1, 4.1]	[-0.11, 0.11]	[-0.031, 0.031]	[-0.11, 0.012]
$\frac{\Lambda}{\sqrt{C_{\varphi q}^{(3)}}}$	0.49	3.1	5.7	4.1
$C_{\varphi u}$	[-38., 38.]	[-0.51, 0.51]	[-0.45, 0.45]	[-0.63, 0.25]
$\frac{\Lambda}{\sqrt{C_{\varphi u}}}$	0.16	1.4	1.5	1.5
$C_{\varphi d}$	[-84., 84.]	[-0.82, 0.82]	[-0.71, 0.71]	[-0.91, 0.13]
$\frac{\Lambda}{\sqrt{C_{\varphi d}}}$	0.11	1.1	1.2	1.4
$C_{\varphi \ell}^{(1)}$	[-18., 18.]	[-0.094, 0.094]	[-0.060, 0.060]	[-0.19, 0.41]
$\frac{\Lambda}{\sqrt{C_{\varphi \ell}^{(1)}}}$	0.23	3.3	4.1	1.8
$C_{\varphi \ell}^{(3)}$	[-4.1, 4.1]	[-0.060, 0.060]	[-0.022, 0.022]	[-0.13, 0.055]
$\frac{\Lambda}{\sqrt{C_{\varphi \ell}^{(3)}}}$	0.49	4.1	6.7	3.3
$C_{\varphi e}$	[-5.7, 5.7]	[-0.16, 0.16]	[-0.046, 0.046]	[-0.41, 0.79]
$\frac{\Lambda}{\sqrt{C_{\varphi e}}}$	0.42	2.5	4.6	1.3
$C_{\ell \ell}$	[-7.7, 7.7]	[-0.039, 0.039]	[-0.026, 0.026]	[-0.084, 0.02]
$\frac{\Lambda}{\sqrt{C_{\ell \ell}}}$	0.36	5.1	6.2	4.4

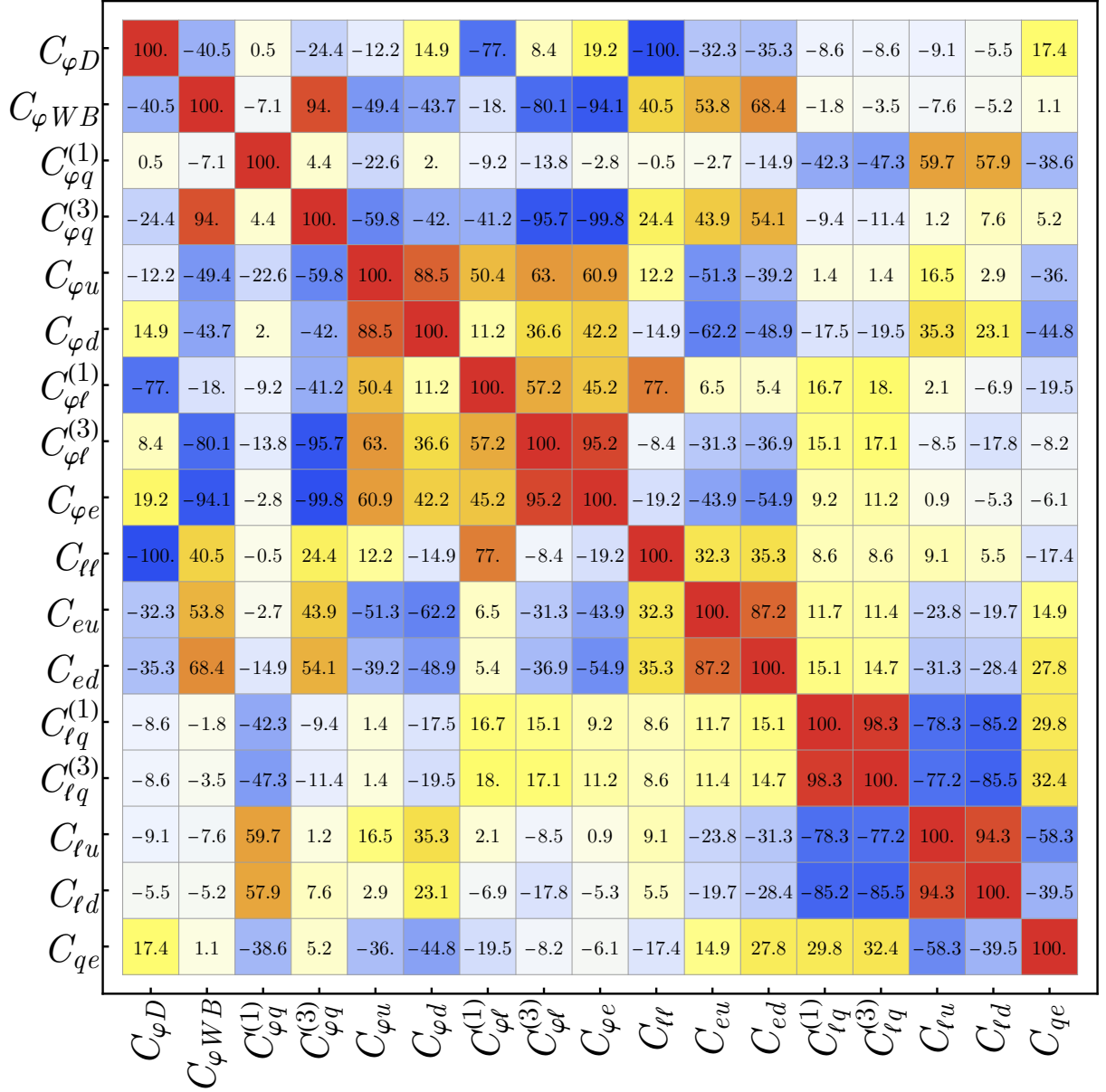


Figure 4.11. Correlation matrix of the 17d joint LHeC fit of Wilson coefficients.

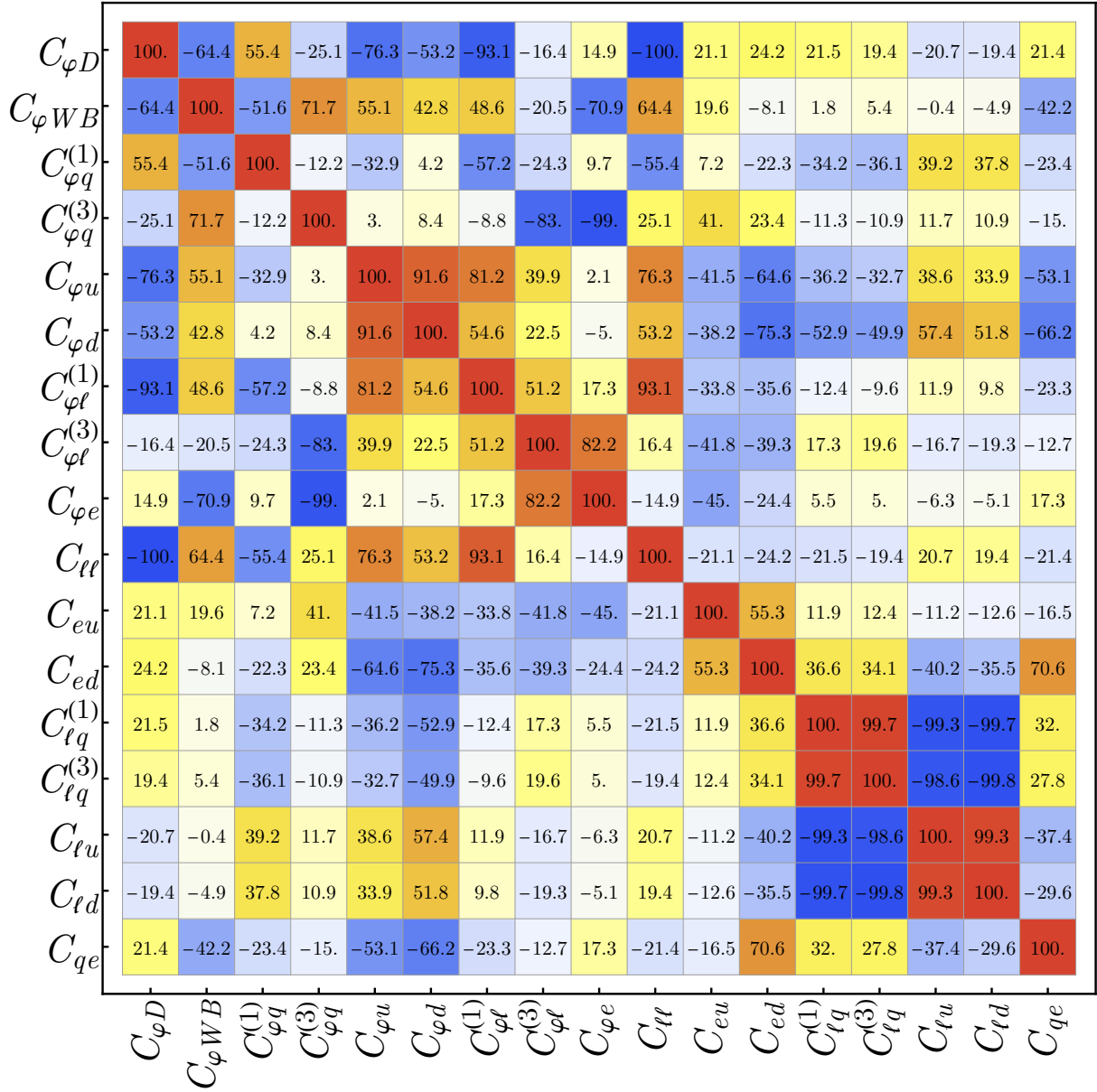


Figure 4.12. The same as in Figure 4.11 but for FCC-eh.

cases. Second, the  $ffV$  operators are only weakly correlated with the four-fermion operators, which justifies treating them separately in some analyses. Third, the EIC bounds barely reach 0.5 TeV and remain much weaker than those from the LHeC or the EW fits.

Two-dimensional projections of selected Wilson coefficients are shown in Figures 13 and 14. These include both non-marginalized and marginalized fits at 95 percent confidence level. The EW fits are taken from [92] for comparison. In most cases, the LHeC provides stronger bounds than the EW data. The FCC-eh pushes them even further. The EW fits show tight correlations and degeneracies, but these are largely resolved when using DIS data from the LHeC or FCC-eh. On the other hand, the EIC fits are by far the weakest and contribute little to probing the  $ffV$  parameter space.

#### 4.6. Coda

This chapter explored the potential of the EIC, LHeC, and FCC-eh to probe BSM physics using the SMEFT framework. We focused on NC DIS cross sections at the LHeC and FCC-eh, and PV observables at the EIC, following the general strategy used in earlier studies. Our analysis included the full set of dimension-6 operators that can modify the DIS amplitude, namely the semi-leptonic four-fermion operators and the vertex corrections to fermion couplings with neutral EW gauge bosons.

We examined a wide range of machine configurations, varying in energy, beam polarization, and lepton species. The results show that the EIC can reach effective scales up to 3 TeV, the LHeC up to 13 TeV, its joint runs up to 14 TeV, and the FCC-eh up to 18 TeV. No single configuration gives full coverage of the SMEFT parameter space. Multiple runs with different polarization and species setups are necessary. Most importantly, future DIS

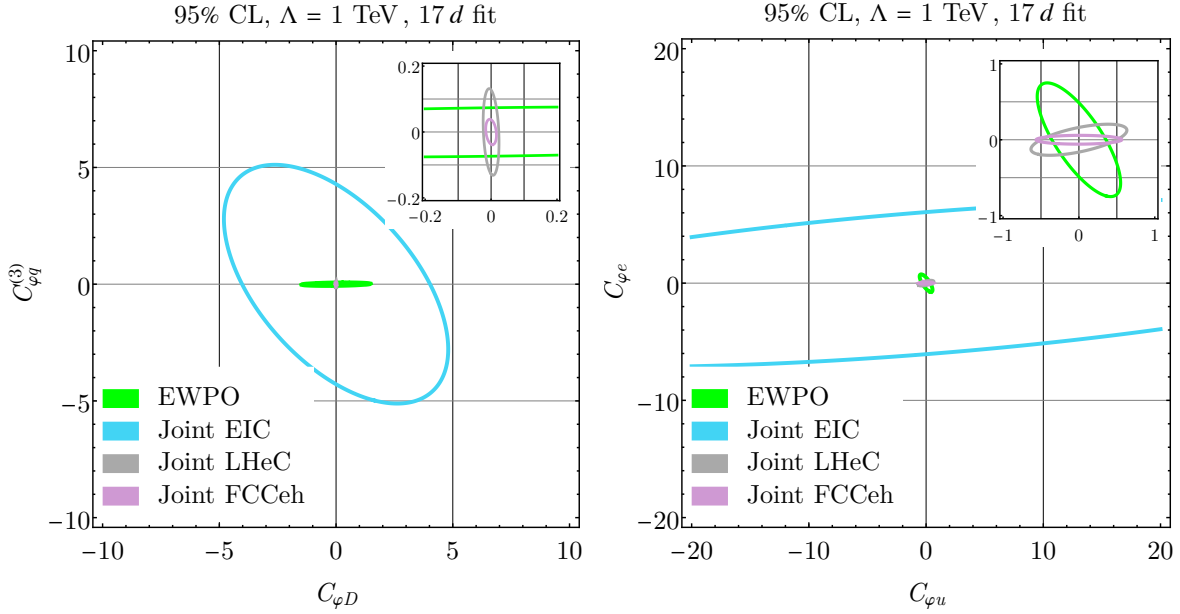


Figure 4.13. Marginalized 95% confidence level ellipses in the two-parameter fits of  $C_{\varphi D}$  and  $C_{\varphi e}^{(3)}$  (left) and  $C_{\varphi \ell}^{(1)}$  and  $C_{\varphi e}$  (right) at  $\Lambda = 1$  TeV.

The joint EIC, LHeC, and FCC-eh fits are shown, along with the EWPO fit adapted from [92].

data can lift the flat directions that limit EW precision fits. Among the three machines, the LHeC and FCC-eh provide the strongest reach.

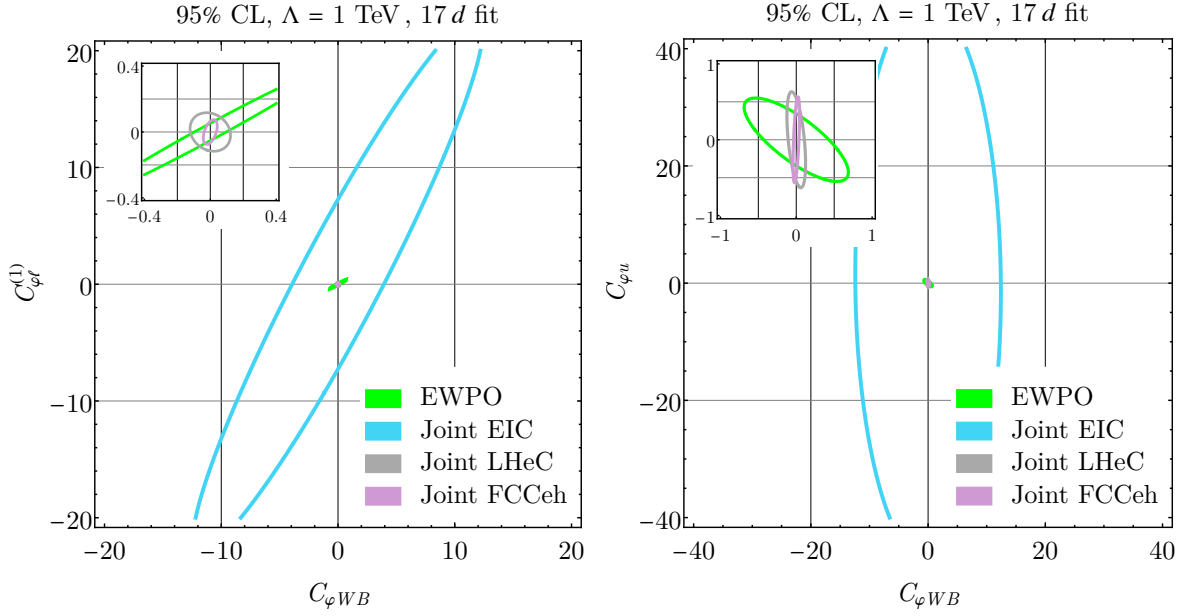


Figure 4.14. Marginalized 95% confidence level ellipses in the two-parameter fits of  $C_{\varphi D}$  and  $C_{\varphi e}$  (left) and  $C_{\varphi \ell}^{(1)}$  and  $C_{\varphi e}$  (right) at  $\Lambda = 1$  TeV. The joint EIC, LHeC, and FCC-eh fits are shown, along with the EWPO fit from [92].



## CHAPTER 5

**Transcendental Étude in  $e^-$  Minor**

*Church Painter: Why should one always make people happy? It might be a good idea to scare them once in a while.*

*Jöns: Then they'll close their eyes and refuse to look.*

*Church Painter: They'll look. A skull is more interesting than a naked woman.*

*Jöns: If you do scare them...*

*Church Painter: Then they think.*

*Jöns: And then?*

*Church Painter: They'll become more scared.*

Ingmar Bergman, *The Seventh Seal*

This chapter demonstrates how measuring single transverse-spin asymmetries at the Future Circular Collider operating in the electron-positron mode can enhance the sensitivity to the electron Yukawa coupling. We show that using transversely polarized electron, in both the  $b\bar{b}$  and semi-leptonic  $WW$  final states, the significance can increase up to three times compared to inclusive cross section methods. If positrons are also longitudinally polarized even at just 30%, the significance and improve by a factor of five or more. The method takes advantage of the quantum interference between the Higgs signal and the continuum background, providing a more precise measurement. This approach is also applicable to other  $WW$  and  $ZZ$  final states.

### 5.1. Prelude

The discovery of the Higgs boson marked a monumental milestone in our understanding of the Standard Model (SM) and the mechanism by which elementary particles acquire mass. However, despite over a decade of experimental progress, some of the most fundamental aspects of Higgs physics remain elusive, for example the coupling of the Higgs boson to the electron, which in the SM is predicted to be proportional to the electron mass and therefore vanishingly small. The electron Yukawa coupling,  $y_e$ , is the smallest among all SM fermion Yukawa couplings, with a value  $y_e^{\text{SM}} = \sqrt{2}m_e/v \approx 2.9 \times 10^{-6}$ , where  $v$  is the Higgs vacuum expectation value. Probing this coupling directly would represent one of the most stringent tests of the SM Higgs mechanism.

The current experimental bounds on  $y_e$  are orders of magnitude above the SM expectation. The most recent limits, derived from Drell-Yan production processes at the Large Hadron Collider (LHC), constrain  $|y_e| \leq 260|y_e^{\text{SM}}|$  at 95% confidence level [124, 7]. Even with the full High-Luminosity LHC dataset, this bound is projected to improve to no better than  $|y_e| \leq 120|y_e^{\text{SM}}|$  [64]. The primary obstacle is that the Higgs coupling to electrons enters the cross section quadratically and is heavily suppressed by the electron mass, making it extremely challenging to isolate from large SM backgrounds.

The electron Yukawa coupling can be accessed more directly at future electron-positron colliders operating near the Higgs resonance. The proposed Future Circular Collider in electron-positron mode (FCC-ee) is a promising candidate for such measurements. By running at  $\sqrt{s} = 125$  GeV with a planned integrated luminosity of  $10 \text{ ab}^{-1}$ , the FCC-ee can produce Higgs bosons in the  $s$ -channel via electron-positron annihilation. Several dedicated studies [106, 89, 88] have investigated the prospects of measuring  $y_e$  at

the FCC-ee, employing inclusive cross-section measurements and multivariate techniques. The most complete analysis to date suggests that an upper bound of  $|y_e| \leq 1.6|y_e^{\text{SM}}|$  could be achievable, provided that exceptional control over beam energy spread and detector resolution is maintained.

However, the inclusive cross-section approach suffers from intrinsic limitations. The cross section for s-channel Higgs production is proportional to  $y_e^2$ , further compounding the suppression associated with the electron mass. Additionally, large backgrounds from electroweak continuum processes challenge the statistical significance of such a measurement. This motivated the search for alternative observables that could provide enhanced sensitivity to  $y_e$ .

A promising avenue lies in exploiting the spin degrees of freedom of the initial-state electrons. Single transverse-spin asymmetries, well-studied in the context of deep-inelastic scattering and quantum chromodynamics (QCD), exhibit a chiral suppression proportional to the electron mass [139]. More importantly, these asymmetries arise from interference between the Higgs signal and the SM continuum background, leading to a dependence linear in  $y_e$  rather than quadratic. This crucial feature opens up a new strategy for probing the electron Yukawa coupling.

The theoretical framework underpinning these asymmetries is rooted in the discrete symmetries of the SM amplitudes. Transverse single-spin asymmetries are odd under the combined transformation of parity and naive time reversal. In the processes considered here, namely  $e^-e^+ \rightarrow b\bar{b}$  and  $e^-e^+ \rightarrow W^-W^+$ , the asymmetries emerge from the imaginary part of the interference between the Higgs-mediated amplitude and the continuum background. This imaginary part is generated primarily near the Higgs resonance

through Dyson resummation effects. As a result, the asymmetries exhibit a characteristic dependence on the azimuthal angle between the transverse spin direction of the electron and the final-state particles. Properly weighting the measured events with this angular dependence allows the signal to be isolated from background contributions.

Our study builds upon these theoretical insights and systematically investigates the potential of transverse spin asymmetries to probe the electron Yukawa coupling at the FCC-ee. We consider two final states: the  $b\bar{b}$  channel, which is clean and well-understood, and the semi-leptonic  $W^-W^+$  channel, which benefits from favorable kinematics and reduced QCD background. For each process, we construct asymmetry observables under various beam polarization configurations and assess their statistical significance. Our analysis accounts for realistic experimental effects, including beam energy spread and initial-state radiation, and incorporates optimized kinematic cuts to enhance sensitivity.

A key result of this analysis is that the use of transverse spin asymmetries significantly improves the prospects of observing the electron Yukawa coupling at the FCC-ee. In the  $b\bar{b}$  channel, we find that the significance can be improved by a factor of five compared to the inclusive cross-section analysis. In the semi-leptonic  $W^-W^+$  channel, the improvement reaches a factor of six, with the significance approaching the threshold for observation of the SM electron Yukawa coupling. These results illustrate the power of spin asymmetries as precision probes of fundamental SM parameters.

The broader theoretical implication of this work is that quantum interference effects, long known to play a subtle role in collider observables, can be harnessed strategically to

access otherwise inaccessible parameters. By coupling this approach with beam polarization, precision measurements at future colliders can extend beyond conventional limits, probing the chiral structure of the SM with unprecedented sensitivity.

This chapter is organized as follows. In Section 5.2, we introduce the processes considered, outline our theoretical framework clearly and explicitly, and define all relevant observables. Analytical expressions for the transverse-spin asymmetries are presented in Section 5.3. In Section 5.4, we explain how beam energy spread and initial-state radiation effects are incorporated into our analysis. Our primary sensitivity results and numerical estimates are provided and discussed in detail in Section 5.5. We conclude in Section 5.6.

## 5.2. Structure of the cross section

We consider the following processes of interest:

$$(5.1) \quad e^-(p_a) + e^+(p_b) \rightarrow b(p_1) + \bar{b}(p_2),$$

$$(5.2) \quad e^-(p_a) + e^+(p_b) \rightarrow W(p_{12}) + W(p_{34}) \rightarrow \ell(p_1) + \nu(p_2) + j(p_3) + j(p_4).$$

We refer to these processes simply as the  $b\bar{b}$  and  $WW$  processes, respectively. Analytical calculations are carried out with FeynArts [111] and FeynCalc [146, 145, 144, 138] using the unitary SM model. The only modification we introduce is the generalized projectors for the spinors [43]:

$$(5.3) \quad u_a \bar{u}_a = (\not{p}_a + m_e) \mathbb{P}_a, \quad v_b \bar{v}_b = (\not{p}_b - m_e) \mathbb{P}_b,$$

Here, the generalized projector for a particle/antiparticle in a spin state  $\lambda$  having the spin 4-vector  $s$  is given by

$$(5.4) \quad \mathbb{P}^\pm(\lambda, s) = \frac{1 \pm \lambda \gamma_5 \not{s}}{2}.$$

Then we have  $\mathbb{P}_a = \mathbb{P}^+(\lambda_a, s_a)$  and  $\mathbb{P}_b = \mathbb{P}^-(\lambda_b, s_b)$ . The electron and positron spin vectors are given respectively by  $s_a = (0, 1, 0, 0)$  and  $s_b = \frac{1}{m_e}(|\mathbf{p}_b|, E_b \hat{\mathbf{p}}_b)$ , where we use bold-face font to denote 3-vectors. The latter can also be expressed in terms of the incoming momenta as  $s_b = c_a p_a + c_b p_b$ , where

$$(5.5) \quad c_a = -\frac{2m_e}{\mathcal{E}\sqrt{\mathcal{E}^2 - 4m_e^2}}, \quad c_b = -\frac{2m_e^2 - \mathcal{E}^2}{m_e \mathcal{E}\sqrt{\mathcal{E}^2 - 4m_e^2}},$$

so that  $p_a^2 = p_b^2 = m_e^2$ ,  $s_a^2 = s_b^2 = -1$ ,  $p_a \cdot s_a = p_b \cdot s_b = 0$ ,  $s_a \cdot s_b = 0$ , and  $p_a \cdot s_b = \frac{\mathcal{E}\sqrt{\mathcal{E}^2 - 4m_e^2}}{2m_e}$ .

Here,  $\mathcal{E}$  denotes the center-of-mass (c.m.) energy.

Once we derive the squared amplitude, the cross section is obtained by

$$(5.6) \quad \sigma = F \int \text{dLIPS} \quad |\overline{\mathcal{A}}|^2,$$

where  $F = 1/(2\mathcal{E}^2)$  is the flux factor in the small- $m_e$  approximation and  $|\overline{\mathcal{A}}|^2$  is the squared amplitude with the spin sum over the outgoing fermions. The Lorentz-invariant phase space (LIPS) is given by

$$(5.7) \quad \int \text{dLIPS} = \frac{1}{32\pi^2} \int \text{d}\Omega$$

for the  $b\bar{b}$  process and

$$(5.8) \quad \int d\text{LIPS} = \frac{1}{2(4\pi)^8 \mathcal{E}^2} \int_0^\mathcal{E} dm_{12} \, 2m_{12} \int_0^{\mathcal{E}-m_{12}} dm_{34} \, 2m_{34} \sqrt{\lambda(\mathcal{E}^2, m_{12}^2, m_{34}^2)} \\ \times \int d\Omega_{12} \int d\Omega_1 \int d\Omega_3$$

for the  $WW$  process. Here,  $m_{12(34)}$  is the invariant mass of the  $W$  boson that decays into the lepton-neutrino pair (jets).

Let us detail the geometries of the  $b\bar{b}$  and  $WW$  processes and provide the expressions for various 4-momenta involved. Figures 5.1 and 5.2 depict said geometries for the  $b\bar{b}$  and  $WW$  processes, respectively. In the latter, the momenta with asterisk (circle) are the ones as measured in the rest frame of the  $W$  boson that decays into the lepton-neutrino pair (jets). We emphasize that the momenta  $\mathbf{p}_1$  and  $\mathbf{p}_2$  for the  $b\bar{b}$  process and  $\mathbf{p}_{12}$  and  $\mathbf{p}_{34}$  are not restricted to the  $xz$  plane just because the drawings indicate so; that is, these vectors have nontrivial azimuthal angles  $\varphi$  and  $\varphi_{12}$ , respectively.

The 4-momenta for the external particles in the  $b\bar{b}$  process are given by  $p_i = (E_i, \mathbf{p}_i)$  for  $i = a, b, 1, 2$ , where

$$(5.9) \quad E_{a/b} = \frac{\mathcal{E}^2 + m_{a/b}^2 - m_{b/a}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{a/b}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_a^2, m_b^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{a/b} = \pm(0, 0, 1),$$

$$(5.10) \quad E_{1/2} = \frac{\mathcal{E}^2 + m_{1/2}^2 - m_{2/1}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{1/2}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_1^2, m_2^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{1/2} = \pm(s_\theta c_\varphi, s_\theta s_\varphi, c_\theta).$$

where  $m_a = m_b = m_e$ ,  $m_1 = m_2 = m_b$ , and  $s$  and  $c$  are the sine and cosine functions of the angle denoted in the subscript, respectively.

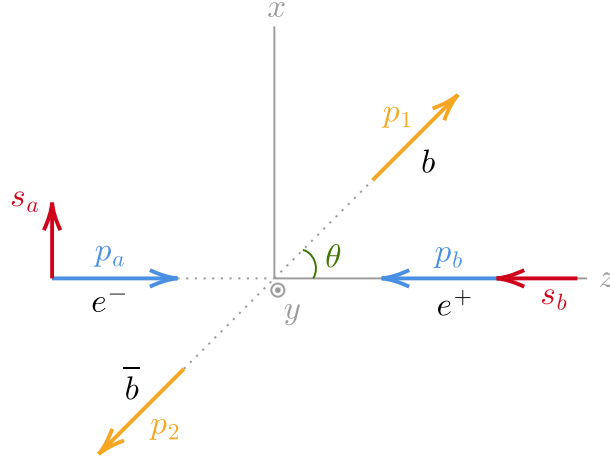


Figure 5.1. The collision of a transversely-polarized electron beam with a longitudinally-polarized positron beam for the  $b\bar{b}$  process in the center-of-mass frame of the incoming particles. Note that the vectors  $\mathbf{p}_i$  for  $i = 1, 2$  are not restricted to the  $xz$  plane.

