Neutral-Current Electroweak Physics and SMEFT Studies at the EIC

Radja Boughezal\textsuperscript{1}, Alexander Emmert\textsuperscript{2}, Tyler Kutz\textsuperscript{3}, Sonny Mantry\textsuperscript{4}, Michael Nyč\textsuperscript{2}, Frank Petriello\textsuperscript{1,5}, Kağan Şimşek\textsuperscript{5}, Daniel Wiegand\textsuperscript{5}, Xiaochao Zheng\textsuperscript{2}

\textsuperscript{1} Argonne National Laboratory, Lemont, IL, USA
\textsuperscript{2} University of Virginia, Charlottesville, VA, USA
\textsuperscript{3} Massachusetts Institute of Technology, Cambridge, MA, USA
\textsuperscript{4} University of North Georgia, Dahlonega, GA, USA
\textsuperscript{5} Northwestern University, Evanston, IL, USA

We study the potential for precision electroweak (EW) measurements and beyond-the-Standard Model (BSM) searches using cross-section asymmetries in neutral-current (NC) 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 ten-fold luminosity upgrade of the EIC beyond its initial decade of running. We use the SM effective field theory (SMEFT) framework to parameterize BSM effects and focus on semi-leptonic 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 EIC Comprehensive Chromodynamics Experiment (ECCE) 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.

CONTENTS

I. Introduction 2

II. Neutral-current DIS Measurements at the EIC
   A. Deep Inelastic Scattering and the SMEFT Formalism 3
   B. Measurement of Parity-Violating Asymmetries at the EIC 3
   C. Measurement of Lepton-Charge Asymmetries at the EIC 10

III. Projection of Parity-Violation and Lepton-Charge Asymmetry Data
   A. ECCE Detector Configuration for Inclusive Neutral-Current Study 10
   B. Simulation with Fast Smearing 11
   C. Event Selection 12
   D. Integrated Luminosity 12
   E. Statistical Uncertainty Projection for Parity-Violating Asymmetries 13
   F. Statistical and QED Uncertainty Projection for Lepton-Charge Asymmetries 13
   G. Projection for High-Luminosity EIC 14

IV. Pseudodata Generation and the Uncertainty Matrix
   A. Pseudodata for Parity-Violating Asymmetries 15
   B. Pseudodata for Lepton-Charge Asymmetries 15
   C. Uncertainty Matrix 15
   D. Comparison of uncertainty components 15
     1. Individual uncertainty components 15
     2. Total uncertainties for nominal luminosity vs. high luminosity 19

V. Extraction of the SM Weak Mixing Angle 20

VI. Framework for the SMEFT Analysis
   A. Data Generation and Selection 22
   B. Structure of the SMEFT asymmetry corrections 22
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 under design. 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, $A^{(e)}_{\text{PV}}$, or the proton (ion) spin-flip, $A^{(p(D))}_{\text{PV}}$. 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 above, it also has 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 $A^{(e)}_{\text{PV}}$ and $A^{(p(D))}_{\text{PV}}$. In addition to determining the BSM reach of PV observables, we consider the reach of the lepton-charge asymmetry $A^{(p(D))}_{\text{LC}}$ 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-6 operator basis of SMEFT for on-shell fields has been completely classified (there is a dimension-5 operator that violates 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-TeV 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 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 Section III 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 Section III we present data simulation based on the design of the 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. Generation of pseudodata, as well as the uncertainty matrix, are presented in Section IV followed by extractions of the EW mixing angle in Section V. In Section 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 Section VII. We also show an example fit in which six Wilson coefficients are turned on simultaneously, in order to demonstrate that EIC data is capable of removing all degeneracies in the semi-leptonic four-fermion operator parameter space. We conclude in Section 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 formulae to include contributions from SMEFT operators, which encode new physics at an energy level \(\Lambda\) that lies well beyond the electroweak scale. We denote a lepton scattering off a nucleus as:

\[ \ell(k) + H(P) \rightarrow \ell(k') + X , \]

where \(\ell\) 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:

\[ s = (P + k)^2 , \quad Q^2 = -(k - k')^2 , \]
\[ x = \frac{Q^2}{2P \cdot (k - k')} , \]
\[ y = \frac{P \cdot (k - k')}{P \cdot k} , \]
\[ W^2 = (P + k - k')^2 , \]

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 single-photon exchange, single-Z-boson exchange, and the SMEFT contact interactions. The SMEFT Lagrangian that describes these contact interactions has the form:

\[ \mathcal{L}_{\text{SMEFT}} = \frac{1}{\Lambda^2} \sum_r C_r \mathcal{O}_r + \cdots , \tag{7} \]

where the summation index \( r \) runs over the set of dimension-6 SMEFT operators and the ellipsis denotes SMEFT operators of mass-dimension greater than 6. We restrict our analysis to include only the effects of dimension-6 SMEFT operators since the higher-dimensional operators are formally suppressed by additional powers of \( E^2/\Lambda^2 \), 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-6 operator and \( C_r \) is the corresponding (dimensionless) Wilson coefficient arising from integrating out the new-physics degrees of freedom at the scale \( \Lambda \). 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-6 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 data sets such as precision Z-pole observables [11,12] and we neglect them here. The above assumptions leave us with the seven Wilson coefficients associated with the listed operators which 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) after electroweak symmetry breaking, which is a customary parameterization (see e.g. [14]):

\[ \mathcal{L}_{\text{SMEFT}} = \frac{1}{\Lambda^2} \sum_r \tilde{C}_r \left\{ \sum_f \bar{e} \gamma^\mu (\tilde{c}_{V,f} - \tilde{c}_{A,f} \gamma_5) e \bar{q} \gamma_\mu (\tilde{c}_{V,f} - \tilde{c}_{A,f} \gamma_5) q + \cdots \right\} , \tag{8} \]

where the specific values of the vector and axial-vector couplings---\( \tilde{c}_{V,q} \) and \( \tilde{c}_{A,q} \), 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 \( \tilde{c}_{V,q}, \tilde{c}_{A,q} \). We specify in Table I the exact definitions that we use. These couplings are analogous to the vector and axial-vector couplings, \( g_A^V \) and \( g_A^A \), of the Z-boson but are instead generated from integrating out UV physics associated with the scale \( \Lambda \).

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

\[ \mathcal{M} = \mathcal{M}_\gamma + \mathcal{M}_Z + \mathcal{M}_r , \tag{9} \]

where \( \mathcal{M}_\gamma, \mathcal{M}_Z, \) and \( \mathcal{M}_r \) denote the contributions from single-photon exchange, 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-6 SMEFT operators that scale as \( 1/\Lambda^2 \) are kept, the total amplitude squared can be written as:

\[ |\mathcal{M}|^2 = |\mathcal{M}_\gamma|^2 + 2 \mathcal{M}_\gamma \mathcal{M}_Z + \mathcal{M}_{ZZ} + 2 \mathcal{M}_\gamma \mathcal{M}_r + 2 \mathcal{M}_Z \mathcal{M}_r , \tag{10} \]

where \( |\mathcal{M}_\gamma|^2, |\mathcal{M}_Z|^2, 2 \mathcal{M}_\gamma \mathcal{M}_Z = \mathcal{M}_\gamma^2 + \mathcal{M}_Z^2 + \mathcal{M}_{\gamma Z}, 2 \mathcal{M}_\gamma \mathcal{M}_r = \mathcal{M}_\gamma^2 + \mathcal{M}_r^2 + \mathcal{M}_\gamma \mathcal{M}_r + \mathcal{M}_\gamma \mathcal{M}_r^*, 2 \mathcal{M}_Z \mathcal{M}_r = \mathcal{M}_Z^2 + \mathcal{M}_Z \mathcal{M}_r + \mathcal{M}_Z \mathcal{M}_r^*, \) and \( 2 \mathcal{M}_Z \mathcal{M}_r = \mathcal{M}_Z^2 + \mathcal{M}_Z \mathcal{M}_r + \mathcal{M}_Z \mathcal{M}_r^* \). These denote the amplitudes of the single-photon exchange, single-Z-boson exchange, the interference between the single-photon and single-Z-boson exchange, the interference between the single-photon exchange and the
SMEFT, and the interference between the single-Z-boson exchange and SMEFT, respectively. Here, we ignore the $|\mathcal{M}_{\ell Z}|^2$ contribution since it scales as $1/\Lambda^4$, formally the same size as contributions from dimension-8 SMEFT operators interfering with the SM.

