

Neutral-current electroweak physics and SMEFT studies at the EIC

Radja Boughezal,¹ Alexander Emmert², Tyler Kutz³, Sonny Mantry⁴, Michael Nycz,² Frank Petriello,^{1,5} Kağan Şimşek⁵, Daniel Wiegand,⁵ and Xiaochao Zheng²

¹Argonne National Laboratory, Lemont, Illinois 60439, USA

²University of Virginia, Charlottesville, Virginia 22904, USA

³Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

⁴University of North Georgia, Dahlonega, Georgia 30597, USA

⁵Northwestern University, Evanston, Illinois 60208, USA



(Received 28 April 2022; accepted 27 June 2022; published 13 July 2022)

We study the potential for precision electroweak (EW) measurements and beyond-the-Standard Model (BSM) searches using cross-section asymmetries in neutral-current deep inelastic scattering at the electron-ion collider (EIC). Our analysis uses a complete and realistic accounting of systematic errors from both theory and experiment and considers the potential of both proton and deuteron beams for a wide range of energies and luminosities. We also consider what can be learned from a possible future positron beam and a potential tenfold luminosity upgrade of the EIC beyond its initial decade of running. We use the SM effective field theory (SMEFT) framework to parametrize BSM effects and focus on semileptonic four-fermion operators, whereas for our precision EW study, we determine how well the EIC can measure the weak mixing angle. New features of our study include the use of an up-to-date detector design of the EIC Comprehensive Chromodynamics Experiment and accurate running conditions of the EIC, the simultaneous fitting of beam polarization uncertainties and Wilson coefficients to improve the sensitivity to SMEFT operators, and the inclusion of the weak mixing angle running in our fit template. We find that the EIC can probe BSM operators at scales competitive with and in many cases exceeding LHC Drell-Yan bounds while simultaneously not suffering from degeneracies between Wilson coefficients.

DOI: [10.1103/PhysRevD.106.016006](https://doi.org/10.1103/PhysRevD.106.016006)

I. INTRODUCTION

The Standard Model (SM) of particle physics currently describes all known laboratory phenomena. All particles predicted by the SM have now been found after the discovery of the Higgs boson at the Large Hadron Collider (LHC). No new particles beyond those present in the SM have been discovered and no appreciable deviation from SM predictions has been conclusively observed. Despite the enormous success of this theory, it contains numerous shortcomings. It does not contain an explanation of the dark matter observed in the universe or of the baryon-antibaryon asymmetry, and it does not describe neutrino masses. It additionally suffers from several aesthetic issues, such as the hierarchy problem and an extreme hierarchy of fermion Yukawa couplings. Even the sectors of the theory that have been experimentally successful still contain unsatisfying and poorly understood features. For example, the exact composition of the proton

spin in terms of the spin and orbital angular momentum of its constituent quarks and gluons is still poorly known.

Numerous experimental programs that attempt to address these residual issues in our understanding of nature are either running or being designed. Our focus in this manuscript will be on the Electron-Ion Collider (EIC) to be built at Brookhaven National Laboratory in Upton, New York. The EIC will be a particle accelerator that collides electrons with protons and nuclei in the intermediate-energy range between fixed-target scattering facilities and high-energy colliders. It will provide luminosity orders of magnitude higher than HERA, the only electron-proton collider operated to date. It will also be the first lepton-ion collider with the ability to polarize both the electron and the proton (ion) beams and the first collider with a fast spin-flip capacity. These unique design features will allow direct extraction of parity-violating (PV) asymmetries in the electroweak neutral-current (NC) scattering cross section associated with either the electron, $A_{\text{PV}}^{(e)}$, or the proton (ion) spin flip, $A_{\text{PV}}^{(p(D))}$. Experimental uncertainties from effects such as luminosity measurement and detector acceptance or efficiency will be substantially reduced due to these capabilities.

Although the EIC was designed primarily to explore outstanding issues in QCD such as the proton spin mentioned

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

above, it also has a strong potential to probe several aspects of precision electroweak (EW) and beyond-the-SM (BSM) physics. It can measure the value of the weak mixing angle over a wide range of momentum transfer complementary to Z -pole measurements and low-energy determinations. The possibility of polarizing both electron and proton/ion beams gives it unique handles on BSM physics. Our goal in this manuscript is to provide a detailed accounting of the EW and BSM potential of the EIC with a realistic simulation of anticipated experimental uncertainties. We explore the use of the asymmetries $A_{\text{PV}}^{(e)}$ and $A_{\text{PV}}^{(p(D))}$. In addition to determining the BSM reach of PV observables, we consider the reach of the lepton-charge asymmetry $A_{\text{LC}}^{(p(D))}$ at the EIC for the first time, assuming a positron beam will become available in the future.

Since no new particles beyond the SM have so far been discovered, we adopt the Standard Model effective field theory (SMEFT) for our BSM studies (for a review of the SMEFT, see Ref. [1]). The SMEFT contains higher-dimensional operators formed by using SM fields, assuming all new physics is heavier than both SM states and the accessible collider energy. The leading dimension-six operator basis of SMEFT for on-shell fields has been completely classified (there is a dimension-five operator that violates the lepton number, which we do not consider here) [2–4]. We find that the EIC can probe the full spectrum of SMEFT operators to the few-tera-electron-volt level or beyond. The wide variety of observables possible at the EIC, which include several asymmetries with either proton or ion beams, ensure that no flat directions remain in the Wilson coefficient parameter space, unlike at the LHC in the neutral-current Drell-Yan process [5–7]. Our analysis of the determination of the weak mixing angle, assuming a realistic annual luminosity and accounting for experimental and theoretical uncertainties to the best level that can be reached at the pre-EIC running stage, finds good precision for this fundamental SM parameter in a kinematic region not explored before. The precision will continue to improve as data are accumulated from decades-long running of the EIC.

Our paper is organized as follows. In Sec. II, first we provide a complete description of deep inelastic scattering (DIS) that includes both SM contributions and the SMEFT extensions. The DIS cross sections that account for both electron and hadron polarizations are provided in both structure-function and parton-model languages. We follow this theoretical framework by presenting a basic strategy to measure different polarization components of the cross sections and to form PV asymmetries at the EIC. Measurement of the lepton-charge (LC) asymmetry is also discussed. In Sec. III, we present data simulation based on the design of the EIC Comprehensive Chromodynamics Experiment (ECCE) detector (recently endorsed as the reference design for EIC detector 1 by the EIC Detector Proposal Advisory Panel [8]) using a fast-smearing method

and event-selection criteria, followed by projections of statistical precision for PV and LC asymmetries based on the planned annual luminosity of the EIC. The generation of pseudodata, as well as the uncertainty matrix, is presented in Sec. IV, followed by extractions of the EW mixing angle in Sec. V. In Sec. VI, we provide an extensive description of our SMEFT analysis framework, with representative results on the fits of single and two Wilson coefficients given in Sec. VII. We also show an example fit in which six Wilson coefficients are turned on simultaneously, in order to demonstrate that EIC data are capable of removing all degeneracies in the semileptonic four-fermion operator parameter space. We conclude in Sec. VIII. In Appendix A, we present novel analysis methods to simultaneously fit PV asymmetries and the beam polarization or LC asymmetries and the luminosity difference between e^+ and e^- runs. A complete collection of all the SMEFT fit results of single and two Wilson coefficients from this study is given in Appendix B.

II. NEUTRAL-CURRENT DIS MEASUREMENTS AT THE EIC

A. Deep inelastic scattering and the SMEFT formalism

In this section, we give a brief overview of the formalism of DIS and the SMEFT. In particular, we generalize the SM DIS cross-section and asymmetry formulas to include contributions from SMEFT operators, which encode new physics at an energy level Λ that lies well beyond the electroweak scale. We denote a lepton scattering off a nucleus as

$$\ell(k) + H(P) \rightarrow \ell(k') + X, \quad (1)$$

where ℓ stands for an electron or positron, the hadron H stands for either the proton (p) or the deuteron (D), and X denotes the final-state hadronic system. The four-momenta of the initial and final leptons and the initial hadron are denoted as k , k' , and P , respectively. Using the momenta of the initial- and final-state leptons and the initial-state hadron, one can define the following Lorentz-invariant kinematic variables:

$$s = (P + k)^2, \quad (2)$$

$$Q^2 = -(k - k')^2, \quad (3)$$

$$x = \frac{Q^2}{2P \cdot (k - k')}, \quad (4)$$

$$y = \frac{P \cdot (k - k')}{P \cdot k}, \quad (5)$$

$$W^2 = (P + k - k')^2, \quad (6)$$

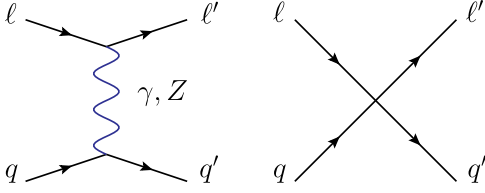


FIG. 1. The Feynman diagrams for $\ell + H \rightarrow \ell + X$ at the parton level from one-boson exchange (left) and SMEFT contact interactions (right).

where s is the center-of-mass energy squared, Q^2 is the negative of the lepton four-momentum transfer squared, the Bjorken- x variable is the longitudinal hadron momentum fraction carried by the struck parton, the inelasticity parameter y gives the fractional energy loss of the lepton in the hadron rest frame, and W gives the invariant mass of the final-state hadronic system X . The kinematic variables x , y , s , and Q^2 are related to each other via $Q^2 = xy(s - M^2)$, where M is the mass of the proton or deuteron.

The diagrams in Fig. 1 show the partonic tree-level processes that contribute to Eq. (1). These are the contributions to the total tree-level amplitude from the single-photon exchange, the single-Z-boson exchange, and the SMEFT contact interactions. The SMEFT Lagrangian that describes these contact interactions has the form

$$\mathcal{L}_{\text{SMEFT}} = \frac{1}{\Lambda^2} \sum_r C_r \mathcal{O}_r + \dots, \quad (7)$$

where the summation index r runs over the set of dimension-six SMEFT operators and the ellipsis denotes SMEFT operators of a mass dimension greater than 6. We restrict our analysis to include only the effects of dimension-six SMEFT operators since the higher-dimensional operators are formally suppressed by additional powers of E^2/Λ^2 , where E is the typical energy scale of the scattering process. Although these effects can be important for Drell-Yan production at the LHC [9,10], the low energy of the EIC renders them negligible in this analysis. \mathcal{O}_r denotes the r th dimension-six operator and C_r is the corresponding (dimensionless) Wilson coefficient arising from integrating out the new-physics degrees of freedom at the scale Λ . These Wilson coefficients can be constrained through a comparison of SMEFT predictions with precision measurements of various processes studied in a variety of experiments across a wide range of energy scales.

The subset of dimension-six operators that we consider in our analysis of DIS is given in Table I. We note that there are additional SMEFT operators but they are known to be far better bounded through other datasets such as precision Z-pole observables [11–13], and we neglect them here. The above assumptions leave us with the seven Wilson coefficients associated with the listed operators that enter the predictions for DIS cross sections and asymmetries.

As seen in Table I, the SMEFT operators \mathcal{O}_r are expressed in terms of the basis of SM fields before electroweak symmetry breaking. For the purposes of DIS phenomenology below the electroweak scale, it is useful to rewrite these SMEFT operators in the vector and axial-vector basis using Dirac fields that describe the massive electrons (e) and quarks (q_f) after electroweak symmetry breaking, which is a customary parametrization (see, e.g., [14]):

$$\mathcal{L}_{\text{SMEFT}} = \frac{1}{\Lambda^2} \sum_r \tilde{C}_r \left\{ \sum_f \bar{e} \gamma^\mu (c_{V_r}^e - c_{A_r}^e \gamma_5) e \bar{q}_f \gamma^\mu \times (c_{V_r}^f - c_{A_r}^f \gamma_5) q_f \right\} + \dots, \quad (8)$$

where the specific values of the vector and axial-vector couplings— $c_{V_k}^{e,q}$ and $c_{A_k}^{e,q}$, respectively—for the r th SMEFT operator follow from the corresponding chiral and flavor structure of the SMEFT operators. The coefficients \tilde{C}_r are related to the C_r by an overall factor and can be fixed by comparing Eqs. (7) and (8). There is freedom to always redefine the \tilde{C}_r by absorbing an overall factor into the couplings $c_{V_r}^{e,q}$ and $c_{A_r}^{e,q}$. We specify in Table I the exact definitions that we use. These couplings are analogous to the vector and axial-vector couplings, $g_{V_k}^{e,q}$ and $g_{A_k}^{e,q}$, of the Z-boson but are instead generated from integrating out UV physics associated with the scale Λ .

As seen in Fig. 1, the total tree-level amplitude can be decomposed into three contributions:

$$\mathcal{M} = \mathcal{M}_\gamma + \mathcal{M}_Z + \mathcal{M}_r, \quad (9)$$

where \mathcal{M}_γ , \mathcal{M}_Z , and \mathcal{M}_r denote the contributions from the single-photon exchange, the single-Z-boson exchange, and the SMEFT operators, respectively. In particular, $\mathcal{M}_r = \sum_i \mathcal{M}_i$, where the summation index i runs over the amplitudes arising from the SMEFT operators listed in Table I. Up to leading order in the SMEFT power counting, where only dimension-six SMEFT operators that scale as $1/\Lambda^2$ are kept, the total amplitude squared can be written as

$$|\mathcal{M}|^2 = \mathcal{M}_{\gamma\gamma} + 2\mathcal{M}_{\gamma Z} + \mathcal{M}_{ZZ} + 2\mathcal{M}_{\gamma r} + 2\mathcal{M}_{Zr}, \quad (10)$$

where $\mathcal{M}_{\gamma\gamma} = |\mathcal{M}_\gamma|^2$, $\mathcal{M}_{ZZ} = |\mathcal{M}_Z|^2$, $2\mathcal{M}_{\gamma Z} = \mathcal{M}_\gamma^* \mathcal{M}_Z + \mathcal{M}_\gamma \mathcal{M}_Z^*$, $2\mathcal{M}_{\gamma r} = \mathcal{M}_\gamma^* \mathcal{M}_r + \mathcal{M}_\gamma \mathcal{M}_r^*$, and $2\mathcal{M}_{Zr} = \mathcal{M}_Z^* \mathcal{M}_r + \mathcal{M}_Z \mathcal{M}_r^*$. These denote the amplitudes of the single-photon exchange, single-Z-boson exchange, the interference between the single-photon and the single-Z-boson exchange, the interference between the single-photon exchange and the SMEFT, and the interference between the single-Z-boson exchange and the SMEFT, respectively. Here, we ignore the $|\mathcal{M}_r|^2$ contribution since it scales as $1/\Lambda^4$, formally the same size as contributions from dimension-eight SMEFT operators interfering with the SM.

TABLE I. List of SMEFT operators relevant to DIS in the basis of SM fields before electroweak symmetry breaking and reexpressed in the vector and axial-vector current basis after electroweak symmetry breaking: $C_r \mathcal{O}_r = \tilde{C}_r \sum_f \bar{e} \gamma^\mu (c_{V_r}^e - c_{A_r}^e \gamma_5) e \bar{q}_f \gamma^\mu (c_{V_r}^f - c_{A_r}^f \gamma_5) q_f + \dots$. The coefficients $c_{V,A}^f$ give the chiral structure of each operator.

C_r	\mathcal{O}_r	\tilde{C}_r	$c_{V_r}^e$	$c_{A_r}^e$	$c_{V_r}^u$	$c_{A_r}^u$	$c_{V_r}^{d,s}$	$c_{A_r}^{d,s}$
$C_{\ell q}^{(1)}$	$\mathcal{O}_{\ell q}^{(1)} = (\bar{L}_L \gamma^\mu L_L)(\bar{Q}_L \gamma_\mu Q_L)$	$C_{\ell q}^{(1)}/4$	1	1	1	1	1	1
$C_{\ell q}^{(3)}$	$\mathcal{O}_{\ell q}^{(3)} = (\bar{L}_L \gamma^\mu \tau^I L_L)(\bar{Q}_L \gamma_\mu \tau^I Q_L)$	$C_{\ell q}^{(3)}/4$	1	1	-1	-1	1	1
C_{eu}	$\mathcal{O}_{eu} = (\bar{e}_R \gamma^\mu e_R)(\bar{u}_R \gamma_\mu u_R)$	$C_{eu}/4$	1	-1	1	-1	0	0
C_{ed}	$\mathcal{O}_{ed} = (\bar{e}_R \gamma^\mu e_R)(\bar{d}_R \gamma_\mu d_R)$	$C_{ed}/4$	1	-1	0	0	1	-1
$C_{\ell u}$	$\mathcal{O}_{\ell u} = (\bar{L}_L \gamma^\mu L_L)(\bar{u}_R \gamma_\mu u_R)$	$C_{\ell u}/4$	1	1	1	-1	0	0
$C_{\ell d}$	$\mathcal{O}_{\ell d} = (\bar{L}_L \gamma^\mu L_L)(\bar{d}_R \gamma_\mu d_R)$	$C_{\ell d}/4$	1	1	0	0	1	-1
C_{qe}	$\mathcal{O}_{qe} = (\bar{Q}_L \gamma^\mu Q_L)(\bar{e}_R \gamma_\mu e_R)$	$C_{qe}/4$	1	-1	1	1	1	1

For the hadron-level cross sections and asymmetries, these different contributions will give rise to corresponding structure functions. In particular, in addition to the usual structure functions encountered in the SM DIS, new structure functions corresponding to SMEFT contributions arise. Thus, including the SMEFT contributions, the DIS differential cross section takes the general form

$$\frac{d^2\sigma}{dx dy} = \frac{2\pi y \alpha^2}{Q^4} \left\{ \eta^\gamma L_{\mu\nu}^\gamma W_{\gamma}^{\mu\nu} + \eta^{\gamma Z} L_{\mu\nu}^{\gamma Z} W_{\gamma Z}^{\mu\nu} + \eta^Z L_{\mu\nu}^Z W_Z^{\mu\nu} + \sum_r \xi^{\gamma r} L_{\mu\nu}^{\gamma r} W_{\gamma r}^{\mu\nu} + \sum_r \xi^{Zr} L_{\mu\nu}^{Zr} W_{Zr}^{\mu\nu} \right\}, \quad (11)$$

where α is the electromagnetic fine structure constant and $L_{\mu\nu}^{\gamma, \gamma Z, Z, \gamma r, Zr}$ and $W_{\mu\nu}^{\gamma, \gamma Z, Z, \gamma r, Zr}$ are the leptonic and hadronic tensors, respectively. The first three terms on the right-hand side (RHS) correspond to the SM contributions from $\mathcal{M}_{\gamma\gamma}$, $2\mathcal{M}_{\gamma Z}$, and \mathcal{M}_{ZZ} , respectively, and the last two sets of terms correspond to the contributions from the SMEFT operators, i.e., $2\mathcal{M}_{\gamma r}$ and $2\mathcal{M}_{Zr}$, respectively. For completeness, below we collect some useful results to make the form of the cross section explicit. The dimensionless coefficients $\eta^{\gamma, \gamma Z, Z}$, $\xi^{\gamma r}$, and ξ^{Zr} are given by

$$\begin{aligned} \eta^\gamma &= 1, \\ \eta^{\gamma Z} &= \frac{G_F M_Z^2}{2\sqrt{2}\pi\alpha} \frac{Q^2}{Q^2 + M_Z^2}, \\ \eta^Z &= (\eta^{\gamma Z})^2, \\ \xi^{\gamma r} &= \frac{\tilde{C}_r}{4\pi\alpha} \frac{Q^2}{\Lambda^2}, \\ \xi^{Zr} &= \eta^{\gamma Z} \frac{\tilde{C}_r}{4\pi\alpha} \frac{Q^2}{\Lambda^2}, \end{aligned} \quad (12)$$

where $G_F = 1.1663787(6) \times 10^{-5} \text{ GeV}^{-2}$ is the Fermi constant and $M_Z = 91.1876 \pm 0.0021 \text{ GeV}$ [14] is the mass of the Z boson. The leptonic tensors in Eq. (11) are

$$\begin{aligned} L_{\mu\nu}^\gamma &= 2[k_\mu k'_\nu + k'_\mu k_\nu - k \cdot k' g_{\mu\nu} - i\lambda_e \epsilon_{\mu\nu\alpha\beta} k^\alpha (k')^\beta], \\ L_{\mu\nu}^{\gamma Z} &= -(g_V^e - \lambda_e g_A^e) L_{\mu\nu}^\gamma, \\ L_{\mu\nu}^Z &= (g_V^e - \lambda_e g_A^e)^2 L_{\mu\nu}^\gamma, \\ L_{\mu\nu}^{\gamma r} &= (c_{V_r}^e - \lambda_e c_{A_r}^e) L_{\mu\nu}^\gamma, \\ L_{\mu\nu}^{Zr} &= -(c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) L_{\mu\nu}^\gamma, \end{aligned} \quad (13)$$

where $\lambda_e = \pm 1$ denotes the lepton helicity. For positrons, one flips the sign of all the g_A^e and $c_{A_r}^e$ terms and the overall sign of $L^{\gamma Z}$ and $L^{\gamma r}$ above. Using these identities for the leptonic tensors, Eq. (11) can be written more explicitly as

$$\begin{aligned} \frac{d^2\sigma}{dx dy} &= \frac{2\pi y \alpha^2}{Q^4} L_{\mu\nu}^\gamma \left\{ \eta^\gamma W_{\gamma}^{\mu\nu} - \eta^{\gamma Z} (g_V^e - \lambda_e g_A^e) W_{\gamma Z}^{\mu\nu} \right. \\ &\quad \left. + \eta^Z (g_V^e - \lambda_e g_A^e)^2 W_Z^{\mu\nu} + \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) W_{\gamma r}^{\mu\nu} \right. \\ &\quad \left. - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) W_{Zr}^{\mu\nu} \right\}. \end{aligned} \quad (14)$$

Based on the general Lorentz-tensor structure, the available four-momenta, and the nucleus spin vector, S^μ , numerous hadronic tensors are parametrized in terms of structure functions as

$$\begin{aligned} W_{\mu\nu}^j &= \left(-g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2} \right) F_1^j + \frac{\hat{P}_\mu \hat{P}_\nu}{(P \cdot q)} F_2^j \\ &\quad + \frac{i\epsilon_{\mu\nu\alpha\beta}}{2(P \cdot q)} [P^\alpha q^\beta F_3^j + 2q^\alpha S^\beta g_1^j] \\ &\quad - \frac{S \cdot q}{(P \cdot q)} \left[\frac{\hat{P}_\mu \hat{P}_\nu}{P \cdot q} g_4^j + \left(g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2} \right) g_5^j \right], \end{aligned} \quad (15)$$

where $\hat{P}_\mu \equiv P_\mu - q_\mu (P \cdot q)/q^2$. The index j denotes the possibilities $\{\gamma, \gamma Z, Z, \gamma r, Zr\}$, and $F_{1,2,3}^j$ and $g_{1,4,5}^j$ denote various unpolarized and polarized nuclear structure functions, respectively. We omit two additional possible

Lorentz structures in the hadronic tensor, typically denoted as the polarized structure functions g_2 and g_3 , since these terms give a contribution to the cross section that is suppressed by M^2/Q^2 when contracted with the leptonic tensor. Therefore, we do not consider the structure functions $g_{2,3}$ in the rest of our analysis. The nucleus spin vector S^μ satisfies the constraints $S^2 = -M^2$ and $S \cdot P = 0$. For longitudinal polarization, it takes the canonical form $S^\mu = \lambda_H(|\mathbf{p}|, E \frac{\mathbf{p}}{|\mathbf{p}|})$, where $\lambda_H = \pm 1$ is the nucleon helicity and $P^\mu = (E, \mathbf{p})$ is the nucleon four-momentum.

Based on the structure of the cross section in Eq. (14), in conjunction with the form of the hadronic tensor in Eq. (15), it becomes useful to define the following combinations of structure functions that also include the SMEFT contributions

$$\begin{aligned} F_i &= F_i^{\text{SM,NC}} + F_i^{\text{SMEFT}}, \\ g_i &= g_i^{\text{SM,NC}} + g_i^{\text{SMEFT}}, \end{aligned} \quad (16)$$

where the SM contributions are given by the commonly known NC structure functions

$$\begin{aligned} F_i^{\text{SM,NC}} &= F_i^\gamma - \eta^{\gamma Z}(g_V^e - \lambda_e g_A^e) F_i^{\gamma Z} + \eta^Z (g_V^e - \lambda_e g_A^e)^2 F_i^Z, \\ g_i^{\text{SM,NC}} &= g_i^\gamma - \eta^{\gamma Z}(g_V^e - \lambda_e g_A^e) g_i^{\gamma Z} + \eta^Z (g_V^e - \lambda_e g_A^e)^2 g_i^Z, \end{aligned} \quad (17)$$

and the SMEFT contributions are given by

$$\begin{aligned} F_i^{\text{SMEFT}} &= \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) F_i^{\gamma r} - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) F_i^{Zr}, \\ g_i^{\text{SMEFT}} &= \sum_r \xi^{\gamma r} (c_{V_r}^e - \lambda_e c_{A_r}^e) g_i^{\gamma r} - \sum_r \xi^{Zr} (c_{V_r}^e - \lambda_e c_{A_r}^e) (g_V^e - \lambda_e g_A^e) g_i^{Zr}. \end{aligned} \quad (18)$$

The parton-model expressions for the SM structure functions are summarized below. We also provide the corresponding expressions for the structure functions arising from the interference of the SM with the SMEFT operators:

$$\begin{aligned} [F_2^\gamma, F_2^{\gamma Z}, F_2^Z, F_2^{\gamma r}, F_2^{Zr}] &= x \sum_f [Q_f^2, 2Q_f g_V^f, g_V^{f2} + g_A^{f2}, 2Q_f c_{V_r}^f, 2(g_V^f c_{V_r}^f + g_A^f c_{A_r}^f)] (q_f + \bar{q}_f), \\ [F_3^\gamma, F_3^{\gamma Z}, F_3^Z, F_3^{\gamma r}, F_3^{Zr}] &= \sum_f [0, 2Q_f g_A^f, 2g_V^f g_A^f, 2Q_f c_{A_r}^f, 2(g_V^f c_{A_r}^f + g_A^f c_{V_r}^f)] (q_f - \bar{q}_f), \\ [g_1^\gamma, g_1^{\gamma Z}, g_1^Z, g_1^{\gamma r}, g_1^{Zr}] &= \frac{1}{2} \sum_f [Q_f^2, 2Q_f g_V^f, g_V^{f2} + g_A^{f2}, 2Q_f c_{V_r}^f, 2(g_V^f c_{V_r}^f + g_A^f c_{A_r}^f)] (\Delta q_f + \Delta \bar{q}_f), \\ [g_5^\gamma, g_5^{\gamma Z}, g_5^Z, g_5^{\gamma r}, g_5^{Zr}] &= \sum_f [0, Q_f g_A^f, g_V^f g_A^f, Q_f c_{A_r}^f, g_V^f c_{A_r}^f + g_A^f c_{V_r}^f] (\Delta q_f - \Delta \bar{q}_f), \end{aligned} \quad (19)$$

where $q_f(x, Q^2)$ and $\Delta q_f(x, Q^2)$ are unpolarized and polarized parton distribution functions (PDFs) of quark flavor f , respectively, and Q_f denotes the electric charge in units of the proton charge e . In the parton model, at leading order (LO), one has for the structure functions the Callan-Gross relations $F_2^i = 2x F_1^i$ and $g_4^i = 2x g_5^i$ for $i = \gamma, \gamma Z, Z, \gamma r, Zr$. For an ion beam (or nuclear target), the neutron PDFs can be related to the proton PDFs by assuming isospin symmetry for the valence quarks

$$\begin{aligned} q_{u/n}(x, Q^2) &= q_{d/p}(x, Q^2), \\ q_{d/n}(x, Q^2) &= q_{u/p}(x, Q^2), \\ \Delta q_{u/n}(x, Q^2) &= \Delta q_{d/p}(x, Q^2), \\ \Delta q_{d/n}(x, Q^2) &= \Delta q_{u/p}(x, Q^2), \end{aligned} \quad (20)$$

while the charm and strange sea quark PDFs are assumed to be identical for the proton and the neutron:

$$\begin{aligned} q_{s/n}(x, Q^2) &= q_{s/p}(x, Q^2), \\ q_{c/n}(x, Q^2) &= q_{c/p}(x, Q^2), \\ \Delta q_{s/n}(x, Q^2) &= \Delta q_{s/p}(x, Q^2), \\ \Delta q_{c/n}(x, Q^2) &= \Delta q_{c/p}(x, Q^2). \end{aligned} \quad (21)$$

For the deuteron, an isoscalar bound state of a proton and a neutron, the PDFs can be constructed from the proton and neutron PDFs as

$$\begin{aligned} q_{f/D}(x, Q^2) &= \frac{1}{2} [q_{f/p}(x, Q^2) + q_{f/n}(x, Q^2)], \\ \Delta q_{f/D}(x, Q^2) &= \frac{1}{2} [\Delta q_{f/p}(x, Q^2) + \Delta q_{f/n}(x, Q^2)], \end{aligned} \quad (22)$$

for quark flavor f .

In terms of the generalized structure functions in Eq. (16), which include dependence on the electron

helicity, λ_e , as seen in Eqs. (17) and (18), one can write the cross section for given electron and nucleon helicities, including SMEFT operator contributions, as

$$\frac{d^2\sigma(\lambda_e, \lambda_H)}{dx dy} = \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 F_1 + (1-y)F_2 - \lambda_e \frac{y}{2} (2-y)x F_3 + \lambda_e \lambda_H (2-y)xy g_1 - \lambda_H (1-y)g_4 - \lambda_H xy^2 g_5 \right\}, \quad (23)$$

where we ignore the electron mass and all target-mass correction terms that are proportional to M^2/Q^2 .

To connect to experimentally measured observables, it is convenient to write the scattering cross section of Eq. (23) as the sum of four components that depend on the spin direction of the initial electron and hadron: $d\sigma_0$, $d\sigma_e$, $d\sigma_H$, and $d\sigma_{eH}$, where each $d\sigma$ represents the differential cross section as $d^2\sigma/(dx dy)$. The quantity $d\sigma_0$ is the unpolarized cross section, $d\sigma_e$ and $d\sigma_H$ denote the cross-section differences between initial electron and hadron states of opposite helicity, respectively, and $d\sigma_{eH}$ is the cross-section difference between initial electron and hadron states with the same and opposite helicities defined in the center-of-mass frame. These quantities can be formed by using Eq. (23) as

$$\begin{aligned} d\sigma_0 &= \frac{1}{4} [d\sigma|_{\lambda_e=+1, \lambda_H=+1} + d\sigma|_{\lambda_e=+1, \lambda_H=-1} + d\sigma|_{\lambda_e=-1, \lambda_H=+1} + d\sigma|_{\lambda_e=-1, \lambda_H=-1}], \\ d\sigma_e &= \frac{1}{4} [d\sigma|_{\lambda_e=+1, \lambda_H=+1} + d\sigma|_{\lambda_e=+1, \lambda_H=-1} - d\sigma|_{\lambda_e=-1, \lambda_H=+1} - d\sigma|_{\lambda_e=-1, \lambda_H=-1}], \\ d\sigma_H &= \frac{1}{4} [d\sigma|_{\lambda_e=+1, \lambda_H=+1} - d\sigma|_{\lambda_e=+1, \lambda_H=-1} + d\sigma|_{\lambda_e=-1, \lambda_H=+1} - d\sigma|_{\lambda_e=-1, \lambda_H=-1}], \\ d\sigma_{eH} &= \frac{1}{4} [d\sigma|_{\lambda_e=+1, \lambda_H=+1} - d\sigma|_{\lambda_e=+1, \lambda_H=-1} - d\sigma|_{\lambda_e=-1, \lambda_H=+1} + d\sigma|_{\lambda_e=-1, \lambda_H=-1}], \end{aligned} \quad (24)$$

and can be computed in conjunction with Eqs. (16), (17), and (18).

The SM contributions to the DIS cross sections with the target-mass terms omitted are

$$\begin{aligned} \frac{d^2\sigma_0}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 [F_1^\gamma - g_V^e \eta_{\gamma Z} F_1^{\gamma Z} + (g_V^e{}^2 + g_A^e{}^2) \eta_Z F_1^Z] \right. \\ &\quad + (1-y) [F_2^\gamma - g_V^e \eta_{\gamma Z} F_2^{\gamma Z} + (g_V^e{}^2 + g_A^e{}^2) \eta_Z F_2^Z] \\ &\quad \left. - \frac{xy}{2} (2-y) [g_A^e \eta_{\gamma Z} F_3^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_3^Z] \right\}, \\ \frac{d^2\sigma_e}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ xy^2 [g_A^e \eta_{\gamma Z} F_1^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_1^Z] + (1-y) [g_A^e \eta_{\gamma Z} F_2^{\gamma Z} - 2g_V^e g_A^e \eta_Z F_2^Z] \right. \\ &\quad \left. + \frac{xy}{2} (2-y) [g_V^e \eta_{\gamma Z} F_3^{\gamma Z} - (g_V^e{}^2 + g_A^e{}^2) \eta_Z F_3^Z] \right\}, \\ \frac{d^2\sigma_H}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ (2-y)xy [g_A^e \eta_{\gamma Z} g_1^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_1^Z] \right. \\ &\quad + (1-y) [g_V^e \eta_{\gamma Z} g_4^{\gamma Z} - (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_4^Z] \\ &\quad \left. - xy^2 [g_V^e \eta_{\gamma Z} g_5^{\gamma Z} - (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_5^Z] \right\}, \\ \frac{d^2\sigma_{eH}}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \left\{ (2-y)xy [g_1^\gamma - g_V^e \eta_{\gamma Z} g_1^{\gamma Z} + (g_V^e{}^2 + g_A^e{}^2) \eta_Z g_1^Z] \right. \\ &\quad \left. - (1-y) [g_A^e \eta_{\gamma Z} g_4^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_4^Z] - xy^2 [g_A^e \eta_{\gamma Z} g_5^{\gamma Z} - 2g_V^e g_A^e \eta_Z g_5^Z] \right\}. \end{aligned} \quad (25)$$

The SMEFT contributions are

$$\begin{aligned}
\frac{d^2\sigma_0^{\text{SMEFT}}}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \sum_r \left[xy^2 (c_{V_r}^e \xi_{\gamma r} F_1^{\gamma r} - (c_{V_r}^e g_V^e + c_{A_r}^e g_A^e) \xi_{Zr} F_1^{Zr}) \right. \\
&\quad + (1-y) (c_{V_r}^e \xi_{\gamma r} F_2^{\gamma r} - (c_{V_r}^e g_V^e + c_{A_r}^e g_A^e) \xi_{Zr} F_2^{Zr}) \\
&\quad \left. + \frac{xy}{2} (2-y) (c_{A_r}^e \xi_{\gamma r} F_3^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_3^{Zr}) \right], \\
\frac{d^2\sigma_e^{\text{SMEFT}}}{dx dy} &= -\frac{4\pi\alpha^2}{xyQ^2} \sum_r \left[xy^2 (c_{A_r}^e \xi_{\gamma r} F_1^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_1^{Zr}) \right. \\
&\quad + (1-y) (c_{A_r}^e \xi_{\gamma r} F_2^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} F_2^{Zr}) \\
&\quad \left. + \frac{xy}{2} (2-y) (c_{V_r}^e \xi_{\gamma r} F_3^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} F_3^{Zr}) \right], \\
\frac{d^2\sigma_H^{\text{SMEFT}}}{dx dy} &= -\frac{4\pi\alpha^2}{xyQ^2} \sum_r [xy(2-y) (c_{A_r}^e \xi_{\gamma r} g_1^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_1^{Zr}) \\
&\quad + (1-y) (c_{V_r}^e \xi_{\gamma r} g_4^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_4^{Zr}) \\
&\quad + xy^2 (c_{V_r}^e \xi_{\gamma r} g_5^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_5^{Zr})], \\
\frac{d^2\sigma_{eH}^{\text{SMEFT}}}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} \sum_r [xy(2-y) (c_{V_r}^e \xi_{\gamma r} g_1^{\gamma r} - (c_{A_r}^e g_A^e + c_{V_r}^e g_V^e) \xi_{Zr} g_1^{Zr}) \\
&\quad + (1-y) (c_{A_r}^e \xi_{\gamma r} g_4^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_4^{Zr}) \\
&\quad + xy^2 (c_{A_r}^e \xi_{\gamma r} g_5^{\gamma r} - (c_{V_r}^e g_A^e + c_{A_r}^e g_V^e) \xi_{Zr} g_5^{Zr})]. \tag{26}
\end{aligned}$$

If a positron beam becomes available at the EIC, one can measure both e^+H and e^-H cross sections and study the differences. Neglecting target-mass terms again and writing the SM and SMEFT contributions all together, we have

$$\begin{aligned}
\frac{d^2\sigma_0^{e^+}}{dx dy} - \frac{d^2\sigma_0^{e^-}}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} g_A^e [xy(2-y) (\eta_{\gamma Z} F_3^{\gamma Z} - 2g_V^e \eta_Z F_3^Z)] \\
&\quad - \frac{8\pi\alpha^2}{xyQ^2} \sum_r \frac{xy}{2} (2-y) c_{A_r}^e (\xi_{\gamma r} F_3^{\gamma r} + 2g_V^e \xi_{Zr} F_3^{Zr}), \\
\frac{d^2\sigma_e^{e^+}}{dx dy} - \frac{d^2\sigma_e^{e^-}}{dx dy} &= -\frac{4\pi\alpha^2}{xyQ^2} g_A^e [2(1-y) (\eta_{\gamma Z} F_2^{\gamma Z} - 2g_V^e \eta_Z F_2^Z) + 2xy^2 (\eta_{\gamma Z} F_1^{\gamma Z} - 2g_V^e \eta_Z F_1^Z)] \\
&\quad + \frac{8\pi\alpha^2}{xyQ^2} \sum_r [(1-y) c_{A_r}^e (\xi_{\gamma r} F_2^{\gamma r} + 2g_V^e \xi_{Zr} F_2^{Zr}) + xy^2 c_{A_r}^e (\xi_{\gamma r} F_1^{\gamma r} + 2g_V^e \xi_{Zr} F_1^{Zr})], \\
\frac{d^2\sigma_H^{e^+}}{dx dy} - \frac{d^2\sigma_H^{e^-}}{dx dy} &= -\frac{4\pi\alpha^2}{xyQ^2} g_A^e [2xy(2-y) (\eta_{\gamma Z} g_1^{\gamma Z} - 2g_V^e \eta_Z g_1^Z)] \\
&\quad + \frac{8\pi\alpha^2}{xyQ^2} \sum_r xy(2-y) c_{A_r}^e (\xi_{\gamma r} g_1^{\gamma r} - 2g_V^e \xi_{Zr} g_1^{Zr}), \\
\frac{d^2\sigma_{eH}^{e^+}}{dx dy} - \frac{d^2\sigma_{eH}^{e^-}}{dx dy} &= \frac{4\pi\alpha^2}{xyQ^2} g_A^e [2(1-y) (\eta_{\gamma Z} g_4^{\gamma Z} - 2g_V^e \eta_Z g_4^Z) + 2xy^2 (\eta_{\gamma Z} g_5^{\gamma Z} - 2g_V^e \eta_Z g_5^Z)] \\
&\quad - \frac{8\pi\alpha^2}{xyQ^2} \sum_r [(1-y) c_{A_r}^e (\xi_{\gamma r} g_4^{\gamma r} + 2g_V^e \xi_{Zr} g_4^{Zr}) + xy^2 c_{A_r}^e (\xi_{\gamma r} g_5^{\gamma r} + 2g_V^e \xi_{Zr} g_5^{Zr})]. \tag{27}
\end{aligned}$$

In this study, we focus on measurements of both parity-violating and lepton-charge asymmetries. The parity-violating asymmetry can be formed either by comparing right-handed and left-handed electron scattering from unpolarized hadrons, referred to as the ‘‘unpolarized PV asymmetry,’’

$$A_{\text{PV}}^{(e)} \equiv \frac{d\sigma_e}{d\sigma_0}, \quad (28)$$

or by comparing unpolarized electron scattering off right-handed and left-handed hadrons, referred to as the ‘‘polarized PV asymmetry,’’

$$A_{\text{PV}}^{(H)} \equiv \frac{d\sigma_H}{d\sigma_0}. \quad (29)$$

If a positron beam becomes available in the future, the lepton-charge asymmetry, defined as the unpolarized DIS cross-section asymmetry between electron and positron beams,

$$A_{\text{LC}}^{(H)} = \frac{d\sigma_0^{e^+} - d\sigma_0^{e^-}}{d\sigma_0^{e^+} + d\sigma_0^{e^-}}, \quad (30)$$

will provide additional constraints on SMEFT interactions. On the other hand, the double-spin asymmetry, $A_{\text{PV}}^{(eH)} \equiv d\sigma_{eH}/d\sigma_0$, is the primary observable to study the nucleon spin structure but is not within the scope of this work. Similarly, a complete list of lepton-charge asymmetries that includes lepton-polarization dependence can be found in [15], but they provide similar constraints to SM and SMEFT studies as the unpolarized asymmetry defined in Eq. (30) and are not discussed in this work.