The components of the 4-momenta relevant to the  $WW$  process are given similarly by

$$(5.11) \quad E_{a/b} = \frac{\mathcal{E}^2 + m_{a/b}^2 - m_{b/a}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{a/b}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_a^2, m_b^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{a/b} = \pm(0, 0, 1),$$

$$(5.12) \quad E_{12/34} = \frac{\mathcal{E}^2 + m_{12/34}^2 - m_{34/12}^2}{2\mathcal{E}},$$

$$(5.13) \quad |\mathbf{p}_{12/34}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_{12}^2, m_{34}^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{12/34} = \pm(s_{\theta_{12}}c_{\varphi_{12}}, s_{\theta_{12}}s_{\varphi_{12}}, c_{\theta_{12}}),$$

$$(5.14) \quad E_{12}^* = m_{12}, \quad \mathbf{p}_{12}^* = \mathbf{0},$$

$$(5.15) \quad E_{1/2}^* = \frac{m_{12}^2 + m_{1/2}^2 - m_{2/1}^2}{2m_{12}},$$

$$(5.16) \quad |\mathbf{p}_{1/2}^*| = \frac{\sqrt{\lambda(m_{12}^2, m_1^2, m_2^2)}}{2m_{12}}, \quad \hat{\mathbf{p}}_{1/2}^* = \pm(s_{\theta_1}c_{\varphi_1}, s_{\theta_1}s_{\varphi_1}, c_{\theta_1}),$$

$$(5.17) \quad E_{34}^\circ = m_{34}, \quad \mathbf{p}_{34}^\circ = \mathbf{0},$$



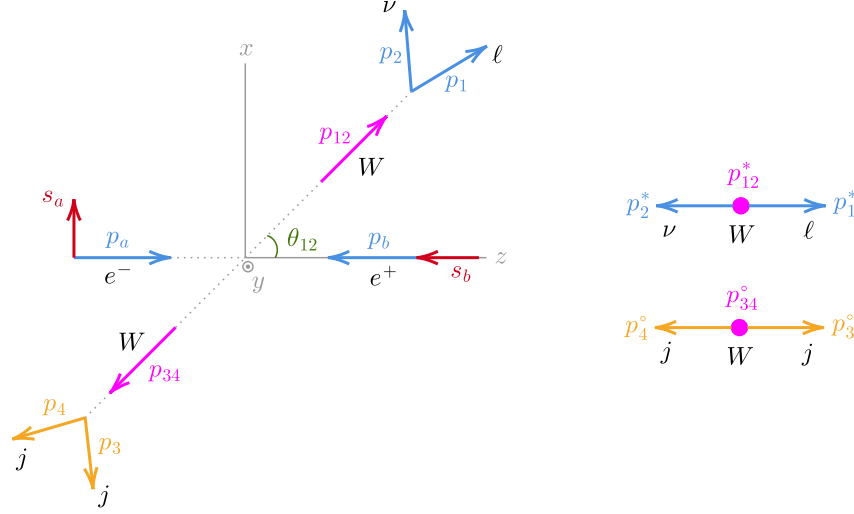


Figure 5.2. The collision of a transversely-polarized electron beam with a longitudinally-polarized positron beam for the  $WW$  process in the center-of-mass frame of the incoming particles on the left. On the right, we draw the decays of the  $W$  bosons into a lepton-neutrino pair or two jets in their respective center-of-mass frames. Note that the vectors  $\mathbf{p}_i$  for  $i = 12, 34, 1, 2, 3, 4$  are not restricted to the  $xz$  plane.

$$(5.18) \quad E_{3/4}^{\circ} = \frac{m_{34}^2 + m_{3/4}^2 - m_{4/3}^2}{2m_{34}},$$

$$(5.19) \quad |\mathbf{p}_{3/4}^{\circ}| = \frac{\sqrt{\lambda(m_{34}^2, m_3^2, m_4^2)}}{2m_{34}}, \quad \hat{\mathbf{p}}_{3/4}^{\circ} = \pm(s_{\theta_3}c_{\varphi_3}, s_{\theta_3}s_{\varphi_3}, c_{\theta_3}).$$

where  $m_a = m_b = m_e$  and  $m_1 = m_2 = m_3 = m_4 = 0$ . The  $p_i$  for  $i = 1, 2, 3, 4$  can then be obtained with a Lorentz boost from their respective frames into the c.m. frame of  $p_a$  and  $p_b$ . For the “12” system, we have

$$(5.20) \quad p_{12}^* = \Lambda_{12} p_{12} = B_{12} R_{12,y} R_{12,z} p_{12},$$

where

$$(5.21) \quad B_{12} = \begin{pmatrix} \gamma_{12} & 0 & 0 & -\gamma_{12}\beta_{12} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_{12}\beta_{12} & 0 & 0 & \gamma_{12} \end{pmatrix},$$

$$(5.22) \quad R_{12,y} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & c_{\theta_{12}} & 0 & -s_{\theta_{12}} \\ 0 & 0 & 1 & 0 \\ 0 & s_{\theta_{12}} & 0 & c_{\theta_{12}} \end{pmatrix},$$

$$(5.23) \quad R_{12,z} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & c_{\varphi_{12}} & s_{\varphi_{12}} & 0 \\ 0 & -s_{\varphi_{12}} & c_{\varphi_{12}} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which we can solve for  $\gamma_{12}$  and  $\beta_{12}$  to find

$$(5.24) \quad \gamma_{12} = \frac{E_{12}}{m_{12}}, \quad \beta_{12} = \frac{|\mathbf{p}_{12}|}{E_{12}}.$$

For the “34” system, we have

$$(5.25) \quad p_{34}^{\circ} = \Lambda_{34} p_{34} = B_{34} R_{34,y} R_{34,z} p_{34},$$

where

$$(5.26) \quad B_{34} = \begin{pmatrix} \gamma_{34} & 0 & 0 & -\gamma_{34}\beta_{34} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_{34}\beta_{34} & 0 & 0 & \gamma_{34} \end{pmatrix}, \quad R_{34,y} = R_{12,y}, \quad R_{34,z} = R_{12,z},$$

which we can solve for  $\gamma_{34}$  and  $\beta_{34}$  to find

$$(5.27) \quad \gamma_{34} = \frac{E_{34}}{m_{34}}, \quad \beta_{34} = -\frac{|\mathbf{p}_{34}|}{E_{34}}.$$

This allows us to write  $p_{1/2} = \Lambda_{12}^{-1} p_{1/2}^*$  and  $p_{3/4} = \Lambda_{34}^{-1} p_{3/4}^\circ$ . Finally, we expand each fermion momentum with respect to the corresponding fermion mass as

$$(5.28) \quad p_i = p_i^{(0)} + m_i^2 p_i^{(2)}$$

for  $i = a, b, 1, 2$  for the  $b\bar{b}$  process and  $i = a, b$  for the  $WW$  process. Next, we present an intuitive discussion on the derivation of many-particle LIPS. Our intention is not to state that this is how we derive the LIPS, but rather to provide a generalizable way to construct the many-particle LIPS from scratch iteratively.

We note that our expressions are consistent with the general formula

$$(5.29) \quad \text{LIPS}_n = \frac{1}{2(4\pi)^{2n-3}} \frac{\mathcal{E}^{2n-4}}{\Gamma(n)\Gamma(n-1)}$$

after integration for the case of massless outgoing particles. We also confirm that our numbers for the unpolarized cross sections for the background  $b\bar{b}$  and  $WW$  processes in the absence of any cuts match our MADGRAPH [24] simulations, as well as the relevant

numbers quoted in Table 1 of [88], verifying not only our expressions for the LIPS but also for the 4-momenta.

For the  $b\bar{b}$  process, we keep the mass of the  $b$  quark to first order and for the  $WW$  process, we take the outgoing fermions to be massless. For both, we keep the mass of the electron to first order. We keep the vertex factors in closed form so as to distinguish between the fermion masses that derive from the equation of motion and from the Yukawa couplings.

Our observable is the polarization asymmetry,  $A = N/D$ , where  $N$  is the differences between cross-section measurements with distinct incoming beam polarizations, which are denoted by  $\sigma^{\lambda_a\lambda_b}$  where  $\lambda_a$  and  $\lambda_b$  can take on values  $+1$ ,  $0$ , and  $-1$ , and  $D$  is the sum of said cross-section measurements. We investigate various polarization asymmetries constructed as follows:

- The double-polarization asymmetry (DP):

$$(5.30) \quad N = \frac{1}{4}(\sigma^{++} - \sigma^{+-} - \sigma^{-+} + \sigma^{--}),$$

$$(5.31) \quad D = \frac{1}{4}(\sigma^{++} + \sigma^{+-} + \sigma^{-+} + \sigma^{--}).$$

- The single-polarization asymmetry with an unpolarized positron beam,  $\lambda_b = 0$  (SP<sup>0</sup>):

$$(5.32) \quad N = \frac{1}{2}(\sigma^{+0} - \sigma^{-0}),$$

$$(5.33) \quad D = \frac{1}{2}(\sigma^{+0} + \sigma^{-0}).$$

- The single-polarization asymmetry with a left-handed positron beam,  $\lambda_b = +1$  (SP<sup>+</sup>):

$$(5.34) \quad N = \frac{1}{2}(\sigma^{++} - \sigma^{-+}),$$

$$(5.35) \quad D = \frac{1}{2}(\sigma^{++} + \sigma^{-+}).$$

- The single-polarization asymmetry with a right-handed positron beam,  $\lambda_b = -1$  (SP<sup>-</sup>):

$$(5.36) \quad N = \frac{1}{2}(\sigma^{+-} - \sigma^{--}),$$

$$(5.37) \quad D = \frac{1}{2}(\sigma^{+-} + \sigma^{--}).$$

We note that the angular weight functions  $\sin(\varphi)$  and  $\pm \sin(\varphi_{12})$  are introduced to the phase space whilst we form the numerator of the asymmetry of the  $b\bar{b}$  and  $WW$  processes, respectively, where the sign of the latter is opposite to the sign of the electric charge of the outgoing lepton,  $\ell^\mp$ . To see this, we need to understand the analytical expressions for the numerators of the asymmetries, which is what we do next.

### 5.3. Analytical calculations

We start with the  $b\bar{b}$  process. In general, we write

$$(5.38) \quad N = F \int \text{dLIPS} \ w \sum_c N^c,$$

$$(5.39) \quad D = F \int \text{dLIPS} \ \sum_c D^c,$$

where  $w$  is a weight function and the summation runs over all the interference channels. For the  $b\bar{b}$  process, we have  $c \in \{h, \gamma, Z, h\gamma, hZ, \gamma Z\}$ . To avoid clutter, we present the expressions only for the primary observable of interest, which is the  $SP^-$  asymmetry. We have

$$(5.40) \quad N^h = 0,$$

$$(5.41) \quad N^\gamma = -6\mathcal{E}^3 \Delta_\gamma^2 m_e \sin(2\theta) \cos(\varphi) C_{bb\gamma}^2 C_{ee\gamma}^2,$$

$$(5.42) \quad N^Z = -3\mathcal{E}^3 m_e \sin(\theta) \cos(\varphi) (\Delta_Z^i{}^2 + \Delta_Z^r{}^2) C_{eeZ}^L (C_{eeZ}^L + C_{eeZ}^R) \\ \times (\cos(\theta) (C_{bbZ}^L{}^2 + C_{bbZ}^R{}^2) + C_{bbZ}^L{}^2 - C_{bbZ}^R{}^2),$$

$$(5.43) \quad N^{h\gamma} = 12\mathcal{E}^3 m_b \Delta_\gamma \sin(\theta) C_{bbh} C_{bb\gamma} C_{eeh} C_{ee\gamma} (\sin(\varphi) \Delta_h^i + \cos(\varphi) \Delta_h^r),$$

$$(5.44) \quad N^{hZ} = 6\mathcal{E}^3 m_b \sin(\theta) C_{bbh} C_{eeh} C_{eeZ}^L (C_{bbZ}^L + C_{bbZ}^R) \\ \times (\sin(\varphi) (\Delta_h^i \Delta_Z^r - \Delta_h^r \Delta_Z^i) + \cos(\varphi) (\Delta_h^i \Delta_Z^i + \Delta_h^r \Delta_Z^r)),$$

$$(5.45) \quad N^{\gamma Z} = -3\mathcal{E}^3 \Delta_\gamma m_e \sin(\theta) C_{bb\gamma} C_{ee\gamma} (\cos(\theta) (C_{bbZ}^L + C_{bbZ}^R) + C_{bbZ}^L - C_{bbZ}^R) \\ \times (\sin(\varphi) \Delta_Z^i (C_{eeZ}^R - C_{eeZ}^L) + \cos(\varphi) \Delta_Z^r (3C_{eeZ}^L + C_{eeZ}^R)),$$

and

$$(5.46) \quad D^h = 3\mathcal{E}^4 C_{bbh}^2 C_{eeh}^2 (\Delta_h^i{}^2 + \Delta_h^r{}^2),$$

$$(5.47) \quad D^\gamma = \frac{3}{2} \mathcal{E}^4 \Delta_\gamma^2 (\cos(2\theta) + 3) C_{bb\gamma}^2 C_{ee\gamma}^2,$$

$$(5.48) \quad D^Z = \frac{3}{4} \mathcal{E}^4 (\Delta_Z^i{}^2 + \Delta_Z^r{}^2) C_{eeZ}^L{}^2 (4 \cos(\theta) (C_{bbZ}^L - C_{bbZ}^R) (C_{bbZ}^L + C_{bbZ}^R) \\ + (\cos(2\theta) + 3) (C_{bbZ}^L{}^2 + C_{bbZ}^R{}^2)),$$

$$(5.49) \quad D^{h\gamma} = 24\mathcal{E}^2 m_b \Delta_\gamma m_e \cos(\theta) C_{bbh} C_{bb\gamma} C_{eeh} C_{ee\gamma} \Delta_h^r,$$

$$(5.50) \quad D^{hZ} = 6\mathcal{E}^2 m_b m_e \cos(\theta) C_{bbh} C_{eeh} (\Delta_h^i \Delta_Z^i + \Delta_h^r \Delta_Z^r) (C_{bbZ}^L + C_{bbZ}^R) (C_{eeZ}^L + C_{eeZ}^R),$$

$$(5.51) \quad D^{\gamma Z} = \frac{3}{2} \mathcal{E}^4 \Delta_\gamma C_{bb\gamma} C_{ee\gamma} \Delta_Z^r C_{eeZ}^L \\ \times (4 \cos(\theta) (C_{bbZ}^L - C_{bbZ}^R) + (\cos(2\theta) + 3) (C_{bbZ}^L + C_{bbZ}^R)).$$

Here,  $\Delta_h$ ,  $\Delta_\gamma$ , and  $\Delta_Z$  are the propagators, given by

$$(5.52) \quad \Delta_h = \frac{1}{\mathcal{E}^2 - m_h^2 + im_h \Gamma_h}, \quad \Delta_\gamma = \frac{1}{\mathcal{E}^2}, \quad \Delta_Z = \frac{1}{\mathcal{E}^2 - m_Z^2 + im_Z \Gamma_Z},$$

$\Delta^{r/i}$  are the real/imaginary parts of said propagators, and numerous  $C$  factors are the coupling strengths of indicated interactions in the subscripts.

We are interested in isolating the electron Yukawa coupling, denoted by  $C_{eeh}$ , in (5.40)-(5.45) near the Higgs resonance. With small  $\Delta_h^r$  and small  $\Delta_Z^i$ , we see that all the channels go like  $\cos(\varphi)$ , except for the Higgs- $Z$  boson interference, which is proportional to  $\sin(\varphi)$  due to the coupling of the imaginary part of the Higgs propagator to the real part of the  $Z$ -boson propagator. This channel can be isolated with the weight function  $w = \sin(\varphi)$ .

We investigate the on-shell  $WW$  production to have a decent understanding of the full, more complicated  $WW$  process in regards to determining a proper weight function. Thus, the process of interest is now

$$(5.53) \quad e^-(p_a) + e^+(p_b) \rightarrow W(p_1) + W(p_2).$$

At tree level, this process takes places via three  $s$ -channel diagrams with the Higgs, photon, and the  $Z$ -boson emission, as well as a  $t$ -channel diagram with a neutrino exchange. Here,

the  $p_i$  ( $i = a, b, 1, 2$ ) have components

$$(5.54) \quad E_{a/b} = \frac{\mathcal{E}^2 + m_{a/b}^2 - m_{b/a}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{a/b}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_a^2, m_b^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{a/b} = \pm(0, 0, 1),$$

$$(5.55) \quad E_{1/2} = \frac{\mathcal{E}^2 + m_{1/2}^2 - m_{2/1}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{1/2}| = \frac{\sqrt{\lambda(\mathcal{E}^2, m_1^2, m_2^2)}}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{1/2} = \pm(s_\theta c_\varphi, s_\theta s_\varphi, c_\theta),$$

where  $m_a = m_b = m_e$  and  $m_1 = m_2 = m_W$ . We focus on the numerators of the various asymmetries of interest. We write

$$(5.56) \quad N = F \int d\text{LIPS} \sum_c N^c,$$

where  $c = h, \gamma, Z, \nu, h\gamma, hZ, h\nu, \gamma Z, \gamma\nu, Z\nu$ . To avoid clutter, we present the expressions only for the primary observable of interest, which is the  $\text{SP}^+$  asymmetry. We have

$$(5.57) \quad N^h = 0,$$

$$(5.58) \quad \begin{aligned} N^\gamma &= \frac{1}{4m_W^4} (\mathcal{E} \Delta_\gamma^2 m_e \sin(2\theta) \cos(\varphi) (\mathcal{E}^2 - 4m_W^2) \\ &\times (-4\mathcal{E}^2 m_W^2 + 12m_W^4 + \mathcal{E}^4) C_{ee\gamma}^2 (-C_{\gamma WW}^2)), \\ N^Z &= \frac{1}{8m_W^4} (\mathcal{E} m_e \sin(2\theta) \cos(\varphi) (28\mathcal{E}^2 m_W^4 - 8\mathcal{E}^4 m_W^2 - 48m_W^6 + \mathcal{E}^6) \end{aligned}$$

$$(5.59) \quad \times C_{ZWW}^2 (\Delta_Z^{i2} + \Delta_Z^{r2}) (-C_{eeZ}^R) (C_{eeZ}^L + C_{eeZ}^R),$$

$$(5.60) \quad N^\nu = 0,$$

$$(5.61) \quad \begin{aligned} N^{h\gamma} &= \frac{1}{4m_W^4} (\mathcal{E}^2 \Delta_\gamma \sin(\theta) \sqrt{\mathcal{E}^2 - 4m_W^2} (\mathcal{E}^4 - 12m_W^4) C_{eeh} C_{ee\gamma} \\ &\times C_{hWW} (-C_{\gamma WW}) (\cos(\varphi) \Delta_h^r - \sin(\varphi) \Delta_h^i)), \end{aligned}$$

$$N^{hZ} = \frac{1}{4m_W^4} (\mathcal{E}^2 \sin(\theta) \sqrt{\mathcal{E}^2 - 4m_W^2} (\mathcal{E}^4 - 12m_W^4) (-C_{eeh}))$$



$$(5.62) \quad \times C_{hWW}C_{ZWW}C_{eeZ}^R(\sin(\varphi)(\Delta_h^r\Delta_Z^i - \Delta_h^i\Delta_Z^r) + \cos(\varphi)(\Delta_h^i\Delta_Z^i + \Delta_h^r\Delta_Z^r)),$$

$$(5.63) \quad N^{h\nu} = 0,$$

$$(5.64) \quad N^{\gamma Z} = \frac{1}{8m_W^4}(\mathcal{E}\Delta_\gamma m_e \sin(2\theta)(28\mathcal{E}^2 m_W^4 - 8\mathcal{E}^4 m_W^2 - 48m_W^6 + \mathcal{E}^6)C_{ee\gamma} \\ \times (-C_{\gamma WW})C_{ZWW}(\sin(\varphi)\Delta_Z^i(C_{eeZ}^R - C_{eeZ}^L) + \cos(\varphi)\Delta_Z^r(C_{eeZ}^L + 3C_{eeZ}^R))),$$

$$(5.65) \quad N^{\gamma\nu} = \frac{1}{16m_W^4}(\mathcal{E}\Delta_\gamma\Delta_\nu m_e \sin(\theta) \cos(\varphi)C_{ee\gamma}C_{\gamma WW}(\mathcal{E}\sqrt{\mathcal{E}^2 - 4m_W^2} \\ \times (-\cos(2\theta)(-6\mathcal{E}^2 m_W^2 + 8m_W^4 + \mathcal{E}^4) - 2\mathcal{E}^2 m_W^2 + 16m_W^4 \\ + \mathcal{E}^4) - 4\cos(\theta)m_W^2(-10\mathcal{E}^2 m_W^2 + 24m_W^4 + \mathcal{E}^4))C_{ff'W^2}),$$

$$(5.66) \quad N^{Z\nu} = \frac{1}{16m_W^4}(\mathcal{E}\Delta_\nu m_e \sin(\theta)C_{ZWW}(4\cos(\theta)m_W^2(-10\mathcal{E}^2 m_W^2 + 24m_W^4 + \mathcal{E}^4) \\ + \mathcal{E}\sqrt{\mathcal{E}^2 - 4m_W^2}(\cos(2\theta)(-6\mathcal{E}^2 m_W^2 + 8m_W^4 + \mathcal{E}^4) + 2\mathcal{E}^2 m_W^2 \\ - 16m_W^4 - \mathcal{E}^4))C_{ff'W^2}(-C_{eeZ}^R)(\sin(\varphi)\Delta_Z^i + \cos(\varphi)\Delta_Z^r)).$$

Here,  $\Delta_h$ ,  $\Delta_\gamma$ ,  $\Delta_Z$ , and  $\Delta_\nu$  are the propagators given by

$$(5.67) \quad \Delta_h = \frac{1}{\mathcal{E}^2 - m_h^2 + im_h\Gamma_h}, \quad \Delta_\gamma = \frac{1}{\mathcal{E}^2},$$

$$(5.68) \quad \Delta_Z = \frac{1}{\mathcal{E}^2 - m_Z^2 + im_Z\Gamma_Z}, \quad \Delta_\nu = \frac{1}{(p_a - p_1)^2},$$

and  $\Delta^{r/i}$  are the real/imaginary parts of said denominators. Numerous  $C$  factors denote the coupling strengths of interactions indicated in the subscripts and are given in the appendix. One should be scrupulous with the collider energy here because now we cannot simply set  $\mathcal{E} = m_h$  because of the two outgoing  $W$  bosons. Thus, what we have here serves only as a crude estimation of the bigger picture. Nevertheless, we observe that in

all these expressions,  $w(\varphi) = \sin(\varphi)$  is again a good weight function that will isolate the electron Yukawa coupling,  $C_{eeh}$ , near the Higgs resonance, i.e. with negligible  $\Delta_h^r$ . Be that as it may, this choice of the angular weight does not perfectly isolate the Higgs- $Z$  boson interference as in the case of the  $b\bar{b}$  process; namely, we now have contributions from the Higgs-photon and Higgs-neutrino interference amplitudes, as well. One could try to find a weight function that would eliminate said contributions; however, it is a nontrivial task due to the  $\cos(\theta)$  term in the denominator of the neutrino propagator. Even in the case where the  $h\nu$  interference could be removed, we still observe that the  $h\gamma$  and  $hZ$  channels have identical dependence on  $\theta$ . This means that any attempt to exterminate the  $h\gamma$  interference by introducing an angular weight that also depends on  $\theta$  would annihilate the  $hZ$  interference, as well. As we observe later, the chosen weight function when generalized to the full, more involved  $WW$  process serves its purpose nearly perfectly.

#### 5.4. Dilution of the signal

To make a realistic estimate of the potential significance at an FCC, especially for a study that needs the beam collision energy to be tuned near the Higgs resonance, it is important to include the effects of beam energy spread and initial-state radiation. We account for the beam-spreading (BS) and initial-state-radiation (ISR) effects with the following convolution:

$$(5.69) \quad \sigma(\mathcal{E}_{\text{coll}}) = \int_{-\infty}^{\infty} d\hat{\mathcal{E}} \frac{d\mathcal{L}(\mathcal{E}_{\text{coll}}, \hat{\mathcal{E}}, \delta)}{d\hat{\mathcal{E}}} \int_0^1 dx \, f(x, \hat{\mathcal{E}}) \sigma(\sqrt{x}\hat{\mathcal{E}}),$$

where the BS is characterized by a relativistic Voigtian function [130]

$$(5.70) \quad \frac{d\mathcal{L}(\mathcal{E}_{\text{coll}}, \hat{\mathcal{E}}, \delta)}{d\hat{\mathcal{E}}} = \frac{1}{\sqrt{2\pi}\delta^2} \exp \left[ -\frac{(\hat{\mathcal{E}} - \mathcal{E}_{\text{coll}})^2}{2\delta^2} \right].$$

Here,  $\mathcal{E}_{\text{coll}}$  is the collider energy and  $\delta$  is the c.m. energy spread. In the meantime, we use the Jadach-Ward-Was ISR function [120, 121] for its popularity in similar studies [106, 81]:

$$(5.71) \quad f(x, \hat{\mathcal{E}}) = \exp \left[ \frac{\beta_e}{4} + \frac{\alpha}{\pi} \left( -\frac{1}{2} + \frac{\pi^2}{3} \right) \right] \frac{\exp(-\gamma\beta_e)}{\Gamma(1 + \beta_e)} \\ \times \beta_e (1-x)^{\beta_e-1} \left[ 1 + \frac{\beta_e}{2} - \frac{1}{2}(1-x)^2 \right],$$

with  $\gamma$  being the Euler-Mascheroni constant and

$$(5.72) \quad \beta_e = \frac{2\alpha}{\pi} \left[ \ln \left( \frac{\hat{\mathcal{E}}^2}{m_e^2} \right) - 1 \right].$$

We carry out numerical integrations using VEGAS [133, 132]. For the BS convolution, we perform the integration from  $\hat{\mathcal{E}}_{\text{min}}$  to  $\hat{\mathcal{E}}_{\text{max}}$  by transforming the integral to one over the unit interval via the following change of variables:

$$(5.73) \quad \hat{\mathcal{E}} = \sqrt{2}\delta\hat{\mathcal{F}} + \mathcal{E}_{\text{coll}}, \quad \hat{\mathcal{F}} = \tan(\hat{\psi}),$$

$$(5.74) \quad \hat{\psi} = 2 \arctan(\hat{\mathcal{G}}), \quad \hat{\mathcal{G}} = (\hat{\mathcal{G}}_{\text{max}} - \hat{\mathcal{G}}_{\text{min}})v_{\hat{\mathcal{E}}} + \hat{\mathcal{G}}_{\text{min}},$$

where  $v_{\hat{\mathcal{E}}} \in [0, 1]$  and

$$(5.75) \quad \hat{\mathcal{G}}_{\text{max/min}} = \tan \left( \frac{\hat{\psi}_{\text{max/min}}}{2} \right),$$

$$(5.76) \quad \hat{\psi}_{\max/\min} = \arctan(\hat{\mathcal{F}}_{\max/\min}),$$

$$(5.77) \quad \hat{\mathcal{F}}_{\max/\min} = \frac{\hat{\mathcal{E}}_{\max/\min} - \mathcal{E}_{\text{coll}}}{\sqrt{2}\delta}.$$

These changes of variables introduce the Jacobian

$$(5.78) \quad J_{\hat{\mathcal{E}}} = \left[ \sqrt{2}\delta \right] \left[ \sec(\hat{\psi})^2 \right] \left[ \frac{2}{\hat{\mathcal{G}}^2 + 1} \right] \left[ \hat{\mathcal{G}}_{\max} - \hat{\mathcal{G}}_{\min} \right].$$

Here, we assume  $\hat{\mathcal{E}}_{\min} = 110$  GeV and  $\hat{\mathcal{E}}_{\max} = 140$  GeV. Our analysis shows that these limits are practically equivalent to integrating  $\hat{\mathcal{E}}$  over the entire real line, for the integrands are highly peaked around the Higgs mass.

As for the ISR convolution, we perform the following change of variables to stabilize the Monte Carlo integration routine:

$$(5.79) \quad x = 1 - e^{-\alpha_x u}, \quad u = \tan(y), \quad y = \frac{\pi}{2}v_x,$$

where  $v_x \in [0, 1]$ . This introduces the Jacobian

$$(5.80) \quad J_x = \left[ \alpha_x e^{-\alpha_x u} \right] \left[ \sec(y)^2 \right] \left[ \frac{\pi}{2} \right].$$

Our analysis shows that  $\alpha_x = 4.5$  yields the highest stability by essentially flattening the ISR function to provide uniformity as a means of importance sampling.

### 5.5. Sensitivity estimates

The experimental reconstruction of the asymmetry is given by

$$(5.81) \quad A^{\text{exp}} = \frac{1}{P_{e^-} P_{e^+}} \frac{N_N}{N_D}$$

when both incoming beams are polarized and

$$(5.82) \quad A^{\text{exp}} = \frac{1}{P_{e^-}} \frac{N_N}{N_D}$$

when only the electron beam is polarized. Here,  $P_{e^\mp}$  is the electron/positron beam polarization reach at the collider, and  $N_N = \eta L N$  and  $N_D = \eta L D$  are the event counts, where  $\eta$  is the acceptance/efficiency and  $L$  is the integrated luminosity. In the limit of small asymmetry, the error in  $A^{\text{exp}}$  is given by

$$(5.83) \quad \delta A^{\text{exp}} = \frac{\delta P_{e^-}}{P_{e^-}} A^{\text{exp}} \oplus \frac{\delta P_{e^+}}{P_{e^+}} A^{\text{exp}} \oplus \frac{1}{P_{e^-} P_{e^+}} \frac{1}{\sqrt{N_D}}$$

when both beams are polarized and

$$(5.84) \quad \delta A^{\text{exp}} = \frac{\delta P_{e^-}}{P_{e^-}} A^{\text{exp}} \oplus \frac{1}{P_{e^-}} \frac{1}{\sqrt{N_D}}$$

when the positron beam is unpolarized. The significance is then defined as  $\mathcal{S} = A^{\text{exp}} / \delta A^{\text{exp}}$ .

We assume  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  [19, 40], and 3% relative uncertainties in the beam polarization reaches, as well as  $L = 10 \text{ ab}^{-1}$  and  $\eta = 80\%$  (100%) for the  $b\bar{b}$  ( $WW$ ) process in accordance with [88]. We assume the input scheme  $\{G_F, m_W, m_Z\}$ . We have

$$(5.85) \quad G_F = 1.1663787 \times 10^{-5} \text{ GeV}^{-2},$$

$$(5.86) \quad m_h = 125 \text{ GeV}, \quad \Gamma_h = 4.1 \times 10^{-3} \text{ GeV},$$

$$(5.87) \quad m_W = 80.379 \text{ GeV}, \quad \Gamma_W = 2.085 \text{ GeV},$$

$$(5.88) \quad m_Z = 91.1876 \text{ GeV}, \quad \Gamma_Z = 2.4952 \text{ GeV},$$

$$(5.89) \quad m_e = 0.511 \times 10^{-3} \text{ GeV}, \quad m_b = 3.105 \text{ GeV},$$

and the derived parameters are

$$(5.90) \quad \alpha = G_F \left( \frac{\pi}{\sqrt{2}m_W^2 \left(1 - \frac{m_W^2}{m_Z^2}\right)} \right)^{-1}, \quad e = \sqrt{4\pi\alpha},$$

$$(5.91) \quad c_W = \frac{m_W}{m_Z}, \quad s_W = \sqrt{1 - c_W^2},$$

$$(5.92) \quad g_L^f = \frac{T_3^f - Q_f s_W^2}{s_W c_W}, \quad g_R^f = -\frac{Q_f s_W}{c_W}.$$

The coupling strengths are given as follows:

$$(5.93) \quad C_{ffh} = -\frac{em_f}{2s_W m_W},$$

$$(5.94) \quad C_{ff\gamma} = -eQ_f, \quad C_{ffZ}^L = eg_L^f, \quad C_{ffZ}^R = eg_R^f, \quad C_{ff'W} = \frac{e}{\sqrt{2}s_W},$$

$$(5.95) \quad C_{hWW} = \frac{em_W}{s_W}, \quad C_{\gamma WW} = -e, \quad C_{ZWW} = \frac{ec_W}{s_W}.$$

The Cabibbo-Kobayashi-Maskawa matrix is taken to be the identity. The width parameter in the BS function is set to the Higgs decay width:

$$(5.96) \quad \delta = \Gamma_h.$$

The collider energy is set to the Higgs mass:

$$(5.97) \quad \mathcal{E}_{\text{coll}} = m_h.$$

Now we assume only an invariant-mass cut  $x\hat{\mathcal{E}}^2 > m_{\text{inv,cut}}^2$ , which we refer to as the level-0 cut, denoted by  $\mathcal{C}^0$ :

$$(5.98) \quad \mathcal{C}^0 : m_{\text{inv,cut}} = 120 \text{ GeV}.$$

In the absence of any other cuts, we obtain the significance values presented in Table 5.1. For the  $WW$  process, we note that there are two different  $WW$  processes, depending on the sign of the electric charge of the outgoing lepton. We have  $e^-e^+ \rightarrow WW \rightarrow \ell^- \bar{\nu}_\ell u_i \bar{d}_i$  and  $e^-e^+ \rightarrow WW \rightarrow \ell^+ \nu_\ell \bar{u}_i d_i$ . We assume that the outgoing lepton could be any of  $e^\mp$ ,  $\mu^\mp$ , and  $\tau^\mp$ , and we restrict ourselves to the first two quark generations, namely  $i = 1, 2$ . We form the observables separately for each of these processes by taking six copies of each to account for three lepton and two quark generations whilst forming the event counts  $N_N$  and  $N_D$ . Subsequently, we use (5.83) or (5.84) with the resultant value of experimentally constructed asymmetry. Then, we compute the significances separately for the two processes, combining them in quadrature at the end.

