Topological heavy-flavor tagging and intrinsic bottom at the Electron-Ion Collider

Thomas Boettcher*

Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA

(Received 8 March 2024; accepted 1 May 2024; published 23 May 2024)

Heavy-flavor hadron production, in particular bottom hadron production, is difficult to study in deep-inelastic scattering (DIS) experiments due to small production rates and branching fractions. To overcome these limitations, a method for identifying heavy-flavor DIS events based on event topology is proposed. Based on a heavy-flavor jet tagging strategy developed for the LHCb experiment, this algorithm uses displaced vertices to identify decays of heavy-flavor hadrons. The algorithm's performance at the Electron-Ion Collider is demonstrated using simulation, and it is shown to provide discovery potential for nonperturbative intrinsic bottom quarks in the proton.

DOI: 10.1103/PhysRevD.109.092010

I. INTRODUCTION

The possible existence of nonperturbative "intrinsic" heavy quarks in the proton was first proposed shortly after the discovery of heavy quarks themselves [1]. Intrinsic heavy quarks are predicted to arise from a $|uudQ\bar{Q}\rangle$ component of the proton's wave function, where $Q\bar{Q}$ denotes a heavy quark-antiquark pair. Various models predict the intrinsic contribution to the heavy-quark parton distribution functions (PDFs), including models inspired by light-front quantum chromodynamics (LFQCD) [1] and fluctuations of the proton into heavy meson-baryon pairs [2]. These models generally agree that intrinsic heavy quarks carry a large fraction x of the proton's momentum, resulting in valencelike heavy quarks. This can be seen in Fig. 1, which shows LFQCD-inspired models of intrinsic charm (IC) and intrinsic bottom (IB) [3].

Experimental searches for IC have been carried out in both fixed-target deep-inelastic scattering (DIS) and high-energy hadron collisions. Charm structure function data from the European Muon Collaboration (EMC) experiment [5] and studies of Z-boson production in association with charm-quark jets ($Z + c$) by the LHCb experiment [6,7] are expected to be particularly sensitive probes of IC. The LHCb experiment has also searched for evidence of IC in charm production and charge asymmetry measurements in fixed-target proton-nucleus collisions [8,9]. Intrinsic heavy flavor is typically characterized by the average momentum carried by the intrinsic heavy quarks, $\langle x\rangle_{\mathrm{IC,IB}}$, at an initial energy scale $Q_0 = m_c$. The NNPDF collaboration performed a global analysis including EMC and LHCb $Z + c$ data [10]. The analysis claimed 3σ evidence for nonzero IC with $\langle x\rangle_{\mathrm{IC}} \approx 1\%$. A global analysis based on the CT18 PDF fit omitted the LHCb and EMC measurements due to difficulties with theoretical interpretation. The resulting fit mildly prefers nonzero IC, with $\langle x\rangle_{\mathrm{IC}} \approx 0.5\%$ [3]. Yet another global analysis excluded percent-level IC at the 4σ level [2]. The Electron-Ion Collider (EIC), under construction at Brookhaven National Laboratory, is expected to produce in excess of 100 times more data than previous collider DIS facilities, allowing for detailed studies of the charm quark PDF [11]. Recent studies indicate that the EIC will be able to conclusively observe or exclude percent-level IC in the proton [12,13].

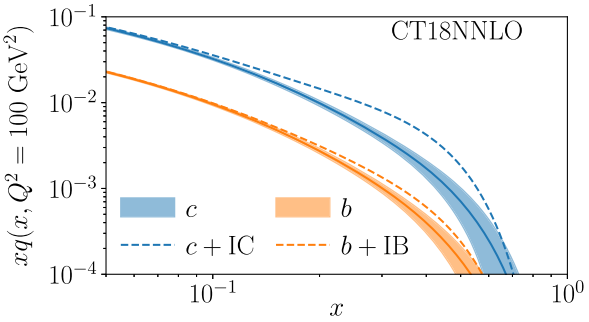

FIG. 1. Intrinsic charm and bottom PDFs. The baseline PDFs are from the CT18NNLO PDF set [4]. The shaded regions show the 68% confidence-level regions. The $c + \mathrm{IC}$ PDF is from CT18FC [3]. The $b + \mathrm{IB}$ PDF is obtained by scaling the intrinsic component of the CT18FC charm PDF by m_c^2/m_b^2 and adding the result to the baseline b PDF.

*boettcts@ucmail.uc.edu

In contrast to the experimental and theoretical interest in IC, the possibility of intrinsic bottom quarks in the proton has received relatively little attention (see Ref. [14] for a review). The size of the intrinsic heavy-quark contribution to the proton PDF is expected to scale as $1/m_Q^2$, where m_Q is the heavy quark mass, suppressing IB by an order of magnitude relative to IC [15]. As a result, both the absolute size of the IB contribution and its size relative to the perturbative b-quark PDF are smaller than the analogous IC contributions. The b-hadron cross section in DIS is also suppressed relative to the c-hadron cross section due to the smaller electric charge of the b quark. Additionally, the largest b-hadron branching fractions to fully reconstructible final states are $\mathcal{O}(10^{-3})$ [16]. As a result of these limitations, little data constraining the b-quark PDF exists. What little data does exist does not probe the valence region, leaving the IB content of the proton almost entirely unconstrained [17]. Consequently, no global analysis of IB in the proton has been performed.