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:

$$\frac{d^2\sigma}{dx\,dy} = \frac{2\pi y a^2}{Q^4} \left\{ \eta^\gamma L_{\mu\nu}^\gamma W_{\mu\nu}^\gamma + \eta^\gamma L_{\mu\nu}^Z W_{\mu\nu}^Z + \eta^Z L_{\mu\nu}^Z W_{\mu\nu}^Z + \sum_r \xi_r^{\gamma r} L_{\mu\nu}^{\gamma r} W_{\mu\nu}^{\gamma r} + \sum_r \xi_r^{Z r} L_{\mu\nu}^{Z r} W_{\mu\nu}^{Z r} \right\},$$

where $\alpha$ is the electromagnetic fine structure constant and $L_{\mu\nu}^{\gamma Z, Z, Z, Z}$ and $W_{\mu\nu}^{\gamma Z, Z, Z, Z}$ 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. $2\mathcal{M}_{\ell r}$ and $2\mathcal{M}_{Z r}$, respectively. For completeness, below we collect some useful results to make the form of the cross section explicit. The dimensionless coefficients $\eta^\gamma, \eta^Z$, $\xi^{\gamma r}$, and $\xi^{Z r}$ are given by:

$$\eta^\gamma = 1,$$

$$\eta^Z = \frac{G_F M_Z^2}{2\sqrt{2} \pi \alpha} \frac{Q^2}{Q^2 + M_Z^2},$$

$$\xi^{\gamma r} = \frac{C_r Q^2}{4\pi \alpha \Lambda^2},$$

$$\xi^{Z r} = \eta^Z \frac{C_r Q^2}{4\pi \alpha \Lambda^2},$$

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

$$L^\gamma_{\mu\nu} = 2\left[ k_{\mu} k_{\nu}^{\prime} + k_{\nu}^{\prime} k_{\mu} - k \cdot k^{\prime} g_{\mu\nu} - i\lambda_e e_{\mu\nu,\beta} k^{\alpha}(k^{\prime})^\beta \right],$$

$$L^Z_{\mu\nu} = -(g_{\nu}^{\prime} - \lambda_e g_A^e) L^\gamma_{\mu\nu},$$

$$L^{\gamma r}_{\mu\nu} = (c_r^e - \lambda_e c_A^e) L^\gamma_{\mu\nu},$$

$$L^{Z r}_{\mu\nu} = -(c_r^e - \lambda_e c_A^e)(g_{\nu}^{\prime} - \lambda_e g_A^e) L^{\gamma r}_{\mu\nu},$$

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

$$\frac{d^2\sigma}{dx\,dy} = \frac{2\pi y a^2}{Q^4} L_{\mu\nu}^{\gamma r} \left\{ \eta^\gamma W_{\mu\nu}^\gamma - \eta^Z (g_{\nu}^{\prime} - \lambda_e g_A^e) W_{\mu\nu}^Z + \eta^Z (g_{\nu}^{\prime} - \lambda_e g_A^e)^2 W_{\mu\nu}^Z + \sum_r \xi_r^{\gamma r} L_{\mu\nu}^{\gamma r} W_{\mu\nu}^{\gamma r} - \sum_r \xi_r^{Z r} L_{\mu\nu}^{Z r} W_{\mu\nu}^{Z r} \right\},$$

where $r = V, A$ and $\mathcal{O}_r$ are given by:

$$C_r^V = \frac{C_r^{V,1}}{4},$$

$$C_r^A = \frac{C_r^{A,1}}{4},$$

$$\mathcal{O}_r^{\gamma r} = (L^\gamma_{\mu\nu} L^r_{\mu\nu})(Q^L_{\gamma r} Q^L_{\mu\nu}),$$

$$\mathcal{O}_r^{Z r} = (L^Z_{\mu\nu} L^r_{\mu\nu})(Q^L_{Z r} Q^L_{\mu\nu}),$$

$$\mathcal{O}_r^{\gamma} = (L^\gamma_{\mu\nu})(Q^L_{\gamma} Q^L_{\mu\nu}),$$

$$\mathcal{O}_r^{Z} = (L^Z_{\mu\nu})(Q^L_{Z} Q^L_{\mu\nu}).$$

| $C_r$ | $\mathcal{O}_r^{\gamma r}$ | $\mathcal{O}_r^{Z r}$ | $\mathcal{O}_r^{\gamma}$ | $\mathcal{O}_r^{Z}$ |
|------|-----------------|-----------------|-----------------|-----------------|
| $C_r^{(2)}$ | $C_r^{(3)}$ | $C_r^{(4)}$ | $C_r^{(5)}$ | $C_r^{(6)}$ |
| $C_r^{(7)}$ | $C_r^{(8)}$ | $C_r^{(9)}$ | $C_r^{(10)}$ | $C_r^{(11)}$ |
| $C_r^{(12)}$ | $C_r^{(13)}$ | $C_r^{(14)}$ | $C_r^{(15)}$ | $C_r^{(16)}$ |

**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: $C_r \mathcal{O}_r = \tilde{C}_r \sum_f e_{\gamma f}^{\mu}(\partial_\mu c_r^e - \partial_\mu c_A^e)(c_r^e + c_A^e)\gamma_f + \cdots$. The coefficients $c_r^e, c_A^e$ give the chiral structure of each operator.
For an ion beam (or nuclear target), the neutron PDFs can be related to the proton PDFs by assuming isospin (LO), one has for the structure functions the Callan-Gross relations

\[
W^{i}_{\mu\nu} = \left( -g_{\mu\nu} + \frac{q_{\mu}q_{\nu}}{q^2} \right) F^{i}_{1} + \frac{\hat{P}_{\mu} \hat{P}_{\nu}}{(P \cdot q)} F^{i}_{2} + \frac{i\epsilon_{\mu\nu\alpha\beta}}{2(P \cdot q)} \left[ P^{\alpha} q^{\beta} F^{i}_{3} + 2q^{\alpha} S^{\beta} q_{\beta} \right]
\]

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 \( F^{i}_{1,2,3} \) and \( g_{i,1,4,5} \) 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 \( S^i \) satisfies the constraints \( S^2 = -M^2 \) and \( S \cdot P = 0 \). For longitudinal polarization, it takes the canonical form \( S^\mu = \lambda_H (|p|, E, p/|p|) \), where \( \lambda_H = \pm 1 \) is the nucleon helicity and \( P^\mu = (E, p) \) 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:

\[
F_i = F_{i,SM,NC} + F_{i,SMEFT},
\]

\[
g_i = g_{i,SM,NC} + g_{i,SMEFT},
\]

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

\[
F_{i,SM,NC} = F_{i}^{w} - \eta^{\gamma Z}(g_{V}^{c} - \lambda_{c}g_{A}^{c})F_{i}^{\gamma Z} + \eta^{\gamma Z}(g_{V}^{e} - \lambda_{e}g_{A}^{e})F_{i}^{\gamma Z},
\]

\[
g_{i,SM,NC} = g_{i}^{w} - \eta^{\gamma Z}(g_{V}^{e} - \lambda_{e}g_{A}^{e})g_{i}^{\gamma Z} + \eta^{\gamma Z}(g_{V}^{e} - \lambda_{e}g_{A}^{e})g_{i}^{\gamma Z},
\]

and the SMEFT contributions are given by:

\[
F_{i,SMEFT} = \sum_{r} \xi^{C_{r}}(c_{V}^{c} - \lambda_{c}c_{A}^{c})F_{i}^{C_{r}} - \sum_{r} \xi^{Z_{r}}(c_{V}^{e} - \lambda_{e}c_{A}^{e})F_{i}^{Z_{r}},
\]

\[
g_{i,SMEFT} = \sum_{r} \xi^{C_{r}}(c_{V}^{e} - \lambda_{e}c_{A}^{e})g_{i}^{C_{r}} - \sum_{r} \xi^{Z_{r}}(c_{V}^{e} - \lambda_{e}c_{A}^{e})g_{i}^{Z_{r}}.
\]

