Dealing with High Background Rates in the STAR Heavy Flavor Tracker in Simulation

Embedding Simulation into Real Events

Jason C. Webb[1,*], *Jérôme* Lauret[2], *Gene* Van Buren[1], *Sooraj* Radhakrishnan[2], and *Xin* Dong[2]

[1]Brookhaven National Laboratory
[2]Lawrence Berkeley National Laboratory

Abstract. The STAR Heavy Flavor Tracker (HFT) has enabled a rich physics program, providing important insights into heavy quark behavior in heavy ion collisions. Acquiring data during the 2014 through 2016 runs at the Relativistic Heavy Ion Collider (RHIC), the HFT consisted of four layers of precision silicon sensors. Used in concert with the Time Projection Chamber (TPC), the HFT enables the reconstruction and topological identification of tracks arising from charmed hadron decays. The ultimate understanding of the detector efficiency and resolution demands large quantities of high quality simulations, accounting for the precise alignment of sensors, and the detailed response of the detectors and electronics to the incident tracks. The background environment presented additional challenges, as simulating the significant rates from pileup events accumulated during the long integration times of the tracking detectors could have quickly exceeded the available computational resources, and the relative contributions from different sources was unknown. STAR has long addressed these issues by embedding simulations into background events directly sampled during data taking at the experiment. This technique has the advantage of providing a completely realistic picture of the dynamic background environment while introducing minimal additional computational overhead compared to simulation of the primary collision alone, thus scaling to any luminosity. We will discuss how STAR has applied this technique to the simulation of the HFT, and will show how the careful consideration of misalignment of precision detectors and calibration uncertainties results in the detailed reproduction of basic observables, such as track projection to the primary vertex. We will further summarize the experience and lessons learned in applying these techniques to heavy-flavor simulations and discuss recent results.

1 Introduction

Monte-Carlo simulations are central to high-energy and nuclear physics experiments, providing the detailed modeling of detector response, signal production and underlying background distributions necessary to produce physics results comparable with theory. Backgrounds pose particular challenges for simulations of experiments, especially when interaction rates exceed detector integration times, and events "pile-up" during read-out. These pile-up events require

---

*e-mail: jwebb@bnl.gov

additional CPU resources and disk space, as they must be simulated on par with the collisions of interest. Detector noise, cavern backgrounds, beam-gas interactions and (in heavy-ion collisions) low-energy particles from ultra-peripheral collisions (UPC) create additional complications, requiring additional effort to tune models and understand relative yields for input into the simulation chain.

The STAR experiment[1] at the Relativistic Heavy Ion Collider (RHIC) has long addressed these challenges by embedding simulations into appropriate background events measured *in-situ* during data taking. This approach conveys several advantages. First, it obviates the need to simulate pileup events and other sources of backgrounds, as they have been directly sampled. Second, there is no need to spend time modeling the backgrounds and understanding their relative contributions in detail. And third, any time dependence in the backgrounds as luminosities change throughout the fill will be properly sampled.

In this paper we will discuss the application of the embedding technique within the context of STAR's heavy flavor physics program. The importance of not only the detector misalignments of precision trackers on hit positions, but also the uncertainties in misalignments and data corrections, will be discussed. We will show how careful treatment of these effects in embedding simulations provide excellent agreement in observables, such as the distance-of-closest approach to the vertex and the relative efficiencies of the trackers, which give confidence that the extracted efficiencies from simulation are correct.

2 The Heavy Flavor Program

Charmed hadron production provides important insights into the hot, dense medium created in heavy ion collisions. As they are created early in the collision, they experience the full evolution of the system. Reconstruction of charmed-hadron decays in heavy ion collisions requires topological reconstruction to reduce combinatoric backgrounds: high-precision tracking detectors must be employed to identify decay products with vertices displaced from the primary interaction. In table 1 we list some particles and decay modes of interest. Typical displacements are on the order of $100\,\mu m$. This informs one of the requirements of the physics program to identify charged kaons at $p_T \geq 750\,MeV$ separated from the primary vertex by $50\,\mu m$.