B. Measurement of parity-violating asymmetries at the EIC

In DIS experiments utilizing an electron beam of polarization P_e and a hadron beam of polarization P_H , the measured differential cross section is

$$d\sigma = d\sigma_0 + P_e d\sigma_e + P_H d\sigma_H + P_e P_H d\sigma_{eH}, \quad (31)$$

where P_e and P_H have the same sign as the respective beam helicities, λ_e and λ_H , and can take the values $-1 \leq P_e, P_H \leq 1$. The various cross-section components in Eq. (31) are given in Eqs. (25).

The PVDIS asymmetry can be formed by flipping the spin direction of either the electron beam or the ion beam. For the EIC, beams of opposite polarizations will be injected into the storage rings alternately, and thus each of the signs of both electron and ion polarizations is flipped periodically on a short timescale. This is in contrast to HERA, where data were taken with positive then negative electron polarization, with such long time intervals in between that runs with opposite electron polarizations are essentially two independent experiments.

We express the measured DIS event counts during a certain beam-helicity state as

$$N^{++} = a_{\text{det}} L^{++} (d\sigma_0 + |P_e^{++}| d\sigma_e + |P_H^{++}| d\sigma_H + |P_e^{++}| |P_H^{++}| d\sigma_{eH}), \quad (32)$$

$$N^{+-} = a_{\text{det}} L^{+-} (d\sigma_0 + |P_e^{+-}| d\sigma_e - |P_H^{+-}| d\sigma_H - |P_e^{+-}| |P_H^{+-}| d\sigma_{eH}), \quad (33)$$

$$N^{-+} = a_{\text{det}} L^{-+} (d\sigma_0 - |P_e^{-+}| d\sigma_e + |P_H^{-+}| d\sigma_H - |P_e^{-+}| |P_H^{-+}| d\sigma_{eH}), \quad (34)$$

$$N^{--} = a_{\text{det}} L^{--} (d\sigma_0 - |P_e^{--}| d\sigma_e - |P_H^{--}| d\sigma_H + |P_e^{--}| |P_H^{--}| d\sigma_{eH}), \quad (35)$$

where L^{ij} stands for the integrated luminosity, and P_e^{ij} and P_H^{ij} are the electron and the proton (or ion) beam polarizations during the corresponding helicity bunch ij . The superscripts $ij = ++, +-, -+, --$ represent the electron and the proton helicity states with their time sequence depending on the helicity pattern of the beam injection. The a_{det} factor represents the detector phase space, acceptance, and efficiency. In the simplest case, if we assume both beam polarizations, the luminosity, and detector efficiency and acceptance do not vary with time, then

$$d\sigma_0 = \frac{1}{4} (d\sigma^{++} + d\sigma^{+-} + d\sigma^{-+} + d\sigma^{--}), \quad (36)$$

$$d\sigma_e = \frac{1}{4|P_e|} (d\sigma^{++} + d\sigma^{+-} - d\sigma^{-+} - d\sigma^{--}), \quad (37)$$

$$d\sigma_H = \frac{1}{4|P_H|} (d\sigma^{++} - d\sigma^{+-} + d\sigma^{-+} - d\sigma^{--}), \quad (38)$$

$$d\sigma_{eH} = \frac{1}{4|P_e||P_H|} (d\sigma^{++} - d\sigma^{+-} - d\sigma^{-+} + d\sigma^{--}), \quad (39)$$

where we define the experimentally measured cross section by $d\sigma^{ij} \equiv N^{ij}/L^{ij}/a_{\text{det}}$. The PVDIS asymmetry due to the electron spin flip can be extracted from data by taking the ratio of the cross sections. Because spin flips of both electron and hadron beams will be carried out at a very short timescale, the factor a_{det} can be assumed constant and cancels out when forming the asymmetry, and we can extract the asymmetry from experimentally measured yields, defined by $Y^{ij} \equiv N^{ij}/L^{ij}$, as

$$A_{\text{PV}}^{(e)} \equiv \frac{d\sigma_e}{d\sigma_0} = \frac{1}{|P_e|} \frac{Y^{++} + Y^{+-} - Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}}, \quad (40)$$

and that due to proton (ion) spin flip can be similarly extracted as

$$A_{\text{PV}}^{(H)} \equiv \frac{d\sigma_H}{d\sigma_0} = \frac{1}{|P_H|} \frac{Y^{++} - Y^{+-} + Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}}. \quad (41)$$

The design of the EIC requires that the point-to-point luminosity uncertainty be at the 10^{-4} level. Therefore, the

dominant experimental uncertainty would come from electron and proton (ion) polarimetry for $A_{PV}^{(e)}$ and $A_{PV}^{(H)}$, respectively.

C. Measurement of lepton-charge asymmetries at the EIC

Unlike PV asymmetries, which can be formed by comparing scattering yields of right-handed vs left-handed electron or hadron scattering on a short timescale, the measurement of the LC asymmetry requires comparison between electron and positron runs and thus relies on two independent cross-section measurements. To reduce the uncertainty in the measurement of $A_{LC}^{(H)}$, we can reverse the polarity of the magnet to minimize the systematic uncertainty due to differences in e^- and e^+ detection. In this case, the main experimental systematic uncertainty will come from the luminosity difference between e^- and e^+ runs, which is assumed to be 2% (relative in luminosity, absolute in $A_{LC}^{(H)}$) in this analysis.

III. PROJECTION OF PARITY-VIOLATION AND LEPTON-CHARGE ASYMMETRY DATA

A. ECCE detector configuration for inclusive neutral-current study

The ECCE detector concept [16] addresses the full EIC science mission as described in the EIC community white paper [17] and the 2018 National Academies of Science (NAS) report [18]. It is simultaneously fully capable, low risk, and cost effective. ECCE strategically repurposes select components of existing experimental equipment to maximize its overall capabilities within the envelope of planned resources. For example, the central barrel of the detector incorporates the storied 1.4-T *BABAR* superconducting solenoid and the sPHENIX barrel hadronic calorimeter, currently under construction.

For EW NC physics studied in this work, we focus on the detection and identification of inclusive scattered electrons, provided by ECCE's tracking system [19] combined with electromagnetic calorimetry [20] in a nearly hermetic coverage. ECCE features a hybrid tracking detector design using three state-of-the-art technologies to achieve high-precision primary and decay-vertex determination, fine momentum tracking, and distance-of-closest-approach resolution in the region $|\eta| \leq 3.5$ with full azimuth coverage. The ECCE tracking detector consists of the Monolithic Active Pixel Sensor (MAPS)-based silicon vertex/tracking subsystem, the μ RWELL tracking subsystem, and the AC-LGAD outer tracker, which also serves as the time-of-flight detector, all optimized by artificial intelligence. For the electromagnetic calorimeter, the system employed by ECCE consists of the PbWO_4 -based electron endcap EM calorimeter (EEMC) for the region $-3.7 < \eta < -1.8$, the SciGlass-based barrel ECal for the region $-1.7 < \eta < 1.3$, and the Pb-scintillator

shashlike-type forward ECal (FEMC, hadron beam direction) that covers roughly $1.3 < \eta < 4$.

For the inclusive DIS kinematics determination, we use single-electron simulations in the full detector to study the measurement of the electron momentum and trajectory, and we characterize the difference between detected and true values as smearing in the electron momentum and polar and azimuthal angles. The smearing can then be applied to simulated events without involving the full detector. This is referred to as fast smearing and is the simulation method adopted here that yields all physics projections provided in this work. On the other hand, other methods that can be used to identify DIS kinematics, such as detecting all hadrons in the final state or detecting both the scattered electron and all hadrons, are not investigated here. Similarly, the use of the electromagnetic calorimeter can improve track identification in part of the phase space, but is not included in this work.

B. Simulation with fast smearing

We use the DJANGO event generator [21] (version 4.6.16 [22]) that includes full electromagnetic and electroweak radiative effects to generate Monte Carlo (MC) events for each of the four beam-energy and two beam-type combinations: $18 \times 275(137)$, $10 \times 275(137)$, 10×100 , and 5×100 GeV for ep (eD) collisions, respectively. For the deuterium ion beam, the energy specified is per nucleon. The fast-smearing method is applied to inclusive electron events in the DJANGO output, and the physics cross section and parity-violating asymmetries are calculated event by event using a modified user routine of DJANGO. The number of scattered DIS electrons is then calculated using the cross-section information and the expected integrated luminosity after correcting for bin migration.

The detector fast smearing is obtained from a single-electron gun simulation. Resolution spectra are determined for 57 evenly spaced bins for the pseudorapidity range $\eta = (-3.5625, 3.5625)$ and 1 GeV wide bins in the transverse momentum, p_T . For each DJANGO-simulated event, smearing in the electron momentum, p , and polar and azimuthal angles θ and ϕ are randomly picked from the corresponding spectrum and applied to the event, which are used to determine the detected kinematics of the event. While the smearing spectra are not exactly Gaussian-shaped, they are fitted with a Gaussian function. The fitted root-mean-square (RMS) values extracted for illustration purposes are displayed in Fig. 2.

Using the fast-smearing method, we generate 20 M total MC events for each of the beam-energy combinations. Of these 20 M, 10 M events are generated to study the kinematic coverage over the full phase space. The remaining 10 M events are generated with $Q_{\min}^2 = 50 \text{ GeV}^2$ for which DIS events have the most impact on the extraction of the weak mixing angle. The drawback of the fast-smearing method is that no selection of the hadronic state is implemented. Methods utilizing hadronic final states such

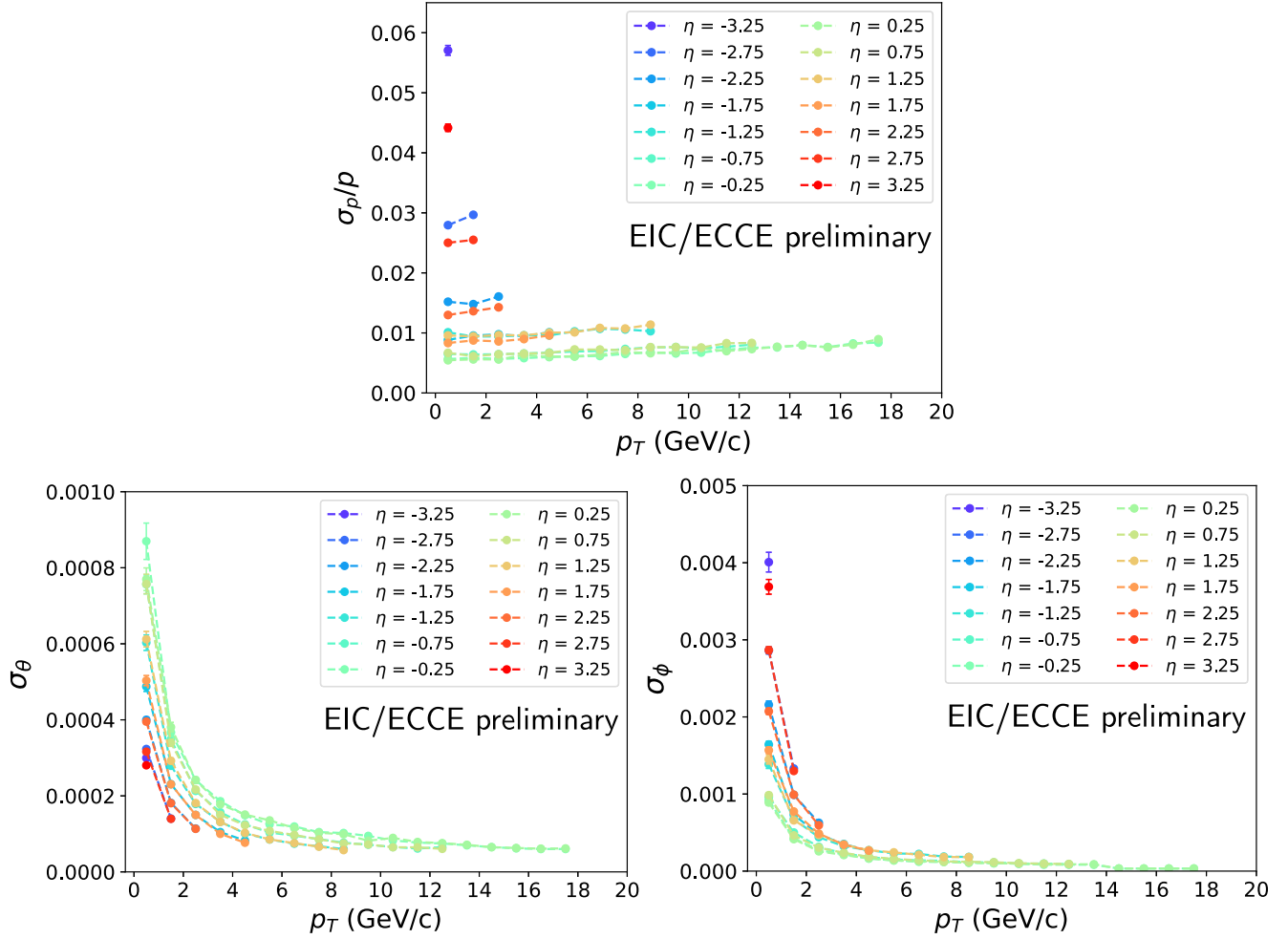


FIG. 2. RMS values for fast-smearing spectra obtained from single electron-gun simulation of the July 2021 concept of ECCE. The unit for σ_θ and σ_ϕ is radians.

as the double-angle method may provide better DIS event identification for a certain kinematic range and thus improve the precision of the analysis.

Bin migration of inclusive scattering electrons due to internal and external radiative effects is studied with fast-smearing simulation and treated using the “ R matrix” unfolding method [23]. Background reactions are studied using the hadronic final state generated by DJANGO (with $Q_{\min}^2 = 1.0 \text{ GeV}^2$), and another Monte Carlo simulation of photoproduction events are generated by PYTHIA (version 6.428, with $Q_{\min}^2 = 0$). All events are passed through the full ECCE simulation. We find that the highest background events occur at high y values. These events are rejected at the event-selection stage; see the next section.

We have also studied how our results change if a simple “theory-only” simulation without a detailed detector simulation is performed. We find two major differences with respect to the current analysis:

- (i) As mentioned in the next section, we use the inelasticity constraint $0.1 < y < 0.9$ in our current simulation. We find that the regions $0.1 < y < 0.2$

and $0.8 < y < 0.9$ are not reliably modeled without a detailed detector simulation. Our theory-only simulation cannot accurately reproduce the expected event counts in this region due to missing detector response effects. We therefore must remove these regions, leading to an effective reduction of statistics for the theory-only simulation.

- (ii) Second, in the $0.2 < y < 0.8$ region considered, the total error is relatively 10% to 30% more optimistic in each bin compared to the full detector simulation, with the 30% differences occurring near the boundaries of the y region.

The net result of these two competing effects is that theory-only bounds are up to 10% more optimistic than those found with a full detector simulation.

C. Event selection

For the 20 M fast-smearing events, event-selection criteria are applied to choose DIS events ($Q_{\text{det}}^2 > 1.0 \text{ GeV}^2$) in order to avoid regions with severe bin migration and unfolding uncertainty ($y_{\text{det}} > 0.1$), to avoid regions with

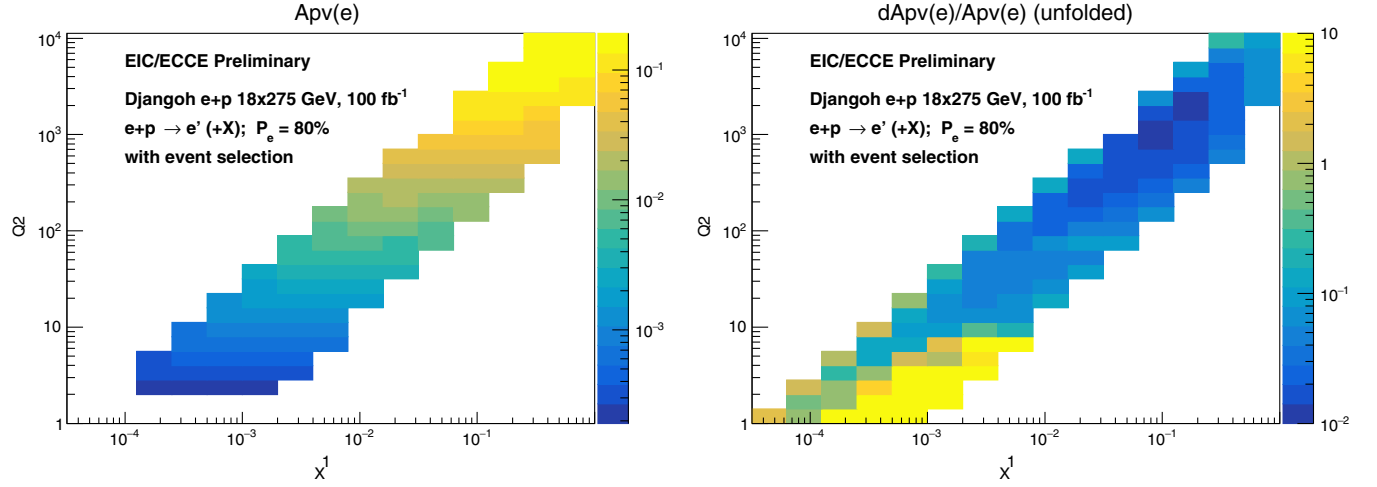


FIG. 3. Projection for $A_{PV}^{(e)}$ (left) and $dA_{PV,stat}^{(e)}/A_{PV}^{(e)}$ after unfolding (right) for 18×275 GeV ep collisions, with event-selection criteria applied. An integrated luminosity of 100 fb^{-1} and an electron polarization of 80% are assumed.

high photoproduction background ($y_{\text{det}} < 0.90$), to restrict events in the main acceptance of the ECCE detector where the fast-smearing method is applicable ($\eta_{\text{det}} > -3.5$ and $\eta_{\text{det}} < 3.5625$), and to ensure high purity of electron samples ($E' > 2.0$ GeV). Here, the subscript “det” implies the variables are calculated using the detected information of the electron. The projected values and statistical uncertainties for $A_{PV}^{(e)}$ and $A_{PV}^{(H)}$ after unfolding are shown in Figs. 3 and 4, respectively, for 18×275 GeV ep collisions with an integrated luminosity of 100 fb^{-1} .

D. Integrated luminosity

To account for realistic running conditions, the annual luminosity—the “high-divergence configuration” value as shown in Table 10.1 of the EIC yellow report (YR) [24], multiplied by 10^7 s—are used. These values are shown in Table II and will be referred to as “nominal luminosity

(NL)” hereafter. As a comparison with the weak mixing angle extraction presented in the YR, we also carry out projections for 100 fb^{-1} 18×275 GeV ep and 10 fb^{-1} 18×137 GeV eD collisions as the “YR reference point.” We abbreviate the ep pseudo-datasets as P1, P2, P3, P4, and P5 and the eD pseudo-datasets as D1, D2, D3, D4, and D5; see Table II. The YR reference point is denoted by P6. Simulated pseudo-datasets with polarized hadrons are indicated as Δ D1–5 and Δ P1–6, while positron datasets are referred to as LD1–5 and LP1–6 (with “L” for lepton charge).

As an exercise, we consider the additional statistical power that could be obtained by a hypothetical future high-luminosity upgrade to the EIC (HL-EIC) that delivers a tenfold increase in the integrated luminosity ($10\times$ higher than those in Table II) for these measurements. As the EIC is not yet built, there is no technical basis to assume that such an upgrade is possible. We choose the factor of $10\times$

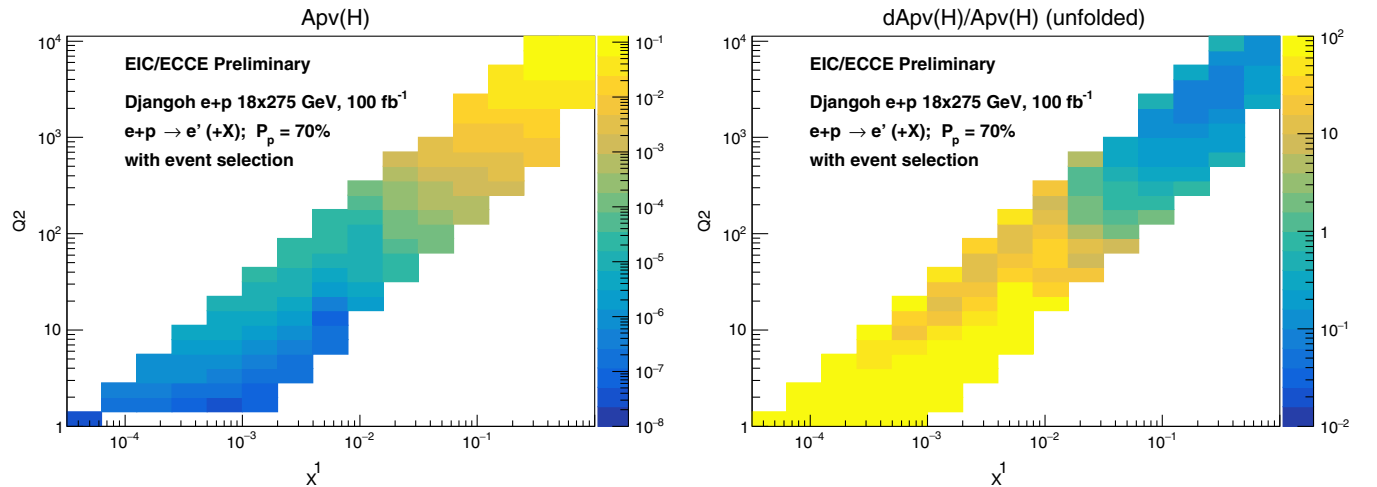


FIG. 4. Projection for $A_{PV}^{(p)}$ (left) and $dA_{PV,stat}^{(p)}/A_{PV}^{(p)}$ after unfolding (right) for 18×275 GeV ep collisions, with event-selection criteria applied. An integrated luminosity of 100 fb^{-1} and a proton polarization of 70% are assumed.

TABLE II. Beam energy, beam type, and the corresponding nominal annual luminosity assumed for the EIC in our analysis. P6 is the YR reference setting.

D1	5 GeV × 41 GeV eD , 4.4 fb ⁻¹	P1	5 GeV × 41 GeV ep , 4.4 fb ⁻¹
D2	5 GeV × 100 GeV eD , 36.8 fb ⁻¹	P2	5 GeV × 100 GeV ep , 36.8 fb ⁻¹
D3	10 GeV × 100 GeV eD , 44.8 fb ⁻¹	P3	10 GeV × 100 GeV ep , 44.8 fb ⁻¹
D4	10 GeV × 137 GeV eD , 100 fb ⁻¹	P4	10 GeV × 275 GeV ep , 100 fb ⁻¹
D5	18 GeV × 137 GeV eD , 15.4 fb ⁻¹	P5	18 GeV × 275 GeV ep , 15.4 fb ⁻¹
		P6	18 GeV × 275 GeV ep , 100 fb ⁻¹

luminosity increase to explore the sensitivity of the measurements we study in this paper, without making a comment as to the feasibility of such an upgrade. These projections will be denoted with a “high luminosity (HL)” label hereafter.

E. Statistical uncertainty projection for parity-violating asymmetries

For a given value of integrated luminosity, the statistical uncertainty of an asymmetry measurement is

$$dA_{\text{stat,measured}} = \frac{1}{\sqrt{N}}, \quad (42)$$

where N is the total number of events detected, assumed to be approximately equally divided between the two scattering types—between left- and right-handed electron beams, between left- and right-handed proton (ion) beams, or between positron and electron runs. The unfolding process increases the statistical precision only slightly for the region where the relative statistical uncertainty on the asymmetry is most precise.

If the asymmetry originates from polarization (as for the case of PV asymmetries), one must correct for the beam polarization:

$$dA_{\text{PV,stat}}^{(e)} = \frac{1}{|P_e|} \frac{1}{\sqrt{N}} \quad \text{and} \quad dA_{\text{PV,stat}}^{(H)} = \frac{1}{|P_H|} \frac{1}{\sqrt{N}}. \quad (43)$$

For $A_{\text{PV}}^{(e)}$ projections, an electron beam polarization of $P_e = 80\%$ with relative 1% systematic uncertainty from the electron polarimetry is assumed. Similarly, for $A_{\text{PV}}^{(H)}$ projections, a proton (ion) beam polarization of $P_H = 70\%$ with relative 2% systematic uncertainty from the proton (ion) polarimetry is used. An illustration of the relative precision of PV asymmetries is provided in Figs. 3 and 4. The statistical uncertainty of $A_{\text{PV}}^{(H)}$ is rather large because of the much smaller size of $A_{\text{PV}}^{(H)}$ than $A_{\text{PV}}^{(e)}$.

F. Statistical and QED uncertainty projection for lepton-charge asymmetries

As described in Sec. II C, to measure the lepton-charge asymmetry $A_{\text{LC}}^{(H)}$, one can reverse the polarity of the magnet to minimize the systematic uncertainty due to differences in

e^- and e^+ detection. In this case, the main experimental systematic uncertainty would come from the luminosity difference between e^- and e^+ runs, which is assumed to be 2% (relative in luminosity, absolute in A_{LC}) in this analysis. If the detector magnet polarity is reversed, then the detection of DIS positrons would be very similar to that of DIS electrons and all the data simulations, event selections, unfolding, etc., described in Sec. III B apply. The statistical uncertainty in A_{LC} is thus determined by the luminosity of the e^+ run, which we assume to be one-tenth of that of the electron beam. Note that beam polarization and thus polarimetry uncertainties do not affect A_{LC} measurements.

The EW physics reach of A_{LC} is further clouded by the difference in e^- vs e^+ DIS cross sections due to higher-order QED effects. We calculate the value of A_{LC} using DJANGO version 4.6.19 in both the Born LO (that includes one-boson exchange only) and next-to-leading-order (NLO) radiated mode (that includes higher-order EW and QED effects); see Fig. 5. The difference of NLO minus Born values is taken as an estimate of QED NLO effects, and the uncertainty is assumed to be 5% relative.

Because of the moderate Q^2 reach of the EIC, the 2% absolute uncertainty from luminosity measurement is a dominating systematic effect for the uncertainty of $A_{\text{LC}}^{(H)}$. In Appendix A 1, we present a method to simultaneously fit the luminosity term with SMEFT coefficients; however, we find this method yields 15% to 20% weaker SMEFT constraints.

G. Projection for high-luminosity EIC

In addition to the nominal luminosity expected for the EIC, we also carry out projections considering the possibility of an additional tenfold increase in the annual luminosity beyond EIC’s initial phase of running, the so-called HL-EIC. Assuming all experimental systematic effects remain the same, we scale the projected statistical uncertainty of asymmetry observables described in the previous section by a factor of $1/\sqrt{10}$. For beam energies with lower luminosity (hence larger statistical uncertainty) or asymmetries of smaller sizes such as $A_{\text{PV}}^{(H)}$, the tenfold increase in luminosity will push the physics reach one step further. On the other hand, for beam energies with already high luminosity and for observables where systematic effects dominate over the statistical ones, such as $A_{\text{PV}}^{(e)}$ for 10×275 GeV ep

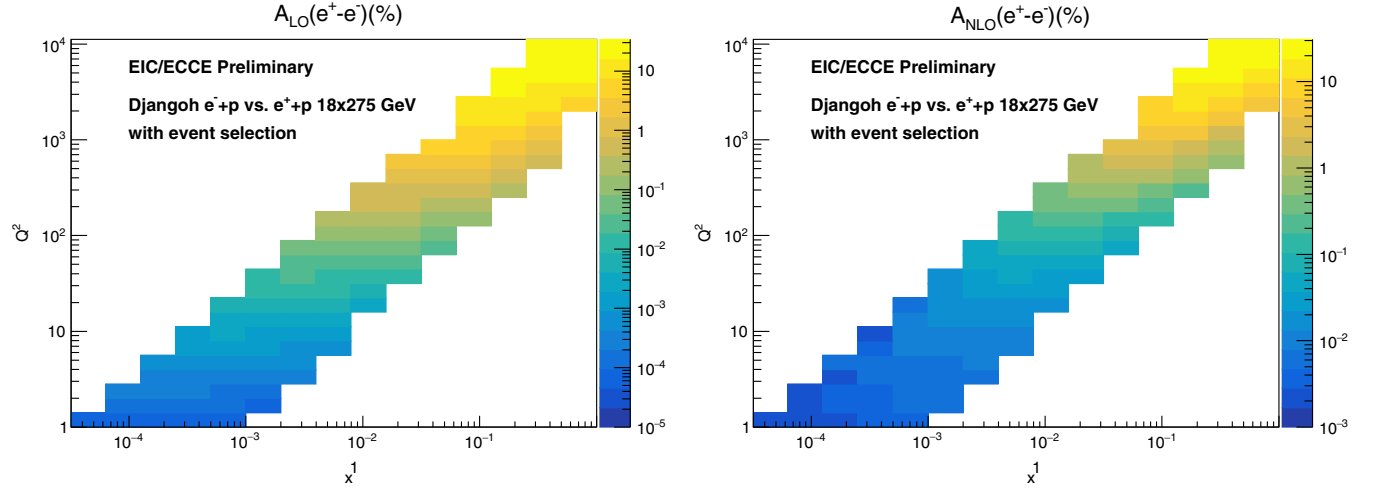


FIG. 5. Calculation for A_{LC} at the Born (LO) (left) and NLO (right) levels for e^+p vs e^-p collisions at 18×275 GeV. The LO calculation includes only the γZ interference term, which is of main interest of this study. The NLO calculation includes box diagrams, which introduces a large QED effect to the asymmetry and is effectively a background to the EW and SMEFT study presented here.

and 10×137 GeV eD collisions and $A_{LC}^{(H)}$, the impact from the luminosity increase of HL-EIC on the physics reach is marginal.

IV. PSEUDODATA GENERATION AND THE UNCERTAINTY MATRIX

A. Pseudodata for parity-violating asymmetries

We discuss first the case of the two PV asymmetries: polarized electron asymmetries with unpolarized hadrons, $A_{PV}^{(e)}$, and polarized hadron asymmetries with unpolarized electrons, $A_{PV}^{(H)}$. The experimental uncertainties are from

three sources: statistical, σ_{stat} ; experimental systematic, σ_{sys} , which is mainly due to particle background, also including other imperfections of the measurement, and is assumed to be fully uncorrelated; and beam polarimetry, σ_{pol} , which is assumed to be fully correlated within data of the same \sqrt{s} and beam type.

For the b th bin, with given \sqrt{s} , x , and Q^2 values and using the nominal PDF set under consideration, first we compute the theoretical SM prediction, $(A_{PV})_{\text{SM},0,b}^{\text{theo}}$. Combining the given uncertainties in quadrature separately for uncorrelated and correlated ones, we obtain a pseudoexperimental asymmetry value by

$$(A_{PV})_b^{\text{pseudo}} = (A_{PV})_{\text{SM},0,b}^{\text{theo}} + r_b \sqrt{\sigma_{\text{stat},b}^2 + \left[(A_{PV})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{sys}}}{A} \right)_b \right]^2} + r' \sqrt{\left[(A_{PV})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{pol}}}{A} \right)_b \right]^2}, \quad (44)$$

where r_b and r' are random numbers chosen from a normal distribution of mean 0 and standard deviation 1. Note that the correlated errors are incorporated using a single random number, r' , across all the bins. The systematic uncertainties are $\sigma_{\text{sys}}/A = 1\%$, $\sigma_{\text{pol}}/A = 1\%$ for $A_{PV}^{(e)}$, and $\sigma_{\text{pol}}/A = 2\%$ for $A_{PV}^{(H)}$.

B. Pseudodata for lepton-charge asymmetries

We consider next unpolarized electron-positron asymmetries with unpolarized hadrons, namely the LC asymmetries.

The uncertainties used in the data generation are from three sources: statistical, σ_{stat} ; experimental systematic, σ_{sys} , which is mainly due to background and is assumed to be fully uncorrelated; luminosity difference between e^+ and e^- runs, σ_{lum} , which is fully correlated within data of the same \sqrt{s} and ion beam type; and higher-order QED effects, $\sigma_{\text{QED NLO}}$, taken as 5% of the difference between the calculated NLO and Born (LO) A_{LC} values.

In analogy with Eq. (44), for the LC asymmetries, we write

$$(A_{LC})_b^{\text{pseudo}} = (A_{LC})_{\text{SM},0,b}^{\text{theo}} + r_b \sqrt{\sigma_{\text{stat}}^2 + \left[(A_{LC})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{sys}}}{A} \right)_b \right]^2} + \sigma_{\text{QED NLO},b}^2 + r' \sigma_{\text{lum},b}. \quad (45)$$

C. Uncertainty matrix

The uncertainty matrix, Σ^2 , for a given dataset with N_{bin} bins is an $N_{\text{bin}} \times N_{\text{bin}}$ symmetric matrix. It consists of two parts, which we call Σ_0^2 and Σ_{pdf}^2 :

$$(\Sigma^2)_{bb'} = (\Sigma_0^2)_{bb'} + (\Sigma_{\text{pdf}}^2)_{bb'}. \quad (46)$$

The first part of the matrix, Σ_0^2 , is constructed using all the uncertainty components (statistical, systematic, polarimetry or luminosity, and QED NLO if present) other than the PDF uncertainties. All the uncertainties that enter Σ_0^2 must be absolute; relative uncertainties are converted to absolute ones by multiplying with the theoretical SM prediction, $A_{\text{SM},0,b}^{\text{theo}}$, computed using the central member of the PDF set taken into account. The first part of the matrix then takes the form

$$\Sigma_0^2 = \begin{pmatrix} \sigma_1^2 & \rho_{12} \tilde{\sigma}_1 \tilde{\sigma}_2 & \cdots & \rho_{1N_{\text{bin}}} \tilde{\sigma}_1 \tilde{\sigma}_{N_{\text{bin}}} \\ & \sigma_2^2 & \cdots & \rho_{2N_{\text{bin}}} \tilde{\sigma}_2 \tilde{\sigma}_{N_{\text{bin}}} \\ & & \ddots & \vdots \\ & & & \sigma_{N_{\text{bin}}}^2 \end{pmatrix}_{\text{sym}}, \quad (47)$$

where, for the PV asymmetries, we have for the diagonal elements

$$\sigma_b^2 = \sigma_{\text{stat},b}^2 + \left[(A_{\text{PV}})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{sys}}}{A} \right)_b \right]^2 + \left[(A_{\text{PV}})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{pol}}}{A} \right)_b \right]^2, \quad (48)$$

and for the off-diagonal elements

$$\tilde{\sigma}_b = (A_{\text{PV}})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{pol}}}{A} \right)_b. \quad (49)$$

For the LC asymmetries, we have for the diagonal elements

$$\sigma_b^2 = \sigma_{\text{stat},b}^2 + \left[(A_{\text{LC}})_{\text{SM},0,b}^{\text{theo}} \left(\frac{\sigma_{\text{sys}}}{A} \right)_b \right]^2 + \sigma_{\text{lum},b}^2 + \sigma_{\text{QED NLO},b}^2, \quad (50)$$

and for the off-diagonal elements

$$\tilde{\sigma}_b = \sigma_{\text{lum},b}. \quad (51)$$

Here, b and b' are bin numbers, and we assume full correlation for uncertainties originating from beam polarimetry or luminosity: $\rho_{bb'} = 1$ for all b and b' .

