Applying the KF Particle Method to Strange and Open Charm Hadron Reconstruction in the STAR Experiment

Xinyue Ju,[1,2] Yue-Hang Leung,[2] Sooraj Radhakrishnann,[3,2] Petr Chaloupka1,[4] Xin Dong,[2] Yury Fisyak,[5] Pavol Federic,[6] Ivan Kisel,[7,8,9] Hongwei Ke,[5] Michal Kocan,[4] Spyridon Margetis,[3] Aihong Tang,[5] Iouri Vassiliev,[9] Yifei Zhang,[1] Xianglei Zhu,[10] and Maksym Zyzak[9]

[1]*University of Science and Technology of China, Hefei, Anhui Province 230026, China*
[2]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[3]*Kent State University, Kent, OH 44242, USA*
[4]*Czech Technical University in Prague, Prague, Czech Republic*
[5]*Brookhaven National Laboratory, Upton, NY 11973, USA*
[6]*Nuclear Physics Institute of the Czech Academy of Sciences, Prague, Czech Republic*
[7]*Goethe-Universität Frankfurt, Frankfurt am Main, Germany*
[8]*Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany*
[9]*GSI Helmholtzzentrum für Schwerionenforschung GmbH, Darmstadt, Germany*
[10]*Tsinghua University, Beijing, China*

We apply KF Particle, a Kalman Filter package for secondary vertex finding and fitting, to strange and open charm hadron reconstruction in heavy-ion collisions in the STAR experiment. Compared to the conventional helix swimming method used in STAR, the KF Particle method improves the reconstructed Λ, Ω and D^0 significance considerably. At the same time, we demonstrate that Monte Carlo simulation with the STAR detector responses can well reproduce the topological variable distributions reconstructed in real data using the KF Particle method, therefore retaining good control on the reconstruction efficiency uncertainties for strange and open charm hadrons measurements in heavy-ion collisions.

Keywords: Heavy-iron collisions, secondary vertex finding, Kalman Filter

I. INTRODUCTION

In high-energy particle and nuclear physics experiments, strange and heavy flavor hadrons have unique roles in studying the electroweak and strong interactions in the Standard Model [1–3]. These particles are mostly short-lived particles, and their ground state particles, such as K^0_S, Λ, D^0, and Λ_c^+, have a proper lifetime ($c\tau$) varying from tens of micrometers to several centimeters [4]. Experimentally reconstructing their decay positions and separating them from collision vertices would be necessary to achieve precision measurements [5, 6]. This becomes extremely critical in high energy heavy-ion experiments at RHIC and the LHC where thousands of particles are produced from the collision vertex. Secondary vertex reconstruction can significantly reduce the combinatorial background in these collisions while at the same time it also involves a finite reconstruction efficiency, especially for low momentum particles [5, 6]. Therefore one would need to consider a balance between the combinatorial background and the reconstruction efficiency for the particle of interest to achieve the best experimental measurement precision.

The STAR detector at RHIC is a general purpose detector dedicated for heavy-ion experiments [7]. The main tracking subsystem, the Time Projection Chamber (TPC) [8], provides a pointing resolution of $\sim$mm to the collision vertex for charged tracks which allows the topological separation of strange hadron weak decay positions from the primary collision point. A high resolution silicon detector, the Heavy Flavor Tracker (HFT), operated in 2014-2016, improves the charged track pointing resolution to be better than ~ 50 μm for 750 MeV/c charged kaon tracks [9]. This enables the topological reconstruction of various open charm hadron decays in heavy-ion collisions [5, 10].

Traditionally, the secondary vertex reconstruction in STAR has been conducted by searching for the closest distance of approach (DCA) points of two charged track helices, called the helix swimming method (HS). Then the decay position is taken to be the middle of the two DCA points. This method has shown good performance in reconstructing strange and open charm hadrons in heavy-ion collisions [5, 6]. Figure 1 shows a sketch of key topological variables used in this method: DCA of daughters particles to the primary vertex, DCA between two daughter particles, decay length from the decay vertex position to the primary vertex, θ - the angle between the interested particle momentum vector and the decay length vector, and/or the DCA between the interested particle helix and the primary vertex. The calculations are conducted based on the mathematical helix model for daughter tracks. No experimental estimated uncertainties are included in this reconstruction method.

Recently, within STAR, an experimentally estimated error matrix on track helix fitted parameters has been made available in offline analysis software infrastructure. At the same time, the KF Particle package, a Kalman Filter method for secondary vertex finding and fitting utilizing the estimated track helix error matrices, has been deployed for STAR offline analysis. The goal is to improve the secondary particle reconstruction with constraints provided by the additional knowledge on the error matrices of various topological variables.

This paper reports the result of applying the KF Particle method to the reconstruction of strange (Λ,Ω^-) and open charm (D^0) hadrons in heavy-ion collisions at the STAR

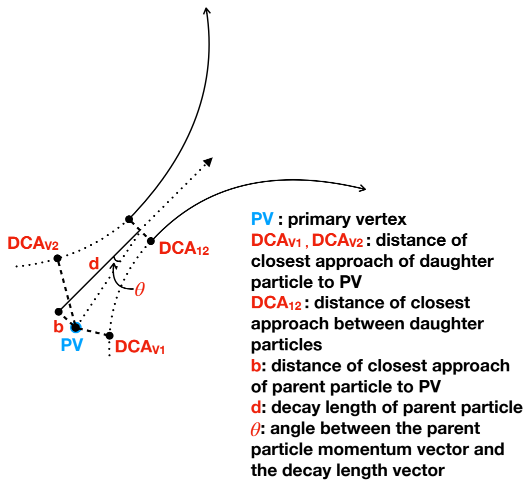

Fig. 1. Sketch of key topological variables used by the helix swimming method.

experiment. A Toolkit for Multi-Variate Analysis (TMVA) package deployed in ROOT [11] has been used to optimize topological selection cuts for the best signal significance in both helix swimming and KF Particle methods. The paper is organized as follows: Sec. II describes how the KF Particle method handles secondary particle reconstruction and fitting. The application of the KF Particle method to the STAR data are discussed in Sec. III. We compare the optimized signal performance from the helix swimming method and the KF Particle method. We also compare various topological variable distributions from the KF Particle method obtained in real data and Monte Carlo simulations. Finally, we summarize our findings in Sec. IV.