Particle	Decay Channel	Branching Ratio	$c\tau\,\mu m$	Mass GeV/c^2
D^0	$K^-\pi^+$	3.8%	123	1.8645
D^+	$K^-\pi^+\pi^+$	9.5%	312	1.8694
D_s^+	$K^+K^-\pi^+$	5.2%	150	1.9683
	$\pi^+\pi^+\pi^-$	1.2%		
Λ_C^+	$pK^-\pi^+$	5.0%	59.9	2.2865

Table 1: A few charmed hadron decay channels accessible with the HFT, with branching ratios, mass and mean decay length.

To achieve the required pointing resolution, STAR utilizes four detector technologies, summarized in table 2, to identify track candidates and provide increasingly precise position constraints during a combinatorial kalman fit [2]. The main tracking detector in STAR is a large acceptance time projection chamber (TPC) [3]. It is responsible for track identification, providing *mm*-scale projection uncertainty at the inner detectors. Two layers of silicon strip detectors follow – the silicon strip tracker (SST) [4] and the intermediate silicon tracker (IST). These detectors further improve the pointing resolution, allowing for a reasonable search window for hits in the two layers of the high precision $20\mu m \times 20\mu m$ silicon pixel detector

(PXL) [5] located closest to the beam. These four inner layers are collectively referred to as the heavy flavor tracker (HFT) [6].

Subsystem	Hit Distance from Interaction Point	Pointing resolution to next layer	Integration Time
TPC	$60 < r < 200\ cm$	$1\ mm$	$40\ \mu s$
SST	$23\ cm$	$300\ \mu m$	fast
IST	$14\ cm$	$250\ \mu m$	fast
PXL	$8\ cm$ $2.8\ cm$	$< 30\ \mu m$	$185.6\ \mu s$

Table 2: Detector parameters

As noted in table 2, the TPC and PXL detectors integration times are significant compared to the ~50 kHz interaction rate typical during the 2014 and 2016 AuAu runs.[1] At these luminosities, we expect ~2 extra minimum bias collisions in the TPC and ~10 extra collisions in the PXL detector. The SST and IST are both fully read out between bunch crossings, which helps to mitigate pile-up. The proximity of the first pixel layer to the beamline brings an additional source of background hits in heavy-ion runs. Ultra-peripheral collisions between the gold nuclei produce a flux of low-energy electrons ($\sim 70\ MeV$), contributing hits at a rate comparable to those from pile-up tracks. These background hits pose two concerns for tracking, which simulation must ultimately address. First, real tracks in the TPC might pickup the wrong hits in the HFT, leading to inefficiencies. And second, pileup tracks in the TPC might pickup accidental hits in the HFT, giving rise to unwanted background tracks.

3 Simulation Strategies

The role of simulation in heavy-flavor analyses is to provide reliable estimates of the inefficiencies in single track reconstruction and the association of tracks to a displaced vertex. As discussed above, understanding the detector performance for tracks with $p_T \geq 750\ MeV$ is a requirement imposed on simulation by the physics program. We will discuss two approaches to the problem: (1) pure Monte Carlo, in which all aspects of the events including primary event generation, detector noise and backgrounds must be modeled and simulated; and (2) embedding, in which primary events are simulated and embedded within real events representative of the detector environment sampled during data taking by STAR.

3.1 Pure Monte Carlo

In pure Monte Carlo, we try to account for all aspects of experimental environment using simulations. Hijing [7] is utilized to simulate the primary event, providing the set of particles and their kinematics as input to a GEANT [8] simulation of the STAR detector. In addition to the primary event, additional minimum bias events are simulated and distributed within the time integration windows of the TPC and pixel detectors, accounting for pile-up. Energy deposits in the active elements of detectors are digitized, adding in the effects of detector noise at this stage. Finally, the effects of UPC electrons are modeled by adding in additional random hits to the first layer of the pixel detector. While the rate of pile-up events can be calculated directly from the instantaneous luminosity of the data, the relative contribution of UPC electrons cannot be so easily determined. Therefore, the rate of UPC hits was tuned to match observables in the Monte Carlo to the data.

---

[1]During polarized *pp* runs, this can be as much as 200x greater.

It is crucial to compare simulations to data to demonstrate that the tracking and vertex association efficiencies can be correctly determined. Figure 1 shows a comparison between the 2014 AuAu data (red points) and simulations (black histograms) for two observables which are sensitive to those quantities. The left panel shows the HFT matching ratio. This is the number of tracks reconstructed in the event with HFT hits divided by the total number of tracks found in the event by the TPC. In the absence of pile-up and UPC backgrounds (blue points) it indicates the relative efficiencies of HFT+TPC tracking to the TPC alone. The right panel shows the distribution of the 2d distance-of-closest-approach (DCA_{xy}) to the vertex, indicating how well we can simulate pointing back to the vertex.

