Higgs Production and $H^0 \to c \bar{c}$ at a Future Muon-Ion Collider

Pavan Ahluwalia,1,2 Justine Choi,1,3 Simba Masters,1 Austin Mullins,1 Noah Randall,1 and Stephen J. Sekula1,4,5,*

1Department of Physics, Southern Methodist University, Dallas, TX 75275, USA
2Harmony School of Innovation, Fort Worth, TX 76123, USA
3Highland Park High School, Highland Park, TX 75205, USA
4SNOLAB, Lively, ON P3Y 1N2, Canada
5Department of Physics, Engineering Physics, and Astronomy, Queen’s University, Kingston, ON K7L 3N6, Canada

We investigate the production of Higgs bosons in neutral- and charged-current interactions at a future Muon-Ion Collider (MuIC) with emphasis on $H^0 \to c \bar{c}$ decay. In particular, we focus on the leading production mechanisms and explore the impact of a range of Parton Distribution Function (PDF) scenarios, with emphasis on the role of heavy intrinsic quark flavor. We emphasize estimates of non-resonant di-charm background production and explore the impact of this background on a possible future measurement of $H^0 \to c \bar{c}$ at a future Muon-Ion Collider facility.

I. INTRODUCTION

The Electron-Ion Collider at Brookhaven National Laboratory [1, 2] will open an entirely new window on hadron structure through the use of high-intensity, high-energy electron and ion collisions. A recent proposal for later upgrading the EIC to a Muon-Ion Collider (MuIC) has been presented [3] and a case is being built within the community for such an upgrade path beyond the nominal EIC program [4]. A MuIC would not only provide a useful step forward toward the eventual idea of a dedicated muon collider, but itself would facilitate a novel physics program that complements the High-Luminosity Large Hadron Collider (HL-LHC) and other future accelerator complexes (c.f. Ref. [5]). In particular, the center-of-mass energy possible in such a machine would facilitate a program that includes Higgs physics.

In this note, we explore in additional detail the production of Higgs particles at such a collider. We focus specifically on $H^0 \to c \bar{c}$ and di-charm production that would serve as a background to the Higgs reconstruction. The decay $H^0 \to c \bar{c}$ is a key part of the ongoing LHC program (c.f. Refs [6, 7]) but poses a challenge in the LHC proton-collider environment. Mechanisms for production of the Higgs boson at a lepton-hadron collider are well-understood in the context of past such collider studies. A MuIC operated at, for example, a reasonable center-of-mass energy ($E_{\text{cm}} = 960$ GeV, $E_{\mu} = 275$ GeV, $\sqrt{s} \approx 1000$ GeV, achievable without a complete overhaul of the EIC complex) would allow for single-Higgs production dominated by charged-current (CC) and neutral-current (NC) deep-inelastic scattering (DIS) involving light-flavor quarks. This is illustrated in Fig. 1.

The cross-sections for the above processes were estimated in Ref. [4] for a $960$ GeV $\times 275$ GeV muon-proton collider. Estimates were made at leading order (LO) in the matrix element and next-to-leading order (NLO) in the parton distribution function (PDF). The matrix element was computed using MADGRAPH 3.3.1 and the sm-no_b_mass model (a "five-flavor scheme" in which the quarks contributing to the proton are assumed to all be massless except the top quark). The PDF set used was PDF4LHC15_nlo_mc_pdfas. A number of kinematic selections are set implicitly by MADGRAPH; these are disabled (e.g. the minimum jet $p_T$ is set to 0 instead of 20 GeV, etc.). For certain processes (c.f. Section II C) we do employ generator-level kinematic criteria.

For zero beam polarization, the authors of Ref. [4] determined the CC (NC) DIS cross-section to be $65 \text{ fb}$ (12 fb). In addition, the authors did a dedicated simulation of a top-associated Higgs production channel that is strongly influenced by the bottom quark content in the proton and the modeling of the proton’s gluon content (which can lead to $g^* \to b \bar{b}$ splitting). They estimated the cross-section for this process (Fig. 2) to be $0.0158 \text{ fb}$ (also at zero polarization).

The authors of the study also estimated the uncertainties associated with these production mechanisms owing to the scale and the combined PDF and $\alpha_s$ uncertainties. They estimated the scale uncertainty in the CC and NC DIS processes to be at the level of 6% and the combined PDF and $\alpha_s$ uncertainty to be at the level of 1%. For the top-Higgs associated production mechanism, these were estimated to be at the level of $1 - 2\%$ and 14%, respectively.

In this note, we reproduce some of the key results in

*Corresponding author: stephen.sekula@snolab.ca
Ref. [4] and explore in more detail the effect of PDFs on various production mechanism (especially as regards the modeling of heavy flavor content in the proton) as well as a dedicated study of the $H \rightarrow \tau$ final state. All simulations for this study were run using MadGraph [8, 9], Pythia 8.306 [10, 11], and Delphes 3 [12].

To facilitate the simulation of lepton-hadron collisions, we implement a custom calculation of $Q^2$ that utilizes the initial- and final-state leptons to ascertain the momentum transfer-squared to the proton. This is documented in Appendix B and is based on conversations with some of the authors of Ref. [4]. In addition, to facilitate the hadronization stage we employ a newer model of color flow for the beam remnant that is implemented in PYTHIA8, also documented in the appendix.

![Figure 1: Leading-order diagram for the production of a single final-state Higgs particle in charged-current (left) and neutral-current (right) muon-proton DIS.](image1)

![Figure 2: Leading-order diagram for the production of a single final-state Higgs particle in top-Higgs associated production via charged-current muon-proton DIS.](image2)

II. SIMULATION AND MODELLING

A. Cross-Section Estimations

The Snowmass MuIC white paper [4] considered Higgs production using the MadGraph model sm-no_b_mass, which implements a leading-order (LO) version of the
SM Lagrangian and sets the b-quark mass to zero in the parton-level interactions, effectively implementing a "5-flavor scheme." The Higgs Effective Field Theory (HEFT) is also provided in MadGraph, and the FeynRules package provides three alternative SM implementations that are stated to be NLO in QCD corrections [13]. The latter models come in three forms, denoted SM_NLO, SM5F_NLO, and SM4F_NLO for "...QCD corrections with the top and bottom yukawa [sic] renormalised in the on-shell scheme," "...QCD corrections in the 5 flavour scheme (massless b but non-vanishing bottom yukawa [sic] renormalised in the MSbar)," and "...QCD corrections in the 4 flavour scheme (massive b but bottom yukawa [sic] renormalised in the MSbar)," respectively [13].