II. KF PARTICLE METHOD

The Kalman Filter (KF) [12] is a recursive method for the analysis of linear discrete dynamic systems described by a vector of parameters, which is called the state vector $\boldsymbol{r}$, according to a series of measurements observed over time. It produces estimates of unknown vector parameters with high accuracy and is widely used in tracking and data prediction tasks.

In particle experiments, the Kalman filter can be used to solve different tasks, such as track finding, particle reconstruction, and event vertex reconstruction [13]. In particular, the KF Particle package, which utilizes the Kalman filter for the reconstruction of short-lived particles and vertex finding, has been developed and is now applied to data analysis in STAR.

In the KF Particle framework, each particle is described by a state vector with eight parameters [14]: $\boldsymbol{r} = (x, y, z, p_x, p_y, p_z, E, s)$, where (x, y, z) is the position of the particle, (p_x, p_y, p_z) is the momentum, E is the energy of the particle, and $s = l/p$, with l being the length of the trajectory in the laboratory coordinate system and p the particle total momentum. This natural particle parametrization makes the algorithm independent on the geometry of the detector system. The reconstructed state vector and its covariance matrix ($\boldsymbol{C}$) contain all necessary information about the particle, which allows one to handily calculate physical quantities such as its momentum, energy, and lifetime with their accuracy, and also the χ^2 values during the reconstruction, i.e. estimate the quality of reconstruction.

To simplify the calculation, the momentum and energy of the mother particle are calculated from the sum of all daughter particles and only the vertex position is fitted. After transporting a daughter particle to the current estimation of the decay vector $(r_k, \boldsymbol{C}_k)$, the state vector of this daughter particle can be taken as a measurement $(m_k, \boldsymbol{V}_k)$ of the mother particle's state vector. Using the residual ζ_k between r_k and m_k and the Kalman gain matrix $\boldsymbol{K}_k$ calculated from the $\boldsymbol{C}_k$ and $\boldsymbol{V}_k$, the estimation of mother particle's vector can be updated $(r_{k+1}, \boldsymbol{C}_{k+1})$ according to the formula 1.

$$\zeta_k = r_k - m_k,\ r_{k+1} = r_k + \boldsymbol{K}_k\zeta_k,\ \boldsymbol{C}_{k+1} = \boldsymbol{C}_k - \boldsymbol{K}_k\boldsymbol{C}_k^{'} \tag{1}$$

The χ^2-criterion of this estimation can be obtained at the same time. By conducting this process on all daughter tracks, a basic filtering algorithm is formed. A full description of the algorithm and the mathematical justification can be found in Ref. [14, 15]. Here we briefly outline the scheme of the short-lived particle reconstruction, also shown in Fig. 2:

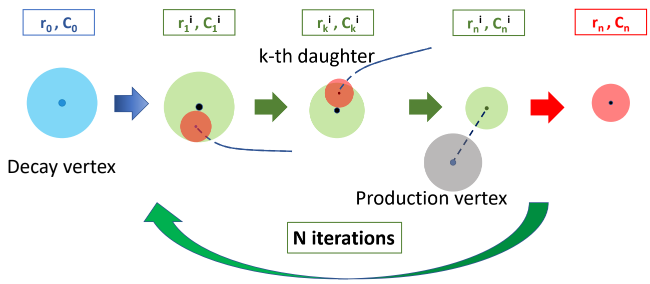

Fig. 2. Basic diagram of short-lived particle reconstruction with the KF Particle package.

1. Sort the final state particles into primary and secondary according to its χ^2 to collision vertex.
2. Choose an initial secondary decay point, often to be the DCA point to the collision vertex from the first daughter track. Set the mother particle initial parameters ($\boldsymbol{r}_0$, $\boldsymbol{C}_0$), $\boldsymbol{C}_0$ is often set as an infinite diagonal matrix.
3. Extrapolate the k-th daughter particle to the point of the closest approach with the current estimation of the decay point and update its parameters
4. Correction of the decay vertex according to k-th daughter particle and adding the 4-momentum of the daughter particle to the 4-momentum of the mother particle.

5. Loop over all n daughter particles and calculate an optimum estimation of the decay vector and its covariance matrix ($\boldsymbol{r_n^i}$, $\boldsymbol{C_n^i}$) and the χ^2 probabilities.

6. If the production vertex of the mother particle (usually the primary vertex) is known, transport the mother particle to it, then filter with the production vertex's position and calculate the χ^2 probabilities of the origination from the production vertex.

7. Set $\boldsymbol{r_n^i}$ and $\boldsymbol{C_n^i}$ as the mother particle's initial parameters and repeat steps 3-6 N times.

8. Finalize the precision of the mother particle parameters ($\boldsymbol{r_n}$, $\boldsymbol{C_n}$).

Compared to the traditional helix swimming method, the KF Particle method enjoys several important advantages:

- Usage of the daughter particle track parameters covariance matrices adds information about the detector performance and the track reconstruction quality that improves the mother particle reconstruction accuracy and efficiency.
- Statistical criteria are calculated and used for background rejection.
- The natural and simple interface allows to the reconstruction of complicated decay chains [15].
- Usage of parallel programming provides high computing speed for the above rather complicated calculations.

III. APPLICATION TO DATA

We apply the KF Particle method to the reconstruction of strange (Λ, Ω^-) and open charm (D^0) hadrons, using data collected by the STAR experiment. Recent experimental datasets of Au+Au collisions at $\sqrt{s_{\rm NN}}$ = 27 GeV (for Λ, Ω^-) and 200 GeV (for D^0), which contain the error matrices information of tracks parameters were used in this analysis.

