sPH-HF-2021-001
April 22, 2021

First MDC1 Results from Heavy Flavor Topical Group

Sebastian Tapia Araya[1], Cameron Dean[2], Jin Huang[3], Hideki Okawa[4], and Zhaozhong Shi[5]

[1]Iowa State University
[2]Los Alamos National Laboratory
[3]Brookhaven National Laboratory
[4]Fudan University
[5]Massachusetts Institute of Technology

Abstract

This note details the first studies of simulation data from the First Mock Data Challenge (MDC1) in sPHENIX undertaken by the Heavy Flavor (HF) Topical Group (TG) in 2021. One hundred million heavy flavor signal events are generated using Pythia 8.3, simulated through Geant4, and the HF resonance reconstruction is performed by KFParticle, adapted for use within the Fun4All framework. Decays of D^0, $D^{\pm}$, $D_s^{\pm}$ and $D^{*\pm}$ mesons are analyzed in various final states and initial strategies for prompt/non-prompt separation are demonstrated. More analyses and results are expected from this MDC1 simulation dataset.

1 Introduction

This is a brief technical note summarizing the first set of performance plots from the Heavy Flavor (HF) Topical Group (TG) using the KFParticle package in the First Mock Data Challenge (MDC1) in sPHENIX, which was presented at the third Software and Computing Review in March 2021 [1]. Since the review, the members of the HF TG are continuing to analyze the MDC1 simulation samples with regular updates at the TG meetings. However, several ongoing studies, e.g. Refs. [2, 3], are beyond the scope of this note, and will be summarized in upcoming notes from the HF TG.

MDC1 exercise in the HF TG represents use of the complete Day-1 analysis chain for the first time: the event generation with Pythia 8.3, simulation and digitization with the full detector description, reconstruction with A Common Tracking Software (ACTS) [4] and KFParticle, and analysis over inclusive HF signal samples. The simulation samples focus on signals for Day-1 measurements (i.e. HF hadron resonances). Specifically, two samples are generated and analyzed in this note as below:

- 50 million $c\bar{c}$ events, which correspond to approximately 0.2 pb^{-1} or four-day data taking [5]. Event statistics of a similar size will also be used for the online production monitoring.
- 50 million $b\bar{b}$ events, corresponding to 30 pb^{-1} and six times the integrated luminosity of the Minimum Bias data in p+p collisions in the first three-year run plan [5].

Both samples are generated using the HF MDC1 tune[1] for the Pythia 8.3 event generator [6, 7] which simulates 200 GeV p+p collisions. The events are produced with single event multiplicity and no pile-up arising from collisions with different bunch crossings. An additional filter is applied to only select events containing at least one charm or bottom quark within the rapidity of $|y| < 1.5$ with no further requirements that the pairs decay to specific final states.

We would like to emphasize that the light flavor background simulation is still being produced, and not yet included in this study. Therefore, this note explores the reconstruction of the HF signals using the KFParticle reconstruction chain, while the background is not fully represented in the figures.

2 KFParticle package

Heavy flavor reconstruction during the MDC1 was performed using the KFParticle package developed for use at CBM [8] and adapted to be used within the Fun4All framework [9][2]. The KFParticle package, as produced for CBM, consists of a set of classes which can store particle information as a 7-element vector with an associated 7×7 covariance matrix. The

[1] https://github.com/sPHENIX-Collaboration/MDC1/blob/17b5f9143e02b31545d10dcb3f235a3eafb8dec9/submit/HF_pp200_signal/rundir/phpythia8_HF_MDC1.cfg

[2] https://github.com/sPHENIX-Collaboration/coresoftware/tree/master/offline/packages/KFParticle_sPHENIX

information stored for each particle is the 3-position and four-momentum,

$$\mathrm{P}(\overrightarrow{x}, \overrightarrow{p}) = (x, y, z, p_x, p_y, p_z, p_E) \tag{1}$$

where, for stable final state particles, p_E represents the particles mass.

