Analysis Note:
Measurement of D^0-meson production in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}} = 200\,\mathrm{GeV}$

Xiaolong Chen, Xin Dong, Mustafa Mustafa, Guannan Xie, Yifei Zhang, Long Zhou

June 5, 2017

Abstract

Heavy quarks nuclear modification factor (R_{AA}) has been proposed as an important measurement to study the flavor dependence of partons energy loss in the medium, and eventually to help in extracting the medium transport, drag and diffusion coefficients.

We report the first measurement of efficiency corrected spectrum for D^0 from HFT through the hadronic channels ($D^0(\bar{D}^0) \to K^{\mp}\pi^{\pm}$), the Nuclear Modification Factor (R_{AA}) of D^0 mesons in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. The precision was much better compared to previous published one. The new R_{AA} results show significant suppression at high p_T which means the strong interaction between charm quark and the medium and loose energy. While the D^0 R_{AA} shows quite similar trend as light hadrons.

1 Introduction

Heavy quarks nuclear modification factor (R_{AA}) has been proposed as an important measurement to study the flavor dependence of partons energy loss in the medium, and eventually to help in extracting the medium transport, drag and diffusion coefficients. There are lots of theoretical calculations for the energy losses for different flavor particles. Fig. **??** shows the jet flavor tomography level crossing pattern of nuclear modification factors at middle rapidity of π, D, B, e from CUJT calculations for central Au + Au 200 GeV collisions. As clearly see the mass hierarchy of the different flavor energy loss.

The hadronic channels allow to fully reconstruct the charmed hadrons and do not suffer from the complications in the semi-leptonic decays, however the measurement can be challenging due to large combinatorial backgrounds and lower branching ratios. One approach is to use the decay topology to reduce this background by distinguishing between tracks that come from the collision itself (primary vertex) and those from a secondary decay vertex. This requires the detectors must be able to resolve differences on the order of tens of microns. Heavy Flavor Tracker (HFT) is essence the right detector for this mission.

Fig. **??** shows the R_{AA} of D^0, π, $h^{\pm}$ from various measurements. A significant suppression is clearly seen at the high p_T range for both light hadrons and charmed hadrons both in RHIC energy and LHC energy. The enhancement observed in the intermediate p_T range from STAR can be described by the models including coalescence of charm and light quarks, even though the uncertainties are still large in the low transverse momentum range. It will be critical to precise measure the low p_T structure.

2 Datasets and Event Selection

The dataset used in this analysis is P16id production of 2014 Au+Au 200 GeV data. This is the first year of physics running the new STAR HFT Detector. The analysis uses picoDst which is produced from MuDst.

The Minimum-Bias (MinBias) trigger is defined as a coincidence between the two VPDs, and an online collision vertex cut. Moreover, a pile-up protection at the trigger level was applied for the data taking. In this analysis, the MinBias trigger, denoted as "vpdmb-5-p-nobsmd" and "vpdmb-5-p-nobsmd-hlt", is used. The triggers used in this analysis are listed in Table **??**.
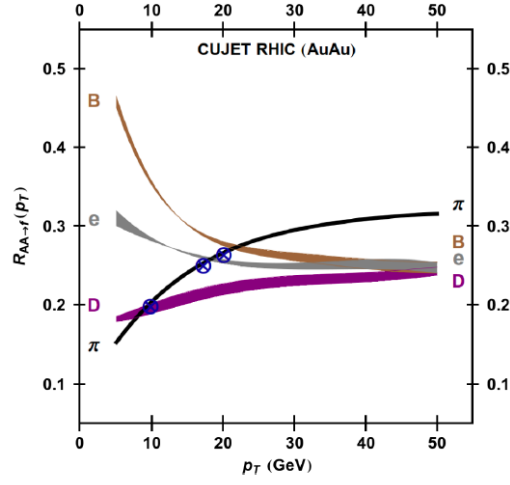
Figure 1: Jet flavor tomography level crossing pattern of nuclear modification factors at middle rapidity of π, D, B, e calculations for central Au + Au 200 GeV collisions.
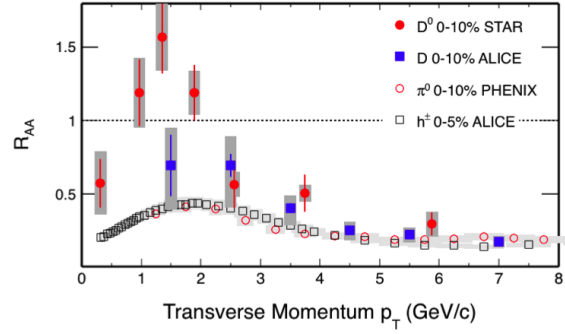

Figure 2: (upper) D^0, π, $h^\pm$ R_{AA} from different measurements. (bottom) v_2 of D and $h^\pm$ from ALICE.

Table 1: Triggers ID used in this analysis from run14

Trigger ID	description
450050	vpdmb-5-p-nobsmd-hlt
450060	vpdmb-5-p-nobsmd-hlt
450005	vpdmb-5-p-nobsmd
450015	vpdmb-5-p-nobsmd
450025	vpdmb-5-p-nobsmd

Events used in this analysis are required to have a valid collision vertex V_z (primary vertex) within 6 cm of the TPC center along z direction (the beam direction) to ensure a uniform TPC acceptance and make sure the most tracks are within the PiXeL (PXL) detector coverage. The PXL detector is about 20 cm along the z direction, and the radius of the inner layer is about 2.8cm and outer layer is about 8 cm. Furthermore, the distance between the V_z constructed by TPC and the vertex constructed by VPD (V_z^{VPD}, fast detector) is within 3 cm to reject the bad events. A radial length less than 2 cm for the vertex is required to reject the events from the beam hitting the beam pipe. After event selection, $\sim$875 million MinBias events are used for this analysis. Table **??** lists the event selection criterion.

Table 2: Event selection in Au+Au collisions at 200 GeV for D^0.

Event Selection Criteria
$!(|V_x| == 0 \ \&\& \ |V_y| == 0 \ \&\& \ |V_z| == 0)$
$|V_z| < 6$ cm
$|V_r| < 2$ cm
$|V_z - V_z^{VPD}| < 3$ cm

2.1 Centrality Definition

The centrality for Run14 200GeV Au+Au collisions MinBias sample is based on gRefMult. The gRefMult is defined as the number of global tracks with $|\eta| < 0.5$, no less than 10 TPC hits, and Distance of Closest Approach (DCA) to primary vertex less than 3 cm with some correction according to V_z and luminosity. The centrality definition according to the corrected gRefMult is listed in Table **??**. This is decided by comparing the measured gRefMult distribution with the Glauber model simulation.

The basic procedure for centrality definition have three steps. First, need the quality assurance (QA) for the data set and remove those outlier runs. The second step would be correct the V_z and luminosity dependence for the reference multiplicity (gRefmult). The last step would be compare our data with Glauber MC simulation and determine the centrality classification.

For the QA, several variables are used for the outlier selection, such as Refmult (primary track multiplicity), gRefmult (global track multiplicity), TofRefmult (tof track multiplicity) and etc. In the Fig. **??** shows the $\langle$gRefmult$\rangle$ as a function of run index for the QA, and there are several outliers are identified. Those dashed lines are the 4$\times$RMS range, beyond those range, the runs are identified as bad run. And several iterations are did until all the runs are within these 4$\times$RMS range.

In the Fig. **??** shows the $\langle$HFT $p_T\rangle$, which is the mean p_T of HFT tracks, as a function of run index for the QA, and we can clearly see there is a deep before the run index $\sim$520 which corresponding to the run number 15107008. And also we saw the same structure in the Fig. **??**, which shows the average of HFT matching Ratio in the p_T range between 0.7 to 0.8 GeV/c and Fig. **??** shows this HFT matching ratio in the high p_T range. So, basically those runs before day 107 were taken out for this analysis, since it will complicate our efficiency calculation.

This deep was identified later on with a lot of effort, it was due to the firmware issue. And more details can be found in the STAR documents below.

https://drupal.star.bnl.gov/STAR/system/files/STAR_PXL_Firmware_Issue_Solved_Final_Report_Oct3_v2.pdf

Fig. **??** and Fig. **??** show the normalized gRefmult distribution for several different V_z range from -6 cm to 6 cm. The shape are quite different for VpdMB5 trigger in Fig. **??** while the Fig. **??** shows the same plots for VpdMB30 trigger (V_z within range from -30 cm to 30 cm). This difference was explained by that the online Vpd vertex cut have a negative offset and the Vpd resolution has centrality dependence. As for the most central collisions, the resolution will be better than the most peripheral collisions. That is the reason we saw more events in the negative V_z range and more central events for VpdMB5 trigger compared to VpdMB30 trigger.
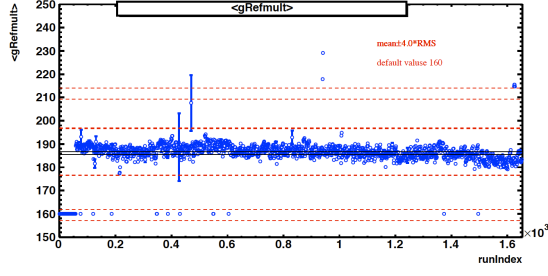
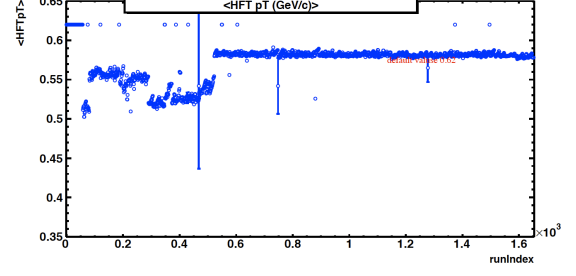Figure 3: The mean value of gRefmult ($\langle$gRefmult$\rangle$) as a function of run index from QA.

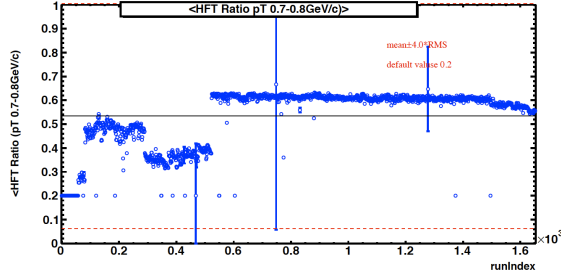Figure 4: The mean value of p_T for HFT matched track ($\langle$HFT $p_T\rangle$) as a function of index.

Figure 5: The mean value of HFT matching ratio ($\langle$HFT $p_T\rangle$) as a function of run index at the transverse momentum range $0.7 < p_T < 0.8$ GeV/c.
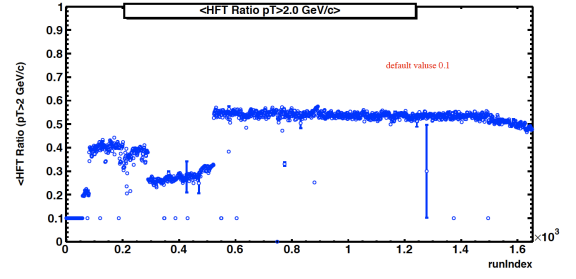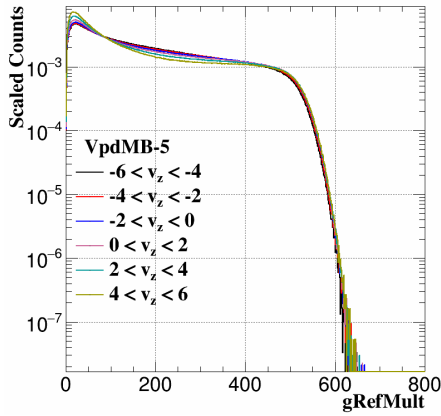

Figure 6: The mean value of HFT matching ratio ($\langle$HFT $p_T\rangle$) as a function of run index at the transverse momenaum range $p_T > 2.0$ GeV/c.

Figure 7: Normalized gRefmult distribution for VpdMB5 trigger along different V_z range
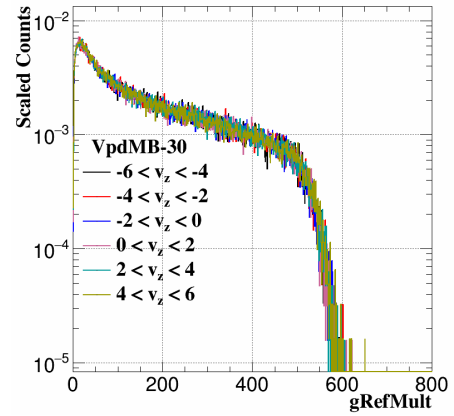

Figure 8: Normalized gRefmult distribution for VpdMB30 trigger along different V_z range

4

As the gRefmult have the luminosity dependence (related to the TPC tracking efficiency have luminosity dependence), we need to take out this effect by doing ZdcX (Zdc coincidence rate) correction. Fig. **??** shows the mean value of gRefmult (<gRefmult>) distribution as a function of ZdcX. There was a clear slope for this distribution as shown by the fitting parameters. Here the fitting function is Eq. **??**,

$$f_{ZdcX} = p0 + p1 * ZdcX \tag{1}$$

The goal of this correction is try to flatten this ZdcX dependence to take out of the luminosity effect. So here the correction factor was shown by Eq. **??**.

$$f_{ZdcX} = \frac{1}{1 + p1/p0 * ZdcX} \tag{2}$$

After the ZdcX correction, this <gRefmult> is flat as shown by Fig. **??**.