A. Λ reconstruction

Λ particles are reconstructed via the decay channel $\Lambda \to p + \pi^-$, which has a branching ratio of 69.2% [4]. Λ particles decay with a proper decay length of c$\tau \simeq 79$ mm after they are produced in Au+Au collisions. Protons and pions are identified by ionization energy loss in the TPC gas. Practically, charged tracks with $|n\sigma_{\rm X}| < 3$ for any interested particle X are selected, where $n\sigma_{\rm X}$ is defined by the following:

$$n\sigma_{\rm X} = \frac{1}{\sigma_{\rm X}} \log \frac{\langle {\rm dE/dx} \rangle_{\rm measured}}{\langle {\rm dE/dx} \rangle_{\rm X}^{\rm Bischel}}, \quad (2)$$

where $\langle dE/dx \rangle_{\rm measured}$ is the average energy loss per unit length, measured by the time projection chamber (TPC) of the STAR detector; $\langle dE/dx \rangle_{\rm X}^{\rm Bischel}$ is the expected energy loss $\langle dE/dx \rangle$ for a certain particle species X (in this case, proton or pion), and $\sigma_{\rm particle}$ is the $\langle dE/dx \rangle$ resolution measured by the TPC (typically $\simeq 8\%$ [8]). For each proton or pion track, we require a minimum of 15 hits in the TPC to ensure good track quality.

Using data collected by the STAR experiment from Au+Au collisions at $\sqrt{s_{\rm NN}} = 27$ GeV, Λ particles are reconstructed using the KF Particle method, and various kinematic and topological variables such as mass, p_T, decay length, etc. are calculated. As shown in Fig. 3, clear Λ mass peaks are seen in the invariant mass $m_{p\pi^-}$ distributions in the p_T range of 0.4 to 6 GeV/c.

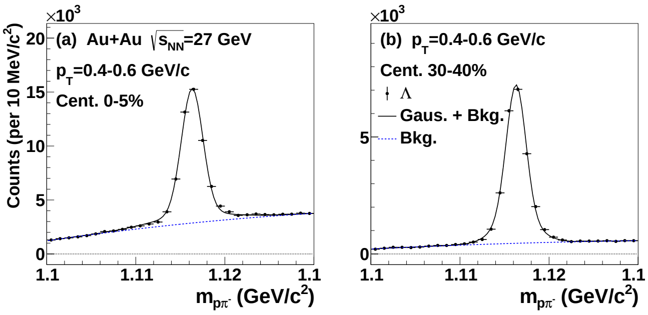

Fig. 3. $p\pi$ invariant mass distributions with $p_T = 0.4$–0.6 GeV/c in Au+Au collisions at $\sqrt{s_{\rm NN}} = 27$ GeV with centrality $0-5\%$(left), and $30-40\%$(right). Black data points depict all unlike-sign $p\pi$ pair distributions while the blue lines depict the combinatorial background distributions estimated via side-band fitting.

To ensure that the KF Particle method can be reliably used for the extraction of physical yields, we also applied the KF Particle method to a Monte Carlo (MC) simulated sample generated using an embedding technique detailed as follows.

Simulated Λ particles with a flat p_T and rapidity distribution are propagated through a GEANT3 [16] simulation of the STAR TPC. The Λ particles decay inside the simulated detector and the electronic signals originating from the decay particles are mixed with those from a given event from real data. The number of simulated Λs particles is 5% of the measured charged particle multiplicity of the event in which the simulated particles are embedded, and the simulated Λs all originate from the primary vertex of that event. The combined electronic signals are then processed with STAR tracking software, which is also used for real data processing. The KF Particle package is then deployed to the resultant tracks for Λ reconstruction.

We compare the performance of KF Particle on real data and MC simulation samples. The topological variables listed below are used in the selection of Λ candidates during the KF Particle reconstruction.

Comparisons of these variables between data and MC simulation for Λ candidates with $0.4 \le p_T \le 1.2$ GeV and centrality between $0-10\%$ are shown in Fig. 4. In general, distributions of such topological variables from data are well described by MC simulations for all centrality and p_T.

In order to achieve optimal significance of the Λ signal, the **T**oolkit for **M**ulti**v**ariate data **A**nalysis is used. TMVA is a

Table 1. Topological variables for Λ reconstruction.

variable	description
$\chi^2_{prim,\pi}$	χ^2 deviation of π track to the primary vertex
$\chi^2_{prim,p}$	χ^2 deviation of p track to the primary vertex
χ^2_{topo}	χ^2 of primary vertex to the reconstructed Λ
$\chi^2_{p-\pi}$	χ^2 of daughter particle (p-π) fit
d_Λ	decay length of Λ
$d_\Lambda/\sigma d_\Lambda$	decay length normalized by its uncertainty

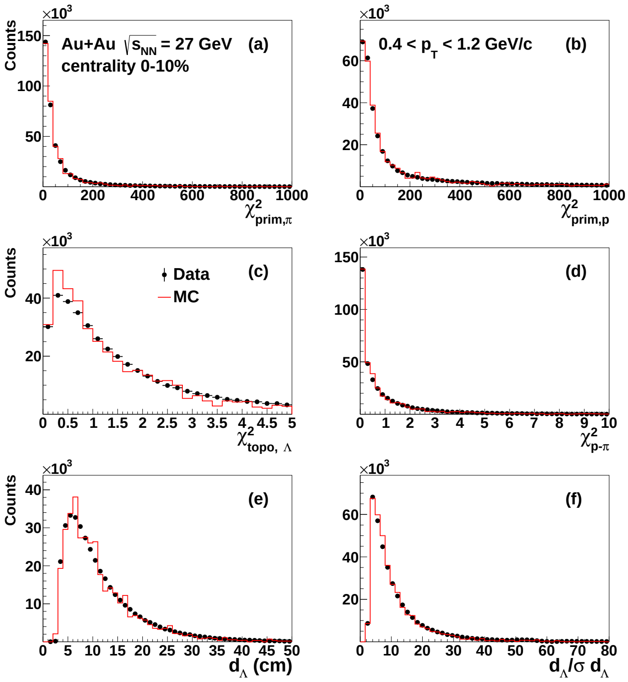

Fig. 4. Key topological variables used in KF Particle method for Λ reconstruction. Data and MC simulations are compared.