The second part of the uncertainty matrix, Σ_{pdf}^2 , is built using the same procedure for both PV and LC asymmetries by taking into account differences between the theoretical

SM asymmetry prediction computed at the nominal PDF member, $A_{\text{SM},0}^{\text{theo}}$, and theoretical SM asymmetry predictions evaluated at all other members of the PDF set under consideration, $A_{\text{SM},m}^{\text{theo}}$, where $m = 1, 2, \dots, N_{\text{PDF}}$ with N_{PDF} the total number of PDF sets or replicas available. For Hessian-based PDF sets, the diagonal and off-diagonal elements can be collectively written as

$$(\Sigma_{\text{PDF}}^2)_{bb'}^{\text{Hessian}} = \frac{1}{4} \sum_{m=1}^{N_{\text{PDF}}/2} (A_{\text{SM},2m,b}^{\text{theo}} - A_{\text{SM},2m-1,b}^{\text{theo}}) \times (A_{\text{SM},2m,b'}^{\text{theo}} - A_{\text{SM},2m-1,b'}^{\text{theo}}). \quad (52)$$

For replica-based PDF sets, this expression becomes

$$(\Sigma_{\text{PDF}}^2)_{bb'}^{\text{replica}} = \frac{1}{N_{\text{PDF}}} \sum_{m=1}^{N_{\text{PDF}}} (A_{\text{SM},m,b}^{\text{theo}} - A_{\text{SM},0,b}^{\text{theo}}) \times (A_{\text{SM},m,b'}^{\text{theo}} - A_{\text{SM},0,b'}^{\text{theo}}). \quad (53)$$

D. Comparison of uncertainty components

We present in this section the various uncertainty components that enter the SMEFT analysis. We also investigate the total uncertainties combined in quadrature that contribute to the diagonal entries of the uncertainty matrix.

1. Individual uncertainty components

We begin by considering the individual components of the uncertainties. We investigate the effects of sea quarks in the analysis by defining a valence-only approximation for the PDFs. The tag ud in the plot labels implies the valence-only approximation, in which only up- and down-quark contributions are considered in the hadronic cross section, whereas uds indicates that up, down, strange, and their antiquarks are taken into account. Note that for the datasets involving unpolarized deuteron with the ud tag, there will be no uncertainty from PDFs since deuteron PDFs, defined in terms of proton and neutron PDFs using isospin symmetry, cancel when analytically forming asymmetries in the valence-only approximation. Note also that for experimental systematic uncertainties other than those from beam polarimetry, both 1% and 2% values are shown in all figures of this section, although the 1% value is used in the results presented.

Figure 6 shows the comparison of the uncertainty components for the dataset D4 in the ud and uds scenarios. As for the PDFs, we use NNPDF3.1 NLO [25] in the unpolarized case and NNPDFPOL1.1 [26] in the polarized case throughout. Only the (x, Q^2) region relevant for SMEFT analysis is shown, although the full region is used for the extraction of the weak mixing angle. The x axis of these plots is ordered by bin number; these are ordered first from low to high Q^2 , and then from small to

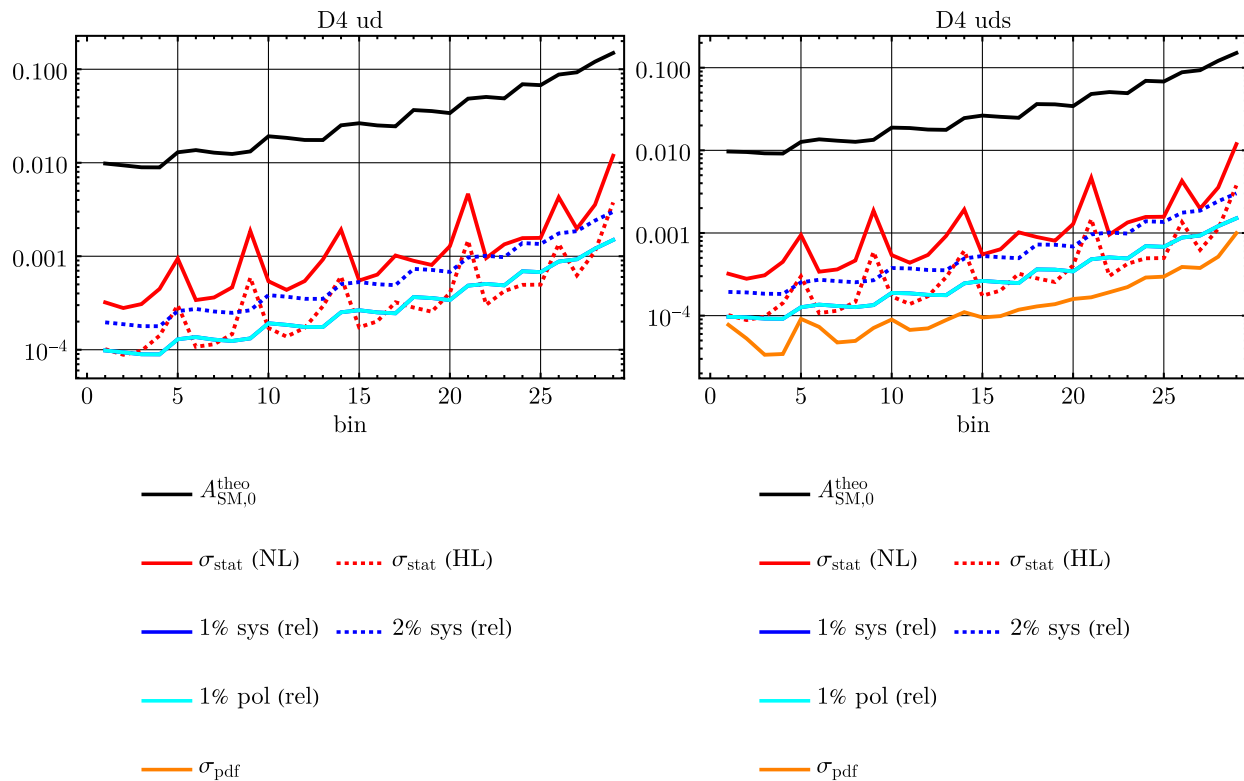


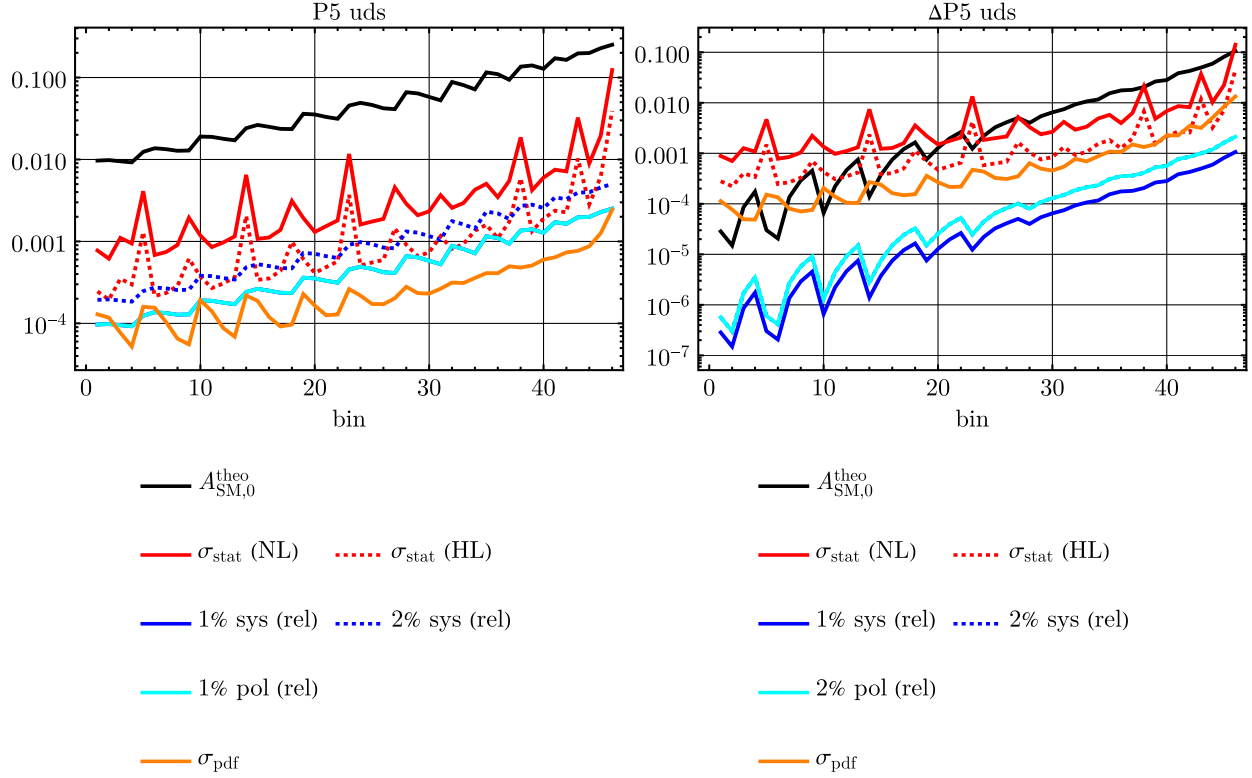
FIG. 6. Comparison of the uncertainty components for the dataset D4 in the valence-only scenario (ud) and with the contributions from the sea quarks (uds). Here, “NL” refers to the currently planned annual luminosity of the EIC, while “HL” refers to a potential tenfold luminosity upgrade.

large x within each Q^2 bin, leading to the observed oscillatory behavior. When we turn on the sea quark contributions, the unpolarized deuteron datasets receive nonzero but highly suppressed PDF uncertainties, indicating that the assumption of deuteron PDFs completely canceling is a reasonably good approximation. The right panel shows that even after including sea quarks, the PDFs are still the smallest uncertainty component. This indicates that potentially poorly determined sea quark and strange quark distributions have little effect on this analysis. The largest single uncertainty component is the statistical uncertainty (shown as a dark red line). This is larger than both the 1% beam polarization uncertainty (light blue line) and either of the 1% or 2% uncorrelated systematic uncertainty assumptions (solid and dotted blue lines, respectively). When we switch to the high-luminosity (HL-EIC) scenario (dotted red line), the statistical uncertainty becomes comparable to the systematic ones. All uncertainties are significantly smaller than the predicted values of the asymmetry, shown as the solid black line in the plots.

In Fig. 7, we display the different contributions to the diagonal entries of the uncertainty matrix of the datasets P5 and $\Delta P5$. The pattern of uncertainties for P5 is very similar to that observed for D4. The statistical ones are the largest single uncertainty source, while the PDFs are the smallest.

Assuming high luminosity, the statistical uncertainties become comparable to the anticipated systematic ones. The pattern is different for $\Delta P5$: the statistical uncertainties are largest for all bins, even assuming high luminosity. The PDF uncertainties are also non-negligible, consistent with the expectation that spin-dependent PDFs are not known as precisely as the spin-independent ones. The anticipated experimental systematic uncertainties are negligible for all bins.

Finally, we show in Fig. 8 the individual uncertainties for the electron-positron asymmetry dataset LP5. The error budget is different for this scenario compared to PV asymmetries. Since both beams are unpolarized, there is no uncertainty related to beam polarization. However, since electron and positron runs occur with different beams, there is the possibility of a significant overall luminosity difference between the two runs that can lead to an apparent asymmetry. We assume an absolute 2% uncertainty, 2 times the luminosity uncertainty requirement of [27]. Finally, we consider the possible errors arising from higher-order QED corrections that may differentiate between electron and positron scattering. We estimate this uncertainty by taking 5% of the difference between the Born-level and NLO QED results, obtained by using DJANGO. The two largest sources of uncertainty throughout the entire kinematic range are the luminosity and statistical uncertainties.


 FIG. 7. Uncertainty components for the datasets P5 and $\Delta P5$.

PDFs, higher-order QED, and anticipated systematic uncertainties are all significantly smaller.

Summarizing all the figures presented in this section, we can make the following main points:

- (i) The expected statistical uncertainties are the dominant ones for the nominal EIC luminosity. If a high-luminosity (HL-EIC) upgrade becomes realistic, they become comparable to experimental systematic uncertainties for PV asymmetries of the unpolarized hadron, $A_{PV}^{(e)}$.
- (ii) PDF uncertainties are nearly irrelevant for the asymmetries of unpolarized hadrons, $A_{PV}^{(e)}$. They become significant, second to statistical uncertainties, for PV asymmetries of polarized hadrons, $A_{PV}^{(H)}$.

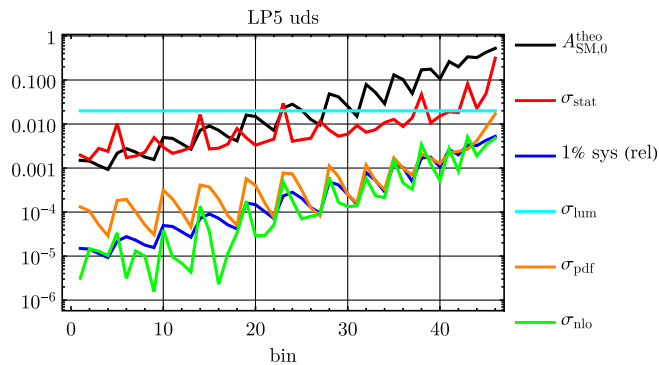
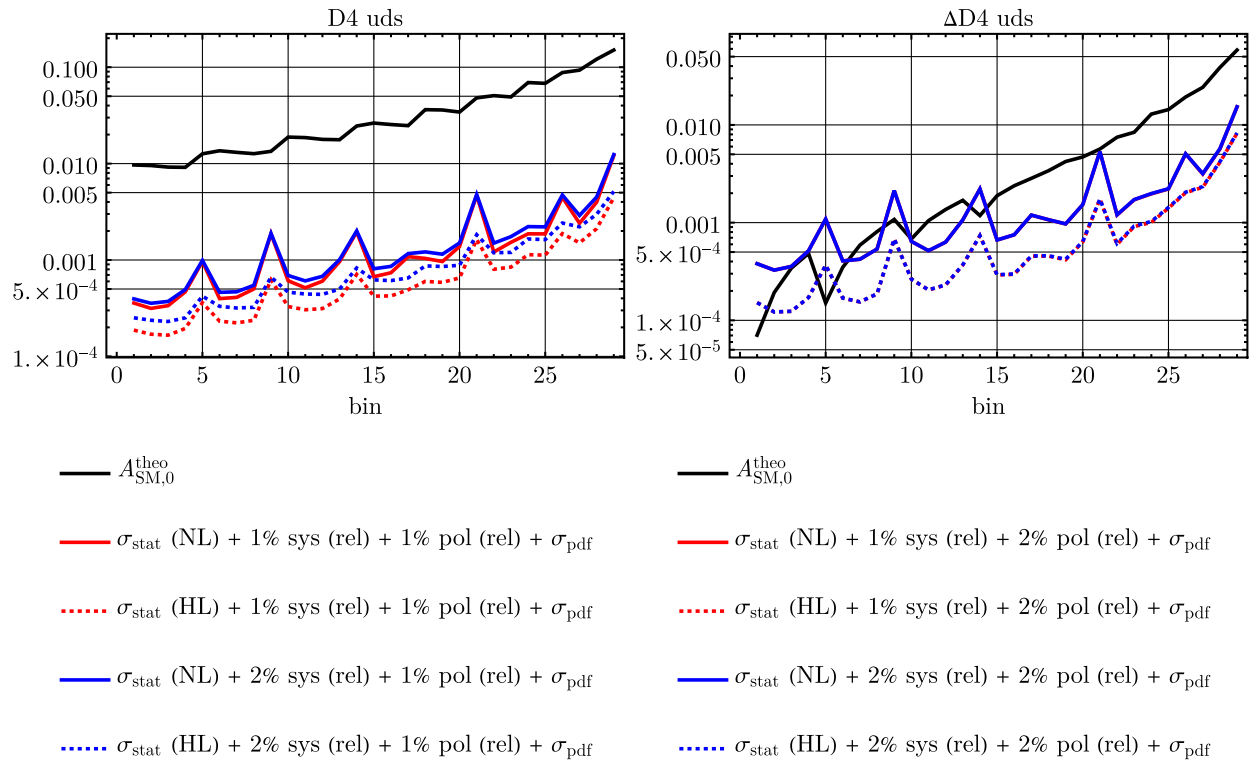
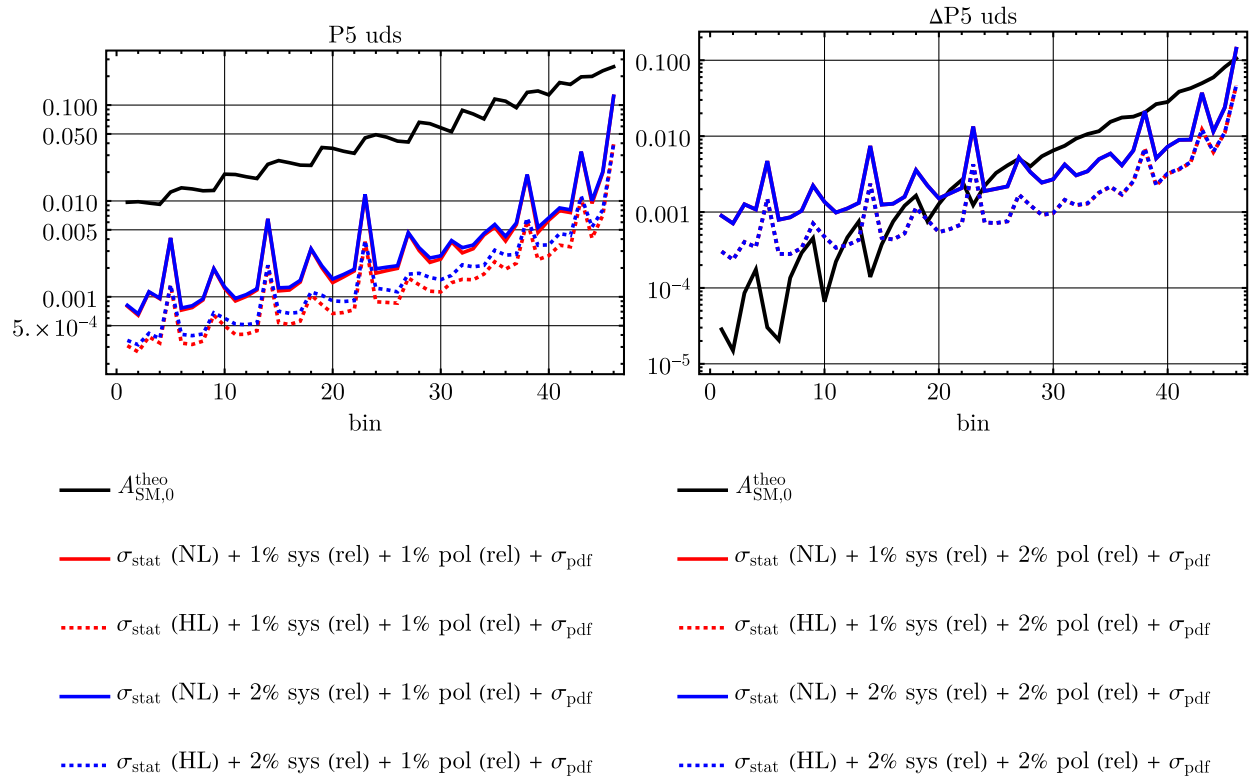


FIG. 8. The same as in Fig. 7 but for LP5.

- (iii) The luminosity effect dominates over the statistical uncertainty for the majority of the phase space in the case of electron-positron asymmetries, $A_{LC}^{(H)}$, particularly at low x and low Q^2 . On the other hand, uncertainties from missing higher-order QED corrections are expected to be small.

2. Total uncertainties for nominal luminosity vs high luminosity

We now investigate the total uncertainties for the datasets D4, $\Delta D4$, P5, and $\Delta P5$. We consider four different scenarios: the nominal annual luminosity planned for the EIC or a potential high-luminosity upgrade beyond the initial phase of the EIC run, combined with 1% or 2% relative experimental systematic uncertainties due to particle background. We show the results in Figs. 9 and 10. We observe first that the dominant uncertainty component in all cases is the statistical one. The four uncertainty scenarios, namely 1% or 2% systematic uncertainties combined with nominal or high luminosity, can, in fact, be reduced to just the luminosity comparison, i.e., nominal vs high. Next, for both D4 and P5, the asymmetry $A_{PV}^{(e)}$ is measured to percent level throughout the considered phase space. This is not the case for the polarized sets $\Delta D4$ and $\Delta P5$. Particularly in the $\Delta P5$ scenario at low Q^2 , the anticipated errors are larger than the asymmetry for all choices of systematic error and luminosity. Only in the very high


 FIG. 9. Total uncertainties combined in quadrature for the datasets D4 and $\Delta D4$ in the uds scenario.

 FIG. 10. The same as in Fig. 9 but for P5 and $\Delta P5$.

Q^2 bins does a measurement of the asymmetry $A_{\text{PV}}^{(H)}$ become meaningful.

Our evaluation of the uncertainties indicates that using 1% or 2% relative systematic uncertainties makes practically no difference, as the total errors are mostly dominated by the statistical uncertainties for the PV asymmetries or the luminosity difference for the LC asymmetries. We also show that one can take into account the contribution of only the valence quarks to the asymmetries or include the sea quarks up to strange flavor and its antiquark, both of which lead to the same size of PDF errors for the datasets under consideration. In our best-fit analyses, we thus focus on the datasets with 1% relative systematic uncertainty and nominal luminosity in the $u\bar{d}s$ scenario as our main datasets. Comparisons are performed to the ones having high luminosity, keeping the rest of the configuration the same.

An important issue to address is whether a joint fit of PDFs and Wilson coefficients would change the potential of the EIC to probe the SMEFT parameter space that we find in this draft. This issue has been studied for both HERA and LHC datasets in the literature [28,29], where it is found that the interplay between PDFs and Wilson coefficients can become a significant challenge for some future high-luminosity measurements at the LHC. It is beyond the scope of this paper to consider such a joint fit, so we can only speculate regarding the exact answer to this question. However, we can make the following points that are supported by the uncertainty plots in this section of this manuscript:

- (i) For the unpolarized deuteron datasets, the PDF uncertainties are an order of magnitude smaller than the statistical uncertainties for the nominal luminosity and 3 to 5 times smaller than the high-luminosity statistical uncertainties, as shown in Fig. 6. We therefore expect that a joint fit of the PDFs and Wilson coefficients would not greatly affect the bounds obtained here. The PDFs are

already determined sufficiently well from other experiments for the purposes of this analysis.

- (ii) The same statement holds for the P5 unpolarized datasets for both nominal and high luminosities, as shown in the left panel of Fig. 7. This plot indicates that taking ratios to form asymmetries, as we do in this study, greatly reduces the dependence on PDFs.
- (iii) Looking at the right panel of Fig. 7, we see that the polarized proton PDF error becomes comparable to the statistical error at high Q^2 for the high-luminosity dataset. In this case, a joint fit of polarized PDFs and Wilson coefficients will be especially important. We note that the bounds from the polarized hadron datasets are generically much weaker than those for the unpolarized hadron sets (see Fig. 13 for example), since the polarized asymmetry is much smaller than the unpolarized one. We believe that our main point regarding the EIC sensitivity to SMEFT Wilson coefficients is mostly unaffected by this point.

V. EXTRACTION OF THE SM WEAK MIXING ANGLE

The weak mixing angle, often written as $\sin^2 \theta_W$, is a fundamental parameter of the SM and has been measured in experiments ranging from atomic parity violation at electron volt energy levels to high-energy colliders at the Z pole [30–32]. The EIC will provide constraints on $\sin^2 \theta_W$ in the intermediate-energy range that resides between the reach of fixed-target and collider facilities.

For the extraction of the weak mixing angle, we focus on $A_{\text{PV}}^{(e)}$, where $\sin^2 \theta_W$ enters through the electron coupling $g_{V,A}^e$ and the corresponding quark couplings in the structure functions. We also include the one-loop renormalization group evolution [33] of $\sin^2 \theta_W$ in the $\overline{\text{MS}}$ scheme, including the relevant particle thresholds that arise between $\mu = M_Z$ and $\mu = \sqrt{Q^2}$. Including target-mass correction terms, we can write

$$A_{\text{PV}}^{(e)} = \frac{|P_e| \eta_{\gamma Z} [g_A^e 2y F_1^{\gamma Z} + g_A^e (\frac{2}{xy} - \frac{2}{x} - \frac{2M^2 xy}{Q^2}) F_2^{\gamma Z} + g_V^e (2-y) F_3^{\gamma Z}]}{2y F_1^\gamma + (\frac{2}{xy} - \frac{2}{x} - \frac{2M^2 xy}{Q^2}) F_2^\gamma - \eta_{\gamma Z} [g_V^e 2y F_1^{\gamma Z} + g_V^e (\frac{2}{xy} - \frac{2}{x} - \frac{2M^2 xy}{Q^2}) F_2^{\gamma Z} + g_A^e (2-y) F_3^{\gamma Z}]}, \quad (54)$$

where M is the nucleon mass. Note that given the moderate Q^2 values of the EIC, the pure-Z contribution to the structure functions is omitted for the precision relevant to our analysis.

A single pseudo-dataset is generated using a reference value of $\sin^2 \theta_W = 0.231$ at the Z pole, and the uncertainties in $A_{\text{PV}}^{(e)}$ in each (x, Q^2) bin are obtained from simulation studies. Comparing the theory prediction to the pseudodata, a best-fit value and uncertainty projection for $\sin^2 \theta_W$ at

the Z pole are obtained by minimizing the χ^2 function defined as

$$\chi^2 = [\mathcal{A}^{\text{pseudodata}} - \mathcal{A}^{\text{theory}}]^T [(\Sigma^2)^{-1}] [\mathcal{A}^{\text{pseudodata}} - \mathcal{A}^{\text{theory}}], \quad (55)$$

where \mathcal{A} is a dimension- N_{bin} vector with N_{bin} the total number of (x, Q^2) bins, Σ^2 is the uncertainty matrix of dimension $N_{\text{bin}} \times N_{\text{bin}}$, described in Sec. IV C, and $\sin^2 \theta_W$

TABLE III. Projected PVDIS asymmetry and fitted results for $\sin^2\theta_W$ using ep collision data and the nominal annual luminosity. Here, $\langle Q^2 \rangle$ denotes the value averaged over all (x, Q^2) bins, weighted by $(dA/A)_{\text{stat}}^{-2}$ for each bin. The electron beam polarization is assumed to be 80% with a relative 1% uncertainty. The total (“tot”) uncertainty is from combining all of statistical, 1% systematic (background), 1% beam polarization, and PDF uncertainties evaluated using three different PDF sets. The rightmost column is for comparison with the YR.

Beam type and energy	$ep\ 5 \times 100$	$ep\ 10 \times 100$	$ep\ 10 \times 275$	$ep\ 18 \times 275$	$ep\ 18 \times 275$
Label	P2	P3	P4	P5	P6
Luminosity [fb^{-1}]	36.8	44.8	100	15.4	(100 YR ref)
$\langle Q^2 \rangle$ [GeV^2]	154.4	308.1	687.3	1055.1	1055.1
$\langle A_{PV} \rangle$ ($P_e = 0.8$)	-0.00854	-0.01617	-0.03254	-0.04594	-0.04594
$(dA/A)_{\text{stat}}$	1.54%	0.98%	0.40%	0.80%	(0.31%)
$(dA/A)_{\text{stat+syst(bg)}}$	1.55%	1.00%	0.43%	0.81%	(0.35%)
$(dA/A)_{1\% \text{pol}}$	1.0%	1.0%	1.0%	1.0%	(1.0%)
$(dA/A)_{\text{tot}}$	1.84%	1.42%	1.09%	1.29%	(1.06%)
Experimental					
$d(\sin^2\theta_W)_{\text{stat+syst(bg)}}$	0.002032	0.001299	0.000597	0.001176	0.000516
$d(\sin^2\theta_W)_{\text{stat+syst+pol}}$	0.002342	0.001759	0.001297	0.001769	0.001244
With PDF					
$d(\sin^2\theta_W)_{\text{tot,CT18NLO}}$	0.002388	0.001807	0.001363	0.001823	0.001320
$d(\sin^2\theta_W)_{\text{tot,MMHT2014}}$	0.002353	0.001771	0.001319	0.001781	0.001270
$d(\sin^2\theta_W)_{\text{tot,NNPDF31}}$	0.002351	0.001789	0.001313	0.001801	0.001308

to be fitted enters $\mathcal{A}^{\text{theory}}$. The PDF portion of the uncertainty matrix is evaluated using the PDF sets CT18NLO [34] (LHAPDF [35] ID 14400-14458), MMHT2014nlo_68c1 [36] (ID 25100-25150), and NNPDF31_nlo_as_0118 [25] (ID 303400-303500).

Our results for $\sin^2\theta_W$ are shown in Tables III and IV for five energy and nominal-annual-luminosity combinations

of ep and eD collisions, respectively. These results are illustrated in Fig. 11. The inner error bars show the combined uncertainty from statistical and 1% uncorrelated experimental systematics (due to particle background); the median error bars show the experimental uncertainty that includes statistical, 1% uncorrelated experimental systematics, and 1% electron polarimetry. The outermost error

TABLE IV. Projected PVDIS asymmetry and fitted results for $\sin^2\theta_W$ using eD collision data and the nominal annual luminosity. The uncertainty evaluation is the same as Table III.

Beam type and energy	$eD\ 5 \times 100$	$eD\ 10 \times 100$	$eD\ 10 \times 137$	$eD\ 18 \times 137$	$eD\ 18 \times 137$
Label	D2	D3	D4	D5	N/A
Luminosity [fb^{-1}]	36.8	44.8	100	15.4	(10 YR ref)
$\langle Q^2 \rangle$ [GeV^2]	160.0	316.9	403.5	687.2	687.2
$\langle A_{PV} \rangle$ ($P_e = 0.8$)	-0.01028	-0.01923	-0.02366	-0.03719	-0.03719
$(dA/A)_{\text{stat}}$	1.46%	0.93%	0.54%	1.05%	(1.31%)
$(dA/A)_{\text{stat+bg}}$	1.47%	0.95%	0.56%	1.07%	(1.32%)
$(dA/A)_{\text{syst,1\%pol}}$	1.0%	1.0%	1.0%	1.0%	(1.0%)
$(dA/A)_{\text{tot}}$	1.78%	1.38%	1.15%	1.46%	(1.66%)
Experimental					
$d(\sin^2\theta_W)_{\text{stat+bg}}$	0.002148	0.001359	0.000823	0.001591	0.001963
$d(\sin^2\theta_W)_{\text{stat+bg+pol}}$	0.002515	0.001904	0.001544	0.002116	0.002414
With PDF					
$d(\sin^2\theta_W)_{\text{tot,CT18}}$	0.002558	0.001936	0.001566	0.002173	0.00247
$d(\sin^2\theta_W)_{\text{tot,MMHT2014}}$	0.002527	0.001917	0.001562	0.002128	0.002424
$d(\sin^2\theta_W)_{\text{tot,NNPDF31}}$	0.002526	0.001915	0.001560	0.002127	0.002423

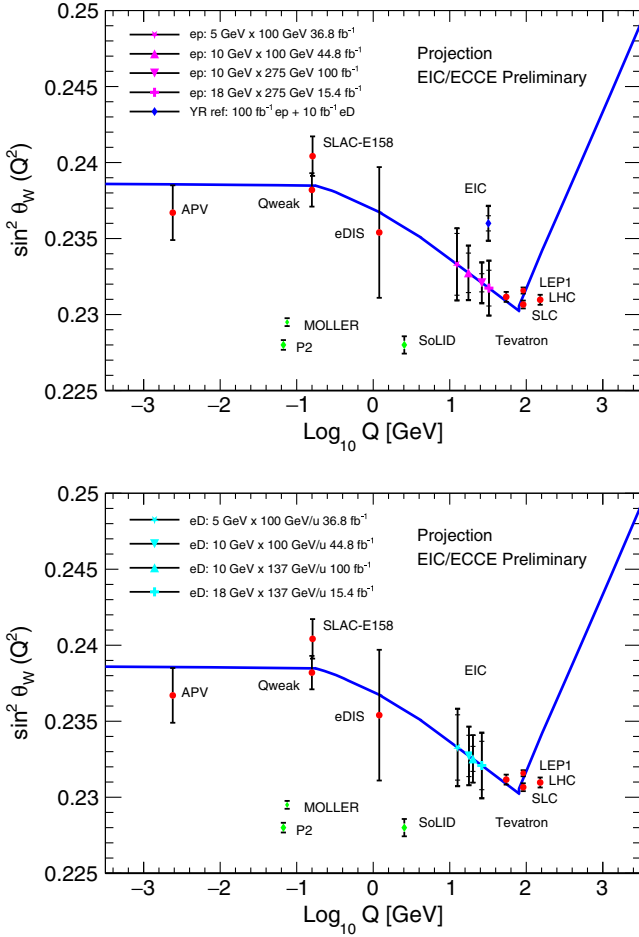


FIG. 11. Projected results for $\sin^2 \theta_W$ using ep (top, solid magenta markers) and eD (bottom, solid cyan markers) collision data and the nominal annual luminosity given in Table 10.1 of the yellow report [24], along with existing world data (red solid circles) and near-future projections (green diamonds); see text for details. Data points for Tevatron and LHC are shifted horizontally for clarity. The script used to produce this plot is inherited from [43]. The scale dependence of the weak mixing angle expected in the SM (blue curve) is defined in the modified minimal subtraction scheme ($\overline{\text{MS}}$ scheme) [33].

bars, which almost coincide with the median error bars, include all the above and the PDF uncertainty evaluated using the set CT18NLO. Results evaluated with the sets MMHT2014 and NNPDF3.1NLO are similar. Along with our projection with the EIC annual nominal luminosity, we show the “YR reference point” (blue diamond), obtained from combining $100 \text{ fb}^{-1} 18 \times 275 \text{ GeV } ep$ and $10 \text{ fb}^{-1} 18 \times 137 \text{ GeV } eD$ pseudodata. Also shown are the expected precisions from near-future P2 [37], MOLLER [38], SoLID [39], and PVDIS [40,41] experiments, respectively, that will dominate the landscape of low- to medium-energy scales.

We note that our results have larger uncertainties than in the YR [24], which fits PDFs and $\sin^2 \theta_W$ simultaneously using the Jefferson Lab Angular Momentum (JAM)

framework [42], possibly due to using realistic detector simulation and accurate running conditions. On the other hand, we find that PDF uncertainties are likely not the dominant ones for the EIC projections, but the electron polarization is, for the settings where the integrated luminosity approaches 100 fb^{-1} . Consequently, upgrading the luminosity of the EIC does not bring significant improvement on the uncertainty of $\sin^2 \theta_W$, and therefore we do not show our fitting results for the tenfold luminosity upgrade.

Our results show that the EIC will provide a determination of $\sin^2 \theta_W$ at an energy scale that bridges higher-energy colliders with low- to medium-energy SM tests. Additionally, data points of different \sqrt{s} values of the EIC can be combined or the Q^2 dependence of the EW parameter can be explored, depending on the run plan of the EIC. Furthermore, one could study the exploratory potential of the EIC beyond the scope of a single SM parameter, and we provide results using the SMEFT framework in the next section.

VI. FRAMEWORK FOR THE SMEFT ANALYSIS

A. Data generation and selection

We use the procedure described in Sec. III to determine the uncertainty of our data projection and the uncertainty matrix. We consider both ep and eD collisions and concentrate on the two highest-energy settings listed in Table II. Because collisions with higher center-of-mass energy are more sensitive to SMEFT operators, we choose four data families with the two highest \sqrt{s} to focus on:

- 10 GeV \times 137 GeV eD 100 fb^{-1} : D4, Δ D4, LD4,
- 18 GeV \times 137 GeV eD 15.4 fb^{-1} : D5, Δ D5, LD5,
- 10 GeV \times 275 GeV ep 100 fb^{-1} : P4, Δ P4, LP4,
- 18 GeV \times 275 GeV ep 15.4 fb^{-1} : P5, Δ P5, LP5.