Reconstructed vertices are represented by a sub-set of the particle vectors as they have no associated momentum and hence are described by a 3-element vector,

$$\mathrm{V}(\overrightarrow{x}) = (x, y, z) \tag{2}$$

and associated 3×3 covariance matrix. With this information, it is possible to reconstruct secondary vertices and decaying particles using conservation of momentum and energy. By using the covariance matrices and a Kalman filter, the properties of these new vertices and particles can be refined to account for uncertainties arising from reconstruction effects.

In the original KFParticle package, the particle hypothesis for the final state tracks is assigned at the creation of each object due to *a priori* particle identification (PID) information being available in those experiments. At the time of producing the MDC1 samples, there was no such PID information available. Thus, new methods were developed to reduce the amount of combinations that could satisfy the specified kinematics of any decay a user wished to study.

This new framework allows users the flexibility to define ranges or limits for various topological properties of heavy flavor decays such as the invariant mass ranges of parents or resonances, the transverse momentum, p_T, of any particle in the decay chain or the quality of the track or vertex reconstruction, χ^2/nDoF among other variables. A simplified example of a b-hadron decay and its associated topological variables associated which are available to analysers is shown in Figure 1.

Within sPHENIX, KFParticle is used with a modular design philosophy. This means that users can create as many selection objects as they desire to reconstruct different heavy flavor decays. These selection objects can either be chained to create reconstructions of increasing complexity, used to trigger other reconstruction packages (or trigger **by** other reconstruction packages), used in online processing to create standard reconstructed containers or used fully offline to pass files filled with tuples which can then be further refined by users.

Each reconstruction will proceed in a similar manner, dependent on the particles required by the analyst being recognized by the package which checks the user input against a class which contains various particle names, PDG IDs and masses. For each event, as long as there exists at least one primary vertex and track,[3] then the package will find any candidates that satisfy a users selection criteria using the following steps:

1. The reconstructed tracks and primary vertices are unpacked and transformed into KFParticle objects,

2. Tracks are selected by their p_T, $\chi^2(p_T)$, track χ^2/nDoF and minimum impact parameter (IP) χ^2 with respect to **all** primary vertices,

[3] It is possible to have events with no vertices and/or tracks if a users analysis file has used other packages which can write out subsets of the sPHENIX vertex and track maps before KFParticle is called.

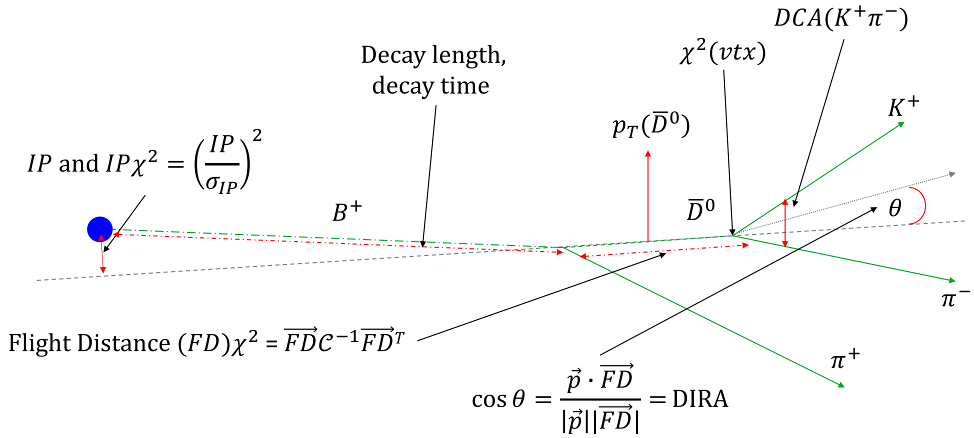

Figure 1: A simplified example of the heavy flavor decay, $B^+ \to \bar{D}^0(\to K^+\pi^-)\pi^+$, emphasizing the topological cuts that can be applied to separate this decay from background candidates. The primary vertex is given in blue, the decay is given in green, measurable parameters are given in red and projections of the decay are given in grey. IP refers to the impact parameter of a particle with respect to the primary vertex, DCA refers to the distance of closest approach between two particles and DIRA stands for the Direction Angle made by a particles momentum and flight direction.