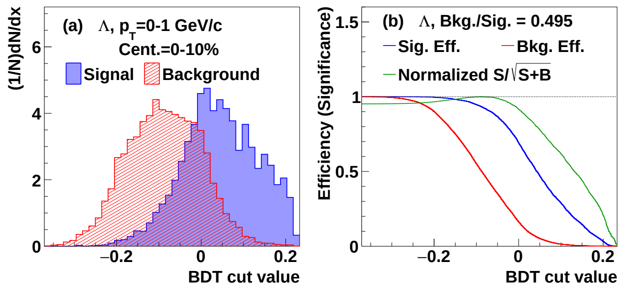

Fig. 5. (left) BDT response value distributions for signal (blue) and background (red) Λ candidates in the p_T range $0-1$ GeV/c for $0-10\%$ centrality. (right) Efficiency for signal (blue) and background Λ candidates (red) in the p_T range $0-1$ GeV/c for $0-10\%$ centrality as a function of the cut value placed on the BDT response value. The significance (green) achieves its maximum value when the cut value is -0.09.

family of supervised learning algorithms that can be used for differentiating signals and backgrounds. For more details, see Ref. [11]. A signal sample and a background sample are prepared as input for training. The signal samples are obtained from a GEANT3 simulation as described above. For the background sample, we select sideband ($3\sigma < |m_{p\pi} - m_{\Lambda,\mathrm{PDG}}| < 6\sigma$) $p-\pi$ pairs in the real data around the Λ mass peak, where σ is the width of the Λ mass peak, and $m_{p\pi}$ are $m_{\Lambda,\mathrm{PDG}}$ are the masses of the $p-\pi$ pair and the Λ baryon from the PDG respectively. These signal and background samples are then further divided into different p_T and centrality classes. We use the **B**oosted **D**ecision **T**ree method for training. Decision tree learning takes a set of input features and splits the input data recursively based on those features. In our case, the input features are the topological variables listed in Tab. 1 and the input data are the signal and background samples depending on these variables. Boosted decision trees combine multiple trees to strengthen the differentiation power for a detailed discussion, see Ref. [17]. The training takes into account the correlations between the different topological variables and collapses them into a single value, referred to as the BDT response value.

The BDT response value distributions from the signal and background samples for Λ candidates with $p_T = 0-1$ GeV/c and centrality 0–10% are shown in the left panel of Fig. 5. We observe that the BDT response values for the signal and background are significantly different from each other and thus serve as a good measure for differentiating between the signal and background. In order to select a BDT response cut value to optimize the significance $S/\sqrt{S+B}$, where S stands for signal counts and B stands for background counts, we use the TMVA package to first calculate the signal and background efficiency as a function of the BDT response cut value, $\varepsilon_S(\mathrm{BDT\ cut})$ and $\varepsilon_B(\mathrm{BDT\ cut})$, using the signal and background samples respectively. The signal and background efficiencies for Λ candidates in the p_T range 0–1 GeV/c centrality are shown in Fig. 5. The estimated Significance can then be calculated from Eq. 3:

$$Sig.(\mathrm{BDT\ cut}) = \frac{\mathrm{S_0}\varepsilon_\mathrm{S}(\mathrm{BDT\ cut})}{\sqrt{\mathrm{S_0}\varepsilon_\mathrm{S}(\mathrm{BDT\ cut}) + \mathrm{B_0}\varepsilon_\mathrm{B}(\mathrm{BDT\ cut})}}, \tag{3}$$

where S_0 and B_0 are the number of signal and background counts where no BDT response value cut is applied. These numbers are obtained from real data directly, and the calculated significance as a function of the cut value applied on the BDT response value for Λ candidates in the p_T range 0–1 GeV/c centrality is also shown in the right panel of Fig. 5. We find that a cut value of -0.09 maximizes the significance, and this cut value is chosen for this analysis. This procedure is then repeated for each p_T and centrality bin. In general, as the signal-to-background ratio decreases, a stricter BDT selection cut is necessary to optimize the significance.

We extract the number of signal and background counts for each p_T and centrality bin using the tuned BDT cuts obtained as explained above. We then use the standard helix swimming method used in previous STAR analyses [6], the cuts are also tuned by the BDT and extract the corresponding number of signal and background counts, and compare the significance

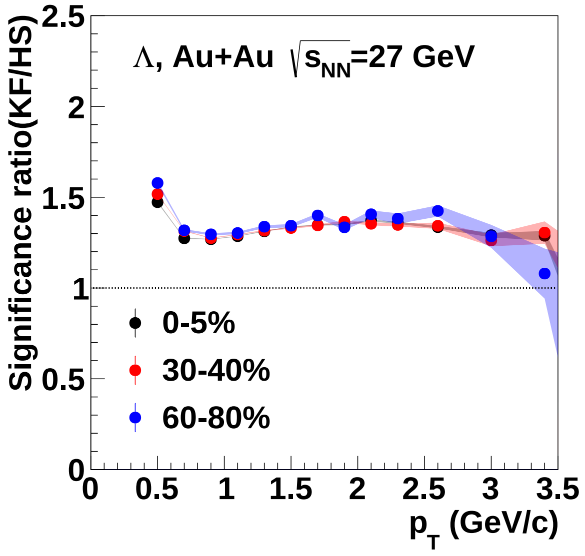

Fig. 6. Ratio of significance for Λ particles using the KF Particle method in conjunction with BDT training over those using the helix swimming(HS) method as a function of p_T of the Λ particles for centrality selection $0-5\%$(red), $30-40\%$(blue) and $60-80\%$(magenta).

obtained using these two methods. The track quality and particle identification cuts are chosen to be identical to each other for a fair comparison. The ratios of the significance as a function of p_T for three different centrality selections are shown in Fig. 6. The increase in significance is approximately independent of centrality, $\approx 30\%$ in the p_T range 1–3 GeV/c, and increases at low p_T to $\approx 50\%$. In conclusion, this demonstrates that the KF Particle method gives a larger significance for Λ signal extraction in Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV with the STAR experiment.