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:

\[
\left[ F_2^{2}, F_2^{Z}, F_2^{Z}, F_2^{z}, F_2^{Z}, F_2^{Z} \right] = \sum_{f} \left[ Q_{f}^{2}, 2Q_{f}g_{f}^{c}, g_{f}^{e}, g_{f}^{e}, 2Q_{f}c_{f}^{c}, 2g_{f}^{c}c_{f}^{c} + g_{f}^{e}c_{f}^{e} \right] (q_{f} + \bar{q}_{f}),
\]

\[
\left[ \bar{F}_3^{2}, F_3^{Z}, F_3^{Z}, F_3^{z}, F_3^{Z}, F_3^{Z} \right] = \sum_{f} \left[ 0, 2Q_{f}g_{f}^{c}, 2g_{f}^{e}, 2Q_{f}c_{f}^{c}, 2g_{f}^{c}c_{f}^{c} + g_{f}^{e}c_{f}^{e} \right] (q_{f} - \bar{q}_{f}),
\]

\[
\left[ g_1^{w}, g_1^{Z}, g_1^{Z}, g_1^{z}, g_1^{Z}, g_1^{Z} \right] = \sum_{f} \left[ 0, 2Q_{f}g_{f}^{c}, 2g_{f}^{e}, 2Q_{f}c_{f}^{c}, 2g_{f}^{c}c_{f}^{c} + g_{f}^{e}c_{f}^{e} \right] (\Delta q_{f} + \Delta \bar{q}_{f}),
\]

\[
\left[ g_2^{w}, g_2^{Z}, g_2^{Z}, g_2^{z}, g_2^{Z}, g_2^{Z} \right] = \sum_{f} \left[ 0, Q_{f}g_{f}^{c}, g_{f}^{e}, g_{f}^{e}, Q_{f}c_{f}^{c}, g_{f}^{c}c_{f}^{c} + g_{f}^{e}c_{f}^{e} \right] (\Delta q_{f} - \Delta \bar{q}_{f}),
\]

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 \( F_{i}^{Z} = 2xF_{i}^{1} \) and \( g_{i}^{Z} = 2xg_{i}^{1} \) 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:

\[
q_{u/n}(x, Q^2) = q_{d/p}(x, Q^2),
\]

\[
q_{d/n}(x, Q^2) = q_{u/p}(x, Q^2),
\]

\[
\Delta q_{u/n}(x, Q^2) = \Delta q_{d/p}(x, Q^2),
\]

\[
\Delta q_{d/n}(x, Q^2) = \Delta q_{u/p}(x, Q^2),
\]
while the charm and strange sea quark PDFs are assumed to be identical for the proton and the neutron:

\[ q_{s/n}(x, Q^2) = q_{s/p}(x, Q^2), \]
\[ q_{c/n}(x, Q^2) = q_{c/p}(x, Q^2), \]
\[ \Delta q_{s/n}(x, Q^2) = \Delta q_{s/p}(x, Q^2), \]
\[ \Delta q_{c/n}(x, Q^2) = \Delta q_{c/p}(x, Q^2). \]

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:

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

for quark flavor \( f \).

In terms of the generalized structure functions in Eq. (16), which include dependence on the electron helicity, \( \lambda_e \), as seen in Eqs. (17) and (18), one can write the cross section for given electron and nucleon helicities, including SMEFT contributions, as:

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

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:

\[
d\sigma_0 = \frac{1}{4} \left[ d\sigma|_{\lambda_e = +1, \lambda_H = +1} + d\sigma|_{\lambda_e = +1, \lambda_H = -1} + d\sigma|_{\lambda_e = -1, \lambda_H = +1} + d\sigma|_{\lambda_e = -1, \lambda_H = -1} \right],
\]
\[
d\sigma_e = \frac{1}{4} \left[ d\sigma|_{\lambda_e = +1, \lambda_H = +1} + d\sigma|_{\lambda_e = +1, \lambda_H = -1} - d\sigma|_{\lambda_e = -1, \lambda_H = +1} - d\sigma|_{\lambda_e = -1, \lambda_H = -1} \right],
\]
\[
d\sigma_H = \frac{1}{4} \left[ d\sigma|_{\lambda_e = +1, \lambda_H = +1} - d\sigma|_{\lambda_e = +1, \lambda_H = -1} + d\sigma|_{\lambda_e = -1, \lambda_H = +1} - d\sigma|_{\lambda_e = -1, \lambda_H = -1} \right],
\]
\[
d\sigma_{eH} = \frac{1}{4} \left[ d\sigma|_{\lambda_e = +1, \lambda_H = +1} - d\sigma|_{\lambda_e = +1, \lambda_H = -1} - d\sigma|_{\lambda_e = -1, \lambda_H = +1} + d\sigma|_{\lambda_e = -1, \lambda_H = -1} \right],
\]

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:

\[
\frac{d^2 \sigma_0}{dx \, dy} = \frac{4 \pi \alpha^2}{xyQ^2} \left\{ xy^2 \left[ F_1^e - g_V^e \eta_F Z F_1^e + (g_V^e + g_A^e)^2 \eta_Z F_1^e \right] \right.
+ (1-y) \left[ F_2^e - g_V^e \eta_F Z F_2^e + (g_V^e + g_A^e)^2 \eta_Z F_2^e \right] \right.
- \frac{xy}{2} (2-y) \left[ g_A^e \eta_F Z F_3^e - 2 g_V^e g_A^e \eta_Z F_3^e \right] \right\},
\]
\[
\frac{d^2 \sigma_e}{dx \, dy} = \frac{4 \pi \alpha^2}{xyQ^2} \left\{ xy^2 \left[ 2 g_A^e \eta_F Z F_1^e - 2 g_V^e g_A^e \eta_Z F_1^e \right] + (1-y) \left[ g_A^e \eta_F Z F_2^e - 2 g_V^e g_A^e \eta_Z F_2^e \right] \right.
+ \frac{xy}{2} (2-y) \left[ g_V^e \eta_F Z F_3^e - (g_V^e + g_A^e)^2 \eta_Z F_3^e \right] \right\},
\]
\[
\frac{d^2 \sigma_H}{dx \, dy} = \frac{4 \pi \alpha^2}{xyQ^2} \left\{ (2-y) \, xy \left[ 2 g_A^e \eta_F Z g_1^e - 2 g_V^e g_A^e \eta_Z g_1^e \right] \right.
+ (1-y) \left[ g_V^e \eta_F Z g_1^e - (g_V^e + g_A^e)^2 \eta_Z g_4^e \right] \right\}. 
\]
\[
\frac{d^2 \sigma_{eH}}{dx \, dy} = 8 \pi^2 \alpha^2 \frac{\sigma_{eH}}{xyQ^2} \left \{ (2 - y) \frac{dy}{2} \left ( \frac{c_1}{g_1} \right )^2 g_{1r}^2 + (g_{1r}^2 + g_A^2) \eta z g_{1z}^2 \right \}.
\]

The SMEFT contributions are:

\[
\begin{align*}
\frac{d^2 \sigma_{eH}^{\text{SMEFT}}}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} \sum_r \left \{ xy^2 (c_1^e \xi_{1r} F_1^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z F_1^{2z}) \\
&\quad + (1 - y)(c_1^e \xi_{1r} F_2^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z F_2^{2z}) \\
&\quad + \frac{xy}{2} (2 - y)(c_1^e, \xi_{1r} F_3^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z F_3^{2z}) \right \}, \\
\frac{d^2 \sigma_{eH}^{\text{SMEFT}}}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} \sum_r \left \{ xy^2 (c_1^e \xi_{1r} g_1^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_1^{2z}) \\
&\quad + (1 - y)(c_1^e \xi_{1r} g_2^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_2^{2z}) \\
&\quad + \frac{xy}{2} (2 - y)(c_1^e, \xi_{1r} g_3^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_3^{2z}) \right \}, \\
\frac{d^2 \sigma_{eH}^{\text{SMEFT}}}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} \sum_r \left \{ xy^2 (c_1^e \xi_{1r} g_1^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_1^{2z}) \\
&\quad + (1 - y)(c_1^e \xi_{1r} g_2^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_2^{2z}) \\
&\quad + \frac{xy}{2} (2 - y)(c_1^e, \xi_{1r} g_3^{2r} - (c_1^e g_1^e + c_A^e g_A^e) \xi z g_3^{2z}) \right \}.
\end{align*}
\]