For the highest \sqrt{s} but lower-luminosity set D5, Δ D5, P5, and Δ P5, we consider two scenarios: the nominal luminosity as indicated above and in Table II, and the high luminosity option denoted with an “HL” label with tenfold higher statistics.

We use Eq. (44) to generate $N_{\text{exp}} = 1000$ pseudo-datasets for each of the data families. We then impose the following selection criteria on the bin points, x and Q^2 , and the inelasticity, y :

$$x < 0.5, \quad Q^2 > 100 \text{ GeV}^2, \quad 0.1 < y < 0.9. \quad (56)$$

These restrictions are designed to remove large uncertainties from nonperturbative QCD and nuclear dynamics that occur at low Q^2 and high x , where sensitivity to SMEFT effects is anyway expected to be reduced. We note that the

condition on y is already applied in the data generation and unfolding stages described in Sec. III C.

B. Structure of the SMEFT asymmetry corrections

In the computation of SMEFT asymmetry values, A_{SMEFT} , we use the central member of the PDF set under consideration. We use the PDF sets NNPDF31_nlo_as_0118 [25] and NNPDFpol11_100 [26] for the computation of unpolarized and polarized PV asymmetries, namely $A_{\text{PV}}^{(e)}$ and $A_{\text{PV}}^{(H)}$, respectively. We factor out the UV cutoff scale from all the seven Wilson coefficients, $C_r \rightarrow C_r/\Lambda^2$, and set $\Lambda = 1$ TeV. We turn on only one or two Wilson coefficients at a time and set the remaining ones to zero and linearize the SMEFT expressions with respect to the Wilson coefficient(s) of interest. SMEFT asymmetry expressions then generically take the form

$$A_{\text{SMEFT}}(x, Q^2, C) = A_{\text{SM},0}^{\text{theo}}(x, Q^2) + C\delta(x, Q^2) \quad (57)$$

or

$$A_{\text{SMEFT}}(x, Q^2, C_1, C_2) = A_{\text{SM},0}^{\text{theo}}(x, Q^2) + C_1\delta_1(x, Q^2) + C_2\delta_2(x, Q^2). \quad (58)$$

Comparing Eq. (57) to (44) or (45), we see that at the end of a multirun analysis, the distribution of the best-fit values for any single Wilson coefficient should be a Gaussian centered at the origin.

C. Best-fit analysis of Wilson coefficients

Generating pseudodata values, $A_{\text{SM}}^{\text{pseudo}}$, and obtaining the SMEFT asymmetry expressions, A_{SMEFT} , we define a χ^2 test statistic as

$$\chi^2 = \sum_{b=1}^{N_{\text{bin}}} \sum_{b'=1}^{N_{\text{bin}}} [A_{\text{SMEFT},b} - A_{\text{SM},b}^{\text{pseudo}}][(\Sigma^2)^{-1}]_{bb'} \times [A_{\text{SMEFT},b'} - A_{\text{SM},b'}^{\text{pseudo}}], \quad (59)$$

where N_{bin} is the number of bins in a given dataset. Generically, it looks like

$$\chi^2(C) = k_0 + k_1 C + k_2 C^2 \quad (60)$$

for a single-parameter fit of Wilson coefficient C or

$$\chi^2(C_1, C_2) = k_{00} + k_{10}C_1 + k_{01}C_2 + k_{11}C_1C_2 + k_{20}C_1^2 + k_{02}C_2^2 \quad (61)$$

for a two-parameter fit of Wilson coefficients C_1 and C_2 . The χ^2 function is minimized with respect to C or to C_1 and C_2 . This gives us the best-fit values, \bar{C} or \bar{C}_1 and \bar{C}_2 .

We obtain the inverse square of the error of the single-parameter best-fit value via

$$\frac{1}{\sigma_C^2} = \frac{1}{2} \frac{d^2\chi^2}{dC^2} \quad (62)$$

evaluated at \bar{C} . The inverse covariance matrix, V^{-1} , of the two-parameter fit is constructed in such a way that its ij th component is given by

$$(V^{-1})_{ij} = \frac{1}{2} \frac{\partial^2\chi^2}{\partial C_i \partial C_j} \quad (63)$$

for $i, j = 1, 2$, evaluated at the best-fit values of C_1 and C_2 . Inverting V^{-1} , we obtain the individual errors and the correlation of the fit:

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix}_{\text{sym}}. \quad (64)$$

I. Averaging over multiple pseudo-datasets

When we repeat N_{exp} times the single-parameter best-fit analysis described in Sec. VI C, we obtain N_{exp} best-fit values, \bar{C}_e , with corresponding uncertainties, $\sigma_{C,e}$, for each pseudoexperiment e . The mean of the best-fit values is obtained by averaging individual best-fit values weighted by the inverse square of the uncertainties,

$$\bar{C} = \left(\sum_{e=1}^{N_{\text{exp}}} \frac{1}{\sigma_{C,e}^2} \right)^{-1} \left(\sum_{e=1}^{N_{\text{exp}}} \frac{1}{\sigma_{C,e}^2} \bar{C}_e \right), \quad (65)$$

and the average uncertainty of this mean value is obtained via

$$\frac{1}{\sigma_C^2} = \frac{1}{N_{\text{exp}}} \sum_{e=1}^{N_{\text{exp}}} \frac{1}{\sigma_{C,e}^2}. \quad (66)$$

When we repeat N_{exp} times the two-parameter best-fit analysis on Wilson coefficients described in Sec. VI C, we obtain N_{exp} pairs of best-fit values, $\bar{C}_{1,e}$ and $\bar{C}_{2,e}$, and inverse covariance matrices, $(V^{-1})_e$, for each pseudoexperiment, e . The best-fit values are averaged similarly to the one-dimensional case but with the inverse square of uncertainties replaced by inverse covariance matrices:

$$\begin{pmatrix} \bar{C}_1 \\ \bar{C}_2 \end{pmatrix} = \left[\sum_{e=1}^{N_{\text{exp}}} (V^{-1})_e \right]^{-1} \left[\sum_{e=1}^{N_{\text{exp}}} (V^{-1})_e \begin{pmatrix} \bar{C}_{1,e} \\ \bar{C}_{2,e} \end{pmatrix} \right]. \quad (67)$$

The average inverse covariance matrix of the resulting best fit is calculated using

$$V^{-1} = \frac{1}{N_{\text{exp}}} \sum_{e=1}^{N_{\text{exp}}} (V^{-1})_e. \quad (68)$$

We note the presence of the factor $1/N_{\text{exp}}$ in Eqs. (66) and (68). Without it, we would be effectively increasing the luminosity of the corresponding central dataset by a factor N_{exp} . We avoid this by including this factor in computing the average uncertainty or inverse covariance matrix.

2. Definition of confidence intervals

The result of a single-parameter multirun fit can be expressed as

$$\frac{(C - \bar{C})^2}{\sigma_C^2} = \Delta\chi^2, \quad (69)$$

and hence we can express the fitted result and the uncertainty of coefficient C as

$$C = C_{\text{best}} \pm \sqrt{\Delta\chi^2} \sigma_C. \quad (70)$$

For a two-parameter multirun fit, the ellipse equation reads

$$\begin{pmatrix} C_1 - \bar{C}_1 \\ C_2 - \bar{C}_2 \end{pmatrix}^T V^{-1} \begin{pmatrix} C_1 - \bar{C}_1 \\ C_2 - \bar{C}_2 \end{pmatrix} = \Delta\chi^2 \quad (71)$$

in the (C_1, C_2) plane.

The $\Delta\chi^2$ values that determine the size of the best-fit region for an arbitrary confidence level are well-known. For 95% C.L., we have $\Delta\chi^2 = 3.841, 5.991,$ and 7.815 for one-, two-, and three-parameter fits, respectively.

3. Combination of best fits from distinct datasets

Suppose we have two datasets, say T1 and T2, from which we obtain the single-parameter best-fit values of Wilson coefficient, C , written as \bar{C}_{T1} and \bar{C}_{T2} , together with the errors $\sigma_{C,T1}$ and $\sigma_{C,T2}$. Assuming these datasets can be treated as uncorrelated to a good approximation, we obtain

the combined best-fit value and the corresponding uncertainty by using Eqs. (65) and (66) with slight modifications. First, the summation index e now runs from 1 to 2, representing the number of datasets combined. Second, the factor $1/N_{\text{exp}}$ should be removed from Eq. (66) because we now have indeed two independent, uncorrelated measurements. This method can be generalized to the combination of the best-fit values from more than two datasets, such as different beam energies, and to the case of multi-parameter fits in a straightforward manner.

4. Simultaneous fit of Wilson coefficients and beam polarization or luminosity difference

We observe in Sec. IV D that experimental uncertainties such as the beam polarization and luminosity difference between e^+ and e^- runs can be limiting factors for some of the datasets. When the data statistical uncertainty is very precise, there is the possibility that one can use the data themselves to constrain these systematic effects. We present in Appendix A 1 a method to simultaneously fit the SMEFT coefficient(s) and the luminosity difference for the LC asymmetries and in Appendix A 2 a method to simultaneously fit the SMEFT coefficient(s) and the beam polarization for PV asymmetries.

VII. SMEFT FIT RESULTS

A. Fits of single Wilson coefficients

In this section, we discuss the 95% C.L. intervals for the Wilson coefficients in single-parameter fits averaged over 1000 pseudoexperiments. The bounds on the Wilson coefficient C_{eu} across numerous datasets are representative and exhibit the common features of fits of single Wilson coefficients. We therefore show only the bounds on C_{eu} to illustrate the main observations and include the remaining Wilson coefficients in Appendix B 1. Figure 12 displays the

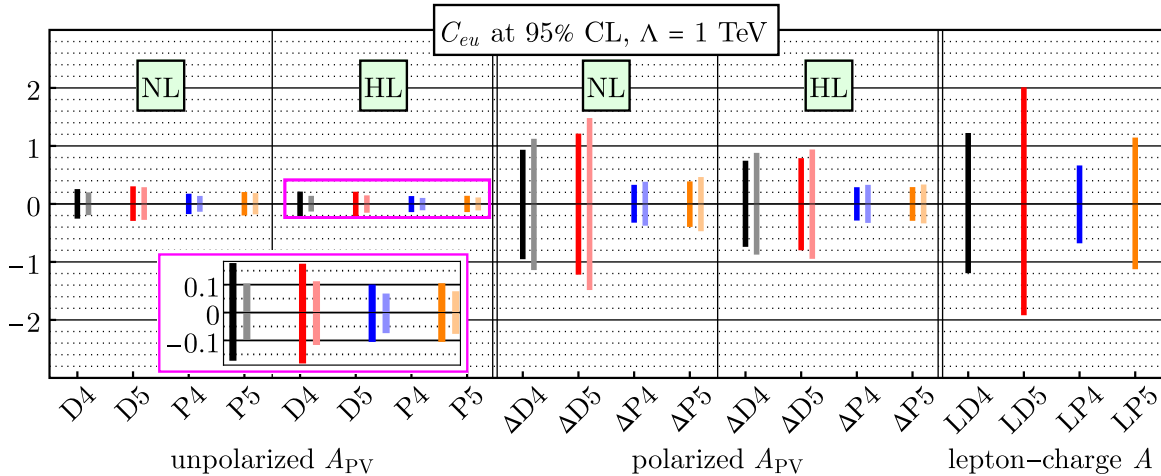


FIG. 12. The 95% C.L. bounds of C_{eu} from single-parameter fits (darker) and from the $(1 + 1)$ -parameter fits with beam polarization as an additional fitting parameter (lighter) using the families of datasets D4, D5, P4, and P5 at $\Lambda = 1$ TeV.

95% C.L. intervals of C_{eu} for the four data families in which we are primarily interested in this paper. The intervals are grouped by asymmetries, namely electron PV asymmetries, $A_{pV}^{(e)}$, of unpolarized hadrons (“unpolarized A_{pV} ”); hadron PV asymmetries, $A_{pV}^{(H)}$, with unpolarized electrons (“polarized A_{pV} ”); and unpolarized electron-positron asymmetries, $A_{LC}^{(H)}$, of unpolarized hadrons (“lepton-charge A ”). PV asymmetries are then grouped into two, showing the fits in the nominal- and high-luminosity scenarios. In each block of intervals, there are four double lines in the case of PV asymmetries and four single lines in LC asymmetries. These four lines correspond to the data families D4 (black and its shade), D5 (red), P4 (blue), and P5 (orange), respectively. The darker of the two lines indicates the bounds from single-parameter fits with the Wilson coefficient C_{eu} , whereas the lighter ones show the bounds on the Wilson coefficient from simultaneous $(1+1)$ -parameter fits with C_{eu} and the beam polarization. We describe the details of the fits involving the beam polarization as an additional free parameter in Appendix A 2.

From Fig. 12, we can extract the following main points:

- (i) Proton asymmetries of all the three types, namely unpolarized and polarized PV asymmetries and LC asymmetries, impose considerably stronger bounds than deuteron.
- (ii) High-energy low-luminosity datasets D5 and P5 lead to slightly weaker bounds than the less-energetic but higher-luminosity ones, D4 and P4, respectively.
- (iii) Unpolarized PV asymmetries, $A_{pV}^{(e)}$, offer much stricter bounds than the polarized ones $A_{pV}^{(H)}$; however, it should be noted that for some Wilson coefficients, unpolarized proton asymmetries yield nearly the same bounds as the corresponding polarized ones.
- (iv) Datasets in the high-luminosity scenario make a noticeable difference in the size of bounds. The

improvement due to increased luminosity is slightly more significant for polarized deuteron asymmetries.

- (v) Bounds from electron-positron asymmetries, $A_{LC}^{(H)}$, are comparable to or looser than the ones from polarized hadron asymmetries. They never offer stricter bounds than high-luminosity hadron PV asymmetries.
- (vi) If the beam polarization is introduced as a new fitting parameter, unpolarized hadron asymmetries give considerably stronger bounds. The improvement is more significant in the high-luminosity scenario. However, the same fitting method yields weaker bounds with polarized hadron asymmetries. We explain this finding in Appendix A 2.

Assuming weak correlations, one can also combine the bounds within a given family of datasets, e.g., D4, Δ D4, and LD4. We find that the resultant bound is never stronger than the strongest one obtained from the individual family members, which is always from the electron PV asymmetry dataset. We note that this observation holds for other Wilson coefficient choices and for the two-dimensional fits in the next section, as well.

In Fig. 13, we present the effective UV cutoff scales, $\Lambda/\sqrt{C_{eu}}$, with $\Lambda = 1$ TeV, corresponding to the bounds shown in Fig. 12. The organization of this plot in terms of asymmetries and datasets is the same as in Fig. 12. Improved bounds on C_{eu} with the addition of the beam polarization to the fits are equivalent to higher-energy scales in the unpolarized PV asymmetries, which are indicated by the lighter columns in the background; on the other hand, weaker bounds from the fits with beam polarization are depicted by the lighter columns in the foreground for the polarized PV asymmetries.

One can observe that scales reaching 3 TeV can be probed with nominal luminosity, while scales exceeding 4 TeV can be probed for other luminosities. We remark that care must be taken in comparing these mass limits with

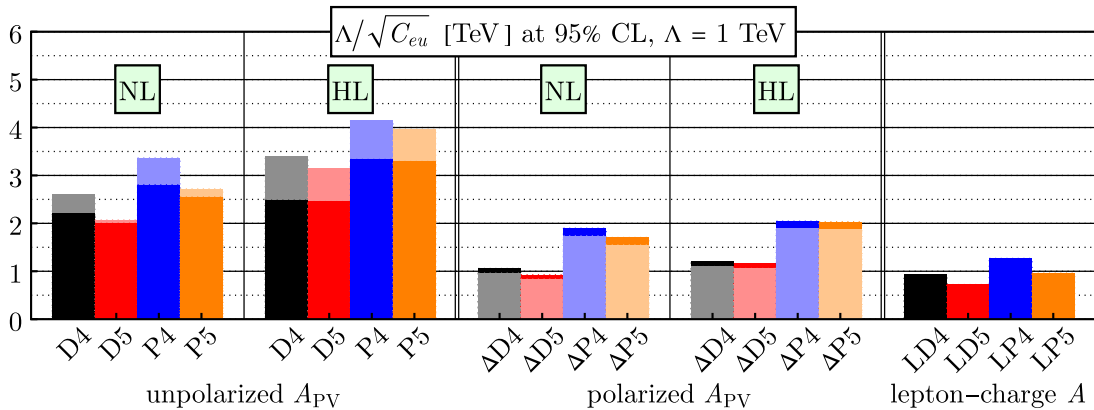


FIG. 13. Effective UV cutoff scales, $\Lambda/\sqrt{C_{eu}}$, defined in terms of the 95% C.L. bounds on the Wilson coefficient C_{eu} and with $\Lambda = 1$ TeV.

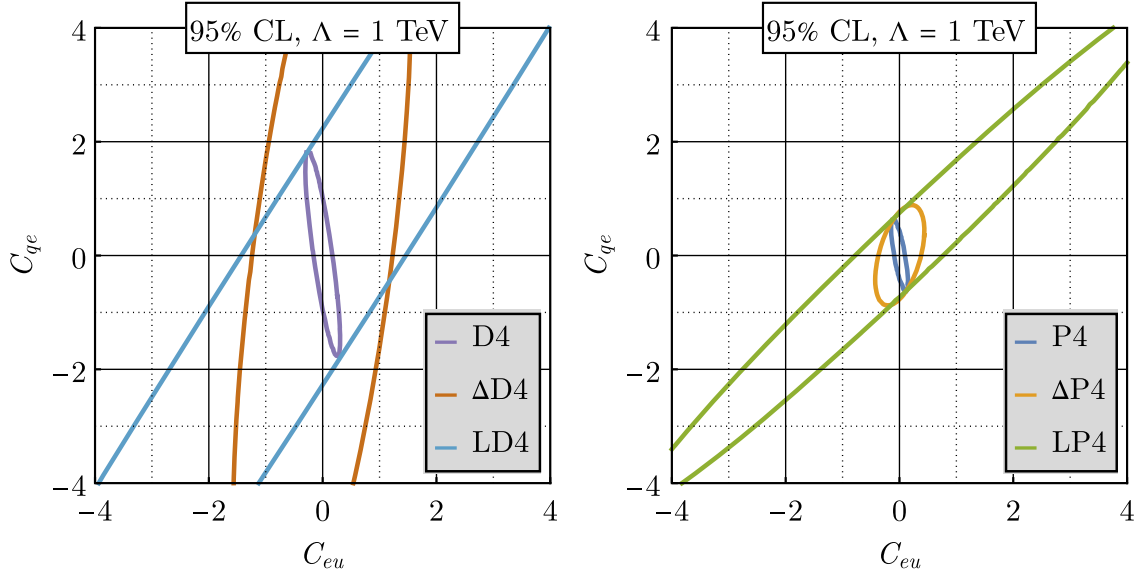


FIG. 14. The 95% C.L. ellipses for the Wilson coefficients C_{eu} and C_{qe} using the families of datasets D4 and P4 in the simultaneous $(2 + 1)$ -parameter fits that include the beam polarization as an additional fitting parameter.

others found in the literature, which sometimes assume a strong coupling limit, equivalent to setting $C_r = 4\pi$, and also maximally constructive interference between different quark contributions. For example, converting our results to the notation of [44] would yield a bound on $\Lambda/\sqrt{C_{eu}}$ of 19 TeV, instead of 3 TeV quoted here, which is only very approximate and is calculated by multiplying 3 TeV by $\sqrt{4\pi}$ and $\sqrt{\sqrt{5}}$, where the latter is to account for the constructive interference between quark contributions, and by another factor to convert 90% C.L. to 95% C.L.

B. Fits of two Wilson coefficients

In this section, we discuss fits on pairs of Wilson coefficients in order to determine how well the EIC can break degeneracies between parameters that occur in the LHC Drell-Yan data [6,45]. We emphasize that the representative examples shown in this section are the results of the simultaneous fits with beam polarization in light of the significantly improved results of the $(1 + 1)$ -parameter fits in the previous section. The description of the beam-polarization fits is presented in Appendix A 2. The complete set of plots of confidence ellipses is given in Appendix B 2.

In Fig. 14, we compare the 95% C.L. ellipses for the pair (C_{eu}, C_{qe}) between the data families D4 and P4. Each asymmetry type gives a distinct correlation pattern, complementary to one another. Electron-positron asymmetries give rise to wide and elongated bandlike ellipses compared to PV asymmetries. As in the case of $(1 + 1)$ -parameter fits, electron PV asymmetries of unpolarized hadrons offer the strongest bounds on the pairs of Wilson coefficients. Comparing deuteron to proton, one can see that proton data are significantly more constraining.

Figure 15 shows the comparison of the simultaneous fit of the Wilson coefficients $(C_{eu}, C_{\ell u})$ projected for the EIC to the corresponding fit with the LHC data adapted from [45]. The LHC fit exhibits a flat direction; i.e., a particular linear combination of the two coefficients cannot be determined. A similar comparison is given in Fig. 16 for the pair $(C_{eu}, C_{\ell q}^{(1)})$, using the nominal- and high-luminosity P4 set of the EIC. We observe that in both

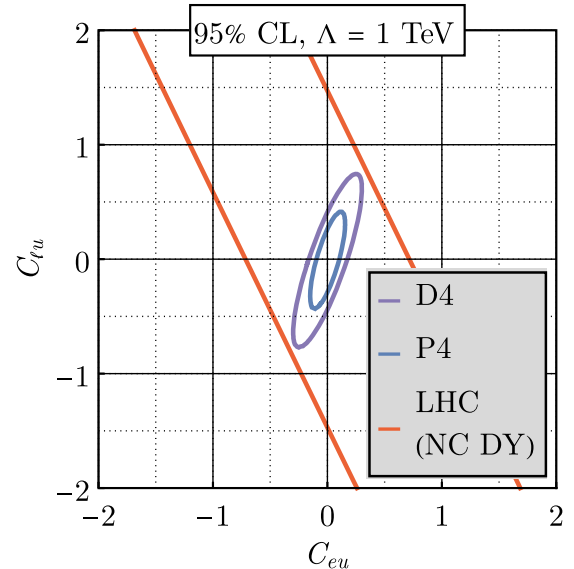


FIG. 15. The 95% C.L. ellipses for the Wilson coefficients C_{eu} and $C_{\ell u}$ using the datasets D4 and P4 in the $(2 + 1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [45].

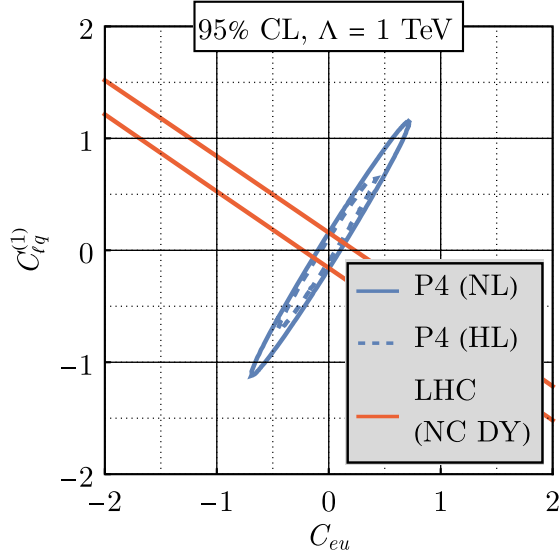


FIG. 16. The 95% C.L. ellipses for the Wilson coefficients C_{eu} and $C_{\ell q}^{(1)}$ using the nominal- and high-luminosity dataset P4 in the $(2+1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding two-parameter fit from the LHC data [6].

figures, projected EIC fits have different correlation patterns from the LHC. More importantly, the EIC projected data show the capability of resolving flat directions and significantly constraining the aforementioned pairs of Wilson coefficients.

Finally, in Fig. 17, we present the fits from the P4 dataset and the LHC adapted from [6] for the pair $(C_{\ell q}^{(1)}, C_{\ell q}^{(3)})$. This figure shows that when the LHC data impose tight bounds on a pair of Wilson coefficients, the EIC preliminary data can introduce far stronger bounds on the same pair of Wilson coefficients. Moreover, the fits from EIC and LHC have distinct correlations, which indicate the complementarity of the EIC to the LHC as a future collider. Treating the projected EIC and the LHC data to be uncorrelated, we also plot the combined fit of the two, which turns out to even more strongly constrain the chosen pair of Wilson coefficients. We remark that the effective UV scales probed with the combined dataset exceed 2 TeV.

It should be noted that there appear flat directions in the fits of certain pairs of Wilson coefficients with the projected EIC data that utilize the deuteron beam. Examples include (C_{eu}, C_{ed}) and $(C_{\ell u}, C_{\ell d})$. We can explain these observations analytically. We find that these pairs always appear in a specific way in asymmetry expressions, for example, $2C_{eu} - C_{ed}$ for electron PV asymmetries with unpolarized deuteron. In all such cases, only one of the data families exhibits this behavior, with the degeneracy broken by another data family.

Our results on the bounds from Wilson coefficients in simultaneous $(2+1)$ -parameter fits with the beam

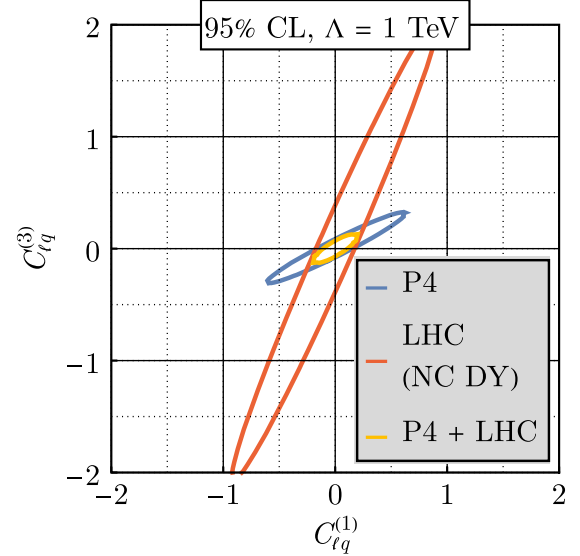


FIG. 17. The 95% C.L. ellipses for the Wilson coefficients $C_{\ell q}^{(1)}$ and $C_{\ell q}^{(3)}$ using the nominal-luminosity dataset P4 in the $(2+1)$ -parameter fit that includes the beam polarization as an additional fitting parameter, compared with the corresponding fit from the LHC data [6] and the combined fit of the two.

polarization as an additional parameter can be summarized as follows:

- (i) Proton asymmetries impose much stricter bounds than deuteron.
- (ii) Unpolarized hadron asymmetries lead to stronger correlations than the polarized ones.
- (iii) The three types of asymmetries of deuteron and proton considered in this work, together with the LHC data, are complementary to each other in the sense that they offer distinct correlation patterns.
- (iv) The projected EIC data are capable of resolving all flat directions that appear in the LHC Drell-Yan data.
- (v) The bounds from the projected EIC data can be much stronger than the LHC data, indicating that the EIC has an important role to play in future probes of the SMEFT.

We can ask what happens when more than two Wilson coefficients are turned on simultaneously. We study this in Sec. B 3, where we turn on six Wilson coefficients. The resulting bounds are 20% to 30% weaker than the ones found here in the $2d$ case. We note that no flat directions appear in these fits, indicating that the EIC can fully probe this parameter space without degeneracies.

VIII. CONCLUSIONS

In this manuscript, we have analyzed the potential of testing the electroweak SM and exploring BSM physics with the future EIC. We have focused on the precision determination of the weak mixing angle over a wide range of momentum transfer and on probes of heavy new physics.

We have provided all the formulas for neutral-current DIS and simulation details that will be needed for future studies of these areas. Our BSM analysis utilizes the model-independent SMEFT framework and focuses on the semileptonic four-fermion operator sector of the theory. We translate our formalism into the DIS language in terms of parity-violating couplings and structure functions to facilitate crosstalk between the high-energy-physics and nuclear communities. We provide a detailed accounting of uncertainties from statistics, experimental systematic effects, beam polarimetry for parity-violating asymmetries, higher-order QED corrections for lepton-charge asymmetries, and finally PDFs. Additionally, we explore simultaneously fitting the beam polarization with the anticipated high-precision parity-violating asymmetry data as a possible analysis technique to improve upon the experimental limitation from beam polarimetry.

Our BSM analysis finds that UV scales in excess of 3 TeV can be probed with the currently planned (nominal) annual luminosity of the EIC, with scales above 4 TeV possible if a tenfold luminosity upgrade becomes available beyond EIC's initial decade of running. For the latter, we focus on studying the physics reach and limitations from sources other than statistical, without comment as to the feasibility of such an upgrade. The most stringent bounds come from polarized electron scattering off of unpolarized protons. Constraints from polarized hadrons, deuterons, and from a possible future positron beam provide important complementary probes. Our complete study of correlations between Wilson coefficients finds that no degeneracies remain upon combining all EIC datasets. This is not the case with LHC Drell-Yan measurements, in which numerous degeneracies exist, and will continue to occur even after LHC's high luminosity running.

This demonstrates that the EIC polarization provides a powerful probe of BSM effects. Although the EIC is primarily thought of as a QCD machine, it is also a powerful probe of potential new physics, in many ways complementary to the higher-energy LHC. We note that current global fits of LHC data, for example to top-quark and Higgs data, probe orthogonal sets of Wilson coefficients [46]. However, the strongest LHC constraints on the semileptonic four-fermion sector of the SMEFT come from the Drell-Yan data. Given that LHC Drell-Yan cross sections are blind to certain combinations of Wilson coefficients, we envision that high-precision EIC data will help remove these degeneracies in global fits, both at the present moment and at the time when high-luminosity LHC Drell-Yan data become available. We hope that our work motivates future studies of the unexpected power of the EIC for new-physics searches.

ACKNOWLEDGMENTS

R. B. is supported by the U.S. Department of Energy (DOE) Contract No. DE-AC02-06CH11357. T. K. is

supported by the DOE Grant No. DE-SC0020240 and the Zuckerman STEM Leadership Program. F. P., K. S., and D. W. are supported by the DOE Grants No. DE-FG02-91ER40684 and No. DE-AC02-06CH11357. X. Z. and M. N. are supported by the DOE Grant No. DE-SC0014434. The authors thank the ECCE Consortium for performing a full simulation of their detector design, for providing up-to-date information on EIC run conditions, and for suggestions and comments on the manuscript. X. Z. thanks H. Spiesberger for suggestions on the use of the DJANGO generator and useful discussions related to the analysis.

APPENDIX A: ADDITIONAL FITS

1. Luminosity difference fits

Since electron and positron data would be taken at different times with different beam configurations, there is the possibility of a significant offset between the absolute luminosities of the two datasets. In the main text, we include this uncertainty in the error matrix as the luminosity error, $\sigma_{\text{lum}} = 0.02$, which is assumed absolute. We study here the possibility of simultaneously fitting this luminosity difference together with the Wilson coefficients.

We fit the pseudodata for the LC asymmetries with an overall shift, A_{lum} , added to the pseudodata. Then, we define the χ^2 test statistics as

$$\chi^2 = \sum_{b=1}^{N_{\text{bin}}} \sum_{b'=1}^{N_{\text{bin}}} [A_{\text{SMEFT},b} - A_{\text{SM},b}^{\text{pseudo}}][(\tilde{\Sigma}^2)^{-1}]_{bb'} \times [A_{\text{SMEFT},b'} - A_{\text{SM},b'}^{\text{pseudo}}], \quad (\text{A1})$$

where we omit the uncertainty in the luminosity difference between e^+ and e^- runs from the uncertainty matrix:

$$\tilde{\Sigma}^2 = \Sigma^2|_{\sigma_{\text{lum}} \rightarrow 0}. \quad (\text{A2})$$

However, we keep the luminosity uncertainty in the pseudodata generation. By introducing the luminosity difference, A_{lum} , as a new parameter, we extend our one-parameter and two-parameter Wilson-coefficient fits to $(1+1)$ - and $(2+1)$ -parameter fits.

We find that there are mild correlations, $|\rho_r| \lesssim 0.4$, between A_{lum} and any C_r in the $(1+1)$ - and $(2+1)$ -parameter fits. In addition, the fitted results for Wilson coefficients have slightly larger uncertainties when the luminosity difference is treated as a fitting parameter. In Fig. 18, we show the 95% C.L. intervals with and without A_{lum} for the Wilson coefficient C_{eu} in all the four LC asymmetry datasets of interest. In Fig. 19, we compare the 95% C.L. ellipses of the Wilson coefficients (C_{eu}, C_{qe}) between the datasets LD4 and LP5 with and without the luminosity difference as a fitted parameter. From these figures, we see that the 95% C.L. bounds on C_{eu} become

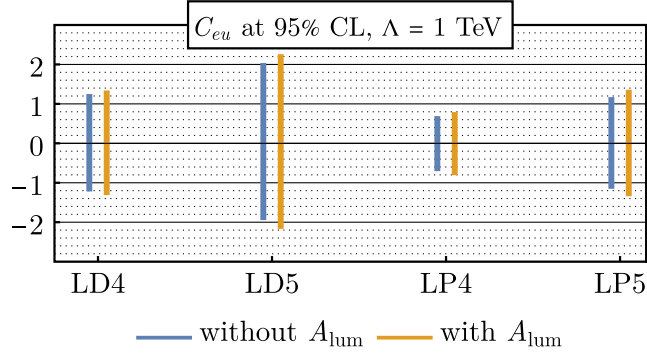


FIG. 18. Comparison of the bounds on the Wilson coefficient C_{eu} with all the LC asymmetry datasets of interest in the absence and presence of the luminosity difference as a new free fitting parameter.

15% to 20% weaker. The difference is less noticeable in the confidence ellipses.

2. Beam polarization fits

In the same spirit as the previous section, we now consider fitting the beam polarization simultaneously with the Wilson coefficients in an attempt to reduce the uncertainty associated with the experimental limitation from beam polarimetry. We fit the pseudodata for the PV asymmetries by including a factor of P in the SMEFT asymmetries. We then define a χ^2 test statistics as

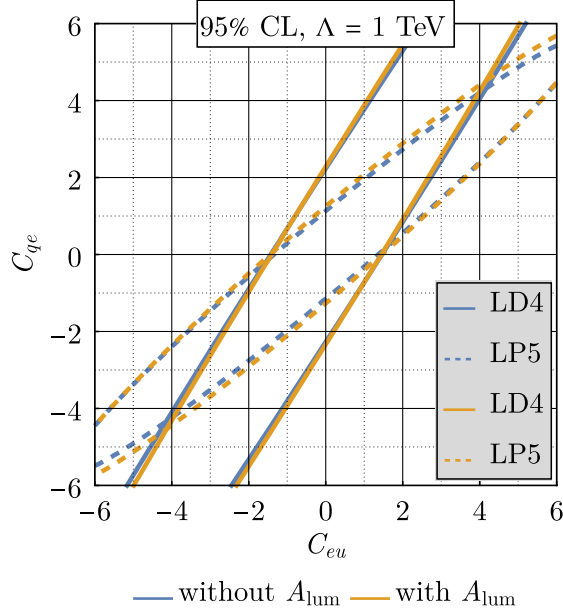


FIG. 19. Comparison of the 95% C.L. ellipses for the Wilson coefficients (C_{eu}, C_{qe}) with the datasets LD4 and LP5 in the absence and presence of the luminosity difference as an additional fitting parameter.

$$\chi^2 = \sum_{b=1}^{N_{\text{bin}}} \sum_{b'=1}^{N_{\text{bin}}} [PA_{\text{SMEFT},b} - A_{\text{SM},b}^{\text{pseudo}}][(\tilde{\Sigma}^2)^{-1}]_{bb'} \times [PA_{\text{SMEFT},b'} - A_{\text{SM},b'}^{\text{pseudo}}] + \frac{(P - \bar{P})^2}{\delta P^2}. \quad (\text{A3})$$

In this approach, we omit the beam polarization uncertainty, σ_{pol} , from the uncertainty matrix because it is now treated as a fitting parameter,

$$\tilde{\Sigma}^2 = \Sigma^2|_{\sigma_{\text{pol}} \rightarrow 0}, \quad (\text{A4})$$

but not during pseudodata generation. The second term on the RHS of Eq. (A3) is added by hand, where \bar{P} and δP are the beam polarization value and its uncertainty provided by the polarimetry, respectively, presumably uncorrelated with the asymmetry measurements. The logic behind this addition is that experimentally, the polarimetry does provide knowledge on the beam polarization, but we hope to obtain a better determination of the polarizations within the uncertainty provided by the polarimetry by fitting data with high statistical precision. As for the beam polarization itself, we use a normalized value of $\bar{P} = 1$ in this study for simplicity. Treating the new term to be the contribution of a new observable, we increase the degrees of freedom of the χ^2 distribution by 1. As in the case of luminosity difference fits, we extend our one- and two-parameter fits of Wilson coefficients to (1 + 1)- and (2 + 1)-parameter simultaneous fits by including the beam polarization as a new parameter.

From (1 + 1)-parameter fits, we find that P and any C_r are rather weakly correlated, $|\rho_r| \lesssim 0.1$, in the polarized hadron datasets, whereas there are strong correlations, $|\rho_r| \gtrsim 0.7$, in the unpolarized hadron asymmetries. We observe similar correlations in the (2 + 1)-parameter fits.

