

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

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

September 29, 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 measurements of  $D^0$ -meson production at mid-rapidity ( $|y| < 1$ ) in Au + Au collisions at  $\sqrt{s_{NN}} = 200$  GeV utilizing the Heavy Flavor Tracker, high resolution silicon detector at the STAR experiment.  $D^0$ -mesons are reconstructed via their hadronic decay channel  $D^0 \rightarrow K^- + \pi^+$  and its charge conjugate via topological reconstruction of  $D^0$  decay vertices. After corrected for the detector acceptance, tracking and geometric selection efficiency, invariant yields of  $D^0$ -mesons are reported in various centrality bins covering a transverse momentum region of  $0 - 8$  GeV/c. Nuclear modification factors ( $R_{CP}$ ) are obtained and compared to various phenomenological model calculations. Physics implications on charm hadron production and charm quark dynamics in the Quark-Gluon Plasma are discussed.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Datasets and Event Selection</b>	<b>4</b>
2.1	Centrality Definition . . . . .	4
<b>3</b>	<b><math>D^0</math> Reconstruction</b>	<b>10</b>
3.1	Daughter Selection . . . . .	10
3.2	Topological Cut Optimization . . . . .	11
3.3	Mixed Event Background . . . . .	17
3.4	Correlated background ‘bump’ for $D^0$ meson . . . . .	17
<b>4</b>	<b>Efficiency Correction</b>	<b>21</b>
4.1	Single Track Efficiency . . . . .	21
4.2	TPC Tracking Efficiency . . . . .	21
4.3	TOF Matching Efficiency . . . . .	22
4.4	PID Cut Efficiency . . . . .	23
4.4.1	$n\sigma_{K/\pi}$ Cut Efficiency . . . . .	23
4.4.2	$1/\beta$ Cut Efficiency . . . . .	26
4.5	Data-driven fast Monte Carlo setup for HFT and Topological Cut Efficiency . . . . .	26
4.5.1	Assumptions . . . . .	26
4.5.2	Ingredients . . . . .	27
4.5.3	Recipe . . . . .	27
4.5.4	$D^0$ Efficiency and Topological Distribution . . . . .	28
4.6	Validation with Full GEANT+Hijing Simulation . . . . .	29
4.7	Hijing Samples Performance . . . . .	31
4.8	Validation Procedures . . . . .	31
4.8.1	Validation Efficiency . . . . .	31
4.8.2	Validation Topological Distributions . . . . .	38
4.9	Secondary Track Contribution . . . . .	38
4.10	Vertex Resolution Contribution . . . . .	42
4.11	Double Counting Correction . . . . .	44
<b>5</b>	<b>Systematic Uncertainties</b>	<b>51</b>
<b>6</b>	<b>Validation with <math>K_s</math> Spectra Measurement ... To Be Added</b>	<b>53</b>
<b>7</b>	<b>Results ... To Be Added</b>	<b>53</b>
<b>8</b>	<b>Re-analysis of Run10/11 data ... To Be Added</b>	<b>53</b>
8.1	Long and Yifei’s re-analysis ... To Be Added . . . . .	53



Figure 1: Jet flavor tomography level crossing pattern of nuclear modification factors at middle rapidity of  $\pi$ , D, B, e calculations for central Au + Au 200 GeV collisions.



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

## 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. 1 shows the jet flavor tomography level crossing pattern of nuclear modification factors at middle rapidity of  $\pi$ , 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. 2 shows the  $R_{AA}$  of  $D^0$ ,  $\pi$ ,  $h^\pm$  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 1.

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 2 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 \text{ cm}$
$ V_r  < 2 \text{ cm}$
$ V_z - V_z^{VPD}  < 3 \text{ 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. 3 shows the  $\langle g\text{Refmult} \rangle$  as a function of run index for the QA, and there are several

outliers are identified. Those dashed lines are the  $4 \times \text{RMS}$  range, beyond those range, the runs are identified as bad run. And several iterations are did until all the runs are within these  $4 \times \text{RMS}$  range.

In the Fig. 4 shows the  $\langle \text{HFT } p_T \rangle$ , 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. 5, which shows the average of HFT matching Ratio in the  $p_T$  range between 0.7 to 0.8  $\text{GeV}/c$  and Fig. 6 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](https://drupal.star.bnl.gov/STAR/system/files/STAR_PXL_Firmware_Issue_Solved_Final_Report_Oct3_v2.pdf)

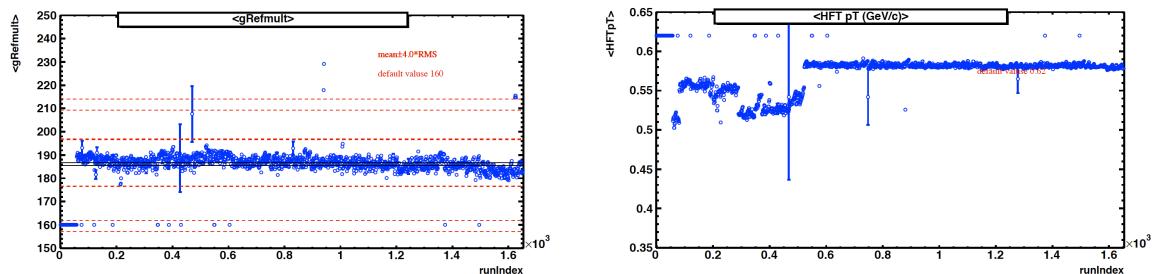


Figure 3: The mean value of gRefmult ( $\langle \text{gRefmult} \rangle$ ) as a function of run index from QA.

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

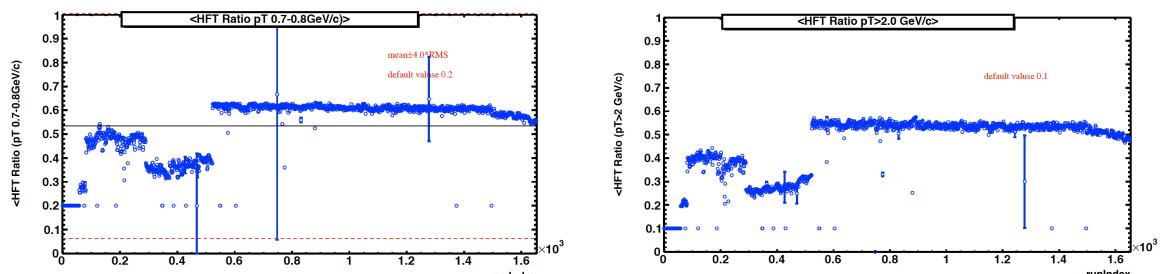


Figure 5: The mean value of HFT matching ratio ( $\langle \text{HFT } p_T \rangle$ ) as a function of run index at the transverse momentum range  $0.7 < p_T < 0.8 \text{ GeV}/c$ . Figure 6: The mean value of HFT matching ratio ( $\langle \text{HFT } p_T \rangle$ ) as a function of run index at the transverse momenauum range  $p_T > 2.0 \text{ GeV}/c$ .

Fig. 7 and Fig. 8 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. 7 while the Fig. 8 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.

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. 9 shows the mean value of gRefmult ( $\langle \text{gRefmult} \rangle$ ) 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. 1,

$$f_{ZdcX} = p0 + p1 * ZdcX \quad (1)$$

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. 2.

$$f_{ZdcX} = \frac{1}{1 + p1/p0 * ZdcX} \quad (2)$$



Figure 7: Normalized gRefmult distribution for Figure 8: Normalized gRefmult distribution for  
VpdMB5 trigger along different  $V_z$  range VpdMB30 trigger along different  $V_z$  range

After the ZdcX correction, this  $\langle g\text{Refmult} \rangle$  is flat as shown by Fig. 10.

For the  $V_z$  dependence correction, we extract the high end point (h) from the fitting of gRefmult tail by the function of Eq. 3.

$$f_x = A * TMath :: Erf(-\sigma * (x - h)) + A \quad (3)$$

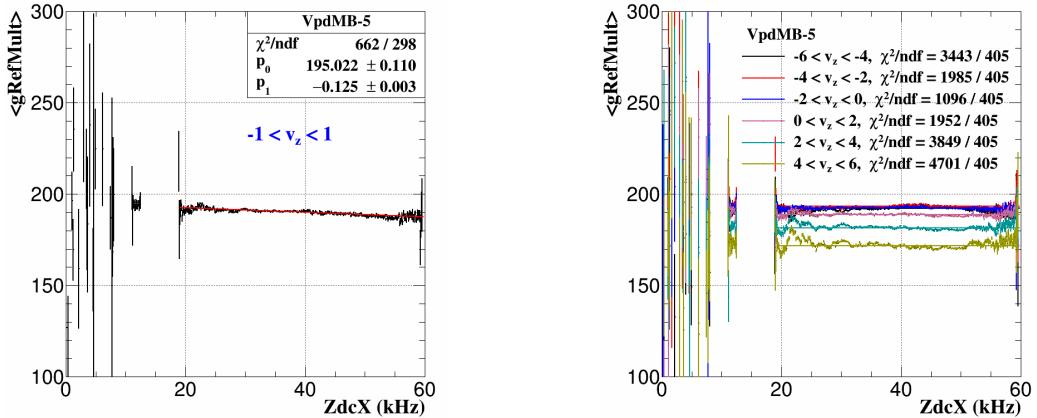


Figure 9:  $\langle g\text{Refmult} \rangle$  as function of ZdcX before Figure 10:  $\langle g\text{Refmult} \rangle$  as function of ZdcX after correction.

Fig. 11 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.

With all the high end point value extracted along  $V_z$  direction, this  $V_z$  dependence was shown on Fig. 12. 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. 4, and then the correction factor was shown by Eq. 5 After the  $V_z$  correction, this high end point is flat as shown by Fig. 13.

$$f_{V_z} = p_0 + p_1 * x + p_2 * x^2 + p_3 * x^3 + p_4 * x^4 + p_5 * x^5 \quad (4)$$

$$f_{V_z} = \frac{p_0}{p_0 + p_1 * x + p_2 * x^2 + p_3 * x^3 + p_4 * x^4 + p_5 * x^5} \quad (5)$$

As shown from Fig. 7 and Fig. 8, 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. 14. After this additional correction, the distributions from VpdMB5 and VpdMB30 are same.



Figure 11: Fitting gRefmult tail distribution in the range of  $1 < V_z < 2$  cm by Eq. 3.

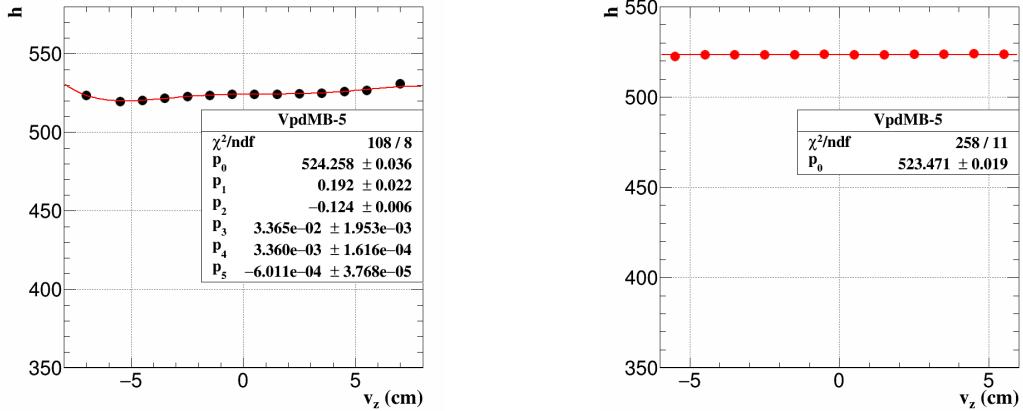


Figure 12: High end point as function of  $V_z$  after Figure 13: High end point as function of  $V_z$  after correction.



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 peripheral events. Fig. 15 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 inefficiency, 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. 16.



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

Figure 16: MC/data gRefmult distribution for Vpd 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. 17. And then the determined centrality classification can be found at Table 3.

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 definition based on gRefMult

### 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  $D^0$ .

#### 3.1 Daughter Selection

$D^0$  have a lifetime of  $c\tau \sim 123\mu\text{m}$ . Thus the global tracks for daughter tracks are used in this analysis. The transverse momentum are required to  $\geq 0.3 \text{ GeV}/c$  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  $\geq 20$  (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 addition, 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  $dE/dx$ ,  $n\sigma_x$  ( $x = \pi, K, p, e$  etc.), defined in Eq. 6, instead of  $dE/dx$  is used in this analysis. Where  $\langle dE/dx \rangle_{\text{measured}}$  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,  $\sigma = 1$ ).

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

Fig. 18 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  $\pi$ ,  $K$ ,  $p$  and  $e$ .

Fig. 19 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  $\pi$ ,  $K$ ,  $p$ .

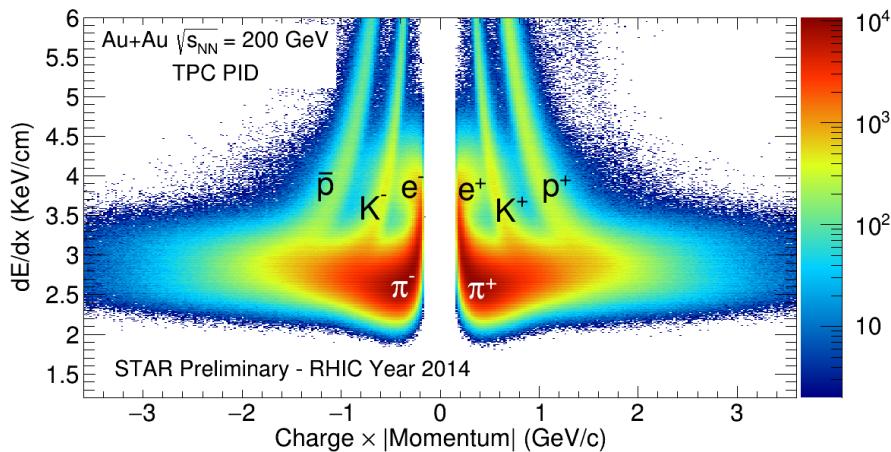


Figure 18: TPC  $dE/dx$  versus charge  $\times$  momentum achieved from Run14 Au+Au 200GeV.

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

- global tracks
- $p_T > 0.3 \text{ GeV}/c$  (next plots are based on 0.3 GeV cut, but the final result central value was from 0.6 GeV cut, same as  $D^0$  v2 analysis)
- $|\eta| < 1$

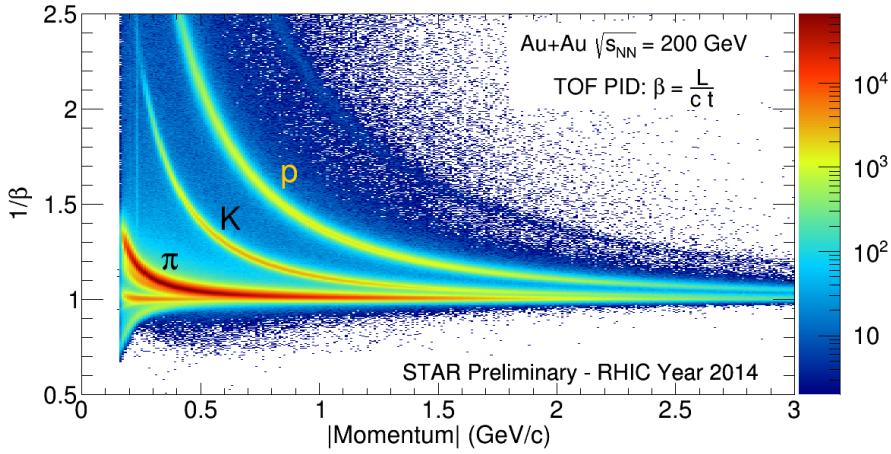


Figure 19: TOF 1/Beta versus momentum achieved from Run14 Au+Au 200GeV.

- $nHitsFit \geq 20$ , 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
- If TOF is available (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 available (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. 20, 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  $\pi$  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  $p_T$  and centrality in order to have the best significance in all the covered  $p_T$  and centrality range. Additionally there is a  $\cos(\theta) > 0.95$  cut to make sure the decay vertex with respect to the primary vertex is roughly close to the same direction as the momentum.

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

Fig. 21 shows an example of the distributions of the 5 geometry variables for signal (blue) and background (red) plotted by the TMVA, for  $p_T$  between 1 and 2 GeV/c and centrality for 0-10%.

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. 23 shows the lowest background efficiency, significance and so on vs. signal efficiency for  $p_T$  between 1 and 2 GeV/c for the centrality 0-10%. We can see that as cuts get tighter, signal and background efficiency both decrease, but background efficiency decreases much faster.

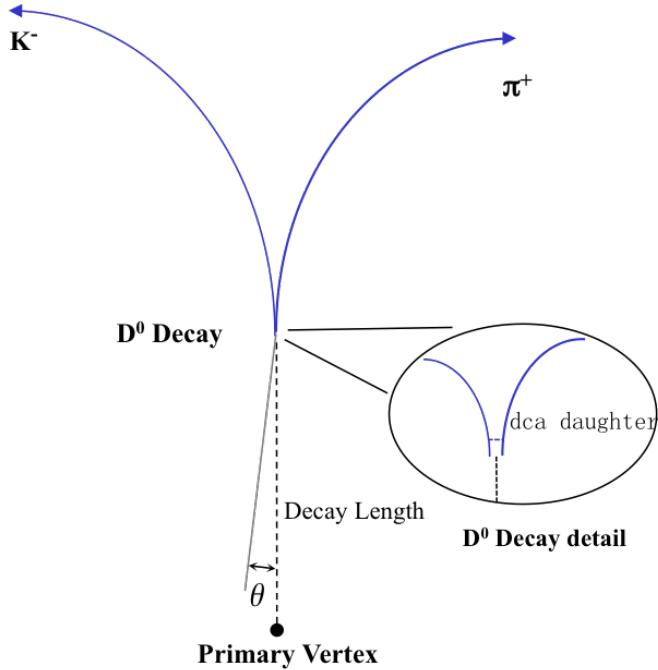


Figure 20: The topology of a  $D^0$  decaying to a kaon and a pion.

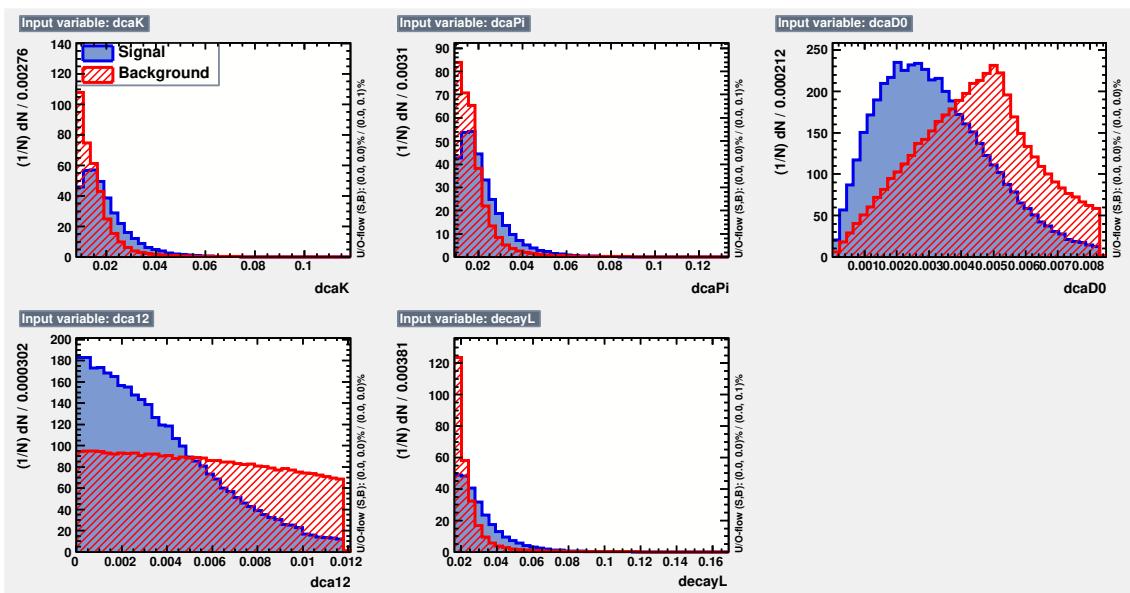


Figure 21: Distributions of the 5 geometry variables for signal (blue) and background (red) from 0-10% and 1-2 GeV/c.