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:

\[
\begin{align*}
\frac{d^2 \sigma_{eH}^+}{dx \, dy} - \frac{d^2 \sigma_{eH}^-}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} g_A \left \{ xy^2 (2 - y) \left ( \eta z F_3^{2z} - 2 g_{1r}^e \eta z g_{1z}^{2r} \right ) \right \} \\
&\quad - 8 \pi^2 \alpha^2 \sum_r \frac{xy}{2} (2 - y) c_A^e (\xi_{1r} F_3^{2r} + 2 g_{1r}^e \xi z F_3^{2z}) \right \}, \\
\frac{d^2 \sigma_{eH}^+}{dx \, dy} - \frac{d^2 \sigma_{eH}^-}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} g_A \left \{ 2(1 - y) \left ( \eta z F_2^{2z} - 2 g_{1r}^e \eta z g_{1z}^{2r} \right ) + xy^2 \left ( \eta z F_3^{2z} - 2 g_{1r}^e \eta z F_1^{2z} \right ) \right \} \\
&\quad + 8 \pi^2 \alpha^2 \sum_r [(1 - y) c_A^e (\xi_{1r} F_2^{2r} + 2 g_{1r}^e \xi z F_2^{2z}) + xy^2 c_A^e (\xi_{1r} F_3^{2r} + 2 g_{1r}^e \xi z F_1^{2z}) \right \}, \\
\frac{d^2 \sigma_{eH}^+}{dx \, dy} - \frac{d^2 \sigma_{eH}^-}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} g_A \left \{ 2xy (2 - y) \left ( \eta z g_1^{2z} - 2 g_{1r}^e \eta z g_{1z}^{2r} \right ) \right \} \\
&\quad + 8 \pi^2 \alpha^2 \sum_r \frac{xy}{2} (2 - y) c_A^e (\xi_{1r} g_1^{2r} + 2 g_{1r}^e \xi z g_1^{2z}) \right \}, \\
\frac{d^2 \sigma_{eH}^+}{dx \, dy} - \frac{d^2 \sigma_{eH}^-}{dx \, dy} &= \frac{4 \pi^2 \alpha^2}{xyQ^2} g_A \left \{ 2(1 - y) \left ( \eta z g_1^{2z} - 2 g_{1r}^e \eta z g_{1z}^{2r} \right ) + xy^2 \left ( \eta z g_3^{2z} - 2 g_{1r}^e \eta z g_{3z}^{2r} \right ) \right \} \\
&\quad - 8 \pi^2 \alpha^2 \sum_r [(1 - y) c_A^e (\xi_{1r} g_1^{2r} + 2 g_{1r}^e \xi z g_1^{2z}) + xy^2 c_A^e (\xi_{1r} g_3^{2r} + 2 g_{1r}^e \xi z g_3^{2z}) \right \}.
\end{align*}
\]

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 unpo-
polarized hadrons, referred to as the “unpolarized PV asymmetry”:

\[ A^{(c)}_{P V} \equiv \frac{d\sigma}{d\sigma_0}, \]  

(28)

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

\[ A^{(H)}_{P V} \equiv \frac{d\sigma_H}{d\sigma_0}. \]  

(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:

\[ A^{(H)}_{L C} = \frac{d\sigma_0^+ - d\sigma_0^-}{d\sigma_0^+ + d\sigma_0^-}, \]  

(30)

will provide additional constraints on SMEFT interactions. On the other hand, the double-spin asymmetry, \( A^{(eH)}_{P V} \equiv d\sigma_{eH}/d\sigma_0 \), 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:

\[ d\sigma = d\sigma_0 + P_e d\sigma_e + P_H d\sigma_H + P_e P_H d\sigma_{eH}, \]  

(31)

where \( P_e \) and \( P_H \) have the same sign as the respective beam helicities, \( \lambda_e \) and \( \lambda_H \), 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 time scale. 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:

\[ N^{++} = a_{det}\mathcal{L}^{++} (d\sigma_0 + |P_e^{++}|d\sigma_e + |P_H^{++}|d\sigma_H + |P_e^{++}||P_H^{++}|d\sigma_{eH}), \]  

(32)

\[ N^{+-} = a_{det}\mathcal{L}^{+-} (d\sigma_0 + |P_e^{+-}|d\sigma_e - |P_H^{+-}|d\sigma_H - |P_e^{+-}||P_H^{+-}|d\sigma_{eH}), \]  

(33)

\[ N^{-+} = a_{det}\mathcal{L}^{-+} (d\sigma_0 - |P_e^{-+}|d\sigma_e + |P_H^{-+}|d\sigma_H - |P_e^{-+}||P_H^{-+}|d\sigma_{eH}), \]  

(34)

\[ N^{--} = a_{det}\mathcal{L}^{--} (d\sigma_0 - |P_e^{--}|d\sigma_e - |P_H^{--}|d\sigma_H + |P_e^{--}||P_H^{--}|d\sigma_{eH}), \]  

(35)

where \( \mathcal{L}^{ij} \) 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:

\[ d\sigma_0 = \frac{1}{4} (d\sigma^{++} + d\sigma^{+-} + d\sigma^{-+} + d\sigma^{--}), \]  

(36)

\[ d\sigma_e = \frac{1}{4|P_e|} (d\sigma^{++} + d\sigma^{+-} - d\sigma^{-+} - d\sigma^{--}), \]  

(37)

\[ d\sigma_H = \frac{1}{4|P_H|} (d\sigma^{++} - d\sigma^{+-} + d\sigma^{-+} - d\sigma^{--}), \]  

(38)

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

(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 very short time-scale, the factor \( a_{\text{det}} \) 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:

\[
A_{PV}^{(c)}(e) = \frac{\sigma_e}{\sigma_0} = \frac{1}{|P_e|} \frac{Y^{++} + Y^{+-} - Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}},
\]

(40)

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

\[
A_{PV}^{(H)} = \frac{\sigma_H}{\sigma_0} = \frac{1}{|P_H|} \frac{Y^{++} - Y^{+-} + Y^{-+} - Y^{--}}{Y^{++} + Y^{+-} + Y^{-+} + Y^{--}}.
\]

(41)

The design of the EIC requires that the point-to-point luminosity uncertainty be at \( 10^{-4} \) level. Therefore, the dominant experimental uncertainty would come from electron and proton (ion) polarimetry for \( A_{PV}^{(c)} \) 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 time-scale, 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 EIC Comprehensive Chromodynamics Experiment (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 µRWell 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 shashlik-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 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 yield 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 Djangoh 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 × 275(137), 10 × 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 Djangoh output, and the physics cross section and parity-violating asymmetries are calculated event-by-event using a modified user routine of Djangoh. 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 pseudo-rapidity range \( \eta = (-3.5625, 3.5625) \) and 1 GeV-wide bins in the transverse momentum, \( p_T \). For each Djangoh-simulated event, smearing in the electron momentum, \( p \), and polar and azimuthal angles \( \theta \) and \( \phi \) 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 \( Q^2_{\text{min}} = 50 \text{ GeV}^2 \) 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 as the double-angle method may provide better DIS event identification for certain kinematic range and thus improve precision of the analysis.

![RMS values for fast-smearing spectra](image)
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 Djangoh (with $Q^2_{\text{min}} = 1.0$ GeV$^2$), and another Monte-Carlo simulation of photoproduction events are generated by Pythia (version 6.428, with $Q^2_{\text{min}} = 0$). 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:

- 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.

- Secondly, 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 ($Q^2_{\text{det}} > 1.0$ GeV$^2$) in order to avoid regions with severe bin migration and unfolding uncertainty ($y_{\text{det}} < 0.1$), to avoid regions with 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 ($y_{\text{det}} > -3.5$ and $y_{\text{det}} < 3.5625$), 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 $A_{\text{PV}}^{(c)}$ and $A_{\text{PV}}^{(H)}$ after unfolding are shown in Figs. 3 and 4, respectively, for $18 \times 275$ 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 11 and will be referred to as “Nominal Luminosity (NL)” hereafter. As a comparison with the weak mixing angle.
EIC/ECCE Preliminary

Djangoh e+p 18x275 GeV, 100 fb⁻¹

e+p → e' (+X); P₆ = 70%
with event selection

| dApv(H)/Apv(H) (unfolded) |
|---------------------------|

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.

Extraction presented in the YR, we also carry out projections for 100 fb⁻¹ 18 × 275 GeV eD collisions as the “YR reference point”. We abbreviate the ep pseudodata sets as P1, P2, P3, P4, and P5 and the eD pseudodata sets as D1, D2, D3, D4, and D5; see Table II. The YR reference point is denoted by P6. Simulated pseudodata sets with polarized hadrons are indicated as ∆D1–5 and ∆P1–6, while positron data sets 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 ten-fold increase in the integrated luminosity (10× 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× 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 an “High Luminosity (HL)” label hereafter.