The experimental challenges of studying b-hadron production in DIS can be partially overcome by using the topology of heavy-flavor hadron decays. This strategy was used by both the H1 and ZEUS experiments at HERA, which used displaced tracks and secondary vertices to identify b-hadrons and extract the $b\bar{b}$ contribution to the proton structure function, $F_2^{b\bar{b}}$ [17–19]. The LHC experiments use a similar strategy to identify heavy-flavor jets. Jets containing heavy-flavor hadrons are identified using the properties of displaced charged-particle vertices [20–23]. Using this strategy, the LHCb experiment is able to identify or "tag" about 60% of jets containing b-hadrons and distinguish between b and c jets. The proposed detector at the EIC is expected to have vertex reconstruction capabilities similar to those of LHCb, enabling a similar strategy for tagging heavy-flavor DIS events [11]. Previous studies have explored the performance of topological charm tagging at the EIC, but studies involving b hadrons have focused on using fully reconstructed decays to study hadronization [24]. Previously explored charm tagging methods rely on the ability to identify charged kaons or count displaced tracks, either in the entire event or clustered into jets [25–27]. In contrast, the algorithm employed by LHCb does not require particle identification and relies only on the topological properties of charged particle vertices.

This paper demonstrates how the LHCb experiment's jet tagging strategy can be applied to study heavy-flavor production at the EIC. Because the LHCb jet-tagging algorithm depends only on the properties of the heavy flavor decay and not on the jet itself, the algorithm can be naturally adapted to identifying heavy-flavor events in DIS. Section II describes the simulation setup used for these studies, and Sec. III describes the heavy-flavor tagging algorithm. Section IV presents the expected sensitivity to IB, and Sec. V summarizes conclusions and discusses additional uses for topological heavy-flavor tagging at the EIC.

II. SIMULATION

The tagging algorithm performance studies were conducted using simulated $e+p$ DIS events generated using the PYTHIA 8.3 generator [28]. The simulation includes both neutral- and charged-current DIS, although the charged-current contribution to the simulated samples is negligible. Simulations were performed for four beam energy configurations: 5×100, 10×100, 10×275, and 18×275 GeV [29], where the first number of each pair is the electron energy and the second is the proton energy. These configurations correspond to $\sqrt{s}=45$, 63, 105, and 141 GeV, respectively. The effect of the EIC's nonzero beam crossing angle is not considered. Heavy-flavor events are defined by the presence of a heavy-flavor hadron. A b event contains a b hadron, whereas a c event contains a c hadron and no b hadron. A light-parton (uds) event contains no c or b hadrons.

The tagging algorithm's performance was studied as a function of the kinematic variables x and Q^2. These variables can be used to calculate the inelasticity $y=Q^2/(xs)$. Topological heavy-flavor tagging methods require a heavy-flavor hadrons to have sufficient momenta to fly detectable distances before decaying. As a result, b-tagging performance will be poor for $Q\lesssim 2m_b$. Consequently, only $Q^2>100$ GeV^2 is considered for this study. For the beam configurations and Q^2 requirement used in this study, the accessible high-x kinematic region relevant for IB corresponds to $0.01\lesssim y\lesssim 0.5$. In this kinematic region, EIC measurements are expected to incur percent-level systematic uncertainties from the finite detector resolution in x and Q^2, as well as the corresponding corrections [30]. The resulting uncertainties are expected to be much smaller than the statistical uncertainties of a b-production measurement. As a result, x and Q^2 were determined at parton level for this study. Furthermore, radiative corrections are expected to be less significant for heavy-quark production than for inclusive DIS and were ignored in this study [27].

The response of a hypothetical EIC detector is modeled according to parametrizations based on the expected performance of the future detector [11]. The momentum and position resolutions are given as functions of transverse momentum (p_T) and pseudorapidity (η), as shown in Table I. Only long-lived charged particles with $p_T>200$ MeV and $|\eta|<2.5$ were considered for this study. A charged particle reconstruction efficiency of 90% was assumed for the entire fiducial region.

The position of the collision vertex, or primary vertex (PV), is determined by smearing the true position of the interaction point. The PV resolution is shown in Ref. [13] and estimated here as $\sigma_{x,y,z}=(10\oplus 30/\sqrt{n})$, where n is

TABLE I. Resolution functions used to smear the generated charged particles to simulate the EIC detector's response. Both p and p_T are in GeV.