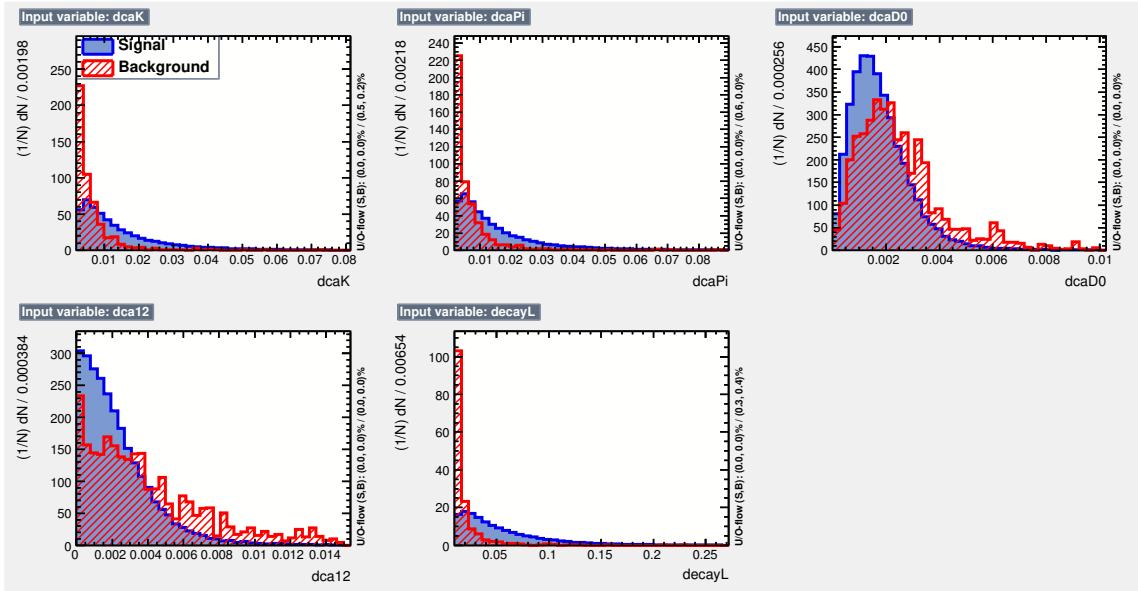


Figure 22: Distributions of the 5 geometry variables for signal (blue) and background (red) from 60-80% and 5-8 GeV/c.

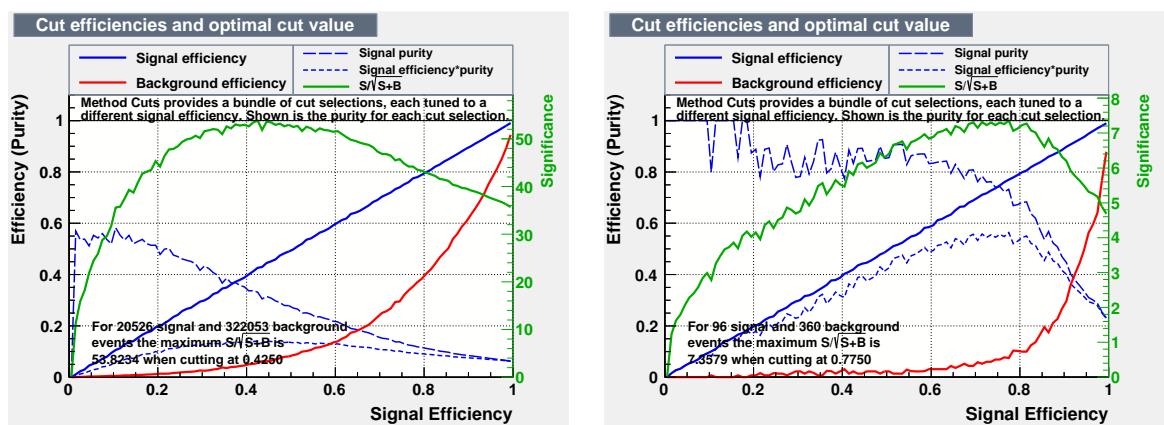


Figure 23: Signal efficiency, lowest background efficiency, significance and so on vs. signal efficiency from 0-10% and 1-2 GeV/c.

Figure 24: Signal efficiency, lowest background efficiency, significance and so on vs. signal efficiency from 60-80% and 5-8 GeV/c.

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

$D^0 p_T$ (GeV/c)	0-0.5	0.5-1	1-2	2-3	3-5	5-8
Centrality 60-80%						
decay length ( $\mu m$ ) >	175	175	187	178	184	187
DCA between 2 daughters ( $\mu m$ ) <	77	77	94	78	81	120
DCA between $D^0$ and PV ( $\mu m$ ) <	76	76	53	54	54	42
DCA between $\pi$ and PV ( $\mu m$ ) >	98	98	83	73	56	34(50)
DCA between $K$ and PV ( $\mu m$ ) >	106	106	69	68	47(50)	31(50)
Centrality 40-60%						
decay length ( $\mu m$ ) >	171	196	210	187	190	214
DCA between 2 daughters ( $\mu m$ ) <	80	83	92	81	94	106
DCA between $D^0$ and PV ( $\mu m$ ) <	72	57	58	49	49	47
DCA between $\pi$ and PV ( $\mu m$ ) >	145	128	72	79	60	51
DCA between $K$ and PV ( $\mu m$ ) >	140	100	75	72	60	42(50)
Centrality 20-40%						
decay length ( $\mu m$ ) >	178	206	221	209	219	240
DCA between 2 daughters ( $\mu m$ ) <	78	73	80	93	96	103
DCA between $D^0$ and PV ( $\mu m$ ) <	66	55	53	46	41	50
DCA between $\pi$ and PV ( $\mu m$ ) >	131	113	99	106	65	52
DCA between $K$ and PV ( $\mu m$ ) >	151	102	104	99	63	42(50)
Centrality 10-20%						
decay length ( $\mu m$ ) >	172	215	252	232	236	237
DCA between 2 daughters ( $\mu m$ ) <	76	78	92	72	86	85
DCA between $D^0$ and PV ( $\mu m$ ) <	63	47	45	46	42	44
DCA between $\pi$ and PV ( $\mu m$ ) >	141	100	74	77	66	52
DCA between $K$ and PV ( $\mu m$ ) >	145	113	94	89	69	48(50)
Centrality 0-10%						
decay length ( $\mu m$ ) >	100	199	227	232	236	255
DCA between 2 daughters ( $\mu m$ ) <	71	64	70	63	82	80
DCA between $D^0$ and PV ( $\mu m$ ) <	62	55	40	40	40	44
DCA between $\pi$ and PV ( $\mu m$ ) >	131	105	93	97	67	55
DCA between $K$ and PV ( $\mu m$ ) >	138	109	82	94	76	54

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

$D^0 p_T$ (GeV/c)	0-0.5	0.5-1	1-2	2-3	3-5	5-8
Centrality 60-80%						
decay length ( $\mu m$ ) >	203	203	206	228	161	216
DCA between 2 daughters ( $\mu m$ ) <	76	76	78	95	76	93
DCA between $D^0$ and PV ( $\mu m$ ) <	58	58	51	36	36	31
DCA between $\pi$ and PV ( $\mu m$ ) >	130	130	130	95	97	86
DCA between $K$ and PV ( $\mu m$ ) >	126	126	116	97	76	39(50)
Centrality 40-60%						
decay length ( $\mu m$ ) >	222	229	269	236	232	182
DCA between 2 daughters ( $\mu m$ ) <	88	59	81	83	73	108
DCA between $D^0$ and PV ( $\mu m$ ) <	56	45	50	37	29	26
DCA between $\pi$ and PV ( $\mu m$ ) >	143	100	72	145	113	95
DCA between $K$ and PV ( $\mu m$ ) >	176	100	123	92	68	56
Centrality 20-40%						
decay length ( $\mu m$ ) >	240	242	268	319	176	338
DCA between 2 daughters ( $\mu m$ ) <	61	43	70	89	66	70
DCA between $D^0$ and PV ( $\mu m$ ) <	62	60	37	38	30	26
DCA between $\pi$ and PV ( $\mu m$ ) >	150	113	63	149	107	105
DCA between $K$ and PV ( $\mu m$ ) >	158	123	133	125	103	53
Centrality 10-20%						
decay length ( $\mu m$ ) >	219	240	213	231	261	399
DCA between 2 daughters ( $\mu m$ ) <	77	49	42	56	53	119
DCA between $D^0$ and PV ( $\mu m$ ) <	56	49	39	36	33	28
DCA between $\pi$ and PV ( $\mu m$ ) >	172	100	144	133	130	92
DCA between $K$ and PV ( $\mu m$ ) >	172	165	119	125	91	57
Centrality 0-10%						
decay length ( $\mu m$ ) >	100	230	268	292	249	225
DCA between 2 daughters ( $\mu m$ ) <	77	44	47	73	60	61
DCA between $D^0$ and PV ( $\mu m$ ) <	55	53	37	27	26	25
DCA between $\pi$ and PV ( $\mu m$ ) >	139	148	93	133	80	88
DCA between $K$ and PV ( $\mu m$ ) >	170	109	117	109	111	59

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

$D^0 p_T$ (GeV/c)	0-0.5	0.5-1	1-2	2-3	3-5	5-8
Centrality 60-80%						
decay length ( $\mu m$ ) >	154	154	163	147	126	140
DCA between 2 daughters ( $\mu m$ ) <	97	97	123	92	93	92
DCA between $D^0$ and PV ( $\mu m$ ) <	83	83	90	77	59	55
DCA between $\pi$ and PV ( $\mu m$ ) >	98	98	80	58	35(50)	20(50)
DCA between $K$ and PV ( $\mu m$ ) >	90	90	73	57	34(50)	20(50)
Centrality 40-60%						
decay length ( $\mu m$ ) >	158	153	172	150	126	164
DCA between 2 daughters ( $\mu m$ ) <	95	84	100	108	130	113
DCA between $D^0$ and PV ( $\mu m$ ) <	69	83	87	77	85	72
DCA between $\pi$ and PV ( $\mu m$ ) >	124	114	72	50	40(50)	35(50)
DCA between $K$ and PV ( $\mu m$ ) >	115	114	67	50	40(50)	35(50)
Centrality 20-40%						
decay length ( $\mu m$ ) >	100	177	177	194	131	150
DCA between 2 daughters ( $\mu m$ ) <	81	90	78	94	137	121
DCA between $D^0$ and PV ( $\mu m$ ) <	75	73	56	53	100	90
DCA between $\pi$ and PV ( $\mu m$ ) >	111	113	89	62	50	35(50)
DCA between $K$ and PV ( $\mu m$ ) >	134	112	74	63	50	35(50)
Centrality 10-20%						
decay length ( $\mu m$ ) >	157	197	222	180	155	189
DCA between 2 daughters ( $\mu m$ ) <	82	86	99	102	121	105
DCA between $D^0$ and PV ( $\mu m$ ) <	71	63	55	52	65	55
DCA between $\pi$ and PV ( $\mu m$ ) >	108	100	74	69	50	37(50)
DCA between $K$ and PV ( $\mu m$ ) >	135	120	70	60	50	35(50)
Centrality 0-10%						
decay length ( $\mu m$ ) >	148	179	215	198	167	206
DCA between 2 daughters ( $\mu m$ ) <	98	87	88	78	100	101
DCA between $D^0$ and PV ( $\mu m$ ) <	78	63	54	45	62	47
DCA between $\pi$ and PV ( $\mu m$ ) >	116	105	93	72	50	41(50)
DCA between $K$ and PV ( $\mu m$ ) >	111	109	80	67	50	35(50)

Fig. 22 and Fig. 24 shows the similar plots for the 5-8 GeV/c in the centrality 60-80%.

The result of the geometry cuts tuned for best significance are shown in Table 4. 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 5 and 6.

For more details can be found in the  $D^0 v_2$  technic 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 7 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. 25 and Fig. 26 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  $m_{K\pi} \in [1.6, 2.1] \text{ GeV}/c^2$ .

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.

### 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)

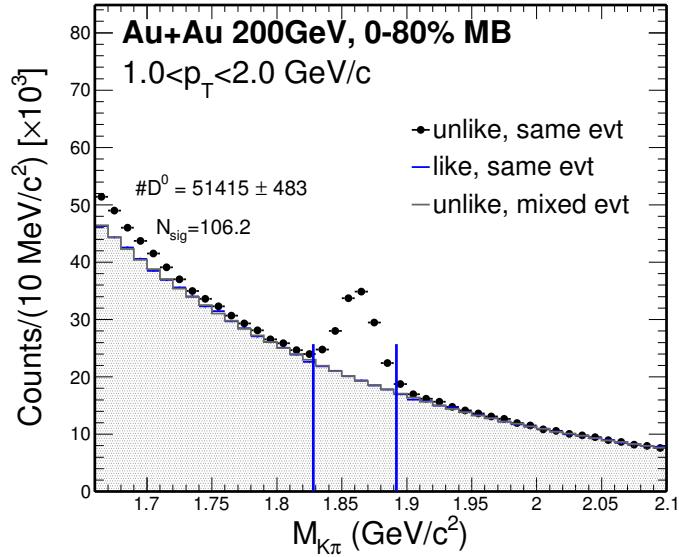


Figure 25: Invariant mass distribution for foreground and two descriptions of combinatorial background in  $1 < p_T < 2 \text{ GeV}/c$ .

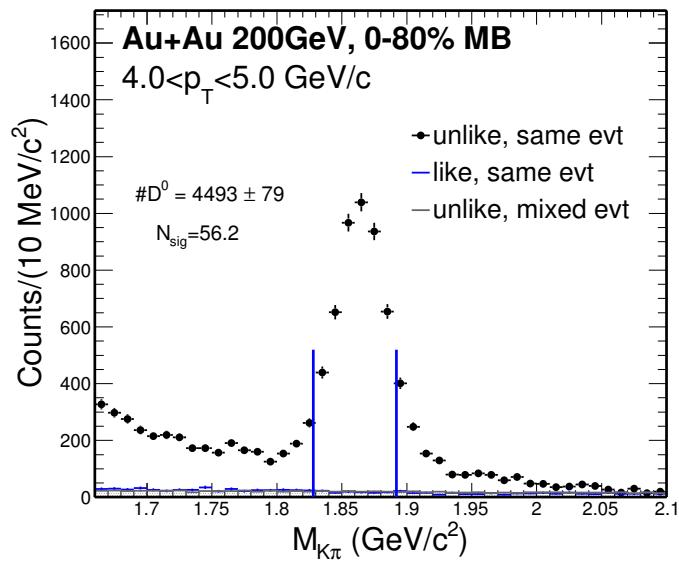


Figure 26: Invariant mass distribution for foreground and two descriptions of combinatorial background in  $4 < p_T < 5 \text{ GeV}/c$ .

- $D^0 \rightarrow K^* - \pi^+ \rightarrow K^- \pi^+ \pi^0$  (B.R. 0.007)
- $D^+ \rightarrow K^- \pi^+ \pi^+$  (B.R.  $0.073 \times 0.415$ )

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

- $f(c \rightarrow D^+) = 0.217$
- $f(c \rightarrow D^0) = 0.523$
- $f(c \rightarrow D_s^+) = 0.095$
- $f(c \rightarrow \Lambda_c^+) = 0.144$
- $f(c \rightarrow D^{*+}) = 0.200$

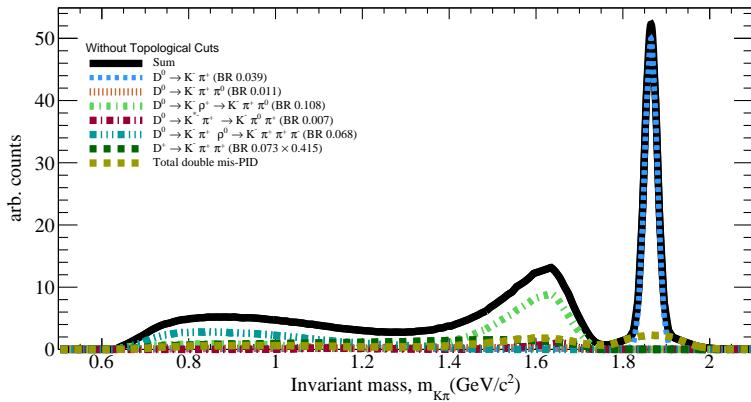


Figure 27: Simulated contribution to the invariant mass spectrum from cocktail without topological cut

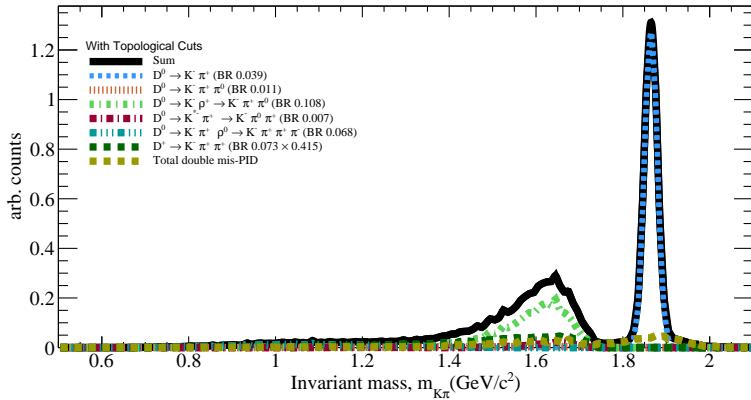


Figure 28: Simulated contribution to the invariant mass spectrum from cocktail with topological cut

Fig. 27 and Fig. 28 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  $1.7 \text{ GeV}/c^2$ .

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. 29 and Fig. 30 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.

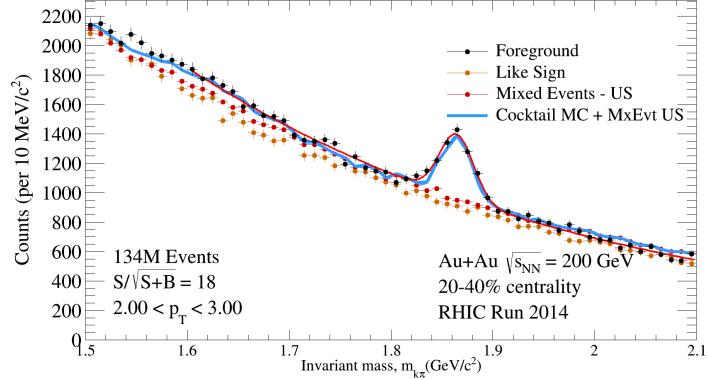


Figure 29: Comparison of  $K\pi$  invariant mass distribution for unlike-sign (US) foreground, like-sign combinatorial background, unlike-sign (US) mixed events combinatorial background, and unlike-sign (US) mixed events combinatorial background + toy montecarlo cocktail for correlated background, for  $2 < p_T < 3$  GeV/ $c$ .