Table: 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 |
| | 18 GeV × 275 GeV ep, 100 fb⁻¹ | P6 | 18 GeV × 275 GeV ep, 100 fb⁻¹ |

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.

E. Statistical Uncertainty Projection for Parity-Violating Asymmetries

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

\[ \frac{dA_{\text{stat,measured}}}{A_{PV}} = \frac{1}{\sqrt{N}}, \]

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 beam, between left- and right-handed proton (ion) beam, 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:

\[ \frac{dA_{PV,stat}^{(e)}}{|P_e|} = \frac{1}{\sqrt{N}}, \quad \text{and} \quad \frac{dA_{PV,stat}^{(H)}}{|P_H|} = \frac{1}{\sqrt{N}}. \]
For $A_{PV}^{(e)}$ projections, an electron beam polarization of $P_e = 80\%$ with relative 1% systematic uncertainty from the electron polarimetry is assumed. Similarly, for $A_{PV}^{(H)}$ 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 $A_{PV}^{(e)}$ is rather large because of the much smaller size of $A_{PV}^{(e)}$ than $A_{PV}^{(e)}$.

F. Statistical and QED Uncertainty Projection for Lepton-Charge Asymmetries

As described in Section [III C], to measure the lepton-charge asymmetry $A_{LC}^{(H)}$, 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 Section [III B] apply. The statistical uncertainty in $A_{LC}$ is thus determined by the luminosity of $e^+$ 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 Djangoh version 4.6.19 in both the Born LO (that includes one-boson exchange only) and 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.

![Diagram](image)

**FIG. 5.** Calculation for $A_{LC}$ at the Born (LO) (left) and NLO (right) levels for $e^+p$ vs. $e^-p$ collisions at $18 \times 275$ GeV. The LO calculation includes only the $\gamma Z$ 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.

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 $A_{LC}^{(H)}$. In Appendix [A1] 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 ten-fold increase in the annual luminosity beyond EIC’s initial phase of running, the so-called high-luminosity EIC (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 $A_{PV}^{(e)}$, the ten-fold 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 $A_{PV}^{(e)}$ for $10 \times 275$ GeV $ep$ and $10 \times 137$ 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, $\sigma_{\text{stat}}$; experimental systematic, $\sigma_{\text{sys}}$, which is mainly due to particle background, also including other imperfections of the measurement, and is assumed to be fully uncorrelated; and beam polarimetry, $\sigma_{\text{pol}}$, which is assumed to be fully correlated within data of the same $\sqrt{s}$ and beam type.

For the $b^\text{th}$ 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 pseudo-experimental asymmetry value by:

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

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. Pseudo Data for Lepton-Charge Asymmetries

We consider next unpolarized electron-positron asymmetries with unpolarized hadrons, namely the lepton-charge (LC) asymmetries. The uncertainties used in the data generation are from three sources: statistical, $\sigma_{\text{stat}}$; experimental systematic, $\sigma_{\text{sys}}$, which is mainly due to background and is assumed to be fully correlated; luminosity difference between $e^+$ and $e^-$ runs, $\sigma_{\text{lum}}$, 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:

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

C. Uncertainty Matrix

The uncertainty matrix, $\Sigma^2$, for a given data set with $N_{\text{bin}}$ bins is an $N_{\text{bin}} \times N_{\text{bin}}$ symmetric matrix. It consists of two parts, which we call $\Sigma^2_0$ and $\Sigma^2_{\text{pdf}}$:

$$(\Sigma^2)_{bb'} = (\Sigma^2_0)_{bb'} + (\Sigma^2_{\text{pdf}})_{bb'}.$$

The first part of the matrix, $\Sigma^2_0$, 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 $\Sigma^2_0$ must be absolute; relative uncertainties are converted to absolute ones by multiplying with the theoretical SM prediction, $(A_{SM})_{\text{theo},0,b}$. Computing the central member of the PDF set taken into account. The first part of the matrix then takes the form:

$$\Sigma^2_0 = \begin{pmatrix}
\sigma_1^2 & \rho_{12} & \tilde{\sigma}_1 & \cdots & \rho_{1N_{\text{bin}}} & \tilde{\sigma}_1 N_{\text{bin}} \\
\rho_{12} & \sigma_2^2 & \cdots & \rho_{2N_{\text{bin}}} & \tilde{\sigma}_2 N_{\text{bin}} \\
\tilde{\sigma}_1 & \cdots & \tilde{\sigma}_2 & \cdots & \tilde{\sigma}_N_{\text{bin}} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\rho_{1N_{\text{bin}}} & \cdots & \rho_{2N_{\text{bin}}} & \cdots & \sigma_{N_{\text{bin}}}^2 \end{pmatrix}_{\text{sym}}.$$

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

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

and for the off-diagonal elements:

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

For the LC asymmetries, we have for the diagonal elements:

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

and for the off-diagonal elements:

\[
\tilde{\sigma}_b = \sigma_{\text{lum},b}.
\]

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, \(\Sigma_{\text{pdf}}^2\), 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, \(\Lambda_{\text{SM},m}^{\text{theo}}\), and theoretical SM asymmetry predictions evaluated at all other members of the PDF set under consideration, \(\Lambda_{\text{SM},m}^{\text{theo}}\), where \(m = 1, 2, \ldots, N_{\text{PDF}}\) with \(N_{\text{PDF}}\) 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:

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

For replica-based PDF sets, this expression becomes:

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

### 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 \(\bar{u}d\) 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 data sets involving unpolarized deuteron with the \(\bar{u}d\) 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.

Fig. 6 shows the comparison of the uncertainty components for the data set D4 in the \(\bar{u}d\) 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 \(Q^2\), then from small to large \(x\) within each \(Q^2\) bin, leading to the observed oscillatory behavior. When we turn on the sea quark contributions, the unpolarized deuteron data sets 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 data sets P5 and ∆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 ∆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 data set 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, two 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 Djangoh. The two largest sources of uncertainty throughout the entire kinematic range are the luminosity and statistical uncertainties. 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:

- 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, $A^{(e)}_\text{PV}$.

- PDF uncertainties are nearly irrelevant for the asymmetries of unpolarized hadrons, $A^{(e)}_\text{PV}$. They become significant, second to statistical uncertainties, for PV asymmetries of polarized hadrons, $A^{(H)}_\text{PV}$.

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

We now investigate the total uncertainties for the data sets D4, ∆D4, P5, and ∆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 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 be in fact reduced to just the luminosity comparison, i.e. nominal vs. high. Next, for both D4 and P5, the asymmetry $A_{PV}^{(c)}$ 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 as our main data case, the anticipated errors are larger than the asymmetry for all choices of systematic error and luminosity. Only in the very high $Q^2$ bins does a measurement of the asymmetry $A_{PV}^{(H)}$ 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 data sets under consideration. In our best-fit analyses, we thus focus on the data sets with 1% relative systematic uncertainty and nominal luminosity in the uds scenario as our main data sets. 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 data sets 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.

2. Total uncertainties for nominal luminosity vs. high luminosity

![Graph showing total uncertainties combined in quadrature for the data sets D4 and ∆D4 in the uds scenario.](image-url)
• For the unpolarized deuteron data sets, the PDF uncertainties are an order of magnitude smaller than the statistical uncertainties for the nominal luminosity and three to five 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 effect the bounds obtained here. The PDFs are already determined sufficiently well from other experiments for the purposes of this analysis.

• The same statement holds for the P5 unpolarized data sets 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.