For the V_z dependence correction, we extract the high end point (h) from the fitting of gRefmult tail by the function of Eq. **??**.
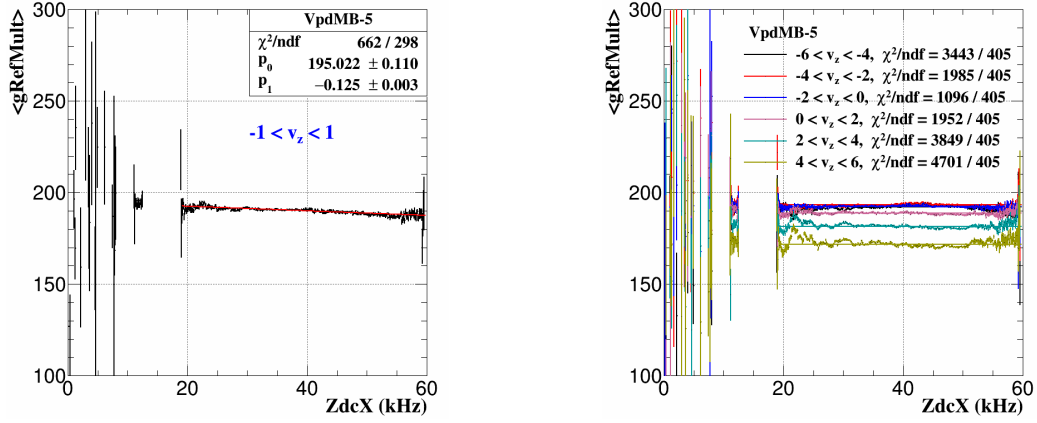
$$f_x = A * TMath :: Erf(-\sigma * (x - h)) + A \tag{3}$$

Figure 9: <gRefmult> as function of ZdcX before correction.

Figure 10: <gRefmult> as function of ZdcX after correction.

Fig. **??** shows an example of the fitting of gRefmult tail in the V_z range from 1 cm to 2 cm. The fitting parameters were shown on the plot.
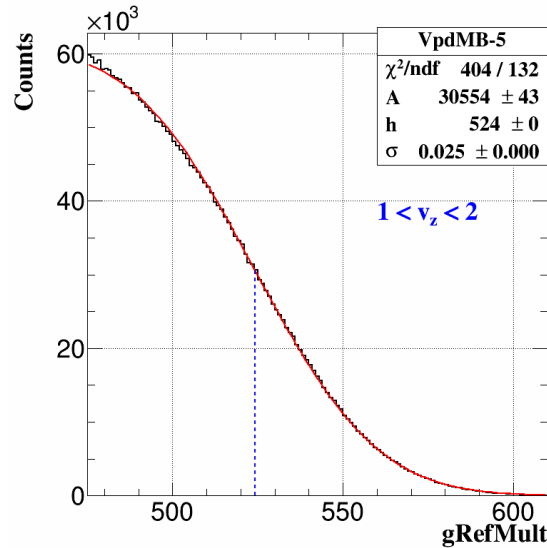

Figure 11: Fitting gRefmult tail distribution in the range of 1<V_z<2 cm by Eq. **??**.

With all the high end point value extracted along V_z direction, this V_z dependence was shown on Fig. **??**. The similar method as ZdcX correction, we need to flatten this V_z dependence, then the data point was fitted by 5th order polynomial function Eq. **??**, and then the correction factor was shown by Eq. **??** After the V_z correction, this high end point is flat as shown by Fig. **??**.

$$f_{V_z} = p0 + p1 * x + p2 * x^2 + p3 * x^3 + p4 * x^4 + p5 * x^5 \qquad (4)$$

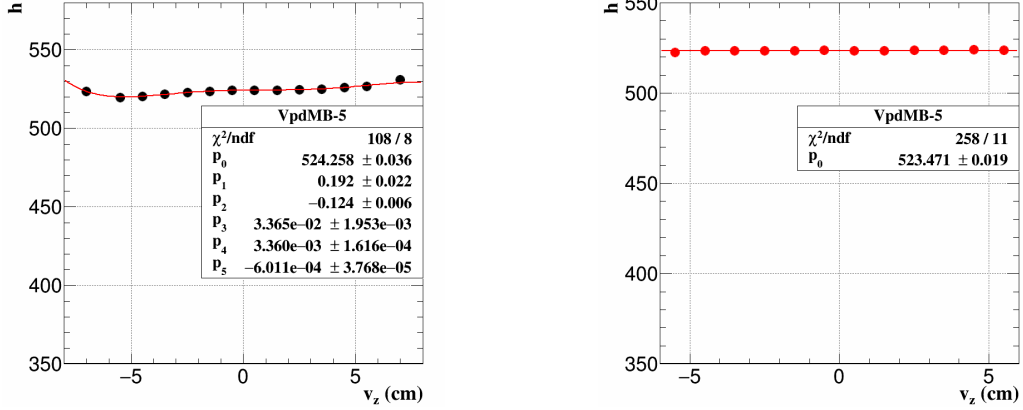$$f_{V_z} = \frac{p0}{p0 + p1 * x + p2 * x^2 + p3 * x^3 + p4 * x^4 + p5 * x^5} \qquad (5)$$

Figure 12: High end point as function of V_z after correction.

Figure 13: High end point as function of V_z after correction.

As shown from Fig. **??** and Fig. **??**, the clear V_z dependence need to avoid for VpdMB5 trigger. So, the centrality definition for VpdMB5 trigger was normalized to VpdMB30. After the ZdcX correction and V_z correction, we directly take it as an additional correction factor for VpdMB5. These correction factor was show in Fig. **??**. After this additional correction, the distributions from VpdMB5 and VpdMB30 are same.
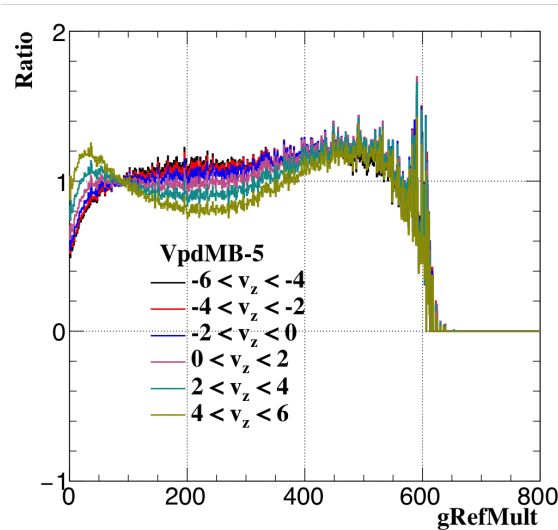

Figure 14: The double ratio of normalized gRefmult distribution from VpdMB5 over VpdMB30, this additional correction factor was try to normalized VpdMB5 to VpdMB30 trigger.

The Vpd MinBias trigger has a trigger efficiency that are lower for periheral events. Fig. **??** shows the gRefmult after V_z and ZdcX correction from data comparison with Glauber MC simulation. In the high end part the agreement was well, but in the low end part, due to this trigger ineffciency, there is clearly discrepancy between data and simulation. To do the measurement

without centrality bias, a weight proportional to inverse trigger efficiency is applied. The weight as a function of corrected gRefMult is shown in Fig. **??**.
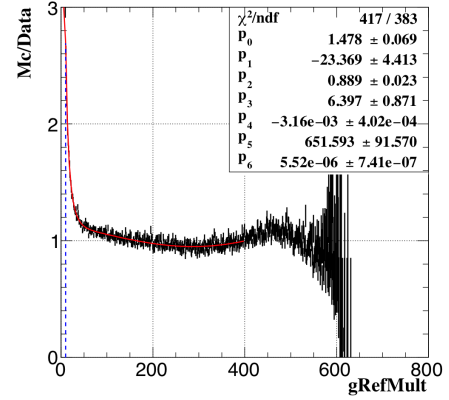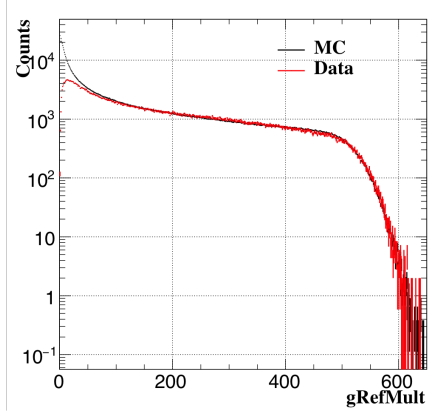

Figure 15: The comparison of corrected gRefMult between Data and Glauber MC. The red line is data and the black line is from Glauber MC.

Figure 16: MC/data gRefmult distribution for Vpd MinBias trigger efficiency correction. The red line fit function is the correction reweight function.
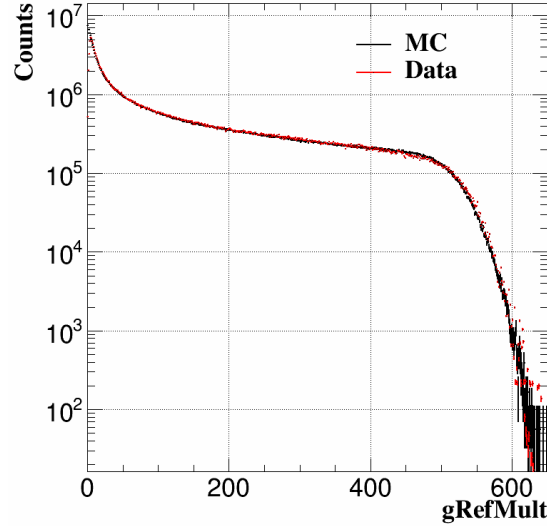

Figure 17: The comparison of corrected gRefMult (after all corrections and reweight) between Data and Glauber MC. The red line is data and the black line is from Glauber MC.

After all these corrections including V_z, ZdcX, Vpd trigger inefficiency, and Vpd resolution for different centralities as discussed before. Final step, the data was compared to Glauber MC simulation as shown in Fig. **??**. And then the determined centrality classification can be found at Table **??**.

3 D^0 Reconstruction

D^0 and $\bar{D^0}$ are reconstructed through the typically hadronic channel $K^{\mp}\pi^{\pm}$ using the topological method. In the following we will describe the daughter selection, the geometry cuts and how they are obtained through the TMVA tuning. We will show the D^0 signals for different p_T bins. We will also discuss some related topics: the mixed event to reconstruct the combinatorial background, and the correlated background source shown as a 'bump' at invariant mass lower than the D^0.

centrality	gRefMult	$\langle N_{coll} \rangle$	$\langle N_{part} \rangle$
75-80%	10-15	10.48	11.82
70-75%	15-21	16.11	16.68
65-70%	21-30	24.59	23.25
60-65%	30-41	36.13	31.15
55-60%	41-56	52.77	41.27
50-55%	56-73	75.36	53.46
45-50%	73-94	105.25	67.93
40-45%	94-119	143.54	84.71
35-40%	119-148	191.83	103.99
30-35%	148-182	253.13	126.52
25-30%	182-221	328.99	152.31
20-25%	221-266	422.49	181.93
15-20%	266-317	537.52	215.98
10-15%	317-376	677.99	254.90
5-10 %	376-443	852.75	299.95
0-5 %	>443	1066.50	348.74

Table 3: Centrality defintion based on gRefMult

3.1 Daughter Selection

D^0 have a lifetime of $c\tau \sim 123\mu$m. Thus the global tracks for daughter tracks are used in this analysis. The transverse momentum are required to ≥ 0.3 GeV/c to ensure that the track can pass through the TPC and have less HFT miss matching, the number of hit points (nHits) along the track is ≥ 20 (of a maximum of 45) to ensure good momentum resolution.

The pion and kaon tracks are identified by combining Time Projection Chamber (TPC) and Time Of Flight detector (TOF). The TPC provides particle identification utilizing the energy loss information dE/dx, different particle species with the same momentum may have different dE/dx. In additional, different particle species with the same momentum have different velocities, thus the TOF can be used to identify different particle species in the dE/dx crossover regions by precise velocity information ($1/\beta = ct/l$). The normalized dE/dx, $n\sigma_x$ (x = π, K, p, e etc.), defined in Eq. **??**, instead of dE/dx is used in this analysis. Where $\langle dE/dx \rangle_{measured}$ and $\langle dE/dx \rangle_x$ represent measured and theoretical dE/dx, and R is the STAR TPC dE/dx resolution (typically $\sim$8%). The $n\sigma_x$ should be close to a standard Gaussian distribution for each corresponding particle species (mean = 0, σ = 1).

$$n\sigma_x = \frac{1}{R} log \frac{\langle dE/dx \rangle_{measured}}{\langle dE/dx \rangle_x} \qquad (6)$$

Fig. **??** shows the TPC energy loss dE/dx information versus momentum achieved from Run14 Au+Au 200GeV, there are several clear bands for different particle species such as π, K, p and e.