	σ_p/p	σ_{xy}	σ_z
$\lvert\eta\rvert < 1$	$0.04p \oplus 1\%$	$30/p_T \oplus 5$ μm	$30/p_T \oplus 5$ μm
$\lvert\eta\rvert < 2.5$	$0.04p \oplus 2\%$	$40/p_T \oplus 10$ μm	$100/p_T \oplus 20$ μm

the number of reconstructed prompt charged particles. Reconstructed particles are classified as prompt based on

$$\chi^2_{\mathrm{DCA,IP}} = \frac{d_x^2}{\sigma_x^2} + \frac{d_y^2}{\sigma_y^2}, \tag{1}$$

where $d_{x,y}$ is the distance of closest approach of the smeared charged particle to the interaction point in the dimension denoted by the subscript, and $\sigma_{x,y}$ is the corresponding detector resolution determined using the parametrizations from Table I. Only the x and y displacements are used in Eq. (1) in order to minimize dependence on the track pointing resolution in the z direction, which rapidly deteriorates at large $|\eta|$ [11]. Furthermore, the quantity defined in Eq. (1) is similar to the impact parameter significance used by the LHCb collaboration to characterize track displacement [20]. Tracks are considered prompt if $\chi^2_{\mathrm{DCA,IP}} < 12$.

III. TAGGING ALGORITHM

The tagging algorithm used for this study is based on the algorithm described in Ref. [20]. The LHCb algorithm constructs secondary vertices (SVs) within jets and uses two boosted decision tree (BDT) classifiers to identify vertices from light-, c-, and b-hadron decays. One BDT is trained to distinguish heavy-flavor SVs from light-hadron SVs, and the other is trained to distinguish between b and c SVs. For this study, the LHCb SV reconstruction algorithm was adapted to the simulated EIC data. Heavy-vs-light and b-vs-c BDTs were trained for a hypothetical EIC detector using variables similar to those used to train the LHCb BDTs.

First, displaced pseudoreconstructed charged particles are combined to form two-track SVs. Charged particle displacement is characterized by χ^2_{DCA}, which is defined as in Eq. (1) but with distances calculated with respect to the smeared PV position instead of the true interaction point. Charged particles are considered displaced if $\chi^2_{\mathrm{DCA}} > 16$. Pairs of displaced charged particles with a distance of closest approach to one another less than 0.2 mm are combined to form two-track SVs. Next, pairs of two-track SVs that share a track are combined to form three-track SVs. Only SVs with $0.4 < m < 5.3$ GeV are considered for merging, where m is the SV mass calculated assuming the charged pion mass for each of the constituent tracks. This merging process is repeated until no SVs share tracks. The resulting SVs can consist of any number of tracks.

To suppress contributions from strange particle decays, two-track SVs are required to have $m > 0.6$ GeV. This requirement removes both $K^0_s \to \pi^+\pi^-$ and $\Lambda \to p\pi^-$ decays. Events containing at least one SV passing these requirements are considered tagged. If an event contains multiple SVs, the SV with the largest p_T is used for further classification.

Tagged events are classified using a pair of BDT classifiers. The first BDT is trained to distinguish heavy-flavor events from uds events ($\mathrm{BDT}_{bc|uds}$), and the second is trained to distinguish between b and c events ($\mathrm{BDT}_{b|c}$). The BDTs use four variables characterizing the SV tag. These include the mass of the SV (m), the number of tracks used to construct the SV (n_{trks}), and the sum of the χ^2_{DCA} of the constituent tracks. In addition, the BDTs use the corrected mass of the SV, which is given by

$$m_{\mathrm{cor}} = \sqrt{m^2 + p_\perp^2} + p_\perp, \tag{2}$$

where $p_\perp$ is the component of the SV momentum perpendicular to its flight direction [31,32]. These variables are chosen because they depend only on the topological properties of the SV and do not depend on the full SV covariance matrix, which is difficult to estimate without a realistic detector simulation and reconstruction algorithms.

The distributions of the BDT input variables are shown in Fig. 2 for the $\sqrt{s} = 63$ GeV beam configuration. Bottom

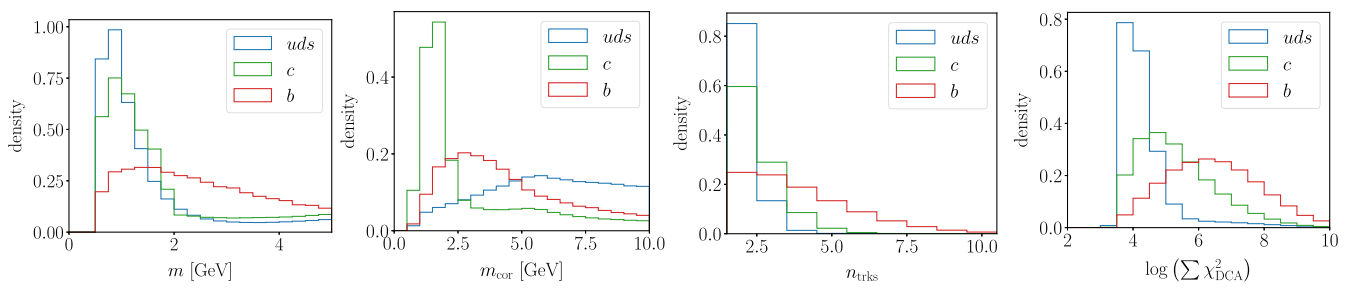

FIG. 2. Distributions of variables used for BDT training from the $\sqrt{s} = 63$ GeV simulated data sample. Each distribution is normalized to unit area.