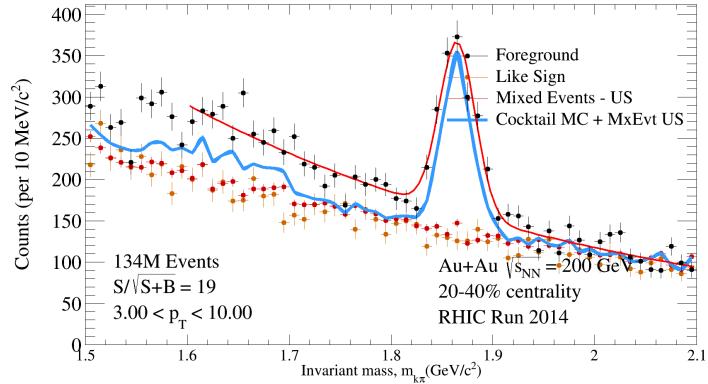


Figure 30: Similar Comparison of  $K\pi$  invariant mass distribution as Fig. 29, for  $3 < p_T < 10$  GeV/ $c$ .

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.

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. 7. 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. 8. Here the PID part are able to factorized as Eq. 9, Eq. 10.

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

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

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

### 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 ( $\varepsilon_{\text{TPC}}$ ) and the TOF matching efficiency ( $\varepsilon_{\text{TOF}}$ ). The particle identification cut efficiency ( $\varepsilon_{\text{PID}}$ ) 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. 11

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

### 4.2 TPC Tracking Efficiency

The TPC tracking efficiency ( $\varepsilon_{\text{TPC}}$ ) 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  $p_T$ ,  $\eta$ , and  $\phi$ , 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 ( $N_{\text{rec}}$ ), satisfying the track quality cuts used in the data analysis, over the number of embedded MC tracks ( $N_{\text{emb}}$ ), as shown in Eq. 12

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

The TPC tracking efficiency in Run14 Au+Au collisions at 200 GeV is shown below. Fig. 31 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. 32 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.

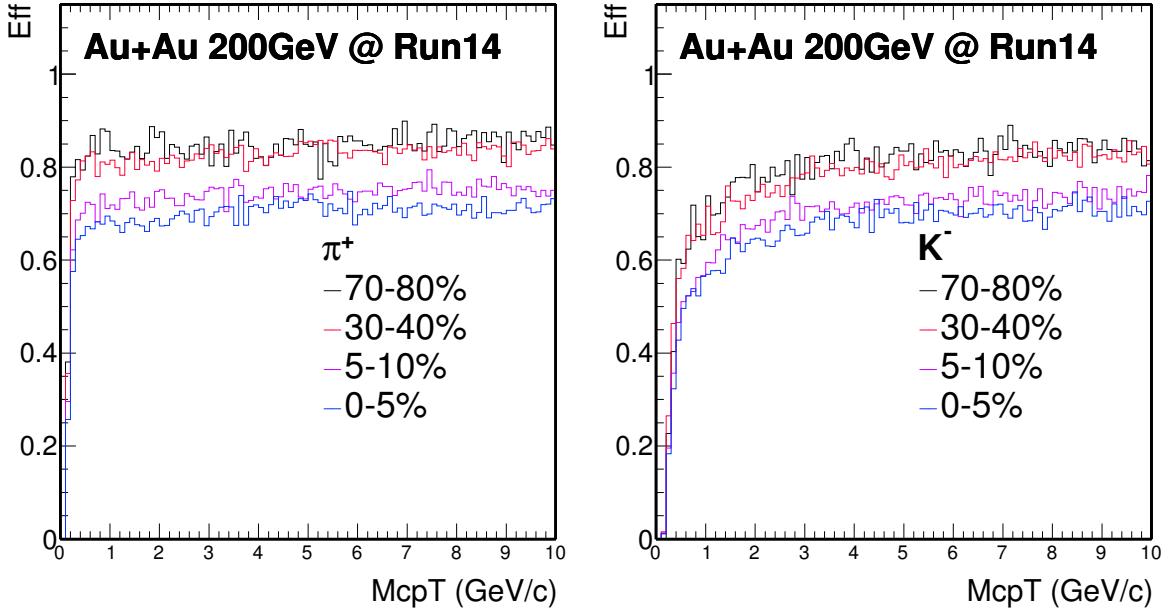


Figure 31: TPC tracking efficiency in Run14 Au+Au collisions at 200 GeV for Pion.

Figure 32: 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 ( $\varepsilon_{TOF}$ ), 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 ( $N_{TPC}$ ).

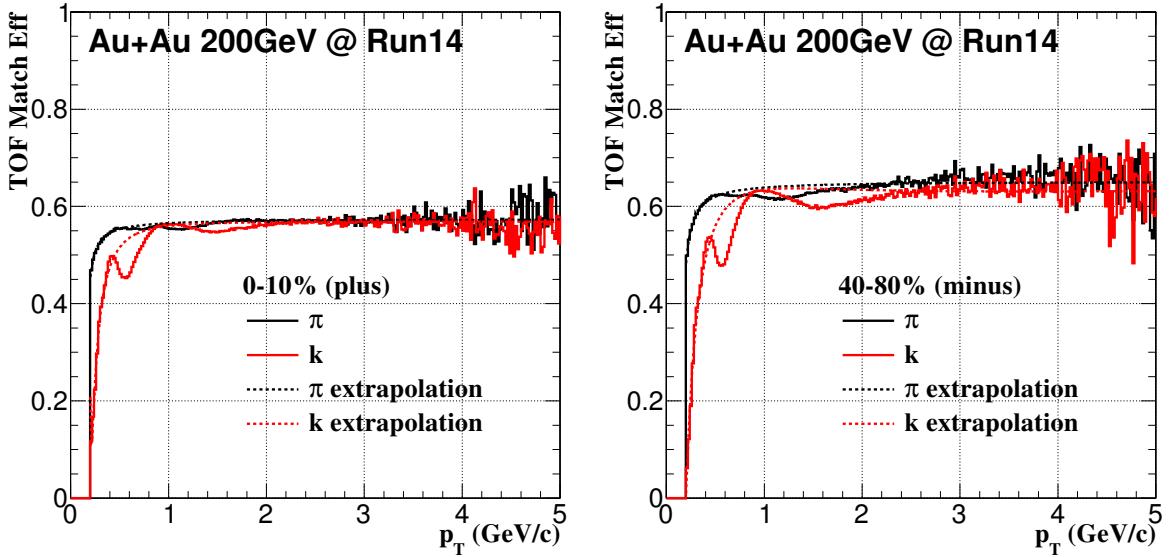


Figure 33: TOF match efficiency in Run14 Au+Au collisions at 200 GeV for positive charge particle in 0-10% centrality. Figure 34: TOF match efficiency in Run14 Au+Au collisions at 200 GeV for negative charge particle in 40-80% centrality.

Fig. 33 shows the TOF match efficiency in Run14 Au+Au collisions at 200 GeV for positive charge particles such as  $\pi^+, K^+$  in the centrality 0-10%. Fig. 34 shows the same plots for negative charge particles in the centrality from 40-80%. For the pion TOF match efficiency, the trend 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. 35 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. 36 and Fig. 37. 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/dx are overlap with each other for kaon and pion, it's not able to distinguish them only use TPC.

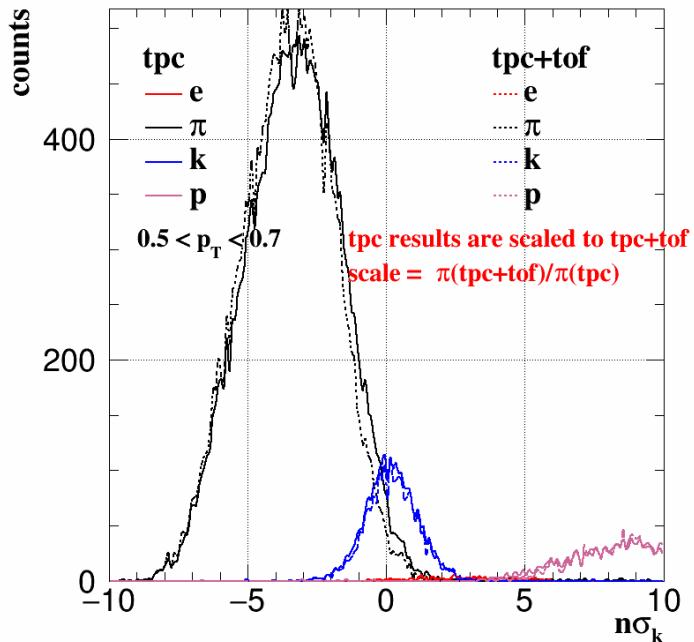


Figure 35:  $n\sigma_K$  distributions for  $0.5 < p_T < 0.7$  GeV/ $c$ . The solid lines are from TPC and dashed lines are from TPC + TOF.

## 4.4 PID Cut Efficiency

The particle identification cut efficiency ( $\varepsilon_{PID}$ ) 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. 38 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. 39 shows the  $KK$  pairs invariant mass distributions. Still with the unlikesign and likesign method, the  $\Phi$  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. 40 and Fig. 41. With these mean and sigma distributions, assuming they follow the Gaussian function, for example, Fig. 42 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.

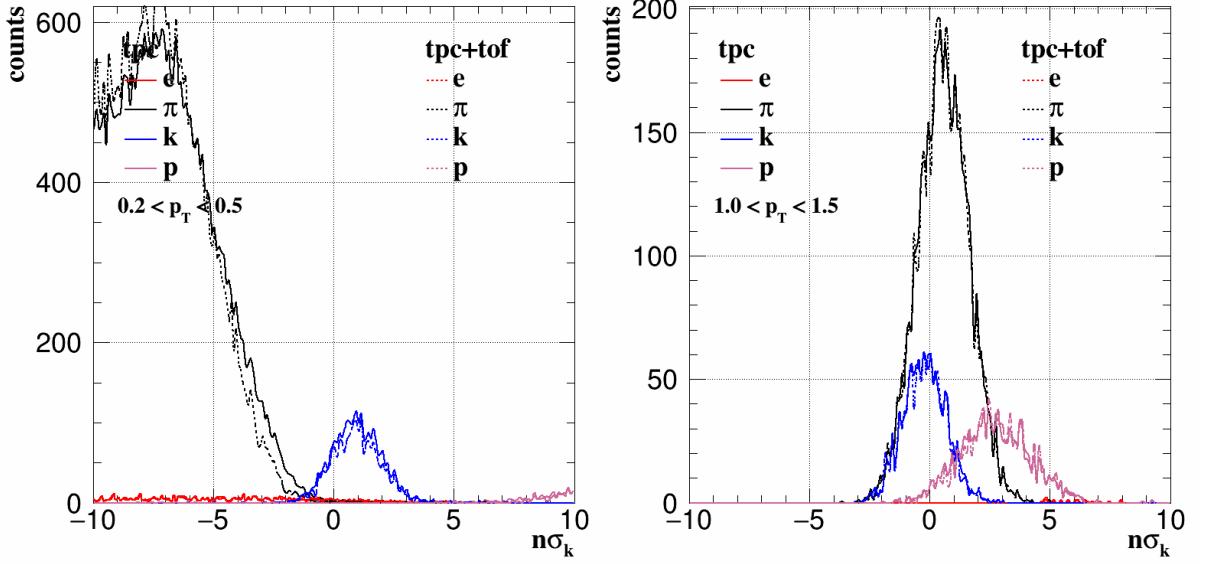


Figure 36:  $n\sigma_K$  distributions for  $0.2 < p_T < 0.5$  GeV/ $c$ . The solid lines are from TPC and dashed lines are from TPC + TOF.

Figure 37:  $n\sigma_K$  distributions for  $1.0 < p_T < 1.5$  GeV/ $c$ . The solid lines are from TPC and dashed lines are from TPC + TOF.

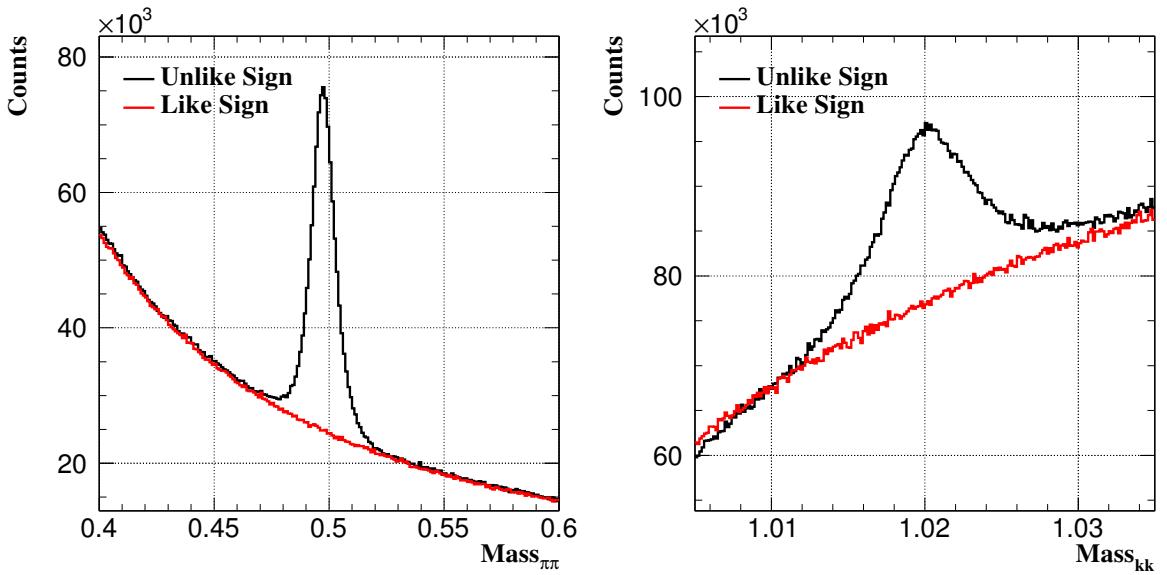


Figure 38: The  $\pi\pi$  pairs invariant mass distributions. The black line is the unlikesign foreground, and the red line is background using like-sign method.

Figure 39: The  $KK$  pairs invariant mass distributions. The black line is the unlikesign foreground, and the red line is background using like-sign method.

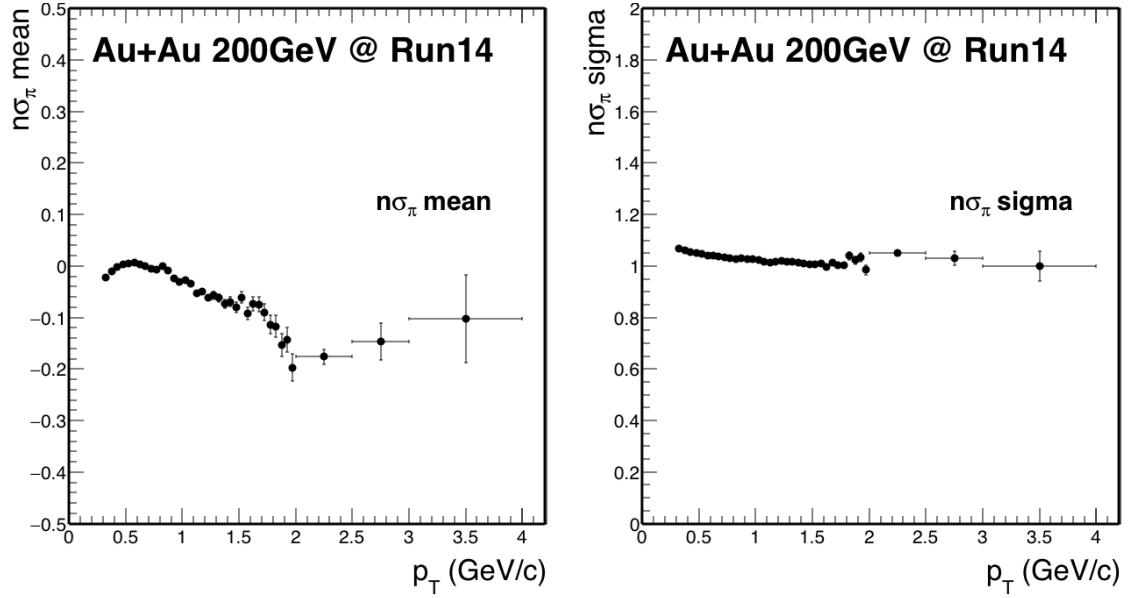


Figure 40: The mean value of  $n\sigma_{\pi}$  distributions vs momentum. The red line is fitted function with polynomial function.  
 Figure 41: The sigma value of  $n\sigma_{\pi}$  distributions vs momentum. The red line is fitted function with polynomial function.

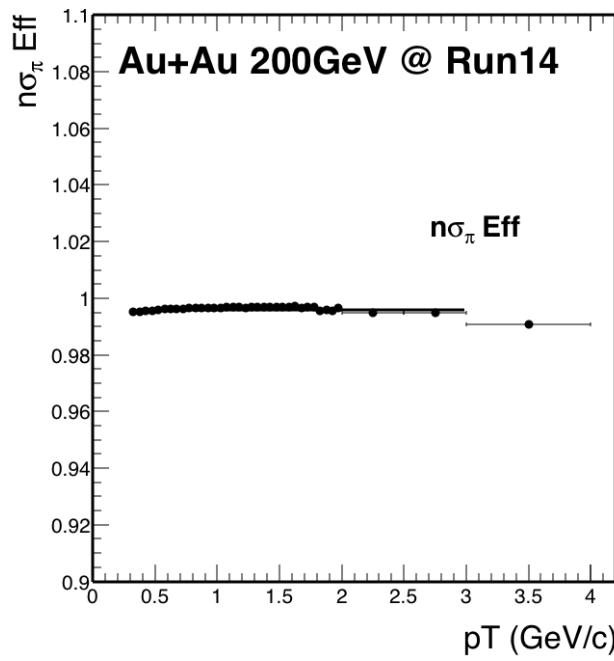


Figure 42:  $n\sigma_{\pi}$  cut efficiency along with momentum. Red line is the fitted polynomial function

#### 4.4.2 $1/\beta$ Cut Efficiency

The  $1/\beta$  cut efficiency is derived from the  $1/\beta$  distributions, two methods are used, one is the bin counting method and another is using Gaussian fitting and calculate the efficiency from the fitting parameters. The default value is from bin counting and the difference between these two methods are quoted as systematic uncertainties as shown on Fig. 43 and Fig. 44 as fitting method. The fluctuations for kaons samples was due to the hadron contaminations at some  $p_T$  range.