For each model, 100,000 events were generated. The PDF used for each was CT18NNLO, unless otherwise specified. In all cases, the proton was redefined in MadGraph to contain b-quarks as well, e.g. define p = b b+ p. We provide additional details of the simulation in Appendix [5]. Results are in Table [1].

Regarding the tH production mechanism it is clear that a MadGraph model employing a 5-flavor scheme is essential to more accurately estimating the cross-section in a MuIC. At the level of uncertainty achievable from systematic on the scale (+0.98%, central scheme (-1.2%), and PDF (+19%) there is no clear difference between SM_NLO-full and SM5F_NLO-full. It is apparent that the level of certainty on the b-quark contribution to the proton is a strong influence on the precision with which this production mechanism can be estimated in advance, as PDF uncertainty dominates this calculation.

| Model Name            | Cross-Section [fb] |
|-----------------------|--------------------|
| p+μ− → H0νμ + jet    | 65.6               |
| p+μ− → H0μ+ + jet    | 11.5               |
| p+μ− → tH0νμ         | 0.000117           |

Table I: Cross-sections determined using MadGraph (MG5 AMC@NLO) and a range of available Standard Model implementations available in the framework or from FeynRules. The PDF set used for all cases is CT18NNLO. In all cases, the pure statistical uncertainly on the estimated cross-section is at the level of 5 × 10⁻² fb.

B. Heavy Flavor, PDFs, and Higgs Production

We investigated the effect on production cross-section resulting from the degree of knowledge of the flavor content of the proton (e.g., at the level of the strange quark and above). Given the role that the bottom quark plays in top-Higgs associated production, the inclusion of heavy flavor in both the underlying model and in the PDF model were of particular interest, should that mechanism be of future interest to a MuIC program.

We compared (Table I) the estimated production cross-sections for the three majors processes — CC DIS, NC DIS, and top-Higgs associated production — using a range of PDF sets. We begin by simply comparing the estimates obtained when using the reference PDF set in Ref. [3] (PDF4LHC15_nlo_mc_pdfs) to our baseline PDF set. We find that we essentially reproduce the results from that original work. We observe that the top-Higgs associated production is enhanced when utilizing the CT18NNLO PDF set, by about 20%. We note that, within the estimated theory uncertainties discussed earlier that this difference is not significant.

We also compare the production cross-sections using a foundational PDF fit — CT14 [14] — that is varied according to a range of assumptions about the fitted number of flavors (NF) and the computation of the PDFs to leading and next-to-leading orders (LO and NLO). We observe no effect on the CC DIS production process. This is compatible with the expectation that this mechanism is dominated by the contributions of valence and the lightest sea quark flavors. This observation is also compatible with the effect on NC DIS Higgs production, which shows negligible changes across the range of assumptions.

The most interesting pattern emerges from the variation of the number of heavy flavors that can contribute to the proton and the tH mechanism. Since this requires a bottom quark PDF, the variations with NF< 5 yield no useful predictions (one cannot model a proton containing a bottom quark when there is no PDF to accommodate this hypothesis, and one cannot produce the Higgs this way without the bottom quark). In the case where all six quark flavors are fitted in the PDF, we observe that increasing the order of the computation of the PDF decreases the production cross-section. This observation is
statistically valid across the different simulation samples, given the level of uncertainty on the yield of Monte Carlo simulation events. However, in a real data sample this difference would not be observable given the current level of theory uncertainty that contributes to knowledge of this process. That uncertainty would be a significant systematic uncertainty on the interpretation of collider data.

| Model Name          | Production Process | Cross-Section [fb] | $tH$ |
|---------------------|--------------------|-------------------|------|
| CT14NNLO            | CC DIS             | 63.9              | 11.0 | 0.0235 |
| PDF4LHC15_nlo_mc_pdfas | NC DIS            | 63.1              | 10.9 | 0.0196 |
| CT14lo              | CC DIS             | 63.6              | 11.2 | 0.0276 |
| CT14lo_NF3          | NC DIS             | 63.8              | 11.0 | N/A    |
| CT14NLO_NF3         | CC DIS             | 63.7              | 10.8 | N/A    |
| CT14NLO_NF4         | NC DIS             | 64.0              | 11.1 | N/A    |
| CT14lo_NF6          | CC DIS             | 63.9              | 11.2 | N/A    |
| CT14NLO_NF6         | NC DIS             | 63.6              | 11.4 | 0.0273 |
|                     |                    |                   |      | 0.0190 |

Table II: Cross-sections determined using MG5_ _AMC@NLO, the Standard Model implementation SM5F_Hcc_NLO available in the framework, and a range of PDF sets at different orders and with different quark flavor assumptions.

C. Background Processes

It is expected that there will be at least a few key background processes in the search for $H^0 \rightarrow \bar{c}c$:

1. Inclusive dijet production: while heavy flavor will be part of this, given charm and bottom jet mis-identification rates it is expected that the much larger production cross-section for inclusive dijets at a MuIC will lead to contamination of the final sample of $H^0 \rightarrow \bar{c}c$ candidates, even if light jet suppression is very good for such algorithms. This should be reducible by increased attention to light-jet rejection performance in any jet identification algorithm.

2. Charm dijet production: heavy flavor dijet production will, of course, directly contaminate the $H^0 \rightarrow \bar{c}c$ sample. This should be reducible by using the kinematics of Higgs decay to discriminate from this population of events, though of course this population will contain real charm jets.

3. $Z^0 \rightarrow \bar{c}c$ events: production of the $Z$ boson followed by decay to charm is expected to be a main, semi-irreducible background. It’s reducible in the sense that the $Z$ dijet mass will be different from the Higgs dijet mass, but given dijet mass resolution at the level of tens of GeV the key to separating these classes of events will depend on detector resolution and jet calibration. A MG5_ _AMC@NLO simulation of this process yields an expected inclusive $Z$ production cross-section of $0.331 \text{ pb (3.0)}$ for CC DIS (NC DIS).