In Table 5.1, we also present the *reference* values of significance computed using the unpolarized signal and background cross-section values as

$$(5.99) \quad \mathcal{S}_{\text{ref}} = \frac{S}{\sqrt{B}} = \frac{\eta L \sigma(e^-e^+ \rightarrow h \rightarrow X)}{\sqrt{\eta L \sigma_{\text{bkg}}(e^-e^+ \rightarrow X)}},$$

where  $X$  is  $b\bar{b}$  for the  $b\bar{b}$  process and  $WW \rightarrow \ell\nu jj$  for the  $WW$  process.

Table 5.1. The sensitivity estimates for the double-polarization and various single-polarization asymmetries, along with the reference observable, are obtained using the level-0 cut  $\mathcal{C}^0$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization values of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider.

Observable	$e^-e^+ \rightarrow b\bar{b}$	$e^-e^+ \rightarrow WW \rightarrow \ell\nu jj$		
		$\ell^-$	$\ell^+$	combined
DP	0.27	0.31	0.31	0.44
SP <sup>0</sup>	0.19	0.38	0.37	0.53
SP <sup>+</sup>	0.11	1.7	1.6	2.4
SP <sup>-</sup>	0.37	0.054	0.046	0.070
Reference	0.11	0.37	0.37	0.53

From [88], one can also estimate the significance of said processes using the unpolarized cross-section values in the absence of any cuts as follows. The cross-section values for the irreducible background processes  $e^-e^+ \rightarrow b\bar{b}$  and  $e^-e^+ \rightarrow WW \rightarrow \ell\nu jj$  with including BS and ISR effects are obtained to be 15 pb and 15 fb, respectively. We emphasize that the reported values in said work are 19 pb and 23 fb, respectively, without the ISR effects. In the meantime, the corresponding Higgs decay cross-section values are 164 ab and 26.5 ab, respectively. This leads to a significance of

$$(5.100) \quad \mathcal{S} = \frac{S}{\sqrt{B}} = \frac{(80\%)(10 \text{ ab}^{-1})(164 \text{ ab})}{\sqrt{(80\%)(10 \text{ ab}^{-1})(15 \text{ pb})}} = 0.12$$



for the  $b\bar{b}$  process and

$$(5.101) \quad \mathcal{S} = \frac{S}{\sqrt{B}} = \frac{(100\%)(10 \text{ ab}^{-1})(26.5 \text{ ab})}{\sqrt{(100\%)(10 \text{ ab}^{-1})(15 \text{ fb})}} = 0.68$$

for the  $WW$  process. Therefore, we remark that in the absence of any cuts, with the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry as our primary observable for the  $b\bar{b}$  ( $WW$ ) process, we can obtain a significance of 3.1 (3.5) times what can be inferred from [88].

Next, we consider cuts on the phase space on top of the invariant-mass cut, which we refer to as the level-1 cuts. Following [88], we impose

$$(5.102) \quad \mathcal{C}^1 : \mathcal{C}^0 \text{ and } 5^\circ < \theta < 175^\circ$$

for the  $b\bar{b}$  process and

$$(5.103) \quad \mathcal{C}^1 : \mathcal{C}^0 \text{ and } E_{j_1, j_2} < 52, 45 \text{ GeV}, E_\ell > 10 \text{ GeV}, \\ E_{\text{miss}} > 20 \text{ GeV}, m_{12} > 12 \text{ GeV}$$

for the  $WW$  process. With the level-1 cuts  $\mathcal{C}^1$ , our new sensitivity estimates are quoted in Table 5.2. Note that at this point, we refrain from imposing any other cuts that might upset the orthogonality of our weight functions with the interference channels not containing the Higgs field in the asymmetry numerator. We emphasize that the aforementioned cuts do not affect the orthogonality of our precious weight functions as  $\theta$  or  $E_{j_1, j_2}$ ,  $E_\ell$ ,  $E_{\text{miss}}$ , and  $m_{12}$  are all independent of  $\varphi$  or  $\varphi_{12}$ .

Next, we investigate additional cuts on the phase space for the observables of interest without spoiling the orthogonality of our weight functions. We start with the polar angle

Table 5.2. The sensitivity estimates for the double-polarization and various single-polarization asymmetries, as well as for the reference observable, are obtained using the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization reaches of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider.

Observable	$e^-e^+ \rightarrow b\bar{b}$	$e^-e^+ \rightarrow WW \rightarrow \ell\nu jj$		
		$\ell^-$	$\ell^+$	combined
DP	0.27	0.22	0.22	0.31
SP <sup>0</sup>	0.19	0.33	0.33	0.47
SP <sup>+</sup>	0.11	1.4	1.4	2.0
SP <sup>-</sup>	0.38	0.086	0.083	0.12
Reference	0.11	0.32	0.32	0.45

$\theta$  in the  $b\bar{b}$  process. The left panel of Figure 5.3 shows the dependence of the sensitivity of the SP<sup>-</sup> asymmetry on the invariant-mass cut and the polar-angle cut. Here,  $\theta_{\text{cut}}$  indicates the percentage of the interval  $[0^\circ, 180^\circ]$  clipped symmetrically from both end points; to illustrate,  $\theta_{\text{cut}} = 20\%$  means  $\theta$  is integrated from  $18^\circ$  to  $162^\circ$ . In the same figure, we also indicate the point with the maximum sensitivity, 0.61, which occurs when  $\theta_{\text{cut}} = 39\%$  and  $m_{\text{inv, cut}} = m_h - \Gamma_h$ , which is the closest we approach the Higgs resonance. The left panel of Figure 5.4 displays the dependence of the sensitivity of the SP<sup>-</sup> asymmetry on the invariant-mass cut for  $\theta_{\text{cut}} = 39\%$  near the Higgs resonance. Once the invariant-mass cut reaches the resonant energy, the sensitivity drops drastically because of the significantly reduced number of event count.

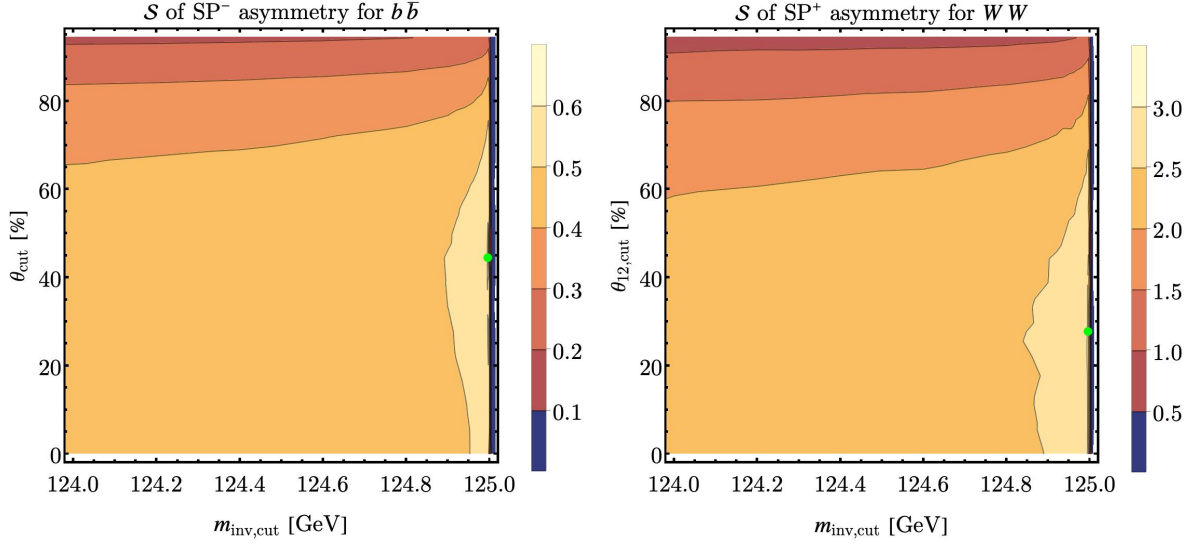


Figure 5.3. The dependence of the sensitivity estimate of the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry on the invariant-mass and polar-angle cuts for the  $b\bar{b}$  ( $WW$ ) process is shown in the left (right) panel. The results are obtained using the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization reaches of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider. The green dot marks the highest sensitivity, achieved at  $m_{\text{inv,cut}} = m_h - \Gamma_h$ .

In light of the  $b\bar{b}$  process, one could argue if a cut on  $\theta_{12}$  would be viable for the  $WW$  process, as well, though the parallel may hardly seem perfect at a first glance because of the  $t$ -channel neutrino diagram. On the right panel of Figure 5.3, we show the dependence of the sensitivity of the  $\text{SP}^+$  asymmetry on the invariant-mass cut and the polar-angle cut. Here,  $\theta_{12,\text{cut}}$  has the same meaning as  $\theta_{\text{cut}}$ . In the same figure, we also indicate the point with the maximum sensitivity, 3.1, which occurs when  $\theta_{12,\text{cut}} = 28\%$  and  $m_{\text{inv,cut}} = m_h - \Gamma_h$ . In the right panel of Figure 5.4, we display the dependence of

the significance of the  $\text{SP}^+$  asymmetry on the invariant-mass cut for  $\theta_{12,\text{cut}} = 28\%$  near the Higgs resonance. The observed behavior is identical to the case of the  $b\bar{b}$  process.

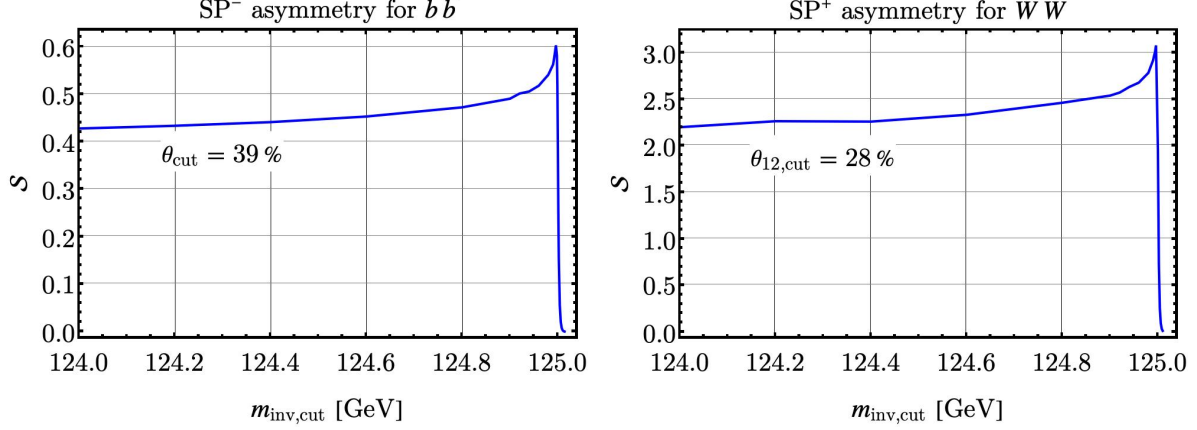


Figure 5.4. The left (right) panel shows how the sensitivity of the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry varies with the invariant-mass cut when the polar-angle cut is fixed at its optimal value,  $\theta_{\text{cut}} = 39\%$  ( $\theta_{12,\text{cut}} = 28\%$ ), for the  $b\bar{b}$  ( $WW$ ) process. The results are based on the level-1 cuts  $\mathcal{C}^1$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization levels of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider.

In Figure 5.5, we display the contribution of each channel to the numerator and denominator of all the asymmetries under consideration for the  $WW$  process, summed over the three generations of  $\ell^\pm$  and two generations of quark jets, to have a better understanding of the wild differences in the sensitivity estimates amongst observables. The black bars indicate the DP asymmetry, red  $\text{SP}^0$ , blue  $\text{SP}^+$ , and green  $\text{SP}^-$ , respectively, with the level-0 cuts  $\mathcal{C}^0$ , whereas the lighter shades in the foreground are the corresponding values after the phase-space, or level-1, cuts  $\mathcal{C}^1$ . The channels not shown contribute by an

amount 0.005 or less in shown units. Note that the numbers for the *reference* observable are just the  $D$  values of the DP asymmetry.

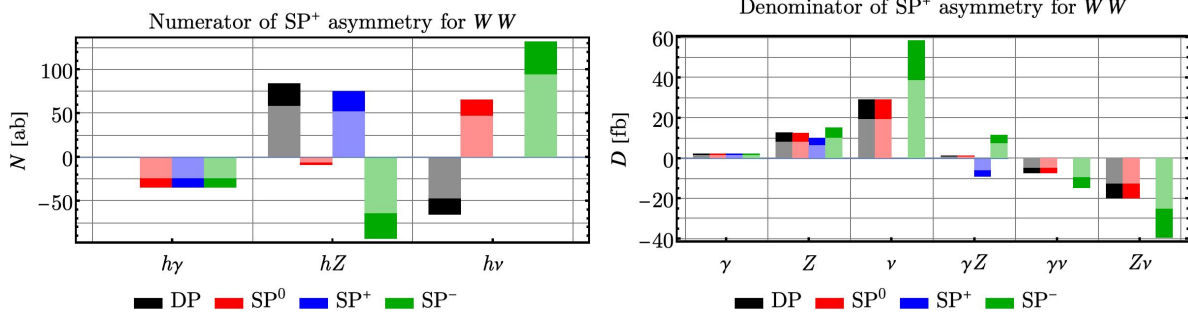


Figure 5.5. The left (right) panel shows the contributions of interference channels to the numerator (denominator) of the SP<sup>+</sup> asymmetry in the  $WW$  process. Results are shown for the level-0 and level-1 cuts,  $\mathcal{C}^0$  and  $\mathcal{C}^1$ , represented by dark and light shading, respectively. Channels contributing less than 0.005 in the displayed units are omitted.

From Figure 5.5, we observe that due to the absence of aggressive competition between the channels with and without the neutrino exchange, the SP<sup>+</sup> asymmetry yields a remarkably high significance value. In this figure, we also see that the negligible contribution of the neutrino channels in the numerator renders the polar-angle cut viable for the  $WW$  process.

In Table 5.3, we present our sensitivity estimates for the  $b\bar{b}$  and  $WW$  processes with the level-1 cuts  $\mathcal{C}^1$  plus the *best* polar-angle cuts at  $m_{\text{inv,cut}} = m_h - (10 \text{ MeV})$ , which we refer to as the level-2 cuts, denoted by  $\mathcal{C}^2$ :

$$(5.104) \quad \mathcal{C}^2 : \mathcal{C}^1 \text{ and } m_{\text{inv,cut}} = m_h - (10 \text{ MeV}) \text{ and } \textit{best} \text{ polar-angle cut.}$$

Table 5.3. Sensitivity estimates for the double-polarization and various single-polarization asymmetries, along with the reference observable, are shown with the corresponding optimal polar-angle cuts. The results are based on the level-2 cuts  $\mathcal{C}^2$ , an integrated luminosity of  $L = 10 \text{ ab}^{-1}$ , and beam polarization values of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider.

Observable	$e^-e^+ \rightarrow b\bar{b}$	$\theta_{\text{cut}} [\%]$	$e^-e^+ \rightarrow WW \rightarrow \ell\nu jj$			
			$\ell^-$	$\ell^+$	combined	$\theta_{12,\text{cut}} [\%]$
DP	0.41	39	0.31	0.31	0.44	22
SP <sup>0</sup>	0.30	33	0.58	0.55	0.80	44
SP <sup>+</sup>	0.17	44	2.1	2.0	2.9	28
SP <sup>-</sup>	0.58	39	0.24	0.22	0.33	67
Reference	0.16	6	0.45	0.45	0.64	6

Our results are summarized on the top and bottom panels of Figure 5.6 for the  $b\bar{b}$  and  $WW$  process, respectively. The black bars indicate the DP asymmetry, red SP<sup>0</sup>, blue SP<sup>+</sup>, green SP<sup>-</sup>, and brown the *reference* observable. The darkest shades in the background are the values obtained with phase-space cuts and the *best* polar-angle cuts, the lighter shades in the foreground no cuts, and the lightest shades in between the phase-space cuts only. The solid yellow lines represent the significance values quoted in [88] for the relevant processes with the BS and ISR effects on top of multivariate analysis.

In Figure 5.7, we picture the dependence of the sensitivity estimates of the asymmetry observables SP<sup>-</sup> and SP<sup>+</sup> for the  $b\bar{b}$  and  $WW$  processes on the left and right panels,

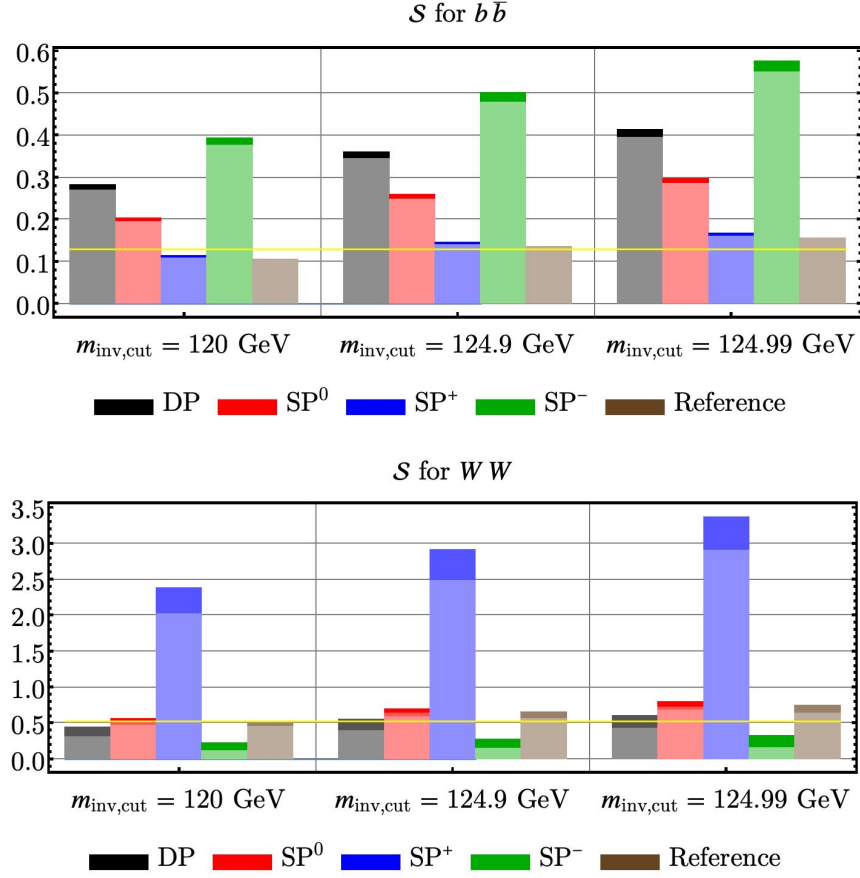


Figure 5.6. Sensitivity estimates for the double-polarization and various single-polarization asymmetries, as well as the reference observable, are shown using the specified invariant-mass cuts at all three cut levels. The results assume an integrated luminosity of  $L = 10 \text{ ab}^{-1}$  and beam polarization reaches of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$  at the collider. The phase-space cuts are  $5^\circ < \theta < 175^\circ$  for the  $b\bar{b}$  channel, and  $E_{j_1,j_2} < 52, 45 \text{ GeV}$ ,  $E_\ell > 10 \text{ GeV}$ ,  $E_{\text{miss}} > 20 \text{ GeV}$ , and  $m_{12} > 12 \text{ GeV}$  for the  $WW$  channels. The corresponding optimal polar-angle cuts are listed in Table 5.3.

respectively, on the c.m. energy spread,  $\delta$ , and the integrated luminosity,  $L$ , with the level-2 cuts  $\mathcal{C}^2$ .

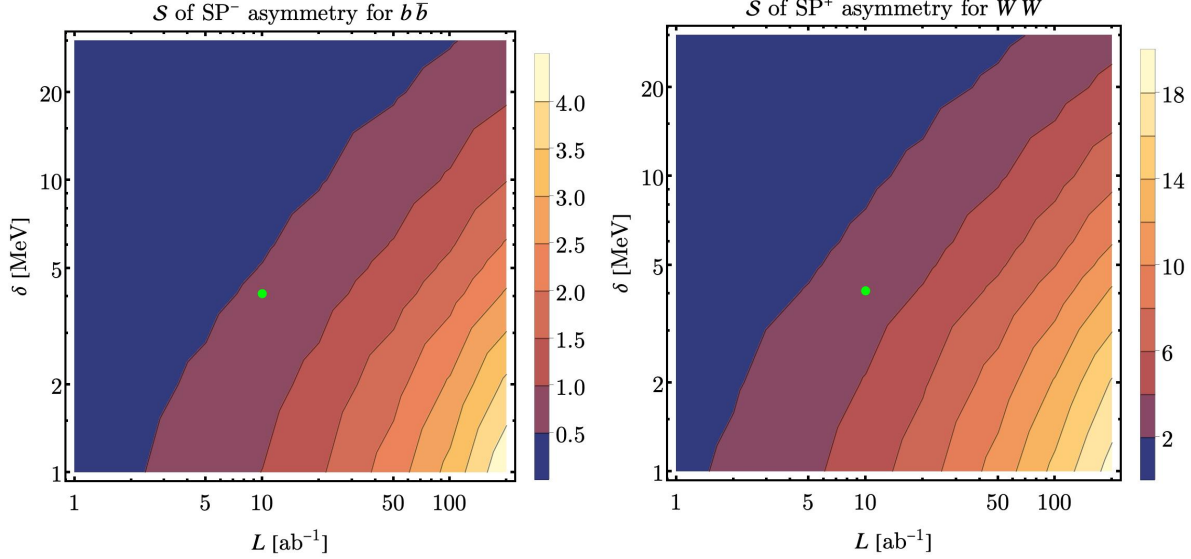


Figure 5.7. The left (right) panel shows how the sensitivity of the  $\text{SP}^-$  ( $\text{SP}^+$ ) asymmetry depends on the center-of-mass energy spread and the integrated luminosity for the  $b\bar{b}$  ( $WW$ ) process. Results are obtained using the level-2 cuts  $\mathcal{C}^2$  and beam polarization values of  $P_{e^-} = 80\%$  and  $P_{e^+} = 30\%$ . The green dot marks the baseline choice of  $\delta = \Gamma_h$  and  $L = 10 \text{ ab}^{-1}$ .

## 5.6. Coda

In this chapter, we explored how transverse-spin asymmetries measured at a future FCC-ee can significantly improve sensitivity to the electron Yukawa coupling compared to traditional, inclusive methods. The electron Yukawa coupling, being the smallest in the SM, comes with unique challenges due to its extremely small predicted value. Directly probing such a tiny quantity demands creative, high-precision experimental strategies.



The approach investigated here uses the quantum interference between the Higgs-mediated amplitude and continuum background processes, allowing sensitivity to scale linearly with the electron Yukawa coupling, rather than quadratically, thus greatly improving measurement potential.

We have shown that transverse-spin asymmetries use this quantum interference to isolate the Higgs-related effects from the substantial backgrounds present in the  $b\bar{b}$  and semi-leptonic  $WW$  final states. For the  $WW$  channel, we found that sensitivity can be improved by nearly a factor of six compared to a conventional inclusive cross-section measurement. Even in the more experimentally challenging  $b\bar{b}$  final state, the gain reaches a factor of five, demonstrating the generality and robustness of our method.

A particularly important result of our analysis is the significant boost in sensitivity achieved using single transverse-spin asymmetries, where only the electron beam is transversely polarized. By carefully exploring optimal phase-space cuts and angular weighting, we showed that SP asymmetries consistently deliver the highest sensitivity improvements. Although double-polarized asymmetries can further enhance sensitivity, it is the SP asymmetry that primarily drives our most substantial gains. Thus, our results strongly suggest experimental efforts at FCC-ee should prioritize single transverse-spin polarization, alongside the precision in angular measurements and phase-space cuts thoroughly explored.

However, achieving such precision at an FCC-ee would involve demanding experimental requirements. High beam polarization, ultra-precise beam energy calibration, and strict control of energy spread and initial-state radiation effects are essential. We carefully included these factors in our sensitivity estimates, demonstrating that even moderate reductions in luminosity or slight increases in beam energy spread still allow substantial

sensitivity gains compared to conventional approaches. This robustness highlights the practical feasibility of our method.

Long story short, transverse-spin asymmetries not only enhance sensitivity to the electron Yukawa coupling but also introduce a new dimension to precision Higgs physics, enabling detailed studies otherwise unreachable. Beyond the numerical improvement, our results clearly show the powerful synergy between collider polarization capabilities and subtle quantum interference effects. If experimentally realized, this measurement strategy could directly confirm the SM prediction for the electron Yukawa coupling, providing crucial insights into one of the theory’s most elusive parameters.

The techniques developed here are widely applicable, extending beyond the specific channels studied. Future experimental analyses at FCC-ee could adapt this method to other diboson final states, such as fully leptonic or fully hadronic decays, thereby further consolidating the electron Yukawa coupling measurement. Ultimately, this approach positions transverse-spin asymmetries as essential tools in future collider physics, setting the stage for unprecedented precision in understanding the Higgs boson and its interactions.

## CHAPTER 6

***CP*-odd to Joy**

*There is this concept called “moment”. You don’t run into it on the bus or at the movies. No one fights anyone over moment. We had this guy in class, who insisted he didn’t believe in moment. I hate moment. It ruins my day.*

Oğuz Atay, *The Disconnected*

We study the sensitivity of the high-luminosity Large Hadron Collider (HL-LHC) to previously unexplored *CP*-odd dimension-8 operators in the Standard Model Effective Field Theory. Focusing on neutral-current Drell–Yan production in association with a real jet, we consider semi-leptonic four-fermion operators involving gluon field strength tensors. These operators do not interfere with the Standard Model in inclusive observables at leading order, but contribute to specific angular structures. We exploit the Collins–Soper moments  $A_6$  and  $A_7$ , which are *CP*-odd and sensitive to such effects, and perform binned analyses in dilepton invariant mass and transverse momentum. Using projected HL-LHC luminosity, we extract constraints on the relevant Wilson coefficients through single and multi-parameter fits. We find that effective scales up to 9 TeV can be probed in single-parameter fits, while simultaneous fits to all relevant operators yield weaker but nontrivial bounds at the TeV scale.

### 6.1. Prelude

A key goal of precision collider phenomenology is to constrain possible effects of new physics through subtle deviations from the Standard Model (SM) predictions. The Standard Model Effective Field Theory (SMEFT) provides a systematic framework for this purpose, parametrizing potential new physics through higher-dimensional operators suppressed by the scale of new physics. While much of the existing SMEFT literature has focused on dimension-6 operators, several classes of dimension-8 operators remain unexplored and may carry unique information about possible ultraviolet (UV) completions of the SM.

One such class involves semi-leptonic four-fermion operators with gluonic field-strength insertions. These operators first appear at dimension-8 and do not interfere with the SM amplitudes in inclusive observables but can contribute significantly to differential distributions. Previous studies have shown that neutral-current Drell-Yan (DY) production, particularly in the high transverse momentum regime, is sensitive to these gluonic dimension-8 effects. Specifically, [45] demonstrated that the doubly-differential DY distribution in dilepton invariant mass and transverse momentum can probe  $CP$ -even semi-leptonic dimension-8 operators involving a gluon field strength tensor. Their analysis revealed that the transverse momentum distribution carries a distinct sensitivity to these operators, offering a valuable diagnostic tool to distinguish between different UV scenarios.

However, said study was limited to  $CP$ -even operators. The  $CP$ -odd counterparts remain uncharted. These  $CP$ -odd dimension-8 operators arise naturally in many UV scenarios, including those with nontrivial  $CP$ -violating dynamics, and can potentially

leave measurable imprints in DY production in association with a real jet (DY $j$ ). Their contributions manifest in specific angular and momentum distributions, providing complementary handles to their  $CP$ -even partners.

The motivation for the present work is to extend the existing analysis to include  $CP$ -odd dimension-8 semi-leptonic operators in the DY $j$  process. This channel is particularly suited for such studies, as the real jet emission allows direct access to the gluon field strength tensor in the hard scattering process. Furthermore, the analytical control over the tree level  $2 \rightarrow 3$  matrix elements enables us to systematically include the effects of these operators and assess their impact on kinematic distributions relevant for current and future measurements at the Large Hadron Collider (LHC).

By constructing the complete leading-order amplitudes for both SM and SMEFT contributions, and performing a detailed numerical analysis, this work aims to quantify the sensitivity of the high-luminosity Large Hadron Collider (HL-LHC) to  $CP$ -odd dimension-8 effects in the DY $j$  production. This effort not only fills a gap in the SMEFT phenomenology landscape but also enhances our ability to diagnose the nature of possible new physics in the event of future deviations from SM predictions.

This chapter is organized as follows. In Section 6.2, we pave the path to the cross section, detailing the process, theoretical calculations, and kinematics. We review the Collins-Soper (CS) frame in Section 6.3. In Section 6.4, we revisit the SMEFT formalism with the operators relevant to our study. The CS moments for the DY production cross section are defined in Section 6.5. Section 6.6 details our numerical calculations, setting up the scene for the HL-LHC simulations in Section 6.7. We present the fit results of the SMEFT parameters in Section 6.8. In Section 6.9, we conclude.