B. Ω Reconstruction

We then turn to Ω baryon. Ω baryons are reconstructed via the decay channel $\Omega \to \Lambda + K^- \to p + \pi^- + \mathrm{K}^-$. Ω particles decay with a proper decay length of $c\tau \simeq 25$ mm [4], and the Λ daughters will decay again soon after. The final daughter tracks are detected by the STAR TPC. Similarly, for each proton, kaon or pion track, we require a minimum of 15 hits to ensure good track quality. We reconstruct the Λ baryons with the KF Particle method first and then treat it as a daughter track to reconstruct the Ω production vertex.

Since the decay topology for Ω baryons is more complicated than that for Λ baryons, more topological variables can be used for training to facilitate the differentiation between the signal and background. The topological variables listed in Tab. 2 are used in the selection of Ω baryon candidates during KF Particle reconstruction.

Similar to the Λ baryon study, we generate a MC sample of reconstructed Ω baryons using a GEANT3 simulation of the STAR TPC. The data-MC comparison of the key topological variables are shown in Fig. 7.

Table 2. Topological variables for Ω reconstruction.

variable	description
$\chi^2_{prim,\pi}$	χ^2 deviation of π track to the primary vertex
$\chi^2_{prim,p}$	χ^2 deviation of p track to the primary vertex
$\chi^2_{prim,K}$	χ^2 deviation of K track to the primary vertex
$\chi^2_{topo,\Lambda}$	χ^2 of primary vertex to the reconstructed Λ
$\chi^2_{p-\pi}$	χ^2 of daughter particle (p-π) fit
χ^2_{topo}	χ^2 of primary vertex to the reconstructed Ω
$\chi^2_{\Lambda-K}$	χ^2 of daughter particle (Λ-K) fit
d_Λ	decay length of Λ
$d_\Lambda/\sigma_{d_\Lambda}$	Λ decay length normalized by its uncertainty
d_Ω	decay length of Ω
$d_\Omega/\sigma d_\Omega$	Ω decay length normalized by its uncertainty

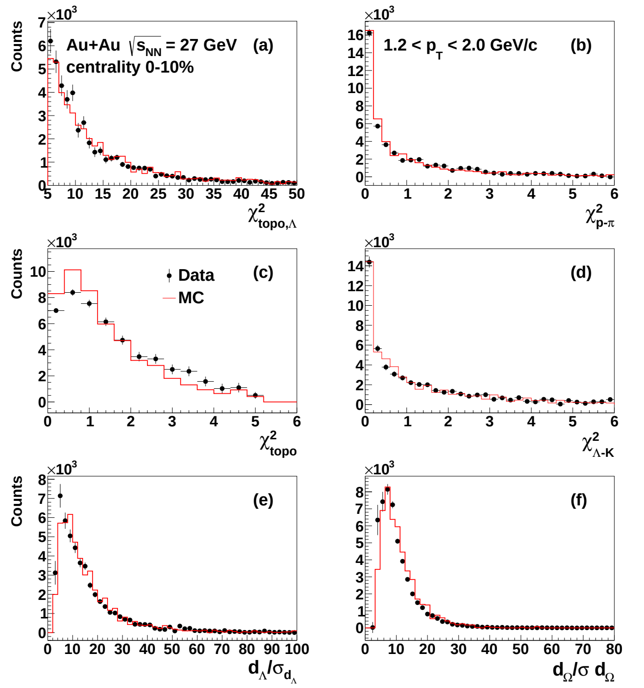

Fig. 7. Key topological variables used in KF Particle method for Ω reconstruction. Data and MC simulations are compared.

We find reasonable agreement between the data and MC simulations, which suggests a proper estimation and usage of the covariance matrix of the Λ daughters and gives us confidence that the KF Particle method may be reliably used for the extraction of Ω baryon yields. We then generate a signal and background sample with the same method as in Λ analysis to supply input for TMVA training using the BDT method. The BDT response value distribution for Ω candidates with $p_T = 1-4$ GeV is shown in the left panel of Fig. 8, and the signal efficiency, background efficiency and significance are shown in Fig. 8. As in the case for Λ analysis, we select the BDT response cut value that optimizes the significance.

This process is repeated for each p_T and centrality bin. The

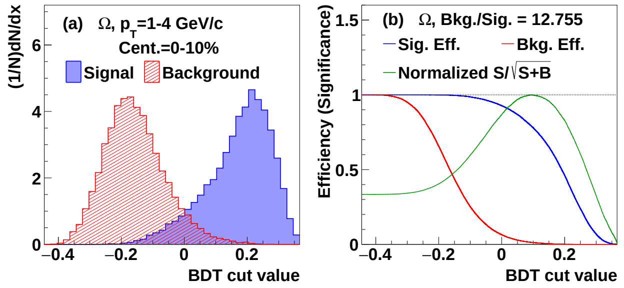

Fig. 8. (left) BDT response value distributions for signal (blue) and background (red) Ω candidates in the p_T range $1-4$ GeV/c for $0-10\%$ centrality. (right) Efficiency for signal (blue) and background Ω candidates (red) in the p_T range $1-4$ GeV/c for $0-10\%$ centrality as a function of the cut value placed on the BDT response value. The significance (green) achieves its maximum value when the cut value is 0.09.