3. n-pronged vertices are constructed from the list of selected tracks based on the track-track distance-of-closest-approach (DCA) and the vertex χ^2,

4. As we have no PID detectors running at sPHENIX, unique PID combinations for the tracks are determined from the users requirements,

5. If intermediate resonances are required, their reconstruction is attempted using all selected tracks which satisfy the vertex requirement and a unique PID assignment for each combination. If the vertex satisfies more than one PID combination then the combination with the lowest invariant mass uncertainty is used as the resonance,

6. All combinations of resonances and tracks that pass the above selections are iterated over to find suitable mother candidates. For each decay vertex, the PID and production vertex (if required) combination that gives the lowest IPχ^2 with respect to all primary vertices (if more than one combination exists) is returned to the user as their candidate. If no production vertex is required by the user, then the candidate with the lowest mass uncertainty is returned.

7. If the user requires intermediate decays, then each intermediate must satisfy flight distance χ^2 and DIRA requirements with respect to the parent particle, and IP and IPχ^2 requirements with respect to all the primary vertices.

Steps 3 to 7 are repeated for every combination of tracks which satisfy all the selection criteria, while Steps 1 to 7 are repeated for every event. The user can specify how to handle the reconstructed candidates, either by appending them back to the sPHENIX node tree and/or writing the candidates into a ROOT TTree. The former option is useful for online event processing, while the latter is commonly used offline by analysts as it gives them immediate access to the candidates.

A typical $D^0 \to K^-\pi^+$ reconstruction object used at sPHENIX would be as in Listing 1.

Listing 1: D^0 example reconstruction

```
KFParticle_sPHENIX *kfparticle =
          new KFParticle_sPHENIX("KFParticleReconstructionObject");

//Track requirements
kfparticle->setMinimumTrackPT(0.1);
kfparticle->setMinimumTrackIPchi2(10);
kfparticle->setMaximumTrackchi2nDOF(2);

//Decay vertex requirements
kfparticle->setMaximumVertexchi2nDOF(2);
kfparticle->setMaximumDaughterDCA(0.3);

//Mother requirements
kfparticle->setMotherName("D0");
kfparticle->setFlightDistancechi2(80);
kfparticle->setMinDIRA(0.8);
kfparticle->setMotherPT(0);
kfparticle->setMinimumMass(1.7);
kfparticle->setMaximumMass(2.0);
kfparticle->setNumberOfTracks(2);
kfparticle->setMotherIPchi2(20);
kfparticle->constrainToPrimaryVertex(true);
kfparticle->getChargeConjugate(true);

//Daughter PID
std::pair<std::string, int> daughterList[99];
daughterList[0] = make_pair("kaon", -1);
daughterList[1] = make_pair("pion", +1);
kfparticle->setDaughters( daughterList );

//Output info
kfparticle->saveOutput(1);
kfparticle->setOutputName("outputData_D02Kpi.root");

se->registerSubsystem(kfparticle);
```