At low p_T both data and simulation over-predict the performance of the HFT to TPC tracking. This arises from TPC pile-up tracks picking up random hits in the HFT. These excess HFT tracks produce the broad tails seen in the DCA_{xy} distributions. While agreement between data and Monte Carlo is good, it was achieved by tuning the relative background contributions in the first pixel layer. The physics program depends on the region where the impact of these backgrounds are becoming large, motivating the need to utilize *in-situ* measurements of the backgrounds directly.
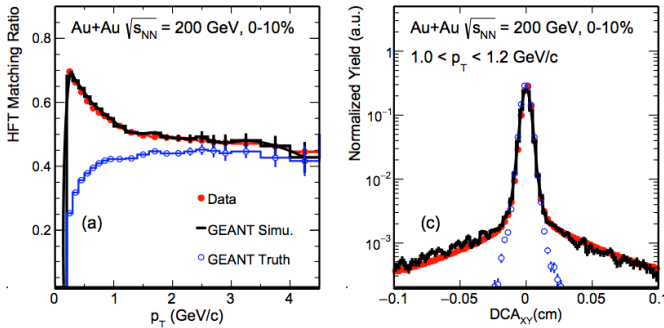

Figure 1. Comparison of single pion production in 200 GeV AuAu collisions from 2014. Left panel shows the relative tracking efficiency of the HFT, defined in the text. Right panel shows the DCA_{xy} of the tracks to the reconstructed vertex.

3.2 Embedding Simulation

As with pure Monte Carlo, embedding utilizes an event generator to simulate the particles produced in the primary interaction, which are then passed through a GEANT simulation of the detector, and energy depositions are digitized. The detector noise and background environment are not simulated, but rather come from dedicated samples accumulated in parallel during real data taking at the experiment. The digitized signals from the simulation are then merged run-by-run and channel-by-channel with these experimentally sampled backgrounds. This ensures that all backgrounds are accounted for in their correct proportions, time-dependent background features during data taking are properly treated, and reduces the computational overhead to simulate the background. This comes at the cost of having to dedicate trigger bandwidth to non-zero-suppressed background samples, and requires careful treatment of detector misalignments at the simulation stage[10].

4 Integrating the HFT into Embedding

Initial embedding studies did not yield satisfactory agreement in either the HFT matching ratio or the DCA_{xy} distributions. The matching ratio diverged at high-p_T in embedding, over predicting the efficiency of HFT tracking. The central peak and tails of the DCA_{xy} distributions were both too narrow, indicating that vertex association was too good in simulation. Hit

digitization procedures for both the HFT and TPC are tightly constrained by cosmic ray measurements, and cannot explain the discrepancy. The embedding techniques that work with a single tracker, and worked with the previous generation of silicon trackers in STAR, were insufficient for the precision provided by the HFT. What was missing was a proper treatment of the *uncertainties* in the calibrations and misalignments which are used to determine the hits positions. These uncertainties can be neglected in simulations of a single detector, but become important in correlating tracks across detector subsystems, especially when the track propagation uncertainty becomes comparable to the sensor size in the experiment.

To treat the uncertainties in the misalignments, an additional smearing of the hit positions was introduced into the pixel simulations. Figure 2 (left) shows the width of the central gaussian in the DCA_{xy} distributions for charged pions in 200 GeV AuAu collisions plotted as a function of p_T, compared with embedded pions. Three different levels of smearing are applied to account for the uncertainties in the misalignments of the HFT. The right panel in figure shows the width of the simulated DCA_{xy} distributions normalized to the data. While there was no perfect match to be found, the best value is 8 μm, which was compatible with the uncertainties estimated in the HFT alignment procedure.
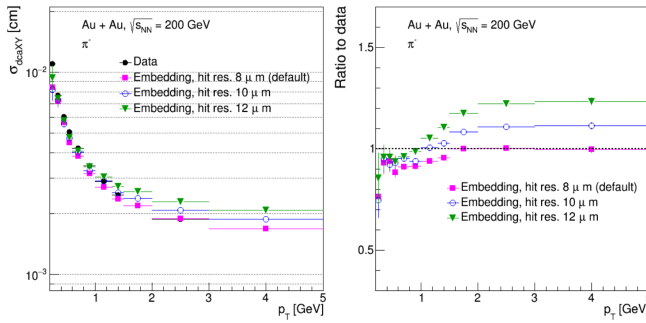