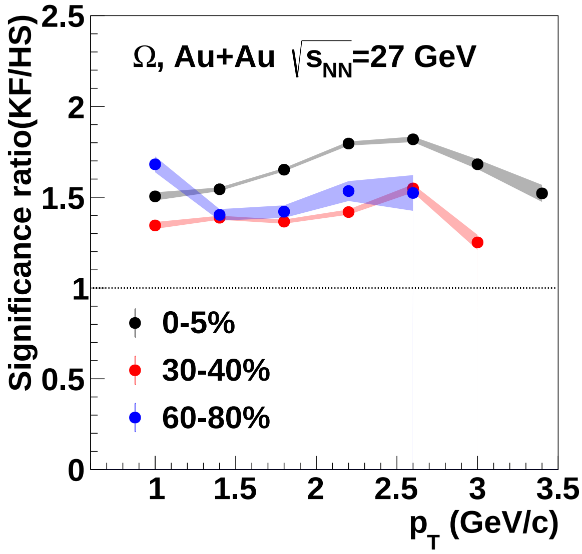

Fig. 10. Ratio of significance for Ω particles using the KF Particle method in conjunction with BDT training over those using the helix swimming(HS) method as a function of p_T of the Ω particles for centrality selection $0-5\%$(red), $30-40\%$(blue) and $60-80\%$(magenta).

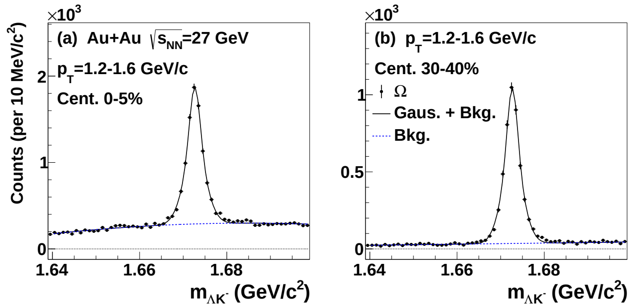

Fig. 9. ΛK invariant mass distributions with $p_T = 1.2-1.6$ GeV/c in Au+Au collisions at $\sqrt{s_{\rm NN}} = 27$ GeV with centrality $0-5\%$(left), and $30-40\%$(right). Black data points depict all unlike-sign $p\pi$ pair distributions while the blue lines depict the combinatorial background distributions estimated via side-band fitting

significances using the optimized BDT response cuts for each p_T and centrality bin are extracted. We then carry out signal extraction using the default helix swimming method, with candidate selection cuts are chosen to be the same as previous Ω analyses at the same collision energy [6, 18]. The signal and background counts using the default helix swimming method are extracted, and the ratio of the significances using these two methods are calculated and shown in Fig. 10. We observe an $\approx 50\%$ increase in significance in the p_T range of $1-4$ GeV/c. This increase is higher than the case for Λ, likely due to the more complex decay topology with two decay vertices reconstructed by KF Particle and larger background. Further studies using KF Particle are underway to extend the low p_T reach beyond 1 GeV/c; however, this is beyond the scope of this paper.

C. D^0 Reconstruction

D^0 particles are reconstructed via the decay channel $D^0 \to K^-\pi^+$ with a proper decay length of $c\tau \simeq 123$ μm [4]. Since this decay length is less than the spatial resolution of the TPC detector, the information from the micro-vertex detector HFT is used to identify the D^0 decay vertex from the primary collision vertex. For each kaon or pion daughter track, we required a minimum of 15 hits in the TPC and a match to the HFT detector with at least 3 hits inside to ensure good track quality. For kaon and pion particle identification, in addition to the requirement of $|n\sigma_\pi| < 3$ and $|n\sigma_{\rm K}| < 2$, we also utilized the information from the Time-of-Flight (TOF) detector by requiring the measured inverse velocity ($1/\beta$) to be within three standard deviations from the expected value when the measurement is available [5]. The topological variables listed in Tab. 3 are used in the selection of D^0 meson candidates in KF Particle reconstruction. $p_{T,\pi}$ and $p_{T,K}$ cut is added here to reject combinatorial background at low p_T.

Table 3. Topological variables for D^0 reconstruction.

variable	description
$\chi^2_{prim,\pi}$	χ^2 deviation of π track to the primary vertex
$p_{T,\pi}$	transverse momentum of π track
$\chi^2_{prim,K}$	χ^2 deviation of K track to the primary vertex
$p_{T,K}$	transverse momentum of K track
χ^2_{topo,D^0}	χ^2 of primary vertex to the reconstructed D^0
$\chi^2_{K,\pi}$	χ^2 of daughter particle (K-π) fit
$L_{D^0}/\sigma_{L_{D^0}}$	D^0 decay length normalized by its uncertainty

Similar to the Λ and Ω baryon study, we generated an MC sample of reconstructed D^0 mesons using a GEANT3 simulation of the STAR TPC and HFT and processed through full detector tracking as was done in the real data reconstruc-

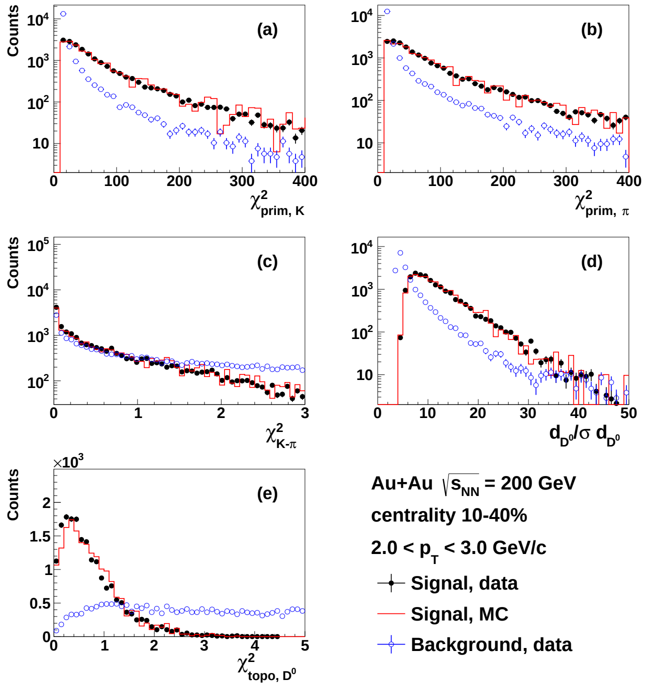

Fig. 11. Key topological variables used in KF Particle method for D^0 reconstruction. Data, MC simulations and background are compared.