• 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 data set. 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 data sets 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 polarized 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 eV 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 $A^{(e)}_{PV}$, where $\sin^2 \theta_W$ enters through the electron coupling $g^e_{V,A}$ 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 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:

$$A_{P_{PV}}^{(e)} = \frac{\left|P_e|\eta_{rZ} \left[ \frac{2}{Q^2} \left( \frac{g_2 g_{F_1}^{YZ} + g_3 \left( \frac{y}{2} - \frac{2M^2 x u}{Q^2} \right) F_2^{\gamma\gamma} + g_5 \left( 2y - \frac{2}{2} \right) F_3^{\gamma\gamma} \right) \right] + 2y F_1^{\gamma\gamma} \right|}{2y F_1^{\gamma\gamma} + \left( \frac{2}{Q^2} \frac{2M^2 x u}{Q^2} \right) F_2^{\gamma\gamma} - \eta_{rZ} \left[ g_2 g_{F_1}^{YZ} + g_5 \left( \frac{y}{2} - \frac{2M^2 x u}{Q^2} \right) F_2^{\gamma\gamma} + g_3 \left( 2y - \frac{2}{2} \right) F_3^{\gamma\gamma} \right]},$$

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 pseudodata set is generated using a reference value of $\sin^2 \theta_W = 0.231$ at the $Z$-pole and the uncertainties in $A_{P_{PV}}^{(e)}$ 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 $\chi^2$ function defined as:

$$\chi^2 = \left( A_{\text{pseudo-data}} - A_{\text{theory}} \right)^T \Sigma_{\text{(2)}-1} \left( A_{\text{pseudo-data}} - A_{\text{theory}} \right),$$

where $A$ is a dimension-$N_{\text{bin}}$ vector with $N_{\text{bin}}$ the total number of $(x, Q^2)$ bins, $\Sigma$ is the uncertainty matrix of dimension $N_{\text{bin}} \times N_{\text{bin}}$, described in Section IV C, and $\sin^2 \theta_W$ to be fitted enters $A_{\text{theory}}$. 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 NNPDF31nlo_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 experimental uncertainty that includes statistical, 1% uncorrelated experimental systematics, 1% correlated systematic experimental systematics, and 1% electron polarimetry. The outermost error 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 NNPDF31NLO are similar. Along with our projection with the EIC annual nominal luminosity, we show the “YR reference point” (blue diamond), obtained from combining 100 fb$^{-1}$ 18 × 275 GeV $ep$ and 10 fb$^{-1}$ 18 × 137 GeV $eD$ pseudodata. Also shown are the expected precision 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 JAM framework \[32\], 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 ten-fold luminosity upgrade.

| Beam type and energy | Label | $ep \ 5 \times \ 100$ | $ep \ 10 \ \times \ 100$ | $ep \ 10 \ \times \ 275$ | $ep \ 18 \ \times \ 275$ |
|---------------------|-------|-------------------------|-------------------------|-------------------------|-------------------------|
|                     | P2    | P3                      | P4                      | P5                      | P6                      |
| Luminosity (fb$^{-1}$) |       |                         |                         |                         |                         |
| $(Q^2)$ (GeV$^2$)     |       |                         |                         |                         |                         |
| $\langle A_{P_{PV}} \rangle (P_e=0.8)$ |       |                         |                         |                         |                         |
| $(dA/A)_{\text{stat}}$ | 1.54% | 0.98%                   | 0.40%                   | 0.80%                   | (0.31%)                 |
| $(dA/A)_{\text{stat+syst}}$ | 1.55% | 1.00%                   | 0.43%                   | 0.81%                   | (0.35%)                 |
| $(dA/A)_{\text{pol}}$ | 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($\sin^2 \theta_W$)$_{\text{stat+syst}}$ | 0.002032 | 0.001299 | 0.000597 | 0.001176 | 0.000516 |
| d($\sin^2 \theta_W$)$_{\text{stat+syst}}$ | 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 |

**TABLE III.** Projected PVDIS asymmetry and fitted results for $\sin^2 \theta_W$ using $ep$ collision data and the nominal annual luminosity. Here, $(Q^2)$ denotes the value averaged over all $(x, Q^2)$ bins, weighted by $(dA/A)^2$ 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.
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 runplan 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 Section 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 × 137 GeV $eD$ 100 fb$^{-1}$: D4, ΔD4, LD4
- 18 GeV × 137 GeV $eD$ 15.4 fb$^{-1}$: D5, ΔD5, LD5
- 10 GeV × 275 GeV $ep$ 100 fb$^{-1}$: P4, ΔP4, LP4
- 18 GeV × 275 GeV $eD$ 15.4 fb$^{-1}$: P5, ΔP5, LP5

For the highest $\sqrt{s}$ but lower-luminosity set D5, ΔD5, P5, and ΔP5, 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 ten-fold higher statistics.

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

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

(56)

These restrictions are designed to remove large uncertainties from non-perturbative QCD and nuclear dynamics that occur at low $Q^2$ and high $x$, where sensitivity to SMEFT effects is anyways expected to be reduced. We note that the condition on $y$ is already applied in the data generation and unfolding stages described in Section III C.

B. Structure of the SMEFT asymmetry corrections

In the computation of SMEFT asymmetry values, $A_{\text{SMEFT}}$, we use the central member of the PDF set under consideration. We use the PDF sets NNPDF31nlo.as.0118 [25] and NNPDFpol11100 [26] for the computation of

| Beam type and energy | $eD$ 5 × 100 | $eD$ 10 × 100 | $eD$ 10 × 137 | $eD$ 18 × 137 | $eD$ 18 × 137 |
|-----------------------|--------------|--------------|--------------|--------------|--------------|
| Label | D2 | D3 | D4 | D5 | N/A |
| Luminosity (fb$^{-1}$) | 36.8 | 44.8 | 100 | 15.4 (10 YR ref) | 687.2 |
| $\langle Q^2 \rangle$ (GeV$^2$) | 160.0 | 316.0 | 403.5 | 687.2 | 687.2 |
| $\langle A_{PV} \rangle$ ($P_e = 0$) | 0.01028 | 0.01923 | 0.02366 | 0.03719 | 0.03719 |
| Experimental | 1.46% | 0.93% | 0.54% | 1.05% (1.31%) | 1.46% (1.66%) |
| with PDF | 1.47% | 0.95% | 0.56% | 1.07% (1.32%) | 1.0% (1.0%) |
| $\Delta A/\Delta$stat | 1.0% | 1.0% | 1.0% | 1.0% | 1.0% |
| $\Delta A/\Delta$stat+bg | 1.0% | 1.0% | 1.0% | 1.0% | 1.0% |
| $\Delta A/\Delta$syst, $\Delta$stat+bg+pol | 1.0% | 1.0% | 1.0% | 1.0% | 1.0% |
| $\Delta A/\Delta$tot, $\Delta$stat+bg+pol | 1.0% | 1.0% | 1.0% | 1.0% | 1.0% |

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.
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 (MS scheme) [33].

unpolarized and polarized PV asymmetries, namely $A^{(c)}_{PV}$ and $A^{(H)}_{PV}$, respectively. We factor out the UV cut-off 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:

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

(57)
or

\[ A_{\text{SMEFT}}(x, Q^2, C_1, C_2) = A_{\text{SM}0}(x, Q^2) + C_1 \delta_1(x, Q^2) + C_2 \delta_2(x, Q^2) . \]  

(58)

Comparing Eq. (57) to (44) or (45), we see that at the end of a multi-run 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 pseudo-data values, \( A^\text{pseudo}_{\text{SM}} \), and obtaining the SMEFT asymmetry expressions, \( A_{\text{SMEFT}} \), we define a \( \chi^2 \) test statistic as:

\[
\chi^2 = \sum_{b=1}^{N_{\text{bin}}} \sum_{b'=1}^{N_{\text{bin}}} \left[ A_{\text{SMEFT},b} - A^\text{pseudo}_{\text{SM},b} \right] \left[ (\Sigma^2)^{-1} \right]_{bb'} \left[ A_{\text{SMEFT},b'} - A^\text{pseudo}_{\text{SM},b'} \right],
\]

(59)

where \( N_{\text{bin}} \) is the number of bins in a given data set. Generically, it looks like:

\[
\chi^2(C) = k_0 + k_1 C + k_2 C^2
\]

(60)

for a single-parameter fit of Wilson coefficient \( C \), or:

\[
\chi^2(C_1, C_2) = k_{00} + k_{10} C_1 + k_{01} C_2 + k_{11} C_1 C_2 + k_{20} C_1^2 + k_{02} C_2^2
\]

(61)

for a two-parameter fit of Wilson coefficients \( C_1 \) and \( C_2 \). The \( \chi^2 \) function is minimized with respect to \( C \) or to \( C_1 \) and \( C_2 \). We obtain the inverse square of the error of the single-parameter best-fit value via:

\[
\frac{1}{\sigma_C^2} = \frac{1}{2} \frac{d^2 \chi^2}{dC^2}
\]

(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:

\[
(V^{-1})_{ij} = \frac{1}{2} \frac{\partial \chi^2}{\partial C_i \partial C_j}
\]