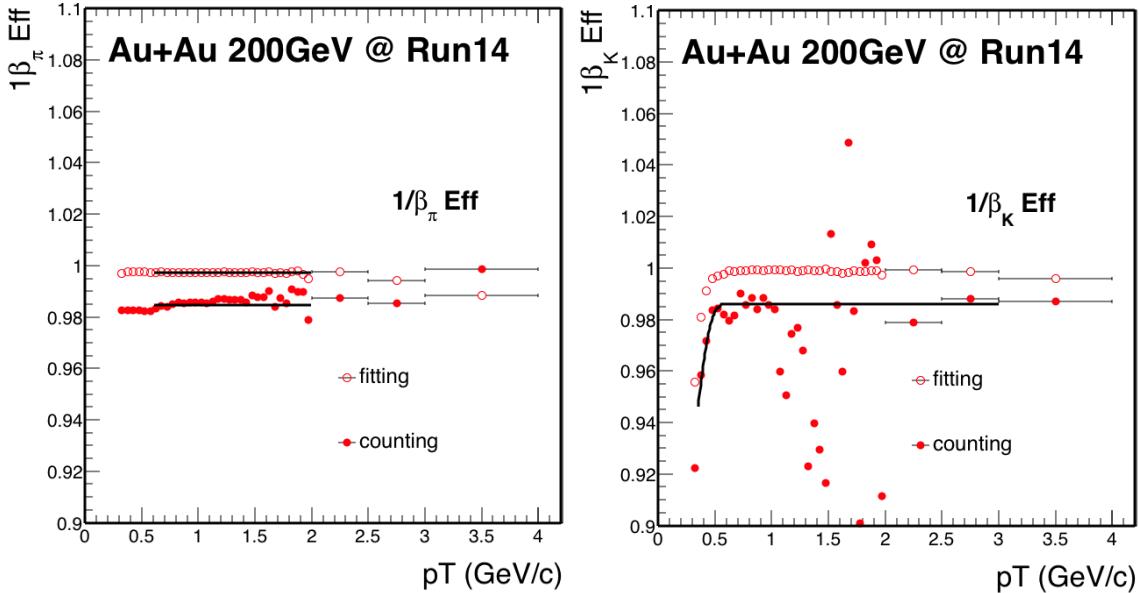


Figure 43:  $1/\beta$  cut efficiency along with momentum for pion. Red line is the fitted function.

Figure 44:  $1/\beta$  cut efficiency along with momentum for kaon. Red line is the fitted function.

#### 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. 7 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.

##### 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:

$$\frac{\text{HFT}}{\text{MC}} = \frac{\text{HFT}}{\text{TPC}} \times \frac{\text{TPC}}{\text{MC}} \quad (13)$$

- Spatial resolution of HFT is encoded in two variables:  $\text{DCA}_{\text{XY}}$  and  $\text{DCA}_Z$  (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,  $p_T$ ,  $\eta$ ,  $\phi$ ,  $V_z$  dependence).
- Extract  $DCA_{XY}$  -  $DCA_Z$  distributions from data. Assuming that data DCA distributions is dominated by primary particles (particle species, centrality,  $p_T$ ,  $\eta$ ,  $V_z$  dependence).
- Extract TPC efficiency and momentum resolution from embedding (particles, centrality and  $p_T$  dependence).

Fig. 45 shows an example of the input HFT match ratio in the certain  $\eta$ ,  $V_z$ ,  $\phi$  and centrality range. The HFT match ratio increase in the low  $p_T$  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  $\eta$  and  $V_z$  since this is effected by HFT acceptance. Fig. 46 shows an example of the  $Dca_{XY}$  vs  $Dca_Z$  distribution in the certain  $p_T$ ,  $\eta$ ,  $V_z$  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  $11 (\phi) \times 10 (\eta) \times 6 (V_z) \times 9 (\text{centrality}) \times 2 (\text{particles})$  1D histograms (36  $p_T$  binning) for HFT match efficiency. There are  $5 (\eta) \times 4 (V_z) \times 9 (\text{centrality}) \times 2 (\text{particles})$  histograms  $\times 19 (p_T)$  2D histograms (144  $\times$  144 Dca binning) for Dca resolution smearing.

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

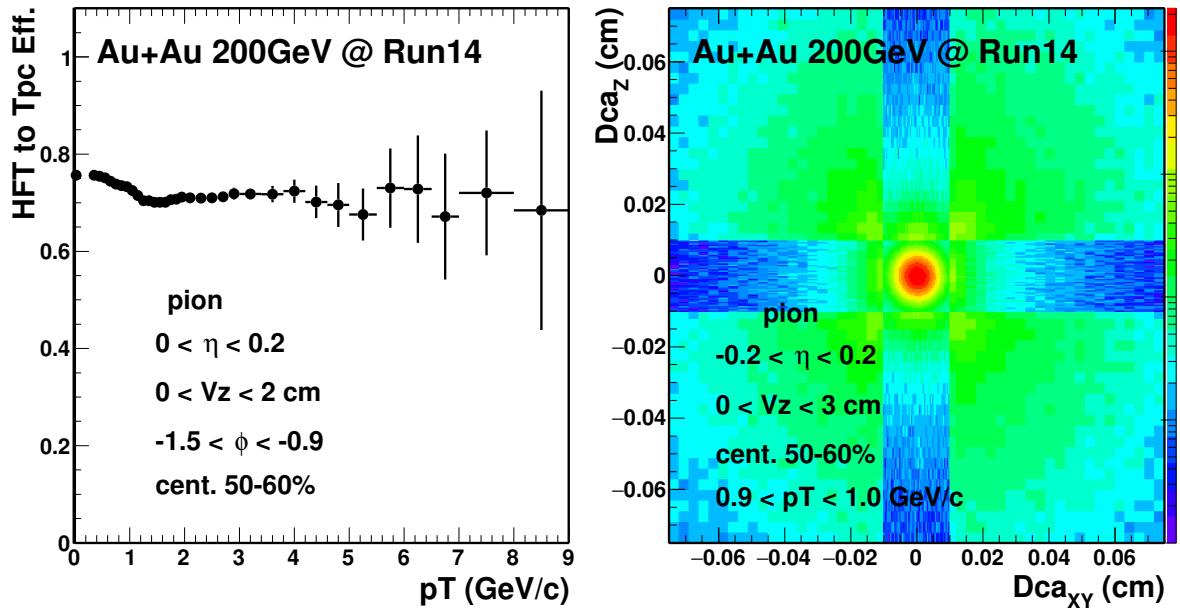


Figure 45: HFT to TPC track match ratio for pion at certain  $\eta, V_z, \phi, \text{centrality}$  range.

Figure 46:  $Dca_{XY}$  vs.  $Dca_Z$  for pion at certain  $\eta, V_z, \text{centrality}$  range.

#### 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.
- 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. 47. First we have the TPC efficiency shown by red marker which is after the  $p_T$ ,  $\eta$  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 seen in the low  $p_T$  part, the topological cut efficiency is really small due to the tight cut as the combinatorial background is huge.

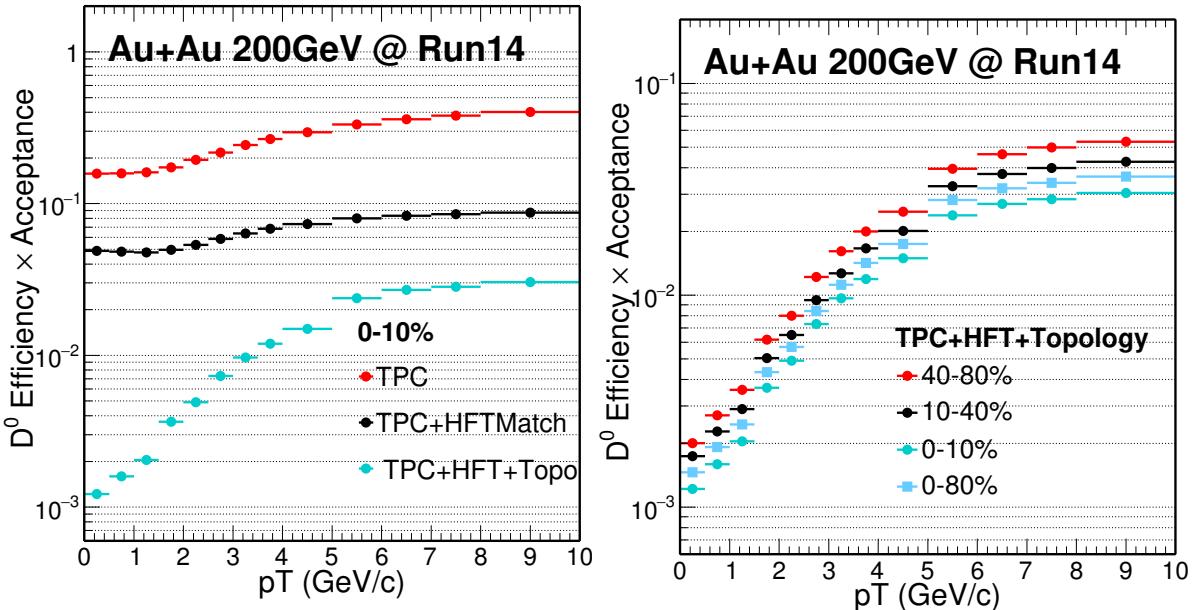


Figure 47:  $D^0$  efficiency step by step from TPC, HFT Ratio, Topological cut in most central 0-10%. Figure 48:  $D^0$  efficiency including TPC, HFT and Topological cut in several centralities.

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. 14. Fig 48 shows the  $D^0$  efficiency for 4 wide centralities after TPC, HFT match and Topological efficiency included.

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

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. 49.  $D^0$  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.

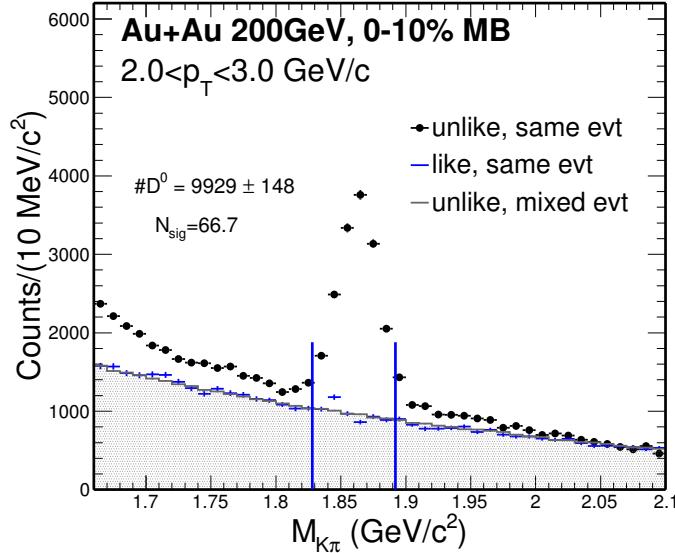


Figure 49:  $D^0$  invariant mass distributions in the  $2 < p_T < 3$  GeV/ $c$ , 0-10% centrality. Black is unlikesign foreground, blue is likesign background and red is mixed event background. The blue vertical lines are the mass window used for the topological comparison.

From Fig. 50 to Fig. 55, these are the topological variables ( $\cos(\theta)$ , decayLength, dcaDaughters, D0DcaToVtx, PionDca and KaonDca) used for the  $D^0$  reconstruction. The distributions from the 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. 13. 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. 56 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. 57 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.

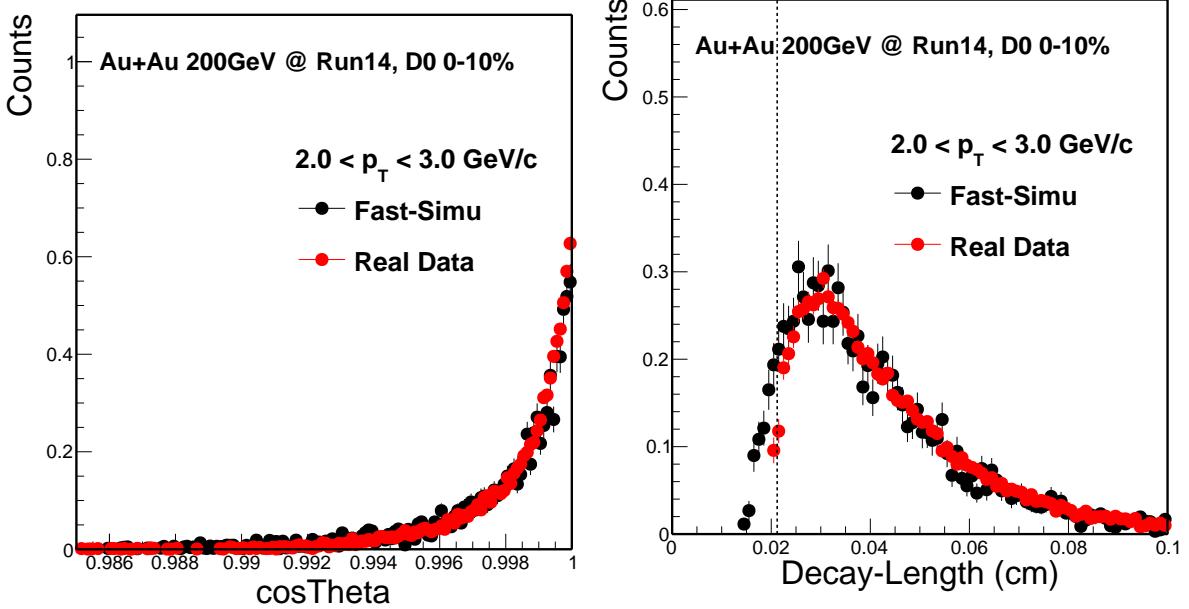


Figure 50:  $D^0$  cosTheta distribution in most central 0-10% between Fast-Simulation and Real Data.

Figure 51:  $D^0$  decay length distribution in most central 0-10% between Fast-Simulation and Real Data.

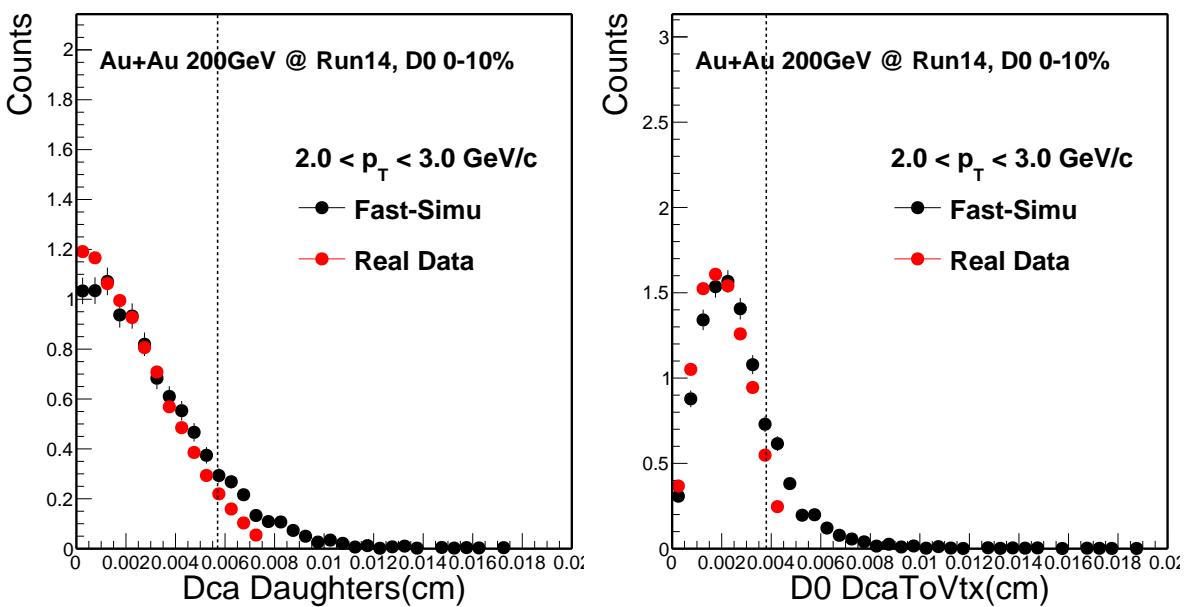


Figure 52:  $D^0$  dcaDaughters distribution in most central 0-10% between Fast-Simulation and Real Data.

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

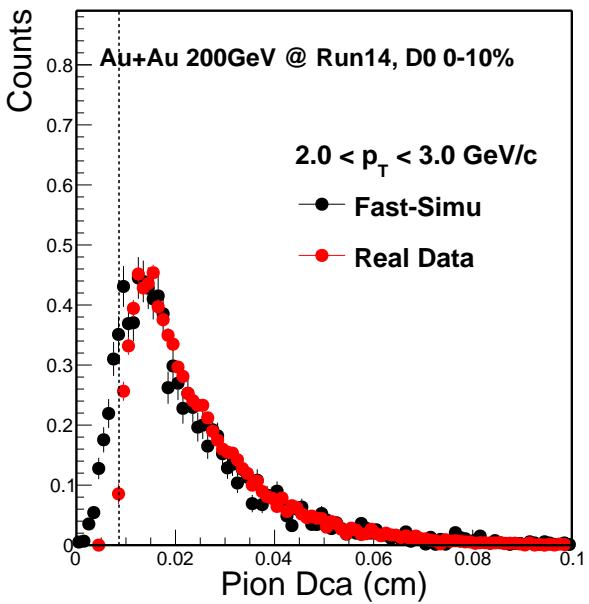


Figure 54:  $D^0$  pionDca distribution in most central 0-10% between Fast-Simulation and Real Data.

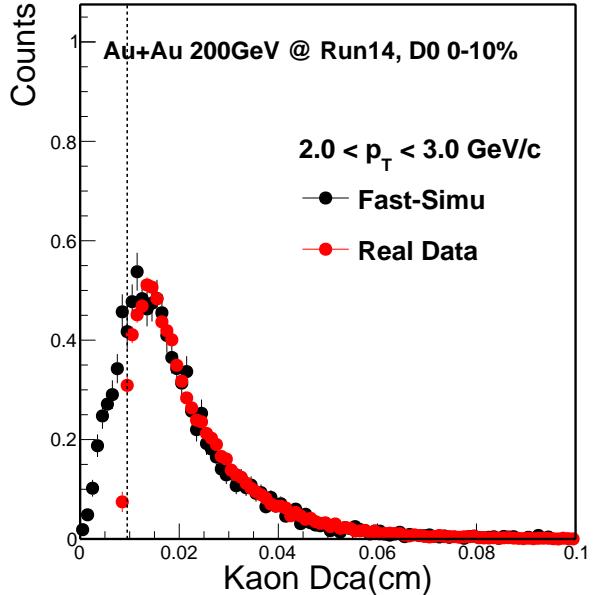


Figure 55:  $D^0$  kaonDca distribution in most central 0-10% between Fast-Simulation and Real Data.

## 4.7 Hijing Samples Performance

The Hijing sample was run through the Full Hijing + GEANT simulation with realistic pileup hits (UPC+MB) in PXL and sensor masking tables. They can provide reasonable performance for the HFT matching ratio and Dca resolution. In total we have  $\sim 45K$  0-10% centrality Hijing events, and for each event is embedded with 20  $D^0$ 's. So in total, we have  $\sim 900K$   $D^0$  for this Hijing sample. The embedded  $D^0$  has small effect on the tracking since the multiplicity is much higher compared to  $20 \times 2 D^0$ , decayed daughters.

As shown in Fig. 58 is the HFT matching ratio comparison between data (red) and Hijing samples (black) in Au+Au 200 GeV/c from 0-10% centrality, the bottom panel is the double ratio of these two HFT matching ratio. The value is around unity, which means the Hijng simulation can well reproduce these matching performance.

Fig. 59 shows the pions Dcaxy comparison between data (red) and Hijing samples (black) in Au+Au 200 GeV/c from 0-10% centrality at  $1.0 < p_T < 1.2\text{GeV}/c$ , the bottom panel is the double ratio of these two Dca distributions. The value is also around unity, which means the Hijng simulation well describes the real data.

## 4.8 Validation Procedures

The idea is simple for this Hijing validation, we have the enriched  $D^0$  Hijing sample. After run through the detector and full GEANT simulation, the  $D^0$  efficiency and topological variables distributions can be extracted. Another procedure is extract the necessary ingredients from Hijing sample for the Fast-Simulation input (Fast-Simulation with Hijing input), such as the TPC Tracking efficiency, the HFT matching ratio and the 2D Dcaxy-Dcaz distributions similar as we used in real data analysis and discussed in the previous section. Then run through the Fast Simulation, as discussed before, the  $D^0$  efficiency and topological variables are also available in this way and can be compared to the first Hijing + GEANT procedure. The workflow is shown in Fig. 60.