tion. The HFT simulator was tuned to reproduce the single track efficiency and DCA pointing resolution observed in real data. However, the consistency in topological variable distributions between data and MC for D^0 signals has yet to be demonstrated. Figure 11 shows the comparison of several key topological variables used in the KF Particle method for D^0 reconstruction between data (black data points) and MC (red histograms) for D^0 signals. We find a reasonable agreement between the data and MC simulations for D^0 signals. Additionally, background distributions also shown in Fig. 11 (blue data points), which are estimated from real data using the side-band method in which the background candidates were selected by requiring the invariant mass of $K\pi$ candidate pairs within $3\sigma < |M_{\rm inv} - M_{D^0}| < 6\sigma$ (σ is the Gaussian width of the D^0 signal). The signal and background can be well distinguished especially on $L_{D^0}/\sigma_{L_{D^0}}$ and χ^2_{topo,D^0} distributions.

We then used the signal sample generated from the MC simulation and the background sample from the real data to conduct the TMVA training with the BDT method to find the topological selection working point for the best signal significance. Figure 8 left panel shows the BDT response value distributions for the D^0 signal and background in the region of $2 < p_T < 3$ GeV/c, and the signal/background efficiencies and the signal significance are shown in the right panel. The signal significance was normalized to its maximum value. We determined the BDT response cut value in order to optimize the significance of D^0 in each p_T and centrality class.

We then applied the optimized BDT selection cuts in the real data analysis. Figure 13 shows the D^0 invariant mass dis-

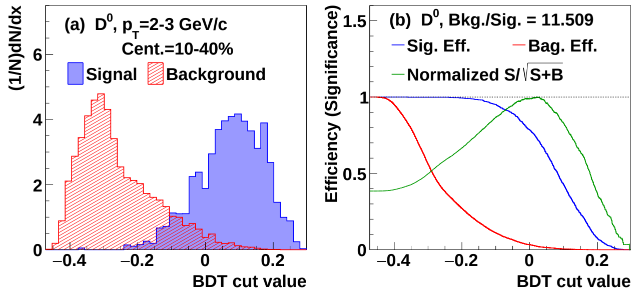

Fig. 12. (left) BDT response value distributions for signal (blue) and background (red) D^0 candidates in the p_T range $2-3$ GeV/c for $10-40\%$ centrality. (right) Efficiency for signal (blue) and background D^0 candidates (red) in the p_T range $2-3$ GeV/c for $10-40\%$ centrality as a function of the cut value placed on the BDT response value. The significance (green) achieves its maximum value when the cut value is 0.05.

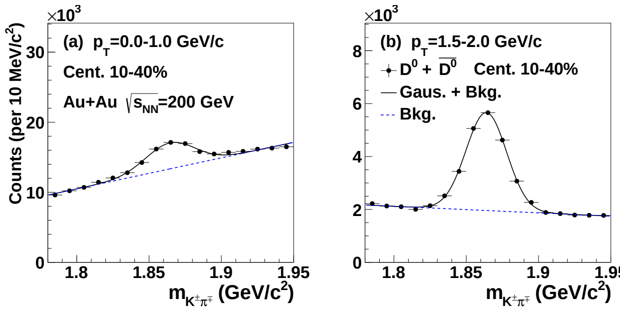

Fig. 13. $k\pi$ invariant mass distributions using the KF Particle method in $10-40\%$ Au+Au collisions at $\sqrt{s_{\rm NN}} = 200$ GeV in the region of p_T<1 GeV/c (left) and 1.5<p_T<2 GeV/c (right). Black data points depict all unlike-sign $k\pi$ pair distributions while the blue lines depict the combinatorial background distributions estimated via side-band fitting

tributions using the KF Particle method in $10-40\%$ Au+Au collisions at $\sqrt{s_{\rm NN}} = 200$ GeV in the regions of p_T<1 GeV/c (left), and 1.5<p_T<2 GeV/c (right), respectively. Red lines depict the function fits to the data with a Gaussian function for the D^0 signal plus a linear background.

Signal significance was then calculated from these distributions for D^0 candidates within a mass window of $|M_{\rm inv} - M_{D^0}| < 3\sigma$ where σ is the D^0 signal width determined by the Gaussian function fit. The background counts were determined based on the linear background function fit results. We compare the significance values from the KF Particle method to the helix swimming (HS) method used in previous analysis [5] and the ratio between the two methods is shown in Fig. 14. The shaded bands indicate statistical uncertainties from this calculation. The comparison demonstrates that the KF Particle method improves the reconstructed D^0 signal significance, especially in the low p_T and more central collisions. In $0-10\%$ central Au+Au collisions at $p_T < 1$ GeV/c, the improvement can be as significant as a factor of ∼3. This is possibly due to the enormous amount of combinatorial background (hundred times signals) in that particular range,

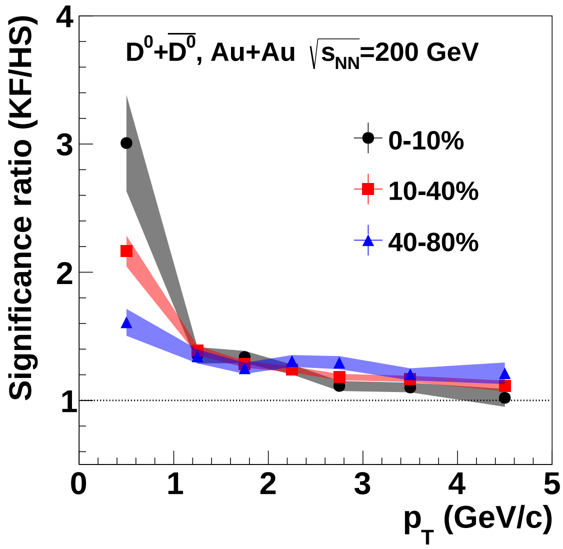

Fig. 14. Ratio of significance for D^0 particles using the KF Particle method in conjunction with BDT training over those using the helix swimming(HS) method as a function of p_T of the D^0 particles for centrality selection $0-10\%$(black), $10-40\%$(red) and $40-80\%$(blue). The shaded bands indicate the statistical uncertainties.

and the cuts based on statistical criteria work well on selecting the particles that originate from a secondary vertex.