It is possible to simulate aspects of the above backgrounds in MADGRAPH and PYTHIA8. Feynman diagrams for the processes utilized in MG5_ _AMC@NLO are shown in Appendices A4 and A5. However, it is challenging to simulate all of these background fully inclusively, as most of the events produced by these underlying mechanisms (especially the non-resonant dijet backgrounds) tend to result in events with extremely soft jets or other factors that make them trivially reducible (e.g. the detector will not see many or most of the products of the collision).

50 GeV, similar to Ref. [4]. We refer to this as "high-mass di-charm." Inclusive di-charm events are essential in revealing the characteristics in missing energy, muon, and jet kinematics that can be used as a first approach to rejecting di-charm events. The high-mass di-charm simulation is essential for a higher statistics sample of this background more likely to pass final selection criteria. Of
Figure 3: Event displays of a CC DIS $H^0 \to c\bar{c}$ event (left) and an NC DIS di-charm event (right), both of which pass all prototype selection criteria described in Section III. Although the perspectives are slightly different, the muon-beam (backward) direction is on the left-most side of each image. In the signal event, the backward two jets (blue cones) are associated with the Higgs decay. In the background event, the two most backward jets were associated with the candidate Higgs when none is present. No reconstructed muons are present in either event.

course, as we restrict the phase space at production we incur the cost of larger theoretical uncertainties. We will return to the challenges of the di-charm background at the conclusion of this note. We estimate the inclusive (high-mass) di-charm production cross-section to be 509,000 fb (97,700 fb). We simulate 200,000 collisions for each of these samples, a yield generally at or far below what would be expected for our target integrated luminosities (e.g., 1 fb$^{-1}$ or 100 fb$^{-1}$).

The systematic uncertainties due to scale, $\alpha_s$, and PDFs are generally larger than these statistical effects (which can be reduced merely by simulating more events of this type). For example, the scale uncertainty is about 20% while the central scheme uncertainty is over 100% in both cases described above and worsens as one further restricts the phase space of the process. As expected, the more exclusive the process the larger are some associated theory uncertainties. Restricting the simulation from the fully inclusive approach (e.g. no limitations on jet $p_T$ or $\eta$) at this level nonetheless retains a sample of events that could be reconstructed by a detector, albeit in a phase space that is only part of the total inclusive rate.

The graphs in this note are generally normalized in one of two ways. When "density" appears on the y-axis, the contents of the figure have been normalized using the area of distribution and the width of the histogram bins. Otherwise, we normalize the samples to their production cross-sections (and branching fractions, where appropriate) and a target luminosity of 100 fb$^{-1}$. The density normalization allows for a comparison of shape differences in samples and populations; the luminosity normalization allows for a sense of how these populations will look in a future data sample.

We note that, unlike in the case of $H^0 \to b\bar{b}$ explicitly studied in Ref. [3], $H^0 \to c\bar{c}$ appears to suffer from at least two major disadvantages at the level of a generator- or pre-selection of events:

1. The branching fraction for $H^0 \to c\bar{c}$ is significantly lower than for $H^0 \to b\bar{b}$ owing to the much smaller Yukawa coupling to charm quarks (a factor of about 20);

2. The probability of non-resonantly pair producing charm quarks in a high-energy collider is much greater than that of producing a pair of bottom quarks, due to the much smaller mass of the charm quark.

These two factors could make the study of $H^0 \to c\bar{c}$ particularly challenging, even in this kind of collider environment. We explore the reconstruction and selection...
of these events to identify the basic challenges of this process after a modest, baseline selection.

1. Visualization of Events

We show DELPHES event displays of the CC DIS $H^0 \to cc$ process and for the NC DIS di-charm process in Fig. 3. The events shown here pass all selection criteria outlined in Section [11] and are considered extremely "signal-like." The $H^0 \to cc$ event has two backward jets that are identified as charm-induced and form a di-jet invariant mass consistent with the Higgs particle. Other jets are present as a result of the proton beam remnant, radiation, etc. The background event is induced by a NC DIS process but no reconstructed muon is present. The two most backward jets are the ones associated with the Higgs candidate. Both events have significant missing transverse momentum. The analysis we outline later is very basic, but these event displays may hint at topological differences in jets that may be exploited by more advanced methods.

D. Jet Reconstruction in Detector Model

The detector model employed in this study is based on the "EIC Generic 3T" model provided in the delphes_eic package [15]. The model is modified, however, in a few key ways:

- The calorimeter design can be changed to try alternatives, including the original EIC Generic calorimeter model [1], a model similar to that proposed for the ATHENA detector concept [10], and a model similar to the CMS calorimeter adopted in Ref. [11]. These are referred to as the "EIC Generic," "ATHENA," and "Snowmass" concepts, respectively. We generally adopt the "Snowmass" configuration in our studies. While it does not appear to provide the best resolution, it provides a somewhat more conservative baseline. All calorimeter models include an electromagnetic and hadronic calorimeter component with identical coverage in pseudorapidity, though not identical cell granularity.

- Jets can be reconstructed using only calorimeter towers ("Calorimeter-only", or "CaloJets") or using particle flow. We allow for both kinds of jet and missing energy reconstruction to facilitate studies of the effect of particle flow algorithms on a future Higgs candidate reconstruction (simply referred to henceforth as "Jets"). Our baseline, unless otherwise stated, uses particle flow (c.f. Ref. [17]). All jet reconstruction approaches use the anti-$k_T$ [18] algorithm.

- A muon efficiency model is added to represent the performance of a future, low-angle muon spectrometer system (intended primarily for beam muon reconstruction in, for example, neutral-current DIS).

LHC experiments employ a range of $R$-parameter choices for anti-$k_T$ jets. For example, standard jets in CMS or ATLAS are typically defined with $R$ parameters of 0.5 or 0.4, respectively [19, 20]. Highly Lorentz-boosted heavy objects decaying to jets will result in larger radius jet clusters, and the typical $R$ parameter for this situation is 1.0 [21]. Substructure within a large-radius jet can be determined by re-running jet clustering inside the object using charged particle tracks, and typically this is done using jets with a narrower $R$ parameter (e.g., 0.2) or a sliding scale that adjusts the $R$ parameter depending on the kinematics of the substructure jet relative to the parent jet. It is not a priori obvious what is the right or best choice for jet reconstruction of $H^0 \to cc$ at a future MuIC.