Fig. **??** shows the TOF 1/Beta information versus momentum achieved from Run14 Au+Au 200GeV, also there are several clear bands for different particle species such as π, K, p .

In summary, next list all the related track selections for D^0 daughters including track quality cut and particle identification cut.

- global tracks

- $p_T > 0.3$ GeV/c

- $|\eta| < 1$

- $nHitsFit \geq 20$, in TPC

- at least one hit in every layer of PXL and IST
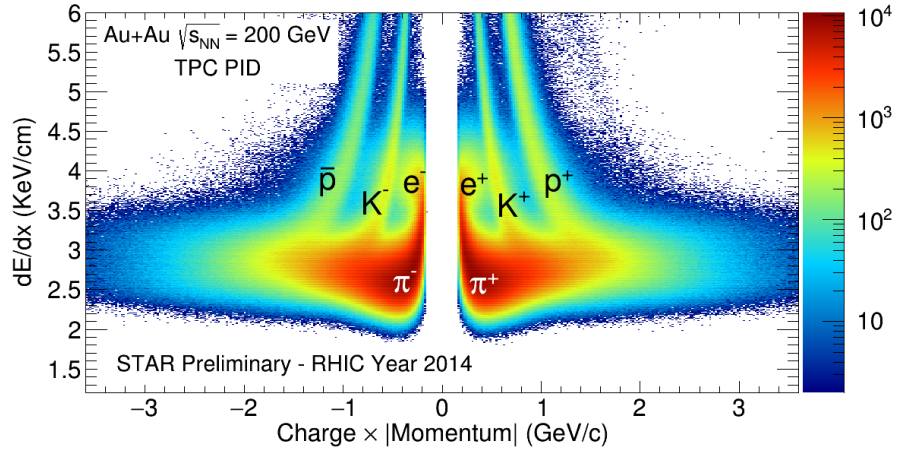
pion PID:

- $|n\sigma_\pi| < 3.0$, based on TPC dE/dx

Figure 18: TPC dE/dx versus charge×momentum achieved from Run14 Au+Au 200GeV.
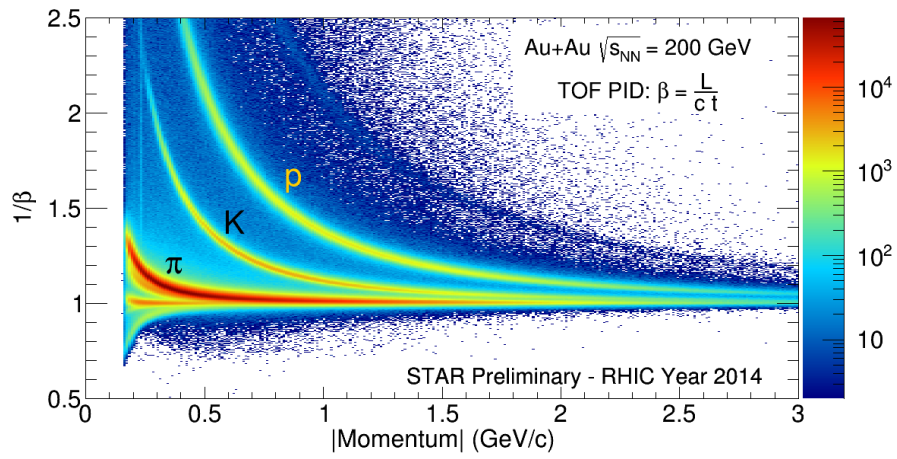

Figure 19: TOF 1/Beta versus momentum achieved from Run14 Au+Au 200GeV.

- If TOF is avaliable (hybrid PID): $|\frac{1}{\beta} - \frac{1}{\beta_{exp}}| < 0.03$

kaon PID:

- $|n\sigma_K| < 2.0$, based on TPC dE/dx

- If TOF is avaliable (hybrid PID): $|\frac{1}{\beta} - \frac{1}{\beta_{exp}}| < 0.03$

3.2 Topological Cut Optimization

The secondary vertex is reconstructed with selected kaon and pion global tracks. In this analysis, the middle point on the Distance of the Closest Approach (DCA) between two daughter tracks is considered as the secondary decay vertex of the candidate D^0. As shown in Fig. ??, 5 geometrical variables are chosen to select D^0 and reject combinatorial background, which is dominated by a pair of tracks directly from the primary vertex: decay length (the distance between the decay vertex and Primary Vertex PV), DCA between the 2 daughters, DCA between the reconstructed D^0 flying path and PV, DCA between the π track and PV, and DCA between the K track and PV. The cuts on these variables are optimized by the Toolkit for Multivariate Data Analysis (TMVA) package. They change according to the D^0 candidate p_T in order to have the best significance in all the covered p_T range. Additionally there is a $cos(\theta) > 0$ cut to make sure the decay vertex with respect to the primary vertex is roughly in the same direction as the momentum.
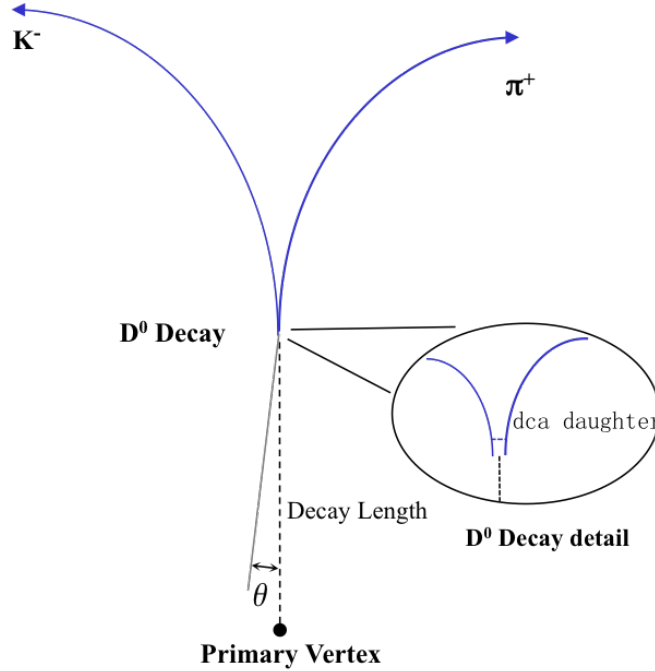

Figure 20: The topology of a D^0 decaying to a kaon and a pion.

The TMVA need signal and background sample input for training. The signal sample is obtained from a toy fast-simulation and the background sample is from real data like sign pairs in D^0 mass window and unlike sign pairs in side bands range.

Fig. ?? shows distributions of the 5 geometry variables for signal (blue) and background (red) plotted by the TMVA, for p_T between 2 and 3 GeV/c.

The 'cuts' option of TMVA is used to tune D^0 cuts. This option randomly sample different cut sets in the variable space, calculate signal and background efficiency for each cut set. Then one cut set with lowest background efficiency at certain signal efficiency. We can then pick the cut set with the best significance according to the signal and background yield corresponding to the whole data set. Fig. ?? shows the lowest background efficiency, significance and so on vs. signal efficiency for p_T between 2 and 3 GeV/c. We can see that as cuts get tighter, signal and background efficiency both decrease, but background efficiency decreases much faster.
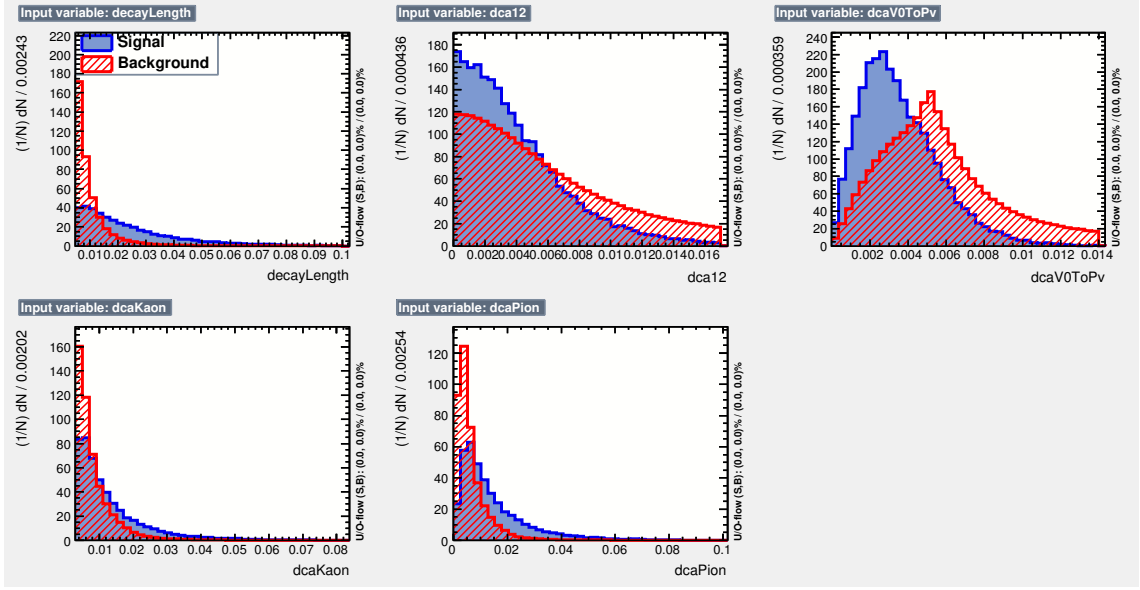
Figure 21: Distributions of the 5 geometry variables for signal (blue) and background (red).
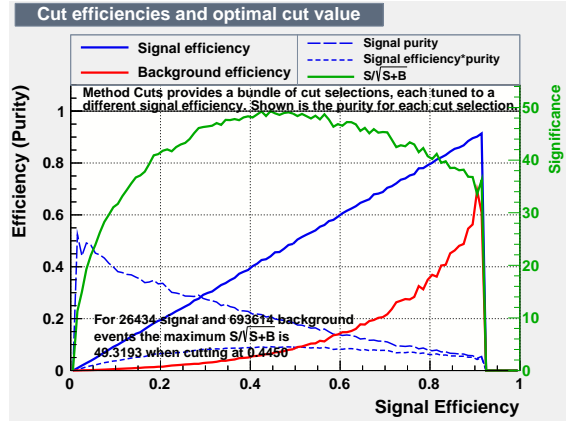

Figure 22: Signal efficiency, lowest background efficiency, significance and so on vs. signal efficiency.

Table 4: Standard geometrical cuts for different D^0 p_T.

D^0 p_T (GeV/c)	0-1	1-2	2-3	3-5	5-10
decay length (μm) >	145	181	212	247	259
DCA between 2 daughters (μm) <	84	66	57	50	60
DCA between D^0 and PV (μm) <	61	49	38	38	40
DCA between π and PV (μm) >	110	111	86	81	62
DCA between K and PV (μm) >	103	91	95	79	58

Table 5: Tight geometrical cuts for different D^0 p_T.

D^0 p_T (GeV/c)	0-1	1-2	2-3	3-5	5-10
decay length (μm) >	144	204	242	245	300
DCA between 2 daughters (μm) <	69	48	44	49	47
DCA between D^0 and PV (μm) <	44	36	31	26	32
DCA between π and PV (μm) >	120	102	118	109	96
DCA between K and PV (μm) >	119	110	109	106	80

Table 6: Loose geometrical cuts for different D^0 p_T.

D^0 p_T (GeV/c)	0-1	1-2	2-3	3-5	5-10
decay length (µm) >	110	168	187	199	180
DCA between 2 daughters (µm) <	77	78	74	68	66
DCA between D^0 and PV (µm) <	72	53	47	42	62
DCA between π and PV (µm) >	92	78	86	65	47
DCA between K and PV (µm) >	105	68	80	66	41

The result of the geometry cuts tuned for best significance are shown in Table **??**. These are the standard cuts used in the D^0 reconstruction to calculate the spectra central value.

For D^0 estimation, another 2 sets of geometry cuts are tuned with TMVA, with 50% and 150% signal efficiency relative to the standard cuts. They do not give the overall best D^0 significance, but for the certain signal efficiency, they are still the cuts with the lowest background efficiency and best D^0 significance. With 50% and 150% signal efficiency relative to the standard cuts, their significance is still about 80% of the standard cuts with the overall best significance. These 2 cuts sets are listed in Table **??** and **??**.

For more details can be found in the D^0 v_2 technicl note, basically we use the same cuts for spectra analysis and v_2 analysis.

3.3   Mixed Event Background

To construct the mixed event background it is important to combine events with some degree of similarity, such as events are classified according to the position of the primary vertex (PV) along the beam-line, the centrality class and the orientation of the event plane. Ten bins of equal width were used for both the event plane ($\Psi \in [-\pi, \pi]$) and the position of the primary vertex($V_z \in [-6, 6]$), as well as nine centrality classes between 0-80%, for a total of 900 event 'categories'.