(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:

\[
V = \begin{pmatrix}
\sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 \\
\rho_{12} \sigma_1 \sigma_2 & \sigma_2^2
\end{pmatrix}_{\text{sym}}.
\]

(64)

### 1. Averaging over multiple pseudodata sets

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

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

(65)

and the average uncertainty of this mean value is obtained via:

\[
\frac{1}{\sigma_{\bar{C}}^2} = \frac{1}{N_{\text{exp}}} \sum_{e=1}^{N_{\text{exp}}} \frac{1}{\sigma_{C,e}^2}.
\]

(66)

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

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

(67)

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

\[
V^{-1} = \frac{1}{N_{\text{exp}}} \sum_{e=1}^{N_{\text{exp}}} (V^{-1})_e.
\]

(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 data set by a factor \(N_{\text{exp}}\). 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 multi-run fit can be expressed as:

\[
\frac{(C - \bar{C})^2}{\sigma_C^2} = \Delta \chi^2,
\]

(69)

and hence we can express the fitted result and the uncertainty of coefficient \(C\) as:

\[
C = C_{\text{best}} \pm \sqrt{\Delta \chi^2} \sigma_C.
\]

(70)

For a two-parameter multi-run fit, the ellipse equation reads:

\[
\begin{pmatrix}
C_1 - \bar{C}_1 \\
C_2 - \bar{C}_2
\end{pmatrix}^T V^{-1} \begin{pmatrix}
C_1 - \bar{C}_1 \\
C_2 - \bar{C}_2
\end{pmatrix} = \Delta \chi^2
\]

(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\% CL, we have \(\Delta \chi^2 = 3.841, 5.991, \text{ and } 7.815 \) for one-, two-, and three-parameter fits, respectively.

3. Combination of best-fits from distinct data sets

Suppose we have two data sets, say \(T_1\) and \(T_2\), 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 data sets 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 data sets 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 data sets, 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 Section 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 data sets. 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% CL intervals for the Wilson coefficients in single-parameter fits averaged over 1000 pseudo-experiments. The bounds on the Wilson coefficient $C_{eu}$ across numerous data sets 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. Fig. 12 displays the 95% CL 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, $A^{(e)}_{PV}$, of unpolarized hadrons ("unpolarized $A_{PV}$"), hadron PV asymmetries, $A^{(H)}_{PV}$, with unpolarized electrons ("polarized $A_{PV}$"), and unpolarized electron-positron asymmetries, $A^{(H)}_{LC}$, 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 indicate 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.

![Fig. 12. 95% CL bounds of $C_{eu}$ from single-parameters fits (darker) and from the (1+1)-parameter fits with beam polarization as an additional fitting parameter (lighter) using the families of data sets D4, D5, P4, and P5 at $\Lambda = 1$ TeV.](image)

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

- Proton asymmetries of all the three types, namely unpolarized and polarized PV asymmetries and LC asymmetries, impose considerably stronger bounds than deuteron.

- High-energy low-luminosity data sets D5 and P5 lead to slightly weaker bounds than the less-energetic but higher-luminosity ones, D4 and P4, respectively.

- Unpolarized PV asymmetries, $A^{(c)}_{PV}$, offer much stricter bounds than the polarized ones $A^{(H)}_{PV}$; however, it should be noted that for some Wilson coefficients, unpolarized proton asymmetries yield nearly the same bounds as the corresponding polarized ones.

- Data sets 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.

- Bounds from electron-positron asymmetries, $A^{(H)}_{LC}$, are comparable to or looser than the ones from polarized hadron asymmetries. They never offer stricter bounds than high-luminosity hadron PV asymmetries.

- 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 the Appendix A 2.

Assuming weak correlations, one can also combine the bounds within a given family of data sets, e.g. D4, ∆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 data set. 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 cut-off 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 data sets 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 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 $\sqrt{5}$, where the latter is to account for the constructive interference between quark contributions, and by another factor to convert 90% CL to 95% CL.

**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% CL 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, band-like 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.
FIG. 14. 95% CL ellipses for the Wilson coefficients $C_{eu}$ and $C_{qe}$ using the families of data sets D4 and P4 in the simultaneous $(2 + 1)$-parameter fits that includes the beam polarization as an additional fitting parameter.

Fig. 15 shows the comparison of the simultaneous fit of the Wilson coefficients $(C_{eu}, C_{ℓu})$ 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 $(C_{eu}, C_{ℓq}^{(1)})$, using the nominal- and high-luminosity P4 set of the EIC. We observe that in both 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 data set and the LHC adapted from [4] for the pair $(C_{ℓq}^{(1)}, C_{ℓq}^{(3)})$. This figure shows that when the LHC data imposes 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 indicates 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 data set 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 polarization
as an additional parameter can be summarized as follows:

- Proton asymmetries impose much stricter bounds than deuteron.
- Unpolarized hadron asymmetries lead to stronger correlations than the polarized ones.
- 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.
- The projected EIC data are capable of resolving all flat directions that appear in the LHC Drell-Yan data.
- 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 Section B3 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 formulae 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 semi-leptonic 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 cross-talk 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 ten-fold 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 data sets. 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 semi-leptonic 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 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.

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 data sets. 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, $A_{\text{lum}}$, added to the pseudodata. Then, we define the $\chi^2$ test statistics as:

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

(A1)

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

$$\tilde{\Sigma}^2 = \Sigma^2 |_{\sigma_{\text{lum}} \to 0} .$$  

(A2)

However, we keep the luminosity uncertainty in the pseudodata generation. By introducing the luminosity difference, $A_{\text{lum}}$, 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, $|\rho| \lesssim 0.4$, between $A_{\text{lum}}$ 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% CL intervals with and without $A_{\text{lum}}$ for the Wilson coefficient $C_{eu}$ in all the four LC asymmetry data sets of interest. In Fig. 19, we compare the 95% CL ellipses of the Wilson coefficients ($C_{eu}, C_{qe}$) between the data sets LD4 and LP5 with and without the luminosity difference as a fitted parameter. From these figures, we see that the 95% CL bounds on $C_{eu}$ become 15 to 20% weaker. The difference is less noticeable in the confidence ellipses.

![Comparison of the bounds on the Wilson coefficient $C_{eu}$ with all the LC asymmetry data sets of interest in the absence and presence of the luminosity difference as a new free fitting parameter.](image)

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

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 $\chi^2$ test statistics as:

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

(A3)

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

$$\tilde{\Sigma}^2 = \Sigma^2 |_{\sigma_{\text{pol}} \to 0} ,$$  

(A4)

but not during pseudodatageneration. The second term on the RHS of Eq. (A3) is added by hand, where $\bar{P}$ and $\delta P$ 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 \( P = 1 \) 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 \( \chi^2 \) distribution by 1. As in the case of luminosity difference fits, we extend our 1- and 2-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 data sets, 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 data sets P4 and \( \Delta P4 \), while Fig. 21 displays the 95% CL ellipse of the Wilson coefficients \( (C_{eu}, C_{qe}) \) for the same data sets in the nominal-luminosity scenario. We find that bounds from unpolarized hadron data sets 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.
FIG. 21. 95% CL ellipse of the Wilson coefficients $C_{eu}$ and $C_{qe}$ for the data sets P4 and ΔP4 in the absence and presence of the beam polarization, $P$, as a new parameter in the fits.

One can explain why the bounds become weaker in the polarized hadron sets by referring to the correlations. Since in these data sets, 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 almost fully independent so that they will not affect each other in the fits. Thus, any increase in the allowed limits of 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% CL intervals and the corresponding effective UV cut-off scales for all the seven Wilson coefficients in single-parameter fits averaged over 1000 pseudo-experiments. We recall the following abbreviations for the EIC preliminary data sets:

- electron PV asymmetries of unpolarized deuteron, $A^{(e)}_{PV}$:
  - D4: $eD$ 10 GeV $\times$ 137 GeV, 100 fb$^{-1}$
  - D5: $eD$ 18 GeV $\times$ 137 GeV, 15.4 fb$^{-1}$

- electron PV asymmetries of unpolarized proton, $A^{(e)}_{PV}$:
  - P4: $ep$ 10 GeV $\times$ 275 GeV, 100 fb$^{-1}$
  - P5: $ep$ 18 GeV $\times$ 275 GeV, 15.4 fb$^{-1}$

- hadron PV asymmetries with unpolarized electron, $A^{(H)}_{PV}$: $\Delta D4$, $\Delta D5$, $\Delta P4$, and $\Delta P5$ with the same energy and luminosity configuration as the corresponding D- and P-sets.