In Fig. 20, we present the allowed intervals of the Wilson coefficient C_{eu} for the nominal- and high-luminosity datasets P4 and $\Delta P4$, while Fig. 21 displays the 95% C.L. ellipse of the Wilson coefficients (C_{eu}, C_{qe}) for the same datasets in the nominal-luminosity scenario. We find that

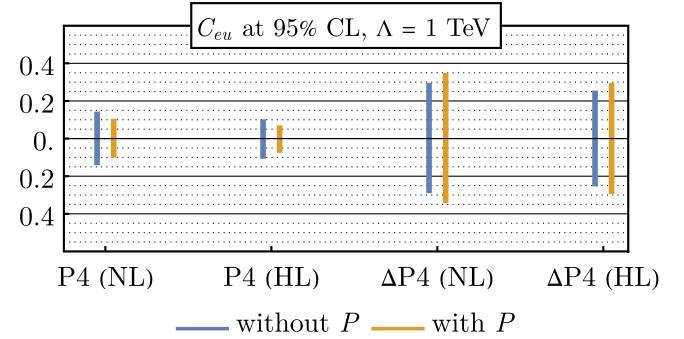


FIG. 20. The 95% C.L. bounds on the Wilson coefficient C_{eu} with the nominal- and high-luminosity datasets P4 and $\Delta P4$ in the absence and presence of the beam polarization, P , as an additional parameter in the fits.

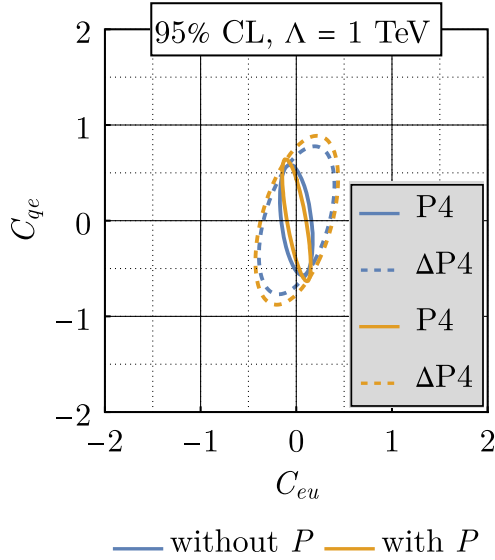


FIG. 21. The 95% C.L. ellipse of the Wilson coefficients C_{eu} and C_{qe} for the datasets P4 and $\Delta P4$ in the absence and presence of the beam polarization, P , as a new parameter in the fits.

bounds from unpolarized hadron datasets become stronger by 30% to 50%, yet the ones from polarized hadron asymmetries become 15% to 20% weaker. The improvement is sharper in the high-luminosity unpolarized hadron sets, whereas the worsening is significant for the nominal-luminosity polarized hadron sets.

One can explain why the bounds become weaker in the polarized hadron sets by referring to the correlations. Since in these datasets, the beam polarization and the Wilson coefficients are found to be weakly correlated, one would naively expect the bounds obtained from single-parameter fits of Wilson coefficients to roughly remain the same on the grounds that P and C_k can be thought of as almost fully independent so that they will not affect each other in the fits. Thus, any increase in the allowed limits of the Wilson coefficient can be attributed to the increase in the number of

parameters fitted, which is reflected as the normalization of the uncertainties of the fit.

APPENDIX B: COMPLETE SET OF FITTED RESULTS ON WILSON COEFFICIENTS

1. Fits of single Wilson coefficients

In this section, we present the 95% C.L. intervals and the corresponding effective UV cutoff scales for all the seven Wilson coefficients in single-parameter fits averaged over 1000 pseudoexperiments. We recall the following abbreviations for the EIC preliminary datasets:

- (1) electron PV asymmetries of unpolarized deuteron, $A_{PV}^{(e)}$:
 - (a) D4: eD 10 GeV \times 137 GeV, 100 fb $^{-1}$
 - (b) D5: eD 18 GeV \times 137 GeV, 15.4 fb $^{-1}$
- (2) electron PV asymmetries of unpolarized proton, $A_{PV}^{(e)}$:
 - (a) P4: ep 10 GeV \times 275 GeV, 100 fb $^{-1}$
 - (b) P5: ep 18 GeV \times 275 GeV, 15.4 fb $^{-1}$
- (3) hadron PV asymmetries with unpolarized electron, $A_{PV}^{(H)}$: $\Delta D4$, $\Delta D5$, $\Delta P4$, and $\Delta P5$ with the same energy and luminosity configuration as the corresponding D- and P-sets.
- (4) unpolarized electron-positron asymmetries of unpolarized hadrons, $A_{LC}^{(H)}$: LD4, LD5, LP4, and LP5 with the same energy configuration as the corresponding D- and P-sets, but with the luminosity of the positron beam assumed to be 10 times smaller than that of the electron beam.

Figures 22–28 display the 95% C.L. bounds of each Wilson coefficient for the four primary data families. As in the main part of the manuscript, the intervals are grouped by asymmetries, namely electron PV asymmetries of unpolarized hadrons, $A_{PV}^{(e)}$ (“unpolarized A_{PV} ”), hadron PV asymmetries with unpolarized electrons, $A_{PV}^{(H)}$ (“polarized A_{PV} ”),

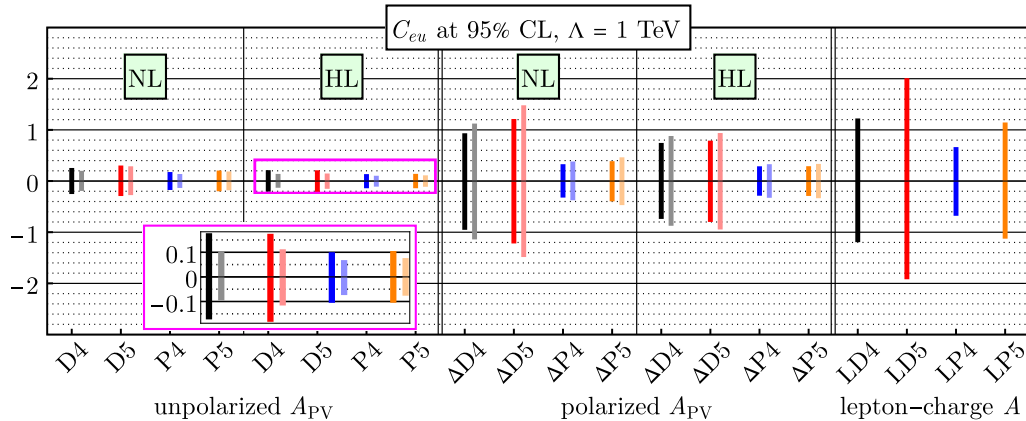


FIG. 22. The 95% C.L. bounds of C_{eu} from one-parameter fits (darker) and from simultaneous (1 + 1)-parameter fits with beam polarization (lighter) using the families of datasets D4, D5, P4, and P5.

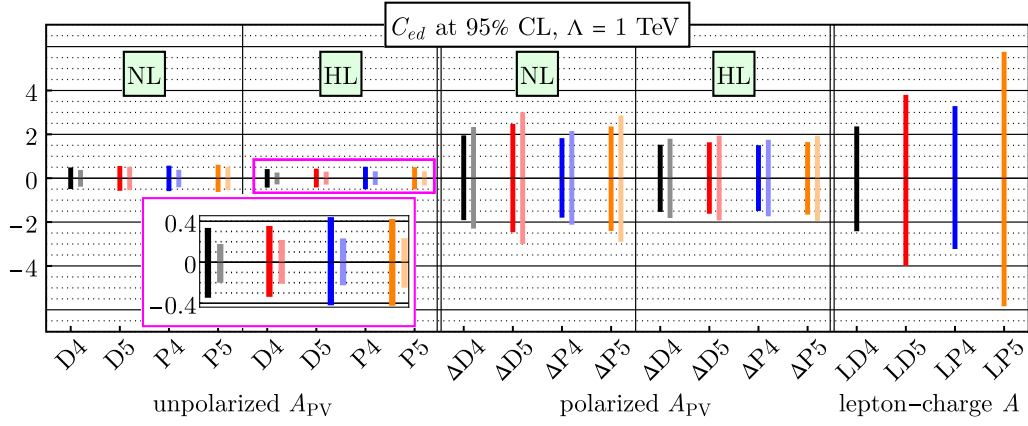


FIG. 23. The same as in Fig. 22 but for C_{ed} .

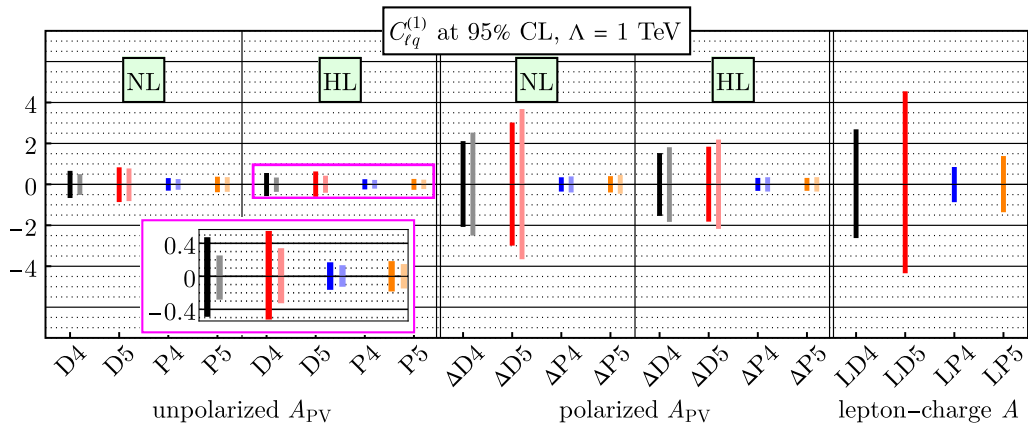


FIG. 24. The same as in Fig. 22 but for $C_{lq}^{(1)}$.

and unpolarized electron-positron asymmetries of unpolarized hadrons, $A_{LC}^{(H)}$ (“lepton-charge A ”). PV asymmetries are then grouped into two, showing the fits in the nominal- and high-luminosity scenarios. The NL refers to the annual integrated luminosity of Table 10.1 of the YR [24]. The HL is assumed to be 10 times higher than the nominal one

and requires a luminosity upgrade of the EIC. In each block of intervals, there are four double lines in the case of PV asymmetries and four single lines in LC asymmetries. These four lines correspond to the data families D4 (black and its shade), D5 (red), P4 (blue), and P5 (orange). The darker of the two lines indicates the bounds from

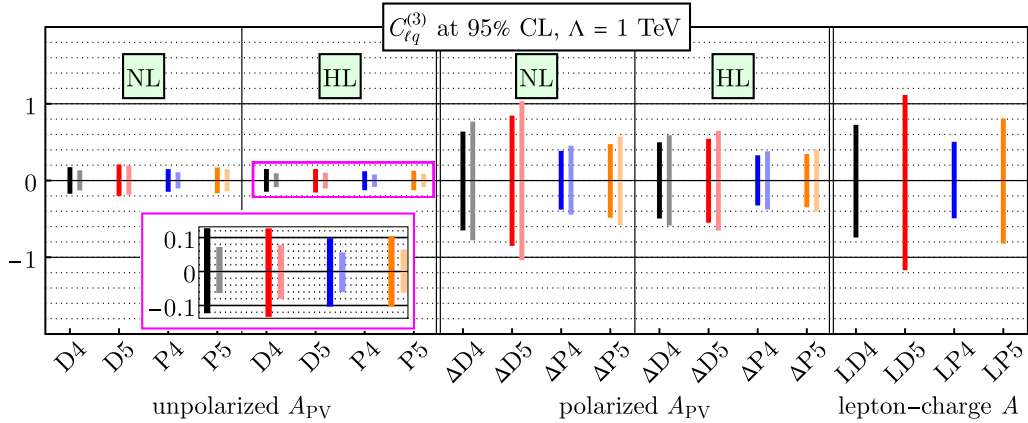


FIG. 25. The same as in Fig. 22 but for $C_{lq}^{(3)}$.

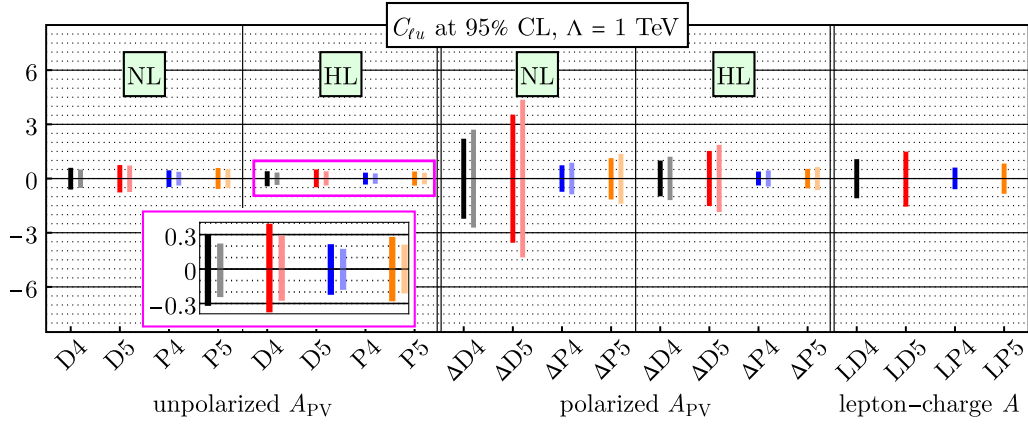


FIG. 26. The same as in Fig. 22 but for $C_{\ell u}$.

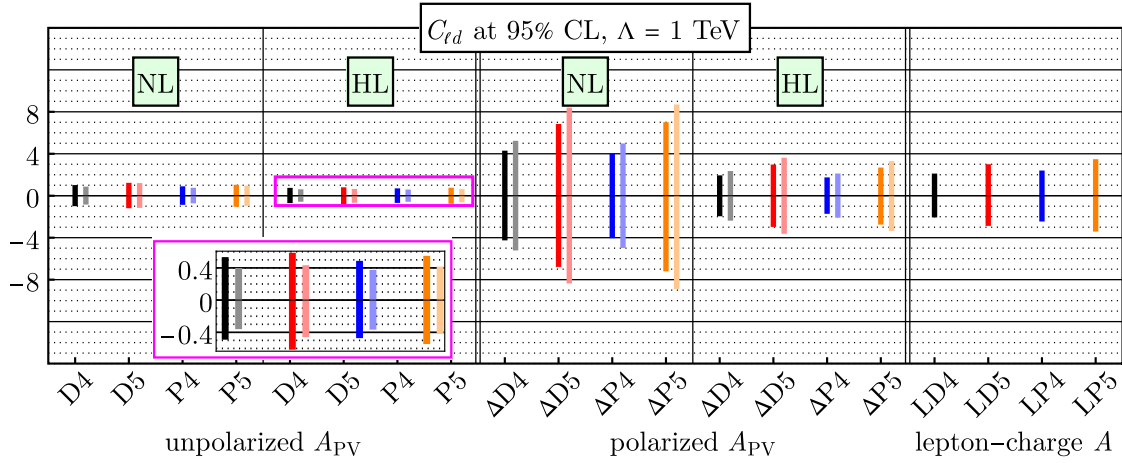


FIG. 27. The same as in Fig. 22 but for $C_{\ell d}$.

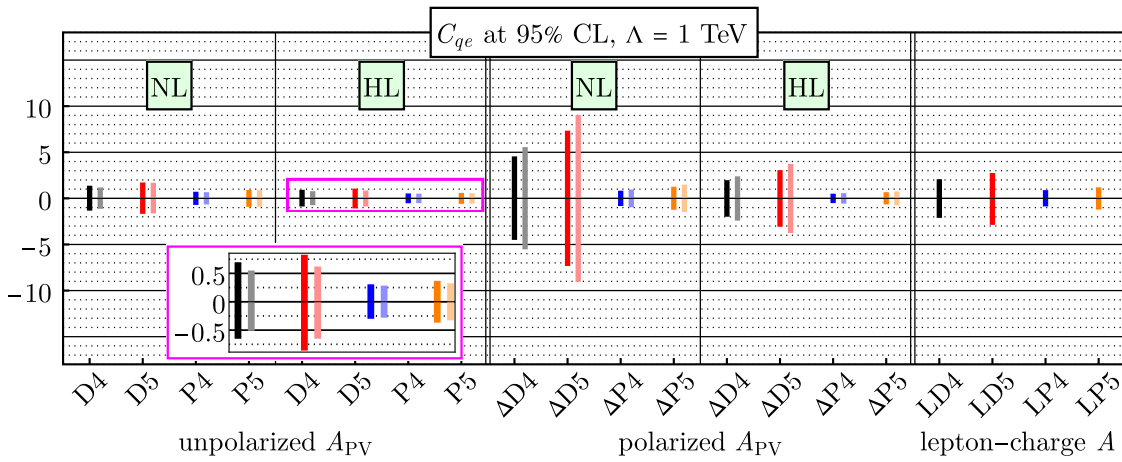


FIG. 28. The same as in Fig. 22 but for C_{qe} .

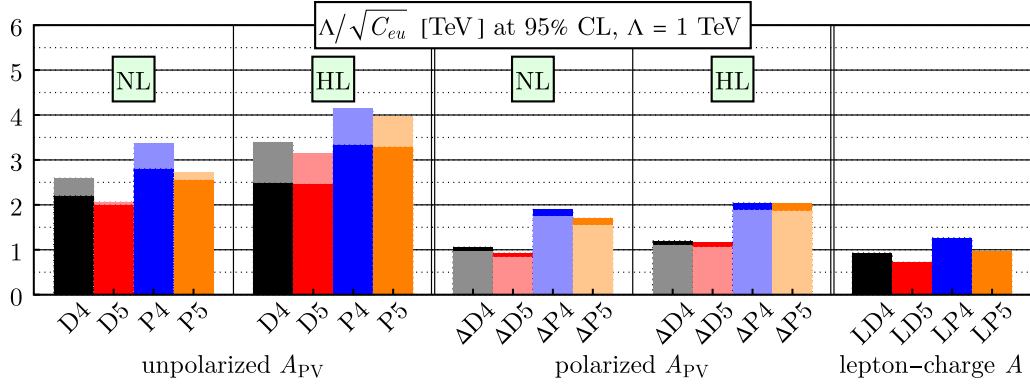


FIG. 29. Effective UV cutoff scales, $\Lambda/\sqrt{C_{eu}}$, defined in terms of the 95% C.L. bounds on the Wilson coefficient C_{eu} with $\Lambda = 1$ TeV. The darker columns in the foreground of unpolarized PV asymmetries and in the background of polarized PV asymmetries indicate the results of single-parameter fits on the Wilson coefficient, C_{eu} . The lighter columns in the background of unpolarized PV asymmetries and in the foreground of polarized PV asymmetries denote the results of simultaneous (1 + 1)-parameter fits of C_{eu} with the beam polarization, P .

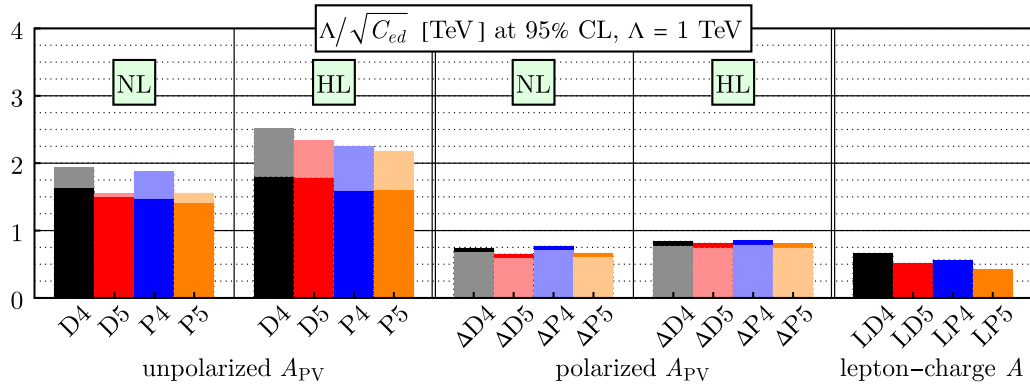


FIG. 30. The same as in Fig. 29 but for C_{ed} .

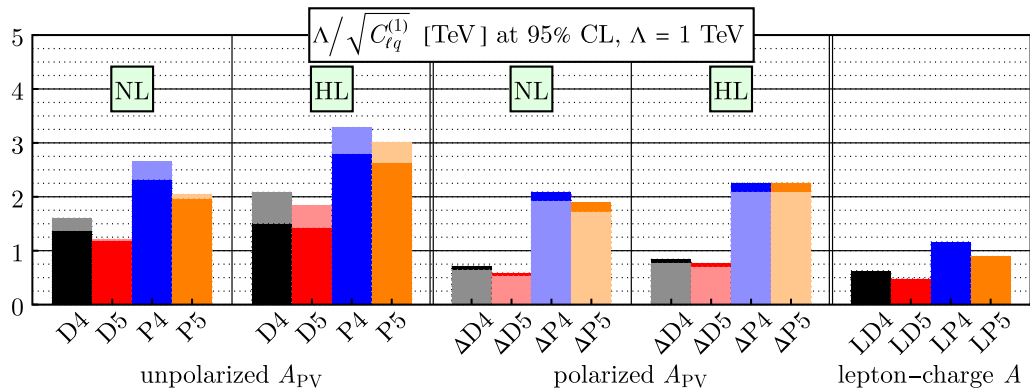


FIG. 31. The same as in Fig. 29 but for $C_{\ell q}^{(1)}$.

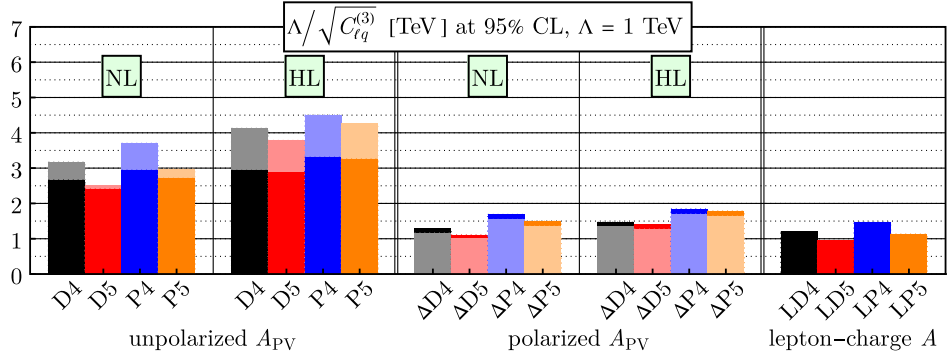


FIG. 32. The same as in Fig. 29 but for $C_{\ell q}^{(3)}$.

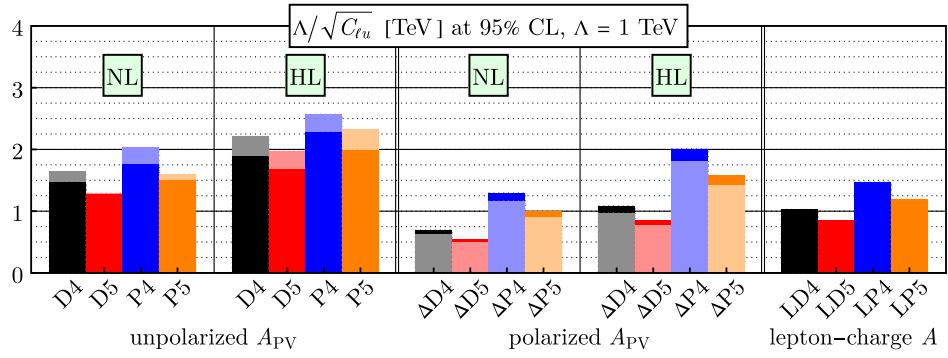


FIG. 33. The same as in Fig. 29 but for $C_{\ell u}$.

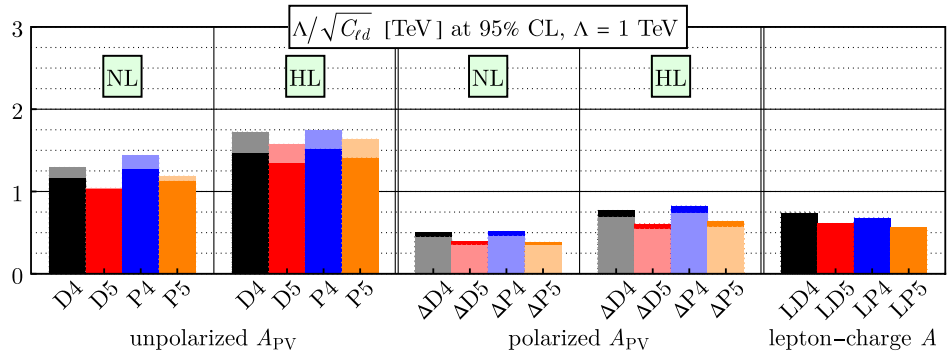


FIG. 34. The same as in Fig. 29 but for $C_{\ell d}$.

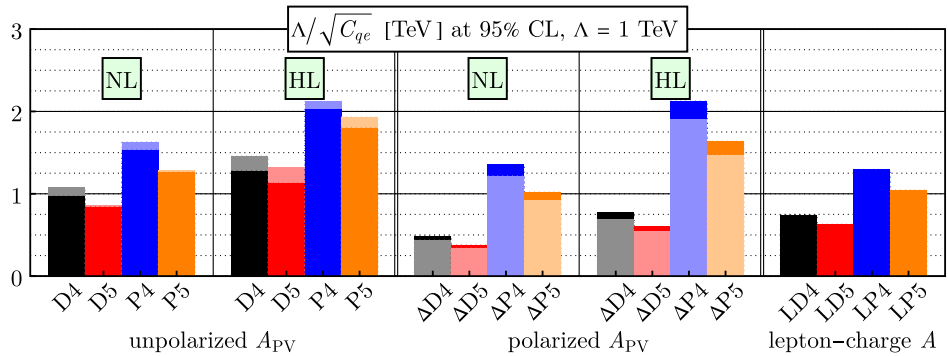


FIG. 35. The same as in Fig. 29 but for C_{qe} .

TABLE V. The 95% C.L. bounds of all the seven Wilson coefficients around zero at $\Lambda = 1$ TeV with the four families of datasets, D4, D5, P4, and P5 in various configurations.

eD				C_{eu}	C_{ed}	$C_{\ell q}^{(1)}$	$C_{\ell q}^{(3)}$	$C_{\ell u}$	$C_{\ell d}$	C_{qe}	
10 GeV \times 137 GeV, 100 fb $^{-1}$	Unpol. A_{PV}	NL	Without P	0.21	0.38	0.54	0.14	0.47	0.75	1.06	
			With P	0.15	0.27	0.39	0.10	0.37	0.60	0.87	
		HL	Without P	0.16	0.31	0.44	0.12	0.28	0.47	0.61	
			With P	0.09	0.16	0.23	0.06	0.21	0.34	0.47	
	Pol. A_{PV}	NL	Without P	0.90	1.82	1.98	0.61	2.08	4.02	4.24	
			With P	1.08	2.20	2.40	0.74	2.57	4.97	5.26	
		HL	Without P	0.69	1.42	1.41	0.46	0.85	1.69	1.69	
			With P	0.83	1.70	1.70	0.56	1.06	2.11	2.11	
	Lepton-charge A				1.16	2.28	2.54	0.70	0.95	1.84	1.81
	18 GeV \times 137 GeV, 15.4 fb $^{-1}$	Unpol. A_{PV}	NL	Without P	0.25	0.45	0.73	0.17	0.62	0.96	1.41
With P				0.23	0.42	0.68	0.16	0.60	0.92	1.37	
		HL	Without P	0.17	0.32	0.50	0.12	0.36	0.56	0.78	
			With P	0.10	0.18	0.30	0.07	0.26	0.40	0.58	
Pol. A_{PV}		NL	Without P	1.17	2.36	2.89	0.82	3.41	6.55	7.05	
			With P	1.43	2.89	3.55	1.00	4.23	8.12	8.75	
		HL	Without P	0.75	1.52	1.71	0.51	1.39	2.72	2.78	
			With P	0.89	1.82	2.06	0.62	1.72	3.37	3.46	
Lepton-charge A				1.92	3.78	4.32	1.11	1.39	2.68	2.53	
ep				C_{eu}	C_{ed}	$C_{\ell q}^{(1)}$	$C_{\ell q}^{(3)}$	$C_{\ell u}$	$C_{\ell d}$	C_{qe}	
10 GeV \times 275 GeV, 100 fb $^{-1}$	Unpol. A_{PV}	NL	Without P	0.13	0.47	0.19	0.12	0.32	0.63	0.43	
			With P	0.09	0.28	0.14	0.07	0.24	0.48	0.38	
		HL	Without P	0.09	0.40	0.13	0.09	0.19	0.44	0.24	
			With P	0.06	0.20	0.09	0.05	0.15	0.33	0.22	
	Pol. A_{PV}	NL	Without P	0.28	1.70	0.23	0.35	0.60	3.79	0.54	
			With P	0.33	2.03	0.27	0.42	0.74	4.72	0.68	
		HL	Without P	0.24	1.39	0.20	0.29	0.25	1.48	0.22	
			With P	0.28	1.63	0.23	0.35	0.31	1.85	0.28	
	Lepton-charge A				0.62	3.15	0.74	0.47	0.46	2.17	0.60
	18 GeV \times 275 GeV, 15.4 fb $^{-1}$	Unpol. A_{PV}	NL	Without P	0.15	0.51	0.26	0.14	0.44	0.79	0.63
With P				0.14	0.41	0.24	0.11	0.39	0.71	0.61	
		HL	Without P	0.09	0.39	0.15	0.09	0.25	0.50	0.31	
			With P	0.06	0.21	0.11	0.06	0.19	0.37	0.27	
Pol. A_{PV}		NL	Without P	0.34	2.27	0.28	0.45	1.01	6.85	0.96	
			With P	0.42	2.77	0.34	0.55	1.24	8.53	1.19	
		HL	without P	0.24	1.54	0.20	0.31	0.40	2.46	0.37	
			With P	0.29	1.83	0.23	0.37	0.50	3.07	0.46	
Lepton-charge A				1.09	5.69	1.25	0.78	0.70	3.19	0.92	

single-parameter fits with the Wilson coefficient C_r , whereas the lighter ones show the bounds on the Wilson coefficient from simultaneous two-parameter fits with C_r and the beam polarization. The details of the fits involving

the beam polarization as an additional parameter are described in Appendix A 2.

In Figs. 29–35, we present the effective UV cutoff scales, $\Lambda/\sqrt{C_r}$, with $\Lambda = 1$ TeV, corresponding to the bounds

shown in Figs. 22–28. The organization of these plots in terms of asymmetries and datasets is the same as in Figs. 22–28. Improved bounds on C_r with the addition of the beam polarization to the fits are equivalent to higher-energy scales in the unpolarized PV asymmetries, which are indicated by the lighter columns in the background; on the other hand, weaker bounds from the fits with beam polarization are depicted by the lighter columns in the foreground for the polarized PV asymmetries.

For completeness, we summarize in Table V the fitting results of all the seven Wilson coefficients with the four families of datasets of interest. The values indicated in this table are the 95% C.L. bounds around zero.

2. Fits of two Wilson coefficients

In this section, we present the complete set of confidence ellipses for all possible pairs of the seven Wilson coefficients

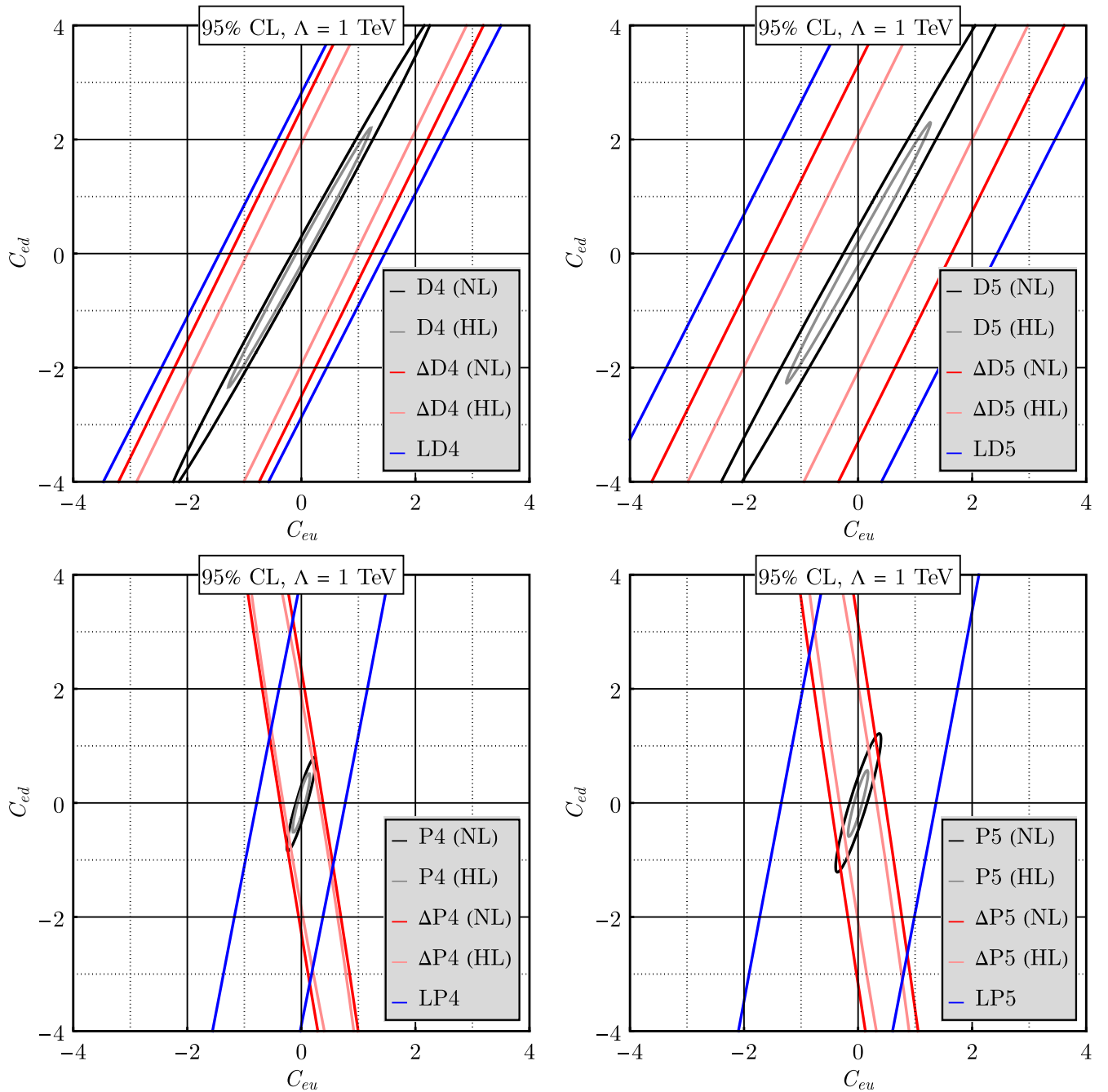


FIG. 36. The 95% C.L. ellipses for the Wilson coefficients C_{eu} and C_{ed} using the families of datasets D4, D5, P4, and P5 at $\Lambda = 1$ TeV.

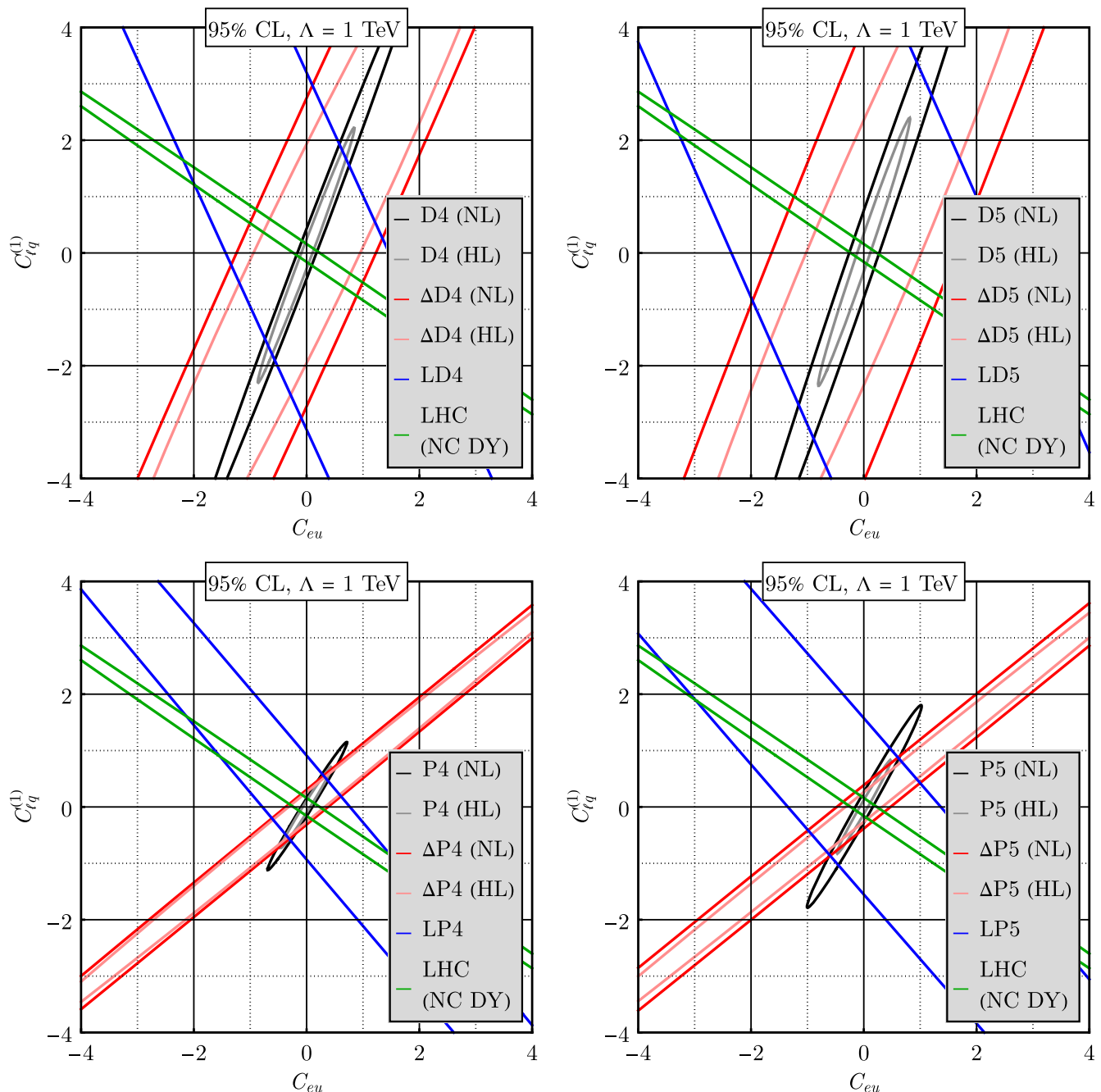


FIG. 37. The same as in Fig. 36 but for C_{eu} and $C_{\ell q}^{(1)}$.

that we consider in this work. The ellipses displayed in Figs. 36–56 are plotted at 95% C.L. and $\Lambda = 1$ TeV.