Table **??** summarizes the important information saved for the event mixing:

Table 7: Summary of information saved for the event mixing

StMixerTrack	StMixerEvent
Origin	PV Origin
Momentum	Magnetic Field
Q-Vector	Event Plane
Track information	Array of mixer tracks
	Array of indices to identified pions
	Array of indices to identified kaons

Fig. **??** and Fig. **??** show the invariant mass distribution for the foreground and two different uncorrelated backgrounds: same event like-sign and mixed event unlike-sign in two p_T bins include 1-2 GeV/c and 4-5 GeV/c. The mixed event backgrounds have been scaled to the foreground using the integration range $m_{K\pi} \in [1.6, 2.1]$ GeV/c^2.

There is good agreement between the two descriptions of the combinatorial background and they appear to provide an adequate description in the vicinity of the D^0 signal and the mixed event backgrounds have improved statistical precision.

It is interesting to observe the presence of an 'excess' in the foreground, relative to all of the background curves, below roughly 1.75 GeV/c^2. This so called bump was investigated using the Data Driven Fast Simulator, and will be covered briefly in the following section.
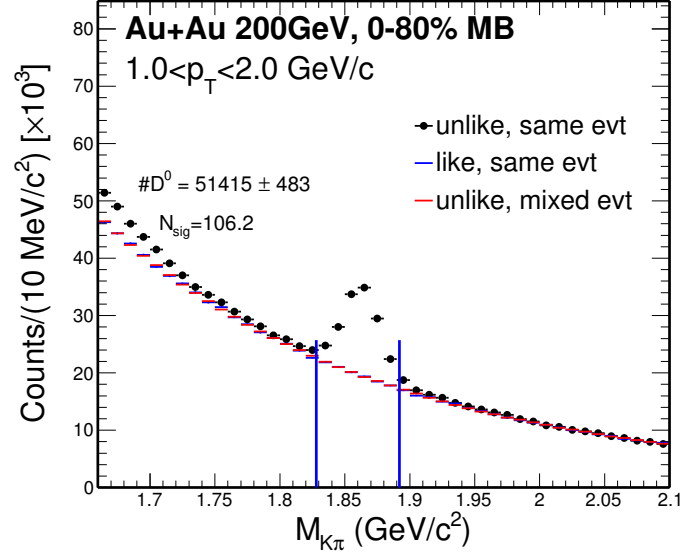
Figure 23: Invariant mass distribution for foreground and two descriptions of combinatorial background in $1 < p_T < 2$ GeV/c.
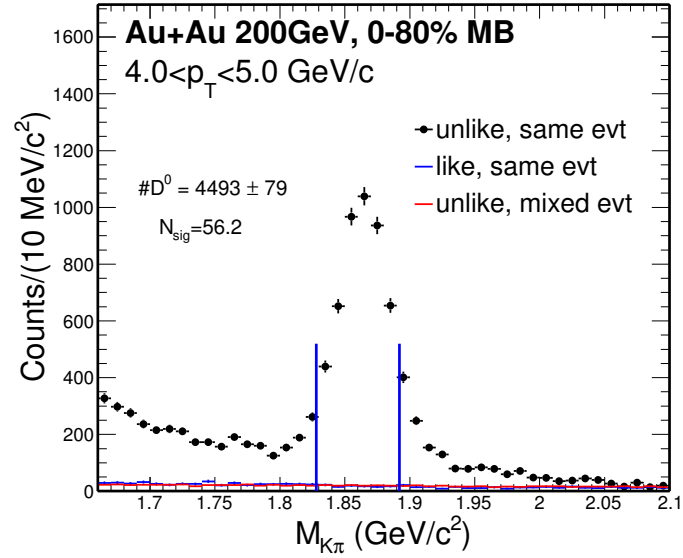

Figure 24: Invariant mass distribution for foreground and two descriptions of combinatorial background in $4 < p_T < 5$ GeV/c.

3.4 Correlated background 'bump' for D^0 meson

The contributions to the invariant mass spectrum from the following D^0 and $D^\pm$ decays were included in a qualitative study of the correlated background:

- $D^0 \rightarrow K^- \pi^+$ (B.R. 0.039)

- $D^0 \rightarrow K^- \pi^+ \pi^0$ (B.R. 0.011)

- $D^0 \rightarrow K^- \rho^+ \rightarrow K^- \pi^+ \pi^0$ (B.R. 0.108)

- $D^0 \rightarrow K^{*-} \pi^+ \rightarrow K^- \pi^+ \pi^0$ (B.R. 0.007)

- $D^+ \rightarrow K^- \pi^+ \pi^+$ (B.R. 0.073×0.415)

The charm fragmentation ratio used is the following from ZEUS Collaboration (arXiv:hep-ex/0508019 - Table 4):

- f$(c \rightarrow D^+) = 0.217$

- f$(c \rightarrow D^0) = 0.523$

- f$(c \rightarrow D_s^+) = 0.095$

- f$(c \rightarrow \Lambda_c^+) = 0.144$

- f$(c \rightarrow D^{*+}) = 0.200$
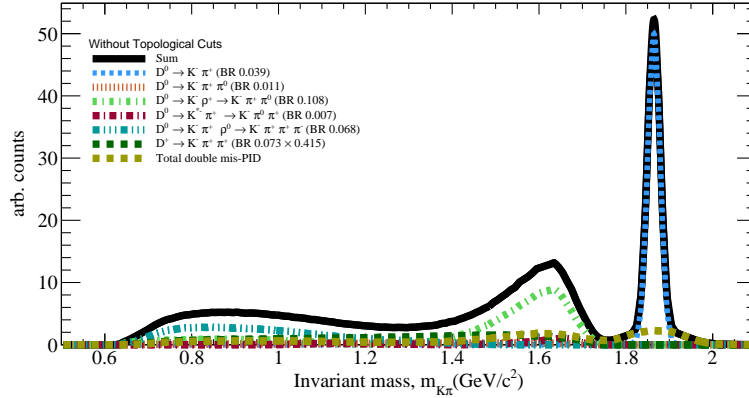
Figure 25: Simulated contribution to the invariant mass spectrum from cocktail without topological cut

Fig. **??** and Fig. **??** show the invariant mass spectrum obtained from the cocktail after scaling by the branching ratio for different decays as well as the fragmentation ratio for the different charmed meson species.

The spectrum is shown before and after the $D^0 \rightarrow K\pi$ topological cuts have been applied. It is clear that the contributions from correlated background can, at least in part, account for the enhancement observed below roughly 1.7 GeV/c^2.

The cocktail simulation was then scaled by fitting the amplitude of the D^0 peak obtained from fast simulator to the signal observed in data, and the cocktail was then added to the mixed event background. Fig. **??** and Fig. **??** shows a comparison between the invariant mass distribution obtained from data and the spectrum obtained by combining the mixed event background and the results from the data-driven Fast-Simulator.

The inclusion of correlated background sources can qualitatively describe the foreground observed, reproducing the location of the bump structure albeit underestimating the degree of enhancement itself. Furthermore, there is likely a finite contribution to the observed bump originating from jet correlations which should be included to improve on the description of the background.
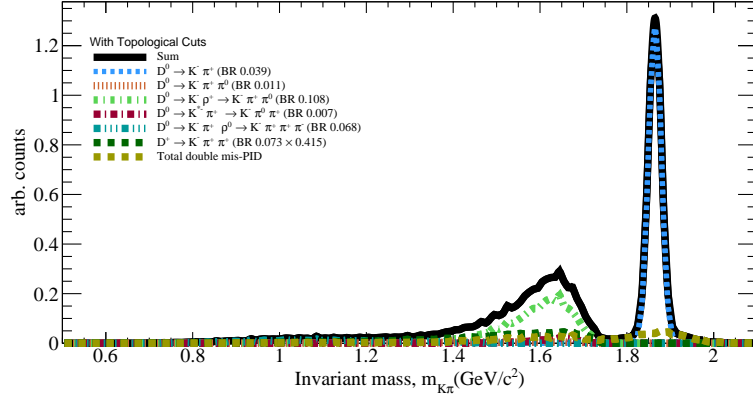
Figure 26: Simulated contribution to the invariant mass spectrum from cocktail with topological cut
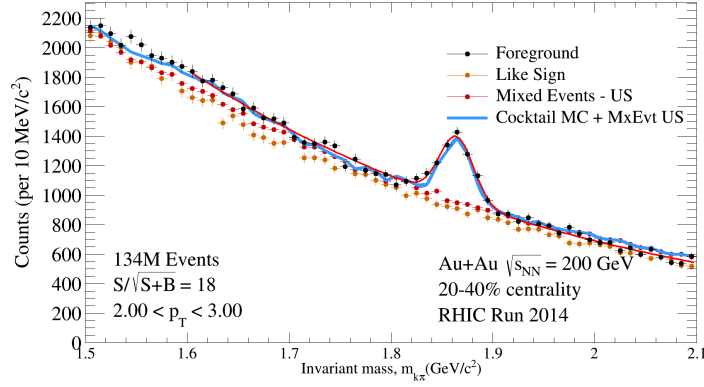

Figure 27: Comparison of $K\pi$ invariant mass distribution for unlike-sign (US) foreground, like-sign combinatorial background, unlike-sign (US) mixed events combinatorial background, and unlike-sign (US) mixed events combinatorial background + toy montecarlo cocktail for correlated background, for $2 < p_T < 3$ GeV/c.
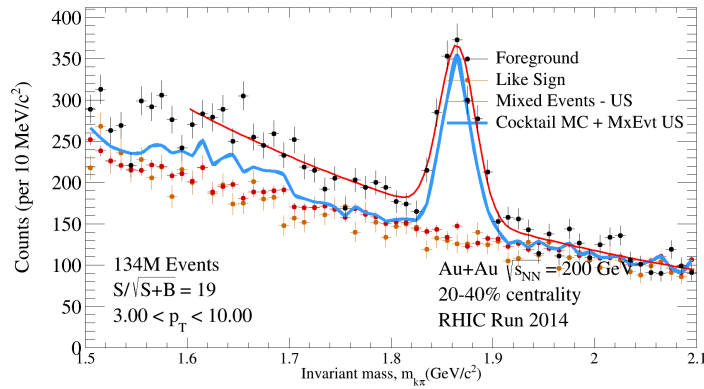

Figure 28: Similar Comparison of $K\pi$ invariant mass distribution as Fig. **??**, for $3 < p_T < 10$ GeV/c.

15

It should also be noted that the studies presented here were done with an early version of the fast simulator which only included the p_T and centrality dependence of sampled distributions, revisiting the studies with more differential distributions should improve on these results.

Nonetheless, the results provide confidence in a qualitative understanding on the sources of the correlated background and, what is more, suggest that the contribution from these source in the D^0 signal range is dominated by double mis-PID, and is nearly negligible as shown in the following sub-section.

4 Efficiency Correction

General idea.....

To obtain the real invariant mass spectrum of D^0 within STAR acceptance ($|\eta_\pi| \leq 1, |\eta_K| \leq 1, |Y_{K\pi}| \leq 1$), the raw spectrum should correct for the efficiency. The $K\pi$ pair efficiency within STAR acceptance is evaluated by folding the TPC related efficiency to the HFT related efficiency as shown on Eq. **??**. For the TPC related tracking efficiency shows on the first term, we use STAR standard Full GEANT simulation. For the HFT related efficiency include the second and third terms which reflect to HFT acceptance and topological cuts, we developed the 'Data-Driven Fast simulation' which will discuss later.

$$\text{Efficiency} \times \text{Acceptance} = \text{TPC Tracking Eff} \otimes \text{HFT Tracking Eff} \otimes \text{Topollogy Cuts} \quad (7)$$

This formula can be written in another way when we consider the particles identification (PID) Eq. **??**. Here the PID part are able to factorized as Eq. **??**, Eq. **??**.

$$\frac{\text{HFT}}{\text{MC}} = \frac{\text{TPC}_{\text{withPID}}}{\text{MC}} \otimes \frac{\text{HFT}_{\text{withPID}}}{\text{TPC}_{\text{withPID}}} \otimes \text{PID} \quad (8)$$

$$\varepsilon(HFT\&PID_{TPC}\&PID_{TOF}) = \varepsilon(HFT|_{PID_{TPC}\&PID_{TOF}}) \times \varepsilon_{PID}(TPC\&TOF) \quad (9)$$

$$\varepsilon_{PID}(TPC\&TOF) = \varepsilon_{PID}(TOF|TPC) \times \varepsilon_{TPC} \quad (10)$$

4.1 Single Track Efficiency

The single track efficiency losses have two contributions, the detector inefficiency and particle identification cuts. The detector efficiency includes the TPC tracking efficiency (ε_{TPC}) and the TOF matching efficiency (ε_{TOF}). The particle identification cut efficiency (ε_{PID}) includes the efficiencies of TOF velocity ($1/\beta$) and the dE/dx selection cuts. So the single track efficiency can be derived by the Eq. **??**

$$\varepsilon = \varepsilon_{TPC} \times \varepsilon_{TOF} \times \varepsilon_{PID} \quad (11)$$