IV. SUMMARY

In summary, we have applied the KF Particle method to the reconstruction of Λ, Ω^- hyperons, and D^0 meson in the STAR experiment. The KF Particle method, by utilizing covariant matrices of tracking parameters, improves the reconstructed Λ (Ω) significance by approximately 30% (50%) compared to the traditional helix swimming method in $\sqrt{s_{\mathrm{NN}}}$ = 27 GeV Au+Au collisions. The improvement in D^0 significance by applying the KF Particle method has a p_T dependence in $\sqrt{s_{\mathrm{NN}}}$ = 200 GeV Au+Au collisions with the largest improvement as significant as a factor of ~3 in p_T<1 GeV/c and $0-10\%$ central collisions. We also demonstrated that Monte Carlo simulation can reproduce the topological variable distributions used in the KF Particle method, and thus establishes KF Particle as a robust method for strange and open charm hadron analyses in the STAR experiment. Since the KF Particle method is independent of the geometry of the detector, it will be useful in other experiments, especially in analyses with a small signal-to-background ratio.

V. ACKNOWLEDGEMENTS

The authors thank the STAR Collaboration, the RHIC Operations Group and RCF at BNL, and the NERSC Center at LBNL for their support. This work was supported in part by the Offices of NP and HEP within the U.S. DOE Office of Science; Authors XXX are supported in part by National Natural Science Foundation of China under Grant No XXX.

[1] P. Koch, B. Muller, J. Rafelski, Strangeness in Relativistic Heavy Ion Collisions, Phys. Rept. 142 (1986) 167–262. doi:10.1016/0370-1573(86)90096-7.

[2] S. Frixione, M. L. Mangano, P. Nason, G. Ridolfi, Heavy quark production, Adv. Ser. Direct. High Energy Phys. 15 (1998) 609–706. arXiv:hep-ph/9702287, doi:10.1142/9789812812667_0009.

[3] X. Dong, Y.-J. Lee, R. Rapp, Open Heavy-Flavor Production in Heavy-Ion Collisions, Ann. Rev. Nucl. Part. Sci. 69 (2019) 417–445. arXiv:1903.07709, doi:10.1146/annurev-nucl-101918-023806.

[4] P. Zyla, et al., Prog. Theor. Exp. Phys. 2020 (2020) 083C01.

[5] J. Adam, et al., Centrality and transverse momentum dependence of D^0-meson production at mid-rapidity in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}}$ = 200 GeV, Phys. Rev. C 99 (3) (2019) 034908. arXiv:1812.10224, doi:10.1103/PhysRevC.99.034908.

[6] J. Adam, et al., Strange hadron production in Au+Au collisions at $\sqrt{s_{NN}}$ =7.7 , 11.5, 19.6, 27, and 39 GeV, Phys. Rev. C 102 (3) (2020) 034909. arXiv:1906.03732, doi:10.1103/PhysRevC.102.034909.

[7] K. Ackermann, et al., STAR detector overview, Nucl. Instrum. Meth. A 499 (2003) 624–632. doi:10.1016/S0168-9002(02)01960-5.

[8] M. Anderson, et al., The Star time projection chamber: A Unique tool for studying high multiplicity events at RHIC, Nucl. Instrum. Meth. A 499 (2003) 659–678. arXiv:nucl-ex/0301015, doi:10.1016/S0168-9002(02)01964-2.

[9] G. Contin, et al., The STAR MAPS-based PiXeL detector, Nucl. Instrum. Meth. A 907 (2018) 60–80. arXiv:1710.02176, doi:10.1016/j.nima.2018.03.003.

[10] L. Adamczyk, et al., Measurement of D^0 Azimuthal Anisotropy at Midrapidity in Au+Au Collisions at $\sqrt{s_{NN}}$=200 GeV, Phys. Rev. Lett. 118 (21) (2017) 212301. arXiv:1701.06060, doi:10.1103/PhysRevLett.118.212301.

[11] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, TMVA: Toolkit for Multivariate Data Analysis, PoS ACAT (2007) 040. arXiv:physics/0703039.

[12] R. E. Kalman, A new approach to linear filtering and prediction problems, J. Basic Eng. 82(1) (1960) 35–45. doi:10.1115/1.3662552.

[13] R. Frühwirth, et al., Data analysis techniques for high-energy physics, 2nd Edition, Cambridge, 2000.

[14] S. Gorbunov, On-line reconstruction algorithms for the CBM and ALICE experiments, PhD Thesisdoi:urn:nbn:de:hebis:30:3-295385.

[15] M. Zyzak, Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR, PhD Thesisdoi:urn:nbn:de:hebis:30:3-414288.

[16] R. Brun, F. Bruyant, M. Maire, A. C. McPherson, P. Zanarini, GEANT 3: user's guide Geant 3.10, Geant 3.11; rev. version, CERN, Geneva, 1987.
URL https://cds.cern.ch/record/1119728

[17] H. Drucker, C. Cortes, Boosting decision trees, in: Proceedings of the 8th International Conference on Neural Information Processing Systems, NIPS'95, MIT Press, Cambridge, MA, USA, 1995, p. 479–485.

[18] L. Adamczyk, et al., Probing parton dynamics of QCD matter with Ω and ϕ production, Phys. Rev. C93 (2) (2016) 021903. arXiv:1506.07605, doi:10.1103/PhysRevC.93.021903.