## 6.2. Structure of the cross section

We are interested in the hadronic process  $pp \rightarrow jV \rightarrow je^-e^+$  at the LHC at TeV energies with both incoming beams unpolarized. In what follows, we use  $e$  and  $\ell$  interchangeably to denote the outgoing lepton. The underlying partonic processes,  $p_r + p_{r'} \rightarrow j + e^- + e^+$  for partons  $p_r$  and  $p_{r'}$ , consist of pair annihilation,

$$(6.1) \quad q_i(p_a) + \bar{q}_i(p_b) \rightarrow g(p_1) + e^-(p_2) + e^+(p_3),$$

and Compton scattering,

$$(6.2) \quad q_i(p_a) + g(p_b) \rightarrow q_i(p_1) + e^-(p_2) + e^+(p_3),$$

$$(6.3) \quad \bar{q}_i(p_a) + g(p_b) \rightarrow \bar{q}_i(p_1) + e^-(p_2) + e^+(p_3).$$

The tree-level Feynman diagrams are illustrated in Fig. 6.1. Here,  $V$  is a neutral electroweak (EW) gauge boson.

We obtain the SM amplitudes using FeynArts [111] and FeynCalc [146, 145, 144, 138] first and add the SMEFT amplitudes later by hand. We denote the propagator denominators by

$$(6.4) \quad D_\gamma = \frac{1}{m_{\ell\ell}^2}, \quad D_Z = \frac{1}{m_{\ell\ell}^2 - m_Z^2}, \quad D_s = \frac{1}{\hat{s}} = \frac{1}{m_{jV}^2}, \quad D_t = \frac{1}{\hat{t}}, \quad D_u = \frac{1}{\hat{u}},$$

where  $\hat{s} = (p_a + p_b)^2 = (p_1 + p_{23})^2 = m_{jV}^2$ ,  $\hat{t} = (p_a - p_1)^2$ , and  $\hat{u} = (p_a - p_{23})^2$  such that  $\hat{s} + \hat{t} + \hat{u} = m_{jV}^2 + \hat{t} + \hat{u} = m_{\ell\ell}^2$ . Here,  $m_{jV}$  is the invariant mass of the  $jV$  system and  $m_{\ell\ell}$  is that of the  $\ell^-\ell^+$  system, and we eliminate one Mandelstam variable

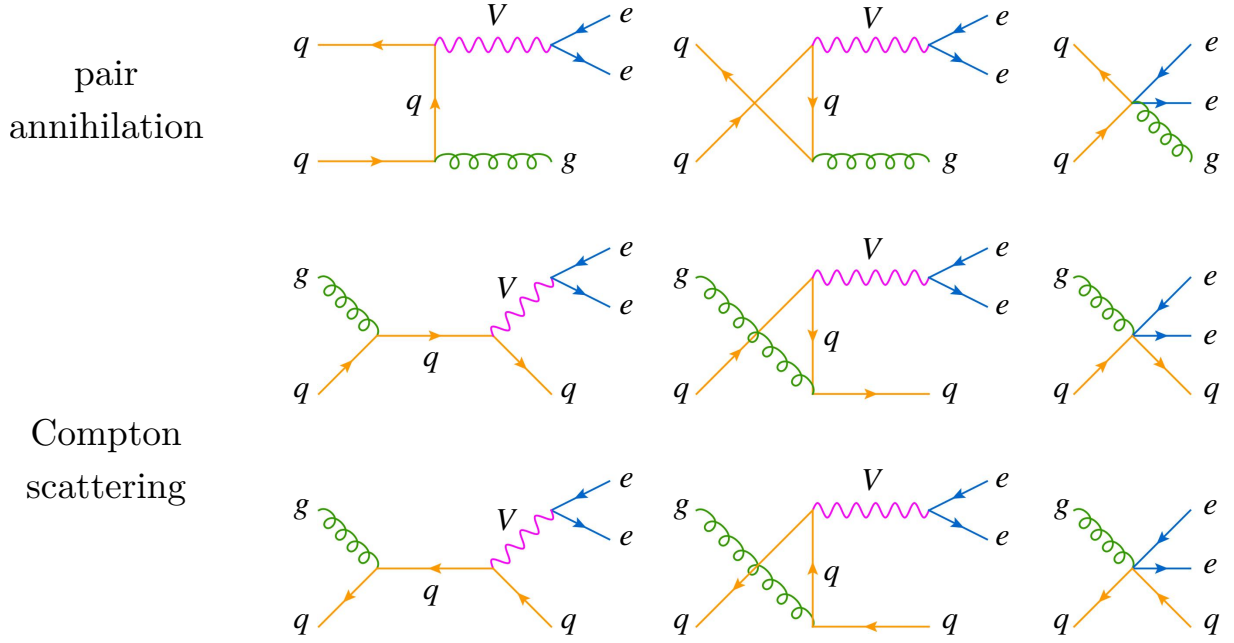


Figure 6.1. The tree-level Feynman diagrams describing the underlying partonic processes.

via  $\hat{u} = m_{\ell\ell}^2 - m_{jV}^2 - \hat{t}$ . We have the following elementary dots:

$$(6.5) \quad p_i^2 = 0, \quad i = a, b, 1, 2, 3; \quad p_{23}^2 = m_{\ell\ell}^2,$$

$$(6.6) \quad p_a \cdot p_b = \frac{\hat{s}}{2} = \frac{m_{jV}^2}{2}, \quad p_a \cdot p_1 = -\frac{\hat{t}}{2}, \quad p_a \cdot p_{23} = -\frac{\hat{u} - m_{\ell\ell}^2}{2},$$

$$(6.7) \quad p_b \cdot p_1 = -\frac{\hat{u}}{2}, \quad p_b \cdot p_{23} = -\frac{\hat{t} - m_{\ell\ell}^2}{2},$$

$$(6.8) \quad p_1 \cdot p_{23} = \frac{\hat{s} - m_{\ell\ell}^2}{2} = \frac{m_{jV}^2 - m_{\ell\ell}^2}{2},$$

$$(6.9) \quad p_2 \cdot p_3 = \frac{m_{\ell\ell}^2}{2}.$$

Since the center-of-mass (c.m.) energy is at a TeV level, we assume massless fermions.

The gluon polarization sum is

$$(6.10) \quad \sum_s \varepsilon_s^{*\mu}(k) \varepsilon_s^\nu(k) = -g^{\mu\nu}.$$

The color simplification is

$$(6.11) \quad T_{bc}^a T_{cb}^a = \text{tr}(T^a T^a) = C_A C_F = N \times \frac{N^2 - 1}{2N} = 4.$$

In the squared amplitude calculation, we have some extra factors, depending on whether the incoming state is a quark or a gluon. The extra factor  $1/(2N)$  is for an incoming quark, where  $1/2$  is for spin-averaging and  $1/N$  is for color-averaging, and the extra factor  $1/(2(N^2 - 1))$  is for the color-averaging of an incoming gluon, where  $1/2$  is for polarization-averaging and  $1/(N^2 - 1)$  is for color-averaging.

The hadronic cross section is given by

$$(6.12) \quad \sigma = \sum_{i=-N_f}^{N_f} \sum_{j=-N_f}^{N_f} \int dx_a dx_b f_i(x_a) f_j(x_b) \sigma^{ij},$$

where the quark flavor numbers are  $(d, u, s, c, b) = (1, 2, 3, 4, 5)$ , the antiquarks get a minus sign, the gluon flavor is simply 0, and

$$(6.13) \quad \sigma^{ij} = F \int \text{dLIPS} |\mathcal{A}^{ij}|^2$$

describes the partonic cross section. Here,

$$(6.14) \quad F = \frac{1}{4\sqrt{(p_a \cdot p_b)^2 - (m_a m_b)^2}} = \frac{1}{2\hat{s}} = \frac{1}{2m_{jV}^2}$$



is the flux factor,  $|\mathcal{A}^{ij}|^2$  is the unpolarized squared amplitude for partonic process  $p_i + p_j$ , and dLIPS is the 3-particle Lorentz-invariant phase space (LIPS):

$$(6.15) \quad \int d\text{LIPS}_3 = \left[ \frac{1}{(2\pi)^{3n-4}} \right]_{n=3} \int dm_{23}^2 \times \left[ \frac{\nabla(\mathcal{E}^2, m_{23}^2, m_1^2)}{8\mathcal{E}^2} \int d\Omega^\bullet \right] \left[ \frac{\nabla(m_{23}^2, m_2^2, m_3^2)}{8m_{23}^2} \int d\Omega^\circ \right],$$

where  $\nabla(x, y, z) = \sqrt{x^2 + y^2 + z^2 - 2(xy + yz + zx)}$  is the square root of the Källén function,  $\mathcal{E} = \sqrt{\hat{s}}$  is the available energy, the angles  $\Omega^\bullet = (\theta^\bullet, \varphi^\bullet)$  are the spherical angles of the particle “23” in the rest frame of the total incoming momentum  $P$  and the angles  $\Omega^\circ = (\theta^\circ, \varphi^\circ)$  are those of the particle 2 in the rest frame of the particle “23”, partitioning the partonic process as  $p_r(p_a) + p_{r'}(p_b) \rightarrow j(p_1) + V(p_{23})$  followed by  $V(p_{23}) \rightarrow e^-(p_2) + e^+(p_3)$ . The frames are discussed in greater detail in the next section. Here, the integration limits of  $m_{23}$  are given by the requirements  $\lambda(\mathcal{E}^2, m_{23}^2, m_1^2) > 0$  and  $\lambda(m_{23}^2, m_2^2, m_3^2) > 0$ . Applying to our case, i.e. with  $\mathcal{E} = \sqrt{\hat{s}} = m_{jV}$ ,  $m_1 = m_2 = m_3 = 0$  and  $m_{23} = m_{\ell\ell}$ , we obtain

$$(6.16) \quad \int d\text{LIPS}_3 = \frac{1}{2048\pi^5} \int_0^{m_{jV}^2} dm_{\ell\ell}^2 \int d\Omega^\bullet \int d\Omega^\circ \left( 1 - \frac{m_{\ell\ell}^2}{m_{jV}^2} \right).$$

We note that our expression agrees with the general formula [128]:

$$(6.17) \quad \text{LIPS}_n = \frac{1}{2(4\pi)^{2n-3}} \frac{\mathcal{E}^{2n-4}}{\Gamma(n)\Gamma(n-1)}.$$

For the partonic subprocess  $p_r(p_a) + p_{r'}(p_b) \rightarrow j(p_1) + V(p_{23})$  in the partonic c.m. frame, the momentum components,  $p^\bullet = (E^\bullet, \mathbf{p}^\bullet)$ , are explicitly given by

$$(6.18) \quad E_{a/b}^\bullet = \frac{\mathcal{E}^2 + m_{a/b}^2 - m_{b/a}^2}{2\mathcal{E}}, \quad |\mathbf{p}_{a/b}^\bullet| = \frac{\nabla(\mathcal{E}^2, m_a^2, m_b^2)}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{a/b}^\bullet = \pm(0, 0, 1),$$

$$(6.19) \quad E_{23/1}^\bullet = \frac{\mathcal{E}^2 + m_{23/1}^2 - m_{1/23}^2}{2\mathcal{E}},$$

$$(6.20) \quad |\mathbf{p}_{23/1}^\bullet| = \frac{\nabla(\mathcal{E}^2, m_{23}^2, m_1^2)}{2\mathcal{E}}, \quad \hat{\mathbf{p}}_{23/1}^\bullet = \pm(s_{\theta^\bullet} c_{\varphi^\bullet}, s_{\theta^\bullet} s_{\varphi^\bullet}, c_{\theta^\bullet}),$$

and for the subsequent leptonic decay  $V(p_{23}) \rightarrow \ell^-(p_2) + \ell^+(p_3)$  in the rest frame of the EW gauge boson, the momentum components,  $p^\circ = (E^\circ, \mathbf{p}^\circ)$ , are explicitly given by

$$(6.21) \quad E_{23}^\circ = m_{23}, \quad \mathbf{p}_{23}^\circ = \mathbf{0},$$

$$(6.22) \quad E_{2/3}^\circ = \frac{m_{23}^2 + m_{2/3}^2 - m_{3/2}^2}{2m_{23}},$$

$$(6.23) \quad |\mathbf{p}_{2/3}^\circ| = \frac{\nabla(m_{23}^2, m_2^2, m_3^2)}{2m_{23}}, \quad \hat{\mathbf{p}}_{2/3}^\circ = \pm(s_{\theta^\circ} c_{\varphi^\circ}, s_{\theta^\circ} s_{\varphi^\circ}, c_{\theta^\circ}).$$

with  $\mathcal{E} = \sqrt{\hat{s}} = m_{jV}$ ,  $m_a = m_b = m_1 = m_2 = m_3 = 0$ , and  $m_{23} = m_{\ell\ell}$ . The two frames are connected by a Lorentz transformation,  $\Lambda^{\circ\bullet}$ , deriving from  $p_{23}^\circ = \Lambda^{\circ\bullet} p_{23}^\bullet$ , with  $\Lambda^{\circ\bullet} = BR_y R_z$ , where  $R_z$  is an azimuthal rotation by  $\varphi^\bullet$ ,  $R_y$  is a polar rotation by  $\theta^\bullet$ , and  $B$  is a boost along the  $z$  axis. By expressing all the momenta in either one of the frames, we also confirm the *symmetry requirements* of the LIPS. With  $s_{ij} = (p_i + p_j)^2$  for  $i, j = 1, 2, 3$ , we must have  $s_{ij}$  integrated over the LIPS giving the same result for all  $i \neq j$ ; *simile*, we must have  $p_a \cdot p_i$  and  $p_b \cdot p_i$  for  $i = 1, 2, 3$  integrated over the LIPS giving

the same result. We find

$$(6.24) \quad \int d\text{LIPS}_3 [s_{12}, s_{13}, s_{23}] = \frac{m_{jV}^4}{768\pi^3},$$

and

$$(6.25) \quad \int d\text{LIPS}_3 [p_a \cdot p_1, p_a \cdot p_2, p_a \cdot p_3, p_b \cdot p_1, p_b \cdot p_2, p_b \cdot p_3] = \frac{m_{jV}^4}{1536\pi^3}.$$

### 6.3. Collins-Soper or: how I learned to stop worrying and love the Frame

With all the momenta expressed in either of the frames presented at the end of the previous section, it is just a cute exercise of bringing all the momenta into a single frame and computing the hadronic cross section; however, we are not going to work with all the angular variables presented in the previous section. To be more precise, the spherical angles  $\Omega^\circ$  are essentially replaced by the CS angles [73, 14],  $\Omega^*$ , but for the leading subprocess, we want to work with the invariant mass of the  $jV$  pair,  $m_{jV}$ , and the rapidity,  $y$ , and transverse momentum,  $p_\top$ , of the  $V = \ell^- \ell^+$  system as measured in the lab frame, i.e. the hadronic c.m. frame.

In the lab frame, we have

$$(6.26) \quad P_{a/b}^\diamond = \frac{\sqrt{s}}{2}(1, 0, 0, \pm 1), \quad p_{a/b}^\diamond = x_{a/b} P_{a/b}^\diamond, \quad p_{23}^\diamond = (E_V, p_\top, 0, p_z).$$

In accordance with [14], we assume that  $V$  is emitted with a positive  $x$  component of the momentum and it is equal to  $p_\top$ . We also assume that the  $jV$  pair is emitted in the  $xz$  plane of lab frame, which is still true in the partonic c.m. frame because the two frames are related by a Lorentz boost along the beam axis.

In the CS frame, we have

$$(6.27) \quad p_{2/3}^* = \frac{m_{\ell\ell}}{2}(1, \pm s_{\theta^*} c_{\varphi^*}, \pm s_{\theta^*} s_{\varphi^*}, \pm c_{\theta^*}),$$

$$(6.28) \quad p_{23}^* = m_{\ell\ell}(1, \mathbf{0}).$$

The transformation from the lab frame to the CS frame is given by a longitudinal boost along the beam axis to make  $p_z = 0$ , followed by a transverse one to make  $p_\tau = 0$  [14]:

$$(6.29) \quad \Lambda^{\star\diamond} = \Lambda_\perp \Lambda_\parallel,$$

where

$$(6.30) \quad \Lambda_\parallel = \begin{pmatrix} \gamma_\parallel & 0 & 0 & -\gamma_\parallel \beta_\parallel \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_\parallel \beta_\parallel & 0 & 0 & \gamma_\parallel \end{pmatrix}, \quad \Lambda_\perp = \begin{pmatrix} \gamma_\perp & -\gamma_\perp \beta_\perp & 0 & 0 \\ -\gamma_\perp \beta_\perp & \gamma_\perp & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

with

$$(6.31) \quad \beta_\parallel = \frac{p_z}{E_V}, \quad \gamma_\parallel = \frac{1}{\sqrt{1 - \beta_\parallel^2}}, \quad \beta_\perp = \frac{p_\tau}{\sqrt{m_{\ell\ell}^2 + p_\tau^2}}, \quad \gamma_\perp = \frac{1}{\sqrt{1 - \beta_\perp^2}}.$$

Using this, we can write, e.g.,  $P_{a/b}^* = \Lambda^{\star\diamond} P_{a/b}^\diamond$ . Note that if we compute  $p_{2/3}^\diamond = \Lambda^{\star\diamond} p_{2/3}^* = \Lambda^{\star\diamond^{-1}} p_{2/3}^*$ , we confirm the well-known relation between the CS polar angle, lepton/antilepton energies and longitudinal momentum components measured in the lab frame, and the

invariant mass and transverse momentum of the  $V = \ell^- \ell^+$  system, which reads

$$(6.32) \quad c_{\theta^*} = \frac{2[(p_2^\diamond)^3(p_3^\diamond)^0 - (p_2^\diamond)^0(p_3^\diamond)^3]}{m_{\ell\ell}\sqrt{m_{\ell\ell}^2 + p_\top^2}}.$$

The rapidity of the  $V = \ell^- \ell^+$  system is defined as

$$(6.33) \quad y = \frac{1}{2} \ln \left( \frac{E_V + p_z}{E_V - p_z} \right).$$

For the  $jV$  pair, we have the usual expression,

$$(6.34) \quad \begin{aligned} y_{jV} &= \frac{1}{2} \ln \left( \frac{(p_1^\diamond + p_{23}^\diamond)^0 + (p_1^\diamond + p_{23}^\diamond)^3}{(p_1^\diamond + p_{23}^\diamond)^0 - (p_1^\diamond + p_{23}^\diamond)^3} \right) \\ &= \frac{1}{2} \ln \left( \frac{(p_a^\diamond + p_b^\diamond)^0 + (p_a^\diamond + p_b^\diamond)^3}{(p_a^\diamond + p_b^\diamond)^0 - (p_a^\diamond + p_b^\diamond)^3} \right) \\ &= \frac{1}{2} \ln \left( \frac{x_a}{x_b} \right), \end{aligned}$$

or

$$(6.35) \quad \frac{x_a}{x_b} = e^{2y_{jV}}.$$

With  $p_{a/b}^\diamond = x_{a/b} P_{a/b}^\diamond$ , we also have

$$(6.36) \quad x_a x_b s = m_{jV}^2.$$

Combining the two, we obtain

$$(6.37) \quad x_{a/b} = \frac{m_{jV}}{\sqrt{s}} e^{\pm y_{jV}},$$

which is not Lorentz-invariant but as long as we define  $x_{a/b}$ ,  $P_{a/b}$ ,  $m_{jV}$ , and  $y_{jV}$  consistently, we can write  $p_{a/b} = x_{a/b}P_{a/b}$  in any frame.

So far, the differential variables of the hadronic cross section are  $x_a$ ,  $x_b$ ,  $m_{\ell\ell}$ ,  $\theta^\bullet$ ,  $\varphi^\bullet$ ,  $\theta^\star$ , and  $\varphi^\star$ . Here,  $\varphi^\bullet$  is immaterial and dropped (but still contributes a factor of  $2\pi$  in the phase space integration), and  $\theta^\star$  and  $\varphi^\star$  are the CS angles. Now, we want to perform a change of variables from  $(x_a, x_b, \theta^\bullet)$  to  $(m_{jV}, y, p_\top)$ .

Note the following. In the parton c.m. frame, we have

$$(6.38) \quad p_{a/b}^\bullet = \frac{m_{jV}}{2}(1, 0, 0, \pm 1),$$

$$(6.39) \quad p_{23/1}^\bullet = (E_{23}^\bullet, \pm |\mathbf{p}_{23}^\bullet| s_{\theta^\bullet}, 0, \pm |\mathbf{p}_{23}^\bullet| c_{\theta^\bullet}).$$

Here, the components of the 4-momenta of the outgoing particles are found from earlier to be

$$(6.40) \quad E_{23/1}^\bullet = \frac{m_{jV}^2 \pm m_{\ell\ell}^2}{2m_{jV}},$$

$$(6.41) \quad |\mathbf{p}_{23}^\bullet| = \frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}.$$

The transformation between the lab frame and the partonic c.m. frame is given by

$$(6.42) \quad \Lambda^{\diamond\bullet} = \begin{pmatrix} \gamma_{ab} & 0 & 0 & -\gamma_{ab}\beta_{ab} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\gamma_{ab}\beta_{ab} & 0 & 0 & \gamma_{ab} \end{pmatrix},$$

with

$$(6.43) \quad \beta_{ab} = -\frac{x_a - x_b}{x_a + x_b}, \quad \gamma_{ab} = \frac{1}{\sqrt{1 - \beta_{ab}^2}}.$$

Since the lab frame and the partonic c.m. frame are related by a  $z$ -boost, we must have

$$(6.44) \quad |\mathbf{p}_{23}^\bullet|_{s_{\theta^\bullet}} = p_\top.$$

This tells us that

$$(6.45) \quad 0 < p_\top < |\mathbf{p}_{23}^\bullet|,$$

or

$$(6.46) \quad 0 < p_\top < \frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}.$$

We also obtain

$$(6.47) \quad c_{\theta^\bullet} = \pm \sqrt{1 - \frac{4m_{jV}^2 p_\top^2}{(m_{jV}^2 - m_{\ell\ell}^2)^2}},$$

where the  $+$  sign is for  $0 < \theta^\bullet < \frac{\pi}{2}$  and the  $-$  sign is when  $\frac{\pi}{2} < \theta^\bullet < \pi$ . Now consider the  $V$  momentum in the partonic c.m. frame and boost it into the lab frame via  $p_{23}^\diamond = \Lambda^{\diamond\bullet} p_{23}^\bullet$  and compute  $y$  with the components obtained after the boost. We obtain

$$(6.48) \quad y = \frac{1}{2} \ln \left( \frac{F^+ x_a}{F^- x_b} \right),$$

where

$$(6.49) \quad F^\pm = m_{jV}^2 + m_{\ell\ell}^2 \pm c_{\theta^\bullet}(m_{jV}^2 - m_{\ell\ell}^2).$$

Together with  $x_a x_b s = m_{jV}^2$ , we arrive at

$$(6.50) \quad x_{a/b} = \sqrt{\frac{F^\mp}{F^\pm}} e^{\pm y} \frac{m_{jV}}{\sqrt{s}}.$$

Noting that  $0 < x_{a/b} < 1$ , we obtain

$$(6.51) \quad -\frac{1}{2} \ln \left( \frac{F^-}{F^+} \frac{s}{m_{jV}^2} \right) < y < \frac{1}{2} \ln \left( \frac{F^+}{F^-} \frac{s}{m_{jV}^2} \right).$$

We can compute the Jacobian of the transformation easily to find

$$(6.52) \quad J = \left| \det \left( \frac{\partial(x_a, x_b, c_{\theta^\bullet})}{\partial(m_{jV}, y, p_\top)} \right) \right| = \frac{8m_{jV}^3 p_\top}{s(m_{jV}^2 - m_{\ell\ell}^2) \sqrt{(m_{jV}^2 - m_{\ell\ell}^2)^2 - 4m_{jV}^2 p_\top^2}},$$

which is independent of the range of  $\theta^\bullet$ . At the end of the day, the hadronic cross section is given by

$$(6.53) \quad \begin{aligned} \sigma(s) = & \sum_{q=1}^{N_f} [2\pi] \int_0^{\sqrt{s}} dm_{jV} \int_0^{m_{jV}} dm_{\ell\ell} \int_0^{\frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}} dp_\top \int_{-\frac{1}{2} \ln \left( \frac{F^-}{F^+} \frac{s}{m_{jV}^2} \right)}^{\frac{1}{2} \ln \left( \frac{F^+}{F^-} \frac{s}{m_{jV}^2} \right)} dy \int d\Omega^\star \\ & \times \left[ \frac{1}{2048\pi^5} (2m_{\ell\ell}) \left( 1 - \frac{m_{\ell\ell}^2}{m_{jV}^2} \right) \right] \left[ \frac{1}{2m_{jV}^2} \right] J \\ & \times \{ f_{q\bar{q}}(x_a, x_b) |\mathcal{A}^{q\bar{q}}|^2 + f_{qg}(x_a, x_b) |\mathcal{A}^{qg}|^2 + f_{\bar{q}g}(x_a, x_b) |\mathcal{A}^{\bar{q}g}|^2 \}, \end{aligned}$$

where  $[2\pi]$  is the  $\varphi^\bullet$  integral,  $\left[ \frac{1}{2048\pi^5} (2m_{\ell\ell}) \left( 1 - \frac{m_{\ell\ell}^2}{m_{jV}^2} \right) \right]$  comes from the LIPS derived in the partonic c.m. frame, and  $\left[ \frac{1}{2m_{jV}^2} \right]$  is the flux factor. As the story unfolds, the order



of integration matters in terms of the efficiency of the numerical integrations. Experimentally, we are interested in  $m_{\ell\ell}$  and  $p_{\top}$ , so that's how we pick our bins. We are given  $m_{\ell\ell} \in [m_{\ell\ell\min}, m_{\ell\ell\max}]$  and  $p_{\top} \in [p_{\top\min}, p_{\top\max}]$  and we integrate  $m_{\ell\ell}$  and  $p_{\top}$  over these intervals, whilst we stay true to the integration limits of the other variables. One way to achieve this is to impose cuts or tacking in a factor of step function; however, the efficiency shows up when we have other cuts, as well. In a Monte-Carlo (MC) simulation, when we require say 1M sampling points, the integrator would omit a large portion of it and we would have to require more and more points to achieve a lower variance, which would increase the run time by insane amounts. Thus, we want to write down the  $m_{\ell\ell}$  and  $p_{\top}$  integrals as the two outermost integrals to pump up the efficiency. We then haven

$$\begin{aligned}
 & \int_0^{\sqrt{s}} dm_{jV} \int_0^{m_{jV}} dm_{\ell\ell} \int_0^{\frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}} dp_{\top} \rightarrow \\
 (6.54) \quad & \int_0^{\sqrt{s}} dm_{\ell\ell} \int_0^{\frac{s - m_{\ell\ell}^2}{2\sqrt{s}}} dp_{\top} \int_{p_{\top} + \sqrt{p_{\top}^2 + m_{\ell\ell}^2}}^{\sqrt{s}} dm_{jV}.
 \end{aligned}$$

We can impose custom limits for  $m_{\ell\ell}$  and  $p_{\top}$  integrals, which makes sure that the MC integration uses a lot bigger portion of the desired sampling points.

As a sanity check, we note that the *amplitudeless* cross-section integrals agree; that is to say, we have

$$\begin{aligned}
 & \int dx_a dx_b d\text{LIPS}_3 F = \\
 & \int_0^1 dx_a \int_0^1 dx_b \int_0^{\sqrt{x_a x_b s}} dm_{\ell\ell} \int_{-1}^1 dc_{\theta^*} \int_0^{\pi} d\theta^* \int_0^{2\pi} d\varphi^* \\
 & \times (2\pi) \left( \frac{1}{2048\pi^5} \right) (2m_{\ell\ell}) \left( 1 - \frac{m_{\ell\ell}^2}{x_a x_b s} \right) \sin(\theta^*) \left( \frac{1}{2x_a x_b s} \right)
 \end{aligned}$$

$$\begin{aligned}
&= \int_0^{\sqrt{s}} dm_{\ell\ell} \int_0^{\frac{s-m_{\ell\ell}^2}{2\sqrt{s}}} dp_{\top} \int_{p_{\top}+\sqrt{p_{\top}^2+m_{\ell\ell}^2}}^{\sqrt{s}} dm_{jV} \\
&\times \int_{-\frac{1}{2}\log\left(\frac{F^+}{F^-}\frac{s}{m_{jV}^2}\right)}^{\frac{1}{2}\log\left(\frac{F^+}{F^-}\frac{s}{m_{jV}^2}\right)} dy \int_0^{\pi} d\theta^* \int_0^{2\pi} d\varphi^* \\
(6.55) \quad &\times (2\pi) \left(\frac{1}{2048\pi^5}\right) (2m_{\ell\ell}) \left(1 - \frac{m_{\ell\ell}^2}{m_{jV}^2}\right) \sin(\theta^*) J\left(\frac{1}{2m_{jV}^2}\right) = \frac{1}{512\pi^3}.
\end{aligned}$$

Essentially, these are the cross-section integrals without the amplitudes or the parton distribution functions (PDFs). Note that in the original expression, we have  $c_{\theta^\bullet}$  running from  $-1$  to  $1$  and we split this integral into two, by emphasizing the sign of the variable, when we switch to the new variables. This means, we need to carry out the integrals with the new variables twice, one with  $\text{sgn} = 1$  and one with  $\text{sgn} = -1$ , where  $\text{sgn}$  is the sign in front in (6.47).