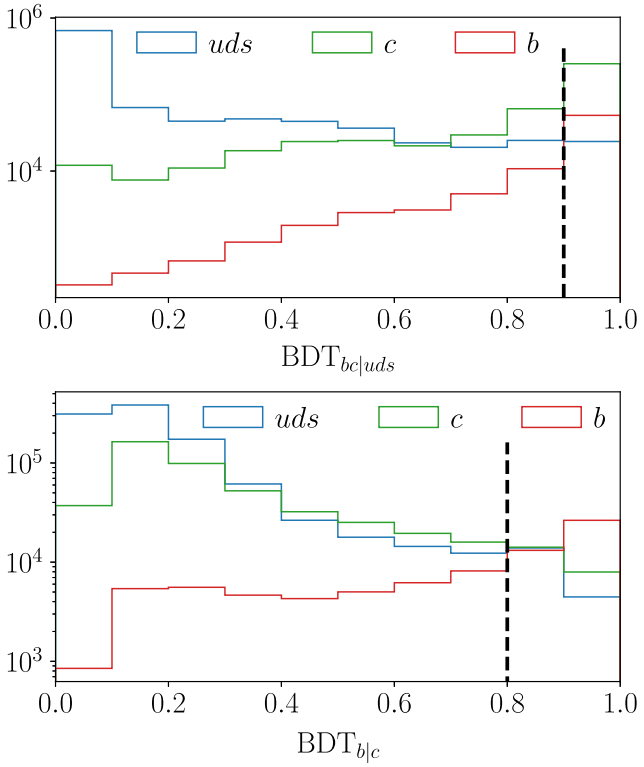

FIG. 3. Response distributions for (top) BDT$_{bc|uds}$ and (bottom) BDT$_{b|c}$ from $\sqrt{s} = 63$ GeV simulation with expected relative normalizations. The dashed lines show the low edges of the signal regions. The $y = 0.1$ and 0.9 contours are shown as dashed lines.

hadrons are more massive and produce more final-state particles than c hadrons, which results in the observed hierarchies in m and n_{trks}. They also have a longer lifetime than c and light hadrons and consequently have larger $\sum\chi^2_{\text{DCA}}$. The corrected mass is particularly powerful for identifying c events because c hadrons typically decay at a single vertex. These decays produce a m_{cor} peak near the mass of the D meson. Bottom hadrons produce more complex decay topologies and a consequently broader m_{cor} distribution than that of charm hadrons. SVs in uds events are made up of combinations of poorly reconstructed prompt tracks. The momenta of these combinations can point far from the PV and produce large corrected masses.

The BDT response distributions are shown in Fig. 3. In an analysis of real data, the composition of the tagged sample could be determined using a two-dimensional template fit to these distributions [33,34]. For this study, the region BDT$_{bc|uds} > 0.9$ and BDT$_{b|c} > 0.8$ was defined as the signal region (SR) for the purpose of estimating statistical uncertainties. The signal region tagging efficiency ϵ_{SR}, defined as the probability that an event is tagged and the SV falls in the BDT SR, is shown in Fig. 4 for the $\sqrt{s} = 63$ GeV configuration. The tagging efficiency ranges from 30%–40% in most kinematically allowed bins and approaches 60% at high Q^2. This efficiency is

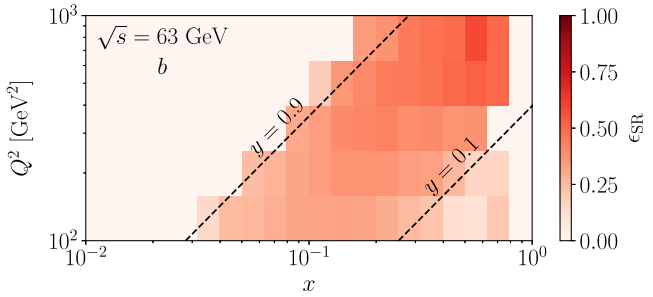

FIG. 4. The signal region tagging efficiency determined from $\sqrt{s} = 63$ GeV simulation. Kinematically forbidden regions are given an efficiency of zero.

consistent with the b-jet tagging efficiency observed by LHCb, which approached 60% at high jet p_{T}. Charm events have a signal-region mistag probability of around 1%, while uds events have a mistag probability of around 10^{-4}. While uds events are the largest background overall, their contribution to the signal region is small. The fast simulation used in this study does not include non-Gaussian misreconstruction effects or secondary particle production from material interactions. Both of these effects will create additional SVs in uds events, but these SVs should still be distinguishable from heavy flavor decays and are expected to make a small contribution to the signal region [20].

The BDT responses can also be used to identify c events. Because c-hadron cross sections are much larger than those of b hadrons, c events are the dominant contribution in the region BDT$_{bc|uds} > 0.9$. For a c event SR defined as BDT$_{bc|uds} > 0.9$ and BDT$_{b|c} < 0.8$, the c-tagging efficiency is around 20% over most of the accessible kinematic region. This performance is similar to those of other topological c-tagging methods developed for the EIC [25].