3 $c, b \to D^0 \to K^{\mp}\pi^{\pm}$ Channel

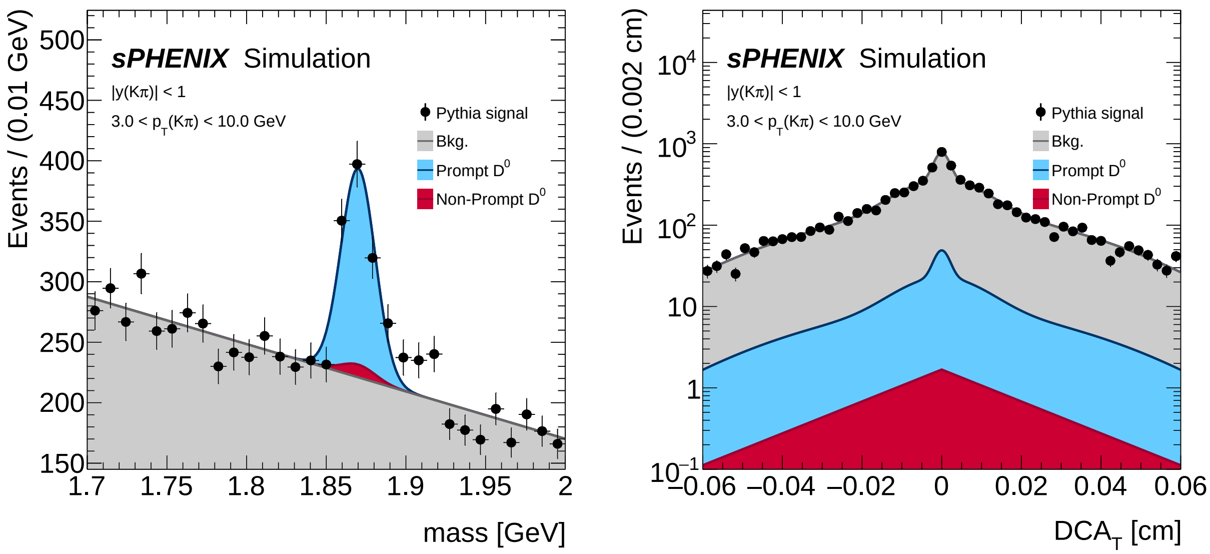

Figure 2: Invariant mass distribution of fully reconstructed D^0 (left) and transverse DCA (right) in the decay channel of $D^0 \to K^-\pi^+$. MDC1 simulation data from Pythia 8.3 is shown for the range $3.0 < p_T < 10.0$ and $|y| < 1.0$. The statistical uncertainty in the simulation data points is denoted as vertical bars. Superimposed on the simulation data are the projections of the fit results.

Figure 2 presents the invariant mass distribution of fully reconstructed D^0 and transverse DCA[4] in the decay channel of $D^0 \to K^-\pi^+$[5] in $c\bar{c}$ (prompt) and $b\bar{b}$ (non-prompt) samples. The invariant mass is fitted with a second-order Chebychev polynomial for background and a Gaussian function for the signal. The same functions are used for prompt and non-prompt components. For the DCA fit, the prompt component is modeled with a triple-Gaussian, and the non-prompt is modeled with a two-sided exponential. The same functions are used for signal and background. The fit is performed with a two-stage unbinned maximum log-likelihood, where we fit the invariant mass first of all by fixing the fraction of signal and secondly the DCA distribution by fixing the b-fraction. The MDC1 $c\bar{c}$ and $b\bar{b}$ signals described in Section 1 are considered in the figure. We consider an additional workflow, which combines two tracks to reconstruct D^0 in the KFParticle framework with very loose selections: track $p_T > 0.5$ GeV, track and vertex $\chi^2/ndf < 4$, the distance between two daughter tracks < 0.03 mm, and the opening angle $< 90°$. No selection at the truth level is applied.