### 4.8.1 Validation Efficiency

The first step is to check the kinematic form different MC decayer such as PYTHIA, Hijing, evtGen and PhaseSpace class from ROOT. Need to make sure the decayer used for Fast-Simulation has the same kinematic as the Hijng. After the basic acceptance cut, such as  $D^0 |y| < 1$ , daughter  $p_T > 0.2 \text{ GeV}/c$  and  $|\eta| < 1$  cut.  $D^0$  is the simple phase space decay, all these decayer give the

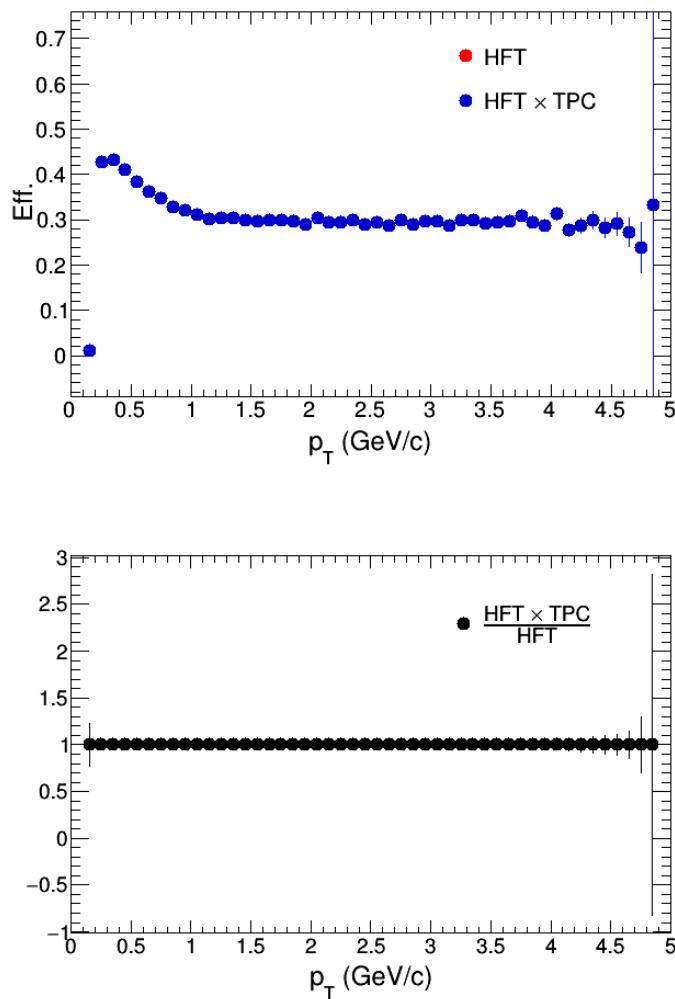


Figure 56: (top)HFT Efficiency Factorization comparison. (bottom) Double Ratio of these factorization.

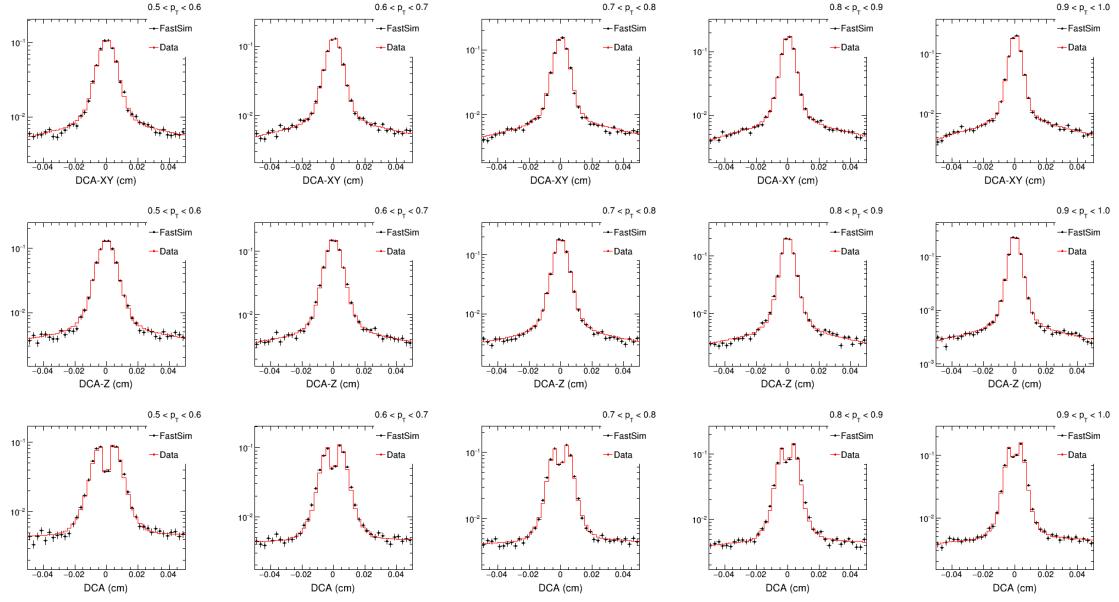


Figure 57: Comparison of Dca between data (red) and Fast-Simulation (black). From top to bottom, the comparison is for Dca<sub>XY</sub>, Dca<sub>Z</sub> and Dca. From left to right the transverse momentum is from low  $p_T$  to high  $p_T$ .

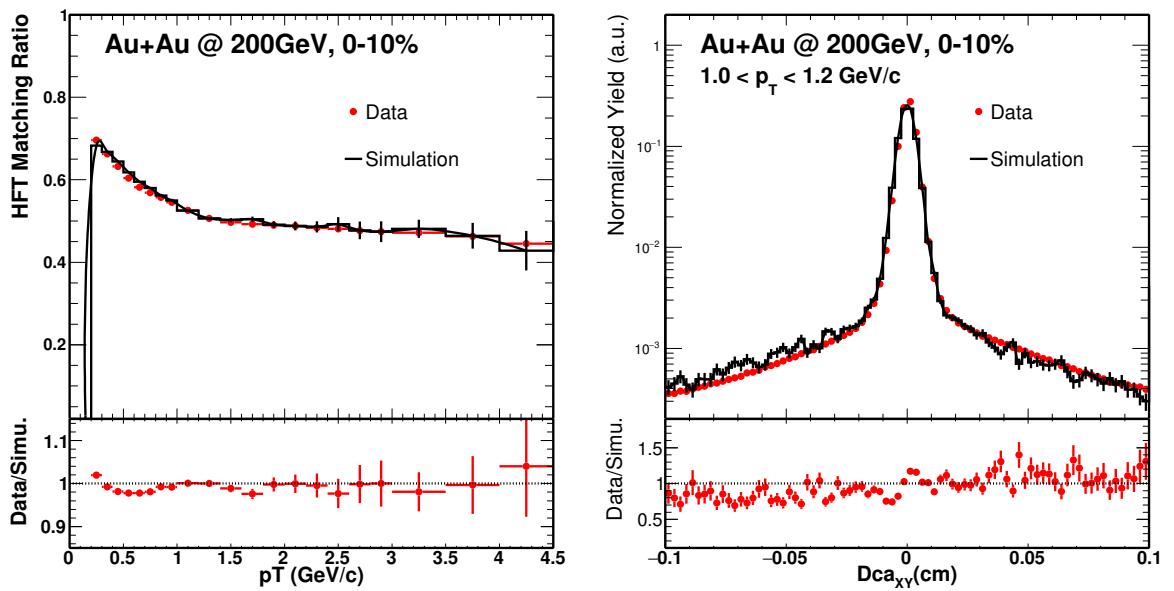


Figure 58: HFT Ratio comparison between data and Hijing simulation in Au+Au 200 GeV/c, 0-10%.

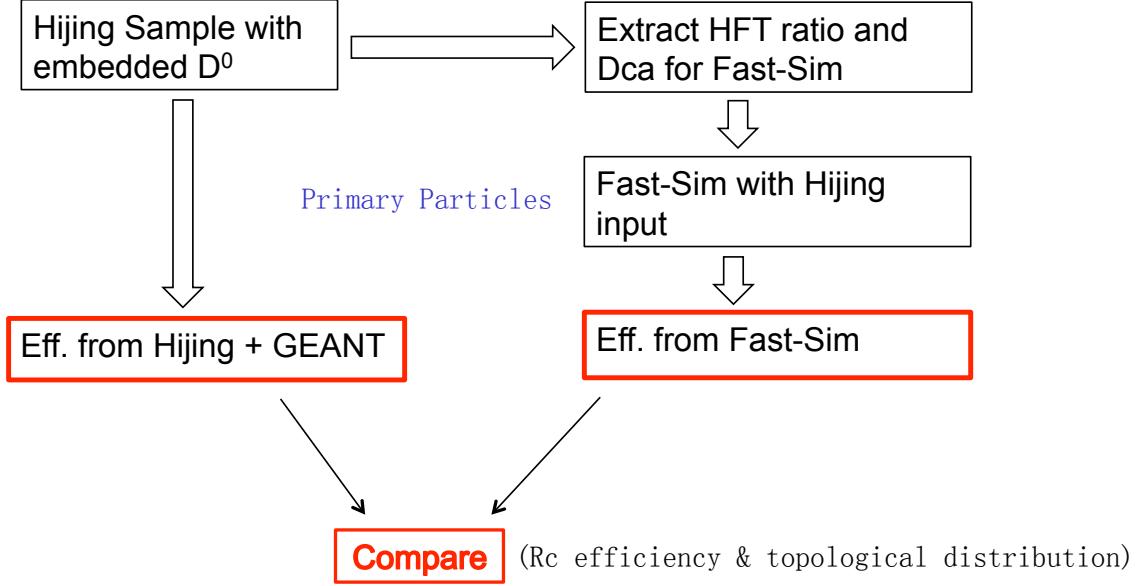


Figure 60: Hijing validation procedure and workflow

same acceptance efficiency as shown in Fig. 61 left panel, the right panel shows the double ratio to PYTHIA. As all the decayer follow the same trend they have the same decay kinematic, so, for our Fast-Simulation decayer we choose PYTHIA for this validation and also for our real data analysis.

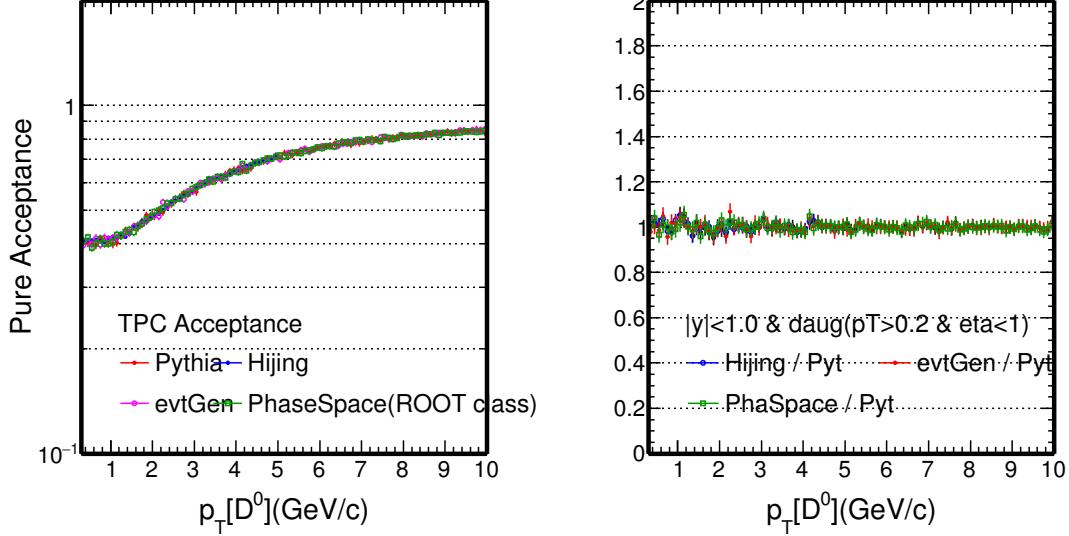


Figure 61:  $D^0$  pure acceptance from different MC decayer, such as PYTHIA, Hijing, evtGen and PhaseSpace class. (right) Double ratio of the acceptance to PYTHIA.

The second step is to check the kinematic with the reconstructed TPC tracking information. Compare to the first step, this one fold in the momentum resolution and the TPC acceptance effect. Fig. 62 left panel shows this efficiency  $\times$  Acceptance comparison between Hijing + GEANT (red) and Fast-Simulation with Hijing input (black), the right panel shows the double ratio to Hijing. As the red line is the fit function and the fit results around  $\sim 1$  shows very good agreement, which means this step is also doing right work.

The next step is to trying to fold in the HFT matching efficiency and this is to consider the HFT acceptance effect. Fig. 63 left panel shows this efficiency  $\times$  Acceptance comparison after TPC and HFT matching between Hijing + GEANT (red) and Fast-Simulation (black), the right

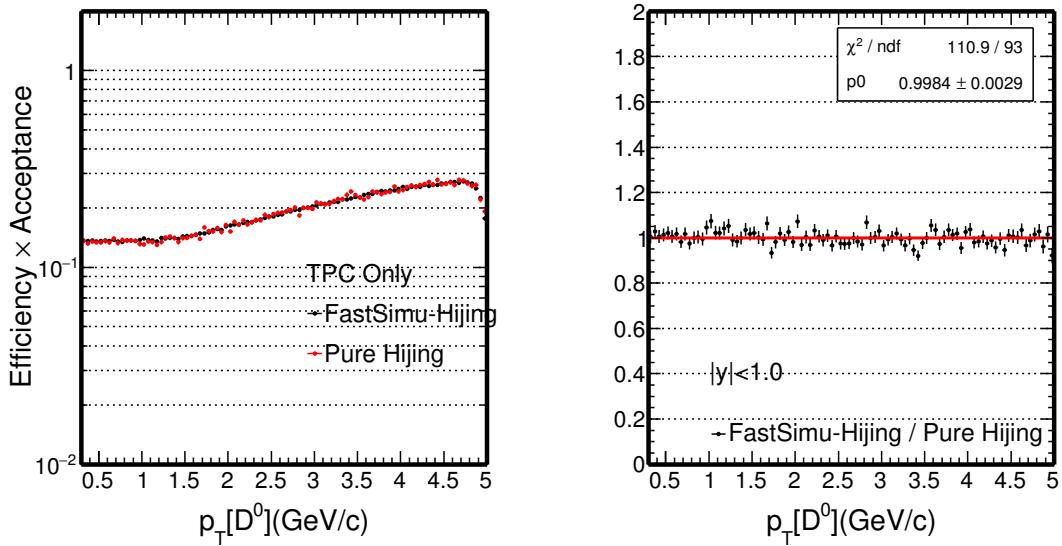


Figure 62: The comparison of  $D^0$  TPC acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input (black). (right) Double ratio of these acceptance to Hijing.

panel shows the double ratio to Hijing. As the red line is the fit function and the fit results around  $\sim 1$  shows good agreement, which means this HFT matching step is also correctly implemented in the package. For the small discrepancy at the high  $p_T$  range, this is purely due to the limited statistics. Since the Hijing sample is time consumption, we do not have enough statistics for the HFT match ratio input. But this problem is not exist for our real data analysis since we have  $\sim 900M$  events which is totally enough and we checked the HFT match ratio, they can extend to a reasonable high  $p_T$  range in real data. We did another small check, use one quart of these Hijing statistics for this validation, and the discrepancy shown here is bigger than the current results, which is another approve of the limited Hijing statistics.

The last step is folding in the topological cuts and then compare between the Hijing and Fast-Simulation. Fig. 64 left panel shows this efficiency  $\times$  Acceptance comparison after TPC, HFT matching and topological cuts between Hijing + GEANT (red) and Fast-Simulation (black), the right panel shows the double ratio to Hijing. Still the red line is the fit function and the fit results around  $\sim 0.93$  shows good agreement, which means this topological variables are well described in the package. For the left panel, there are some twist for this efficiency  $\times$  acceptance at  $p_T \sim 1$  GeV/c and 2 GeV/c, this is due to the topological cuts are different in separate  $p_T$  ranges. As the red points show the efficiency from Hijing + GEANT, the statistics error is larger compared to the Fast-Simulation which shows by black. This is also the reason we use data-driven Fast-Simulation for our efficiency study, it can be easily enlarge the statistics by a factor of 100 or even 1000 compare to the traditional Full GEANT simulation especially for this kind of low efficiency studies. Fig. 65 shows the same plots of the comparison as Fig. 64 with different binning, we merged some binning for statistics concern. After merged the binning, the agreement is even better from the fitting shown on the right panel, the fitting results is  $\sim 0.96$ .

From Hijing + GEANT simulation, we know exactly whether the HFT matched track is real match or mismatch, so we can determine the HFT real matched efficiency  $\times$  acceptance for  $D^0$  reconstruction from Hijing sample. Fig. 66 shows these real matched efficiency  $\times$  acceptance comparison between Hijng + GEANT and our previously Fast-Simulation. The right panel shows the double ratio of these efficiency and fitted with a line, the parameters shows  $\sim 0.98$  which means the (previous) Fast-Simulation can well reproduce the real HFT matched reconstruction efficiency. Fig. 67 shows the same plots of the comparison with different binning.

If we compare with the previous Hijing HFT matched efficiency (not necessary to be real matched), it also indicate that most of the Mis-matched daughter tracks are removed by topological cuts as we said in the assumptions. Fig. 68 shows the different components contributions directly from Hijing, the black one is HFT matched, red one requires all the daughter tracks are real

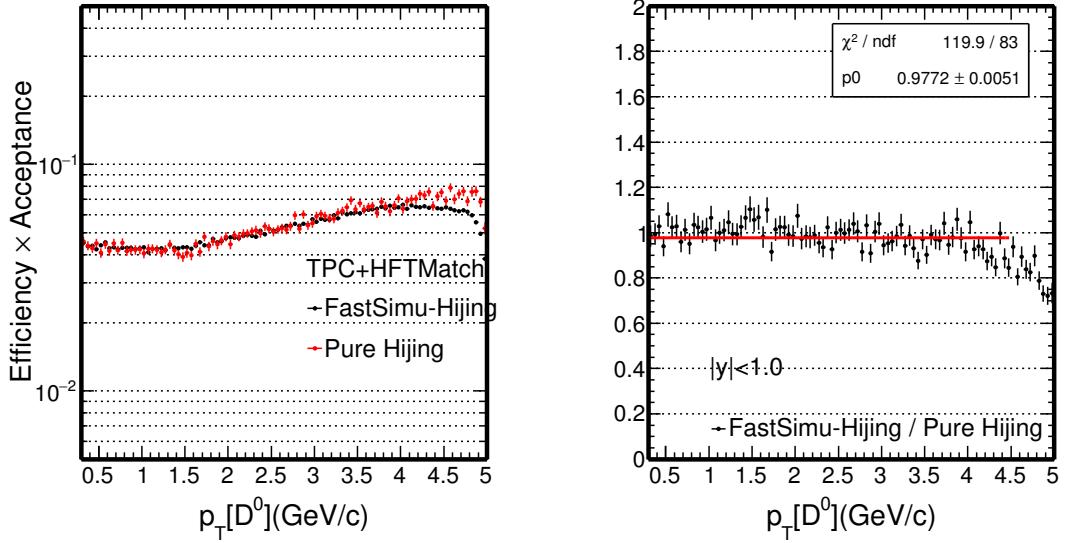


Figure 63: The comparison of  $D^0$  TPC + HFT match acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input (black). (right) Double ratio of these acceptance to Hijing.