IV. INTRINSIC BOTTOM

The $b\bar{b}$ production cross section in unpolarized neutral-current DIS in the kinematic region studied here is given by

$$\frac{\mathrm{d}\sigma^{b\bar{b}}}{\mathrm{d}x\mathrm{d}Q^2} = \frac{2\pi\alpha^2 Y_+}{xQ^4}\left(F_2^{b\bar{b}}(x,Q^2) - \frac{y^2}{Y_+}F_L^{b\bar{b}}(x,Q^2)\right), \quad (3)$$

where α is the fine structure constant, $Y_+ = 1 + (1-y)^2$, and $F_2^{b\bar{b}}$ and $F_L^{b\bar{b}}$ are the b-quark contributions to the proton structure functions [17]. DIS experiments typically report a reduced cross section given by

$$\sigma_r^{b\bar{b}}(x,Q^2) = F_2^{b\bar{b}}(x,Q^2) - \frac{y^2}{Y_+}F_L^{b\bar{b}}(x,Q^2). \quad (4)$$

For the relatively small values of y considered in this study, $\sigma_r^{b\bar{b}}$ is determined primarily by $F_2^{b\bar{b}}$. At leading order (LO) in the strong coupling constant α_{s}, $F_2^{b\bar{b}}$ is proportional to the sum of b and $\bar{b}$ PDFs.

The estimate of the IB contribution to $\sigma_r^{b\bar{b}}$ is based on two observations. First, the intrinsic heavy quark PDFs evolve approximately independently from the other PDFs [15]. This means that the IC contribution to the charm PDF can be approximated as $c_{0+\mathrm{IC}}(x,Q^2) - c_0(x,Q^2)$, where c_0 is the charm PDF from a fit without IC and $c_{0+\mathrm{IC}}$ is from a fit that includes IC. The intrinsic b PDF can then be estimated as

$$b_{\mathrm{IB}}(x,Q^2) = \frac{m_c^2}{m_b^2}(c_{0+\mathrm{IC}}(x,Q^2) - c_0(x,Q^2)). \quad (5)$$

Second, the dominant contribution from intrinsic heavy quarks to the reduced cross section is from the LO contribution to $F_2^{q\bar{q}}$. As a result,

$$\sigma_{r,\mathrm{IB}}^{b\bar{b}}(x,Q^2) \approx \sigma_{r\text{no-IB}}^{b\bar{b}}(x,Q^2) + 2e_b^2 x b_{\mathrm{IB}}(x,Q^2), \quad (6)$$

where $\sigma_{r\text{no-IB}}^{b\bar{b}}$ is the reduced cross section assuming no IB, and e_b is the electric charge of the b quark. The factor of two in front of the IB term accounts for the $\bar{b}$ contribution, assuming the b and $\bar{b}$ PDFs are symmetric. Applying this strategy to calculate $\sigma_{r\mathrm{IC}}^{c\bar{c}}$ reproduces the full next-to-next-to-leading order (NNLO) result to within about 10% in the kinematic region covered by this study, which is sufficiently accurate for the sensitivity estimates performed here. Consequently, the IB contribution can be estimated using only c_0, $c_{0+\mathrm{IC}}$, and $\sigma_{r\text{no-IB}}^{b\bar{b}}$.

The no-IB cross sections for b, c, and uds events were calculated at NNLO in α_s using the YADISM package [35]. The calculations were performed using the zero-mass variable flavor number scheme (ZM-VFNS) and the CT18NNLO PDF set, which was accessed using LHAPDF [4,36]. The no-IC charm PDF c_0 is taken from CT18NNLO, and $c_{0+\mathrm{IC}}$ was taken from CT18FC [3]. CT18FC includes IC using the LFQCD-inspired model of Ref. [1] with $\langle x\rangle_{\mathrm{IC}} \approx 0.5\%$. It should be noted that the IC PDF from CT18FC is smaller than that from other global analyses, and IC normalizations almost three times larger than that used for this study are allowed within the CT18FC 68% confidence interval. Furthermore, the m_c^2/m_b^2 IB scaling is unconfirmed. IB with an order-of-magnitude larger overall normalization has not been excluded by data [15]. In this sense, the IB model used in this study is conservative.

The cross sections are used to calculate expected yields from one year of data taking in each beam configuration according to the integrated luminosities given in Refs. [30,37], which are reproduced in Table II. Signal-region tagging efficiencies for b, c, and uds events were calculated for each beam configuration as described in Sec. III and were used to calculate expected tagged yields. The tagged yields were then used to determine signal significance and expected statistical uncertainties. The IC contribution to the $c\bar{c}$ cross section is included as background in the IB predictions.

TABLE II. Expected annual integrated luminosities for various EIC beam configurations.