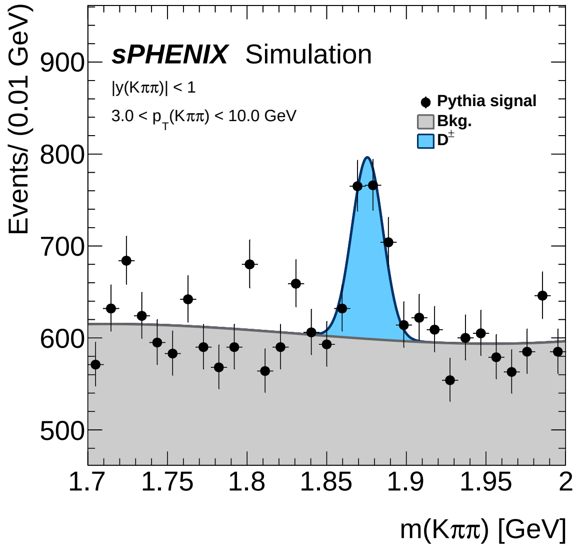

Figure 3: Invariant mass distribution of fully Reconstructed D^+ in the decay channel of $D^+ \rightarrow K^-\pi^+\pi^-$. MDC1 simulation data from Pythia 8.3 is shown for the range $3.0 < p_T < 10.0$ and $|y| < 1.0$. The statistical uncertainty in the simulation data points is denoted as vertical bars. Superimposed on the simulation data are the projections of the fit results.

4 $D^+ \rightarrow K^-\pi^+\pi^+$ Channel

Figure 3 presents the invariant mass distribution of fully reconstructed D^+ in the decay channel of $D^+ \rightarrow K^-\pi^+\pi^-$. The invariant mass is fitted with a second-order Chebychev polynomial for background and a Gaussian function for the signal. The fit is performed with an unbinned maximum log-likelihood. The MDC1 $c\bar{c}$ signals described in Section 1 are considered in the figure. We consider an additional workflow, which combines three tracks to reconstruct D^+ in the KFParticle framework with very loose selections: track $p_T > 0.5$ GeV, track and vertex $\chi^2/ndf < 4$, the distance between two daughter tracks < 0.03 mm, and the open angle $< 90°$. No selection at the truth level is applied.

5 $D^{*+} \rightarrow D^0\pi^+ \rightarrow K^-\pi^+\pi^+$ Channel

Figure 4 presents the difference between the invariant mass distributions (Δm) of fully reconstructed D^{*+} and D^0 in the decay channel of $D^{*+} \rightarrow D^0\pi^+ \rightarrow K^-\pi^+\pi^-$. The invariant mass is fitted with a threshold function $A(\Delta m - m_\pi)^B$ for background and a modified-Gaussian function $e^{-0.5\cdot x^{1+1/(1+0.5\cdot x)}}$ for the signal, where $x = |(\Delta m - m_0)/\sigma|$. The fit is performed with an unbinned maximum log-likelihood with A, B, m_0, and σ as the parameters. The MDC1 $c\bar{c}$ signals described in Section 1 are considered in the figure. We consider an additional workflow, which combines two tracks to reconstruct D^0 adding later a third track to reconstruct D^{*+} in the KFParticle framework with very loose selections: track $p_T > 0.5$ GeV,

[4] Here, the transverse DCA of the D^0 is measured with respect to the primary vertex.

[5] Unless otherwise stated, charge-conjugation is implied throughout.