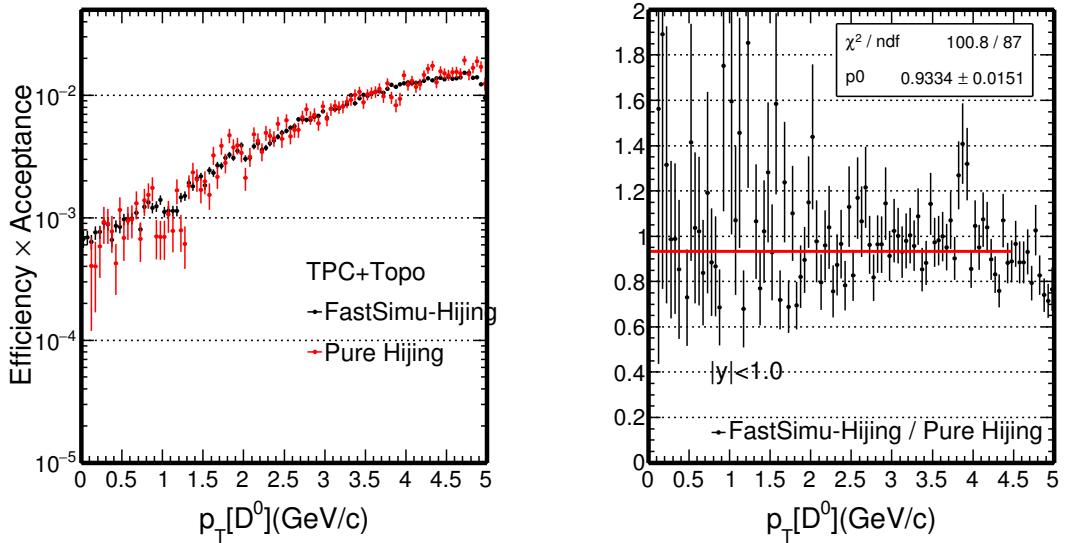


Figure 64: The comparison of  $D^0$  TPC + HFT match + Topological acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input (black). (right) Double ratio of these acceptance to Hijing.

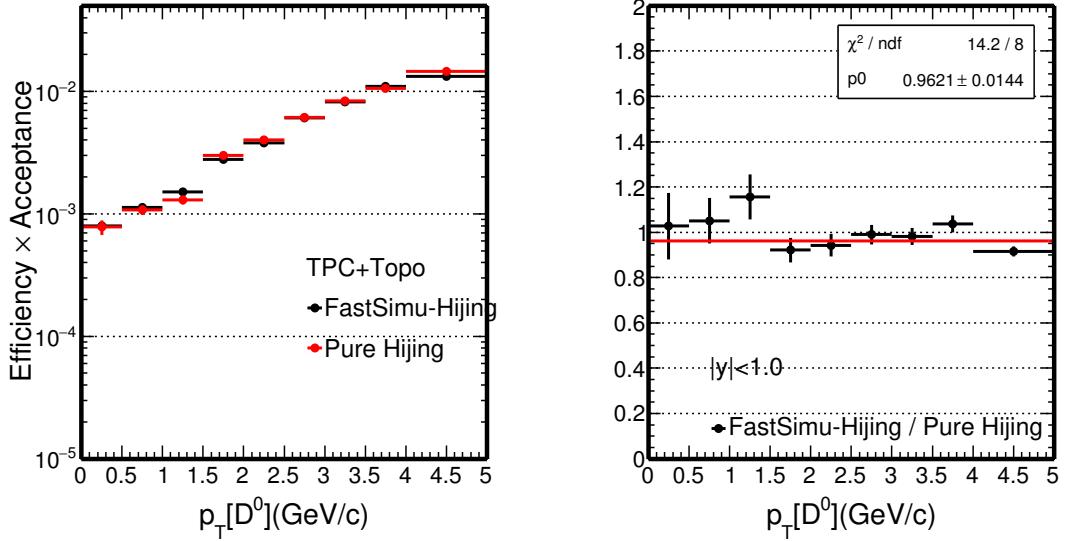


Figure 65: The comparison of  $D^0$  TPC + HFT match + Topological acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input with wide binning (black). (right) Double ratio of these acceptance to Hijing.

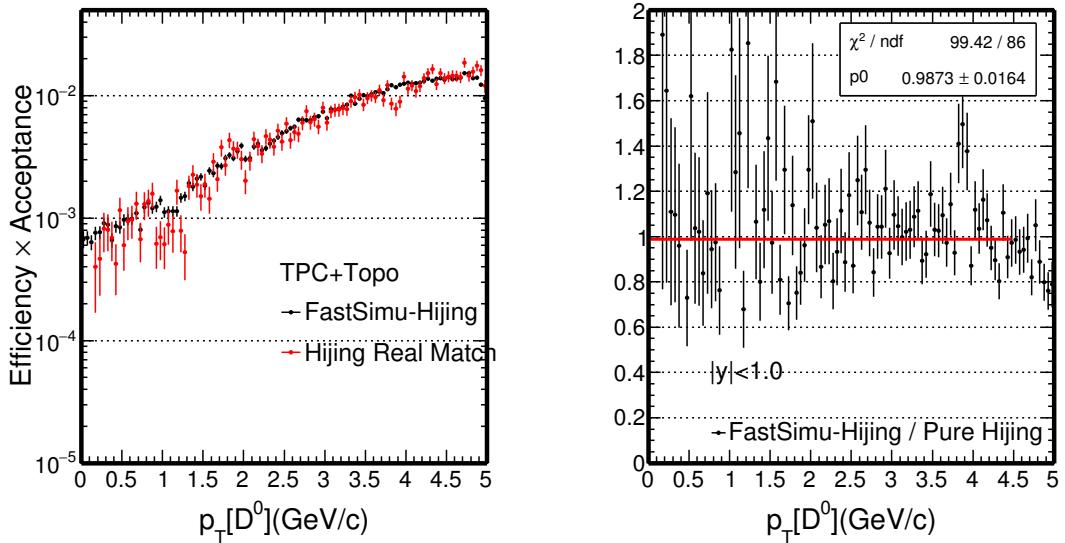


Figure 66: The comparison of  $D^0$  TPC + HFT Real match + Topological acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input (black). (right) Double ratio of these acceptance to Hijing.

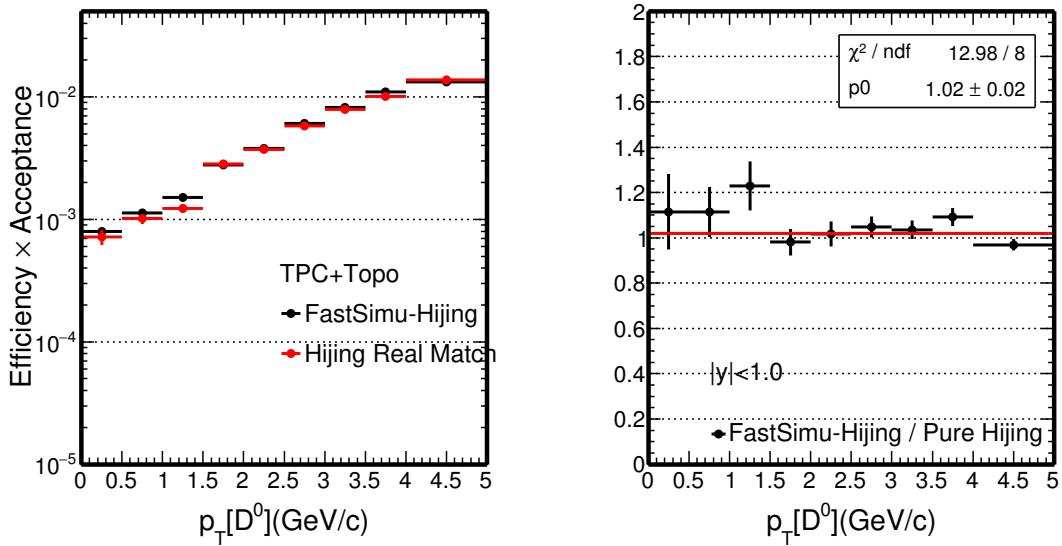


Figure 67: The comparison of  $D^0$  TPC + HFT Real match + Topological acceptance  $\times$  efficiency between Hijing + GEANT (red) and Fast-Simulation with Hijing input with wide binning (black). (right) Double ratio of these acceptance to Hijing.

matched and the blue one shows at least one of the daughter tracks are mis-matched. Right panel shows the relative fraction of the real match and mis-mismatch contribution. As see, most of the mis-matched tracks are removed, but still there are  $\sim 5\%$  contribution from this study.

Above all the discussions in this section 4.8.1, we are confident that the Fast-Simulation method can well reproduce the acceptance and efficiency for this HFT related analysis. The precision as shown on Fig. 65 is good enough for our efficiency study. For the missed-match check, there are  $\sim 5\%$  contributions in the signals. And there is another approval that will describe in the following section.

#### 4.8.2 Validation Topological Distributions

As discussed before, we can extract the topological variables from both Hijing + GEANT and Fast-Simulation relay on those Hijing input. Similar as we did in Sec. ??, we can compare the topological distributions from these two procedures. This will be another evidence that our Fast-Simulation can well reproduce the topological variables which is crucial for these kind of secondary vertex reconstruction analysis.

From Fig. 69 to Fig. 73, these are the topological variables used for the  $D^0$  reconstruction. The topological distributions can be extracted both directly from Hijing + GEANT and from Fast-Simulation relay on Hijing input. The Fast-Simulation part was the same package as we used for the efficiency study before.

As seen, the comparison of topological variables from Hijing have a very good agreement, which means again our Fast-Simulation method can well reproduce the topological variables in Hijing sample just as in the real data case. In another word, the efficiency estimation from this Hijing-Data-Driven Fast-Simulation is reliable. This is the other confident as we discussed in the last part of previous section 4.8.1.

There are two more assumptions which were not answered yet. Here we are trying to discuss a little bit in the following part.

## 4.9 Secondary Track Contribution

The Fast-Simulation is validated with primary track in the procedure Fig. 60. All the tracks for HFT matching ratio and Dca inputted to Fast-Simulation is primary track. Based on the Hijing sample we can study the secondary track contribution since in the real data part we can't

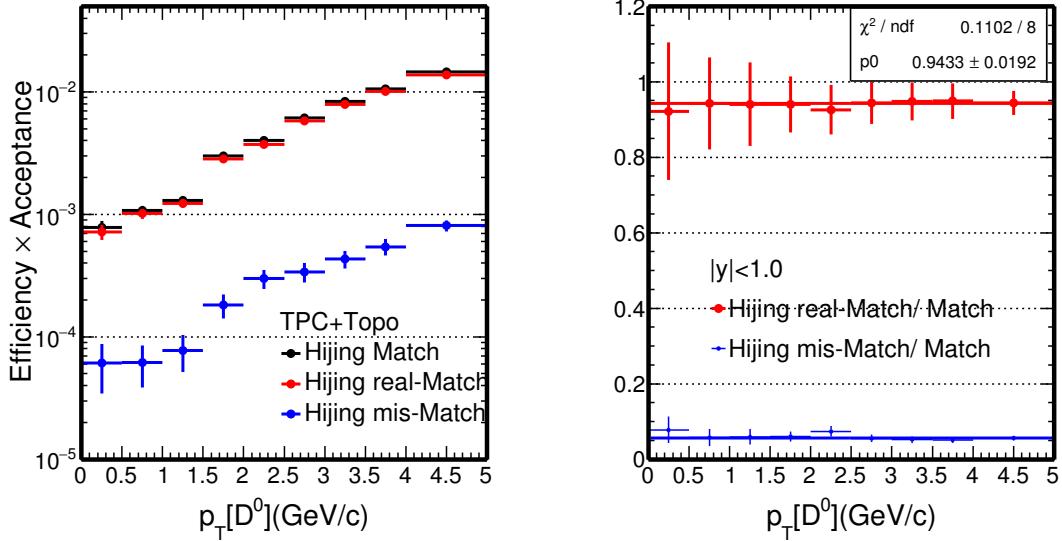


Figure 68: The comparison of  $D^0$  TPC + HFT (real/mis) match + Topological acceptance  $\times$  efficiency for Hijing + GEANT. (right) Double ratio of the components form Hijing.

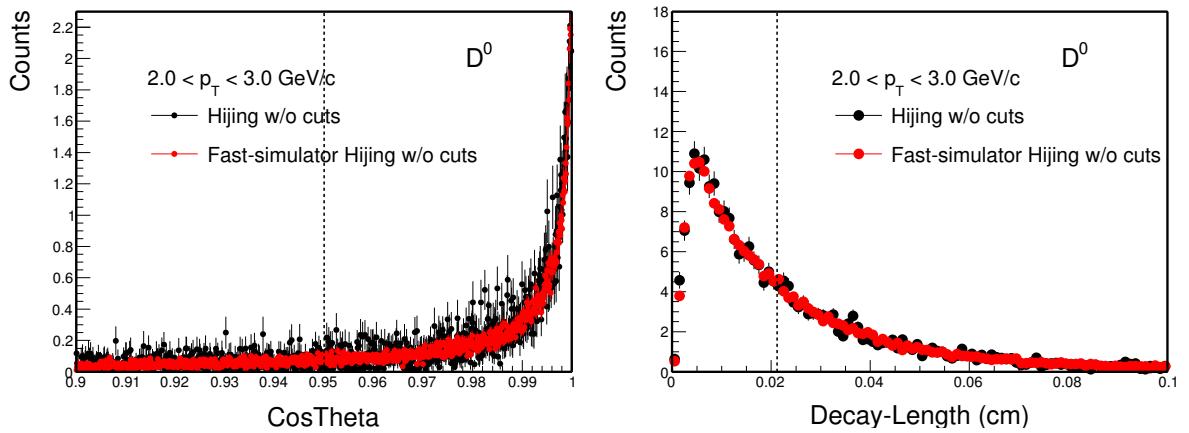


Figure 69:  $D^0$  cosTheta distribution in most central 0-10% between Hijing and Fast-Simulation.

Figure 70:  $D^0$  decay length distribution in most central 0-10% between Hijing and Fast-Simulation.

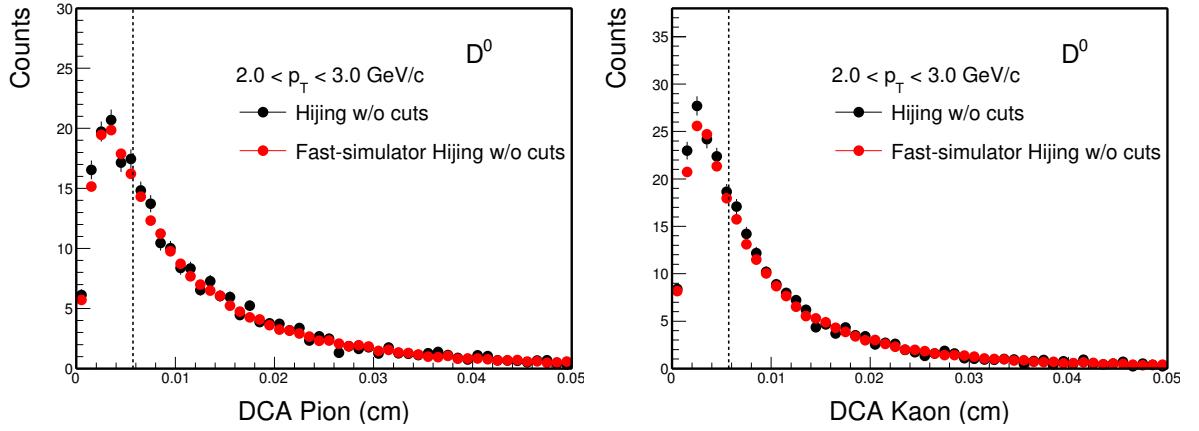


Figure 71:  $D^0$  pions Dca distribution in most central 0-10% between Hijing and Fast-Simulation. Figure 72:  $D^0$  kaons Dca distribution in most central 0-10% between Hijing and Fast-Simulation.

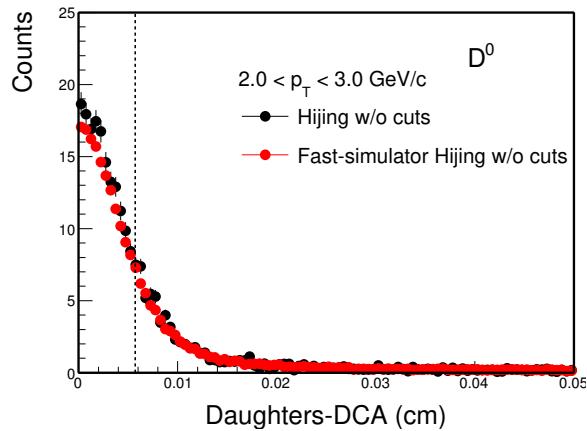


Figure 73:  $D^0$  dcaDaughters distribution in most central 0-10% between Hijing and Fast-Simulation.

distinguish primary track and secondary track. In *Hijing* simulation, we use the start vertex of that track to determine whether it's primary track or secondary track.

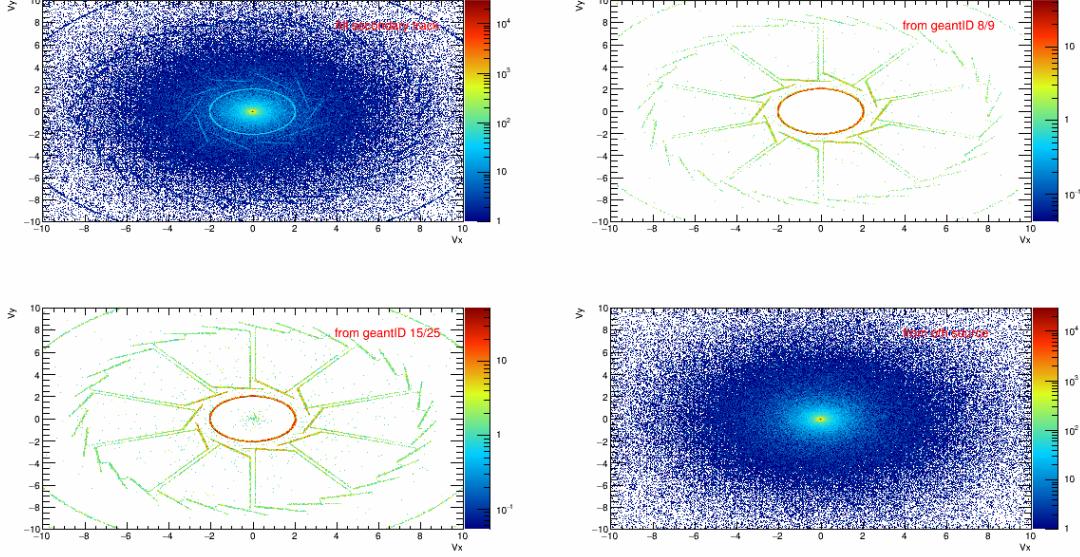


Figure 74: The vertex distribution for Pions from secondary decay. Top left is the overall secondary pion tracks, top right is pion decayed from GeantID=8/9 (which is  $\pi^\pm$ ), bottom left is pion decayed from GeantID=15/25 (which is anti-proton and anti-neutron), bottom right is decayed from other source

Fig. 74 shows the pions vertex distributions from the secondary decay. The first one is the overall secondary pion vertex distributions and we can clearly see some HFT structure. Top right panel is pions decayed from GeantID==8/9 (which is  $\pi^\pm$ ), this part is the knocked out particles with HFT. The bottom left is pion decayed from GeantID==15/25 (which is anti-proton and anti-neutron), this is the normal annihilation particles. The last one bottom right shows the pions decayed from other source.