We reconstructed jets using a few different approaches and selected an $R$ parameter that balanced Higgs mass accuracy and resolution. We scanned across the possibilities $R = 0.4 - 1.0$ in steps of 0.2. To avoid over-constraining the choice to just one possible heavy flavor, we looked at the quality of both di-charm and di-bottom reconstruction. Results are shown in Fig. 4

We hypothesized that a baseline means of selecting the best $R$ parameter was to simply observe the rate at which $H^0 \to b\bar{b}$ and $H^0 \to cc$ simulated events were reconstructed as having two heavy flavor jets. We recognize that the production of charm quarks in the proton remnant affects this hypothesis, as a fraction (at the level of a few percent) of $H^0 \to cc$ events will contain at least 3 charm jets. Nevertheless, most events would be expected to contain only two such heavy flavor jets.

For iterations $R = 0.4$, $R = 0.6$, $R = 0.8$, and $R = 1.0$, events with two $b$-jets occurred with a frequency of $87\%$, $90\%$, $91\%$, and $91\%$, respectively. The statistical uncertainty on these rates is smaller than the reported precision. Di-charm jet production occurred under the same parameters with frequencies of $82\%$, $84\%$, $84\%$, and $85\%$. We observe that for both heavy flavor possibilities
there is a clear few percent-level rise in the rate of di-jet reconstruction as we move from $R = 0.4$ to 0.6. However, after that, the rise is less dramatic and we generally observe that the reconstruction rate of heavy flavor di-jets is flatter for larger jet $R$ parameters. We select a point on the trend that is past the early rise and early in the plateau. Thus, $R = 0.6$ was chosen as the most optimal $R$ value for all future work in this note. A comparison of light and charm jets in CC DIS $H^0 \rightarrow c\bar{c}$ simulation is shown in Fig. 5.

2. Jet Resolution

In the context of our selected jet reconstruction, we can explore the fast-simulation-level expected quality of jets used in the reconstruction of Higgs particles. In particular, the resolution of a jet variable $X$ is defined as $(X_{reco} - X_{true})/X_{true}$, where "reco" labels the variable determined purely from a reconstruction-based approach and "true" labels the variable as determined from a Monte Carlo generator-level set of information. Generator-level jets are defined as jets reconstructed just as they would be from calorimeter information, but instead employing all long-lived stable particles from the simulation before detector-level effects are simulated.

The typical jet energy resolution at a modern experiment, such as ATLAS, varies with the kinematics (e.g. $p_T$) of the jet object. ATLAS obtains a jet energy resolution of about 24% at $p_T = 20$ GeV that improves to about 6% at 300 GeV. The CC DIS production of $H^0 \rightarrow c\bar{c}$ in our simulation results in charm jets with a typical $p_T$ in the range of 30–40 GeV. The energy resolution obtained in our fast simulation is shown in Fig. 6. We estimate it for both CaloJets and Jets. We can clearly see the positive effect of employing particle flow, as this improves significantly the resolution of the energy and its accuracy to the true value from the generator-level jet. The root mean-square (RMS) of the particle flow jet energy resolution is about 21%, consistent with typical LHC jet energy resolution measurements for a similar $p_T$ spectrum.

E. Higgs Jets and Kinematics

We compare the $p_T$, azimuthal angle ($\phi$), and pseudorapidity ($\eta$) between generator-level jets and reconstructed jets. As a basic check, we observed that the distribution of $\phi$ was consistently isotropic, as expected given both a symmetric detector model and a zero crossing angle for the muon and proton beams.
We observe that the jets in $H^0 \rightarrow c\bar{c}$ CC DIS events are typically produced in the backward (muon beam).

We have observed that the jet energy resolution for charm jets is consistent, in this fast simulation, with typical resolutions at existing experiments. We investigate also the mass reconstruction of the dijet system. This is expected to be a key discriminant in any Higgs reconstruction. When comparing jets and CaloJets (Fig. 6) it is clear that particle flow provides a more accurate and precise mass reconstruction.

Our baseline configuration results in a median (peak) dijet mass of 111 GeV (117 GeV for truth-matched charm jets in the CC DIS $H^0 \rightarrow c\bar{c}$ simulation. The standard deviation of the mass peak is about 21 GeV. We would expect to thus be able to separate the $H^0 \rightarrow c\bar{c}$ mass peak from a $Z \rightarrow c\bar{c}$ background mass peak at the level of about 1σ; however, this is without employing standard methods for improving the calibration of jets, such as jet-energy scale and muon-in-jet corrections (which compensate for jet mis-reconstruction and energy loss due to semi-leptonic heavy quark decays). These results are consistent with those obtained using $H \rightarrow b\bar{b}$ decay in Ref. [4].

We also compare the mass reconstruction using the different calorimeter models described earlier. We see no particular advantage in this particular quantity in choosing one calorimeter model over the others (Fig. 7). The importance of calorimeter granularity and coverage will be more apparent in the study of missing momentum.

We adopt a charm-tagging performance for jets similar to that in Ref. [6], with 27%, 8%, and 1.6% c-jet, b-jet, and light-jet efficiency, respectively. This is taken as independent of the jet momentum, though in reality direction, which would be expected from the asymmetric beam kinematics (Fig. 7). The proton remnant would be expected to travel in the forward direction; indeed, separating charm jets from light jets, we observe that the light-jet component of the $\eta$ distribution more often contains a population of jets in the forward direction. Charm jets, primarily resulting from Higgs decay, tend to travel forward. The jet density in $\eta$ peaks around $\eta = -1.5$.

Tracking and calorimetry in a future MuIC detector will need to provide sufficient coverage and pixellation in the backward direction. We show momentum and polar angle coverage plots in Fig. 8, emphasizing the typical region populated by generator-level jets in a collider of this configuration (275x960 $\text{GeV}^2$ protons on muons). The majority of decay products fall at $\eta = [-2.5, 0]$.

Taking this into account with the specified jet parameter $R = 0.6$, a MuIC collider detector will require calorimeter and tracking coverage out to at least $|\eta| = 4.0$, with emphasis for jet reconstruction in the direction of the muon beam.

We recognize that there would be angular and momentum effects that would need to be folded into the development and performance assessment of any charm-tagging algorithm.

**F. Muon Identification**

The production of di-charm final states, from either resonant or non-resonant mechanisms, will occur more often in NC DIS than CC DIS. However, Higgs production is dominated by the CC DIS mechanism. Rejecting
Figure 8: Coverage plot for all generator-level jets (top) and for generator-level charm jets (bottom) illustrating the relationship between the jet momentum $p$ as a function of polar angle $\theta$.