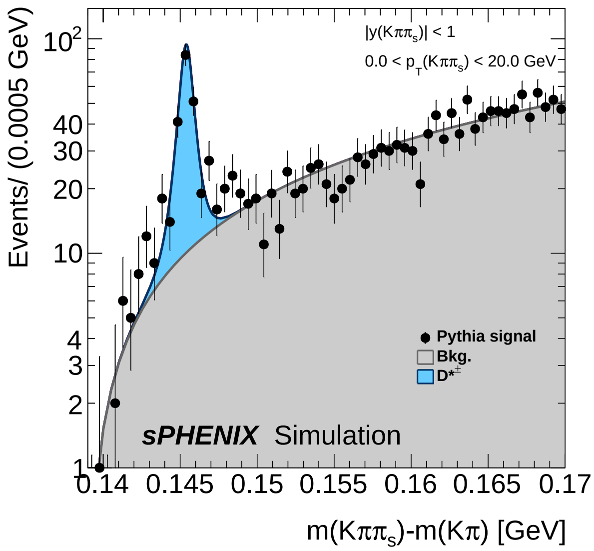

Figure 4: Difference of the invariant mass distribution of fully reconstructed D^{*+} and D^0 in the decay channel of $D^{*+} \to D^0\pi^+ \to K^-\pi^+\pi^-$. MDC1 simulation data from Pythia 8.3 is shown for the range $0.0 < p_T < 20.0$ and $|y| < 1.0$. The statistical uncertainty in the simulation data points is denoted as vertical bars. Superimposed on the simulation data are the projections of the fit results.

track and vertex $\chi^2/ndf < 4$, the distance between D^0 two daughter tracks < 0.03 mm, and the opening angle $< 90°$. The invariant mass of the D^0 candidates is allowed to be between 1.7-2.0 GeV. No selection at the truth level is applied.

6 $D^+_{(s)} \to K^+K^-\pi^+$ and $D^+_s \to \phi\pi^+ \to K^+K^-\pi^+$ Channels

Figure 5 presents the invariant mass distribution of fully reconstructed D^+_s in the decay channels of $D^+_s \to K^+K^-\pi^+$ (left) and $D^+_s \to \phi\pi^+ \to K^+K^-\pi^+$ (right), both fitted with a linear function for background and a Gaussian function for the D-meson signal. The MDC1 $c\bar{c}$ signals described in Section 1 are considered in the figure. We consider an additional workflow, which combines three tracks to reconstruct D^+_s in the KFParticle framework with very loose selections: track $p_T > 1$ GeV, track and vertex $\chi^2/ndf < 40$, the minimum distance between two daughter tracks < 0.1 mm, and the open angle $< 90°$. For this exploratory study, we apply further selections to suppress background using the truth information of the D^+_s candidates by requiring the truth track PDG to match the track PDG ID used in reconstruction, as well as the truth K track $p_T > 0.6$ GeV and truth π track $p_T > 1.1$ GeV. Moreover, for the $D^+_s \to K^+K^-\pi^+$ channel, we require the three $KK\pi$ truth tracks to come from the same truth vertex. For the $D^+_s \to \phi\pi^+ \to K^+K^-\pi^+$ channel, we restrict the truth KK track invariant mass to be 42 MeV within mass peak of the $\phi(1020)$ meson. According to our fit results in Figure 5, the estimated number of D^+_s signal candidates in the $D^+_s \to K^+K^-\pi^+$ channel is 1974 ± 184 (stat.) and in the $D^+_s \to \phi\pi^+ \to K^+K^-\pi^+$ channel is 1182 ± 145 (stat.).