Figure 2. Left panel shows the width of the central gaussian in the DCA_{xy} distribution for charged pions detected at STAR (solid black points) and for simulations embedded into appropriate background events described in text. Right panel shows the ratio of the DCA_{xy} widths from the embedded simulations to the real data.

While the DCA_{xy} distributions could be reasonably tuned to the data by accounting for misalignment uncertainties in the pixel detector, the HFT matching ratio was essentially insensitive to them. Figure 3 shows this ratio as a function of p_T for real data in 200 GeV AuAu events accumulated at STAR, and compares it to embedding simulations. The blue points show the (now default) 8 μm hit smearing established above. Smearing of 12 μm (open magenta circles) had virtually no impact on the matching ratio, and larger values were not compatible with the alignment studies. Single hit efficiencies were the next option, though these were tightly constrained by cosmic ray data to be ~98%. Degrading this to ~95% is shown by the open red circles. Not surprisingly, the impact is seen uniformly across the p_T range, and the data and embedding simulation begins to diverge unacceptably at low p_T.

Having addressed the uncertainties in the HFT, we turned our attention to the TPC. The detailed simulation of the TPC response to the passage of charged particles requires careful accounting of effects which deflect the ionization electrons as they drift to the readout planes. Ionic charge accumulation during RHIC fills from both the primary ionization of the TPC volume and leakage into that volume from the high-gain region deflect the drifting of the ionization electrons both radially and azimuthally[9]. Accounting for these distortions is critical during track reconstruction, but the measurements of these effects come with their own deficiencies and subsequent uncertainties. Fluctuations of the TPC distortions were added to simulation at various levels to account for the impact of these uncertainties, and the best agreement with data was found around ~5%. Figure 4 shows the matching ratio as a
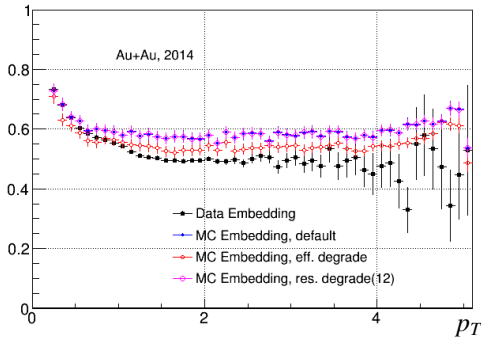
Figure 3. HFT matching ratio in embedding compared with 2014 AuAu data (black points). Comparisons are made to embedding when the hit resolutions and efficiencies are varied.

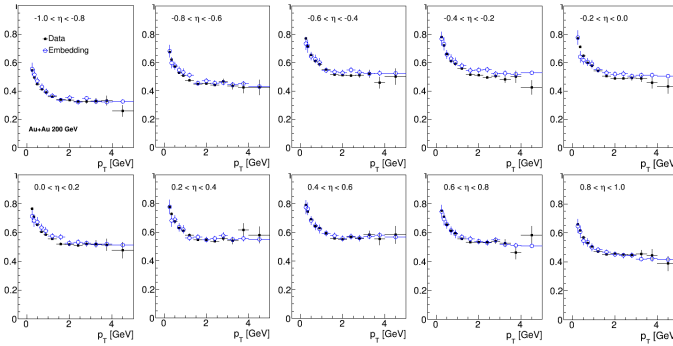function of p_T and pseudorapidity, and the agreement between data and embedding is quite good.

Figure 4. Relative efficiency of the heavy flavor tracker binned in pseudorapidity and p_T. Solid black points denote the real data, open blue circles Hijing simulations (incorporating uncertainties described in text) embedded into unbiased triggers (randomly selected bunch crossings) which capture the background.

As a final validation of the method, simulated 200 GeV AuAu Hijing events, with one or more D_0 particles in the final state, were embedded into zero-bias triggered data accumulated at STAR. Figure 5 compares the distribution of the decay parameters of the identified D_0 candidates. The level of agreement between data and simulation is quite good, and confirms that the embedding simulations are accurately modeling the signals and backgrounds in the data.