The secondary track have kind of different performance compare to the primary track such as the HFT match ratio shown on Fig. 75. The solid circle is the inclusive one for HFT matching ratio, the empty circle is for the primary pions and the solid square is for the secondary pions. All these HFT match ratios are after applying exactly the same cut as real data analysis. The low match ratio for secondary track is reasonable since they are decayed far away from the vertex and most of them do not have three HFT hits. The more contribution from the secondary track, the more difference we observed between inclusive one and primary one. The bottom panel shows the HFT matching double ratio of inclusive one over primary one. For the pions, since the secondary pion have some contributions, we do saw the different between primary one and inclusive one at some certain  $p_T$  range for this HFT match ratio. For the kaons, the relative secondary contribution is small, that's why there is no big difference between primary and inclusive ones as see on Fig. 76.

This secondary track contribution for our efficiency correction need to be taken care, especially for Pions. There are a few percent contributions from our *Hijing* simulation study. In our real data efficiency correction, we took this double ratio from *Hijing* as an additional correction factor for the HFT matching ratio, since the data part can only obtained the inclusive one. After this additional correction, we still be able to obtain the precision like Fig. 65.

For the secondary track Dca contributions, we tested with the inclusive track Dca or primary track Dca. In principle, with the inclusive tracks, they should have slightly broader distributions. But in our test, it seems that these contributions to the final  $D^0$  efficiency is really small. This maybe due to the limited *Hijing* statistics or the slightly Dca difference does not contribute much. But in our real data case, we do not take this secondary Dca contributions as additional correction factor yet.

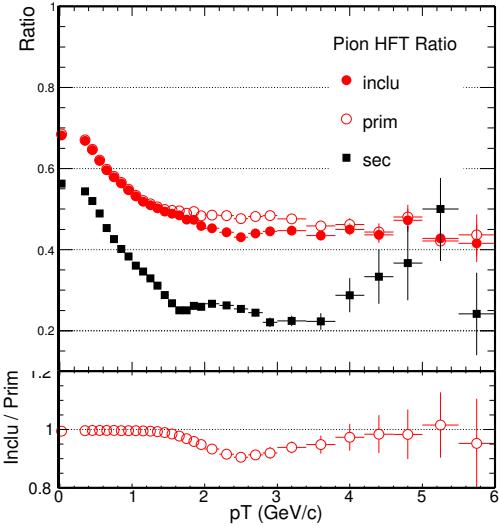


Figure 75: HFT Matching Ratio for Pions, compare between primary track and secondary tracks relay on Hijing. (bottom) The double ratios of inclusive one to primary tracks.

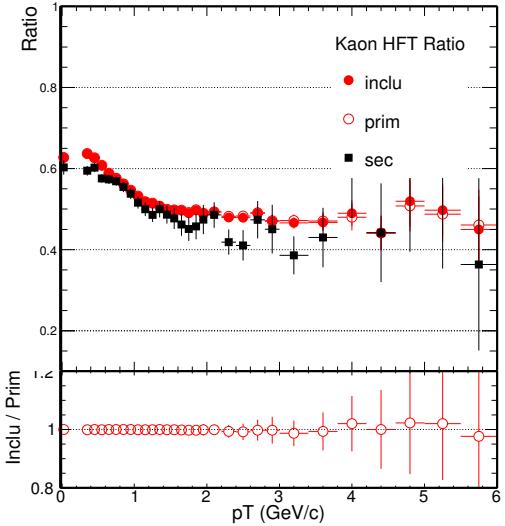


Figure 76: HFT Matching Ratio for Kaons, compare between primary track and secondary tracks relay on Hijing. (bottom) The double ratios of inclusive one to primary tracks.

#### 4.10 Vertex Resolution Contribution

As discussed before, the vertex resolution in peripheral events still have some contributions. If those peripheral events vertices are out of hundreds or dozens of  $\mu\text{m}$  vertex resolution, they are not likely to contribute to the D mesons foreground (maybe not even the background). To correctly count the number of peripheral events we need to understand the vertex resolution. The 2D Dcaxy Dcaz distributions are the only input to the Fast-Simulation for this effect. They contain both the vertices and tracks contribution.

To solve this problem, we need to unfold the vertex resolution from 2D Dca distributions, and this is not straightforward since the vertex resolution contribution could be in the same order of the Dca resolution and this is not reliable (subtracting two numbers that are close to each other have very large uncertainties). There is another way we can rely on to obtain this correction factor, which is the Minimum Bias Hijing simulation sample. From Hijing sample we know the true efficiency for any centrality species, and from the Fast-Simulation we can obtain the efficiency including those vertex effect. The difference can be took as the additional correction factor for real data analysis if this effect is not too big. The MB Hijing sample we used here is from the same setup as we discussed before, and the only difference compared to the 0-10% Hijing sample is the impact parameters (b). But for our centrality selection, we still use grefumlt for both data and simulation.

For the QA of the simulation samples, including HFT matching ratio and Dca resolutions for different centralities, all the details can be found in the next links:

[http://portal.nersc.gov/project/star/xlchen/D0\\_Hijing/QA/all/](http://portal.nersc.gov/project/star/xlchen/D0_Hijing/QA/all/)

The first vertex contribution we checked is for the centrality species 70-80%. Follow the same procedures as we discussed for 0-10%, we did the validation step by step. First is the acceptance check as shown on Fig. 77, the results from Hijing and Fast-Simulation matched very well for this most peripheral events, which is the same as our expectation. The second step is check the kinematic with the reconstructed TPC tracking information. These was involved with the TPC tracking efficiency lost and the momentum resolution. the result in Fig. 78 shows good agreement.

The next step for this peripheral centrality will be fold in the HFT matching efficiency and the results shown in Fig. 79. As seen, the two curves are close with each other and the double ratio was close to unity. The last step will be the topological contributions which also including the vertex contributions. Fig. 80 shows this efficiency  $\times$  Acceptance comparison between Hijing and Fast-Simulation after the TPC, HFT matching and topological cuts for the 70-80% centralities. As seen, due to the involved vertex resolution contribution for this peripheral events, the agreement between Hijing and Fast-Simulation was not good anymore. If we try to fit the double ratio shown

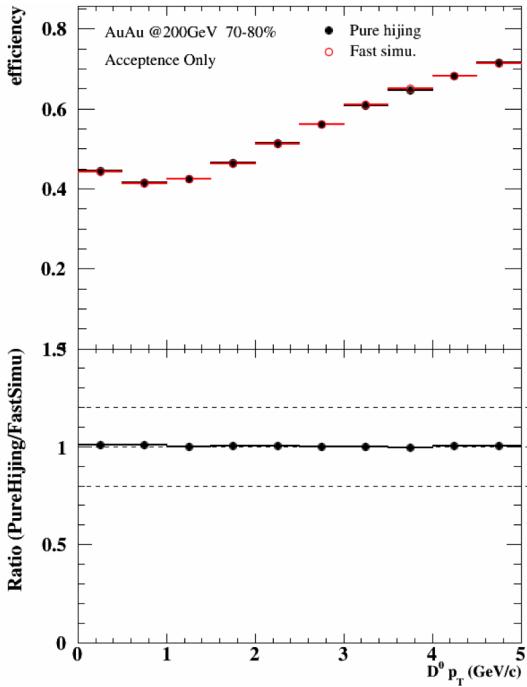


Figure 77: The comparison of  $D^0$  acceptance between Hijing + GEANT (black) and Fast-Simulation with Hijing input (red). (bottom) Double ratio of these acceptance of Hijing to Fast-Simulation.

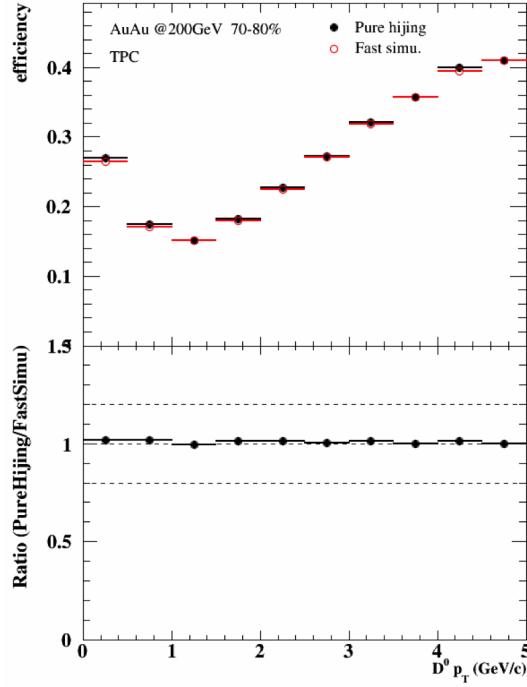


Figure 78: The comparison of  $D^0$  TPC acceptance  $\times$  efficiency between Hijing + GEANT (black) and Fast-Simulation with Hijing input (red). (bottom) Double ratio of these efficiency of Hijing to Fast-Simulation.

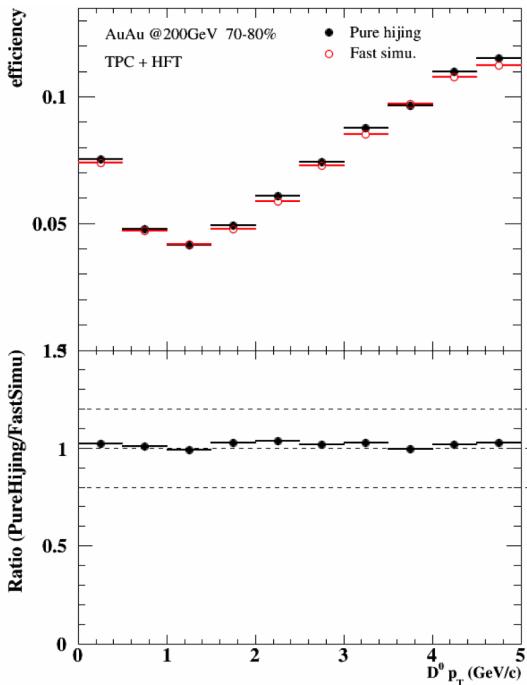


Figure 79: The comparison of  $D^0$  TPC + HFT match acceptance  $\times$  efficiency between Hijing + GEANT (black) and Fast-Simulation with Hijing input (red). (bottom) Double ratio of these acceptance of Hijing to Fast-Simulation.

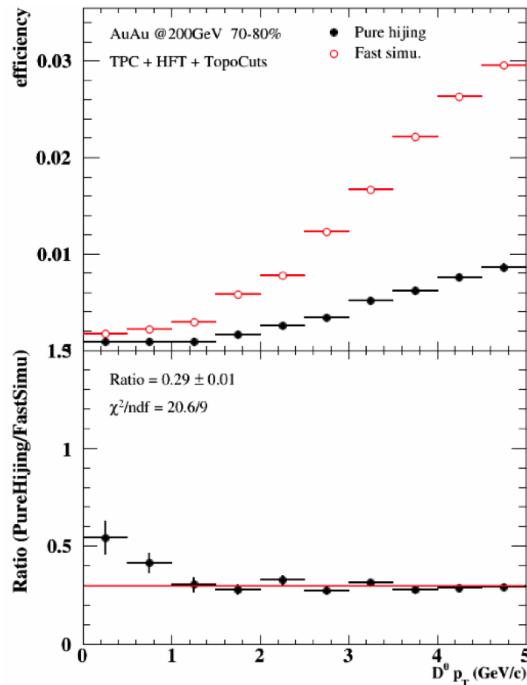


Figure 80: The comparison of  $D^0$  TPC + HFT match + Topological acceptance  $\times$  efficiency between Hijing + GEANT (black) and Fast-Simulation with Hijing input (red). (bottom) Double ratio of these efficiency of Hijing to Fast-Simulation.

in the lower panel, the difference can be as large as a factor of  $\sim 1/0.29$ .

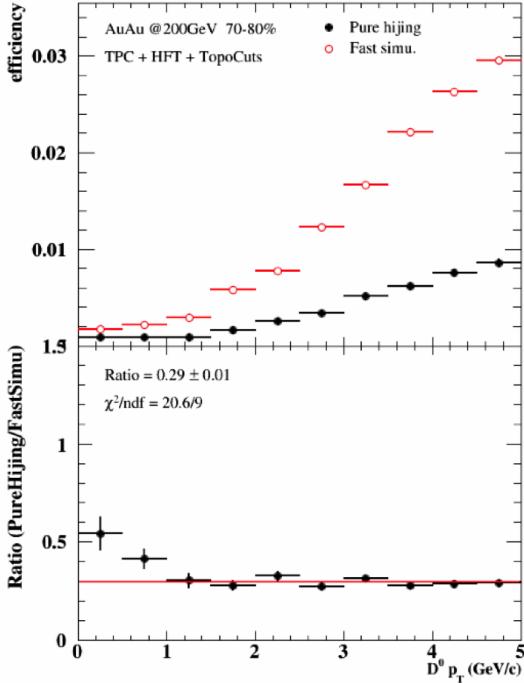


Figure 81: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation for 70-80% centrality (red). (bottom) Double ratio to Fast-Simulation.

For this vertex contributions, we use these double ratios as our additional correction factor for the final results. As we saw and discussed before, the most central events will not suffer this vertex contribution, but only the most peripheral and mid-peripheral events need to consider this effect. The following plots show these correction factor for different centrality species from the most peripheral to most central events.

Fig. 81 and the following figures show the comparison between Hijing and Fast-Simulation from different centralities, in each bottom panel, the fitted results show the expected trend. For the most peripheral events the correction factor is as large as  $\sim 1/0.29$  for centrality 70-80%,  $\sim 1/0.55$  for centrality 60-70%,  $\sim 1/0.77$  for 50-60%,  $\sim 1/0.89$  for 40-50% and  $\sim 1/0.93$  for 30-40%. For the mid central, 20 – 30% centrality, the correction effect is already small, the double ratio is close to 1.

note here,

For the most central collisions, the correction factor can be neglect. This was validated and confirmed by our Hijing samples.

#### 4.11 Double Counting Correction

$D^0$  candidates are reconstructed with pairing  $K^-$  and  $\pi^+$  candidate tracks. When  $D^0$  daughter  $K^-$  is misidentified as a  $\pi^-$  while the other daughter  $\pi^+$  is misidentified as a  $K^+$ , the resulting pair  $K^+\pi^-$  will enter into the distribution for reconstructing  $\bar{D}^0$ . Although the mass assignments are wrong, the pair  $K^+\pi^-$  invariant mass will be still peak around the  $D^0$  region with typically a broader distribution compared to the real signal. When counting the final  $D^0$  candidates, these within the mass selection window will be counted twice. (See also study in previous STAR open charm hadron measurements - STAR notes below).

<https://drupal.star.bnl.gov/STAR/starnotes/private/psn0594>

<https://drupal.star.bnl.gov/STAR/starnotes/private/psn0550>

The double counting issue will certainly affect the obtained  $D^0$  raw yields. In the  $v_2$  analysis, since the doubly counted candidates are still coming from  $D^0$ , this issue should not affect the

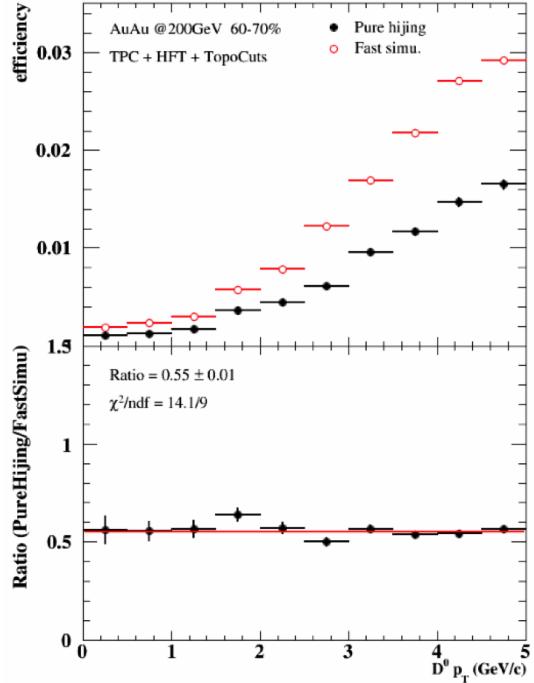


Figure 82: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation for 60-70% centrality (red). (bottom) Double ratio to Fast-Simulation.

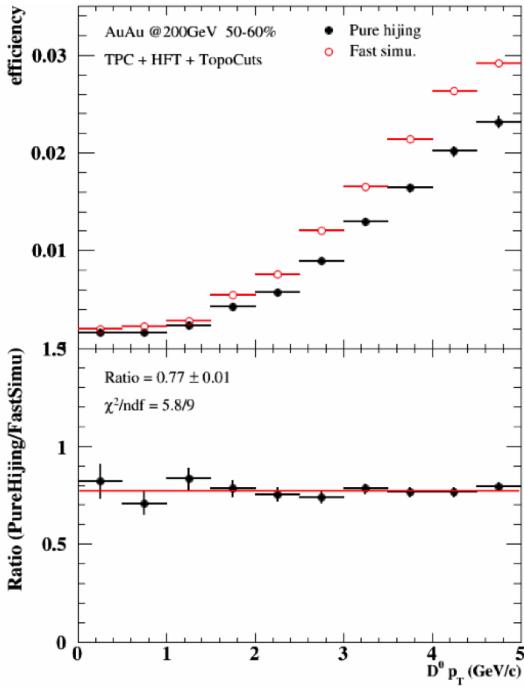


Figure 83: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation (red). (bottom) Double ratio to Fast-Simulation.

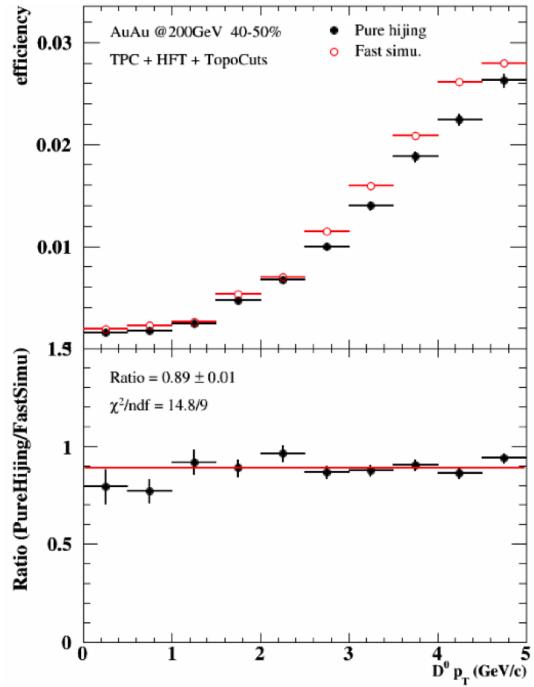


Figure 84: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation (red). (bottom) Double ratio to Fast-Simulation.