As before, we refer to electron PV asymmetries collectively as unpolarized A_{PV} , hadron PV asymmetries as polarized A_{PV} , and electron-positron asymmetries as lepton-charge A . Datasets with the label NL or HL indicate the luminosity: The NL refers to the annual integrated luminosity of Table 10.1 of YR [24]. The HL is assumed to be 10 times higher than the nominal one and requires a luminosity upgrade of the EIC.

Each figure consists of four panels, containing one of the four families of datasets, namely D4, D5, P4, and P5. We show the fits from polarized and unpolarized PV asymmetry datasets in both nominal- and high-luminosity scenarios for comparison. We remark that the ellipses for the polarized and unpolarized PV asymmetry datasets indicate the results of simultaneous fits on Wilson coefficients with the beam-polarization parameter, P , in light of significant improvements in the results with unpolarized PV asymmetries. Moreover, we include for some representative

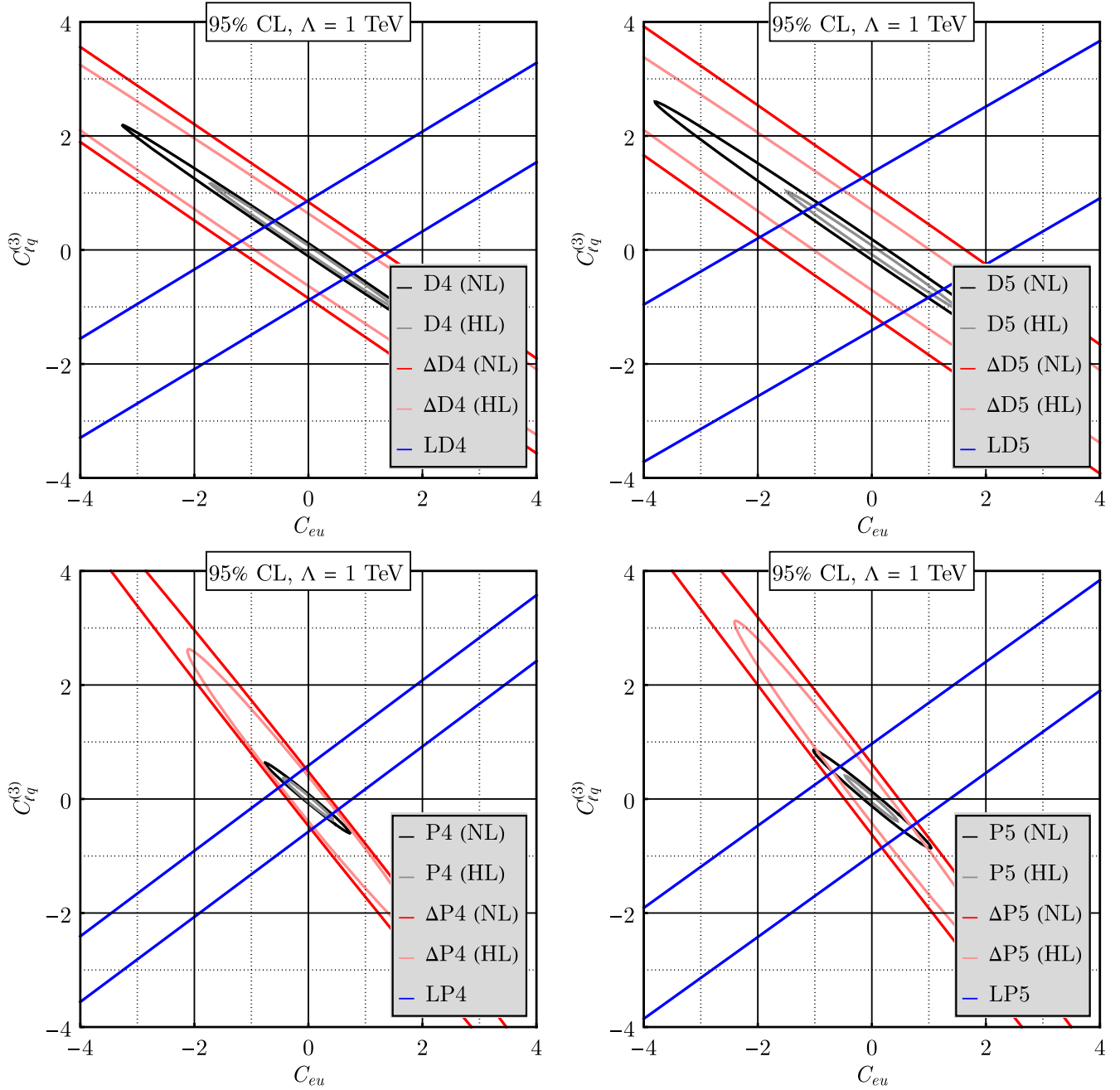


FIG. 38. The same as in Fig. 36 but for C_{eu} and $C_{\ell q}^{(3)}$.

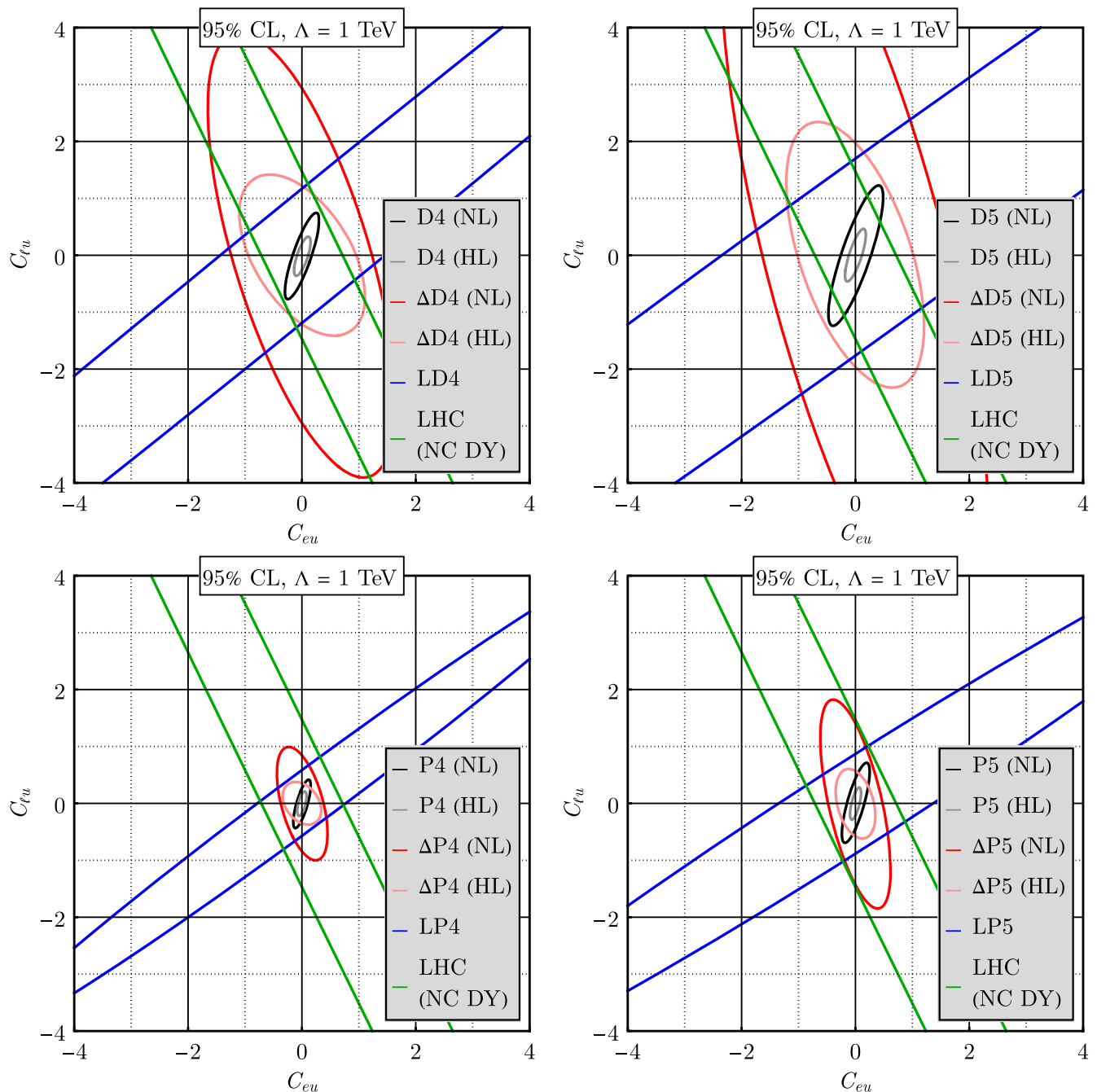
examples fitted results for the LHC Drell-Yan data, adapting from [6] and [45].

For completeness, we show the correlations of Wilson coefficients in Figs. 57–60.

3. Six-dimensional fits

In this section, we discuss projected bounds arising from a fit of six Wilson coefficients. We show projections onto one and two Wilson coefficients from the full 6d hyperellipse. The computational power required for

higher-dimensional fits increases because we increase the number of pseudoexperiments to reflect the required statistics for the fits. We find that $N_{\text{exp}} = 10^3$ pseudoexperiments for 2d fits lead to stable results, meaning the best-fit values and the corresponding bounds do not change for $N_{\text{exp}} \geq 10^3$. This number becomes $N_{\text{exp}} = 10^4$ for 3d fits, $N_{\text{exp}} = 10^5$ for 4d fits, $N_{\text{exp}} = 10^6$ for 5d fits, and $N_{\text{exp}} = 10^7$ for 6d fits. We note that starting from 3d fits, the size and characteristics of multidimensional fits stabilize as we increase


 FIG. 39. The same as in Fig. 36 but for C_{eu} and $C_{\ell u}$.

the number of fitted parameters, in the sense that the bounds do not change further. This indicates that the $6d$ fits stand as a useful indicator of what happens in the complete $7d$ fit of the Wilson coefficients.

To illustrate our discussion, we pick the representative Wilson coefficients C_{eu} , C_{ed} , $C_{\ell q}^{(1)}$, $C_{\ell q}^{(3)}$, $C_{\ell u}$, and C_{qe} and consider the data from ep collisions in the configuration $10 \text{ GeV} \times 275 \text{ GeV}$ with 100 fb^{-1} , which is the nominal-luminosity P4 dataset. In Fig. 61, we compare the bounds

from the original $1d$ fits to the projected bounds from the $6d$ fit of the Wilson coefficients C_{eu} , C_{ed} , $C_{\ell q}^{(1)}$, $C_{\ell q}^{(3)}$, $C_{\ell u}$, and C_{qe} . We observe that the bounds become 25% to 40% weaker as we increase the number of Wilson coefficients fitted. This is due to an interplay between the increased number of fitted parameters and correlations among them. In Figs. 62–66, we compare the confidence ellipses from the initial $2d$ fits to the ones in the two-parameter projections of the $6d$ fit of the aforementioned Wilson

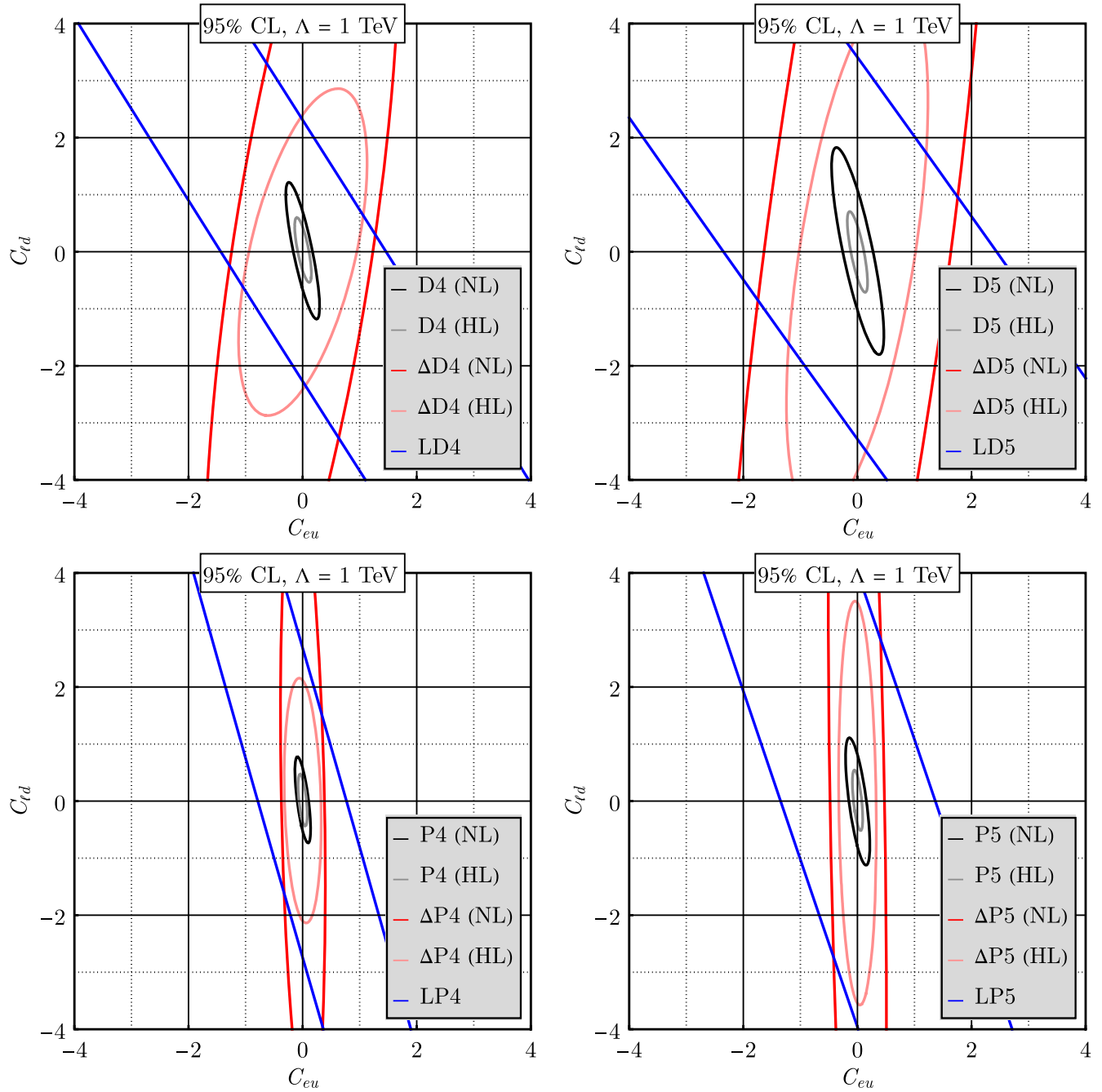


FIG. 40. The same as in Fig. 36 but for C_{eu} and $C_{\ell d}$.

coefficients. We find that the bounds become 20% to 30% weaker as in the case of the comparison of the one-parameter fits and projections, which can be explained by the same reasoning as mentioned above. The other

choices of six Wilson coefficients lead to similar results. We note that no flat directions appear in these fits, indicating that the EIC can fully probe this parameter space without degeneracies.

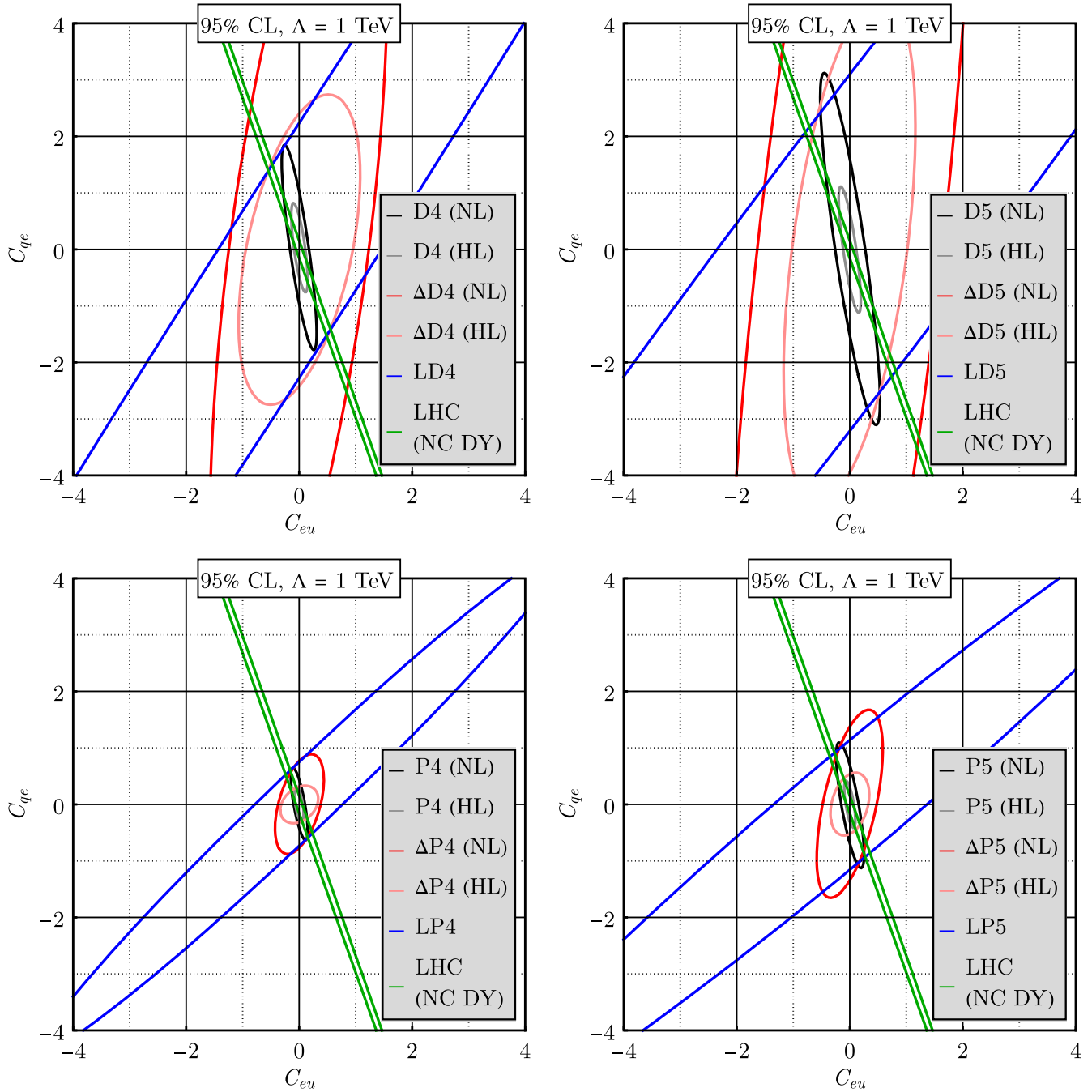


FIG. 41. The same as in Fig. 36 but for C_{eu} and C_{qe} .

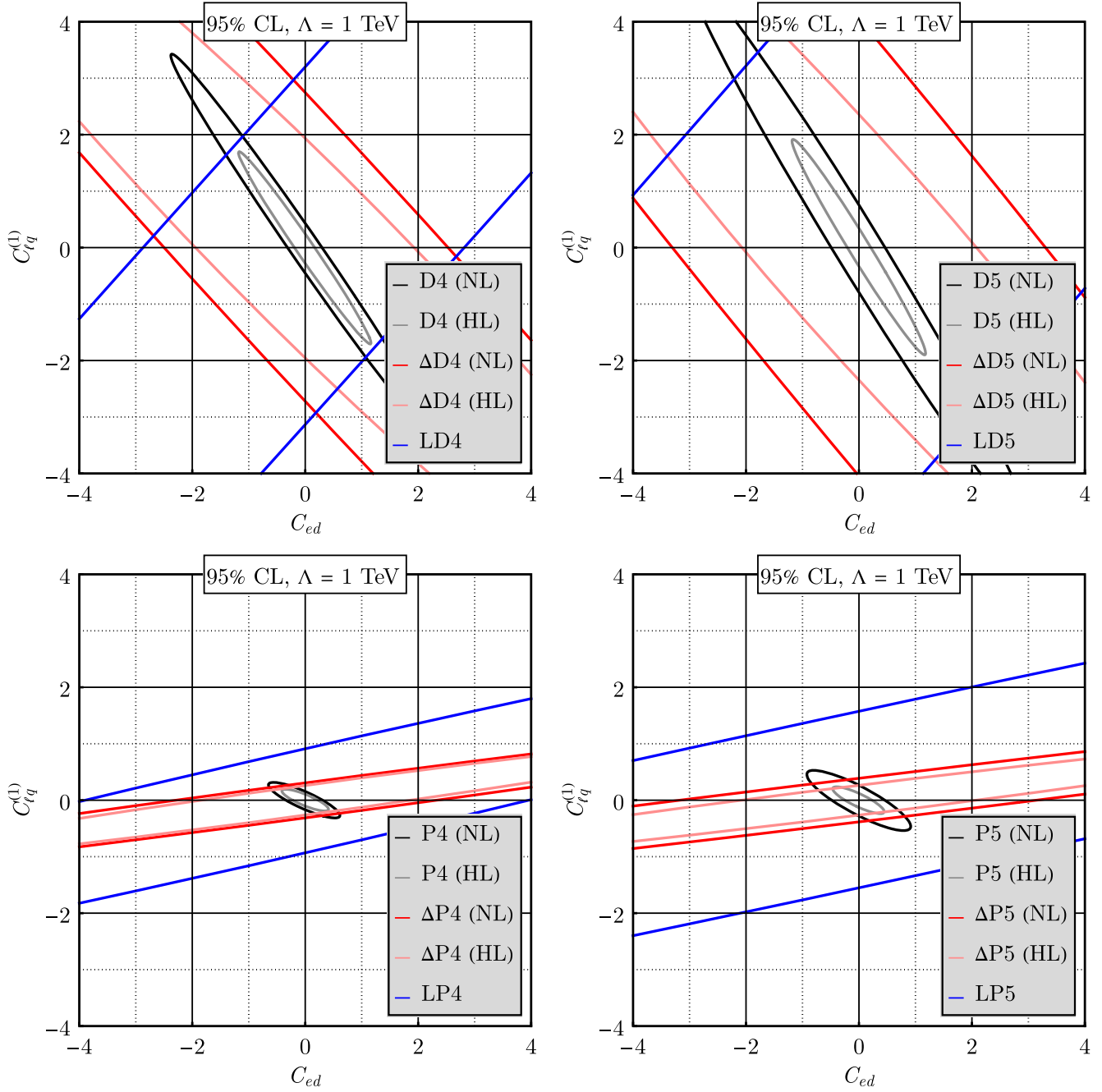


FIG. 42. The same as in Fig. 36 but for C_{ed} and $C_{\ell q}^{(1)}$.

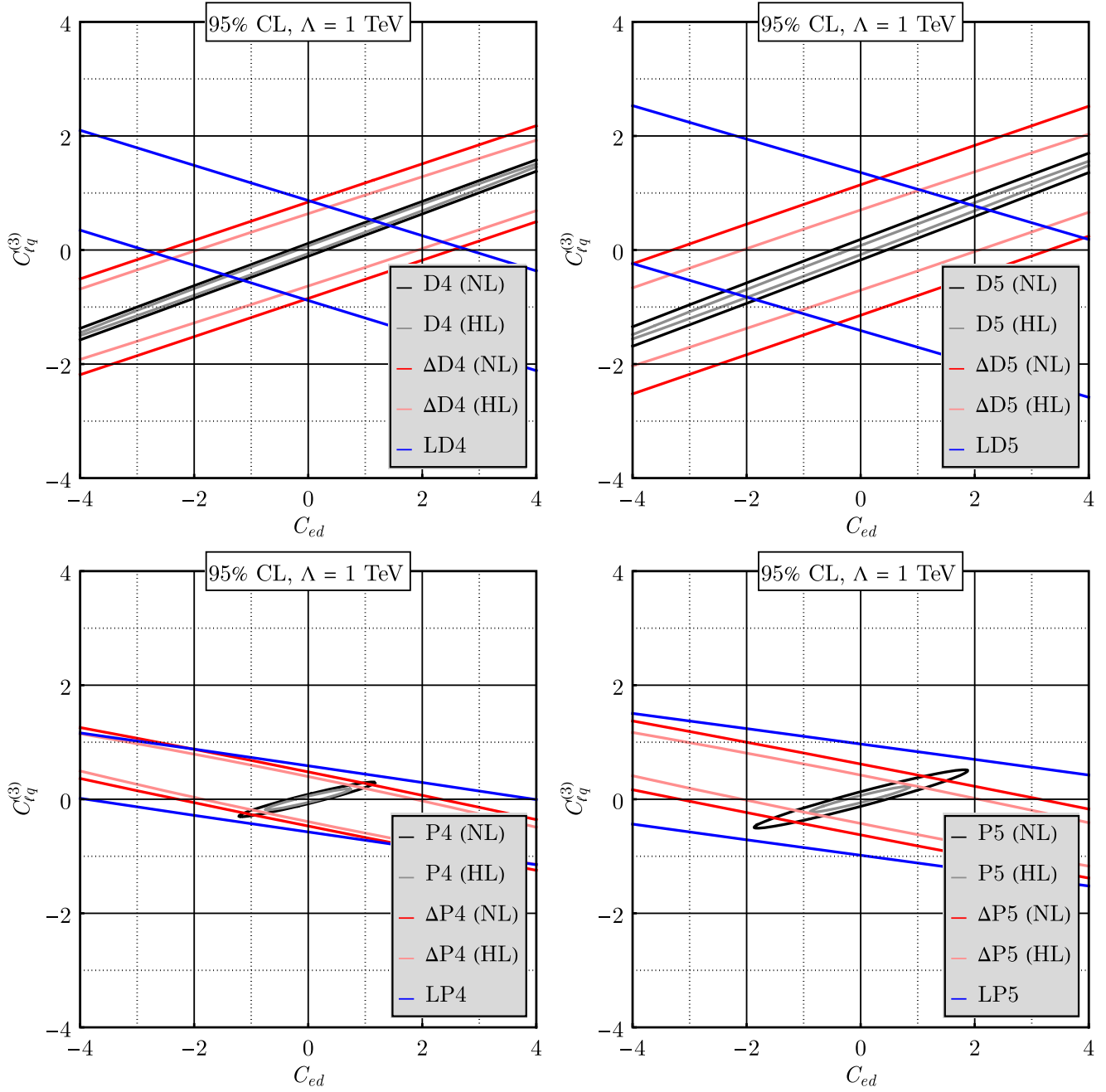


FIG. 43. The same as in Fig. 36 but for C_{ed} and $C_{\ell q}^{(3)}$.

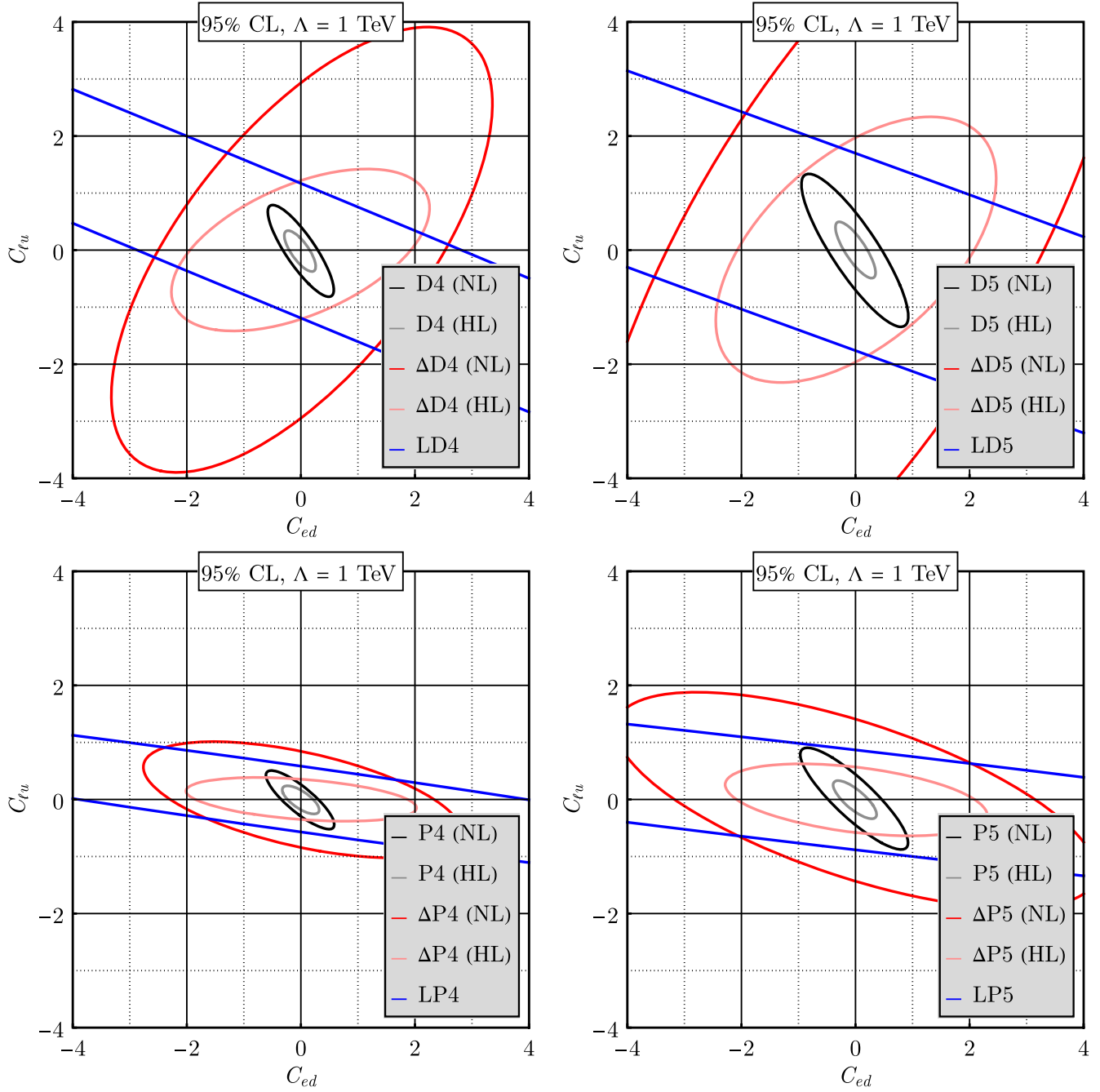


FIG. 44. The same as in Fig. 36 but for C_{ed} and $C_{\ell u}$.

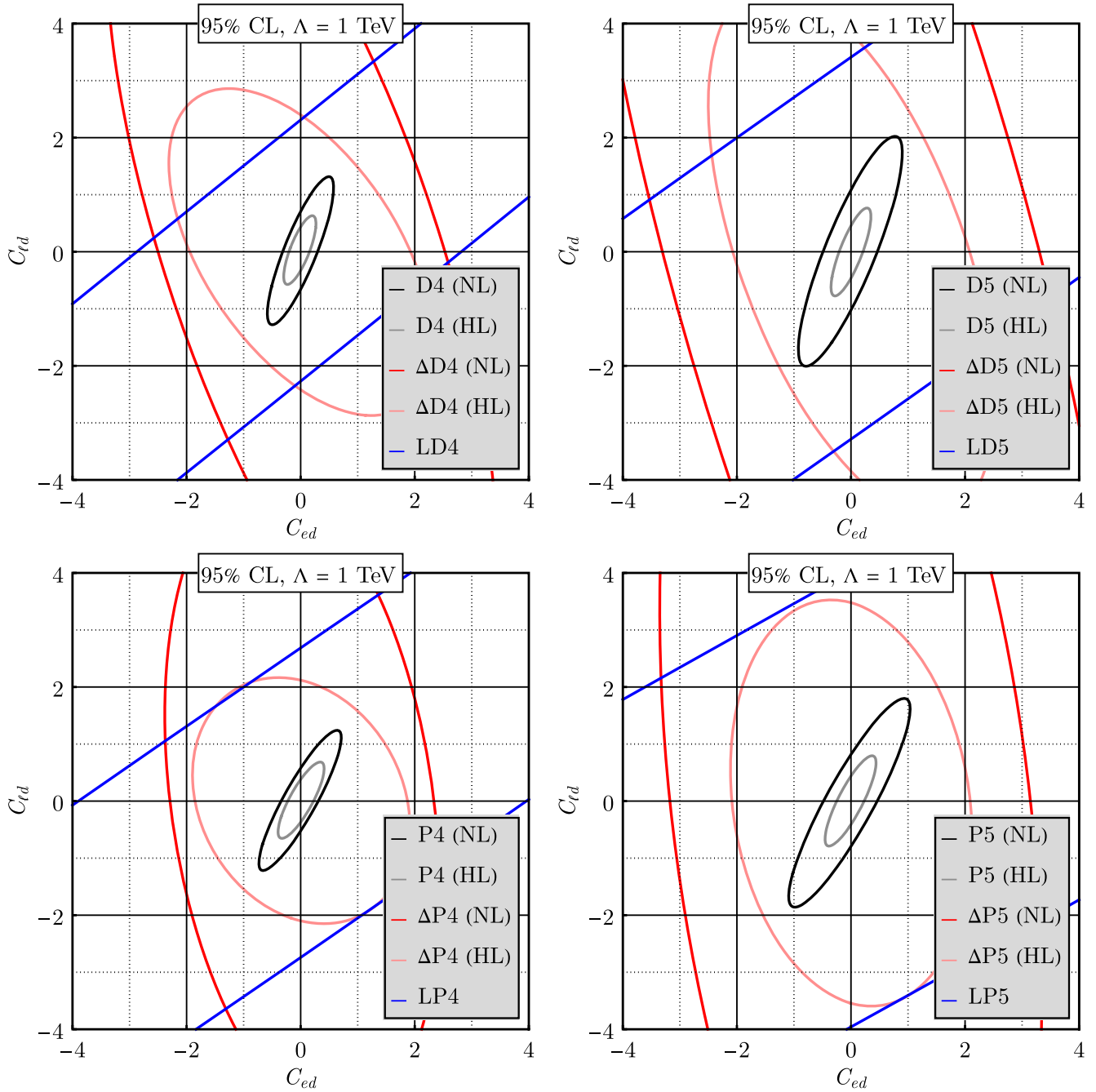


FIG. 45. The same as in Fig. 36 but for C_{ed} and $C_{\ell d}$.