Furthermore, the *symmetry requirements* are confirmed; that is, we have

$$\begin{aligned}
&\int dx_a dx_b \text{dLIPS}_3 F[s_{12}, s_{13}, s_{23}] = \\
&\int_0^{\sqrt{s}} dm_{jV} \int_0^{m_{jV}} dm_{\ell\ell} \int_0^{\frac{m_{jV}^2-m_{\ell\ell}^2}{2m_{jV}}} dp_{\top} \int_{-\frac{1}{2}\log\left(\frac{F^+}{F^-}\frac{s}{m_{jV}^2}\right)}^{\frac{1}{2}\log\left(\frac{F^+}{F^-}\frac{s}{m_{jV}^2}\right)} dy \int_0^{\pi} d\theta^* \int_0^{2\pi} d\varphi^* \\
&\times (2\pi) \left(\frac{1}{2048\pi^5}\right) (2m_{\ell\ell}) \left(1 - \frac{m_{\ell\ell}^2}{m_{jV}^2}\right) \sin(\theta^*) J\left(\frac{1}{2m_{jV}^2}\right) [s_{12}, s_{13}, s_{23}] \\
(6.56) \quad &= \frac{s}{6144\pi^3},
\end{aligned}$$

and

$$\int dx_a dx_b \text{dLIPS}_3 F[p_a \cdot p_1, p_a \cdot p_2, p_a \cdot p_3, p_b \cdot p_1, p_b \cdot p_2, p_b \cdot p_3]$$

$$\begin{aligned}
&= \int_0^{\sqrt{s}} dm_{jV} \int_0^{m_{jV}} dm_{\ell\ell} \int_0^{\frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}} dp_{\top} \int_{-\frac{1}{2} \log\left(\frac{F^-}{F^+} \frac{s}{m_{jV}^2}\right)}^{\frac{1}{2} \log\left(\frac{F^+}{F^-} \frac{s}{m_{jV}^2}\right)} dy \int_0^{\pi} d\theta^* \int_0^{2\pi} d\varphi^* \\
&\times (2\pi) \left( \frac{1}{2048\pi^5} \right) (2m_{\ell\ell}) \left( 1 - \frac{m_{\ell\ell}^2}{m_{jV}^2} \right) \sin(\theta^*) J \left( \frac{1}{2m_{jV}^2} \right) \\
&\times [p_a \cdot p_1, p_a \cdot p_2, p_a \cdot p_3, p_b \cdot p_1, p_b \cdot p_2, p_b \cdot p_3] \\
(6.57) \quad &= \frac{s}{12288\pi^3}.
\end{aligned}$$

As for the momenta, we just write them in some c.m. frame, whichever is the easiest, relate the frames by boosts, and express the components in terms of  $x_a$ ,  $x_b$ ,  $m_{\ell\ell}$ ,  $\theta^\bullet$ ,  $\theta^*$ , and  $\varphi^*$ . We take care of the change of variables only during the numerical integration—we just want to see the CS angular structures of the cross section, and  $\theta^*$  and  $\varphi^*$  do not mix with other variables, so all the change of variables happen inside the structure coefficients, or the moments. *Ergo*, for analytical purposes, all we want is to express the unpolarized squared amplitude in terms of CS angles. Below, we summarize all the momenta:

$$(6.58) \quad p_a^\bullet = \frac{m_{jV}}{2}(1, 0, 0, 1),$$

$$(6.59) \quad p_b^\bullet = \frac{m_{jV}}{2}(1, 0, 0, -1),$$

$$(6.60) \quad p_{23}^\bullet = \left( \frac{m_{jV}^2 + m_{\ell\ell}^2}{2m_{jV}}, \frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}(s_{\theta^\bullet}, 0, c_{\theta^\bullet}) \right),$$

$$(6.61) \quad p_1^\bullet = \left( \frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}, -\frac{m_{jV}^2 - m_{\ell\ell}^2}{2m_{jV}}(s_{\theta^\bullet}, 0, c_{\theta^\bullet}) \right),$$

with

$$(6.62) \quad p_a^* = \Lambda^{\star\diamond} \Lambda^{\diamond\bullet} p_a^\bullet,$$

$$(6.63) \quad p_b^\star = \Lambda^{\star\diamond} \Lambda^{\diamond\bullet} p_b^\bullet,$$

$$(6.64) \quad p_{23}^\star = m_{\ell\ell}(1, \mathbf{0}),$$

$$(6.65) \quad p_1^\star = \Lambda^{\star\diamond} \Lambda^{\diamond\bullet} p_1^\bullet,$$

$$(6.66) \quad p_2^\star = \frac{m_{\ell\ell}}{2}(1, (s_{\theta^\star} c_{\varphi^\star}, s_{\theta^\star} s_{\varphi^\star}, c_{\theta^\star})),$$

$$(6.67) \quad p_3^\star = \frac{m_{\ell\ell}}{2}(1, -(s_{\theta^\star} c_{\varphi^\star}, s_{\theta^\star} s_{\varphi^\star}, c_{\theta^\star})).$$

The Lorentz transformation matrices are as given earlier. As for the components of the EW vector boson in the lab frame, we calculate  $p_{23}^\diamond = \Lambda^{\diamond\bullet} p_{23}^\bullet$  to obtain

$$(6.68) \quad E_V = \frac{(m_{jV}^2 + m_{\ell\ell}^2)(x_a + x_b) + (m_{jV}^2 - m_{\ell\ell}^2)(x_a - x_b)c_{\theta^\bullet}}{4m_{jV}\sqrt{x_a x_b}},$$

$$(6.69) \quad p_\top = \frac{(m_{jV}^2 - m_{\ell\ell}^2)s_{\theta^\bullet}}{2m_{jV}},$$

$$(6.70) \quad p_z = \frac{(m_{jV}^2 + m_{\ell\ell}^2)(x_a - x_b) + (m_{jV}^2 - m_{\ell\ell}^2)(x_a + x_b)c_{\theta^\bullet}}{4m_{jV}\sqrt{x_a x_b}}.$$

#### 6.4. SMEFT formalism

The SMEFT is a model-independent extension of the SM Lagrangian. In this framework, we build operators  $O_k^{(n)}$  of mass dimension  $n > 4$  and introduce Wilson coefficients  $C_k^{(n)}$  as the effective strength of the interaction at a UV scale beyond accessible collider reach:

$$(6.71) \quad \mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{n>4} \frac{1}{\Lambda^{n-4}} \sum_k C_k^{(n)} O_k^{(n)}.$$

In this study, we focus on the case  $n = 8$  and restrict ourselves to the leading-order SMEFT effects, namely we consider only the SM-SMEFT interference at the amplitude level and linearize our observables with respect to the SMEFT parameters.

The dimension-8 SMEFT operators of interest are presented in Table 6.1 [140, 134]. We include the corresponding sibling operators that involve the dual field strength only for completeness. In Table 6.1,  $\ell$  and  $q$  are left-handed SU(2) doublets and  $e$ ,  $u$ , and  $d$

Table 6.1. Dimension-8 four-fermion operators with a gluon field that contribute to the Drell-Yan transverse momentum spectrum classified according to their  $CP$  signature.

$CP$ -even		$CP$ -odd	
$\tilde{O}_{\ell^2 q^2 g}^{(1)}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{q}\gamma^\nu T^a q)\tilde{G}_{\mu\nu}^a$	$O_{\ell^2 q^2 g}^{(1)}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{q}\gamma^\nu T^a q)G_{\mu\nu}^a$
$\tilde{O}_{\ell^2 q^2 g}^{(3)}$	$(\bar{\ell}\tau^i\gamma^\mu\ell)(\bar{q}\tau^i\gamma^\nu T^a q)\tilde{G}_{\mu\nu}^a$	$O_{\ell^2 q^2 g}^{(3)}$	$(\bar{\ell}\tau^i\gamma^\mu\ell)(\bar{q}\tau^i\gamma^\nu T^a q)G_{\mu\nu}^a$
$\tilde{O}_{e^2 u^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{u}\gamma^\nu T^a u)\tilde{G}_{\mu\nu}^a$	$O_{e^2 u^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{u}\gamma^\nu T^a u)G_{\mu\nu}^a$
$\tilde{O}_{e^2 d^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{d}\gamma^\nu T^a d)\tilde{G}_{\mu\nu}^a$	$O_{e^2 d^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{d}\gamma^\nu T^a d)G_{\mu\nu}^a$
$\tilde{O}_{\ell^2 u^2 g}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{u}\gamma^\nu T^a u)\tilde{G}_{\mu\nu}^a$	$O_{\ell^2 u^2 g}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{u}\gamma^\nu T^a u)G_{\mu\nu}^a$
$\tilde{O}_{\ell^2 d^2 g}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{d}\gamma^\nu T^a d)\tilde{G}_{\mu\nu}^a$	$O_{\ell^2 d^2 g}$	$(\bar{\ell}\gamma^\mu\ell)(\bar{d}\gamma^\nu T^a d)G_{\mu\nu}^a$
$\tilde{O}_{q^2 e^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{q}\gamma^\nu T^a q)\tilde{G}_{\mu\nu}^a$	$O_{q^2 e^2 g}$	$(\bar{e}\gamma^\mu e)(\bar{q}\gamma^\nu T^a q)G_{\mu\nu}^a$

are right-handed singlets; in what follows, we switch back to the usual notation that says  $\ell$  ( $q$ ) is a Dirac lepton (quark) that appears in the partonic processes. Note that we are not introducing any operator that affects the  $ffV$  vertices, thus we just borrow the usual SM vertices for the SM amplitude and focus on the aforementioned SMEFT operators for the SMEFT amplitude. To derive the Feynman rules, we assume the convention

$\partial_\mu \rightarrow -ip_\mu$ , outgoing momenta, and multiplication by  $+i$ , which are all consistent with the sign convention of FeynArts. Note that the SMEFT operators have the gluon field strength tensor, which can be taken of the form

$$(6.72) \quad G_{\mu\nu}^a = \partial_\mu g_\nu^a - \partial_\nu g_\mu^a$$

for all practical purposes, for the nonabelian term gives us two gluons, which is not of interest. We have

$$(6.73) \quad G_{\mu\nu}^a = -ip_{g\mu}g_\nu^a + ip_{g\nu}g_\mu^a = -i(p_{g\mu}g_{\nu\rho} - p_{g\nu}g_{\mu\rho})g^{a\rho}$$

in the momentum space, where  $p_g$  is the gluon momentum. Noting that  $\tilde{G}_{\mu\nu}^a = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}G^{a\alpha\beta}$ , we obtain the vertex factors as follows:

$$(6.74) \quad \begin{aligned} V_{\ell^2 \frac{u^2}{d^2} g}^{\rho abc} = & \left\{ \begin{aligned} & (C_{\ell^2 q^2 g}^{(1)} \mp C_{\ell^2 q^2 g}^{(3)})[\gamma^\mu P_L]_\ell [\gamma^\nu P_L]_q + C_{e^2 \frac{u^2}{d^2} g}[\gamma^\mu P_R]_\ell [\gamma^\nu P_R]_q \\ & + C_{\ell^2 \frac{u^2}{d^2} g}[\gamma^\mu P_L]_\ell [\gamma^\nu P_R]_q + C_{q^2 e^2 g}[\gamma^\mu P_R]_\ell [\gamma^\nu P_L]_q \end{aligned} \right\} \\ & \times T_{bc}^a(p_{g\mu}g_{\nu\rho} - p_{g\nu}g_{\mu\rho}) \\ & + \left\{ \begin{aligned} & (\tilde{C}_{\ell^2 q^2 g}^{(1)} \mp \tilde{C}_{\ell^2 q^2 g}^{(3)})[\gamma^\mu P_L]_\ell [\gamma^\nu P_L]_q + \tilde{C}_{e^2 \frac{u^2}{d^2} g}[\gamma^\mu P_R]_\ell [\gamma^\nu P_R]_q \\ & + \tilde{C}_{\ell^2 \frac{u^2}{d^2} g}[\gamma^\mu P_L]_\ell [\gamma^\nu P_R]_q + \tilde{C}_{q^2 e^2 g}[\gamma^\mu P_R]_\ell [\gamma^\nu P_L]_q \end{aligned} \right\} \\ & \times T_{bc}^a(p_{g\alpha}g_{\beta\rho} - p_{g\beta}g_{\alpha\rho})\frac{1}{2}\epsilon_{\mu\nu\alpha\beta}, \end{aligned}$$

where we have now considered the cases  $q = u$  or  $d$  separately because there is a sign difference in the first term of each type of interaction.

### 6.5. Collins-Soper moments

The differential cross section can be parametrized in terms of the CS moments as [15]

$$(6.75) \quad \frac{d\sigma}{d\Omega^*} = \frac{3\sigma}{16\pi} \left[ 1 + c_{\theta^*}^2 + \sum_{m=0}^7 A_m Y_m(\Omega^*) \right],$$

where  $\Omega^* = (\theta^*, \varphi^*)$  are the CS angles, the  $Y_m$  are just orthogonal combos of the spherical harmonics  $Y_1^0$ ,  $Y_1^1 \pm Y_1^{-1}$ ,  $Y_2^0$ ,  $Y_2^1 \pm Y_2^{-1}$ , and  $Y_2^2 \pm Y_2^{-2}$ , namely

$$(6.76) \quad Y_0 = \frac{1}{2}(1 - 3c_{\theta^*}^2), \quad Y_1 = s_{2\theta^*}c_{\varphi^*}, \quad Y_2 = \frac{1}{2}s_{\theta^*}^2c_{2\varphi^*}, \quad Y_3 = s_{\theta^*}c_{\varphi^*}, \quad Y_4 = c_{\theta^*},$$

$$(6.77) \quad Y_5 = s_{\theta^*}^2s_{2\varphi^*}, \quad Y_6 = s_{2\theta^*}s_{\varphi^*}, \quad Y_7 = s_{\theta^*}s_{\varphi^*},$$

and the  $A_m$  are the CS moments. Using the orthogonality of the spherical harmonics, we obtain

$$(6.78) \quad A_0 = \frac{20}{3}\langle Y_0 \rangle + \frac{2}{3}, \quad A_1 = 5\langle Y_1 \rangle, \quad A_2 = 20\langle Y_2 \rangle, \quad A_3 = 4\langle Y_3 \rangle, \quad A_4 = 4\langle Y_4 \rangle,$$

$$(6.79) \quad A_5 = 5\langle Y_5 \rangle, \quad A_6 = 5\langle Y_6 \rangle, \quad A_7 = 4\langle Y_7 \rangle,$$

where we have defined

$$(6.80) \quad \langle Y_m \rangle = \frac{\int Y_m d\sigma}{\sigma}.$$

We note the following.

- We do not obtain novel angular structures. We get them only for the derivative operators involving the lepton current, so the said derivatives can turn into lepton momenta, giving us additional factors of sines and cosines of the CS angles. In

our case, we have the gluon momentum coming into the game, which can be rewritten in terms of the lepton-pair momentum at best, which doesn't contain any angles.

- The Lam-Tung relation [131]  $A_0 = A_2$  is satisfied.
- The moments  $A_5$ ,  $A_6$ , and  $A_7$  are nonzero only for final states with two or more jets within the SM, and in our case they are generated by the SMEFT operators of interest.

## 6.6. Numerical analysis

The coupling strengths are given by

$$(6.81) \quad C_{ff\gamma} = -eQ_f, \quad C_{ffZ}^{L/R} = eg_{L/Rf}, \quad C_{qqg} = -g_s,$$

where

$$(6.82) \quad g_{Lf} = \frac{T_{3f} - s_W^2 Q_f}{s_W c_W}, \quad g_{Rf} = \frac{-s_W^2 Q_f}{s_W c_W},$$

with  $T_{3\ell^-} = T_{3d_i} = -1/2$  and  $T_{3u_i} = 1/2$ . We employ the input scheme  $\{G_F, s_W^2, m_Z\}$ .

We have

$$(6.83) \quad G_F = 1.1663787 \times 10^{-5} \text{ GeV}^{-2}, \quad s_W = 0.23113, \quad m_Z = 91.1876 \text{ GeV},$$

and the derived parameters are

$$(6.84) \quad c_W = \sqrt{1 - s_W^2}, \quad \alpha = \frac{G_F s_W^2 c_W^2 m_Z^2 \sqrt{2}}{\pi}, \quad e = \sqrt{4\pi\alpha},$$



$$(6.85) \quad m_W = \sqrt{\frac{m_Z^2}{2} + \sqrt{\frac{m_Z^4}{4} - \frac{\alpha\pi m_Z^2}{G_F\sqrt{2}}}}.$$

As for the strong coupling, we need 1-loop running  $\alpha_s$ , which has a closed-form solution; however, to be consistent with the PDF set chosen, we call  $\alpha_s$  from the PDF set on the fly.

We use the NNPDF3.1 NLO [32] PDFs through LHAPDF [59] on Python. For the numerical routines, the Vegas module [133, 132] is employed.

We cross-validate our SM results with independent tools. For this purpose, MadGraph [24] is a great choice for snake-lovers and MCFM [60] for old-schoolers. Benchmark values agree within 0.5% among these packages.

The hadronic cross section can be written as

$$(6.86) \quad \sigma = \sigma_{\text{SM}} + \sum_i C_i \sigma_i + \sum_i \tilde{C}_i \tilde{\sigma}_i.$$

Let's define the  $Z$  point as the SM cross section integrated around the  $Z$  point,  $76 < m_{\ell\ell} < 106$  GeV and denote it by  $\sigma_{\text{SM}}^Z$ . Now we can just keep the SM part or activate the Wilson coefficients one at a time by setting  $C_i = 1$  or  $\tilde{C}_i = 1$  (on top of turning off the SM part, so we have a pure SMEFT contribution characterized by some  $C_i$  or  $\tilde{C}_i$ ), and plot the cross section at invariant mass bins beyond the  $Z$  point as a function of  $p_\top$ , normalized with respect to the  $Z$ -point cross section. We call this the ratio to the peak and denote it by  $\sigma_k/\sigma_{\text{SM}}^Z$ , where  $\sigma_k$  is any one of  $\sigma_{\text{SM}}$ ,  $\sigma_i$ , and  $\tilde{\sigma}_i$ . In Figures 6.2 and 6.3, we plot the ratios to the peak for the invariant mass bins [170, 350] GeV and [350, 1000] GeV in conjunction with the  $p_\top$  bins [50, 100], [100, 150], ..., [950, 1000] GeV. In these

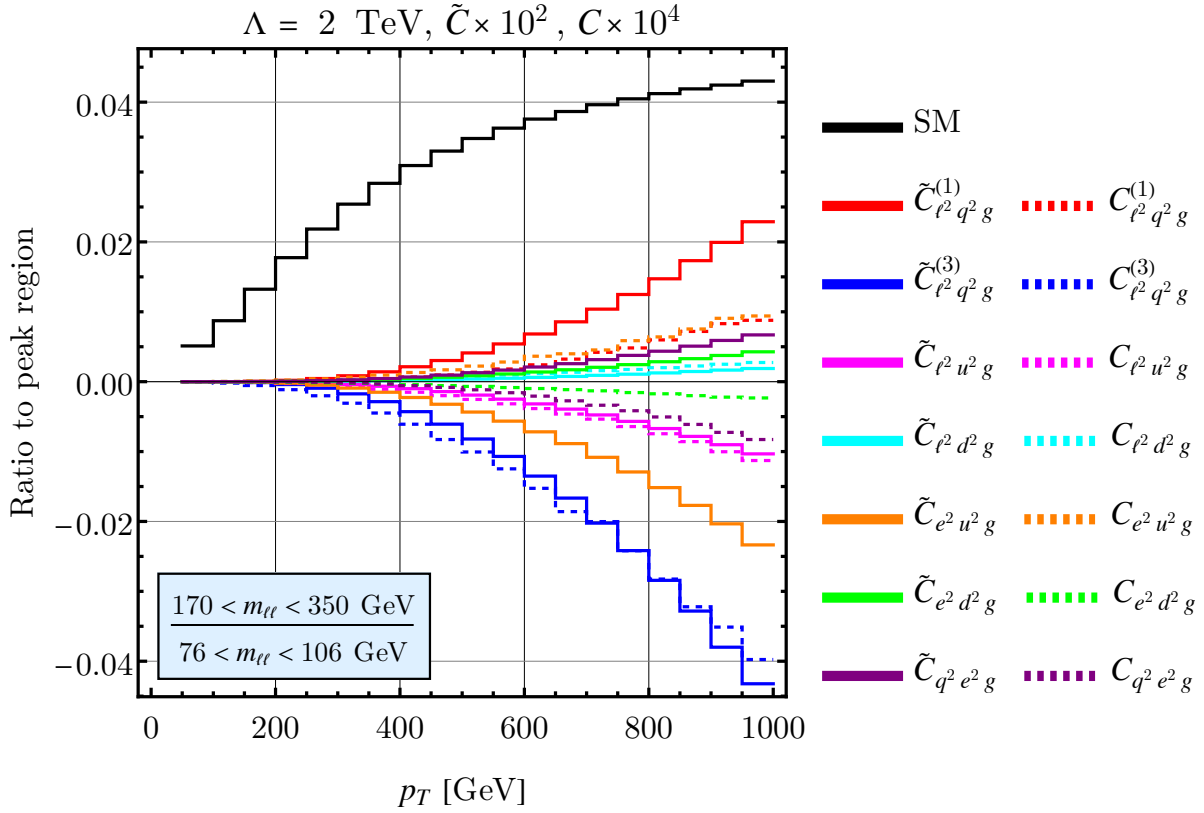


Figure 6.2. SM cross section and SMEFT corrections characterized by Wilson coefficients of interest as functions of dilepton transverse momentum for  $170 < m_{\ell\ell} < 350$  GeV, normalized with the SM cross section at the Z-peak.

figures, the black lines are the SM parts and the color lines are the SMEFT corrections. The solid color lines are the Wilson coefficients corresponding to the  $CP$ -even operators and the dashed ones are those corresponding to the  $CP$ -odd operators. We have a c.m. energy of  $\sqrt{s} = 13$  TeV, we set  $\Lambda = 2$  TeV, and we enhance the SMEFT contributions  $\tilde{\sigma}_i$  by a factor of  $10^2$  and  $\sigma_i$  by  $10^4$ . We avoid a log-log plot to display the SMEFT corrections

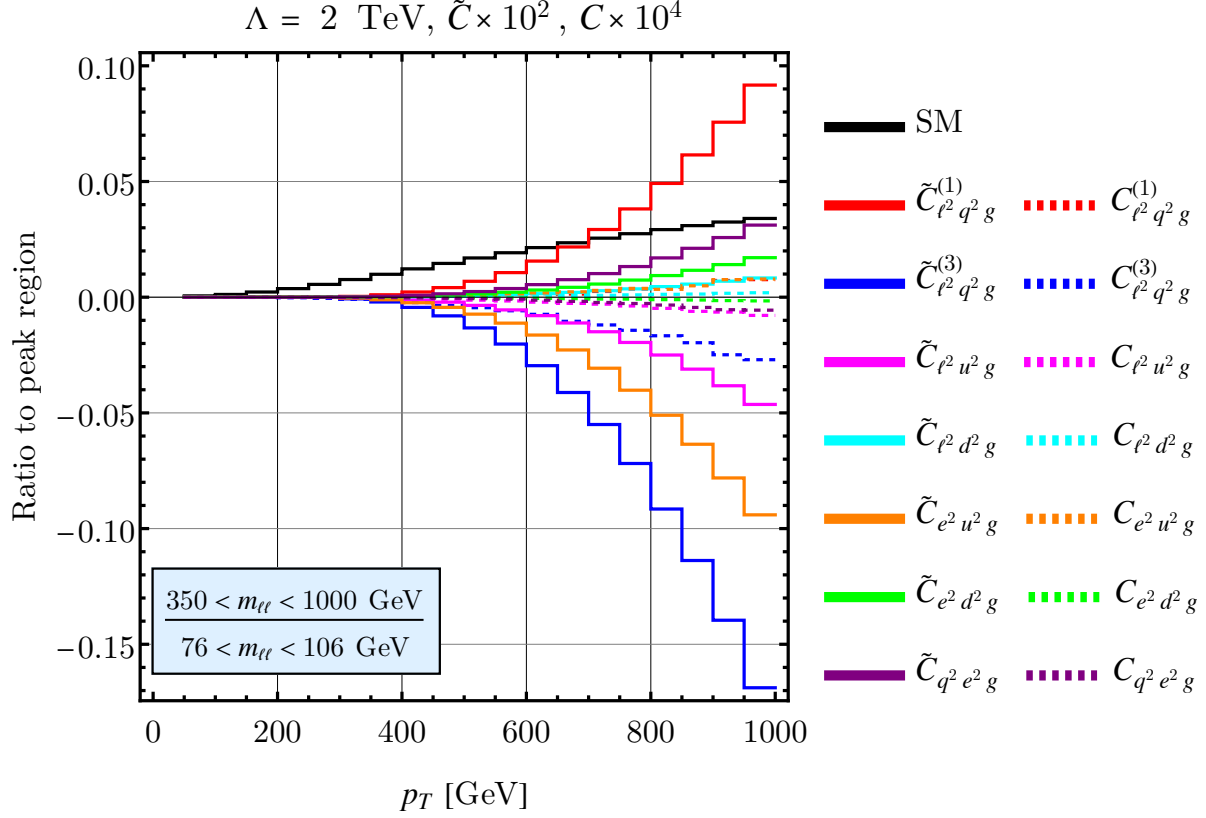


Figure 6.3. The same as in Figure 6.2 but for  $350 < m_{\ell\ell} < 1000 \text{ GeV}$ .

with signs. The numerical integration of the SM part and the high- $p_T$  SMEFT corrections can be performed with 40000 points in 1-2 second, while the low- $p_T$  SMEFT corrections are essentially noise for the integrator with bad statistics. As for the physics going on in these plots, we observe the following:

- The SMEFT corrections are meaningful for higher  $p_T$  values.

- The SMEFT corrections characterized by Wilson coefficients corresponding to the  $CP$ -even operators are significantly larger compared to the  $CP$ -odd operators. This makes sense because the SM Drell-Yan cross section is an  $CP$ -even observable (because the phase space and the squared amplitude are  $CP$ -even,  $\sigma \sim \int |\mathcal{A}|^2 \text{dLIPS}$ ) and the  $CP$ -even operators are expected to contribute to the cross section (because the SM-SMEFT interference amplitude is also  $CP$ -even,  $\sigma \sim \int \text{re}(\mathcal{A}_{\text{SM}}^* \mathcal{A}_{\text{SMEFT}}) \text{dLIPS}$ ). The  $CP$ -odd operators are not *fully* activated until a  $CP$ -odd observable joins the game.
- The SMEFT corrections characterized by the operators involving the interaction of two left-handed currents have the greatest sensitivity to  $p_{\text{T}}$ .

Next, we investigate the SMEFT corrections to the CS moments as functions of  $m_{\ell\ell}$  and  $p_{\text{T}}$ . In doing this, we note that any given CS moment is essentially the ratio of two cross-section integrals, one with a particular angular structure and one with unit weight, both of which can be written as an SM part plus some correction linearly proportional to a Wilson coefficient:

$$(6.87) \quad A = \frac{\int Y \text{d}\sigma}{\sigma} = \frac{N^{(0)} + CN^{(1)}}{D^{(0)} + CD^{(1)}},$$

where  $A$  is a CS moment,  $Y$  is the corresponding angular structure, and  $C$  is some Wilson coefficient, which needs linearizing so as to write

$$(6.88) \quad A = A^{(0)} + CA^{(1)}.$$

Now we focus on our observables of interest, namely the CS moments  $A_6$  and  $A_7$ , and in Figures 6.4 and 6.5, we present the plots demonstrating  $A^{(1)}$  across the same  $m_{\ell\ell}$  and  $p_T$  bins as in the previous figures (because the SM part of these moments are analytically zero). The colors correspond to the same Wilson coefficient. We note that the SMEFT

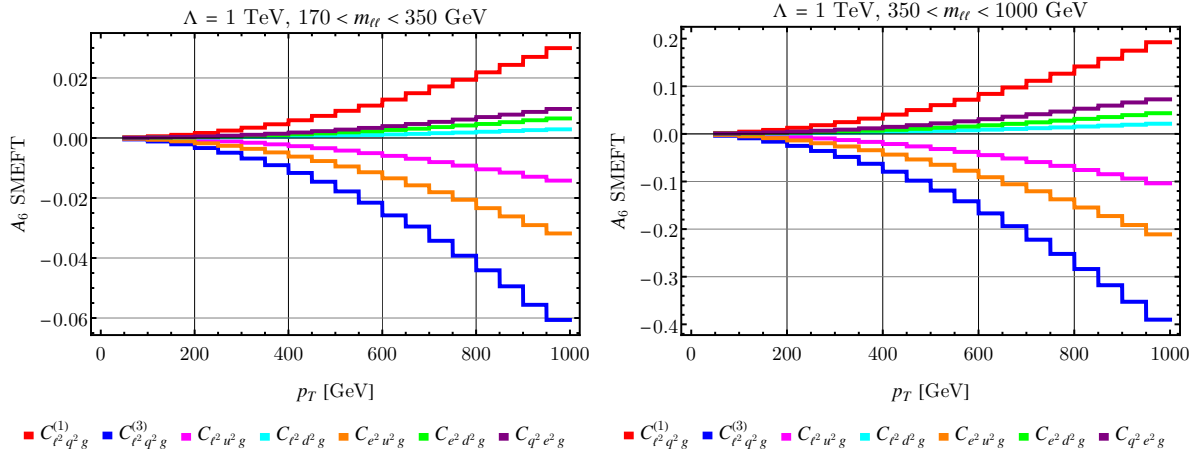


Figure 6.4. SMEFT contributions to  $A_6$  characterized by Wilson coefficients of interest as functions of dilepton transverse momentum at the dilepton invariant mass bins of  $170 < m_{\ell\ell} < 350$  GeV and  $350 < m_{\ell\ell} < 1000$  GeV.

contributions to  $A_5$  are all proportional to  $\Gamma_Z$ . We recall that we must activate the decay width of the  $Z$  boson only near the  $Z$  point, e.g.  $76 < m_{\ell\ell} < 106$  GeV, and for higher bins, these contributions are practically zero. Therefore, even though the angular structure corresponding to  $A_5$  is also  $CP$ -odd, it is negligible in the dilepton invariant masses assumed in our analysis.

### 6.7. HL-LHC simulation

In this section, we crunch in the numbers to simulate high-luminosity LHC (HL-LHC) following [45]. We set the collider energy to  $\sqrt{s} = 14$  TeV and assume the integrated

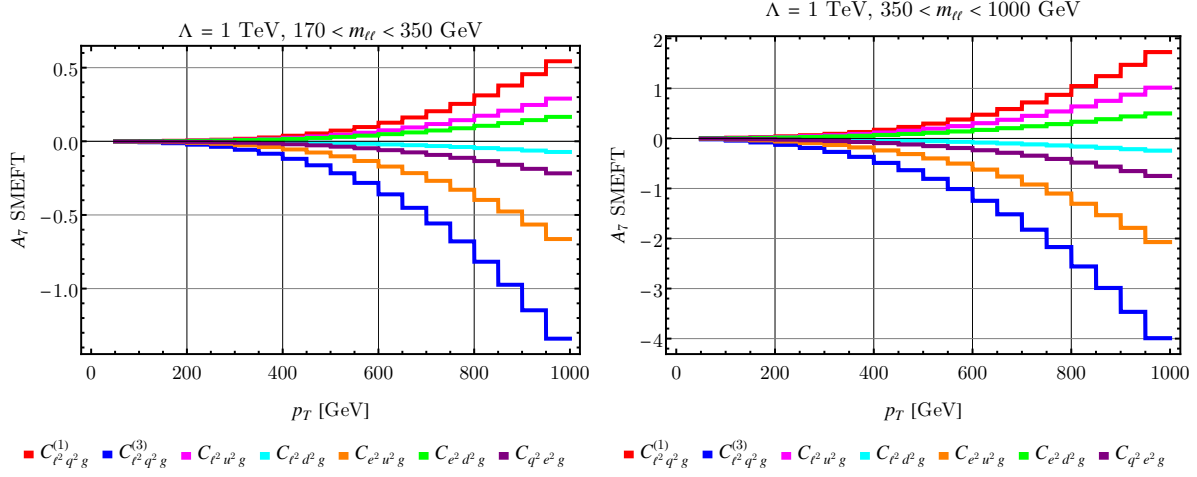


Figure 6.5. The same as in Figure 6.4 but for  $A_7$ .

luminosity to be  $\mathcal{L} = 3 \text{ ab}^{-1}$ . We assume the input scheme  $\{G_F, s_W^2, m_Z\}$ . We employ all the CMS cuts [60] in our calculations:

- Leading electron:  $p_T > 25 \text{ GeV}$
- Subleading electron:  $p_T > 20 \text{ GeV}$
- Both electrons:  $|\eta| < 2.4$
- Jet:  $p_T > 30 \text{ GeV}$ ,  $|y| < 2.4$
- Jet-electron separation:  $\Delta R_{je} > 0.4$  for all jet-electron pairs
- Dilepton system:  $p_T > 100 \text{ GeV}$ ,  $|y| < 2.4$

Given two particle momenta  $k$  and  $k'$ , we have

$$(6.89) \quad k_\top = k_x^2 + k_y^2, \quad \eta_k = \frac{1}{2} \log \left( \frac{|\mathbf{k}| + k_z}{|\mathbf{k}| - k_z} \right),$$

$$(6.90) \quad y_k = \frac{1}{2} \log \left( \frac{E_k + k_z}{E_k - k_z} \right), \quad \phi_k = \arctan \left( \frac{k_y}{k_x} \right),$$

$$(6.91) \quad \Delta\eta_{kk'} = \eta_k - \eta_{k'}, \quad \Delta\phi_{kk'} = \phi_k - \phi_{k'}, \quad \Delta R_{kk'} = \sqrt{\Delta\eta_{kk'}^2 + \Delta\phi_{kk'}^2}.$$

We note that even though our process of interest is  $pp \rightarrow je^-e^+$ , i.e. we have an electron and a positron pair coming out, the CMS language is to refer to these particles simply as *electrons*. We note that all these cuts are required to correctly generate statistical uncertainties.