$\sqrt{s}$ [GeV]	$\mathcal{L}_{\mathrm{int}}$/year [fb^{-1}]
45	61.0
63	79.0
105	100.0
141	15.4

The $\sigma_{r,\mathrm{IB}}^{b\bar{b}}$ results are shown in Fig. 5. To better illustrate the estimated sensitivity to IB, the ratio of the IB results to the baseline are shown in Fig. 6. IB produces an enhancement of up to a factor of 3 in the valence region. The enhancement is most pronounced at low Q^2, where the contribution from perturbative b is smallest. In most of the kinematic bins where IB has a significant effect, the b PDF uncertainties are much larger than the expected statistical uncertainties. Because the no-IB PDF is determined entirely from gluon splitting via DGLAP evolution, these PDF uncertainties reflect uncertainties in the gluon PDF at high x. Consequently, the EIC's sensitivity to IB will depend in part on future constraints on the high-x gluon PDF from both the EIC and the LHC.

Using LHCb's experience as a guide, the largest systematic uncertainties for measurements using this tagging algorithm are likely to arise from the tagging efficiency determination and the BDT template calibration. The LHCb experiment was able to measure its jet tagging efficiency in data to within about 10% and calibrate its templates using dijet calibration samples [20,21]. A similar calibration is possible for $c\bar{c}$ events at the EIC using SV-tagged events containing a fully reconstructed $D^0 \to K^-\pi^+$ decay with a large separation in azimuthal angle ϕ from the tagging SV. The b tagging performance can be studied using events containing two b-like tags with large separations in ϕ. Ultimately, because the same templates are used for efficiency determinations and signal yield extraction, these uncertainties partially cancel in actual measurements. Furthermore, the remaining uncertainty will likely be highly correlated across kinematic bins and should only mildly affect sensitivity to IB. Measurements of heavy flavor production by the H1 and ZEUS collaborations using topological tagging also found that the dominant systematic uncertainties were highly correlated between data points [18,19].

The search for IB will also be complicated by the handling of the b-quark mass in the b PDF evolution. The b-quark pole mass is typically used as a starting scale for generating b quarks perturbatively and is anticorrelated with the b PDF. Variations of m_b within its uncertainties can produce changes in the b PDF comparable to the PDF fit uncertainties in the valence region [38,39]. Varying m_b

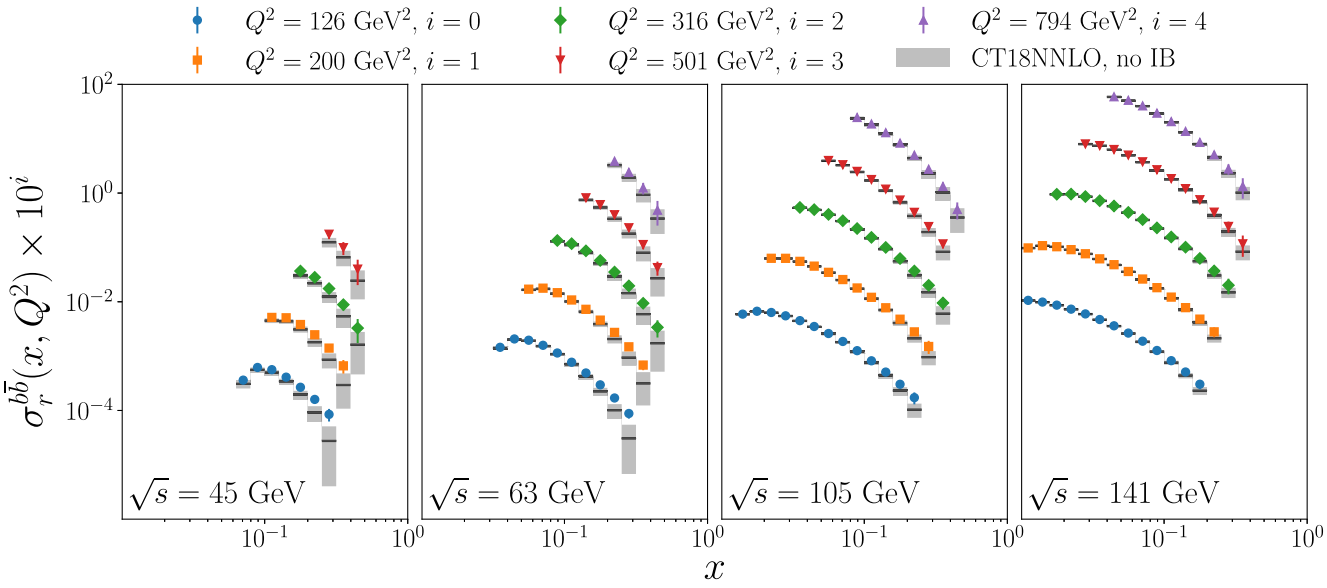

FIG. 5. Reduced cross section predictions. The points show predictions with IB, and the error bars show expected statistical uncertainties. The black lines show predictions without IB, and the shaded boxes show the 68% confidence-level PDF uncertainties. For visual clarity, the result in each Q^2 bin is offset by a factor 10^i, shown in the legend.