Figure 9: Comparison of Particle Flow and Calorimeter Jets for $H \rightarrow c \bar{c}$ Higgs reconstruction.

Figure 10: Dijet invariant mass distribution for CCDIS $H \rightarrow c \bar{c}$ compared between alternative calorimeter models.

The significant production of charm through NC DIS will require attention to muon and missing momentum reconstruction.

A backward muon with high energy will be a potential hallmark of the NC DIS process. At the generator level, muon momentum and pseudorapidity both show the characteristics expected of these "tagging muons" — those scattered from the original beam-beam interaction (Fig. 11). Tagging muons from NC DIS interactions are characteristically in the direction of the muon beam ($\eta < -3$) and very high in momentum, typically with more than 10% of the lepton beam’s original energy. Final-state muons, resulting from the decay of heavy particles, have a much softer momentum spectrum and tend to be much more central in the detector (following more the angular distribution of jets).

Identifying these muons will require the deployment of a far-backward muon system that can detect high-momentum minimum ionizing particles all the way back to $\eta \approx 6$ or lower. LHCb is an excellent example of an experiment with a low-angle muon spectrometer that could serve as a model for a future MuIC detector subsystem with this capability [22, 23]. We assume a conservative LHCb-like performance in the region $\eta = [-5.0, 4.0]$ [24]. This results in 97% (3%) muon (pion) efficiency throughout the region of acceptance for muons.
Figure 11: Generator-level characteristics of muons from the beam-beam interaction (tagging) or from final-state decays of heavy particles (FS).

G. Missing Momentum and Energy in CC DIS

Missing momentum and energy are expected to be key discriminators in the study of $H^0 \rightarrow c \bar{c}$ decay, especially in distinguishing CC and NC DIS production. Missing energy reconstruction in DELPHES is performed using the energy flow objects (tracks and clusters). Missing transverse four-momentum is computed simply using

$$\vec{p}_T^{\text{miss}} = - \sum_i \vec{p}_i \quad (1)$$

where $\vec{p}_i$ is the momentum of an energy flow object. An example of the baseline resolution ($\frac{p_T^{\text{miss, true}}/p_T^{\text{miss, true}}$} is shown in Fig. 12 comparing $E_T^{\text{miss}}$ calculated using raw calorimeter objects and using particle flow objects. As expected, particle flow-based reconstruction provides a superior resolution and fidelity to the true missing transverse momentum.

Figure 12: MET Resolution for calorimeter only and particle flow. Events are from CC DIS signal simulation.

As an assessment for future consideration of different calorimeter granularity designs, we also calculate the $E_T^{\text{miss}}$ for the alternative calorimeter models. A comparison in CC DIS signal simulation is shown in Fig. 13. These include the ATHENA-like model, a baseline EIC detector model, and what we refer to as the Snowmass calorimeter (our baseline for this note). Only particle flow is utilized. We observe that the Snowmass granularity and coverage offers the overall largest resolution of the three designs. For example, in this fast simulation the Snowmass particle-flow-based $E_T^{\text{miss}}$ resolution, determined by the standard deviation of the distribution, is about 33%. In comparison, the resolution in the EIC and ATHENA models (which are similar in granularity) is about 29%. This suggests both that our results from this note will be conservative and that designs with more granularity would benefit future experiments that need stronger $E_T^{\text{miss}}$ resolution.

H. Missing Momentum and Energy in NC DIS

Our fast simulations of CC DIS signal and di-charm background suggests there is little benefit in separating these two populations based solely on the shape of the $E_T^{\text{miss}}$ distribution. However, the distinction between CC DIS signal and NC DIS di-charm background is far more significant. NC DIS events maintain the beam muon and produce little or no real $E_T^{\text{miss}}$, except from final-state
Figure 13: Comparison of the MET Resolution for the different calorimeter models in \( H^0 \rightarrow c\bar{c} \) decay.

physics involving one or more low-energy neutrinos or from mis-reconstruction of the event. The \( E_T^{\text{miss}} \) in these two populations is compared in Fig. 14. As expected, \( E_T^{\text{miss}} \) peaks more sharply toward zero in the NC DIS di-charm background, while it extends to typically much larger values in CC DIS signal simulation.

Figure 14: Comparison of MET between di-charm background and \( H^0 \rightarrow c\bar{c} \) signal in both CC DIS and NC DIS production modes.

We hypothesized that there might be additional benefits in looking at the angular distribution of the MET relative to the di-jet system that forms the Higgs candidate. We recognize that in a collision where only the Higgs and the recoiling neutrino are produced, the \( E_T^{\text{miss}} \) is entirely defined as being opposite the sum of the jet particles and thus is 100% correlated with the di-jet system. However, in collisions where part or all of the beam remnant is present, or where additional radiation or other particles appear in the detector, the correlation will be weakened. Naively, we anticipated that in CC DIS Higgs signal events we should expect to see a strong peak in the azimuthal angle difference between the \( E_T^{\text{miss}} \) and the Higgs candidate around \( \pm \pi \). In NC DIS di-charm events, where there is no real significant source of \( E_T^{\text{miss}} \) and the possibility of a muon in the detector, we anticipated that there would be no or a much weaker correlation.

We show the results of a basic study of this in Fig. 15. As predicted, the difference in the missing energy angle for CC DIS signal peaked at \( \pm \pi \), while the NC DIS di-charm background was somewhat more diffuse. Nevertheless, we still observe strong \( \pm \pi \) values for this angular difference in the transverse plane. We do not exploit this possible distinction in angular correlations for the remainder of this note, but we point out that such a variable (with additional refinement) may be of future use in an analysis that targets extreme separation between CC DIS signal and NC DIS di-charm background.

III. A PROTOTYPE DATA ANALYSIS FOR \( H^0 \rightarrow c\bar{c} \)

We present here the outcome of a very basic event selection and candidate refinement. This is meant to provide a conservative view of the potential for selecting \( H^0 \rightarrow c\bar{c} \) candidates, with emphasis on CC DIS production (the leading cross-section). We apply a basic optimization procedure to set criteria at various stages of the analysis, employing the method of varying a criterion to maximize the following figure of merit \[25\]

\[
\sqrt{2 \left( s + b \right) \ln(1 + s/b) - s}
\]  \hspace{1cm} (2)

where \( s \) (\( b \)) is the number of events in the signal (background) class. This quantity is the asymptotic formula for significance in the limit of large backgrounds, which is a reasonable assumption in this study. The class definitions vary depending on the goal of the selection. We conservatively round any criteria determined from this procedure to avoid over-interpretation of the optimization results on current simulation approaches and statistical limitations (e.g., 66 GeV becomes 70 GeV on an energy criterion).