We assume two types of bins: coarse and fine. The coarse bins presented in Table 6.2 have relative statistical uncertainties smaller than 5% in cross section, whereas the fine bins presented in Table 6.3 have relative statistical uncertainties smaller than 10%.

### 6.8. Fits of Wilson coefficients on HL-LHC pseudodata

Following [45], we assume uncorrelated statistical, uncorrelated systematic, and correlated systematic uncertainties on the experimental side, and correlated PDF and uncorrelated scale uncertainties on the theoretical side when we build the error matrix,  $\mathcal{E}$ :

$$(6.92) \quad \mathcal{E} = \mathcal{E}_{\text{exp}} + \mathcal{E}_{\text{PDF}} + \mathcal{E}_{\text{scale}}.$$

Let's discuss statistical uncertainties. The CS moments are defined by

$$(6.93) \quad A_m = N_m \frac{\int Y_m d\sigma}{\sigma} =: N_m \langle Y_m \rangle, \quad m = 6, 7,$$

where  $Y_m$  is some angular structure and  $N_m$  is some numerical factor,  $N_6 = 5$  and  $N_7 = 4$ .

We now derive the formula for the statistical uncertainties in  $A_m$ , denoted  $\delta A_m^{\text{stat}}$ . We have

$$(6.94) \quad A_m = N_m \langle Y_m \rangle = N_m \left[ \frac{1}{N} \sum_{E=1}^N Y_{m,E} \right],$$

Table 6.2. Coarse bins used in our HL-LHC simulations.

$m_{\ell\ell}$ [GeV]	$p_{\text{T}}$ [GeV]
[300, 360]	[100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 250, 270, 290, 310, 330, 360, 380, 410, 440, 490, 570, 7000]
[360, 450]	[100, 110, 120, 130, 140, 150, 160, 170, 180, 200, 230, 250, 270, 290, 310, 330, 350, 370, 400, 440, 490, 580, 7000]
[450, 600]	[100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 210, 230, 250, 270, 290, 320, 340, 360, 390, 430, 480, 580, 7000]
[600, 800]	[100, 110, 120, 130, 150, 170, 200, 220, 250, 290, 320, 360, 420, 520, 7000]
[800, 1100]	[100, 110, 120, 150, 170, 200, 230, 270, 330, 430, 7000]
[1100, 1500]	[100, 200, 290, 7000]
[1500, 2000]	[100, 7000]
[2000, 2600]	[100, 7000]

where  $E$  is the event number and  $N = \sigma\mathcal{L}$  is the total number of events. This gives us

$$\begin{aligned}
 \text{Var}[A_m] &= \text{Cov}[A_m, A_m] = \text{Cov}\left[\frac{N_m}{N} \sum_E Y_{m,E}, \frac{N_m}{N} \sum_{E'} Y_{m,E'}\right] \\
 (6.95) \quad &= \frac{N_m^2}{N^2} \sum_{E,E'} \text{Cov}[Y_{m,E}, Y_{m,E'}].
 \end{aligned}$$



Table 6.3. Fine bins used in our HL-LHC simulations.

$m_{\ell\ell}$ [GeV]	$p_T$ [GeV]
[300, 360]	[100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 250, 270, 290, 310, 330, 350, 370, 400, 420, 440, 470, 500, 530, 560, 600, 660, 760, 7000]
[360, 450]	[100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 240, 260, 290, 310, 330, 350, 370, 390, 410, 440, 470, 500, 530, 560, 610, 670, 770, 7000]
[450, 600]	[100, 110, 120, 130, 140, 150, 160, 190, 210, 230, 250, 270, 290, 320, 340, 370, 390, 420, 460, 490, 520, 550, 580, 620, 680, 780, 7000]
[600, 800]	[100, 110, 120, 130, 150, 170, 200, 220, 240, 260, 280, 310, 340, 380, 410, 440, 470, 510, 550, 620, 730, 7000]
[800, 1100]	[100, 110, 120, 140, 160, 180, 200, 220, 250, 270, 300, 330, 360, 410, 460, 540, 660, 7000]
[1100, 1500]	[100, 130, 160, 190, 230, 270, 320, 400, 520, 7000]
[1500, 2000]	[100, 210, 330, 7000]
[2000, 2600]	[100, 7000]

Assuming that the  $\{Y_{m,E}\}_{E=1}^N$  are independent and identically distributed (so the cross terms vanish), we obtain

$$\text{Var}[A_m] = \frac{N_m^2}{N^2} \sum_E \text{Cov}[Y_{m,E}, Y_{m,E}] = \frac{N_m^2}{N^2} \sum_E \text{Var}[Y_{m,E}]$$

$$(6.96) \quad = \frac{N_m^2}{N^2} \times N \text{ Var}[Y_m] = \frac{N_m^2}{N} \text{ Var}[Y_m],$$

and hence

$$(6.97) \quad \delta A_m^{\text{stat}} = \frac{N_m}{\sqrt{N}} \delta Y_m,$$

where

$$(6.98) \quad \delta Y_m = \sqrt{\langle Y_m^2 \rangle - \langle Y_m \rangle^2}.$$

Here, we have defined

$$(6.99) \quad \langle Y_m^2 \rangle = \frac{\int Y_m^2 d\sigma}{\sigma}.$$

Bin by bin, we compute statistical uncertainties and we use the minimum of 5% (10%) relative uncertainty in  $A_m$  for the coarse (fine) bins or the one we compute as describe here. We note that a relative uncertainty from the cross section in the denominator reflects back as a relative uncertainty in the CS moment. As for the systematics, we assume 1% uncorrelated and 2% correlated systematic uncertainties in the cross section and hence in the CS moments. In the pseudodata generation (to create statistics), we use only the experimental uncertainties, i.e. uncorrelated statistical (stat), uncorrelated systematic (usys), and correlated systematic uncertainties (csys):

$$(6.100) \quad A_{m,b} = A_{m,b}^{\text{SM}} + r_b \delta A_{m,b}^{\text{stat}} \oplus \delta A_{m,b}^{\text{usys}} + r' \delta A_{m,b}^{\text{csys}},$$

bin by bin,  $b$ , for a given pseudoexperiment. Here,  $r_b, r' \sim \mathcal{N}(0, 1)$  are unit normal variates and while the uncorrelated uncertainties are generated with a different random number for each bin, the correlated ones are generated with a single random number for a given pseudoexperiment so all the bins feel this shift uniformly.

Next, let us discuss the theoretical uncertainties. We start with PDF uncertainties. For replica-based PDF sets, the correlated PDF uncertainty matrix is built such that the  $bb'$  entry reads

$$(6.101) \quad (\mathcal{E}_{\text{PDF}})_{bb'} = \frac{1}{N_{\text{PDF}}} \sum_{\text{mem}=0}^{N_{\text{PDF}}} (A_{m,\text{mem}}^{\text{SM}} - A_{m,0}^{\text{SM}})_b (A_{m,\text{mem}}^{\text{SM}} - A_{m,0}^{\text{SM}})_{b'},$$

where  $A_{m,0(\text{mem})}^{\text{SM}}$  is the SM CS moment  $A_m$  evaluated at the PDF set member 0 (mem) and  $N_{\text{PDF}} = 100$ . The central value corresponds to the PDF member 0. On top of the PDF uncertainties, we also consider variations in renormalization and factorization scales,  $\mu_R$  and  $\mu_F$ , respectively. We introduce uncorrelated scale uncertainties according to

$$(6.102) \quad \frac{1}{2} \leq \frac{\mu_{R,F}}{\mu_0} \leq 2, \quad \frac{1}{2} \leq \frac{\mu_R}{\mu_F} \leq 2,$$

where we have defined

$$(6.103) \quad \mu_0 = \sqrt{p_{\text{T}}^2 + m_{\ell\ell}^2}.$$

Assuming steps of  $1/2$ , we evaluate the CS moments at various scales,  $A_{m,s}$ ,

$$(6.104) \quad (\mu_R, \mu_F) = \left(\frac{1}{2}, \frac{1}{2}\right) \mu_0, \quad (\mu_R, \mu_F) = \left(\frac{1}{2}, 1\right) \mu_0, \quad (\mu_R, \mu_F) = \left(\frac{1}{2}, \frac{3}{2}\right) \mu_0, \quad \dots,$$

$$(6.105) \quad (\mu_R, \mu_F) = \left(2, \frac{3}{2}\right) \mu_0, \quad (\mu_R, \mu_F) = (2, 2) \mu_0$$

that satisfy  $\frac{1}{2} \leq \frac{\mu_R}{\mu_F} \leq 2$ , so we have 12 possibilities. Then, we find the largest variation within this range and form a symmetric scale uncertainty using this largest variation at each bin:

$$(6.106) \quad \delta A_m^{\text{scale}} = \max \{ |A_{m,s} - A_{m,s'}| \}_{s,s'=1}^{12} .$$

The central value corresponds to  $(\mu_R, \mu_F) = (1, 1)\mu_0$ .

It is important to remark that said errors are borrowed from the cross-section analysis of our main reference for the present work, namely [45]. In the current study, our observables of interest are  $A_6$  and  $A_7$ , whose SM predictions are zero. Noting that pseudodata and hence the error matrix is generated using uncertainties based on SM predictions, we see that all the uncertainties are zero here, except for the uncorrelated statistical uncertainties. To be more precise, the SM predictions for  $A_6$  and  $A_7$  are zero, and therefore there cannot be any relative systematic uncertainties (1% or 2% of zero is zero); PDF and scale variations do not touch parity, and therefore PDF and scale uncertainties are zero. Even the term  $\langle Y_m \rangle$  in  $\delta A_m^{\text{stat}}$  is zero. Therefore, the error matrix consists of uncorrelated statistical uncertainties only:

$$(6.107) \quad \mathcal{E} = \text{diag} \{ \delta A_b^{\text{stat}2} \}_{b=1}^{N_{\text{bin}}} .$$

where we now suppress the CS moment index  $m$  to keep the discussion general. In Figures 6.6 and 6.7, we present the error budget plots. Here, we plot the bin-by-bin statistical uncertainties by the black line and the SMEFT contribution to the observables characterized by the indicated Wilson coefficient by the color lines. The bins are sorted

first in increasing order of  $m_{\ell\ell}$  and then within each  $m_{\ell\ell}$  bin, in increasing order of  $p_T$ .

This explains the waves.

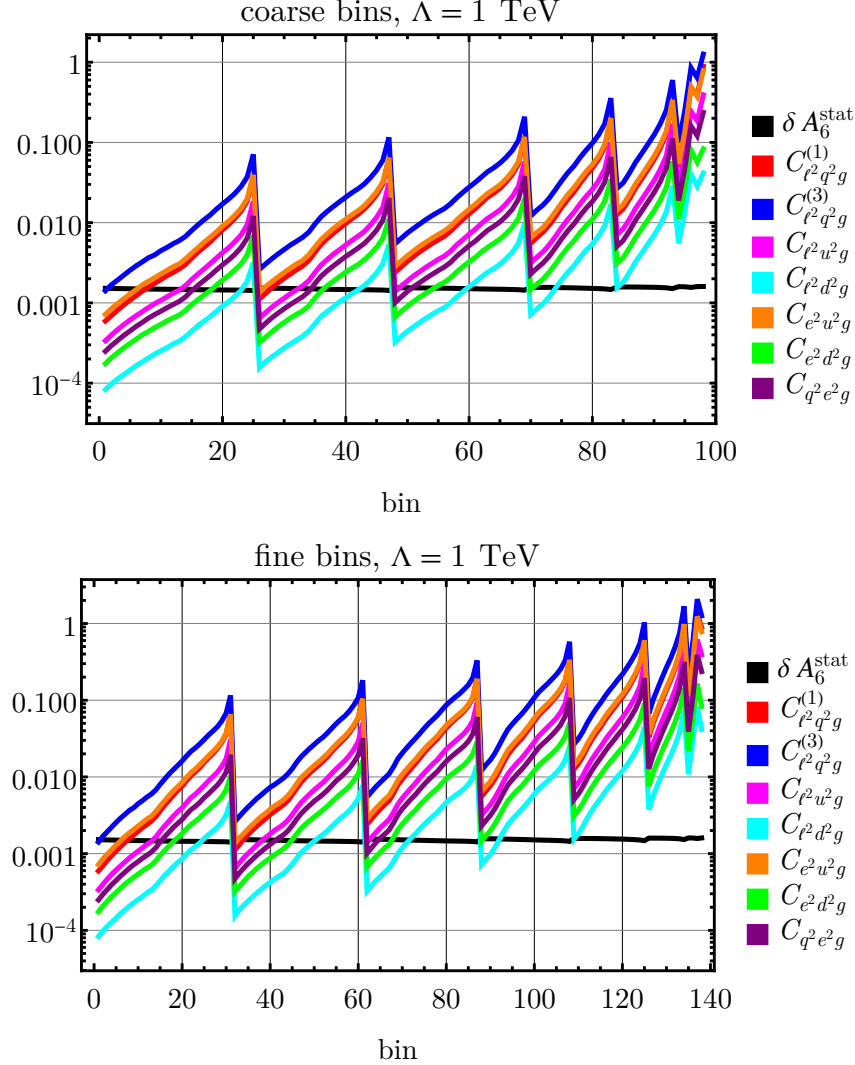


Figure 6.6. The error budget plot for the observable  $A_6$ . The SM value for this observable is zero. The black line is the statistical uncertainty. We also present the SMEFT corrections characterized by Wilson coefficients of interest to compare to the size of the uncertainty.

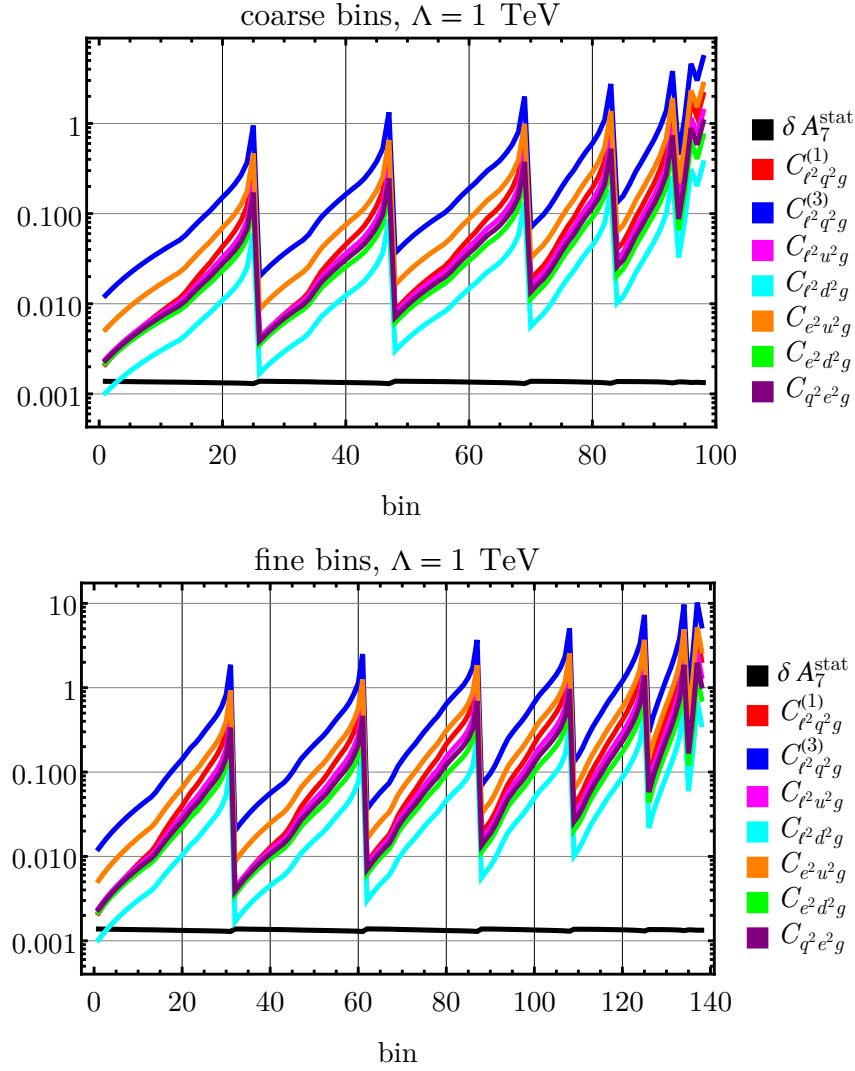


Figure 6.7. The same as in Figure 6.6 but for  $A_7$ .

For the statistical analysis, we define a standard  $\chi^2$  function:

$$(6.108) \quad \chi^2 = \sum_{b=1}^{N_{\text{bin}}} \sum_{b'=1}^{N_{\text{bin}}} \left( \hat{A} - A \right)_b \mathcal{H}_{bb'} \left( \hat{A} - A \right)_{b'},$$

where  $\mathcal{H} = \mathcal{E}^{-1}$  is the inverse uncertainty matrix,  $\hat{A}$  is the SMEFT observable, namely our linearized SMEFT expression for the CS moments  $A_6$  or  $A_7$ , and  $A$  is the pseudodata, which are of the form

$$(6.109) \quad \hat{A}_b = A_b^{\text{SM}} + \sum_{w=1}^{N_W} C_w A_b^{(w)} = \sum_{w=1}^7 C_w A_b^{(w)},$$

$$(6.110) \quad A_b = A_b^{\text{SM}} + r_b \delta A_b^{\text{stat}} = r_b \delta A_b^{\text{stat}}.$$

The  $\chi^2$  function can be expressed in the form

$$(6.111) \quad \chi^2 = k_0 + \sum_{w=1}^7 k_{1,w} C_w + \sum_{w=1}^7 \sum_{w'=1}^7 k_{2,ww'} C_w C_{w'} = k_0 + \mathbf{k}_1 \cdot \mathbf{C} + \mathbf{C} \cdot k_2 \mathbf{C}.$$

The vanishing gradient of this expression gives us the values of the Wilson coefficients that minimize the  $\chi^2$  function,

$$(6.112) \quad \nabla \chi^2(\overline{\mathbf{C}}) = \mathbf{0},$$

and the Hessian evaluated at these values give us the Fisher information matrix,

$$(6.113) \quad \mathcal{F} = \frac{1}{2} \nabla \nabla \chi^2(\overline{\mathbf{C}}).$$

Here, all the derivatives are with respect to the SMEFT parameters. With the quadratic form of the  $\chi^2$  function presented above, we obtain

$$(6.114) \quad \overline{\mathbf{C}} = \frac{1}{2} k_2^{-1} \mathbf{k}_1,$$

and

$$(6.115) \quad \mathcal{F} = k_2.$$

Since real data is not available yet and since our analysis strongly relies on pseudodata, we run multiple pseudoexperiments to generate statistics. Since our observable is linear in SMEFT parameters, the Fisher matrix is constant and we manage to avoid averaging over pseudoexperiments.

We present the nonmarginalized 2-sigma bounds in Figure 6.8, the corresponding effective scales in Figure 6.9, the marginalized 2-sigma bounds in Figure 6.10, and the corresponding effective scales in Figure 6.11. From these figures, we note the following:

- The  $A_6$  fits yield weaker bounds on the Wilson coefficients of interest compared to the  $A_7$  fits. The reason is the competition between the size of the SMEFT corrections to the observables and the anticipated statistical uncertainties, which is the only source of uncertainty in this study.
- The combined fit results resemble the  $A_7$  fits because of its dominance.
- The fine binning leads to bounds twice as strong as the coarse binning.
- In  $1d$  fits, we can obtain effective scales up to 9 TeV. These scales are computed using the 2-sigma bounds on Wilson coefficients as  $\Lambda/C_w^{1/4}$ , with  $\Lambda = 1$  TeV.
- Once we activate all the Wilson coefficients (so as to perform a  $7d$  fit), we see that the allowed intervals grow dramatically. This signals strong interplay between the SMEFT parameters.
- In the  $7d$  fit, the fine binning still leads to stronger bounds but the ratio is now slightly less than 2.



We present representative confidence ellipses in Figures 6.12 and 6.13. Each figure contain four plots. On the top row, we present the case of coarse (left) and fine (bins) in the nonmarginalized fit, and on the bottom row, we have the corresponding marginalized ellipses. From these figures, we note the following:

- The remarks from the *bound* plots apply.
- The  $A_6$  and  $A_7$  fit results are complementary, namely they yield distinct correlations.
- The nonmarginalized ellipses, namely the confidence ellipses of  $2d$  fits yield nearly flat directions, i.e. ellipses elongated wildly.
- The confidence ellipses projected from the  $7d$  fit do not display flat directions, which can be explained by the interplay between the SMEFT parameters.

Finally, the correlation matrices are presented in Figures 6.14–6.16.

### 6.9. Coda

In this study, we have studied the  $CP$ -odd dimension-8 SMEFT operators of semi-leptonic four-fermion interactions coupled to the gluon field strength tensor. The interference of the amplitudes arising from these operators with the SM terms generate  $CP$ -odd observables. To investigate the effects of said operators, we considered the DY production with a single jet at the HL-LHC. We used the anticipated dilepton invariant mass and transverse momentum bins at the CMS, together with the detector cuts. Our observables of interest were the  $CP$ -odd CS moments,  $A_5$ ,  $A_6$ , and  $A_7$ , which are zero at tree level or with a single jet at the SM. However, since  $A_5$  is linear in the  $Z$ -boson decay width, it was negligible at the energy levels considered, we carried out the fitting for  $A_6$  and  $A_7$ . Using

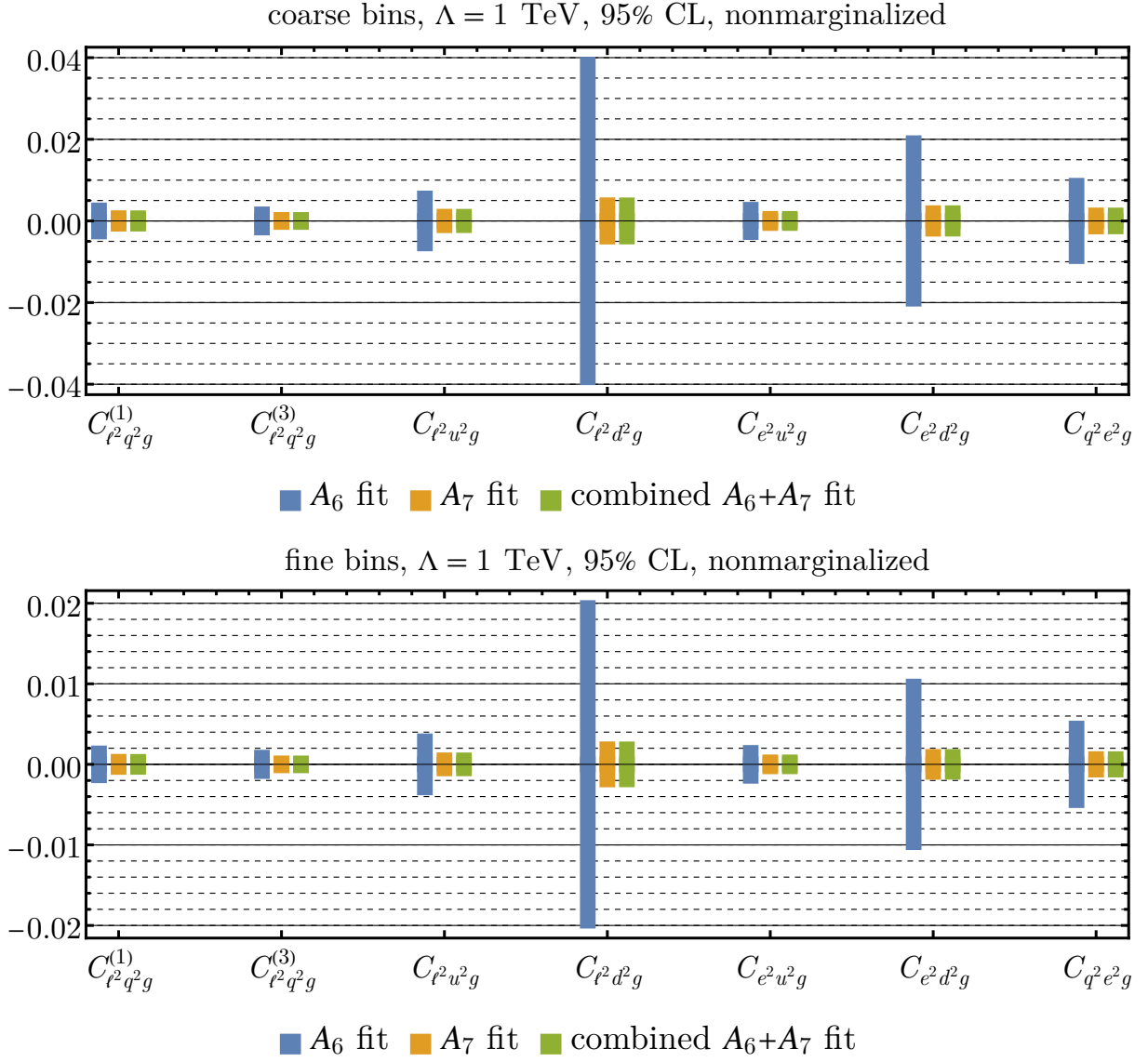


Figure 6.8. 95% CL bounds of  $CP$ -odd Wilson coefficients from single-parameter fits at  $\Lambda = 1$  TeV for the observables  $A_6$  and  $A_7$ , as well as the combined fit assuming independence, with coarse (top) and fine (bottom) bins.

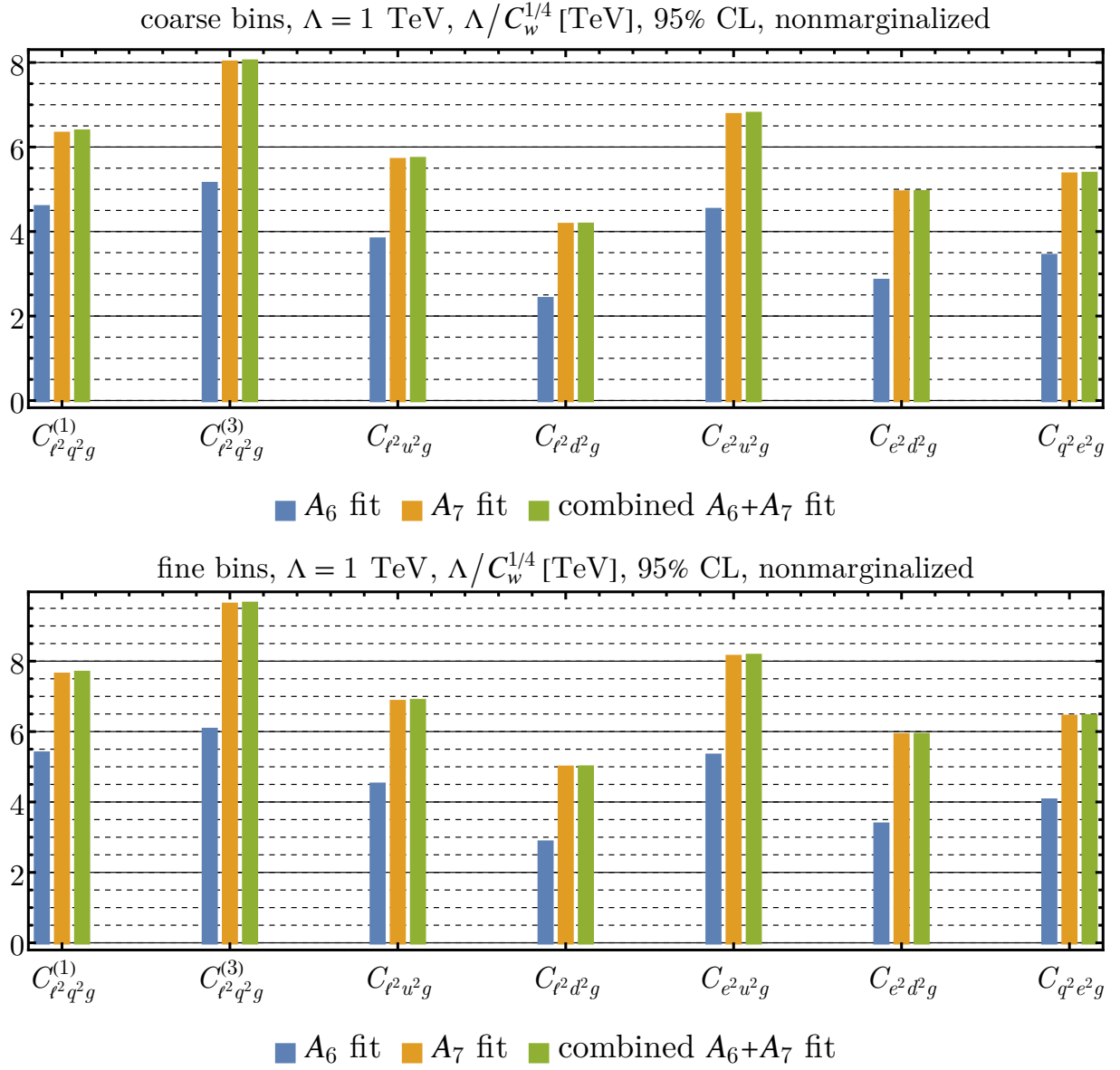


Figure 6.9. Effective UV cut-off scales,  $\Lambda/C_w^{1/4}$ , corresponding to the non-marginalized 95% CL bounds given in Figure 6.8.