4.2 TPC Tracking Efficiency

The TPC tracking efficiency (ε_{TPC}) is evaluated via the standard STAR embedding technique. TPC efficiency including two parts, TPC response and acceptance efficiency. The Monte Carlo (MC) tracks are embedded into the raw data to have a realistic detector occupancy environment. The raw data is randomly sampled over the entire Au+Au minimum-bias data set, while the number of embedded MC tracks is constrained to 5% of the measured multiplicity of the real events to avoid a sizable impact on the realistic TPC tracking efficiency. The MC tracks, with flat p_T, η, and ϕ, are generated and passed through the full GEANT simulation of the STAR detector geometry, and then mixed with the real data . The mixed signals are processed using the same procedures as real data. The quality assurance is made to ensure the MC simulation reproduces the real data before studying the TPC tracking efficiency (Embedding QA). The TPC tracking efficiency is derived by taking the ratio of the number of reconstructed MC tracks (N_{rec}), satisfying the track quality cuts used in the data analysis, over the number of embedded MC tracks (N_{emb}), as shown in Eq. **??**

$$\varepsilon_{TPC} = \frac{N_{rec} \, (nHitsFit \geq 20 \ \& \ dca \leq 1 \ \& \ |\eta| \leq 1 \ \& \ nCommonHits > 10)}{N_{emb} \, (|\eta| \leq 1)} \quad (12)$$

The TPC tracking efficiency in Run14 Au+Au collisions at 200 GeV is shown below. Fig **??** shows the TPC tracking efficiency for pion plus from four different classifications, from up to down, the centrality is from the most peripheral to most central collision. As we see, in the most central top 0-5% collisions, due to the large occupancy the TPC tracking efficiency is much lower than the central one. Fig. **??** shows the same plot for kaon minus. The Kaon can be decay inside TPC, that's the reason the TPC tracking efficiency is lower than pion.
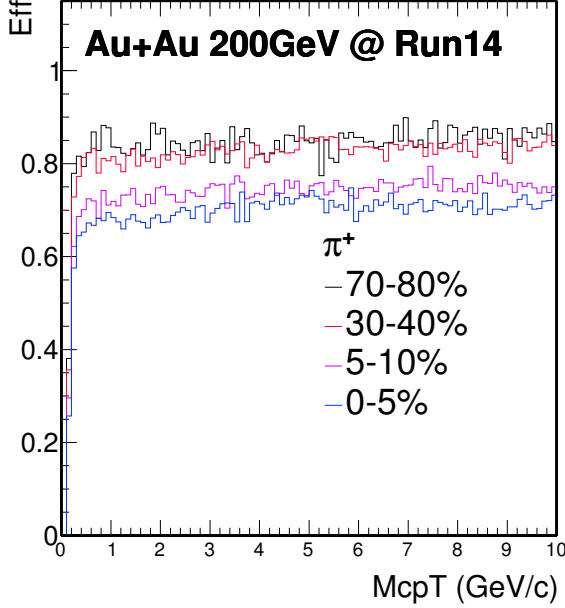
Figure 29: TPC tracking efficiency in Run14 Au+Au collisions at 200 GeV for Pion.
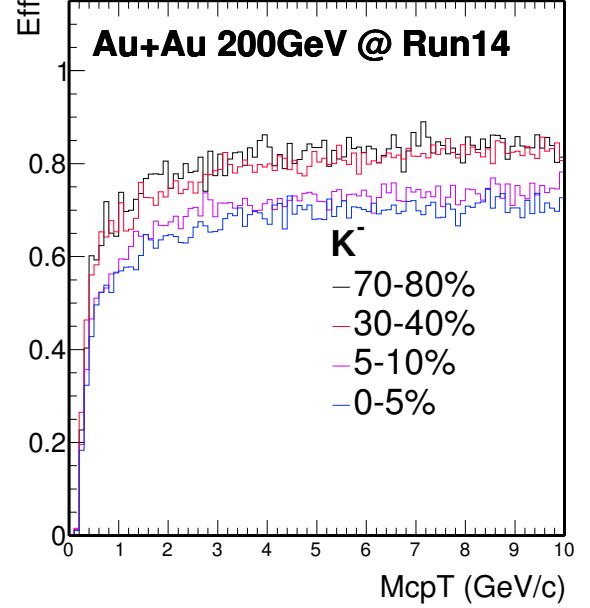
Figure 30: TPC tracking efficiency in Run14 Au+Au collisions at 200 GeV for Kaon.

4.3 TOF Matching Efficiency

For the D^0 analysis, we use the hybrid PID for TOF. Which means when TOF is available we use TOF and TPC, when it's not available we just use TPC. The TOF matching efficiency (ε_{TOF}), including the TOF response and the acceptance difference between the TPC and TOF, is evaluated by the real data. It can be calculated by comparing the number of qualified tracks matched with the TOF (with $\beta > 0$, $N_{matched}$) over the number of qualified tracks (N_{TPC}).

Fig. **??** shows the TOF match efficiency in Run14 Au+Au collisions at 200 GeV for positive charge particles such as π^+, K^+ in the centrality 0-10%. Fig. **??** shows the same plots for negative charge particles in the centrality from 40-80%. For the pion TOF match efficiency, the tread is much smooth compare to kaon. As we see, there are some deep for the TOF match efficiency at some certain p_T range from kaon, such as kaon in the p_T around 0.6 GeV/c. This effect was studied using Hijing simulation, it's found that this is due to the hadron contaminations.

Fig. **??** shows the $n\sigma_K$ distributions from Hijing in the p_T range from 0.5 - 0.7 GeV/c. The solid lines are for particles from TPC , and the dashed lines are for particles also include TOF match information. The total yield was scaled to have the same number of pions for this comparison, since the TOF matching have $\sim$30-40% efficiency lost. In a simple case, if we select kaons with the cut $|n\sigma_K| < 1$, after the requiring of TOF match,the width of this $n\sigma_{K/\pi}$ distribution changed, and the contaminations from pions are reduced. As what we see the dashed black line have less contributions to the kaons peak within $|n\sigma_K| < 1$ compare to the solid black line. We also checked this effect in the other p_T range such as 0.2 - 0.5 GeV/c and 1.0 - 1.5 GeV/c as shown in Fig. **??** and Fig. **??**. This effect is neglectable in the low p_T range 0.2 - 0.5 GeV/c since the TPC can well separate pions and kaon. In the high p_T range, the dE/dex are overlap with each other for kaon and pion, it's not able to distinguish them only use TPC.
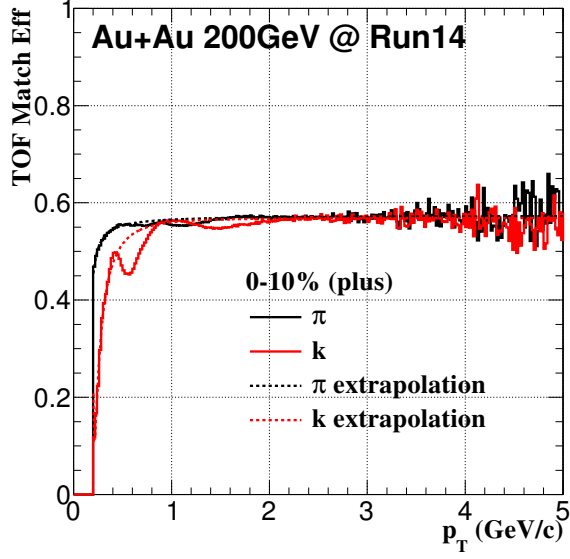
Figure 31: TOF match efficiency in Run14 Au+Au collisions at 200 GeV for positive charge particle in 0-10%.
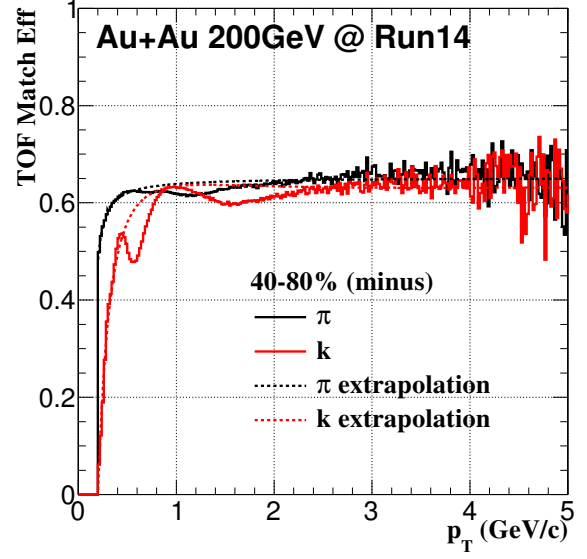
Figure 32: TOF match efficiency in Run14 Au+Au collisions at 200 GeV for negative charge particle in 40-80%.
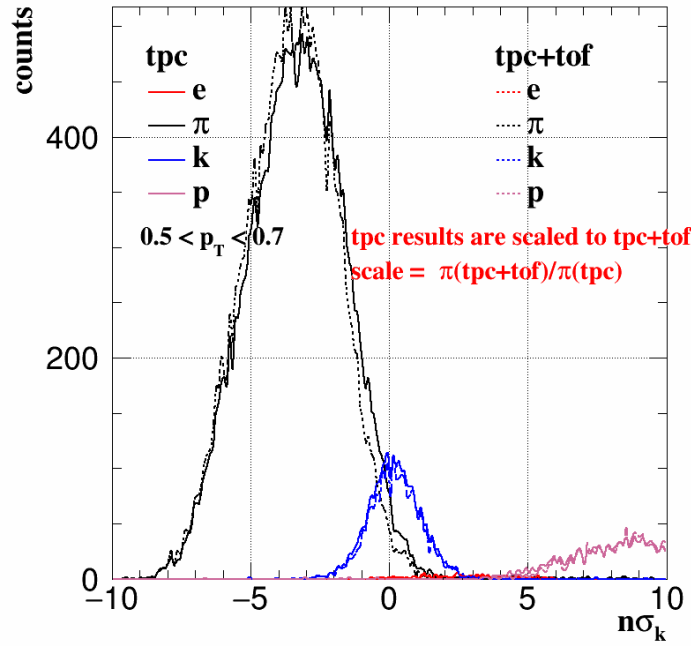

Figure 33: $n\sigma_K$ distributions for $0.5 < p_T < 0.7$ GeV/c. The solid lines are from TPC and dashed lines are from TPC + TOF.
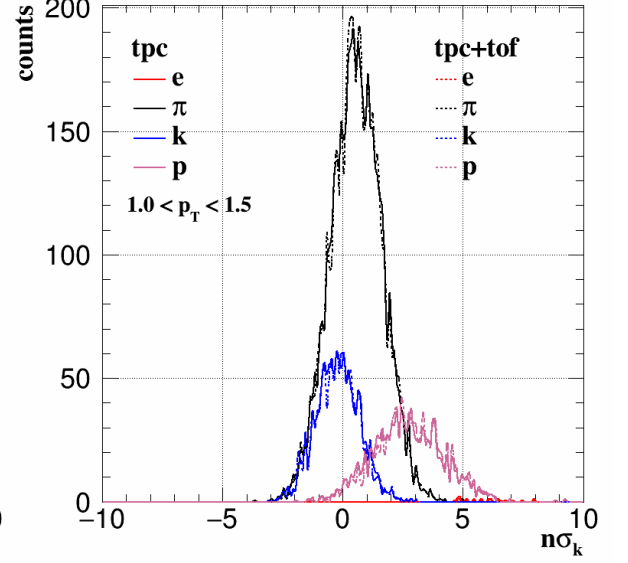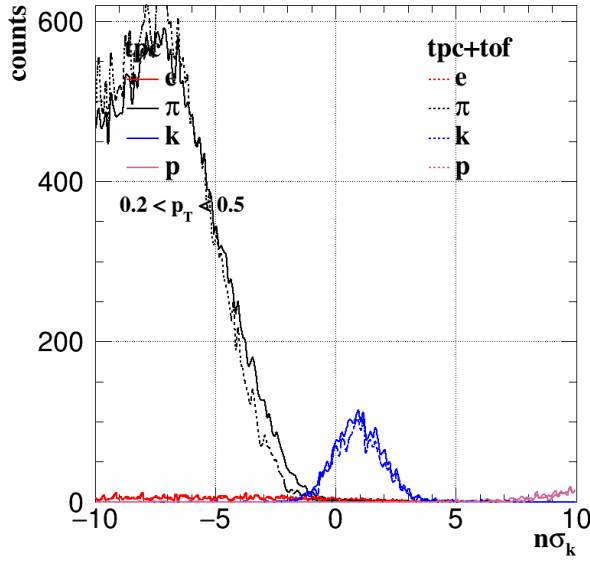
Figure 34: $n\sigma_K$ distributions for $0.2 < p_T < 0.5$ GeV/c. The solid lines are from TPC and dashed lines are from TPC + TOF.

Figure 35: $n\sigma_K$ distributions for $1.0 < p_T < 1.5$ GeV/c. The solid lines are from TPC and dashed lines are from TPC + TOF.