The most basic event characteristics are that the muon-proton collisions, through the CC DIS process, will produce a significant and real (e.g. due to at least one high-momentum neutrino) quantity of missing momentum (Fig. 16). In addition, the event should
not contain evidence of a high-momentum muon at low angle in the backward direction; that signature is a hallmark of the NC DIS process (Fig. 17). Given the lower cross-section for Higgs production in NC DIS and the substantially increased rate of production of non-resonant di-charm and $Z^0 \rightarrow c\bar{c}$ in NC DIS, our goal is to minimize contamination from NC DIS events. We define the signal class as containing luminosity-weighted CC DIS $H^0 \rightarrow c\bar{c}$ events and the background class to contain NC.

Candidate events should then contain at least two jets DIS $Z^0 \rightarrow c\bar{c}$ and inclusive di-charm events. We find that $E_{\text{miss}}^T > 50$ GeV and rejecting events with a backward (bwd) muon satisfying $p_{\text{bwd}}^\mu > 60$ GeV optimizes our criteria. The optimizations are not performed accounting for correlations in the two variables and are only one-sided in each variable (a minimum in $E_{\text{miss}}^T$ and maximum in $p^\mu$). We refer to the selection on MET and the muon veto as the "Event Tag."
minimum $p_T$ to be compatible with the decay of a heavy resonance. This is referred to as the "Jets Selection." Among those jets, there should be at least two charm-tagged jets. These jets should be located primarily in the backward direction. Constraints that emphasize the selection of events consistent with these expectations are referred to as "Charm Dijet Selection."

Finally, we expect the dijet system to form an invariant mass consistent with the Higgs particle. Constraints on the event that emphasize specifically the Higgs signature are referred to as "Higgs Selection."

The selection criteria and their efficiency on the three main samples — $H^0 \rightarrow c\bar{c}$, $Z^0 \rightarrow c\bar{c}$ (CC DIS), and di-charm — are shown in Table III. The samples are weighted to a target luminosity of 100 fb$^{-1}$. We note that the NC DIS di-charm background was the most difficult to simulate in sufficient quantity for this basic study. Whether using the more inclusive sample or the sample where $m_{c\bar{c}} > 50$ GeV at generator-level, we obtain almost the same number of raw events passing all selection criteria: about 6. This means any weighted prediction using these samples have a 40% MC simulation statistical uncertainty and can only be used to guide thinking at this stage.

We also simulated NC DIS $H^0 \rightarrow c\bar{c}$ production (not shown in Table III), but as the analysis was designed to eliminate NC DIS production this resulted in a negligible contribution to the overall signal yield (about 1% of the $H^0 \rightarrow c\bar{c}$ yield). We note that where we expect consistent performance in selection criteria between samples, we do observe it, giving confidence to the underlying simulation models. For example, we expect the event tag criteria to substantially eliminate NC DIS; they do, maintaining CC DIS in any sample at a high level of efficiency while suppressing NC DIS-based samples. We expect that charm identification will perform with about the same efficiency on the $H$ and $Z$ simulation samples, since each contains high-momentum real charm-initiated jets. Indeed, we see about the same performance in both samples (a 5% di-charm efficiency).

Based on our simulations, the most promising target for a future MuIC $H^0 \rightarrow c\bar{c}$ search would be to emphasize the CC DIS production mechanism. This conclusion is based solely on the challenge posed in rejection NC DIS di-charm background from this analysis. Missing momentum and muon identification will be essential ingredients in this effort; the performance on these two physical observables, in turn, is defined by the quality of the calorimeter (and tracking, for particle flow) as well as the coverage of a backward muon system dedicated to beam-muon tagging. Our approach has utilized a calorimeter of reduced granularity compared to alternative models emphasized for the EIC detectors and is therefore expected to provide a much more conservative performance on MET than we think is achievable using the kind of calorimeter designs anticipated for EIC experiments.

Proposed EIC detectors do not emphasize muon reconstruction with low-angle coverage, so the deployment of such a dedicated muon system will be necessary for the MuIC. We see from Table III that the NC DIS di-charm background is of particular concern owning to its substantial cross-section in the jet fiducial region of interest. Reducing this background will rely heavily on efficient event tagging with a strong separation between NC and CC DIS candidate events.

We note that even at early stages of the selection the limited statistics of the di-charm sample are apparent. It was not possible for this basic study to simulate the target luminosity’s worth of inclusive di-charm events, forcing us to apply large weights to these events. The effect of limited statistics becomes greater at increasingly
restrictive selection. In our figures, for any distribution beyond the event tag stage we we utilize the high-mass di-charm simulation. Even here, for the present study we have limited statistics compared to the target luminosity.

We show in Fig. 18 a distribution of the charm-tagged di-jet mass. We recognize first and foremost that it is difficult to draw conclusions about the NC DIS di-charm background; nevertheless, the earlier stages of event selection suggest that even at this most restricted stage one should expect this population to dominate in the current prototype analysis. Additional resources will be needed to simulate sufficient samples of this background and additional approaches, such as more information and/or a correlated multivariate analysis, will be needed to improve selection of signal and rejection of this overwhelming background.

Beyond the NC DIS di-charm background it is clear that work will also be needed to improve the ability to separate Z-boson-induced backgrounds from the $H^0 \rightarrow c\bar{c}$ signal. Separation of these topologies at the LHC has made substantial progress in the past decade, and no doubt a refined approach can be brought to the question of the viability of a future MuIC search for $H^0 \rightarrow c\bar{c}$. In many ways, this measurement at the MuIC may be as challenging as was the original measurement of $H^0 \rightarrow b\bar{b}$ at the LHC. The accomplishment of that task within a decade of starting the LHC program provides lessons for this future-looking issue.

Figure 18: Charm-tagged di-jet mass with all selection criteria applied for a target integrated luminosity of 100 fb$^{-1}$. The NC DIS di-charm background is not reliable owning to the low statistics of the population surviving all selections.