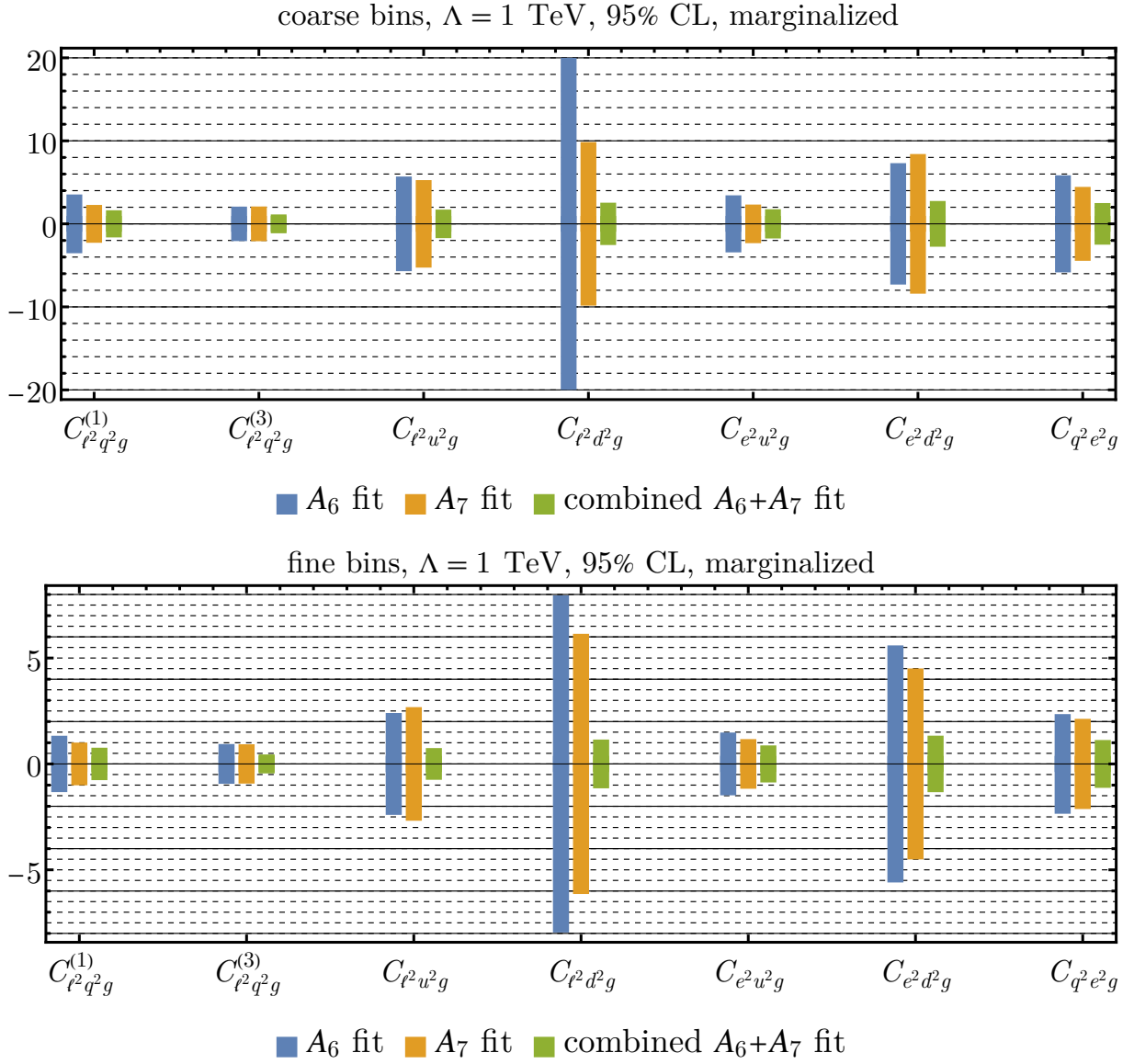


Figure 6.10. 95% CL bounds of  $CP$ -odd Wilson coefficients from multi-parameter fits at  $\Lambda = 1$  TeV for the observables  $A_6$  and  $A_7$ , as well as the combined fit assuming independence, with coarse (top) and fine (bottom) bins.

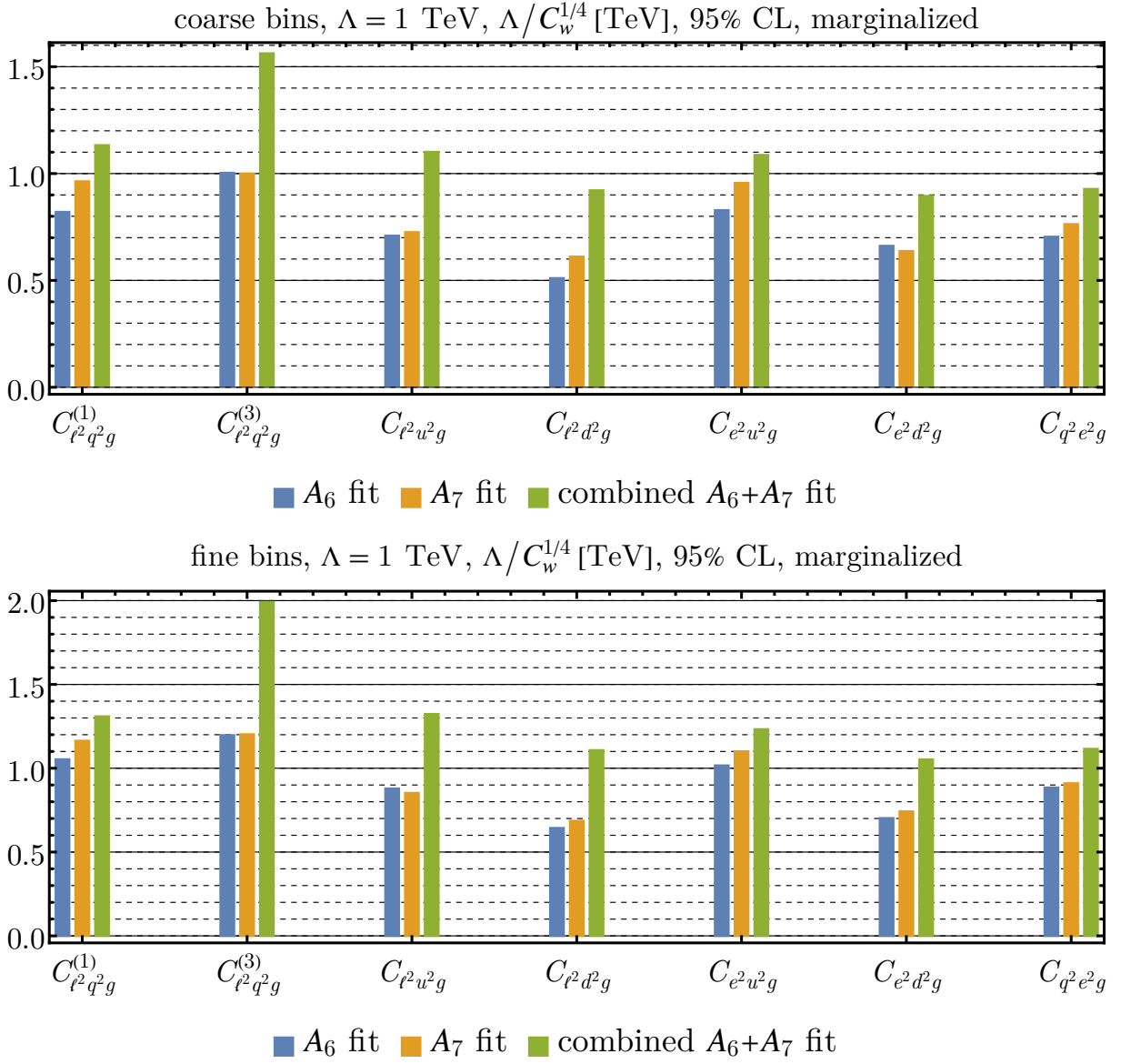


Figure 6.11. Effective UV cut-off scales,  $\Lambda/C_w^{1/4}$ , corresponding to the marginalized 95% CL bounds given in Figure 6.10.

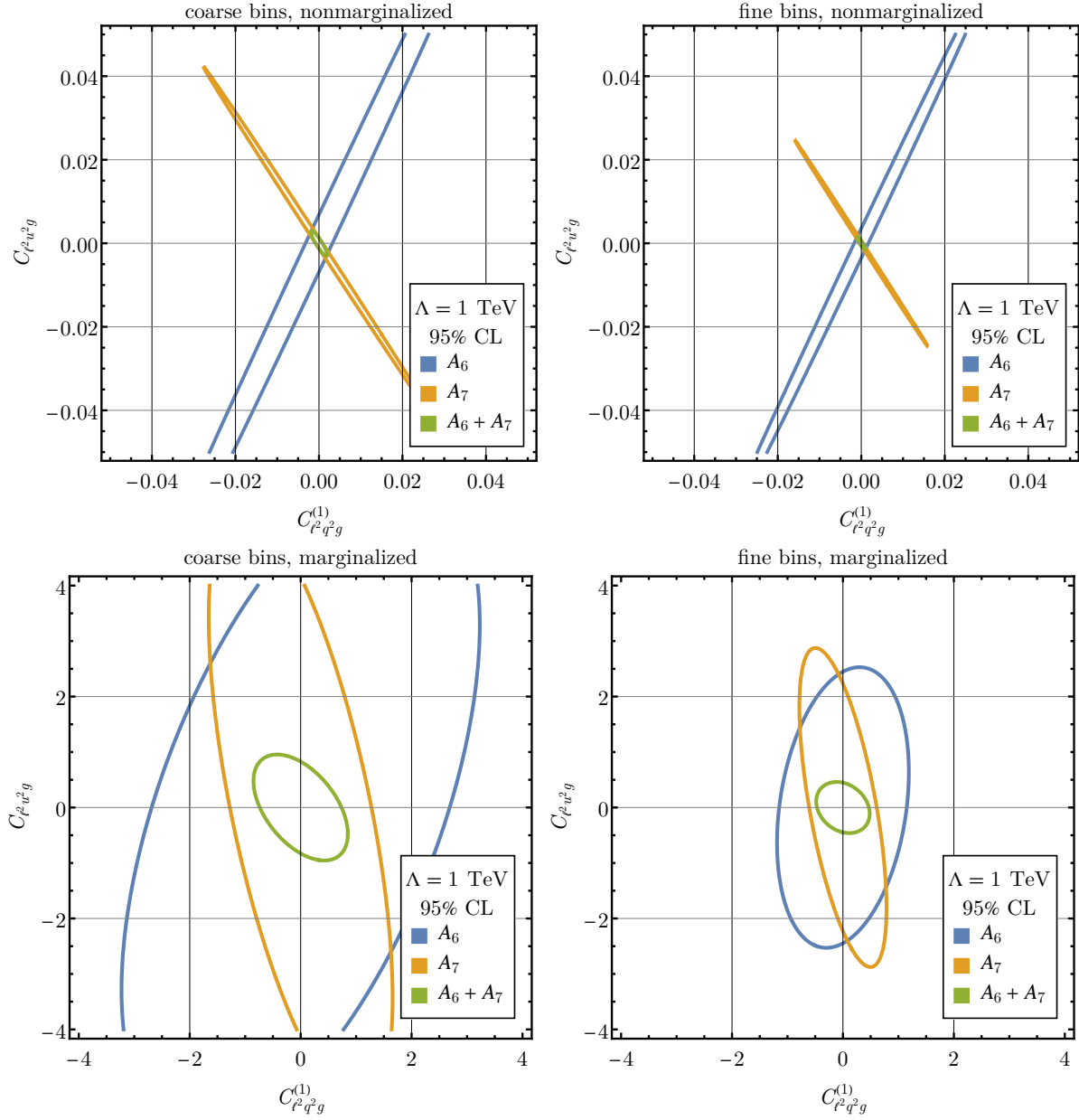


Figure 6.12. Nonmarginalized (top) and marginalized (bottom) 95% CL ellipses for the Wilson coefficients  $C_{\ell^2 q^2 g}^{(1)}$  and  $C_{\ell^2 u^2 g}$  at the HL-LHC for coarse (left) and fine (right) bins from the fits using the observable  $A_6$  and  $A_7$ , as well as the joint fit.

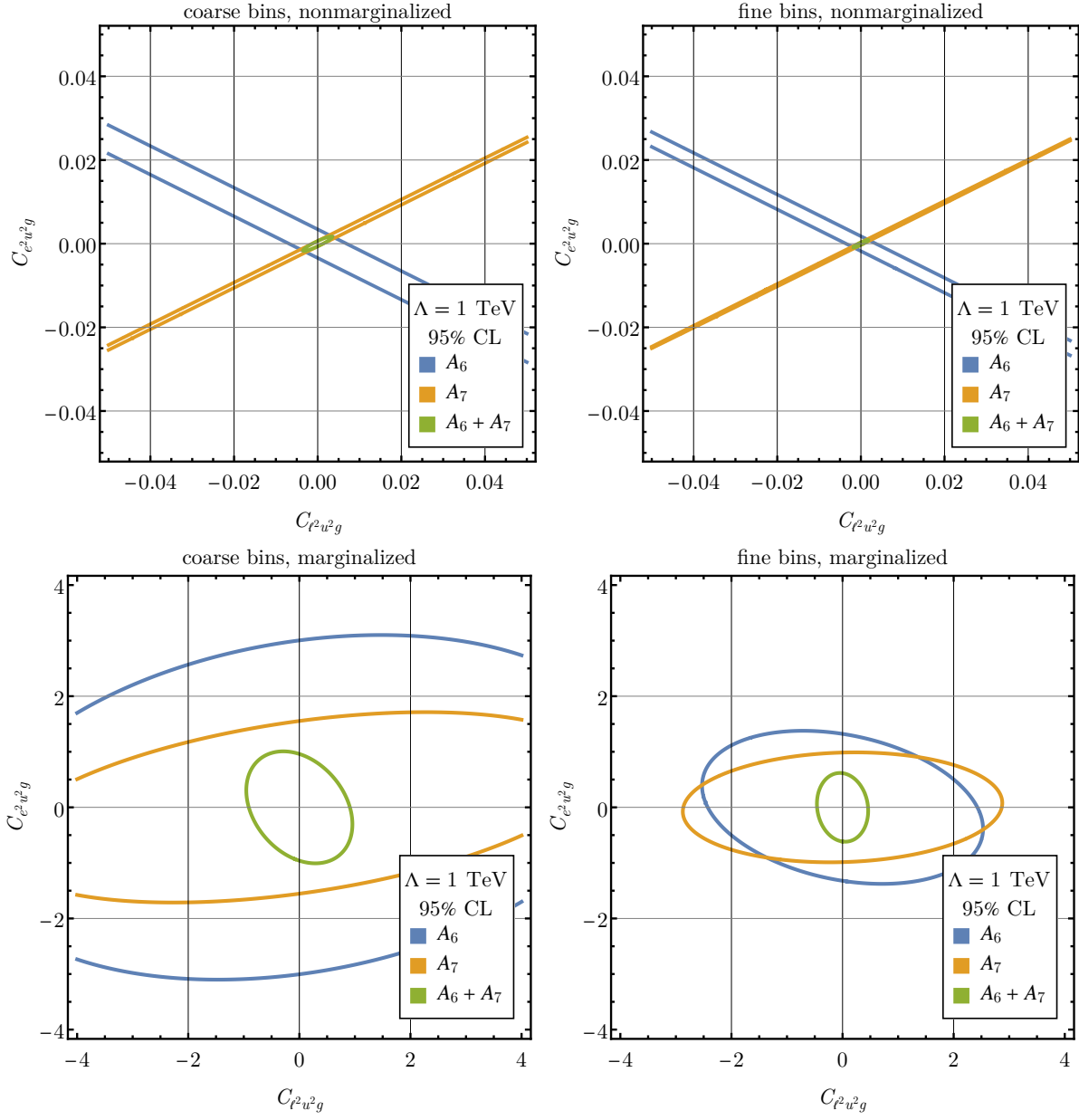


Figure 6.13. The same as in Figure 6.12 but for  $C_{\ell^2 u^2 g}$  and  $C_{e^2 u^2 g}$ .

simulated HL-LHC data with a collider energy of 14 TeV and an integrated luminosity of  $3 \text{ ab}^{-1}$ , we performed a comprehensive analysis for the fits of the SMEFT parameters.

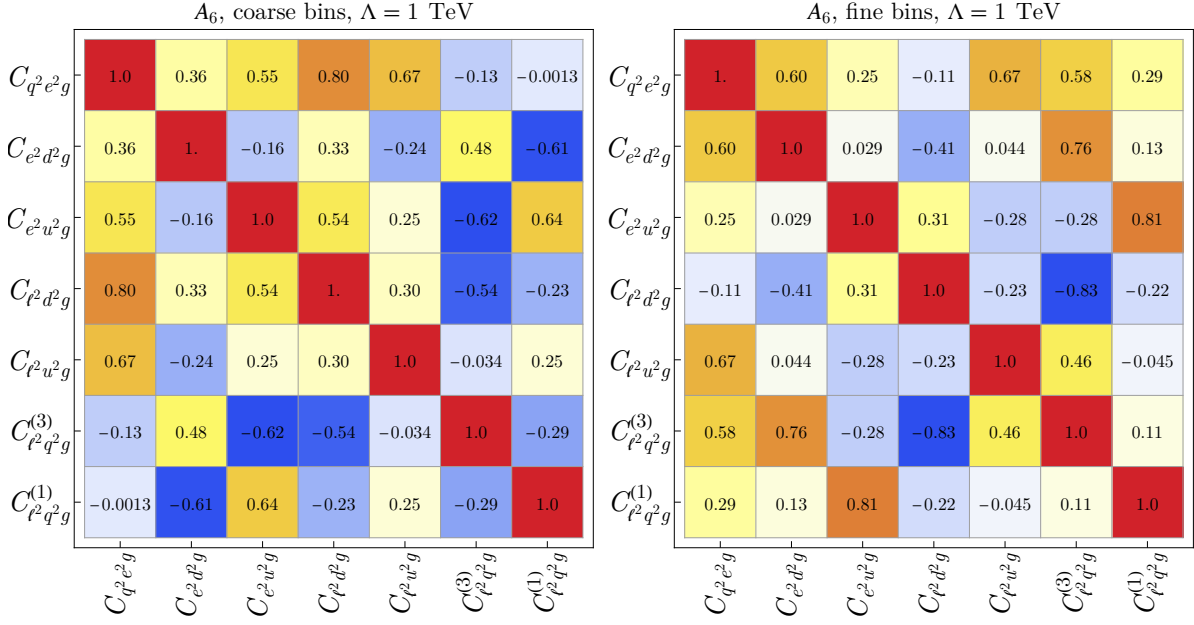


Figure 6.14. Correlation matrix for the fits of  $A_6$  with coarse (left) and fine (right) bins.

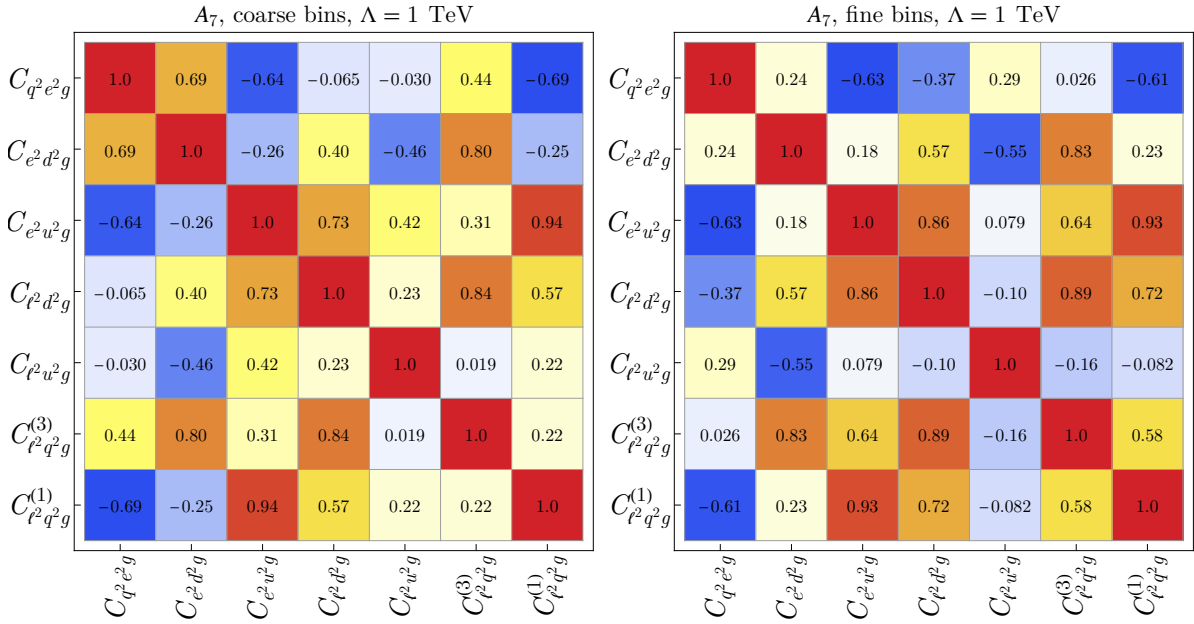


Figure 6.15. The same as in Figure 6.14 but for  $A_7$ .



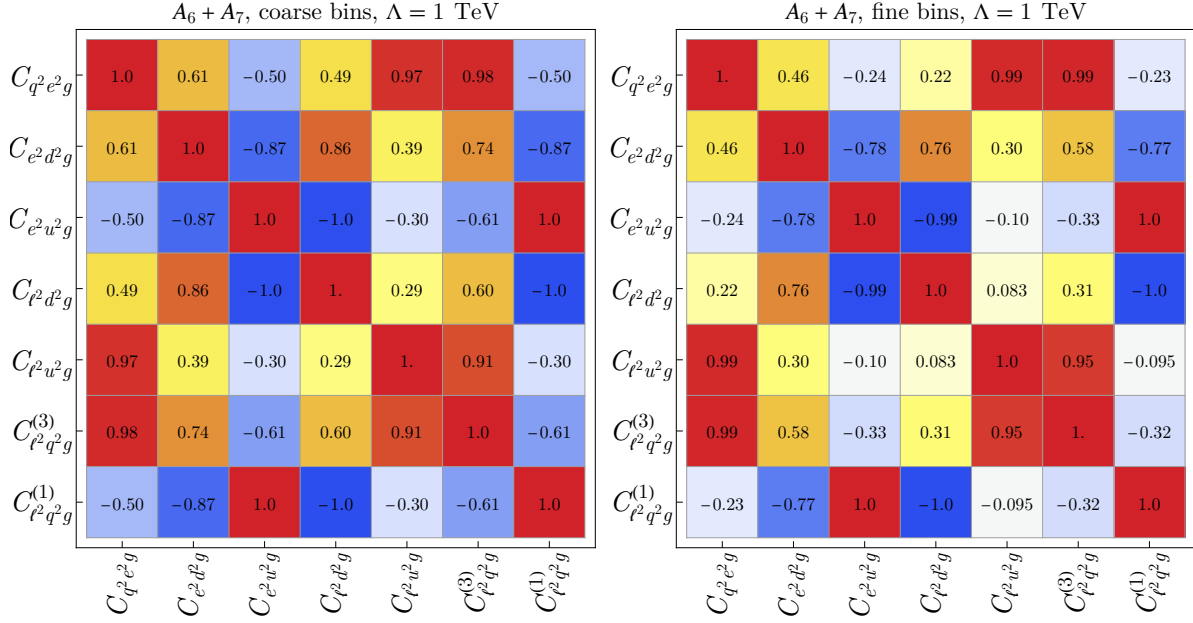


Figure 6.16. The same as in Figure 6.14 but for the joint fit of  $A_6$  and  $A_7$ .

The single-parameter fits yielded bounds at the order of  $\mathcal{O}(0.01)$ , translating into effective UV scales reaching up to approximately 8 TeV with coarse binning and 9 TeV with fine binning. However, when all seven Wilson coefficients were activated simultaneously, these constraints weakened significantly, namely by two to three orders of magnitude, bringing the effective scales down to about 1.5 TeV for the coarse binning and 2 TeV with the fine binning. This substantial shift implies strong interplay among the SMEFT operators. Despite this weakening, our study clearly highlights the capability and importance of using angular observables at the HL-LHC to probe  $CP$ -odd dimension-8 operators. These results set concrete benchmarks, motivating further detailed investigations into subtle signals of  $CP$ -violating new physics.

## CHAPTER 7

### Conclusion

*I think I could stand anything, any suffering, only to be able to say and to repeat to myself every moment, “I exist.” In thousands of agonies, I exist. I’m tormented on the rack, but I exist! Though I sit alone in a pillar, I exist! I see the sun, and if I don’t see the sun, I know it’s there. And there’s a whole life in that, in knowing that the sun is there.*

Fyodor Dostoyevsky, *The Brothers Karamazov*

#### 7.1. Summary of work and findings

This thesis was written in the service of a clear question. Can future colliders, through the tools of precision phenomenology, offer new ways to constrain physics beyond the Standard Model (SM)? That question shaped each project, each calculation, and each decision in the chapters that followed. The analyses were grounded in the projected capabilities of machines of tomorrow still under construction or under consideration. The aim was to understand what these colliders could realistically probe, and how far they could extend current theoretical frameworks when paired with well-defined observables.

Each chapter served a distinct purpose. Together, they contribute to a broader picture of how future collider programs can support precision tests of the SM and guide new physics searches.

Chapter 3 focused on the Electron-Ion Collider and examined whether it could access new physics scenarios that are otherwise difficult to probe. By studying neutral-current (NC) parity-violating observables, we showed that the Electron-Ion Collider (EIC) has the potential to lift degeneracies in the parameter space left open by NC Drell-Yan (DY) process at the Large Hadron Collider (LHC). This required a careful treatment of beam polarization and luminosity uncertainties, which were included as fit parameters for the first time in this context. The study confirmed that with moderate energies and high control over initial states, the EIC could provide access to unique directions in the Standard Model Effective Field Theory (SMEFT) parameter space.

Chapter 4 extended that effort. The operator set was expanded, next-to-leading order corrections from quantum chromodynamics (QCD) were included, and two additional machines, the Large Hadron-electron Collider (LHeC) and the Future Circular Collider (FCC), were incorporated into the analysis. Each collider brings access to a distinct kinematic regime. The central aim was to test whether these machines, when treated collectively, can explore the SMEFT parameter space in a complementary way. The result is promising. Their differences in energy and kinematic coverage offer a route to reducing parameter degeneracies and imposing stronger bounds on several SMEFT parameters relevant to NC deep inelastic scattering. The interplay with global fits based on electroweak (EW) precision observables is nontrivial. These future lepton-hadron colliders are positioned to resolve flat directions that persist in the global fits with Higgs, top, diboson, and  $Z$ -pole measurements. In that sense, they can clarify the underlying structure of new physics constraints and expand the reach of precision phenomenology.

Chapter 5 shifted gears and introduced a new observable in electron-positron collisions at the FCC. The focus was on transverse spin asymmetries and their potential to improve sensitivity to the electron Yukawa coupling. This quantity is notoriously small, and no direct measurement has yet been made. The asymmetries proposed in this study provide a clean and statistically enhanced signal, grounded in expected detector capabilities and polarization configurations. Although this analysis was carried out outside the SMEFT formalism, the overall purpose remained aligned with the rest of the thesis, namely finding observables at future colliders that can isolate small, theoretically clean effects.

Chapter 6 returned to the SMEFT but moved into less explored territory. We studied DY production with an associated jet and investigated CP-violating gluonic operators at dimension eight. This region of parameter space has not received much attention, in part because it lies outside the dominant directions constrained by inclusive processes. Our study showed that with carefully chosen observables, particularly dilepton angular distributions, it is possible to achieve meaningful sensitivity. The backbone analysis is complete. Work continues on matching to ultraviolet models, along with potential connections to spin-2 extensions. These are still being explored with care, but the path forward is now concrete.

Across four chapters, we investigated the potential reach of the EIC, LHeC, FCC, and the high-luminosity LHC. It proposed observables tailored to the structure and strengths of each machine. We introduced methods for resolving degeneracies in global SMEFT fits. We expanded the parameter space considered in precision collider studies. These contributions were not guided by abstraction. They were shaped by collider realities and theoretical clarity. At the end of the day, taxpayers' money is well spent.

## 7.2. Contributions and perspective

This thesis led to several publications:

- R. Boughezal, A. Emmert, T. Kutz, S. Mantry, M. Nycz, F. Petriello, K. Simsek, D. Wiegand, X. Zheng  
*Neutral-current electroweak physics and SMEFT studies at the EIC*  
Phys. Rev. D **106** (2022) 016006, arXiv:2204.07557
- C. Bissolotti, R. Boughezal, K. Simsek  
*SMEFT probes in future precision DIS experiments*  
Phys. Rev. D **108** (2023) 075007, arXiv:2306.05564
- R. Boughezal, F. Petriello, K. Simsek  
*Transverse spin asymmetries and the electron Yukawa coupling at an FCC-ee*  
Phys. Rev. D **110** (2024) 075026, arXiv:2407.12975

We also contributed to:

- *Electron Ion Collider for High Energy Physics*, Snowmass 2021 White Paper, arXiv:2203.13199.
- *SMEFT analysis with LHeC, FCC-eh, and EIC DIS pseudodata*, DIS2023 Proceedings, arXiv:2307.09459

Work continues on the DY plus jet process. The backbone analysis is complete. Further development is ongoing.

The focus of this thesis has been precision and collider phenomenology as tools for exploring new physics. The methods developed are not confined to the examples studied. They offer a modular and transparent approach to interpreting the capabilities of future

machines. The emphasis throughout has been on theoretical consistency, experimental feasibility, and numerical reproducibility. The computational infrastructure created in the process, ranging from simulation and fit pipelines to integration routines and analysis frameworks, is general and can be adapted to future problems.

One of the main goals of this thesis has been to understand what we can say about new physics when we haven't seen any new particles yet. So far, experiments at the LHC and elsewhere have not found direct evidence of particles beyond the SM (BSM). This could mean new physics is either too heavy to produce with current colliders or too weakly coupled to stand out in the data. But just because we do not see new particles does not mean there is nothing there. Heavy particles can still leave indirect effects by slightly changing the behavior of known processes.

This is where SMEFT comes in. It gives us a way to describe how unknown heavy physics would show up as small deviations from SM predictions. These deviations are captured by a set of parameters called Wilson coefficients. Each coefficient tells us how strongly a certain type of new interaction could affect measurable quantities. By putting limits on these coefficients, we are not just testing one model at a time, we are testing broad classes of possible theories all at once. That is what makes SMEFT a powerful tool. It gives us a model-independent way to look for signs of new physics, even if the new particles themselves are out of reach.

These efforts are part of a broader research direction that I plan to continue. The core objective remains the same: identify clean observables, develop realistic predictions, and constrain physics beyond the SM through precision.

### 7.3. Outlook and Future Directions

The next steps in my research will extend the work presented here. The central focus will include precision calculations for top-quark observables, resummation techniques, and the development of global parton distribution functions (PDFs). These projects are conceptually aligned with the work in this thesis. They continue the same logic, structure, and purpose.

My early work explored rare top-quark decays and BSM models. The PhD years added a deep engagement with SMEFT, collider physics, and QCD and EW precision. The growth that followed was shaped by the opportunity to work with and learn from some of the most brilliant particle physicists who guide the field. That collaboration sharpened my thinking, expanded my tools, and deepened my understanding of precision phenomenology.

The technical components are ready. The work in this thesis provided extensive experience in Monte-Carlo routines, statistical analysis, uncertainty quantification, and theoretical interpretation. I developed and maintained several computational frameworks, all structured for modularity and reproducibility. These are now being adapted to include resummation techniques and higher-order corrections at hadron colliders.