4.4 PID Cut Efficiency

The particle identification cut efficiency (ε_{PID}) includes two components: the TOF velocity ($1/\beta$) cut efficiency and dE/dx cut ($n\sigma_{K/\pi}$) efficiency. Pure pions and kaons sample are used to evaluate the TOF velocity cut efficiency and TPC $n\sigma_{K/\pi}$ cut efficiency. Fig. ?? shows the $\pi\pi$ pairs invariant mass distributions. The black line is the unlikesign foreground, and the red line is background using likesign method. With this K_s^0 candidates, we can statistical extract the pure pion sample for the PID efficiency study. Fig. ?? shows the KK pairs invariant mass distributions. Still with the unlikesign and likesign method, the Φ meson candidates are reconstructed, and we can statistical extract the pure kaon sample for the PID efficiency study.

4.4.1 $n\sigma_{K/\pi}$ Cut Efficiency

The $n\sigma_{K/\pi}$ cut efficiency is derived from the Gaussian fit using those pure samples. The $n\sigma_{K/\pi}$ distributions are fitted with Gaussian function, the mean value and sigma value are plotted as Fig. ?? and Fig. ??. With these mean and sigma distributions, assuming they follow the Gaussian function, for example, Fig. ?? depicts the $n\sigma_\pi$ cut efficiency in Run14 Au + Au collisions at 200 GeV for pions. For kaons, we use the same method extracting this $n\sigma_K$ cut efficiency.

4.4.2 $1/\beta$ Cut Efficiency

The $1/\beta$ cut efficiency is also derived from the Gaussian fit using those pure samples. The efficiency is very high since the $1/\beta$ cut is very loose, it's almost $\sim$100% for this analysis as shown on Fig. ?? and Fig. ??.

4.5 Data-driven fast Monte Carlo setup for HFT and Topological Cut Efficiency

As discussed in the beginning of this section, the HFT related efficiency shown on Eq. ?? including two items: HFT acceptance and topological cuts. Since the HFT embedding is not ready yet at that time, we developed the 'Data-Driven Fast simulation' for the HFT related efficiency correction. And this method was validated with full GEANT simulation and will be discuss later.
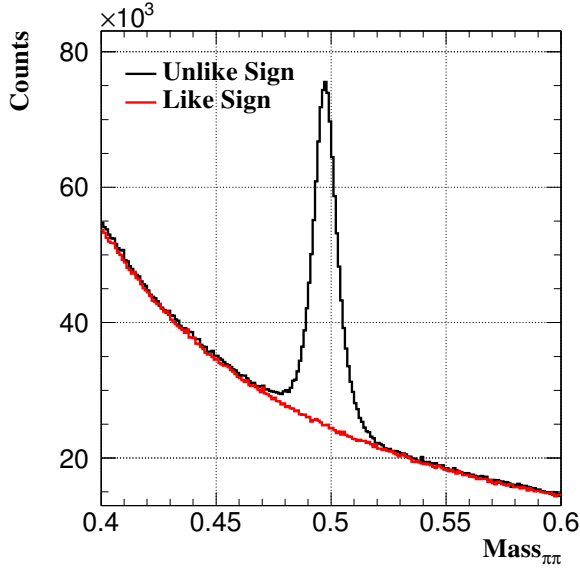
Figure 36: The $\pi\pi$ pairs invariant mass distributions. The black line is the unlikesign foreground, and the red line is background using likesign method.
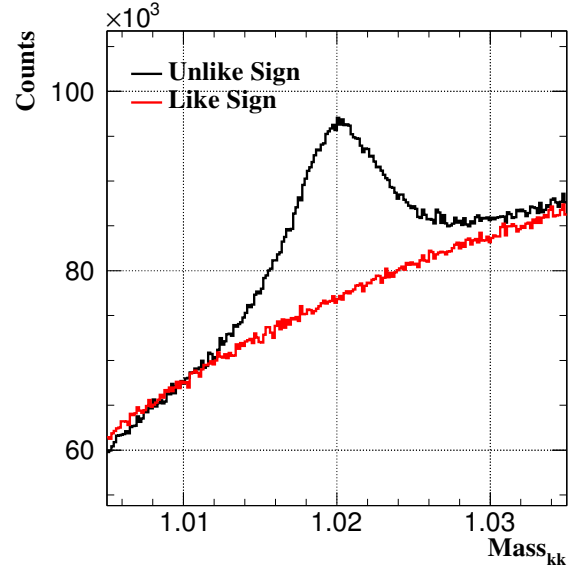

Figure 37: The KK pairs invariant mass distributions. The black line is the unlikesign foreground, and the red line is background using likesign method.
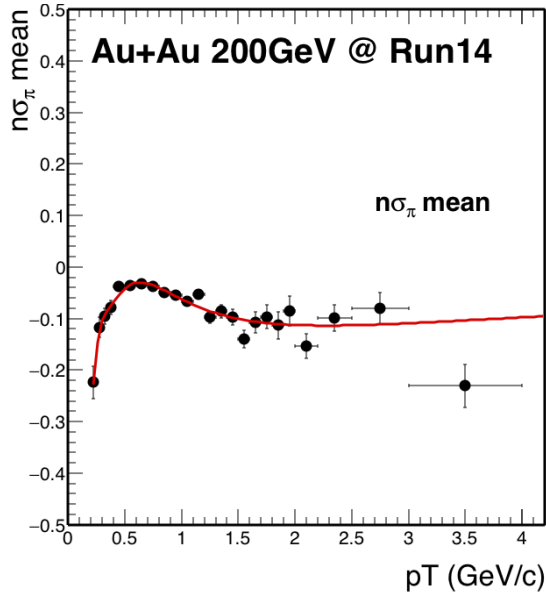

Figure 38: The mean value of $n\sigma_\pi$ distributions vs momentum. The red line is fitted function with polynomial function.
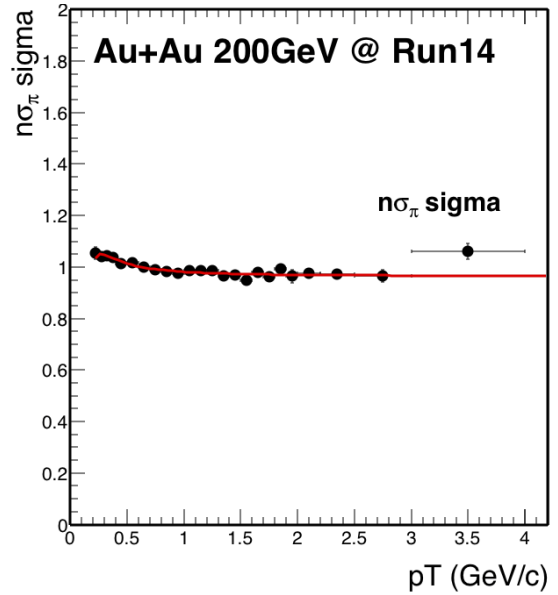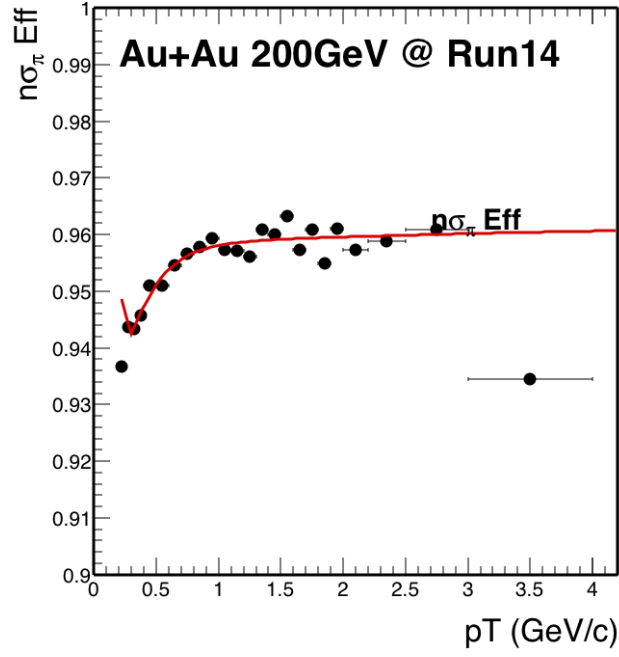

Figure 39: The sigma value of $n\sigma_\pi$ distributions vs momentum. The red line is fitted function with polynomial function.

Figure 40: $n\sigma_\pi$ cut efficiency along with momentum. Red line is the fitted polynomial function
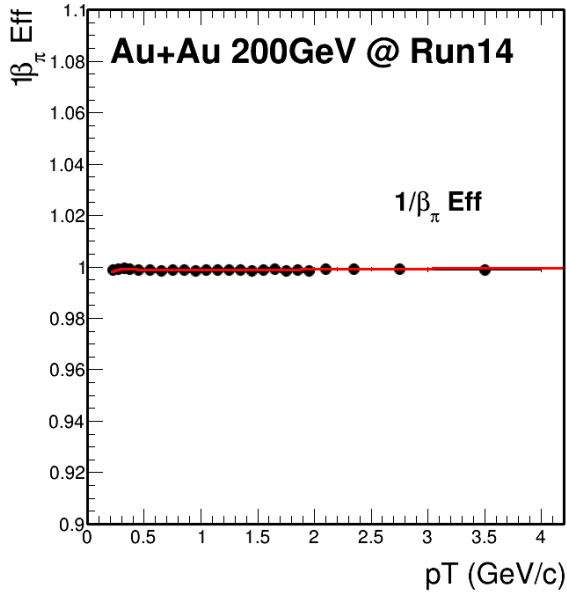

Figure 41: $1/\beta$ cut efficiency along with momentum for pion. Red line is the fitted function.
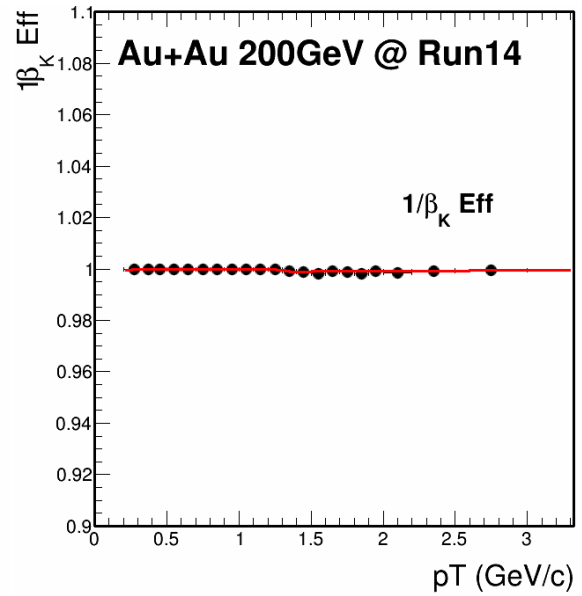

Figure 42: $1/\beta$ cut efficiency along with momentum for kaon. Red line is the fitted function.

4.5.1 Assumptions

Before we discuss the details procedure of the method, it's better to make it clear, this data-driven simulation is based on several assumptions. And the validating will be tested step by step in the later section.

- Factorization of tracking efficiency:

$$\frac{\text{HFT}}{\text{MC}} = \frac{\text{HFT}}{\text{TPC}} \times \frac{\text{TPC}}{\text{MC}} \tag{13}$$

- Spatial resolution of HFT is encoded in two variables: DCA_{XY} and DCA_{Z} (two dimensions correlated).

- Vertex resolution, which is possibly folded in the DCA resolution of single tracks and correlated for kaons and pions, is a negligible, at least for semi-central to central events.

- The contribution of feed-down particles from secondary decays to DCA distributions is negligible.

- Mis-matched daughter tracks are removed by topological cuts.

4.5.2 Ingredients

There are several input ingredients for this fast-simulation package which is extracted from data.

- Extract V_z distributions from data (centrality dependent).

- Extract ratio of HFT matched tracks to TPC tracks from data (This ratio includes all mismatched tracks) (particle species, centrality, p_T, η, ϕ, V_z dependence).

- Extract DCA_{XY} - DCA_{Z} distributions from data. Assuming that data DCA distributions is dominated by primary particles (particle species, centrality, p_T, η, V_z dependence).

- Extract TPC efficiency and momentum resolution from embedding (particles, centrality and p_T dependence).

Fig. ?? shows an example of the input HFT match ratio in the certain η, V_Z, ϕ and centrality range. The HFT match ratio increase in the low pt range due to the high mismatch occupancy and keep flat in the high p_T range. This ratio have a strong dependence on these differential such as η and V_Z since this is effected by HFT acceptance. Fig. ?? shows an example of the Dca_{XY} vs Dca_{Z} distribution in the certain p_T, η, V_Z and centrality range. The axis binning is dynamic binning (non-uniform) since the most central (around 0) part is the dominate part. Limited by the computing memory, the most central part use fine binning and others use the unrefined binning as shown on the plots.

In total, there are 11 (ϕ) × 10 (η) × 6 (V_z) × 9 (centrality) × 2 (particles) 1D histograms (36 p_T binning) for HFT match efficiency. There are 5 (η) × 4 (V_z) × 9 (centrality) × 2 (particles) histograms × 19 (p_T) 2D histograms (144 × 144 Dca binning) for Dca resolution smearing.