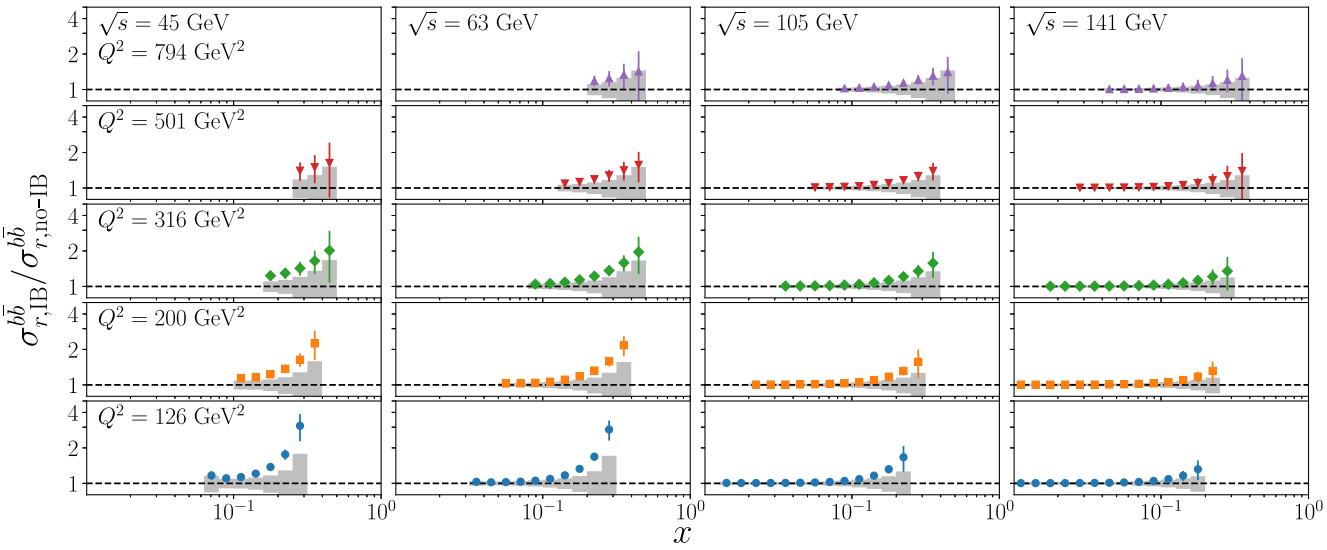

FIG. 6. Ratio of the $b\bar{b}$ reduced cross section with IB to the no IB case. The shaded regions show the 68% confidence interval PDF uncertainties for the no-IB case.

has a much larger effect at low x, however, and data over the broad x range studied here would provide strong constraints on both IB and m_b simultaneously. This data would also provide powerful tests of the heavy-quark scheme used in structure function calculations [40], as well as the kinematic constraints on intrinsic heavy flavor production proposed in Ref. [41]. The $b\bar{b}$ reduced cross section would be particularly sensitive to m_b and the choice of heavy-quark scheme at $Q^2 \gtrsim m_b$. Consequently, these studies would benefit from further development of low-Q^2 b-tagging methods.

V. CONCLUSIONS

Topological b tagging has proven to be a powerful tool for studying QCD in high-energy hadron collisions, and this work demonstrates that these methods are directly applicable to the EIC. The tagging strategy described here has a wide range of potential applications in electron-proton and electron-nucleus scattering. This algorithm could be used to tag heavy-flavor jets, which can be used to study both the structure of nuclei and the hadronization process [25,42]. It could also be used to study heavy

dihadron angular correlations, which provide sensitivity to gluon transverse momentum distributions [26,43]. This tagging strategy can also be used to efficiently tag charm events, potentially expanding the kinematic reach of charm production measurements at the EIC.

This paper also presents the first study of the EIC's ability to probe the b-quark PDF and its sensitivity to intrinsic bottom quarks. The EIC has the potential to observe intrinsic bottom at levels expected from recent global analyses of intrinsic charm in the proton. The observation of intrinsic bottom quarks is crucial for understanding the origin of intrinsic heavy quarks, including possible nonperturbative processes that produce heavy quarks in protons and nuclei. This paper presents the first strategy for observing intrinsic bottom in the near future.

ACKNOWLEDGMENTS

The author thanks Matthew Durham, Philip Ilten, Aleksander Kusina, Paul Newman, Brian Page, Michael Sokoloff, and Michael Williams for helpful discussions and feedback. This material is based upon work supported by the National Science Foundation under Grant No. PHY-2208983. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