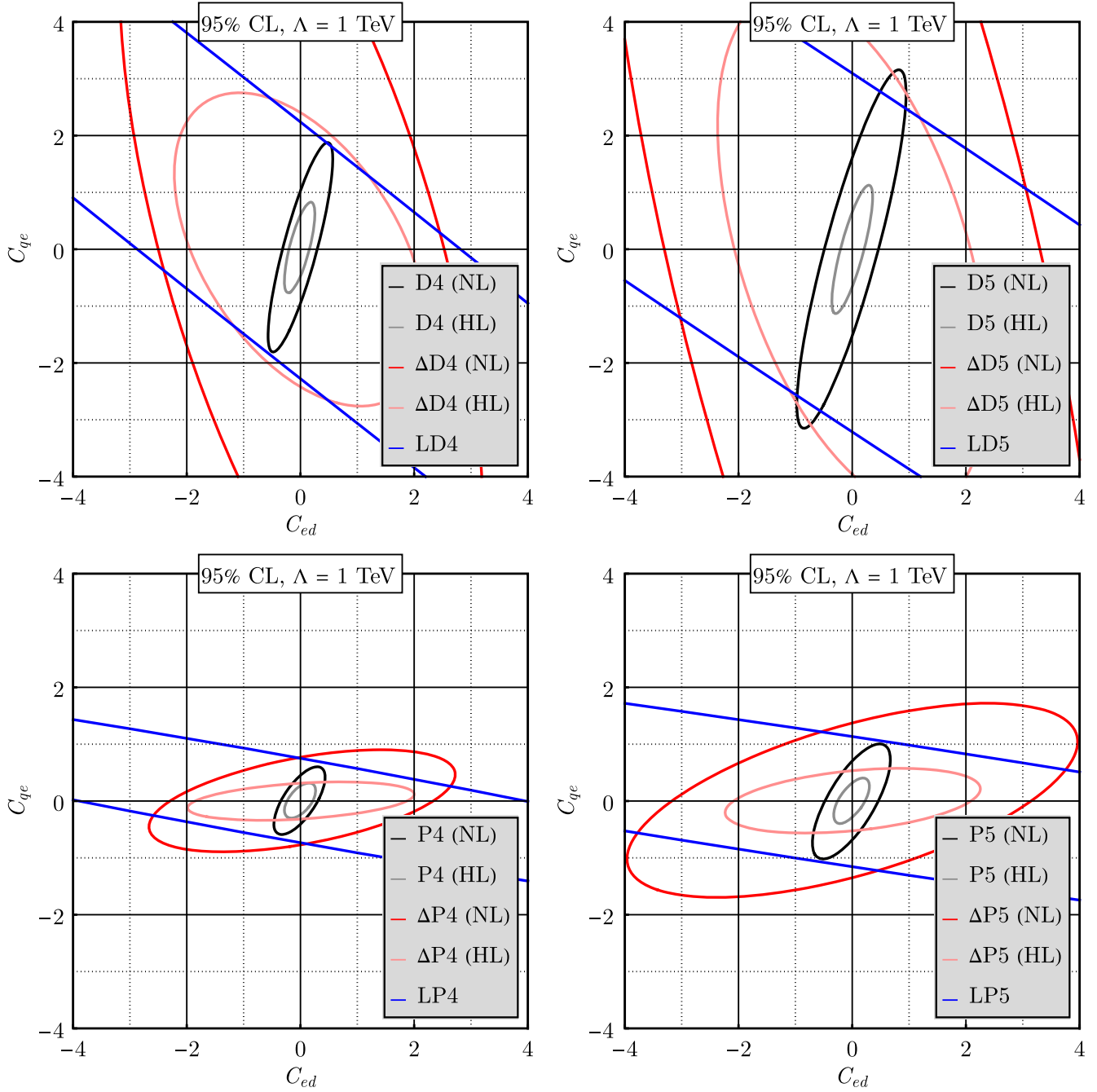


FIG. 46. The same as in Fig. 36 but for C_{ed} and C_{qe} .

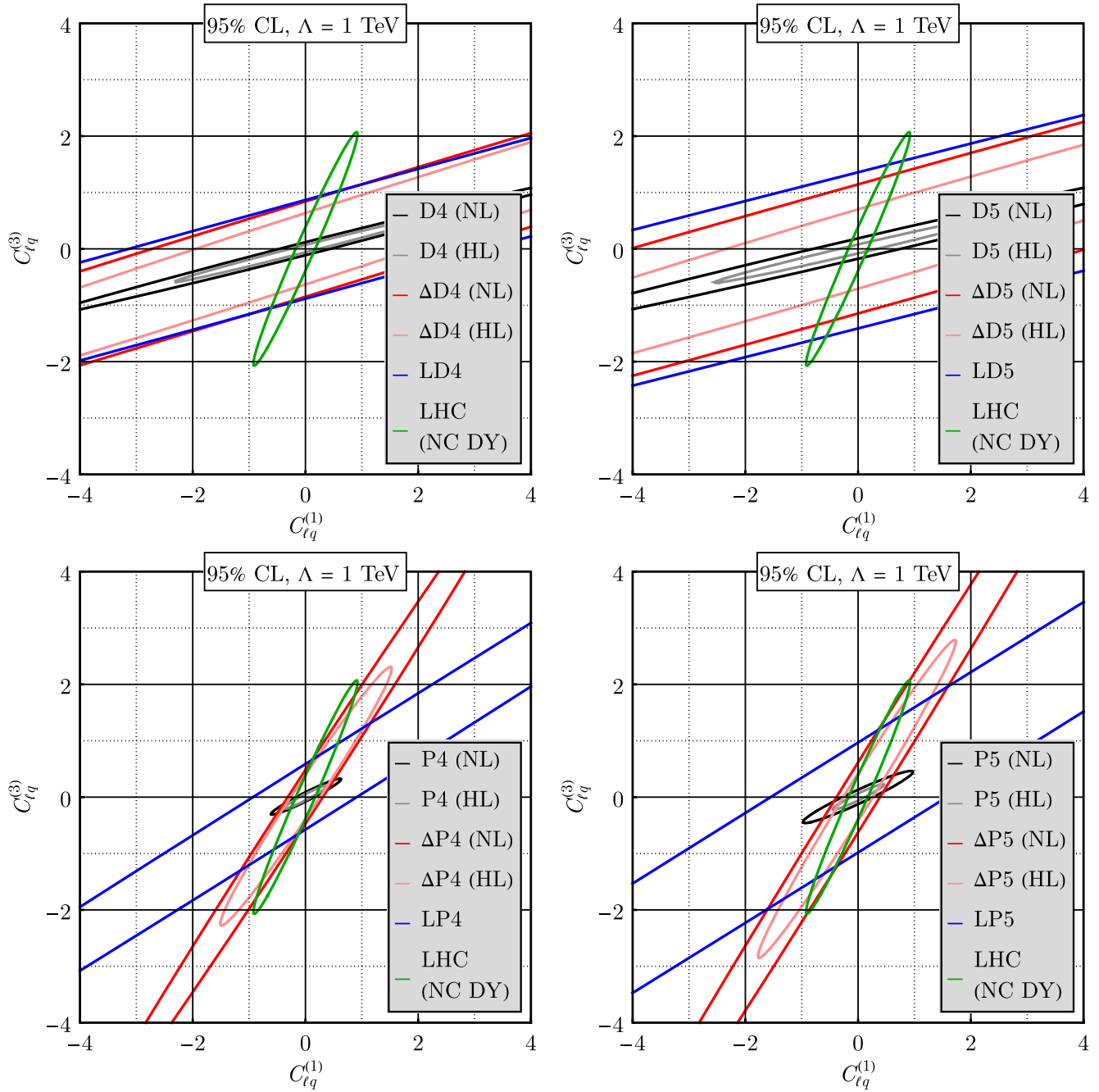


FIG. 47. The same as in Fig. 36 but for $C_{\ell q}^{(1)}$ and $C_{\ell q}^{(3)}$.

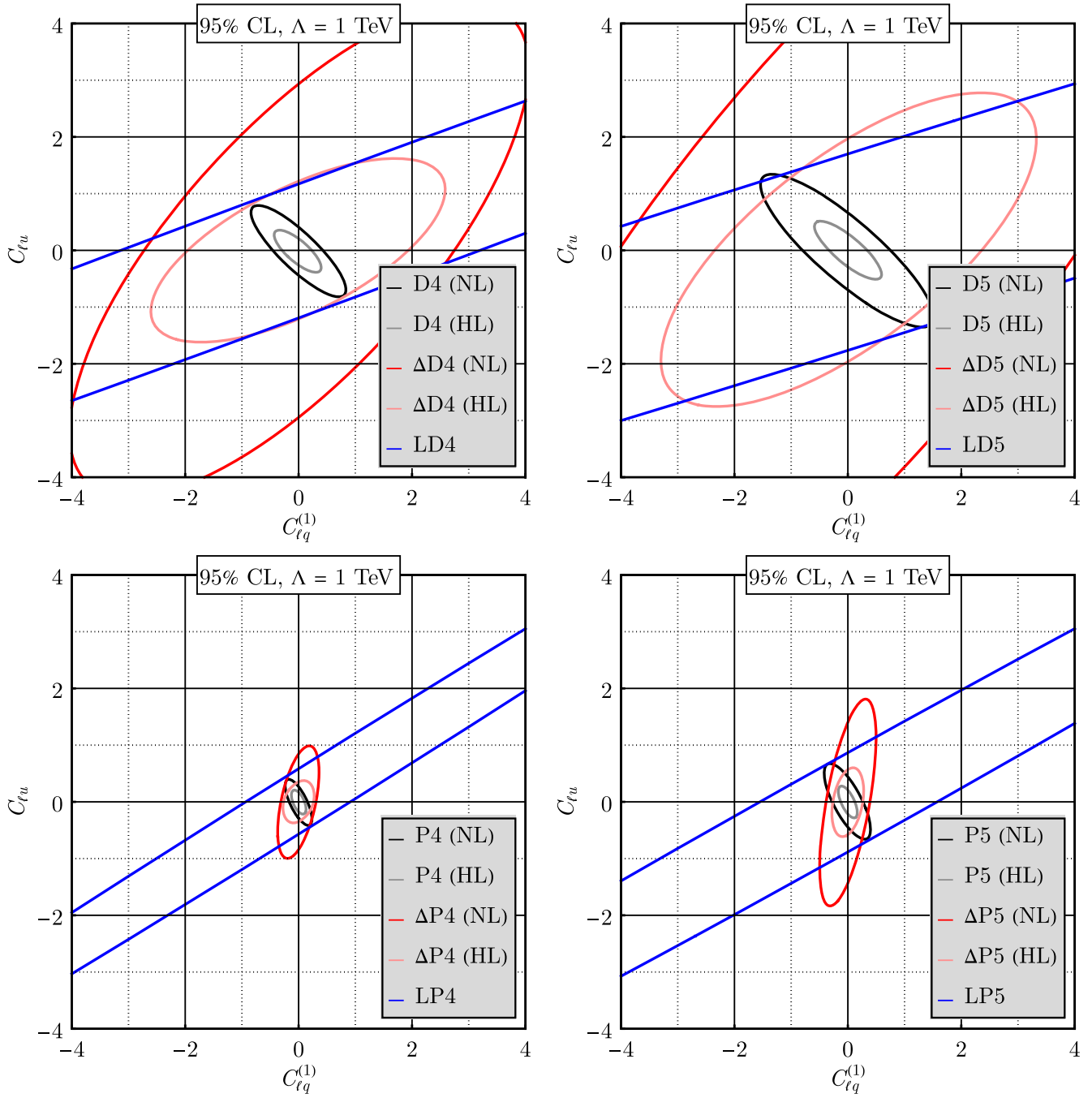


FIG. 48. The same as in Fig. 36 but for $C_{\ell q}^{(1)}$ and $C_{\ell u}$.

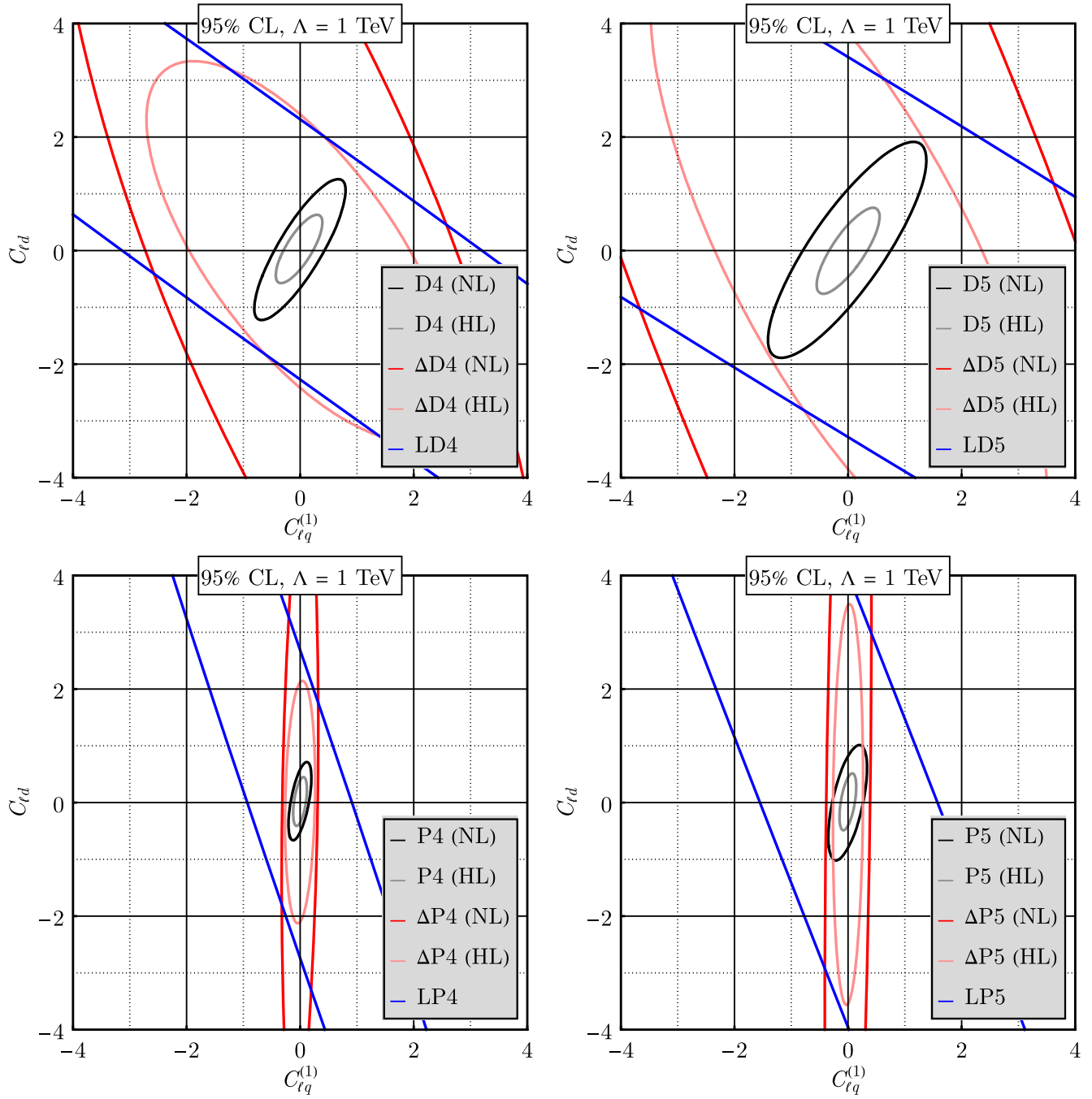


FIG. 49. The same as in Fig. 36 but for $C_{\ell q}^{(1)}$ and $C_{\ell d}$.

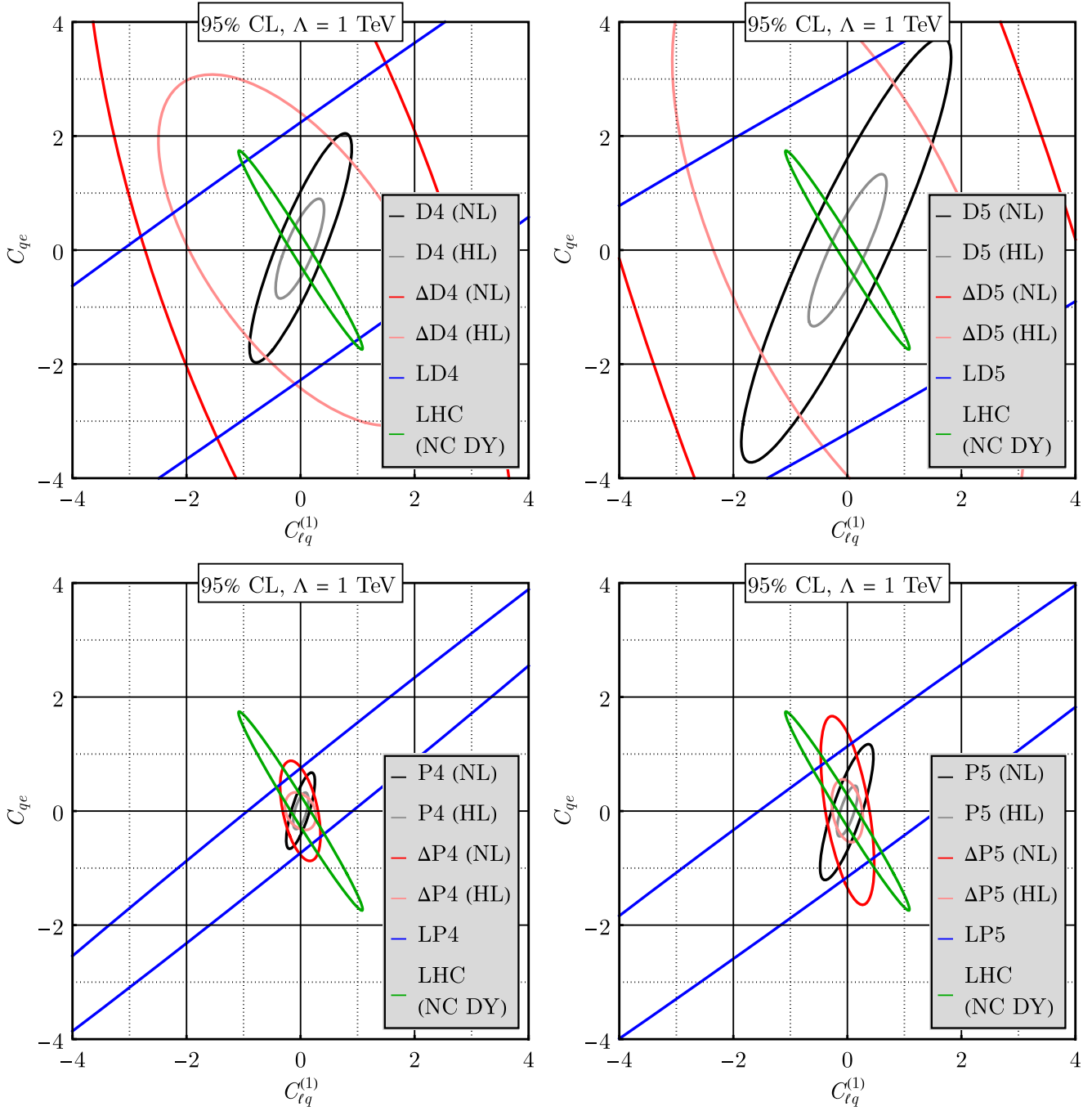


FIG. 50. The same as in Fig. 36 but for $C_{\ell q}^{(1)}$ and C_{qe} .

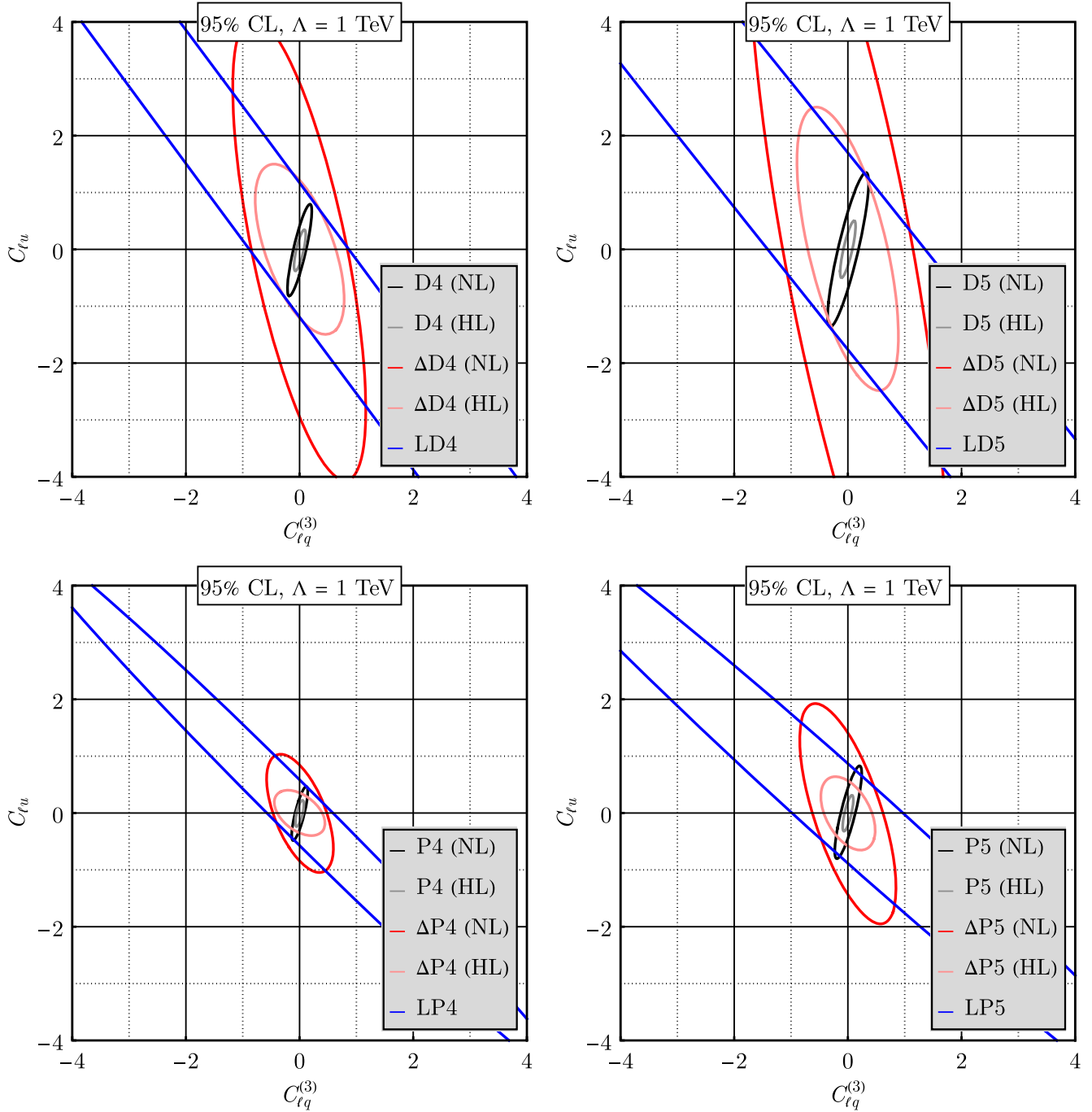


FIG. 51. The same as in Fig. 36 but for $C_{\ell q}^{(3)}$ and $C_{\ell u}$.

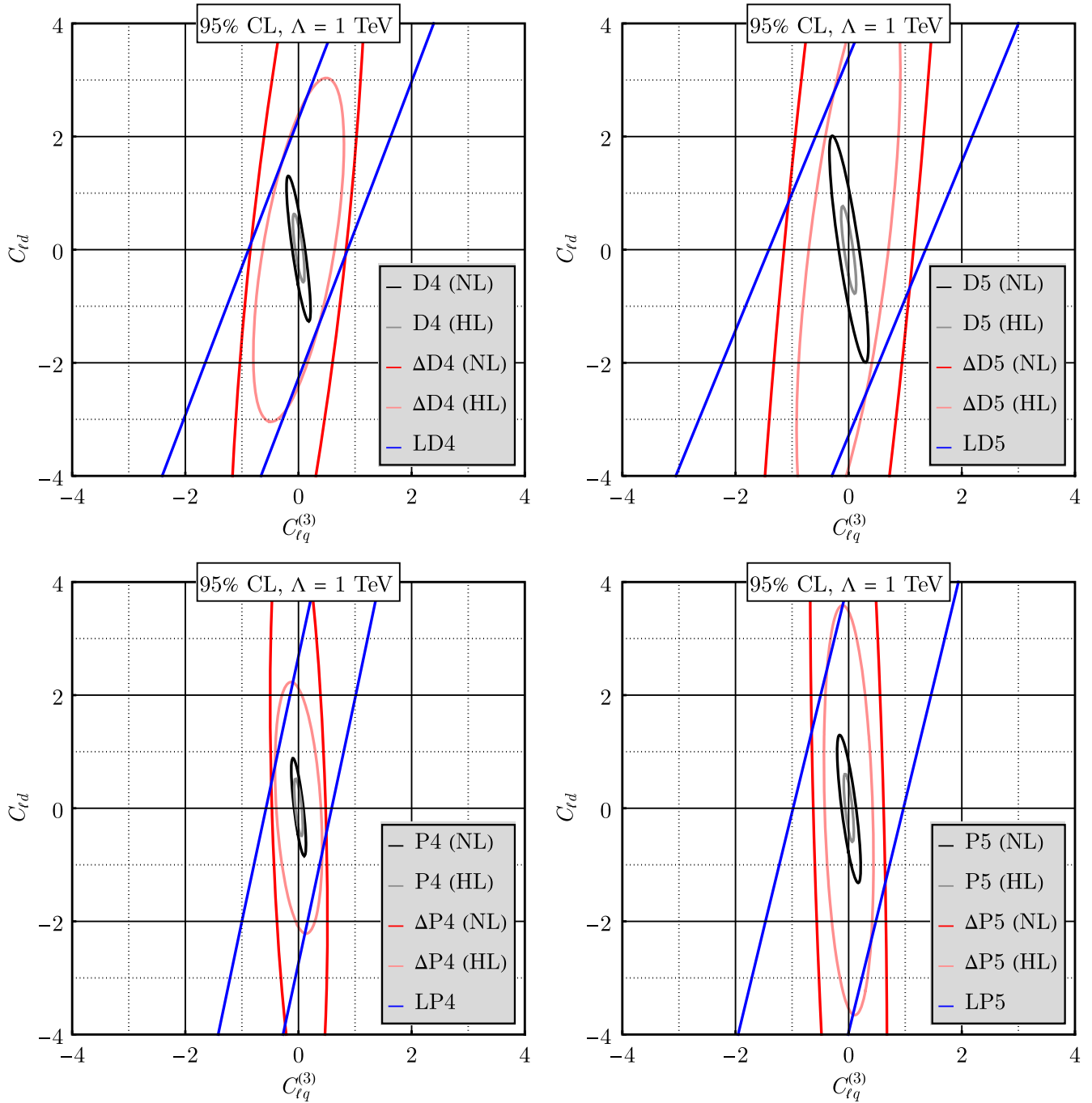


FIG. 52. The same as in Fig. 36 but for $C_{\ell q}^{(3)}$ and $C_{\ell d}$.

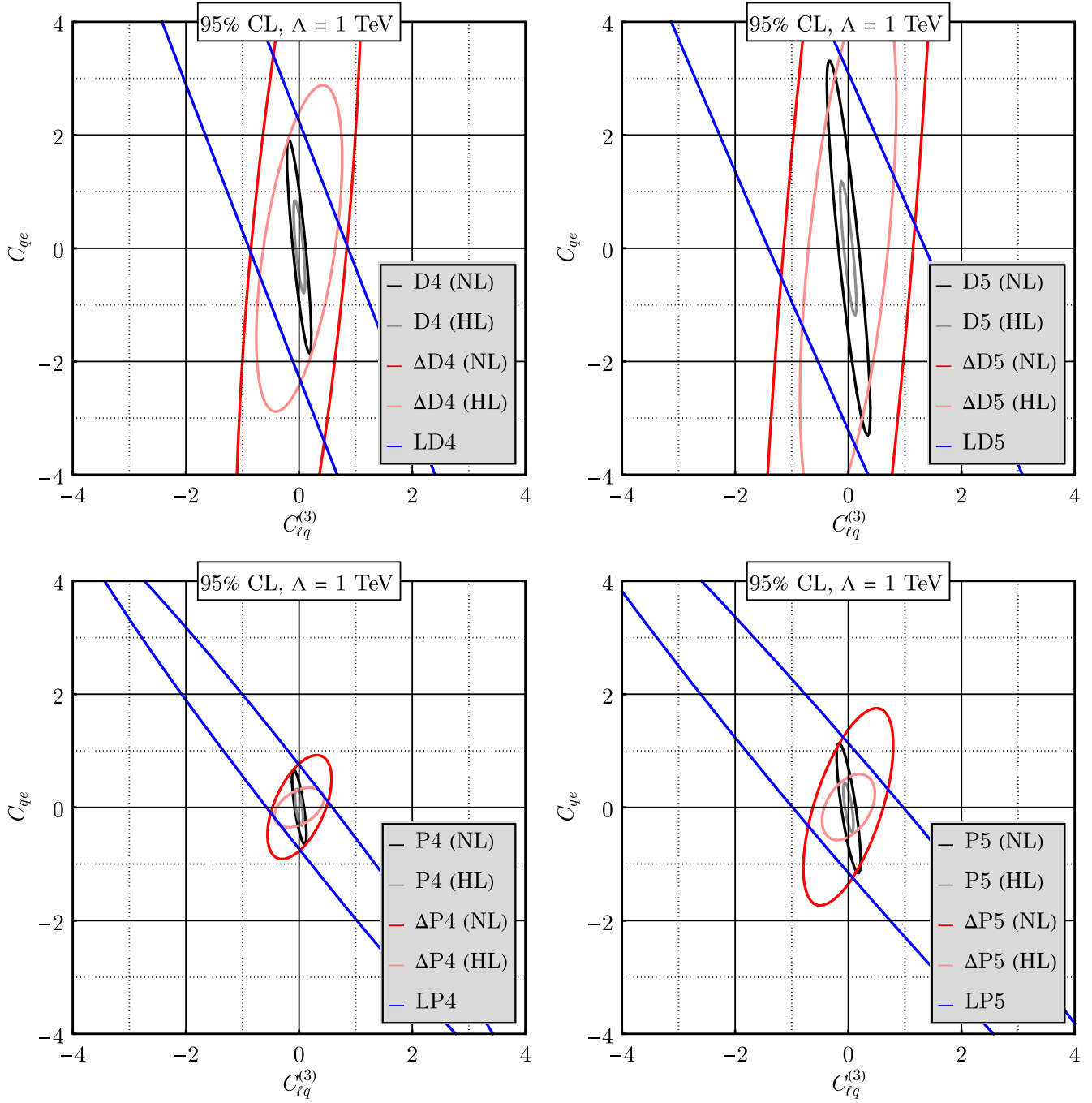


FIG. 53. The same as in Fig. 36 but for $C_{\ell q}^{(3)}$ and C_{qe} .

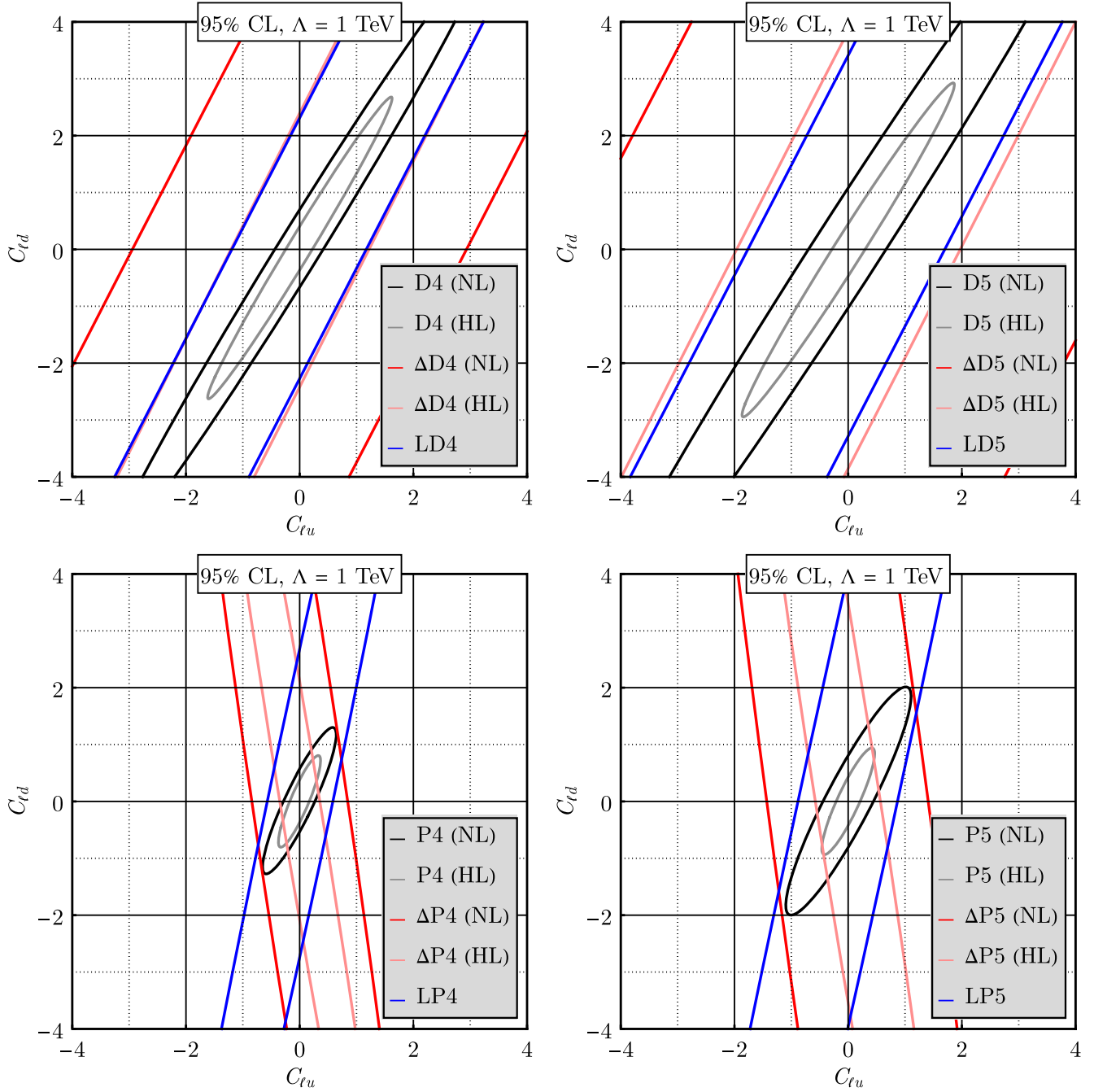


FIG. 54. The same as in Fig. 36 but for $C_{\ell u}$ and $C_{\ell d}$.

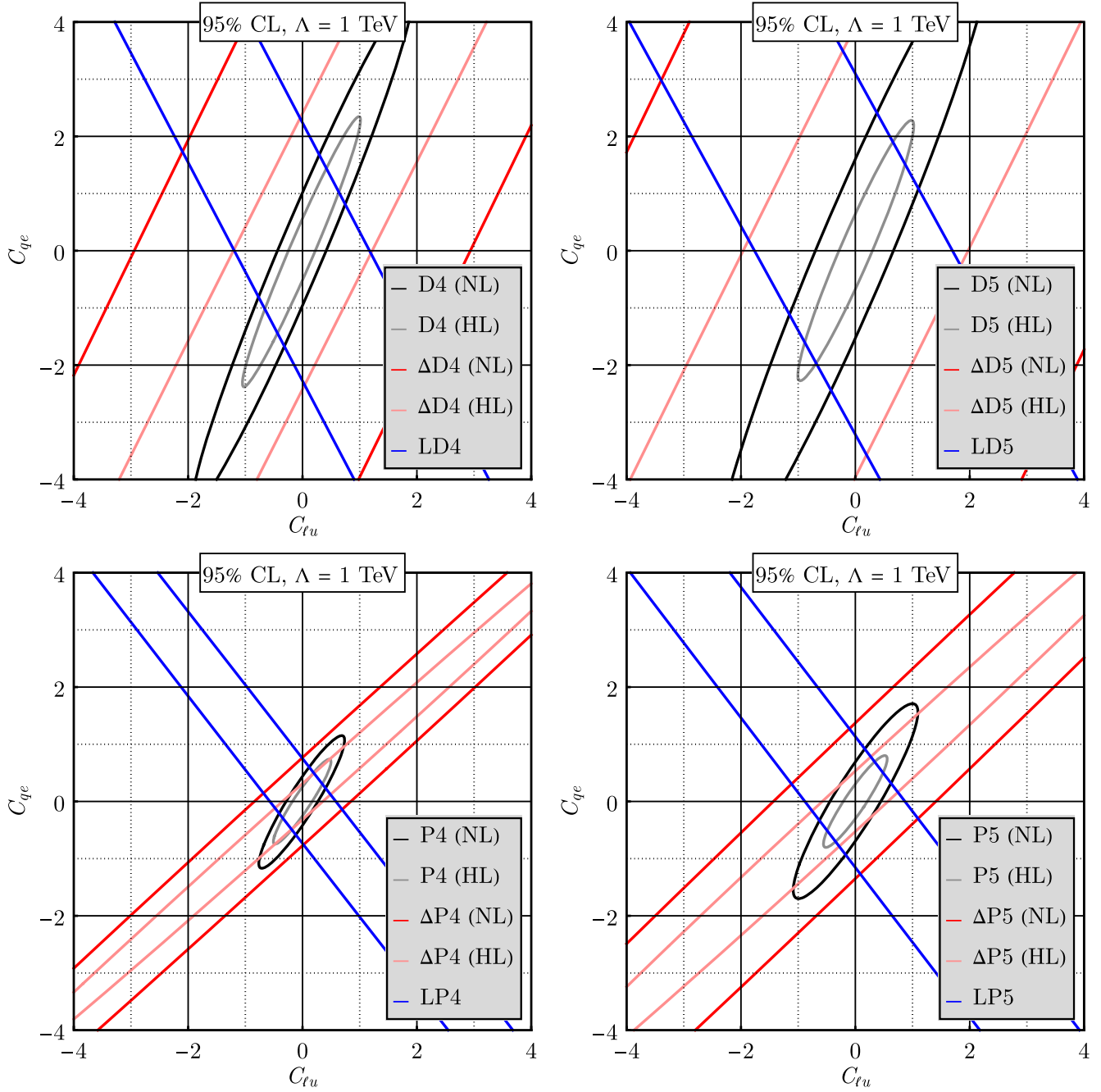


FIG. 55. The same as in Fig. 36 but for $C_{\ell u}$ and C_{qe} .