Effectively, these 1D and 2D histograms encode HFT efficiency, acceptance and spatial resolution performance in Run14 data.

4.5.3 Recipe

After all the input ingredients ready for the fast-simulation, a simple toy MC simulation (PYTHIA/EvtGen) is applied for the efficiency study. The basic recipe is following:

- Sample V_z distribution according to data distribution.

- Generate D^0 flat in p_T and rapidity and decay it.

- Smear momentum according the embedding.

- Smear DCA_{XY} and DCA_{Z} of Kaons and Pions independently according to distributions from data.
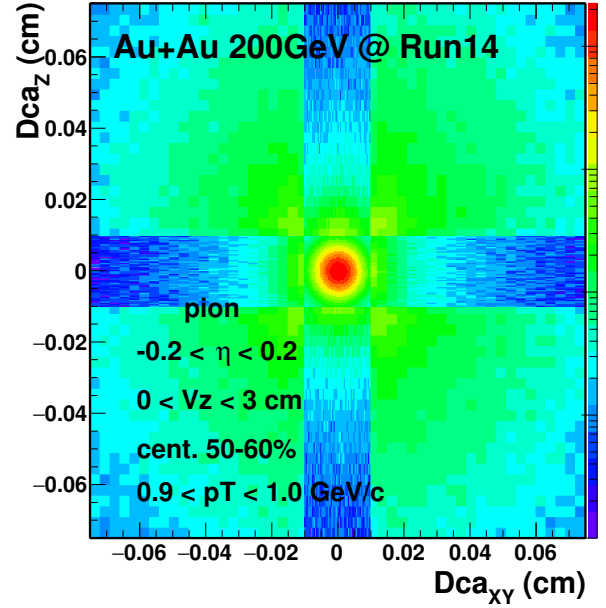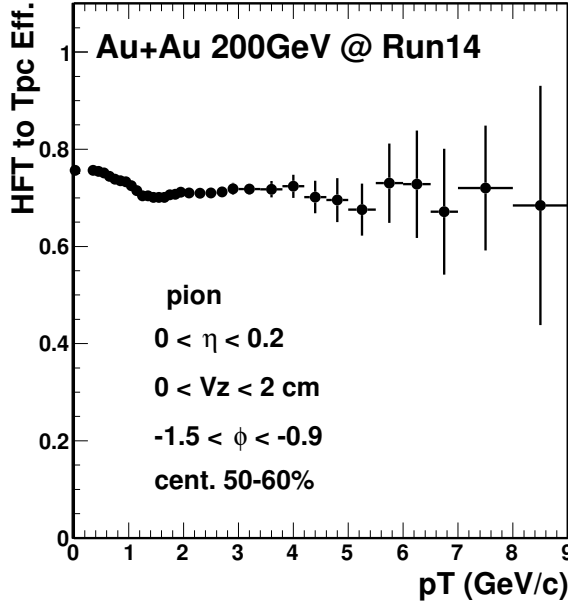
Figure 43: HFT to TPC track match ratio for pion at certain η,V_Z,ϕ,centrality range.

Figure 44: Dca$_{XY}$ vs. Dca$_Z$ for pion at certain η,V_Z,centrality range.

- Apply HFT matching efficiency from HFT ratio.
- Apply TPC reconstruction efficiency.
- Reconstruct D^0

4.5.4 D^0 Efficiency and Topological Distribution

As discussed in the recipe, we obtain the efficiency step by step as shown on Fig. **??**. First we have the TPC efficiency shown by red maker which is after the p_T, η acceptance cut and TPC tracking efficiency from embedding. Then after folding in the HFT matching efficiency, the second item is obtained on black circle. Last step is after the topological cut, as shown by the cyan marker. As see in the low p_T part, the topological cut efficiency is really small due to the tight cut as the combinatorial background is huge.

We study all the efficiencies with small centrality bin width, in total we have 9 centrality bins from our StRefmultCorr class. Since the D^0 production is scaled by the number of binary collisions (N_{bin}), the D^0 is favor produced in more central collisions. So the final efficiencies for the wider centrality bins 0-10%, 10-40%, 40-80% and 0-80% are calculated using N_{bin} as weights, for example, the efficiency in 0-80% is calculated as the following Eq. **??**. Fig **??** shows the D^0 efficiency for 4 wide centralities after TPC, HFT match and Topological efficiency included.

$$\text{Efficiency}_{0-80\%} = \sum_{i=1}^{9} (\text{Efficiency}_i \times N_{bin}^i)/ < N_{bin} > \qquad (14)$$

The Data-Driven Fast-Simulation also provide the topological information, can be used for the comparison with real data. For the real data part, within the D^0 mass window we can statistical subtract the background and extract the pure D^0 topological distributions. The invariant mass plots shown as Fig. **??**. D^0 is in the $2 < p_T < 3$ GeV/c, 0-80% centrality. Black is unlikesign foreground, blue is likesign background and red is mixed event background. The blue vertical lines are the mass window used for the topological comparison. For each topological variable, that corresponding topological cut was removed when reconstruct the D^0 candidate, so that we can compare that variable in a wide range.

From Fig. **??** to Fig. **??**, these are the topological variables ($\cos(\theta)$, decayLength, dcaDaughters, D0DcaToVtx, PionDca and KaonDca) used for the D^0 reconstruction. The distributions from the
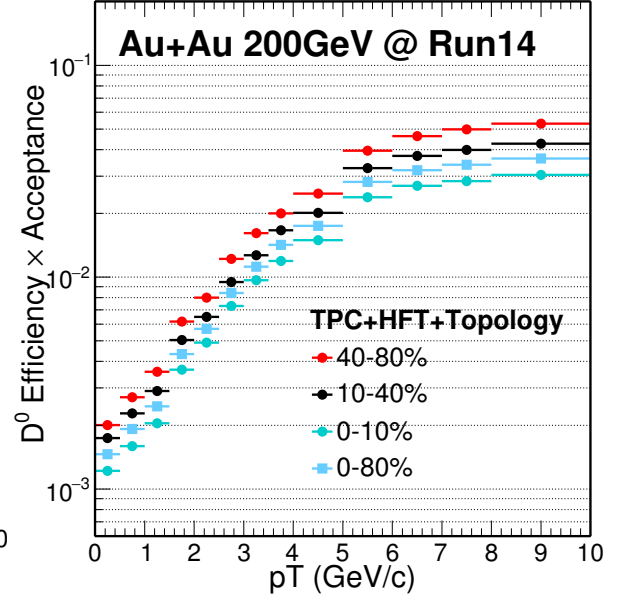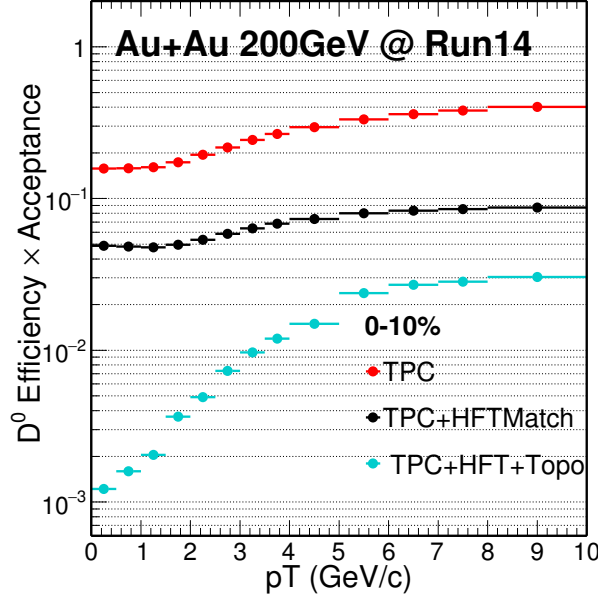
Figure 45: D^0 efficiency step by step from TPC, HFT Ratio, Topological cut in most central 0-10%.

Figure 46: D^0 efficiency including TPC, HFT and Topological cut in several centralities.
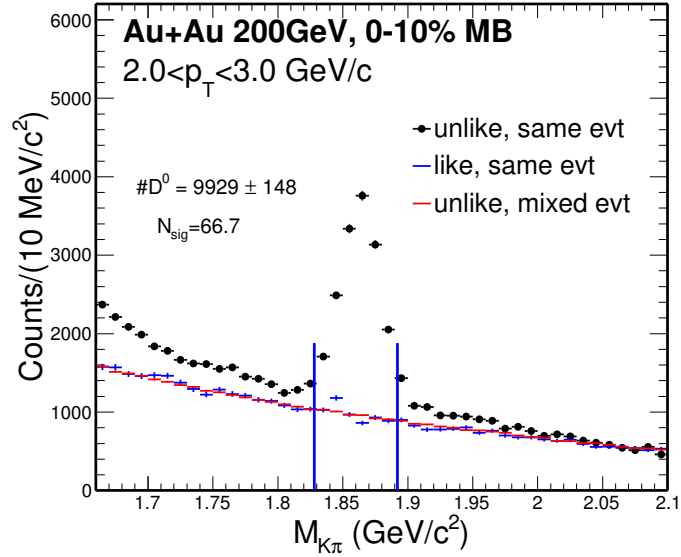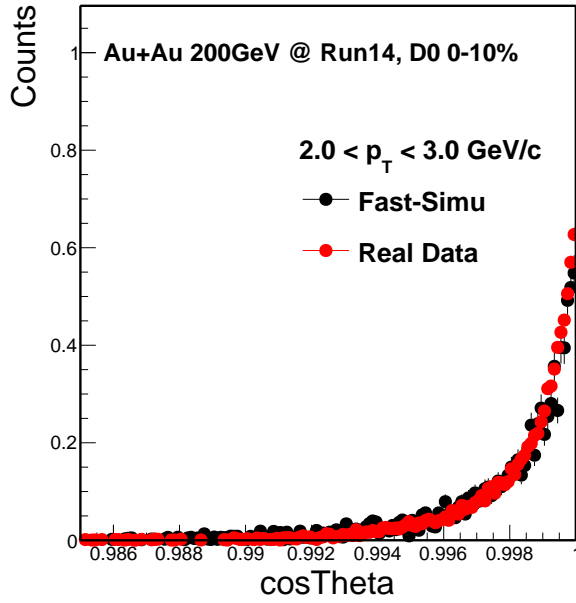

Figure 47: D^0 invariant mass distributions in the $2 < p_T < 3$ GeV/c, 0-10% centrality. Black is unlikesign foreground, blue is likesign background and red is mixed event background. The blue vertical lines are the mass window used for the topological comparison.

Figure 48: D^0 cosTheta distribution in most central 0-10% between Fast-Simulation and Real Data.
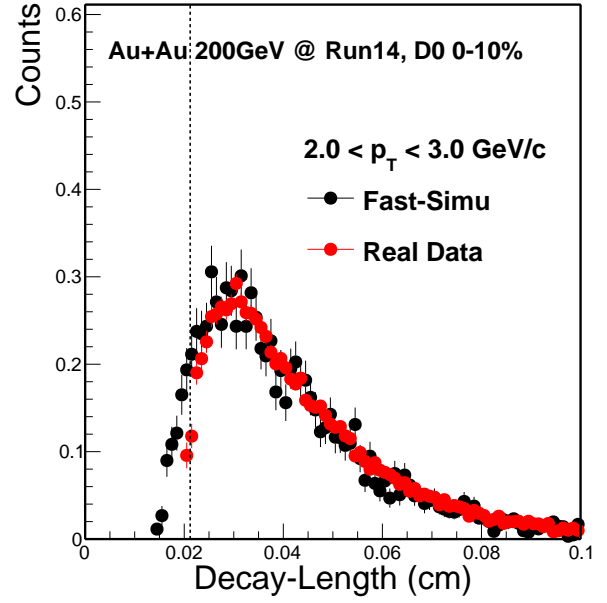

Figure 49: D^0 decay length distribution in most central 0-10% between Fast-Simulation and Real Data.
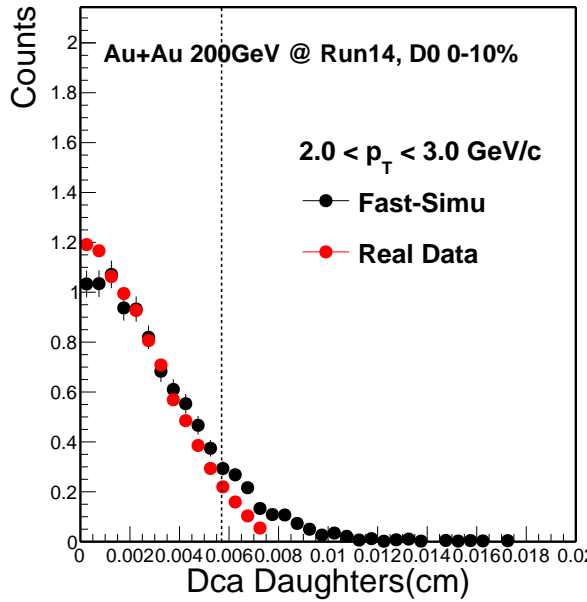

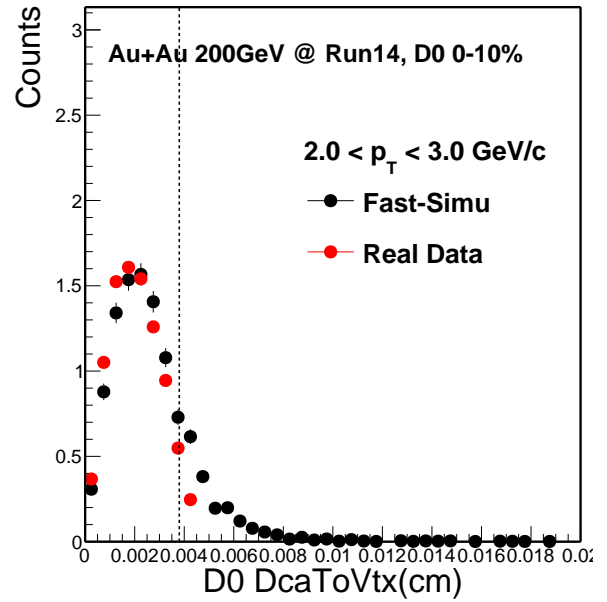Figure 50: D^0 dcaDaughters distribution in most central 0-10% between Fast-Simulation and Real Data.