IV. CONCLUSIONS

We have studied the production of the $H^0 \rightarrow c\bar{c}$ final state at a future MuIC facility using a conservative baseline detector inspired by earlier work for the Snomass community planning exercise. We have assumed rather modest (27%) charm jet-tagging efficiency, which is generally considered to be the minimum level of performance a modern algorithm can manage. We have assumed a more coarse calorimeter than those generally planned for EIC experiments, but in addition to baseline EIC-like detector configurations we have added a low-angle LHCb-like muon system to a fast simulation of an experiment.

We identify NC DIS non-resonant di-charm background as the leading challenge in any MuIC search for $H^0 \rightarrow c\bar{c}$. The second-leading background is expected to arise from $Z^0 \rightarrow c\bar{c}$ production. We have not explored in this note the reality that the $b \rightarrow c$ jet mis-identification rate, typically at the level of a 10% or so, will mean an additional background both from $Z^0 \rightarrow b\bar{b}$ and, more importantly, $H^0 \rightarrow b\bar{b}$ events. Given existing approaches to simultaneous fits of di-jet samples for both $b$ and $c$ final states, we have confidence that should other backgrounds be reduced this one can be resolved through the use of a simultaneous fit of a Higgs candidate sample for both final states, including mis-identification rates assessed on independent samples.

We comment on an element that is missing from this study and that would necessarily be a target for future work. The charm-tagging mis-identification rate for light jets in this study was 1.6%. Naively, the light di-jet mis-identification rate would be than about $2.6 \times 10^{-4}$. However, the light di-jet production cross-section from NC DIS or CC DIS will be much larger than the cross sections studied for background processes in this note. We expect a significant challenge from light di-jet background in a future $H^0 \rightarrow c\bar{c}$ analysis, one that will also be especially difficult to simulate in sufficient quantity and that might benefit from entirely data-driven approaches in a real data analysis. We leave this as a comment for future experimental physicists.

We conclude from this preliminary work that any future MuIC experiment must have good calorimeter granularity and readout capability, good tracking coverage for the purpose of both particle-flow reconstruction and jet acceptance (especially for charm-jet tagging performance at lower angles), and muon identification capability at low angle. Nevertheless, these are necessary but not sufficient ingredients for ultimately selecting $H^0 \rightarrow c\bar{c}$ candidates in data. The challenge of background reduction, rejection, and characterization will provide daunting if such a search is undertaken. We caution the use of $H^0 \rightarrow c\bar{c}$ as a
motivation for a future MuIC program. However, we advocate for future consideration of this channel as part of a healthy MuIC physics analysis portfolio.

V. ACKNOWLEDGEMENTS

We are extremely grateful to Darin Acosta, Osvaldo Miguel Colin, and Xunwu Zuo for valuable conversations and insights about Monte Carlo simulation and underlying physics issues in the production processes for the Higgs. We are extremely grateful to Fred Olness for valuable discussions about PDFs in the study of proton structure and Higgs production. We also acknowledge the support of the Department of Physics at SMU especially for providing facilities to support the research activities of the authors. J. Choi expresses gratitude toward Highland Park High School for its encouragement of academic excellence and especially for its Science and Technology Festival, which resulted in becoming a part of this project. P. Ahluwahlia expresses gratitude toward the Harmony School of Innovation for its encouragement of academic excellence and independent research. We are grateful to the SMU Office of Information Technology and the Center for Research Computation at SMU for access to and support on the ManeFrame II supercomputer cluster.

[1] R. Abdul Khalek et al., Nucl. Phys. A 1026, 122447 (2022), arXiv:2103.05419 [physics.ins-det]
[2] R. Abdul Khalek et al., in 2022 Snowmass Summer Study (2022) arXiv:2203.13190 [hep-ph]
[3] D. Acosta and W. Li, Nucl. Instrum. Meth. A 1027, 166334 (2022) arXiv:2107.02073 [physics.acc-ph]
[4] D. Acosta, E. Barberis, N. Hurley, W. Li, O. M. Colin, D. Wood, and X. Zuo, in 2022 Snowmass Summer Study (2022) arXiv:2203.06258 [hep-ex]
[5] V. Shiltsev and F. Zimmermann, Rev. Mod. Phys. 93, 015006 (2021) arXiv:2003.09084 [physics.acc-ph]
[6] G. Aad et al. (ATLAS), Eur. Phys. J. C 82, 717 (2022) arXiv:2201.11428 [hep-ex]
[7] CMS Collaboration, “Search for higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ tev.” (2022), arXiv:2205.05550 [hep-ex]
[8] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, JHEP 07, 079 (2014) arXiv:1405.0301 [hep-ph]
[9] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, JHEP 07, 185 (2018) [Erratum: JHEP 11, 085 (2021)], arXiv:1804.10017 [hep-ph]
[10] C. Bierlich et al., (2022), arXiv:2203.11601 [hep-ph]
[11] T. Sjostrand, S. Mrenna, and P. Z. Skands, JHEP 05, 026 (2006) arXiv:hep-ph/0603175
[12] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaitre, A. Mertens, and M. Selvaggi (DELPHES 3), JHEP 02, 057 (2014) arXiv:1307.6346 [hep-ex]
[13] C. Duhr, N. Christensen, B. Fuks, and C. Degrande, http://feynrules.irmp.ucl.ac.be/wiki/StandardModel, accessed: 2022-06-07.
[14] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, Phys. Rev. D 93, 033006 (2016) arXiv:1506.07443 [hep-ph]
[15] M. Arratia and S. Sekula, (2021), 10.5281/zenodo.4592887 arXiv:2103.06886 [physics.ins-det]
[16] J. Adam et al. (ATHENA), JINST 17, P10019 (2022) arXiv:2210.09048 [physics.ins-det]
[17] CMS Collaboration, CMS-PAS-PFT-09-001 (2009).
[18] M. Cacciari, G. P. Salam, and G. Soyez, JHEP 04, 063 (2008) arXiv:0802.1189 [hep-ph]
[19] V. Khachatryan et al. (CMS), JINST 12, P02014 (2017) arXiv:1607.03663 [hep-ex]
[20] G. Aad et al. (ATLAS), Eur. Phys. J. C 81, 689 (2021) arXiv:2007.02945 [hep-ex]
[21] G. Aad et al. (ATLAS), Eur. Phys. J. C 81, 334 (2021) arXiv:2009.04986 [hep-ex]
[22] A. A. Alves, Jr. et al., JINST 8, P02022 (2013) arXiv:1211.1346 [physics.ins-det]
[23] L. Anderlini et al., JINST 15, T12005 (2020) arXiv:2008.01579 [hep-ex]
[24] LHCb Collaboration, “Standard set of performance numbers,” https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html
[25] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C 71, 1554 (2011) [Erratum: Eur.Phys.J.C 73, 2501 (2013)] arXiv:1007.1727 [physics.data-an]
Appendix A: MG5\_aMC@NLO Feynman Diagrams

This appendix contains the Feynman diagrams that are employed by MG5\_aMC@NLO in the calculation of the matrix element, and thus the cross-section, for Higgs production and background processes of interest to the effort described in this note.