- unpolarized electron-positron asymmetries of unpolarized hadrons, $A^{(H)}_{LC}$: $LD4$, $LD5$, $LP4$, 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.

Figs. 22–28 display the 95% CL 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, $A^{(e)}_{PV}$ (“unpolarized $A_{PV}$”), hadron PV asymmetries with unpolarized electrons, $A^{(H)}_{PV}$ (“polarized $A_{PV}$”), and unpolarized electron-positron asymmetries of unpolarized hadrons, $A^{(H)}_{LC}$ (“lepton-charge $A$”). PV asymmetries are then grouped into two, showing the fits in the nominal- and high-luminosity scenarios. The nominal luminosity (“NL”) refers to the annual integrated luminosity of Table 10.1 of the YR [24]. The high luminosity (“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 indicate the bounds from 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\textsuperscript{2}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure22}
\caption{95\% CL bounds of $C_{eu}$ from 1-parameter fits (darker) and from simultaneous (1 + 1)-parameter fits with beam polarization (lighter) using the families of data sets D4, D5, P4, and P5.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure23}
\caption{The same as in Fig. 22 but for $C_{ed}$.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure24}
\caption{The same as in Fig. 22 but for $C_{\ell q}^{(1)}$.}
\end{figure}
FIG. 25. The same as in Fig. 22 but for $C_{ℓq}^{(3)}$.

FIG. 26. The same as in Fig. 22 but for $C_{ℓu}$.

FIG. 27. The same as in Fig. 22 but for $C_{ℓd}$. 
In Figs. 29–35, we present the effective UV cut-off 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 data sets 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.

FIG. 29. Effective UV cut-off scales, $\Lambda/\sqrt{C_{eu}}$, defined in terms of the 95% CL 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 $C_{cd}$.
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 $C_{\ell d}$. 
For completeness, we summarize in Table V the fitting results of all the seven Wilson coefficients with the four families of data sets of interest. The values indicated in this table are the 95% CL bounds around zero.

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

| $\epsilon D$ | $C_{eu}$ | $C_{ed}$ | $C_{q_1}^{(1)}$ | $C_{q_1}^{(3)}$ | $C_{eu}$ | $C_{ed}$ | $C_{q_1}$ |
|--------------|----------|----------|-----------------|-----------------|----------|----------|----------|
| unpolarized  | 0.21     | 0.38     | 0.54            | 0.14            | 0.47     | 0.75     | 1.06     |
| A$_{PV}$     | with P   | 0.15     | 0.27            | 0.39            | 0.10     | 0.37     | 0.60     | 0.87     |
|              | 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     |
| high luminosity | 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     |
| polarized A$_{PV}$ | 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     |

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 that we consider in this work. The ellipses are plotted at 95% CL 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. Data sets with the label NL or HL indicate the luminosity:
The nominal luminosity ("NL") refers to the annual integrated luminosity of Table 10.1 of YR [24]. The high luminosity ("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 data sets, namely D4, D5, P4, and P5. We show the fits from polarized and unpolarized PV asymmetry data sets in both nominal- and high-luminosity scenarios for comparison. We remark that the ellipses for the polarized and unpolarized PV asymmetry data sets 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 examples fitted results for the LHC Drell-Yan data, adapting from [6] and [45].

FIG. 36. 95% CL ellipses for the Wilson coefficients $C_{eu}$ and $C_{ed}$ using the families of data sets D4, D5, P4, and P5 at $\Lambda = 1$ TeV.
FIG. 37. The same as in Fig. 36 but for $C_{eu}$ and $C_{ℓq}^{(1)}$. 
FIG. 38. The same as in Fig. 36 but for $C_{eu}$ and $C_{ℓq}^{(3)}$. 

\[ C_{eu} \]

\[ C_{ℓq}^{(3)} \]
FIG. 39. The same as in Fig. 36 but for $C_{eu}$ and $C_{lu}$. 
FIG. 40. The same as in Fig. 36 but for $C_{eu}$ and $C_{ld}$. 
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 $C_{\ell q}^{(1)}$. 

[Graphs showing the relation between $C_{ed}$ and $C_{\ell q}^{(1)}$.]
FIG. 43. The same as in Fig. 36 but for $C_{ed}$ and $C_{e\ell q}^{(3)}$. 

95% CL, $\Lambda = 1$ TeV
FIG. 44. The same as in Fig. 36 but for $C_{ed}$ and $C_{eu}$. 
FIG. 45. The same as in Fig. 36 but for $C_{ed}$ and $C_{ld}$. 
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 $C_{ld}$. 
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 $C_{\ell u}$. 

$C_{\ell q}^{(3)}$ 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 $C_{tu}$ and $C_{td}$. 
FIG. 55. The same as in Fig. 36 but for $C_{\ell u}$ and $C_{qu}$. 
FIG. 56. The same as in Fig. 36 but for $C_{\ell d}$ and $C_{qe}$.

For completeness, we show the correlations of Wilson coefficients in Figs. 57–60.
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.
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 hyper-ellipse. 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 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 $C_{eu}$, $C_{ed}$, $C^{(1)}_{\ell q}$, $C^{(3)}_{\ell q}$, $C_{tu}$, and $C_{qe}$ and consider the data from $ep$ collisions in the configuration 10 GeV $\times$ 275 GeV with 100 fb$^{-1}$, which is the nominal-luminosity P4 data set. In Fig. 61 we compare the bounds from the original 1d fits to the projected bounds from the 6d fit of the Wilson coefficients $C_{eu}$, $C_{ed}$, $C^{(1)}_{\ell q}$, $C^{(3)}_{\ell q}$, $C_{tu}$, 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 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. 62. Comparison of the 95% CL ellipses for the Wilson-coefficient pairs \((C_{eu}, C_{ed})\), \((C_{eu}, C_{e\ell q}^{(1)})\), and \((C_{eu}, C_{e\ell q}^{(3)})\) between the original 2d fits and the projections from the simultaneous fit of \(C_{eu}, C_{ed}, C_{e\ell q}^{(1)}, C_{e\ell q}^{(3)}, C_{eu}, \text{and } C_{e\ell u}\) using the data set P4 at \(\Lambda = 1 \text{ TeV}\).

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

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

**Acknowledgment:** R. B. is supported by the US Department of Energy (DOE) contract DE-AC02-06CH11357. T. K. is supported by the DOE grant DE-SC0020240 and the Zuckerman STEM Leadership Program. F. P., K. Ş., and D. W. are supported by the DOE grants DE-FG02-91ER40684 and DE-AC02-06CH11357. X. Z. and M. N. are supported by the DOE grant DE-SC0014434. The authors would like to 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. would like to thank H. Spiesberger for suggestions on the use of the DJango and useful discussions related to the analysis.

[1] I. Brivio and M. Trott, *Phys. Rept.* **793**, 1 (2019), arXiv:1706.08945 [hep-ph].
[2] W. Buchmuller and D. Wyler, *Nucl. Phys. B* **268**, 621 (1986).
[3] C. Arzt, M. B. Einhorn, and J. Wudka, *Nucl. Phys. B* **433**, 41 (1995), arXiv:hep-ph/9405214.
[4] B. Grzadkowski, M. Iskrzynski, M. Misiak, and J. Rosiek, *J. High Energy Phys.* **10**, 085, arXiv:1008.4884 [hep-ph].
[5] S. Alte, M. König, and W. Shepherd, *J. High Energy Phys.* **07**, 144, arXiv:1812.07575 [hep-ph].
[6] R. Boughezal, F. Petriello, and D. Wiegand, *Phys. Rev. D* **101**, 116002 (2020).
[7] G. Panico, L. Ricci, and A. Walzer, *J. High Energy Phys.* **07**, 086, arXiv:2103.10532 [hep-ph].
[8] E. D. P. A. Panel, Report from the EIC Detector Proposal Advisory Panel (2022).
[9] S. Aioli, R. Boughezal, E. Mereghetti, and F. Petriello, *Phys. Lett. B* **809**, 135703 (2020), arXiv:2003.11615 [hep-ph].
[10] R. Boughezal, E. Mereghetti, and F. Petriello, *Phys. Rev. D* **104**, 095022 (2021), arXiv:2106.05337 [hep-ph].