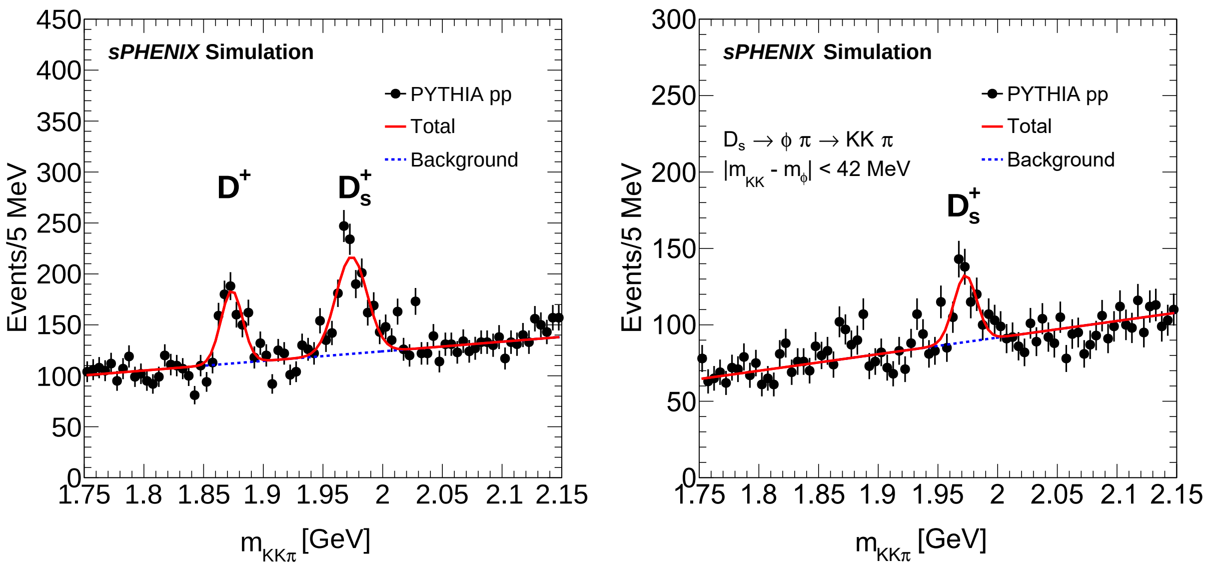

Figure 5: Invariant mass distribution of fully reconstructed D_s^+ in the decay channels of $D_s^+ \to K^+K^-\pi^+$ (left) and $D_s^+ \to \phi\pi^+ \to K^+K^-\pi^+$ (right) fitted with a linear function for background and a Gaussian function for D-meson signal. The MDC1 $c\bar{c}$ signals generated with Pythia 8.3 are considered. The statistical uncertainty in the simulation data points is denoted as vertical bars.

References

[1] C Mironov. First Mock Data Challenge. Presentation at third Software and Computing Review in March 2021, https://indico.bnl.gov/event/11083/contributions/46904/attachments/33233/53346/SCReview_Mar21_Mironov_final.pdf.

[2] Sourav Tarafdar. $\Lambda_c \rightarrow \pi^+ + K^- + p$ from sPHENIX MDC-1 charm and bottom samples. Presentation at Heavy Flavor Topical Group Meeting, March 8, 2021, https://indico.bnl.gov/event/10963/contributions/46251/attachments/33017/52869/HF_update_03_08_21.pdf.

[3] Han-Sheng Li. Prompt-D^0 and non-prompt D^0 DCA distribution. Presentation at Heavy Flavor Topical Group Meeting, March 22, 2021, https://indico.bnl.gov/event/11063/contributions/46920/attachments/33252/53387/HanSheng0322.pdf.

[4] Acts common tracking software. https://acts.readthedocs.io/en/latest/.

[5] sPHENIX Collaboration. sPH-TRG-2020-001: sPHENIX Beam User Proposal 2020. 2017. URL: https://indico.bnl.gov/event/9301/.

[6] Sanghoon Lim, Woohyeong Park. Pythia8 tune in pp 200 gev. Presentation at Heavy Flavor Topical Group Meeting, Dec. 14, 2020, https://indico.bnl.gov/event/10309/contributions/44139/attachments/31909/50542/sPHENIX_HF_shlim_20201215.pdf.

[7] Long Ma, Hideki Okawa. PYTHIA tuning for the p+p 200 GeV simulation. Presentation at Heavy Flavor Topical Group Meeting, Dec. 14, 2020, https://indico.bnl.gov/event/10309/contributions/44140/attachments/31916/50554/HF_group_meeting_20201215.pdf.

[8] Sacha Gorbunov and Ivan Kisel. Reconstruction of decayed particles based on the Kalman filter. *CBM-SOFT-note-2007-003*, May 2007.

[9] Cameron Dean. Event Reconstruction and MDC1. Presentation at sPHENIX 10^{th} Collaboration Meeting, Jan. 22, 2021, https://indico.bnl.gov/event/10568/contributions/45092/attachments/32421/51581/HeavyFlavor_EventRecoAndMDC_C_Dean_20210122.pdf.