[1] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai, Phys. Lett. **93B**, 451 (1980).
[2] T. J. Hobbs, J. T. Londergan, and W. Melnitchouk, Phys. Rev. D **89**, 074008 (2014).
[3] M. Guzzi, T. J. Hobbs, K. Xie, J. Huston, P. Nadolsky, and C. P. Yuan, Phys. Lett. B **843**, 137975 (2023).
[4] T.-J. Hou *et al.*, Phys. Rev. D **103**, 014013 (2021).
[5] J. J. Aubert *et al.* (European Muon Collaboration), Nucl. Phys. **B213**, 31 (1983).
[6] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **128**, 082001 (2022).
[7] T. Boettcher, P. Ilten, and M. Williams, Phys. Rev. D **93**, 074008 (2016).
[8] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **122**, 132002 (2019).
[9] R. Aaij *et al.* (LHCb Collaboration), Eur. Phys. J. C **83**, 541 (2023).
[10] R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, K. Kudashkin, G. Magni, and J. Rojo (NNPDF Collaboration), Nature (London) **608**, 483 (2022).
[11] R. Abdul Khalek *et al.*, Nucl. Phys. **A1026**, 122447 (2022).
[12] R. D. Ball, A. Candido, J. Cruz-Martinez, S. Forte, T. Giani, F. Hekhorn, G. Magni, E. R. Nocera, J. Rojo, and R. Stegeman (NNPDF Collaboration), arXiv:2311.00743.
[13] M. Kelsey, R. Cruz-Torres, X. Dong, Y. Ji, S. Radhakrishnan, and E. Sichtermann, Phys. Rev. D **104**, 054002 (2021).
[14] S. J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger, and R. Vogt, Adv. High Energy Phys. **2015**, 231547 (2015).
[15] F. Lyonnet, A. Kusina, T. Ježo, K. Kovarík, F. Olness, I. Schienbein, and J.-Y. Yu, J. High Energy Phys. 07 (2015) 141.
[16] R. L. Workman *et al.* (Particle Data Group), Prog. Theor. Exp. Phys. **2022**, 083C01 (2022).
[17] H. Abramowicz *et al.* (H1 and ZEUS Collaborations), Eur. Phys. J. C **78**, 473 (2018).
[18] F. D. Aaron *et al.* (H1 Collaboration), Eur. Phys. J. C **65**, 89 (2010).
[19] H. Abramowicz *et al.* (ZEUS Collaboration), J. High Energy Phys. 09 (2014) 127.
[20] R. Aaij *et al.* (LHCb Collaboration), J. Instrum. **10**, P06013 (2015).
[21] R. Aaij *et al.* (LHCb Collaboration), J. Instrum. **17**, P02028 (2022).
[22] G. Aad *et al.* (ATLAS Collaboration), Eur. Phys. J. C **83**, 681 (2023).
[23] A. M. Sirunyan *et al.* (CMS Collaboration), J. Instrum. **13**, P05011 (2018).
[24] C.-P. Wong, X. Li, M. Brooks, M. J. Durham, M. X. Liu, A. Morreale, C. da Silva, and W. E. Sondheim, arXiv:2009.02888.
[25] M. Arratia, Y. Furletova, T. J. Hobbs, F. Olness, and S. J. Sekula, Phys. Rev. D **103**, 074023 (2021).
[26] X. Dong, Y. Ji, M. Kelsey, S. Radhakrishnan, E. Sichtermann, and Y. Zhao, Phys. Rev. D **107**, 074022 (2023).
[27] E. C. Aschenauer, S. Fazio, M. A. C. Lamont, H. Paukkunen, and P. Zurita, Phys. Rev. D **96**, 114005 (2017).
[28] C. Bierlich *et al.*, SciPost Phys. Codebases **2022**, 8 (2022).
[29] Natural units are used throughout this paper.
[30] N. Armesto, T. Cridge, F. Giuli, L. Harland-Lang, P. Newman, B. Schmookler, R. Thorne, and K. Wichmann, Phys. Rev. D **109**, 054019 (2024).
[31] M. Williams, V. Gligorov, C. Thomas, H. Dijkstra, J. Nardulli, and P. Spradlin, The HLT2 topological lines, Report No. LHCb-PUB-2011-002, CERN-LHCb-PUB-2011-002, CERN, 2011, https://cds.cern.ch/record/1323557.
[32] K. Abe *et al.* (SLD Collaboration), Phys. Rev. Lett. **80**, 660 (1998).
[33] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. D **92**, 052001 (2015).
[34] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **115**, 112001 (2015).
[35] A. Candido, F. Hekhorn, G. Magni, T. R. Rabemananjara, and R. Stegeman, arXiv:2401.15187.
[36] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, Eur. Phys. J. C **75**, 132 (2015).

[37] S. Cerci, Z. S. Demiroglu, A. Deshpande, P. R. Newman, B. Schmookler, D. Sunar Cerci, and K. Wichmann, Eur. Phys. J. C **83**, 1011 (2023).
[38] T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 744 (2021).
[39] J. Campbell, T. Neumann, and Z. Sullivan, Phys. Rev. D **104**, 094042 (2021).
[40] A. Accardi *et al.*, Eur. Phys. J. C **76**, 471 (2016).
[41] J. Blümlein, Phys. Lett. B **753**, 619 (2016).
[42] H. T. Li, Z. L. Liu, and I. Vitev, Phys. Lett. B **827**, 137007 (2022).
[43] R. F. del Castillo, M. G. Echevarria, Y. Makris, and I. Scimemi, J. High Energy Phys. 03 (2022) 047.