A related direction is the development of new PDF fitting strategies, especially in kinematic regions relevant for Run II and III of the LHC. This includes examining the interplay between PDF uncertainties and SMEFT parameter extractions. As shown in

Chapters 3 and 4, this interference is not negligible. I plan to contribute to fitting procedures that integrate new physics parametrizations and precision constraints more systematically. These goals will also involve resummation improvements, modified factorization theorems, and potentially new parameter bases.

My first steps in particle physics were in BSM scenarios. I studied rare processes involving the top quark and searched for signals of extra dimensional models. That early work pulled together ideas from many corners and taught me how to move between frameworks. Over the years, my focus shifted toward EW fits, gluon operators, and collider-specific SMEFT observables. Precision and collider physics became the foundation. Along the way, I learned how to build simulations, design computational workflows, and connect theoretical calculations with the kinds of observables collider experiments can access. The next steps will build on this foundation. Resummation for top physics, as well as and global PDF fits, will be part of the next chapter of my life. So will improved uncertainty treatment and closer integration between theoretical inputs and experimental constraints. In that sense, the circle closes. But maybe it was never a circle. Maybe it has always been a helix, moving forward.



## References

- [1] Electron-Ion Collider Achieves Critical Decision 1 Approval. <https://www.energy.gov/science/articles/electron-ion-collider-achieves-critical-decision-1-approval>.
- [2] Future Circular Collider. <https://home.cern/science/accelerators/future-circular-collider#:~:text=experiments,end%20of%20the%2021st%20century>.
- [3] The High-Luminosity Large Hadron Collider. <https://home.cern/press/2022/HL-LHC>.
- [4] The High-Luminosity LHC (HL-LHC). <https://voisins.web.cern.ch/en/high-luminosity-lhc-hl-lhc>.
- [5] AAD, G., ET AL. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* 716 (2012), 1–29.
- [6] AAD, G., ET AL. Measurement of the angular coefficients in  $Z$ -boson events using electron and muon pairs from data taken at  $\sqrt{s} = 8$  TeV with the ATLAS detector. *JHEP* 08 (2016), 159.
- [7] AAD, G., ET AL. Search for the Higgs boson decays  $H \rightarrow ee$  and  $H \rightarrow e\mu$  in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Lett. B* 801 (2020), 135148.
- [8] ABADA, A., ET AL. FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1. *Eur. Phys. J. C* 79, 6 (2019), 474.
- [9] ABDUL KHALEK, R., ET AL. Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report. *Nucl. Phys. A* 1026 (2022), 122447.
- [10] ABDUL KHALEK, R., ET AL. Snowmass 2021 White Paper: Electron Ion Collider for High Energy Physics.

- [11] ABELLEIRA FERNANDEZ, J. L., ET AL. A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector. *J. Phys. G* 39 (2012), 075001.
- [12] ACCARDI, A., ET AL. Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all. *Eur. Phys. J. A* 52, 9 (2016), 268.
- [13] ALBRECHT, A., ET AL. Findings of the Joint Dark Energy Mission Figure of Merit Science Working Group.
- [14] ALCARAZ MAESTRE, J. Details on the Collins-Soper reference frame and lepton angular distributions in electroweak vector boson production at hadron colliders, 2020.
- [15] ALIOLI, S., BOUGHEZAL, R., MEREGHETTI, E., AND PETRIELLO, F. Novel angular dependence in Drell-Yan lepton production via dimension-8 operators. *Phys. Lett. B* 809 (2020), 135703.
- [16] ALIOLI, S., CIRIGLIANO, V., DEKENS, W., DE VRIES, J., AND MEREGHETTI, E. Right-handed charged currents in the era of the Large Hadron Collider. *JHEP* 05 (2017), 086.
- [17] ALIOLI, S., DEKENS, W., GIRARD, M., AND MEREGHETTI, E. NLO QCD corrections to SM-EFT dilepton and electroweak Higgs boson production, matched to parton shower in POWHEG. *JHEP* 08 (2018), 205.
- [18] ALIOLI, S., FARINA, M., PAPPADOPULO, D., AND RUDERMAN, J. T. Catching a New Force by the Tail. *Phys. Rev. Lett.* 120, 10 (2018), 101801.
- [19] ALIPOUR TEHRANI, N., ET AL. FCC-ee: Your Questions Answered. In *CERN Council Open Symposium on the Update of European Strategy for Particle Physics* (6 2019), A. Blondel and P. Janot, Eds.
- [20] ALONSO, R., JENKINS, E. E., MANOHAR, A. V., AND TROTT, M. Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology. *JHEP* 04 (2014), 159.
- [21] ALTARELLI, G., ELLIS, R. K., AND MARTINELLI, G. Large Perturbative Corrections to the Drell-Yan Process in QCD. *Nucl. Phys. B* 157 (1979), 461–497.
- [22] ALTARELLI, G., ELLIS, R. K., MARTINELLI, G., AND PI, S.-Y. Processes Involving Fragmentation Functions Beyond the Leading Order in QCD. *Nucl. Phys. B* 160 (1979), 301–329.

- [23] ALTE, S., KÖNIG, M., AND SHEPHERD, W. Consistent Searches for SMEFT Effects in Non-Resonant Dilepton Events. *JHEP* 07 (2019), 144.
- [24] ALWALL, J., FREDERIX, R., FRIXIONE, S., HIRSCHI, V., MALTONI, F., MATTELAER, O., SHAO, H. S., STELZER, T., TORRIELLI, P., AND ZARO, M. 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.
- [25] AOUBE, R., HURTH, T., RENNER, S., AND SHEPHERD, W. The impact of flavour data on global fits of the MFV SMEFT. *JHEP* 12 (2020), 113.
- [26] APPELQUIST, T., AND CARAZZONE, J. Infrared Singularities and Massive Fields. *Phys. Rev. D* 11 (1975), 2856.
- [27] ARZT, C. Reduced effective Lagrangians. *Phys. Lett. B* 342 (1995), 189–195.
- [28] ARZT, C., EINHORN, M. B., AND WUDKA, J. Patterns of deviation from the standard model. *Nucl. Phys. B* 433 (1995), 41–66.
- [29] BAGLIO, J., DAWSON, S., HOMILLER, S., LANE, S. D., AND LEWIS, I. M. Validity of standard model EFT studies of VH and VV production at NLO. *Phys. Rev. D* 101, 11 (2020), 115004.
- [30] BAGNASCHI, E., ELLIS, J., MADIGAN, M., MIMASU, K., SANZ, V., AND YOU, T. SMEFT analysis of  $m_W$ . *JHEP* 08 (2022), 308.
- [31] BAKER, M. J., FUENTES-MARTÍN, J., ISIDORI, G., AND KÖNIG, M. High- $p_T$  signatures in vector–leptoquark models. *Eur. Phys. J. C* 79, 4 (2019), 334.
- [32] BALL, R. D., ET AL. Parton distributions from high-precision collider data. *Eur. Phys. J. C* 77, 10 (2017), 663.
- [33] BARTOCCI, R., BIEKÖTTER, A., AND HURTH, T. A global analysis of the SMEFT under the minimal MFV assumption. *JHEP* 05 (2024), 074.
- [34] BENEDIKT, M., ET AL. Future Circular Collider Feasibility Study Report: Volume 1, Physics, Experiments, Detectors.
- [35] BENEDIKT, M., ET AL. Future Circular Collider Feasibility Study Report: Volume 2, Accelerators, Technical Infrastructure and Safety.
- [36] BENEDIKT, M., ET AL. Future Circular Collider Feasibility Study Report: Volume 3, Civil Engineering, Implementation and Sustainability.

- [37] BIEKÖTTER, A., CORBETT, T., AND PLEHN, T. The Gauge-Higgs Legacy of the LHC Run II. *SciPost Phys.* 6, 6 (2019), 064.
- [38] BINOSI, D., AND THEUSSL, L. JaxoDraw: A Graphical user interface for drawing Feynman diagrams. *Comput. Phys. Commun.* 161 (2004), 76–86.
- [39] BISWAS, T. Cornering the SMEFT Interactions of Higgs Through Higgs-Photon Production at the LHC. *Springer Proc. Phys.* 304 (2024), 1040–1042.
- [40] BLONDEL, A. Beam polarization (longitudinal vs transverse) and energy calibration requirements. presented at the 3rd Annual Meeting of the Future Circular Collider Study, Berlin, Germany, <https://indico.cern.ch/event/556692/>.
- [41] BLUMLEIN, J., BOOS, E., AND KRYUKOV, A. Leptoquark pair production in hadronic interactions. *Z. Phys. C* 76 (1997), 137–153.
- [42] BLUMLEIN, J., AND RUCKL, R. Production of scalar and vector leptoquarks in  $e^+e^-$  annihilation. *Phys. Lett. B* 304 (1993), 337–346.
- [43] BORODULIN, V. I., ROGALYOV, R. N., AND SLABOSPITSKII, S. R. CORE 3.2 (Compendium of RElations, Version 3.2).
- [44] BOUGHEZAL, R., CHEN, C.-Y., PETRIELLO, F., AND WIEGAND, D. Four-lepton  $Z$  boson decay constraints on the standard model EFT. *Phys. Rev. D* 103, 5 (2021), 055015.
- [45] BOUGHEZAL, R., HUANG, Y., AND PETRIELLO, F. Exploring the SMEFT at dimension eight with Drell-Yan transverse momentum measurements. *Phys. Rev. D* 106, 3 (2022), 036020.
- [46] BOUGHEZAL, R., MEREGHETTI, E., AND PETRIELLO, F. Dilepton production in the SMEFT at  $\mathcal{O}(1/\Lambda^4)$ . *Phys. Rev. D* 104, 9 (2021), 095022.
- [47] BOUGHEZAL, R., PETRIELLO, F., AND WIEGAND, D. Removing flat directions in standard model EFT fits: How polarized electron-ion collider data can complement the LHC. *Phys. Rev. D* 101, 11 (2020), 116002.
- [48] BOUGHEZAL, R., PETRIELLO, F., AND WIEGAND, D. Disentangling Standard Model EFT operators with future low-energy parity-violating electron scattering experiments. *Phys. Rev. D* 104, 1 (2021), 016005.
- [49] BRESÓ-PLA, V., FALKOWSKI, A., AND GONZÁLEZ-ALONSO, M.  $A_{FB}$  in the SMEFT: precision  $Z$  physics at the LHC. *JHEP* 08 (2021), 021.

- [50] BRITZGER, D., KLEIN, M., AND SPIESBERGER, H. Electroweak physics in inclusive deep inelastic scattering at the LHeC. *Eur. Phys. J. C* 80, 9 (2020), 831.
- [51] BRITZGER, D., KLEIN, M., AND SPIESBERGER, H. Precision electroweak measurements at the LHeC and the FCC-eh. *PoS EPS-HEP2021* (2022), 485.
- [52] BRIVIO, I. SMEFTsim 3.0 — a practical guide. *JHEP* 04 (2021), 073.
- [53] BRIVIO, I., BRUGGESSER, S., MALTONI, F., MOUTAFIS, R., PLEHN, T., VRYONIDOU, E., WESTHOFF, S., AND ZHANG, C. O new physics, where art thou? A global search in the top sector. *JHEP* 02 (2020), 131.
- [54] BRIVIO, I., JIANG, Y., AND TROTT, M. The SMEFTsim package, theory and tools. *JHEP* 12 (2017), 070.
- [55] BRIVIO, I., AND TROTT, M. The Standard Model as an Effective Field Theory. *Phys. Rept.* 793 (2019), 1–98.
- [56] BRÜNING, O., SERYI, A., AND VERDÚ-ANDRÉS, S. Electron-Hadron Colliders: EIC, LHeC and FCC-eh. *Front. in Phys.* 10 (2022), 886473.
- [57] BRÜNING, O., AND KLEIN, M. *LHeC: A TeV energy scale lepton-hadron collider using the LHC infrastructure*. ch. Chapter 28, pp. 549–560.
- [58] BUCHMULLER, W., AND WYLER, D. Effective Lagrangian Analysis of New Interactions and Flavor Conservation. *Nucl. Phys. B* 268 (1986), 621–653.
- [59] BUCKLEY, A., FERRANDO, J., LLOYD, S., NORDSTRÖM, K., PAGE, B., RÜFENACHT, M., SCHÖNHERR, M., AND WATT, G. LHAPDF6: parton density access in the LHC precision era. *Eur. Phys. J. C* 75 (2015), 132.
- [60] CAMPBELL, J., AND NEUMANN, T. Precision Phenomenology with MCFM. *JHEP* 12 (2019), 034.
- [61] CARENA, M., DALEO, A., DOBRESCU, B. A., AND TAIT, T. M. P.  $Z'$  gauge bosons at the Tevatron. *Phys. Rev. D* 70 (2004), 093009.
- [62] CARRAZZA, S., DEGRANDE, C., IRANIPOUR, S., ROJO, J., AND UBIALI, M. Can New Physics hide inside the proton? *Phys. Rev. Lett.* 123, 13 (2019), 132001.

- [63] CELADA, E., GIANI, T., TER HOEVE, J., MANTANI, L., ROJO, J., ROSSIA, A. N., THOMAS, M. O. A., AND VRYONIDOU, E. Mapping the SMEFT at high-energy colliders: from LEP and the (HL-)LHC to the FCC-ee. *JHEP 09* (2024), 091.
- [64] CEPEDA, M., ET AL. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr. 7* (2019), 221–584.
- [65] CHATRCHYAN, S., ET AL. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B 716* (2012), 30–61.
- [66] CHEN, C.-Y., DAWSON, S., AND ZHANG, C. Electroweak Effective Operators and Higgs Physics. *Phys. Rev. D 89*, 1 (2014), 015016.
- [67] CIRIGLIANO, V., DEKENS, W., DE VRIES, J., AND MEREGHETTI, E. Constraining the top-Higgs sector of the Standard Model Effective Field Theory. *Phys. Rev. D 94*, 3 (2016), 034031.
- [68] CIRIGLIANO, V., GONZALEZ-ALONSO, M., AND GRAESSER, M. L. Non-standard Charged Current Interactions: beta decays versus the LHC. *JHEP 02* (2013), 046.
- [69] CLARK, D. B., GODAT, E., AND OLNES, F. I. ManeParse : A Mathematica reader for Parton Distribution Functions. *Comput. Phys. Commun. 216* (2017), 126–137.
- [70] COHEN, T., LU, X., AND ZHANG, Z. STReAMlining EFT Matching. *SciPost Phys. 10*, 5 (2021), 098.
- [71] COLEMAN, S. R., WESS, J., AND ZUMINO, B. Structure of phenomenological Lagrangians. 1. *Phys. Rev. 177* (1969), 2239–2247.
- [72] COLLINS, J. *Foundations of Perturbative QCD*, vol. 32. Cambridge University Press, 2011.
- [73] COLLINS, J. C., AND SOPER, D. E. Angular Distribution of Dileptons in High-Energy Hadron Collisions. *Phys. Rev. D 16* (1977), 2219.
- [74] DA SILVA ALMEIDA, E., ALVES, A., ROSA AGOSTINHO, N., ÉBOLI, O. J. P., AND GONZALEZ-GARCIA, M. C. Electroweak Sector Under Scrutiny: A Combined Analysis of LHC and Electroweak Precision Data. *Phys. Rev. D 99*, 3 (2019), 033001.
- [75] DAWSON, S., FONTES, D., HOMILLER, S., AND SULLIVAN, M. Role of dimension-eight operators in an EFT for the 2HDM. *Phys. Rev. D 106*, 5 (2022), 055012.

- [76] DAWSON, S., AND GIARDINO, P. P. Electroweak and QCD corrections to  $Z$  and  $W$  pole observables in the standard model EFT. *Phys. Rev. D* **101**, 1 (2020), 013001.
- [77] DAWSON, S., GIARDINO, P. P., AND ISMAIL, A. Standard model EFT and the Drell-Yan process at high energy. *Phys. Rev. D* **99**, 3 (2019), 035044.
- [78] DAWSON, S., HOMILLER, S., AND LANE, S. D. Putting standard model EFT fits to work. *Phys. Rev. D* **102**, 5 (2020), 055012.
- [79] DE BLAS, J., CIUCHINI, M., FRANCO, E., MISHIMA, S., PIERINI, M., REINA, L., AND SILVESTRINI, L. Electroweak precision observables and Higgs-boson signal strengths in the Standard Model and beyond: present and future. *JHEP* **12** (2016), 135.
- [80] DE BLAS, J., DU, Y., GROJEAN, C., GU, J., MIRALLES, V., PESKIN, M. E., TIAN, J., VOS, M., AND VRYONIDOU, E. Global SMEFT Fits at Future Colliders. In *Snowmass 2021* (6 2022).
- [81] DE BLAS, J., GU, J., AND LIU, Z. Higgs boson precision measurements at a 125 GeV muon collider. *Phys. Rev. D* **106**, 7 (2022), 073007.
- [82] DE FLORIAN, D., AND ROTSTEIN HABARNAU, Y. Polarized semi-inclusive electroweak structure functions at next-to-leading-order. *Eur. Phys. J. C* **73**, 3 (2013), 2356.
- [83] DE FLORIAN, D., AND SASSOT, R. O ( $\alpha$ -s) spin dependent weak structure functions. *Phys. Rev. D* **51** (1995), 6052–6058.
- [84] DE RHAM, C., KUNDU, S., REECE, M., TOLLEY, A. J., AND ZHOU, S.-Y. Snowmass White Paper: UV Constraints on IR Physics. In *Snowmass 2021* (3 2022).
- [85] DEDES, A., MATERKOWSKA, W., PARASKEVAS, M., ROSIEK, J., AND SUXHO, K. Feynman rules for the Standard Model Effective Field Theory in  $R_\xi$ -gauges. *JHEP* **06** (2017), 143.
- [86] DEDES, A., PARASKEVAS, M., ROSIEK, J., SUXHO, K., AND TRIFYLLIS, L. SmeftFR – Feynman rules generator for the Standard Model Effective Field Theory. *Comput. Phys. Commun.* **247** (2020), 106931.
- [87] DEDES, A., ROSIEK, J., RYCKZKOWSKI, M., SUXHO, K., AND TRIFYLLIS, L. SmeftFR v3 – Feynman rules generator for the Standard Model Effective Field Theory. *Comput. Phys. Commun.* **294** (2024), 108943.

- [88] D'ENTERRIA, D., POLDARU, A., AND WOJCIK, G. Measuring the electron Yukawa coupling via resonant s-channel Higgs production at FCC-ee. *Eur. Phys. J. Plus* *137*, 2 (2022), 201.
- [89] DERY, A., FRUGIUELE, C., AND NIR, Y. Large Higgs-electron Yukawa coupling in 2HDM. *JHEP* *04* (2018), 044.
- [90] DI VITA, S., GROJEAN, C., PANICO, G., RIEMBAU, M., AND VANTALON, T. A global view on the Higgs self-coupling. *JHEP* *09* (2017), 069.
- [91] ELIAS-MIRO, J., ESPINOSA, J. R., MASSO, E., AND POMAROL, A. Higgs windows to new physics through d=6 operators: constraints and one-loop anomalous dimensions. *JHEP* *11* (2013), 066.
- [92] ELLIS, J., MADIGAN, M., MIMASU, K., SANZ, V., AND YOU, T. Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory. *JHEP* *04* (2021), 279.
- [93] ELLIS, J., MURPHY, C. W., SANZ, V., AND YOU, T. Updated Global SMEFT Fit to Higgs, Diboson and Electroweak Data. *JHEP* *06* (2018), 146.
- [94] ELLIS, J., SANZ, V., AND YOU, T. Complete Higgs Sector Constraints on Dimension-6 Operators. *JHEP* *07* (2014), 036.
- [95] ELLIS, J., SANZ, V., AND YOU, T. The Effective Standard Model after LHC Run I. *JHEP* *03* (2015), 157.
- [96] ELLIS, R. K., STIRLING, W. J., AND WEBBER, B. R. *QCD and collider physics*, vol. 8. Cambridge University Press, 2 2011.
- [97] ETHIER, J. J., GOMEZ-AMBROSIO, R., MAGNI, G., AND ROJO, J. SMEFT analysis of vector boson scattering and diboson data from the LHC Run II. *Eur. Phys. J. C* *81*, 6 (2021), 560.
- [98] ETHIER, J. J., MAGNI, G., MALTONI, F., MANTANI, L., NOCERA, E. R., ROJO, J., SLADE, E., VRYONIDOU, E., AND ZHANG, C. Combined SMEFT interpretation of Higgs, diboson, and top quark data from the LHC. *JHEP* *11* (2021), 089.
- [99] FALKOWSKI, A. Effective field theory approach to LHC Higgs data. *Pramana* *87*, 3 (2016), 39.



- [100] FALKOWSKI, A., GONZÁLEZ-ALONSO, M., AND MIMOUNI, K. Compilation of low-energy constraints on 4-fermion operators in the SMEFT. *JHEP* 08 (2017), 123.
- [101] FALKOWSKI, A., AND RIVA, F. Model-independent precision constraints on dimension-6 operators. *JHEP* 02 (2015), 039.
- [102] GAO, J., GAO, M., HOBBS, T. J., LIU, D., AND SHEN, X. Simultaneous CTEQ-TEA extraction of PDFs and SMEFT parameters from jet and  $t\bar{t}$  data. *JHEP* 05 (2023), 003.
- [103] GIUDICE, G. F., GROJEAN, C., POMAROL, A., AND RATTAZZI, R. The Strongly-Interacting Light Higgs. *JHEP* 06 (2007), 045.
- [104] GRAF, L., HENNING, B., LU, X., MELIA, T., AND MURAYAMA, H. 2, 12, 117, 1959, 45171, 1170086, ...: a Hilbert series for the QCD chiral Lagrangian. *JHEP* 01 (2021), 142.
- [105] GRÁF, L., HENNING, B., LU, X., MELIA, T., AND MURAYAMA, H. Hilbert series, the Higgs mechanism, and HEFT. *JHEP* 02 (2023), 064.
- [106] GRECO, M., HAN, T., AND LIU, Z. ISR effects for resonant Higgs production at future lepton colliders. *Phys. Lett. B* 763 (2016), 409–415.
- [107] GRELJO, A., IRANIPOUR, S., KASSABOV, Z., MADIGAN, M., MOORE, J., ROJO, J., UBIALI, M., AND VOISEY, C. Parton distributions in the SMEFT from high-energy Drell-Yan tails. *JHEP* 07 (2021), 122.
- [108] GRELJO, A., TIBLOM, H., AND VALENTI, A. New Physics Through Flavor Tagging at FCC-ee.
- [109] GROJEAN, C., MONTULL, M., AND RIEMBAU, M. Diboson at the LHC vs LEP. *JHEP* 03 (2019), 020.
- [110] GRZADKOWSKI, B., ISKRZYNSKI, M., MISIAK, M., AND ROSIEK, J. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP* 10 (2010), 085.
- [111] HAHN, T. Generating Feynman diagrams and amplitudes with FeynArts 3. *Comput. Phys. Commun.* 140 (2001), 418–431.
- [112] HAN, Z., AND SKIBA, W. Effective theory analysis of precision electroweak data. *Phys. Rev. D* 71 (2005), 075009.

- [113] HARLANDER, R. V., KEMPKENS, T., AND SCHAAF, M. C. Standard model effective field theory up to mass dimension 12. *Phys. Rev. D* *108*, 5 (2023), 055020.
- [114] HARTLAND, N. P., MALTONI, F., NOCERA, E. R., ROJO, J., SLADE, E., VRYONIDOU, E., AND ZHANG, C. A Monte Carlo global analysis of the Standard Model Effective Field Theory: the top quark sector. *JHEP* *04* (2019), 100.
- [115] HARTMANN, C., SHEPHERD, W., AND TROTT, M. The  $Z$  decay width in the SMEFT:  $y_t$  and  $\lambda$  corrections at one loop. *JHEP* *03* (2017), 060.
- [116] HELSENS, C. Top physics at Future Circular Colliders. *PoS ICHEP2018* (2019), 660.
- [117] HENNING, B., LU, X., MELIA, T., AND MURAYAMA, H. Hilbert series and operator bases with derivatives in effective field theories. *Commun. Math. Phys.* *347*, 2 (2016), 363–388.
- [118] HENNING, B., LU, X., MELIA, T., AND MURAYAMA, H. 2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT. *JHEP* *08* (2017), 016. [Erratum: *JHEP* *09*, 019 (2019)].
- [119] HORNE, A., PITTMAN, J., SNEDEKER, M., SHEPHERD, W., AND WALKER, J. W. Shift-Type SMEFT Effects in Dileptons at the LHC. *JHEP* *03* (2021), 118.
- [120] JADACH, S., AND KYCIA, R. A. Lineshape of the Higgs boson in future lepton colliders. *Phys. Lett. B* *755* (2016), 58–63.
- [121] JADACH, S., WARD, B. F. L., AND WAS, Z. Coherent exclusive exponentiation for precision Monte Carlo calculations. *Phys. Rev. D* *63* (2001), 113009.
- [122] JENKINS, E. E., MANOHAR, A. V., AND TROTT, M. Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence. *JHEP* *10* (2013), 087.
- [123] JENKINS, E. E., MANOHAR, A. V., AND TROTT, M. Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence. *JHEP* *01* (2014), 035.
- [124] KHACHATRYAN, V., ET AL. Search for a standard model-like Higgs boson in the  $\mu^+\mu^-$  and  $e^+e^-$  decay channels at the LHC. *Phys. Lett. B* *744* (2015), 184–207.
- [125] KIM, T., AND MARTIN, A. Monolepton production in SMEFT to  $\mathcal{O}(1/\Lambda^4)$  and beyond. *JHEP* *09* (2022), 124.

- [126] KLEIN, M., AND RADESCU, V. CERN-LHeC-Note-2013-002PHY. <https://cds.cern.ch/record/1558451>, 2013. Technical Report.
- [127] KLEIN, M., AND YOSHIDA, R. Collider Physics at HERA. *Prog. Part. Nucl. Phys.* *61* (2008), 343–393.
- [128] KRAUSS, F. Quarks and Leptons Lecture Notes: Phase Space.
- [129] KUMAR, P., AND GOMBER, B. The CMS Level-1 Calorimeter Trigger for the HL-LHC. *Instruments* *6*, 4 (2022), 64.
- [130] KYCIA, R. A., AND JADACH, S. Relativistic Voigt profile for unstable particles in high energy physics. *J. Math. Anal. Appl.* *463*, 2 (2018), 1040–1051.
- [131] LAM, C. S., AND TUNG, W.-K. A Systematic Approach to Inclusive Lepton Pair Production in Hadronic Collisions. *Phys. Rev. D* *18* (1978), 2447.
- [132] LEPAGE, G. P. gplepage/vegas: vegas version 6.1.2. Zenodo, May 21, 2024. DOI: 10.5281/zenodo.11238512.
- [133] LEPAGE, G. P. Adaptive multidimensional integration: VEGAS enhanced. *J. Comput. Phys.* *439* (2021), 110386.
- [134] LI, H.-L., REN, Z., SHU, J., XIAO, M.-L., YU, J.-H., AND ZHENG, Y.-H. Complete set of dimension-eight operators in the standard model effective field theory. *Phys. Rev. D* *104*, 1 (2021), 015026.
- [135] LI, X., MIMASU, K., YAMASHITA, K., YANG, C., ZHANG, C., AND ZHOU, S.-Y. Moments for positivity: using Drell-Yan data to test positivity bounds and reverse-engineer new physics. *JHEP* *10* (2022), 107.
- [136] LYONS, L. *Statistics for Nuclear and Particle Physicists*. Cambridge University Press, 1986.
- [137] MAGNI, G., AND GOMEZ-AMBROSIO, R. SMEFT analysis of the electroweak sector: challenges beyond dimension 6. *PoS EPS-HEP2021* (2022), 475.
- [138] MERTIG, R., BOHM, M., AND DENNER, A. FEYN CALC: Computer algebraic calculation of Feynman amplitudes. *Comput. Phys. Commun.* *64* (1991), 345–359.
- [139] METZ, A., SCHLEGEL, M., AND GOEKE, K. Transverse single spin asymmetries in inclusive deep-inelastic scattering. *Phys. Lett. B* *643* (2006), 319–324.

- [140] MURPHY, C. W. Dimension-8 operators in the Standard Model Effective Field Theory. *JHEP* 10 (2020), 174.
- [141] NOCERA, E. R., BALL, R. D., FORTE, S., RIDOLFI, G., AND ROJO, J. A first unbiased global determination of polarized PDFs and their uncertainties. *Nucl. Phys. B* 887 (2014), 276–308.
- [142] PANICO, G., RICCI, L., AND WULZER, A. High-energy EFT probes with fully differential Drell-Yan measurements. *JHEP* 07 (2021), 086.
- [143] POMAROL, A., AND RIVA, F. Towards the Ultimate SM Fit to Close in on Higgs Physics. *JHEP* 01 (2014), 151.
- [144] SHTABOVENKO, V., MERTIG, R., AND ORELLANA, F. New Developments in FeynCalc 9.0. *Comput. Phys. Commun.* 207 (2016), 432–444.
- [145] SHTABOVENKO, V., MERTIG, R., AND ORELLANA, F. FeynCalc 9.3: New features and improvements. *Comput. Phys. Commun.* 256 (2020), 107478.
- [146] SHTABOVENKO, V., MERTIG, R., AND ORELLANA, F. FeynCalc 10: Do multiloop integrals dream of computer codes?
- [147] TORRE, R., RICCI, L., AND WULZER, A. On the W&Y interpretation of high-energy Drell-Yan measurements. *JHEP* 02 (2021), 144.
- [148] TUMASYAN, A., ET AL. Measurement of the Drell-Yan forward-backward asymmetry at high dilepton masses in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP* 2022, 08 (2022), 063.
- [149] TUMASYAN, A., ET AL. Measurement of the mass dependence of the transverse momentum of lepton pairs in Drell-Yan production in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C* 83, 7 (2023), 628.
- [150] VAN BEEK, S., NOCERA, E. R., ROJO, J., AND SLADE, E. Constraining the SMEFT with Bayesian reweighting. *SciPost Phys.* 7, 5 (2019), 070.
- [151] VOGELSANG, W. The Gluonic contribution to  $g(1)\text{-}p(x, Q^{*2})$  in the parton model. *Z. Phys. C* 50 (1991), 275–284.
- [152] WEINBERG, S. Phenomenological Lagrangians. *Physica A* 96, 1-2 (1979), 327–340.
- [153] WELLS, J. D., AND ZHANG, Z. Precision Electroweak Analysis after the Higgs Boson Discovery. *Phys. Rev. D* 90, 3 (2014), 033006.

- [154] WILLEKE, F. Design of the Electron–Ion Collider. *Acta Phys. Polon. Supp.* 16, 7 (2023), 7–A1.
- [155] WORKMAN, R. L., ET AL. Review of Particle Physics. *PTEP 2022* (2022), 083C01.
- [156] ZHANG, C., AND ZHOU, S.-Y. Convex Geometry Perspective on the (Standard Model) Effective Field Theory Space. *Phys. Rev. Lett.* 125, 20 (2020), 201601.
- [157] ZYLA, P. A., ET AL. Review of Particle Physics. *PTEP 2020*, 8 (2020), 083C01.