5 Conclusion

The STAR Heavy Flavor tracker provided the precision tracking required to topologically identify charmed hadron decays, enabling important measurements which probe the evolution of the dense medium created in heavy ion collisions. Embedding simulated events into real data allows us to provide a precise modeling of the large and complicated background environment by measuring backgrounds *in situ* with data taking. By carefully considering the misalignments of the HFT components, the corrections for distortions in the time projection chamber, and accounting for the uncertainties in these quantities during simulation, we were able to provide an excellent match between the observables in the data which are most sensitive to the reconstruction efficiencies for heavy flavor decay products, and the topological reconstruction of the decays themselves. The agreement between data and embedded simulations gives great confidence that we are properly measuring the efficiencies of our detectors, and has enabled us to address even the most challenging of the heavy flavor physics channels.
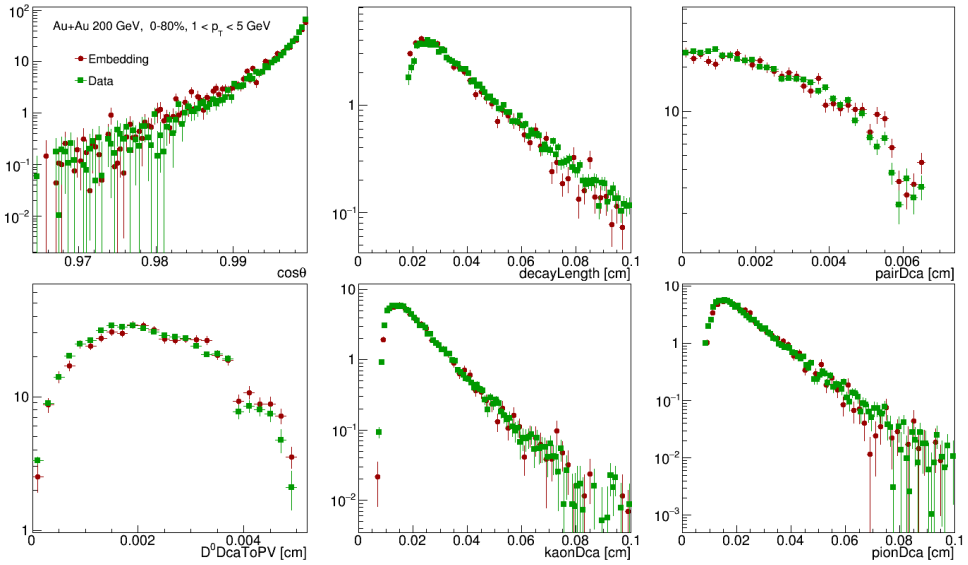
Figure 5: Topological decay parameters for D_0 decays. Green points denote D_0 candidates reconstructed in real data, red points those reconstructed in simulated events embedded into background data.

6 Acknowledgments

References

[1] K.H. Ackermann, *et al.*, Nucl. Instrum. and Meth. **A499**, 624-632 (2003)

[2] J. C. Webb *et al.*, *STAR Reconstruction Improvements for Tracking with the Heavy Flavor Tracker*, J. Phys.: Conf. Ser. **898**, 042035 (2017)

[3] M. Anderson *et al.*, Nucl. Instr. and Meth. A **499**, 699 (2003)

[4] L. Arnold, *et al.*, Nucl. Instr. and Meth. A **499**, 652 (2003)

[5] L. Greiner, *et al.*, Nucl. Instr. and Meth. A **650** 68 (2011)

[6] H. Qiu, *et al.*, Nucl. Phys. A **931**, 1141-1146 (2014)

[7] X. Wang, M. Gyulassy, Comput. Phys. Commun. **83**, 307 (1994)

[8] R. Brun, F. Bruyant, M. Maire, A.C. McPherson and P. Zanarini, *GEANT3* CERN-DD-EE-84-1

[9] G. Van Buren *et al.*, Nucl. Instr. and Meth. **A566**, 22-25 (2006)

[10] J. C. Webb *et al.*, *Representing Misalignments of the STAR Geometry Model using AgML*, J. Phys.: Conf. Ser. **898**, 042036 (2017)