The following is the key for the particles in the diagrams:

- Quarks and gluons are represented by their usual lower-case letters (u, d, c, s, b, t, g);
- Anti-quarks are denoted by the placement of a tilde next to the quark letter;
- Photons are denoted by $a$;
- Z and W bosons are denoted by lower-case versions of their letter names.

1. Charged Current DIS Vector-Boson Fusion (SM5F\_NLO)
Figure 19: Diagrams used in the calculation of the CC DIS VBF process (SM5F_NLO).
Figure 20: Diagrams used in the calculation of the CC DIS VBF process (SM5F_NLO).
2. Neutral Current DIS Vector-Boson Fusion (SM5F_NLO)

Figure 21: Diagrams used in the calculation of the NC DIS VBF process (SM5F_NLO).
Figure 22: Diagrams used in the calculation of the CC DIS VBF process (SM5F_NLO).
3. top-Higgs Vector-Boson Fusion (SM5F_NLO)

![Diagram](image)

Figure 23: Diagrams used in the calculation of the top-Higgs DIS VBF process (SM5F_NLO).
4. Neutral-Current DIS Di-Charm Production (No Higgs Resonance)
Figure 24: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by up quarks in the proton.
Figure 25: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by up quarks in the proton.
Figure 26: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by down quarks in the proton.
Figure 27: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by down quarks in the proton.
Figure 28: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by charm quarks in the proton.
Figure 29: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by charm quarks in the proton.
Figure 30: Diagrams used in the calculation of the NC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by gluons in the proton.
5. Charged-Current DIS Di-Charm Production (No Higgs Resonance)
Figure 31: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by up quarks in the proton.
Figure 32: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by up quarks in the proton.
Figure 33: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by down quarks in the proton.
Figure 34: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by down quarks in the proton.
Figure 35: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by strange quarks in the proton.
Figure 36: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by strange quarks in the proton.
Figure 37: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by charm quarks in the proton.
Figure 38: Diagrams used in the calculation of the CC DIS di-charm (non-resonant) process (SM5F_NLO). The processes emphasized here are induced by charm quarks in the proton.
Appendix B: MG5_aMC@NLO and Pythia8 Configurations

We here document, for the benefit of others who may wish to conduct similar simulations, the configuration of the matrix-element-level and hadronization-level simulation tools. All of these are achieved through the use of the MG5_aMC@NLO interface, and the following code examples are provided assuming all simulation is run through this interface.

1. Generator-level configuration

The following code block illustrates the configuration of a proton-muon collider with the goal of producing the Higgs particle through CC DIS.

```plaintext
set group_subprocesses False
define p = p b b~
define j = j b b~
import model SM5F_NLO
generate p μ− > j vm h
output pmu_275x960_ccdis_vbf_h−SM5F_NLO
```

The order of the initial- and final-state particles is important to the user-defined \(Q^2\) calculation, defined below, as the calculation expects the initial lepton and the final lepton to have fixed positions in the generator string (e.g. the second and fourth positions).

The above configuration separates the subprocesses that result in the same final state; this is not observed to be necessary during signal simulation, but essential to the stability of the background simulation, which involves many more processes. We follow the procedure of Ref. [4] and add bottom quark flavor to the proton definition, allowing it also to be produced in any explicit jet in the final state (e.g. through a flavor-changing or flavor-preserving process). This allows bottom quarks to be controlled by the PDF.

The follow generate strings illustrate how we produced the Higgs also through NC DIS, top-Higgs production, and how we produced di-charm background.

```plaintext
# NC DIS Higgs Production
generate p μ− > j μ− h

# CC DIS di-charm Production, no Higgs
generate p μ− > j μm c c~ /h

# NC DIS di-charm Production, no Higgs
generate p μ− > j μ− c c~ /h
```

2. Production-level configuration

The following code block illustrates the production-level configuration that would be executed to make a larger number of simulated events, having defined the generator-level process. We employ the CT18NNLO PDF, which is LHAPDF ID 14000.

```plaintext
launch pmu_275x960_ccdis_vbf_h−SM5F_NLO
shower=Pythia8 0
set nevents 20000
set ebeam1 275
set ebeam2 960
set polbeam1 0
set polbeam2 0
set pdlabel1 lhapdf
set lhaid 14000
set ptj 5
set ptl 0
set ptjmax −1
set ptlmax −1
set pt_min_pdg {}
set pt_max_pdg {}
set etaj −1
set etal −1
set etalmin 0
set eta_min_pdg {}
set eta_max_pdg {}
set mxx_min_pdg {}
set drjl 0
set drj1max −1
set ptheavy 0
set maxjetflavor 5
set dynamical_scale_choice 0
#set systematics_program none
```

The dynamical scale choice enforces the user-level definition we provide below. The systematic error computation can be disabled if all one needs are generated events (e.g. if one is not interested in a production cycle in knowing the theory uncertainties on the cross-section).

3. \(Q^2\) definition

We modify the file in the output folder for each process named SubProcesses/setscales.f and define the following user-defined scale:

\[
rscale = \max(0d0,−\text{sumdot}(p(0,2),p(0,6),−1d0))
\]

This assumes the the initial and final-state particle are the second and the
4. QCD color flow model in Pythia8

We also modify the color flow model for the beam remnant since the lepton-hadron collisions have been observed to pose a challenge to the default model. Specifically, we modify the MG5_aMC@NLO pythia8_card_default.dat for each simulation process and add the following lines at the end:

```plaintext
BeamRemnants:remnantMode = 1
ColourReconnection:mode = 1
```