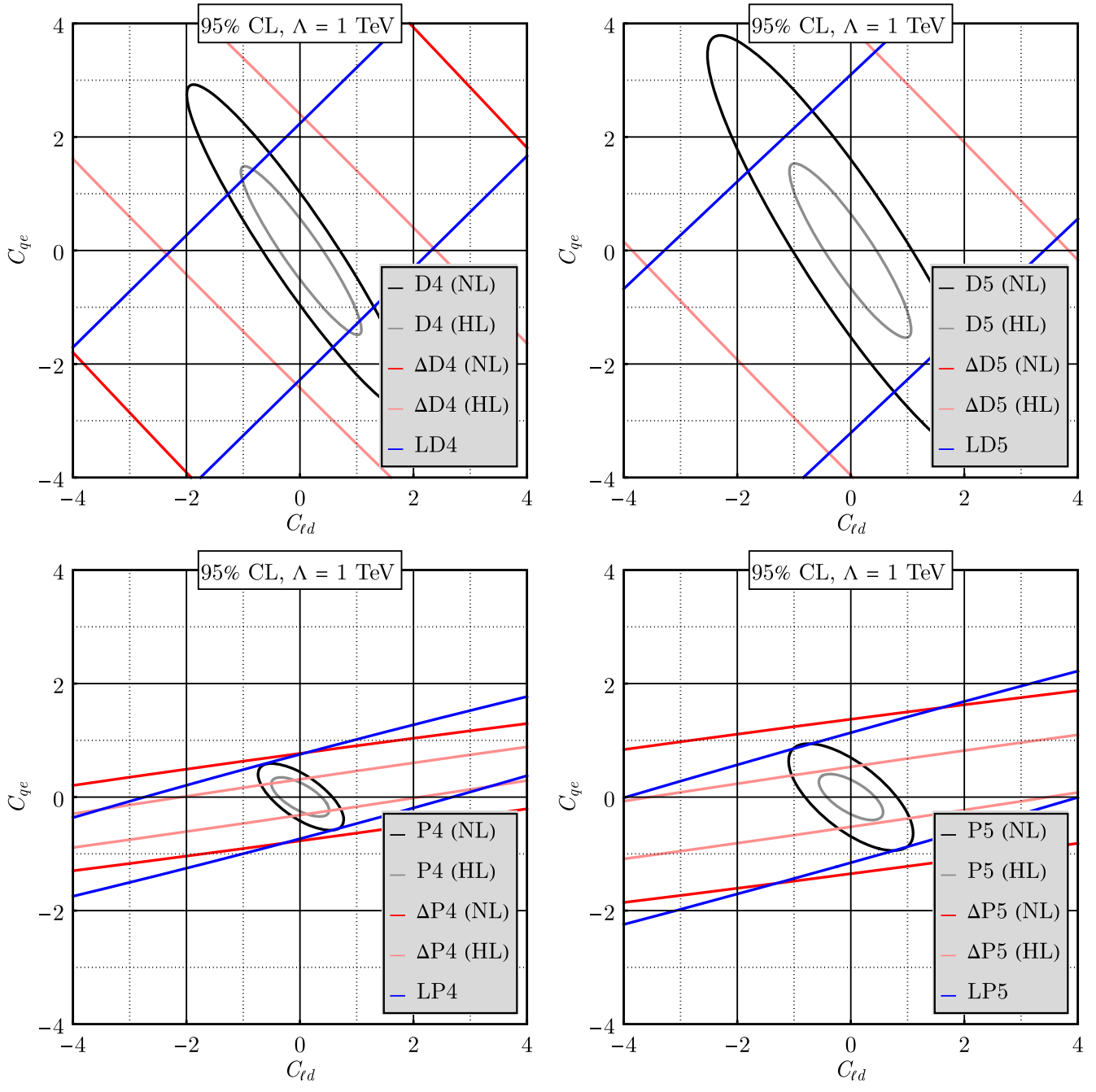


FIG. 56. The same as in Fig. 36 but for $C_{\ell d}$ and C_{qe} .

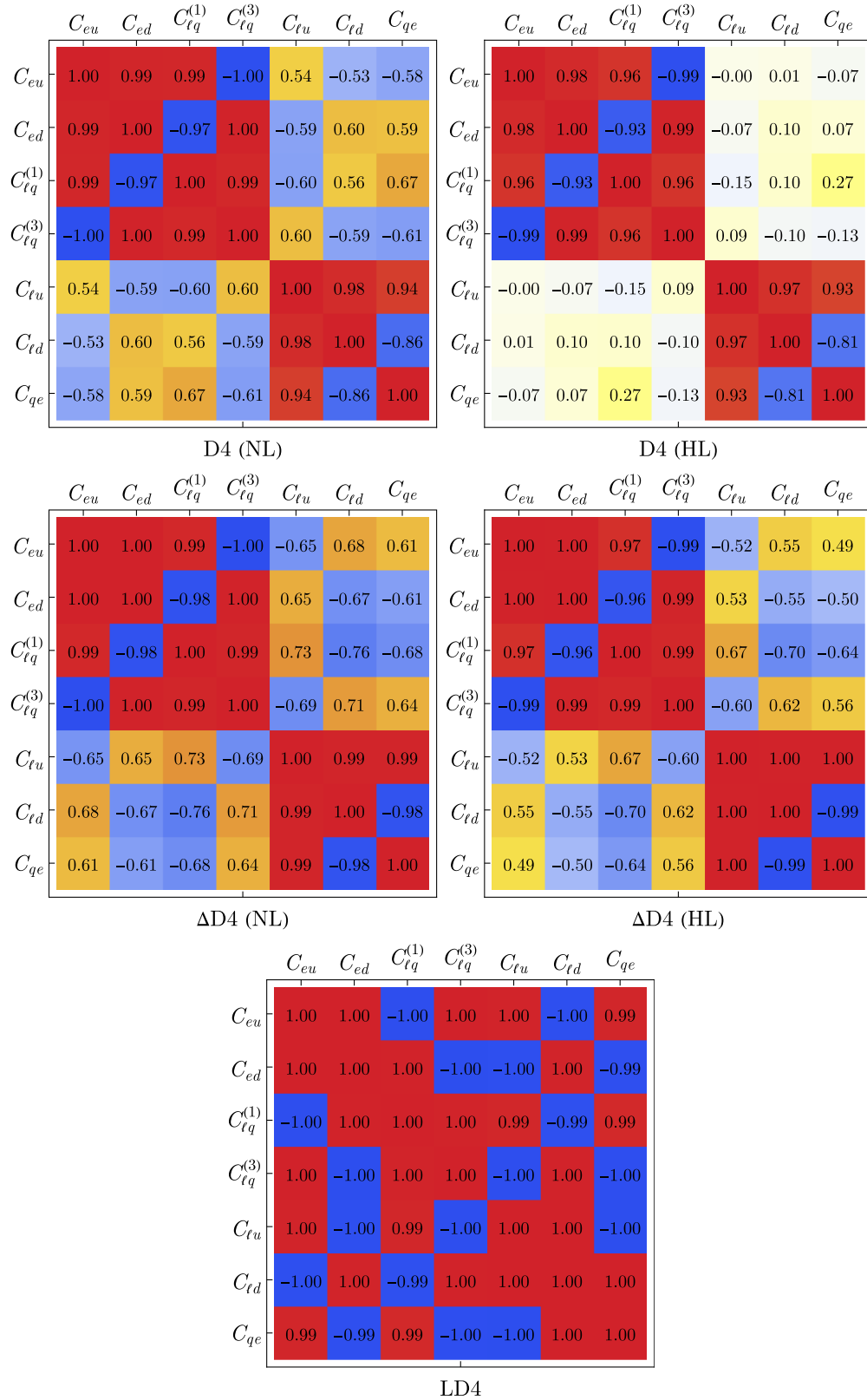


FIG. 57. The correlation table of Wilson coefficients in the D4 data family. The off-diagonal entries are from the results of simultaneous fits of the (2 + 1)-parameter fit of two Wilson coefficients plus the beam-polarization parameter P .

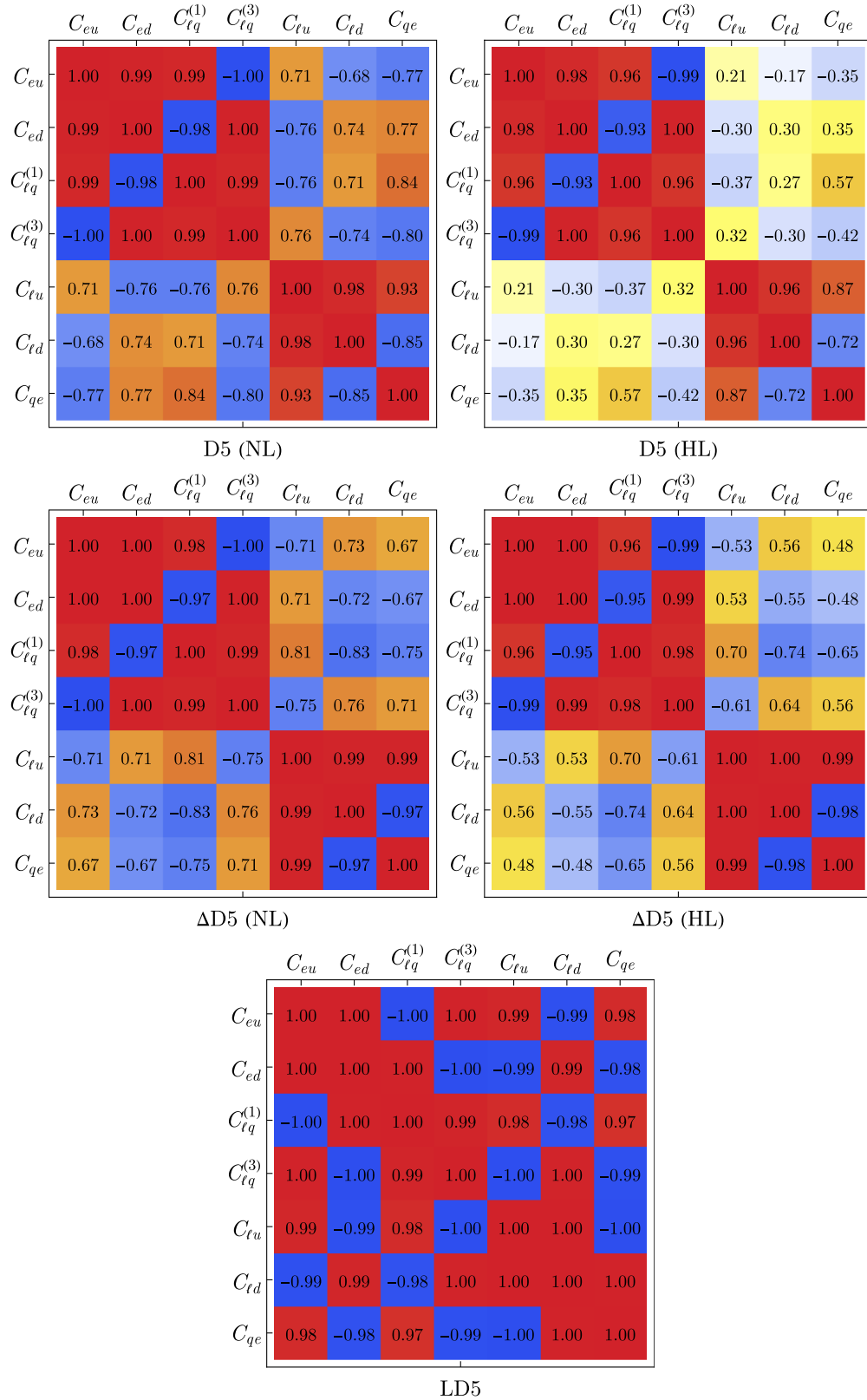


FIG. 58. The same as in Fig. 57 but for D5.

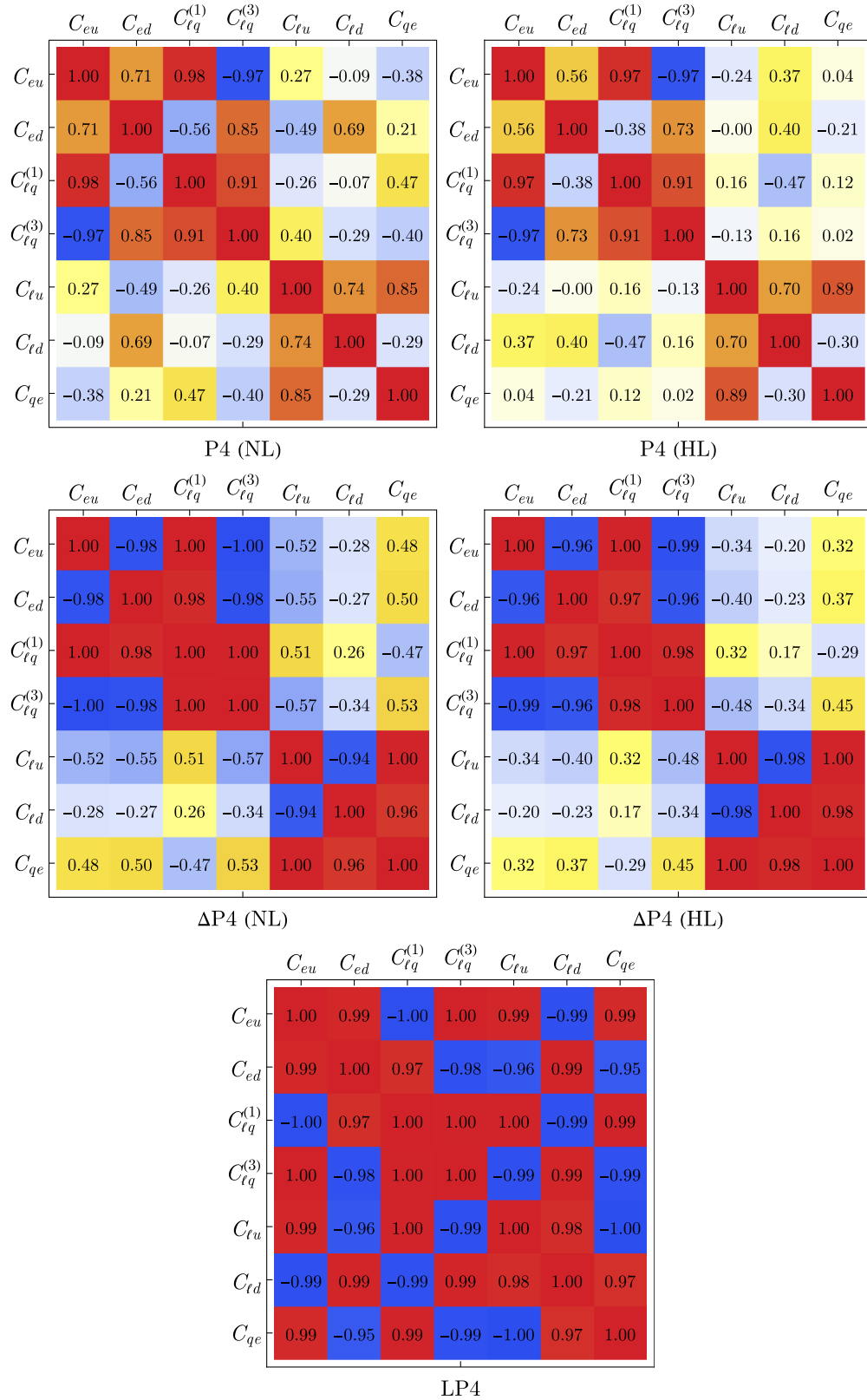


FIG. 59. The same as in Fig. 57 but for P4.

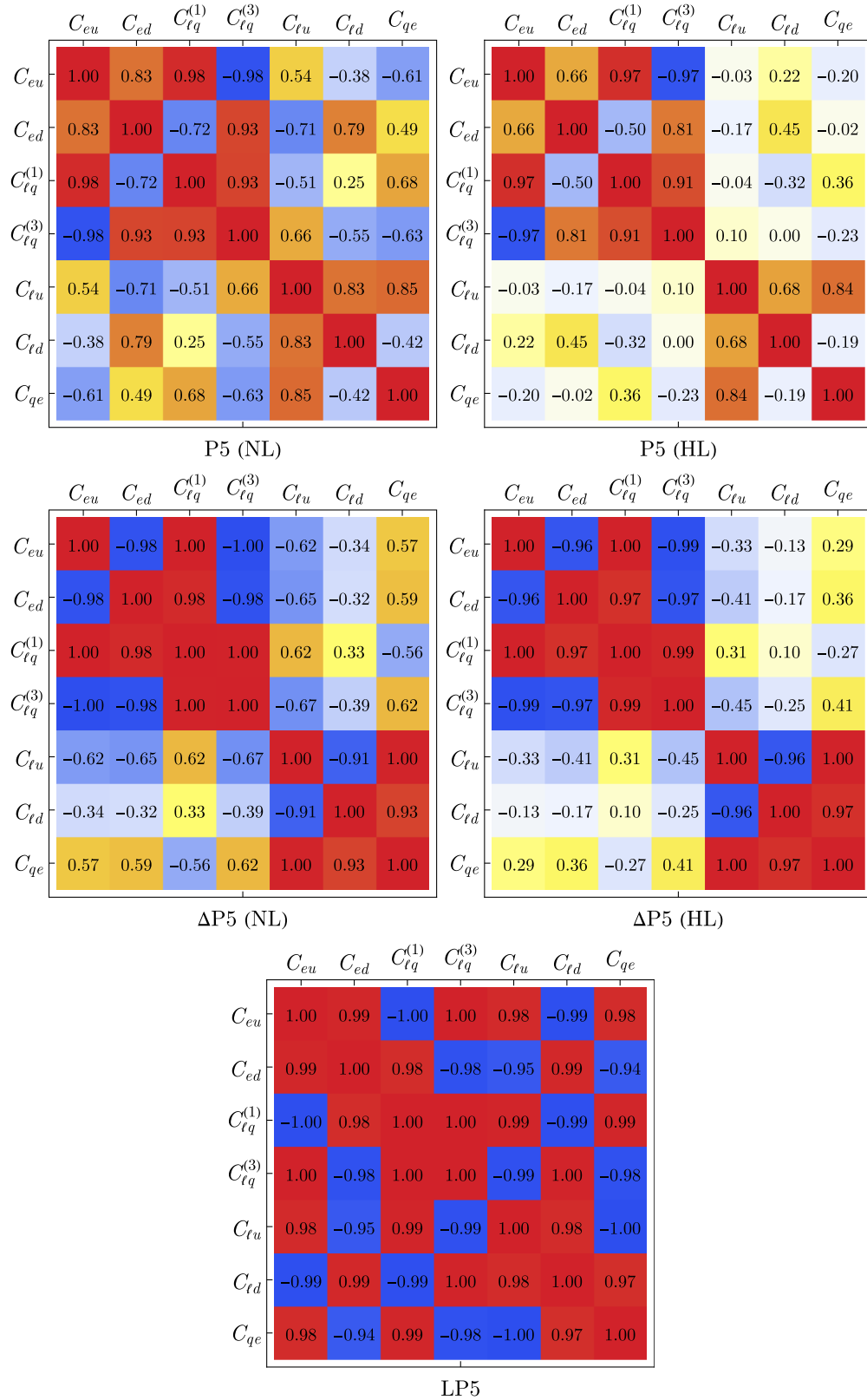


FIG. 60. The same as in Fig. 57 but for P5.

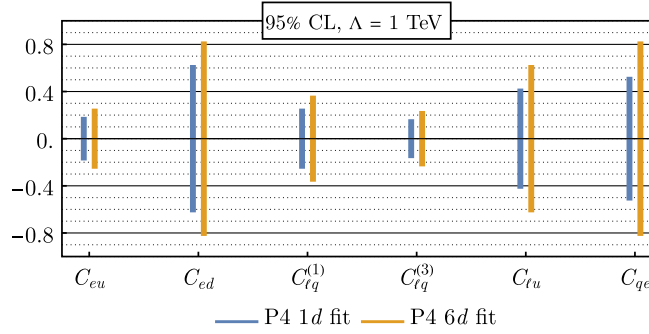


FIG. 61. Comparison of the 95% C.L. bounds on single Wilson coefficients from one-parameter fits to the projections in the 6d fit using the nominal-luminosity dataset P4 at $\Lambda = 1$ TeV.

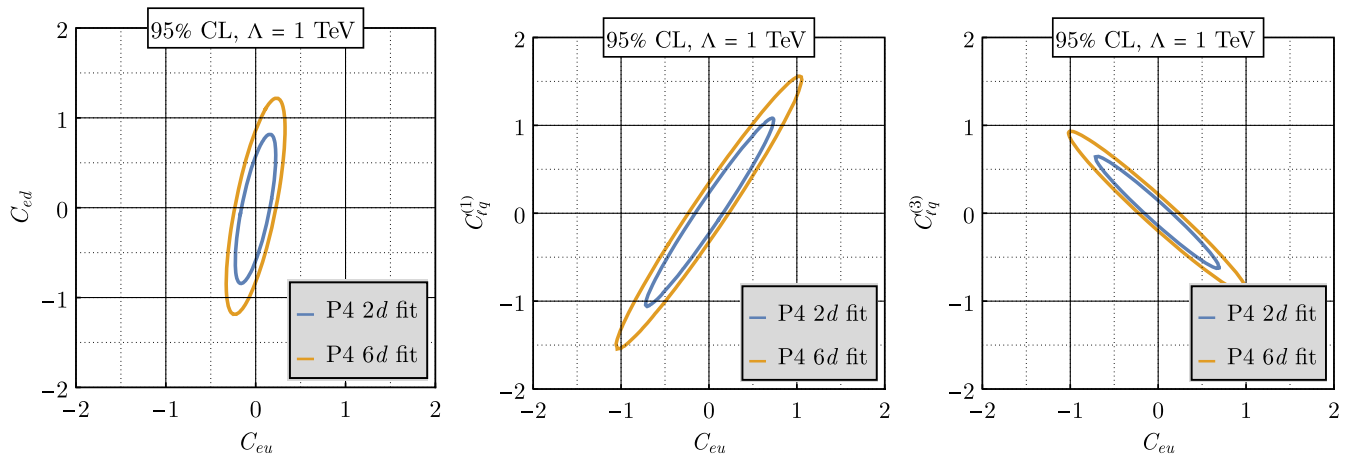


FIG. 62. Comparison of the 95% C.L. ellipses for the Wilson-coefficient pairs (C_{eu}, C_{ed}) , $(C_{eu}, C_{\ell q}^{(1)})$, and $(C_{eu}, C_{\ell q}^{(3)})$ between the original 2d fits and the projections from the simultaneous fit of $C_{eu}, C_{ed}, C_{\ell q}^{(1)}, C_{\ell q}^{(3)}, C_{\ell u}$, and C_{qe} using the dataset P4 at $\Lambda = 1$ TeV.

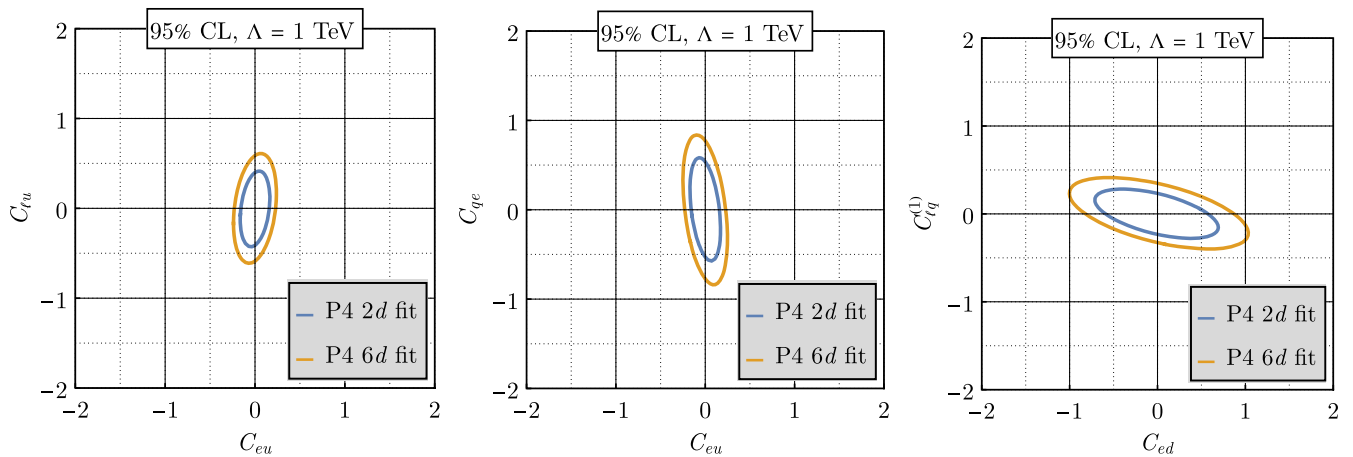


FIG. 63. The same as in Fig. 62 but for $(C_{eu}, C_{\ell u})$, (C_{eu}, C_{qe}) , and $(C_{ed}, C_{\ell q}^{(1)})$.

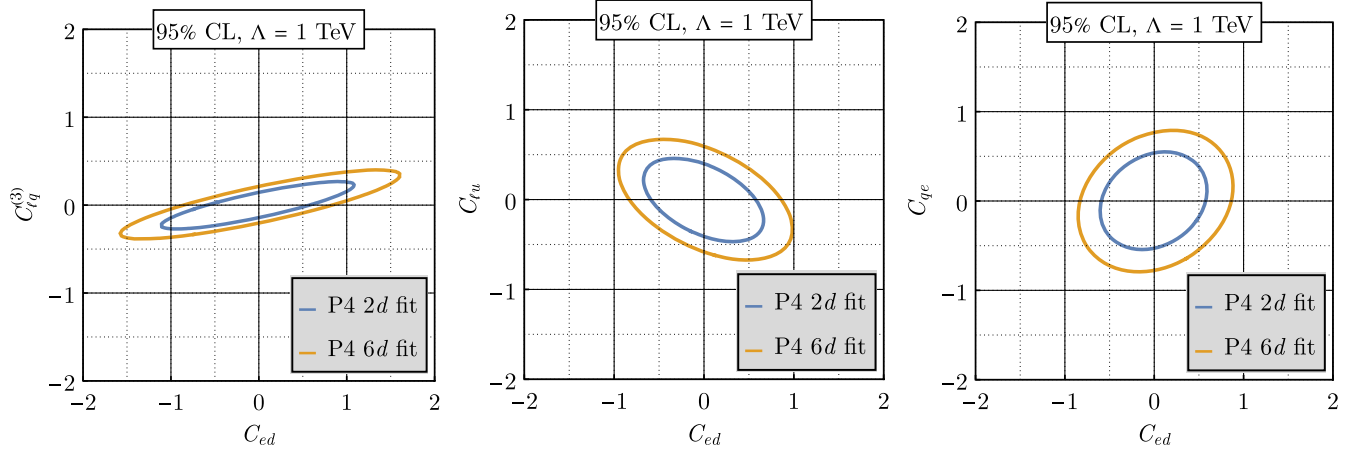


FIG. 64. The same as in Fig. 62 but for $(C_{ed}, C_{\ell q}^{(3)})$, $(C_{ed}, C_{\ell u})$, and (C_{ed}, C_{qe}) .

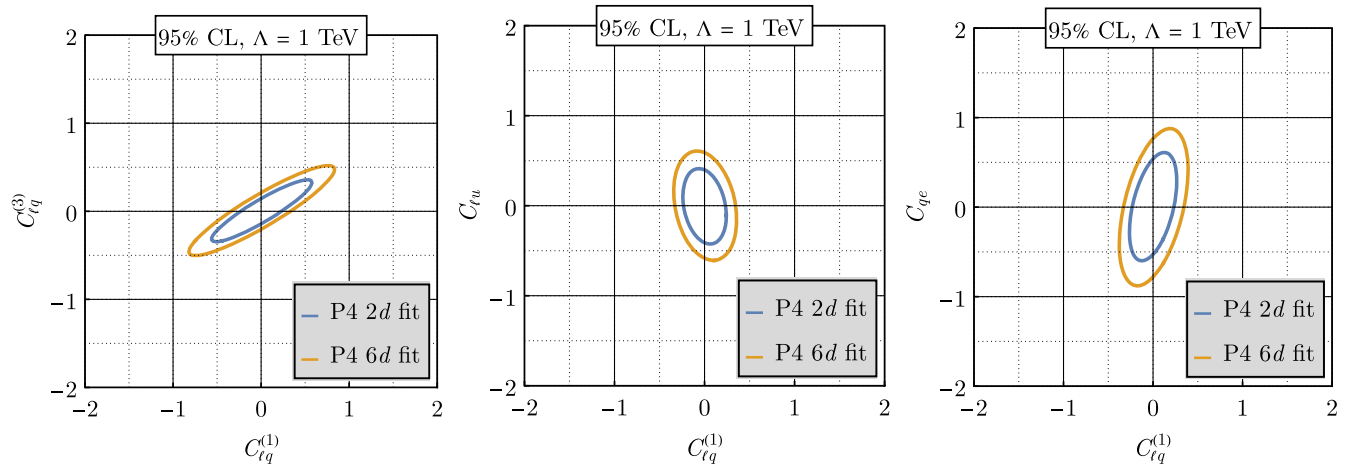


FIG. 65. The same as in Fig. 62 but for $(C_{\ell q}^{(1)}, C_{\ell q}^{(3)})$, $(C_{\ell q}^{(1)}, C_{\ell u})$, and $(C_{\ell q}^{(1)}, C_{qe})$.

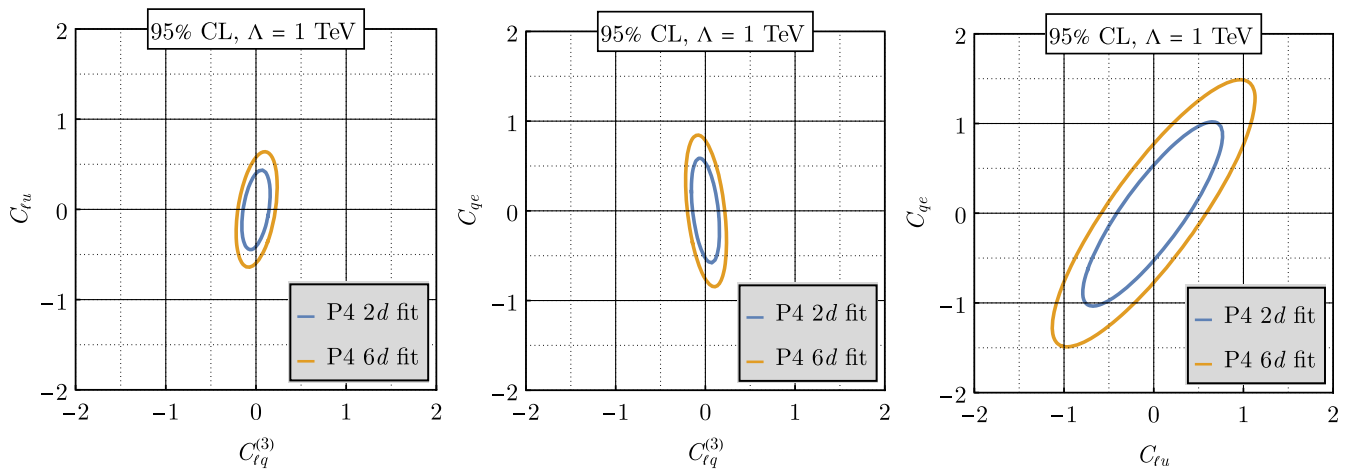


FIG. 66. The same as in Fig. 62 but for $(C_{\ell q}^{(3)}, C_{\ell u})$, $(C_{\ell q}^{(3)}, C_{qe})$, and $(C_{\ell u}, C_{qe})$.

- [1] I. Brivio and M. Trott, *Phys. Rep.* **793**, 1 (2019).
- [2] W. Buchmuller and D. Wyler, *Nucl. Phys.* **B268**, 621 (1986).
- [3] C. Arzt, M. B. Einhorn, and J. Wudka, *Nucl. Phys.* **B433**, 41 (1995).
- [4] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *J. High Energy Phys.* **10** (2010) 085.
- [5] S. Alte, M. König, and W. Shepherd, *J. High Energy Phys.* **07** (2019) 144.
- [6] R. Boughezal, F. Petriello, and D. Wiegand, *Phys. Rev. D* **101**, 116002 (2020).
- [7] G. Panico, L. Ricci, and A. Wulzer, *J. High Energy Phys.* **07** (2021) 086.
- [8] EIC Detector Proposal Advisory Panel, Report from the EIC Detector Proposal Advisory Panel (2022), https://www.bnl.gov/dpamodelmeeting/files/pdf/dpap_report_3-21-2022_final.pdf.
- [9] S. Alioli, R. Boughezal, E. Mereghetti, and F. Petriello, *Phys. Lett. B* **809**, 135703 (2020).
- [10] R. Boughezal, E. Mereghetti, and F. Petriello, *Phys. Rev. D* **104**, 095022 (2021).
- [11] S. Dawson and P. P. Giardino, *Phys. Rev. D* **101**, 013001 (2020).
- [12] J. De Blas, G. Durieux, C. Grojean, J. Gu, and A. Paul, *J. High Energy Phys.* **12** (2019) 117.
- [13] T. Corbett, A. Helset, A. Martin, and M. Trott, *J. High Energy Phys.* **06** (2021) 076.
- [14] P. A. Zyla *et al.* (Particle Data Group), *Prog. Theor. Exp. Phys.* **2020**, 083C01 (2020).
- [15] X. Zheng, J. Erler, Q. Liu, and H. Spiesberger, *Eur. Phys. J. A* **57**, 173 (2021).
- [16] EIC Comprehensive Chromodynamics Experiment Collaboration, Detector Proposal (2021), <https://www.ecce-eic.org/ecce-internal-notes>.
- [17] A. Accardi *et al.*, *Eur. Phys. J. A* **52**, 268 (2016).
- [18] National Academies of Sciences, Engineering, and Medicine, *An Assessment of U.S.-Based Electron-Ion Collider Science* (The National Academies Press, Washington, DC, 2018).
- [19] ECCE Consortium (ECCE Consortium), ECCE Tracking (2021), [ecce-note-det-2021-03](https://www.ecce-eic.org/ecce-internal-notes).
- [20] ECCE Consortium (ECCE Consortium), ECCE Calorimetry (2021), [ecce-note-det-2021-02](https://www.ecce-eic.org/ecce-internal-notes).
- [21] K. Charchula, G. A. Schuler, and H. Spiesberger, *Comput. Phys. Commun.* **81**, 381 (1994).
- [22] <https://github.com/spiesber/DJANGO>.
- [23] S. Schmitt, *EPJ Web Conf.* **137**, 11008 (2017).
- [24] R. Abdul Khalek *et al.*, [arXiv:2103.05419](https://arxiv.org/abs/2103.05419).
- [25] R. D. Ball *et al.* (NNPDF Collaboration), *Eur. Phys. J. C* **77**, 663 (2017).
- [26] E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, and J. Rojo (NNPDF Collaboration), *Nucl. Phys.* **B887**, 276 (2014).
- [27] F. Willeke and J. Beebe-Wang, Electron Ion Collider Conceptual Design Report (2021), <https://www.osti.gov/biblio/1765663-electron-ion-collider-conceptual-design-report>.
- [28] S. Carrazza, C. Degrande, S. Iranipour, J. Rojo, and M. Ubiali, *Phys. Rev. Lett.* **123**, 132001 (2019).
- [29] A. Greljo, S. Iranipour, Z. Kassabov, M. Madigan, J. Moore, J. Rojo, M. Ubiali, and C. Voisey, *J. High Energy Phys.* **07** (2021) 122.
- [30] D. Androić *et al.* (Qweak Collaboration), *Nature (London)* **557**, 207 (2018).
- [31] A. M. Sirunyan *et al.* (CMS Collaboration), *Eur. Phys. J. C* **78**, 701 (2018).
- [32] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **09** (2015) 049.
- [33] J. Erler and M. J. Ramsey-Musolf, *Phys. Rev. D* **72**, 073003 (2005).
- [34] T.-J. Hou *et al.*, [arXiv:1908.11394](https://arxiv.org/abs/1908.11394).
- [35] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *Eur. Phys. J. C* **75**, 132 (2015).
- [36] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, *Eur. Phys. J. C* **75**, 204 (2015).
- [37] D. Becker *et al.*, *Eur. Phys. J. A* **54**, 208 (2018).
- [38] The MOLLER Collaboration, The MOLLER Project, Measurement Of a Lepton Lepton Electroweak Reaction, An Ultra-precise Measurement of the Weak Mixing Angle using Miller Scattering, Conceptual Design Report (2020), https://moller.jlab.org/DocDB/0006/000630/002/MOLLER_CDR_final_aug_4_2020.pdf.
- [39] J. P. Chen, H. Gao, T. K. Hemmick, Z. E. Meziani, and P. A. Souder (SoLID Collaboration), [arXiv:1409.7741](https://arxiv.org/abs/1409.7741).
- [40] P. Souder *et al.* (JLab SoLID-PV DIS, experiment E12-10-007), Precision measurement of parity-violation in deep inelastic scattering over a broad kinematic range (2010).
- [41] J. Erler, C. J. Horowitz, S. Mantry, and P. A. Souder, *Annu. Rev. Nucl. Part. Sci.* **64**, 269 (2014).
- [42] <https://www.jlab.org/theory/jam>.
- [43] Y. X. Zhao, A. Deshpande, J. Huang, K. S. Kumar, and S. Riordan, *Eur. Phys. J. A* **53**, 55 (2017).
- [44] J. Arrington *et al.*, [arXiv:2112.00060](https://arxiv.org/abs/2112.00060).
- [45] R. Boughezal, F. Petriello, and D. Wiegand, *Phys. Rev. D* **104**, 016005 (2021).
- [46] J. J. Ethier, G. Magni, F. Maltoni, L. Mantani, E. R. Nocera, J. Rojo, E. Slade, E. Vryonidou, and C. Zhang (SMEFiT Collaboration), *J. High Energy Phys.* **11** (2021) 089.