Figure 51: D^0 dca to Vertex distribution in most central 0-10% between Fast-Simulation and Real Data.
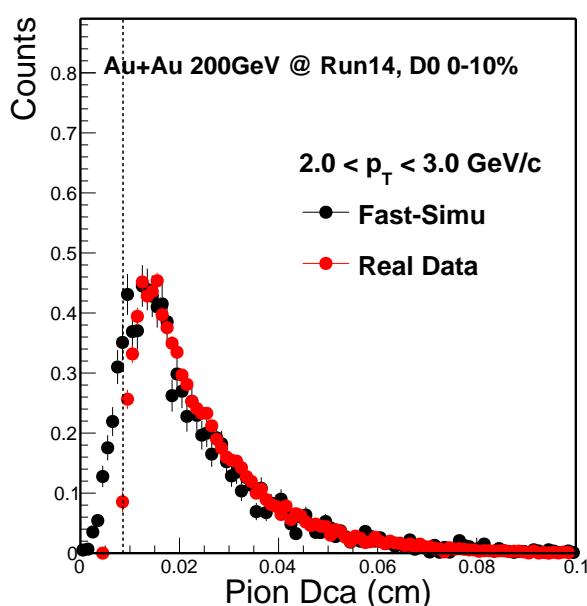
Figure 52: D^0 pionDca distribution in most central 0-10% between Fast-Simulation and Real Data.
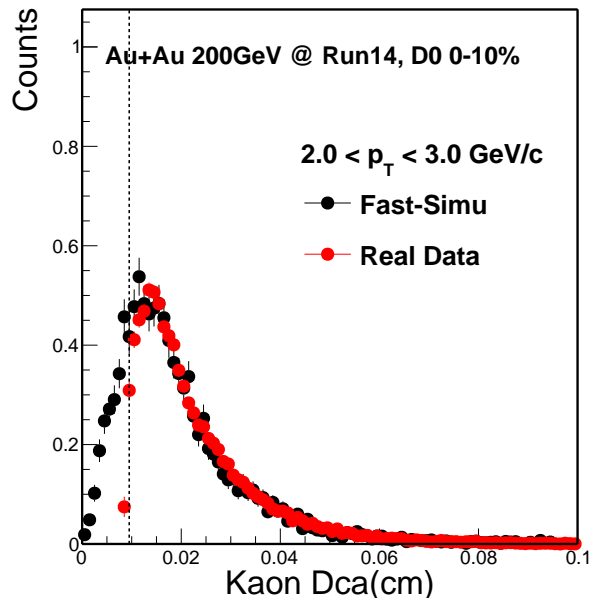
Figure 53: D^0 kaonDca distribution in most central 0-10% between Fast-Simulation and Real Data.

real data part are using mixed event method to statistical subtract the background. The Data-Driven Fast-Simulation part was the package we relayed on for our efficiency study as shown before.

Comparison these topological variables between real data and Fast-Simulation, the agreement is reasonable good, which means our Data-Driven Fast-Simulation method can well reproduce the topological variables in real data. In another word, the efficiency estimation from Data-Driven Fast-Simulation is reliable. Note, there are some small discrepancy such as single track Dca distributions in the low end, that's because in the data analysis part, we already require some minimum cut in order to save the computing resource. There is another method we are going to discuss in the following section, also can be used to validating our Fast-Simulation method.

4.6 Validation with Full GEANT+Hijing Simulation

Before discuss the details of this Hijing validation, it's better to conclude those assumptions we made before. The first assumption is the factorization shown in Eq. ??. Relaying on the Hijing simulation, we have two samples. One is only include TPC tracking, another one include both HFT and TPC in tracking. From the first sample, we can extract the TPC factorized tracking efficiency, and the second sample can be used to extract the overall total efficiency and HFT over TPC factorized efficiency separately. Fig. ?? shows the comparison between the overall efficiency and the multiplied factorization efficiency. The red one is from overall efficiency from the second sample, and the blue one is multiplied efficiency from two components. The bottom panel shows the double ratio of these two efficiency, and they are perfectly factorized as the ratio is flat as unity.

The second assumption is for the spatial resolution, it is encoded in those Dca_{XY} and Dca_Z variables, and they are correlated in the two dimensions. Fig. ?? shows the comparison between the input Dca from real data and output Dca from Fast-Simulation in three dimensions. The first row is Dca in XY plane, the second row is in Z plane and the last row is in the 3-D dimension. From the left to right is the comparison from low p_T to high p_T. As shown the red line is from data and black line is from fast simulation, the agreement is pretty good. For the others assumption, they will be discussed separately in the following section.

4.7 PID Efficiency and Double Counting Correction

The PID efficiency part includes the efficiency from dE/dx or $n\sigma$ cut efficiency and the TOF matching+$1/\beta$ cut efficiency. The dE/dx and $1/\beta$ calibration usually uses a specific set of tracks.
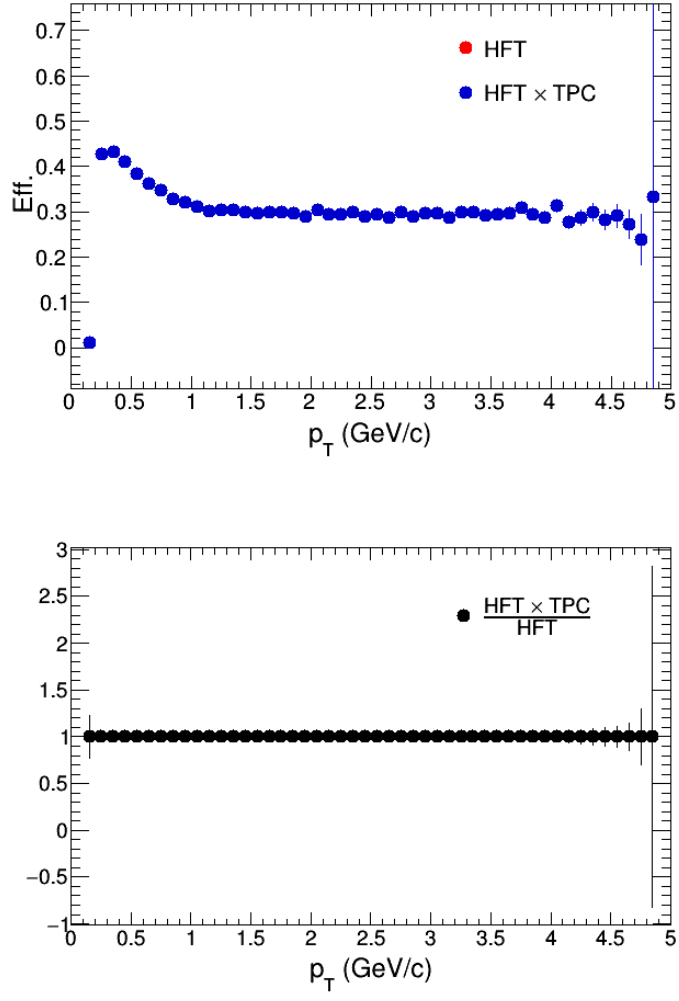
Figure 54: (top)HFT Efficiency Factorization comparison. (bottom) Double Ratio of these factorization.
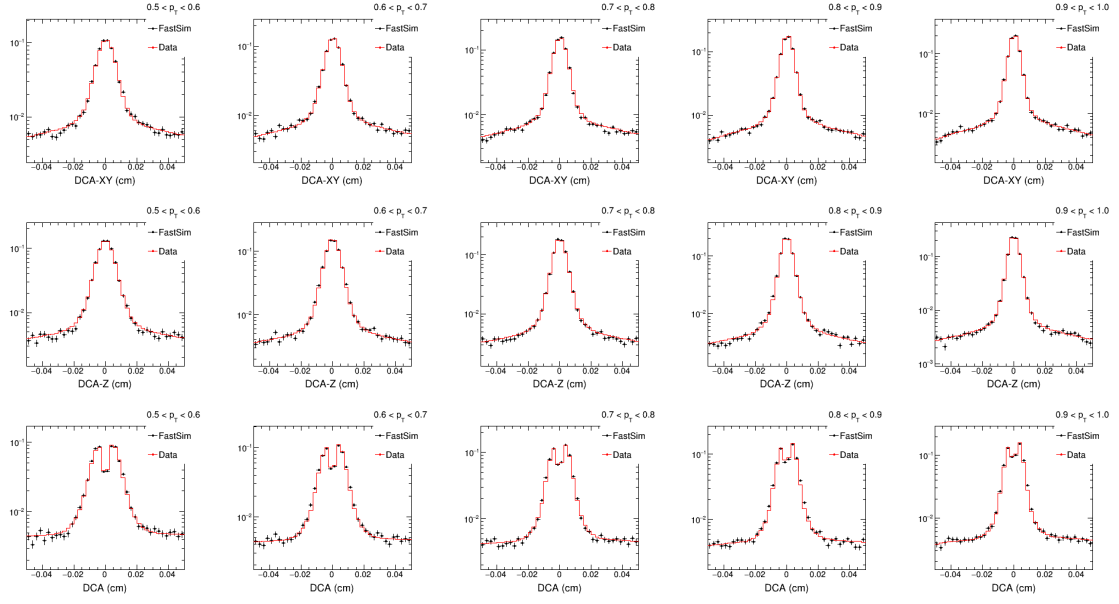
Figure 55: Comparison of Dca between data (red) and Fast-Simulation (black). From top to bottom, the comparison is for Dca_{XY}, Dca_Z and Dca. From left to right the transverse momentum is from low p_T to high p_T.

The performance for identified particles that pass our analysis cut may not necessarily be exactly normal Gaussian distributions with means=0 and widths=1. It is desired to calibrate each individual distributions for identified particles for efficiency correction and mis-PID effect study.

For $n\sigma$ calibration, we followed the same method as described in Ref. [?] to calibrate high p_T pions and protons by selecting daughters from K_S^0 and Λ decays.

4.7.1 dE/dx Calibration

4.8 Vertex Resolution Correction

4.9 Validation with Ks Spectra Measurement

To verify the

5 Results

6 Re-analysis of Run10/11 data

6.1 Long and Yifei's re-analysis
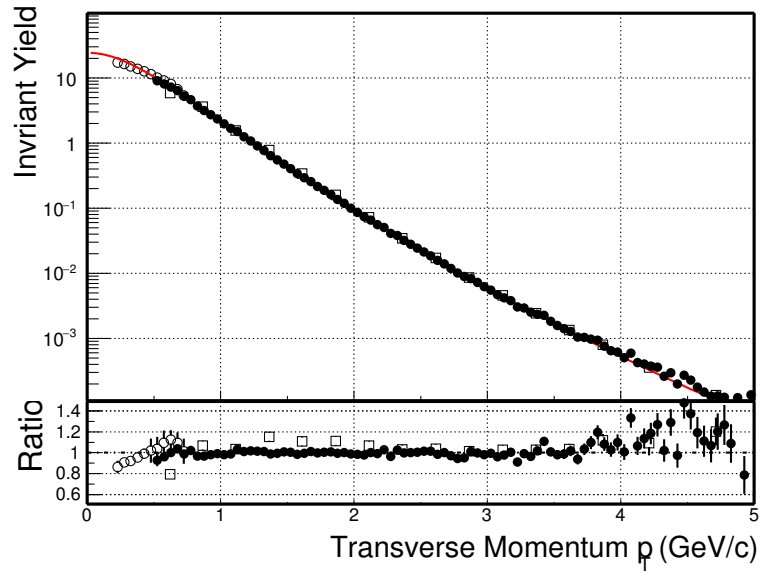
6.2 Xiaolong's re-analysis

7 Run14 TPC analysis

Figure 56: Ks spectra.