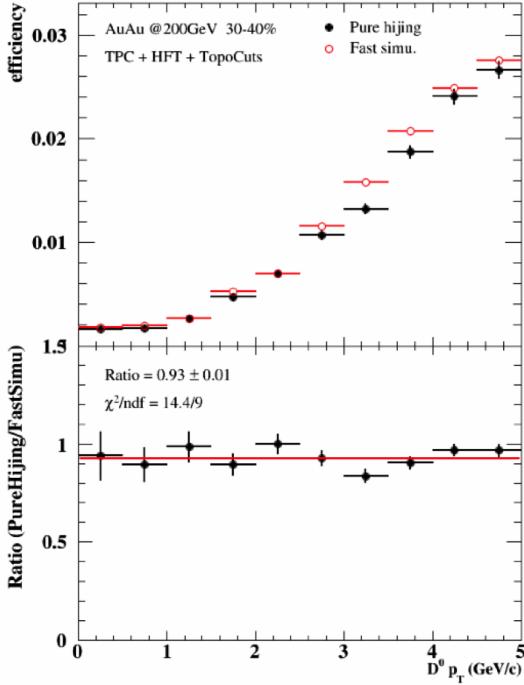


Figure 85: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation (red). (bottom) Double ratio to Fast-Simulation.

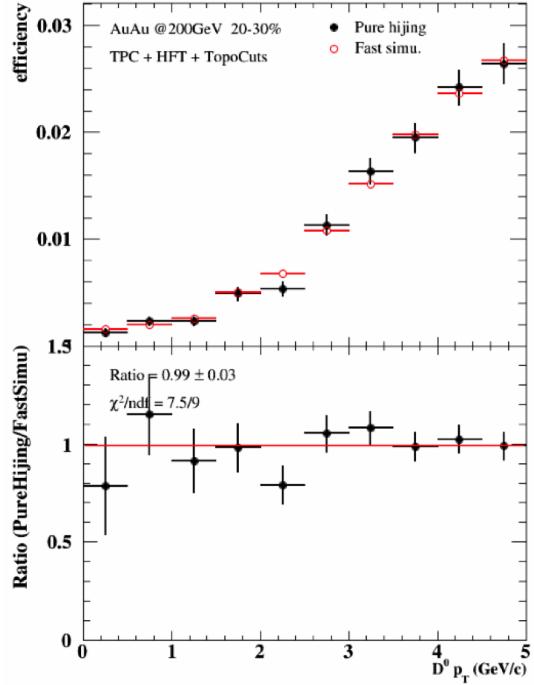


Figure 86: The comparison of  $D^0$  TPC + HFT match + Topological between Hijing (black) and Fast-Simulation (red). (bottom) Double ratio to Fast-Simulation.

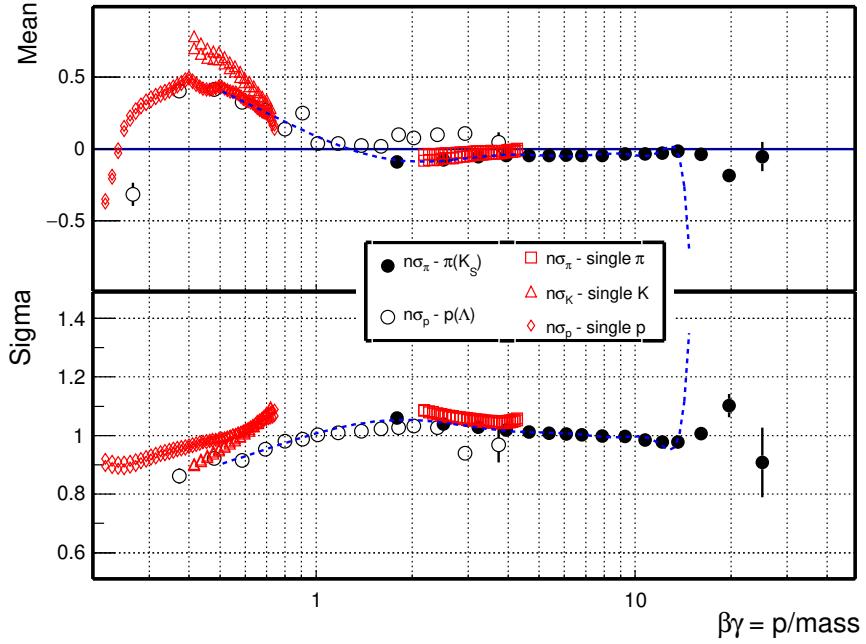


Figure 87: Top 0-10% central Au+Au collisions: extracted  $n\sigma_X$  mean values vs.  $\beta\gamma$  for both methods discussed in the text.

obtained central value of  $v_2$ . However, the statistical errors could be slightly off. For the spectra analysis, we need to consider this effect.

The double counting probability estimation need a precise determination of the PID variable distributions,  $n\sigma_X$  from  $dE/dx$  and  $1/\beta$  from TOF. For  $dE/dx$  calculation, we tried two methods

- 1) Select pure pion and proton samples from weak decays ( $K^S, \Lambda$ )

- 2) Look at the single particle distributions directly and perform multi-component fit in the region where the  $dE/dx$  bands can be separated out.

Figure 87 summarized the extracted  $n\sigma_X$  mean values vs.  $\beta\gamma$  for both methods discussed above. It looks good that in the overlapping region between different particles and different methods, the results look consistent. The dashed blue lines are parametrized function fits to the data points. These will be used to estimated the mis-identification probability.

The PID of  $D^0$  daughters also involves the TOF detector. We also estimated the TOF PID variable  $1/\beta$  distributions and TOF matching/PID efficiency. Figure 88 shows the fit results on the mean and width values for  $1/\beta - 1/\beta_{expected}$  distributions vs. particle momentum for different particles. Similar as  $dE/dx$ , results in the region beyond the TOF PID are not reliable. We use results which are safe in PID for later analysis, which are  $p < 1.5$  GeV/c for pions,  $p < 1$  GeV/c and kaons and  $p < 2$  GeV/c for protons. It is good to see the mean and width values are quite stable in a broad momentum region. At very low momentum, the multiple scattering effect will increase momentum resolution and  $1/\beta - 1/\beta_{expected}$  spread, and track energy loss will also shift the mean of  $1/\beta - 1/\beta_{expected}$  away from 0. But these will not affect the study here since tracks with  $p_T < 0.6$  GeV/c are not used for  $D^0$  reconstruction.

With all these at hand, we can evaluate the PID efficiency. The advantage of using TOF only when available is to keep the highest efficiency as one sees from the plot. The mis-identification probability for pion and kaon daughters can be also evaluated, as shown in Figure 89.

With the misidentification probability, we can reconstruct the invariant mass distributions from doubly mis-PID. The momentum resolution for pion and kaon tracks are chosen to fit to the  $D^0$  signal peak. Figure 90 and Figure 91 show these distributions compared to the signal distributions in different  $D^0 p_T$  regions. The distributions are normalized to the input real  $D^0$  signals.

Figure 92 shows the final estimated double-counting contribution to the real signal with two different calculation methods. The black symbols show the result from directly counting the entries within  $2.5\sigma$  of the  $D^0$  mass window. In real data analysis, we used the side-band distributions to normalize our fit or estimate our background. The blue data points show the result by subtracting also the side-band distributions with the same mass window selection as in event plane method  $v_2$

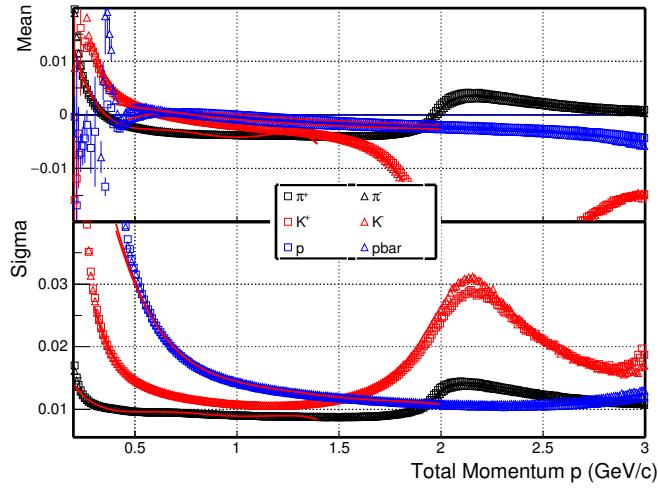


Figure 88: Top 0-10% central Au+Au collisions: extracted  $1/\beta - 1/\beta_{expected}$  mean and width values vs. particle momentum for pions, kaons and protons. Results for pions and kaons at  $p > 1.5$   $\text{GeV}/c$  are beyond the TOF PID capability. The fit results are not reliable.

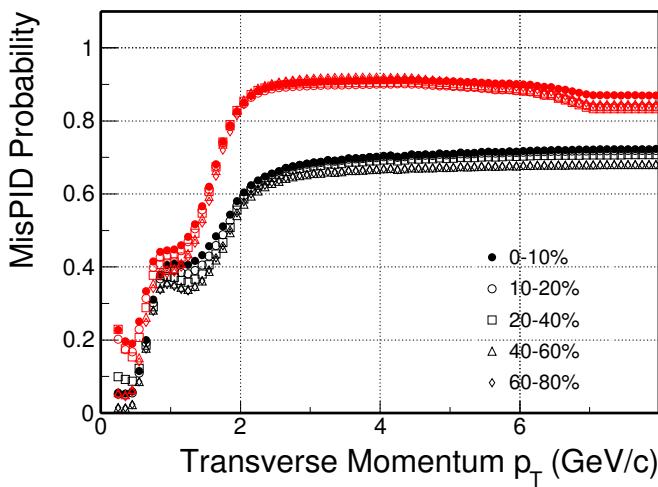


Figure 89: Particle misidentification probability for kaons (red) and pions (black) from different centrality bins in Au+Au collisions.

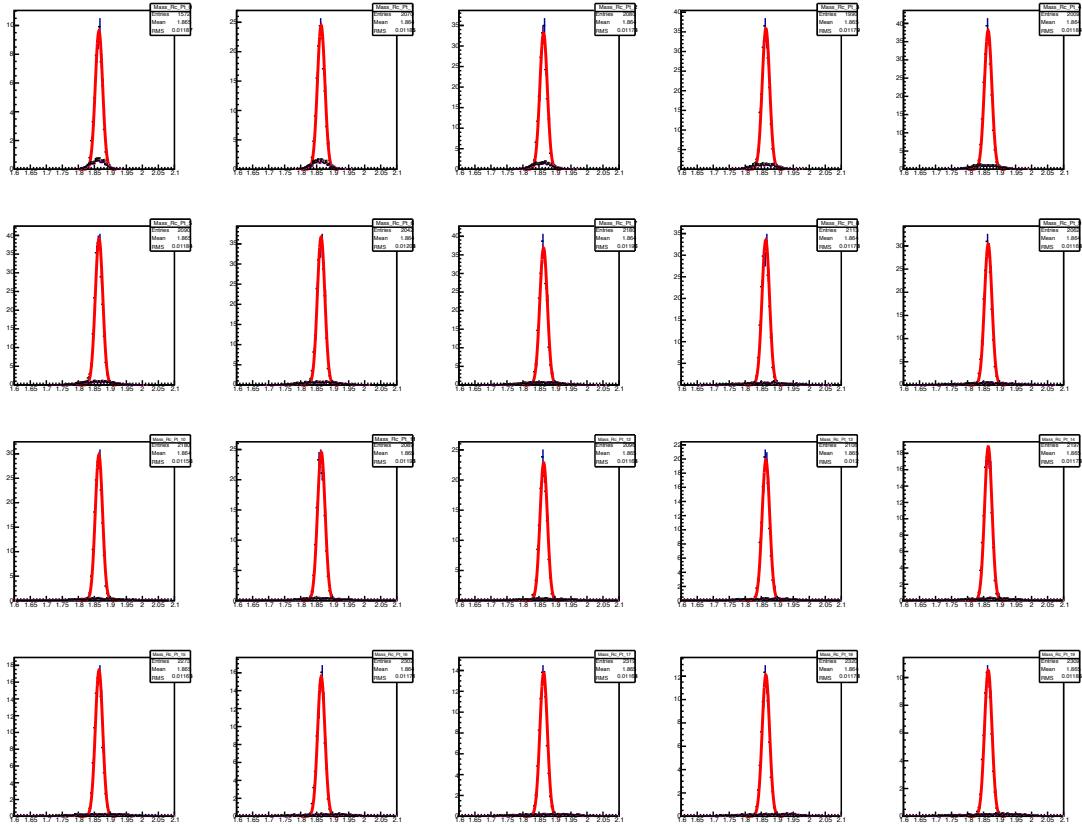


Figure 90: Reconstructed  $K\pi$  invariant mass distributions from clean PID and doubly mis-identification. The relative magnitude is fixed according to the realistic mis-identification probability. From top left to bottom right shows the distributions in  $p_T$  bins 0-0.1 GeV/ $c$ , 0.1-0.2 GeV/ $c$ , ..., 1.9-2.0 GeV/ $c$ .

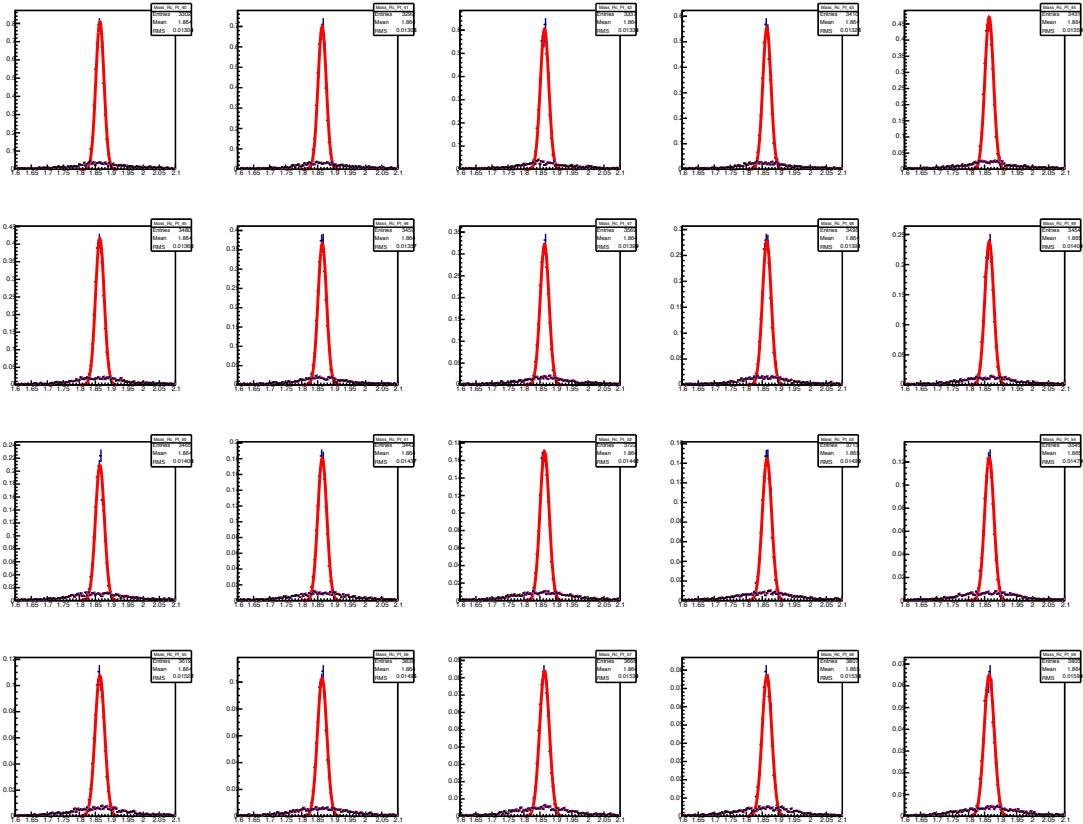


Figure 91: Reconstructed  $K\pi$  invariant mass distributions from clean PID and doubly mis-identification. The relative magnitude is fixed according to the realistic mis-identification probability. From top left to bottom right shows the distributions in  $p_T$  bins 4-4.1  $\text{GeV}/c$ , 4.1-4.2  $\text{GeV}/c$ , ..., 5.9-6.0  $\text{GeV}/c$ .

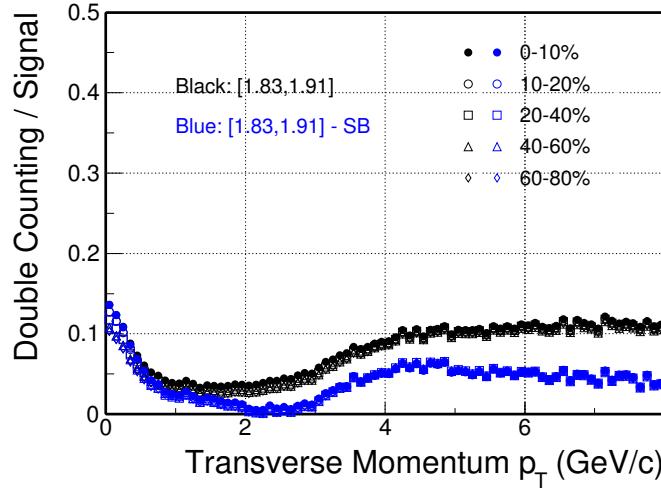


Figure 92: Estimated doubly-counted  $D^0$  fraction to the real signal with two calculation methods from different centrality bins in Au+Au 200 GeV collisions.

calculation.

In general, the influence by double mis-PID is relative small. For the spectra analysis, we use the blue data points in Fig. 92 as the central correction value and quote the range from 0 to the black data points as systematic errors on double-counting correction.

## 5 Systematic Uncertainties

The approach for the  $D^0$  spectrum systematic uncertainties are well studied. Several sources can be contributed to the uncertainties. The first one is coming from the raw yield extraction. We varied the signals extraction methods instead of use binning counting methods, we tried using fitting method, tried vary the fitting range, and also tried using likesign for background to estimate the uncertainties for raw yield extraction.

The second source would be coming from the TPC embedding uncertainties, this one is well studied by comparison the nHits and Dca distributions between data and embedding. Here the real data is without HFT in tracking. This is a self-produced testing sample with the office library and chain option. Fig. 95, Fig. 93, Fig. 96 and Fig. 94 show the nHits and Dca comparison between real data and embedding for 0-80% centrality. Bottom panels show the double ratio, and a fitted line as guidance. As seen, the embedding sample can reproduce the real data reasonable well. Then the systematic uncertainties can be estimate by varying the cuts from  $n\text{Hits} > 20$  to  $n\text{Hits} > 25$ , and  $\text{Dca} > 1.0$  to  $\text{Dca} > 1.5$  as Equ. 15.

$$\begin{aligned} r_{n\text{Hits}} &= (n\text{Hits} > 25)/(n\text{Hits} > 20)_{\text{data}}/(n\text{Hits} > 25)/(n\text{Hits} > 20)_{\text{MC}} \\ r_{\text{Dca}} &= (\text{Dca} < 1.0)/(\text{Dca} < 1.5)_{\text{data}}/(\text{Dca} < 1.0)/(\text{Dca} < 1.5)_{\text{MC}} \end{aligned} \quad (15)$$

Then after varying nHits and Dca, the comparison of data and embedding shown as Fig. ?? for nHits (left) and Dca (right). Fig. ?? shows the comparison of the HFT Matching ratio due to the change of Dca cuts for 0-80% centrality. As seen, due to the change of Dca cut, the contribution for Embedding and HFT matching ratio can be canceled out at some point. So in the final result we do not include this contribution from Dca. The bottom panels show these double ratio, and quote as systematic uncertainties. For the  $R_{CP}$  calculations, at some point they are correlated for central and peripheral collisions and somehow can be canceled out. So, for the  $R_{CP}$  systematic, we use  $r_{n\text{Hits}} - \text{cent}/r_{n\text{Hits}} - \text{peripheral}$  to calculate the uncertainties. The nHits, Dca (not included) and TofPID (as discussed in the pid part) systematic are added up quadratically for single track. For the  $D^0$  calculations, the pion and kaon tracks are added linearly.

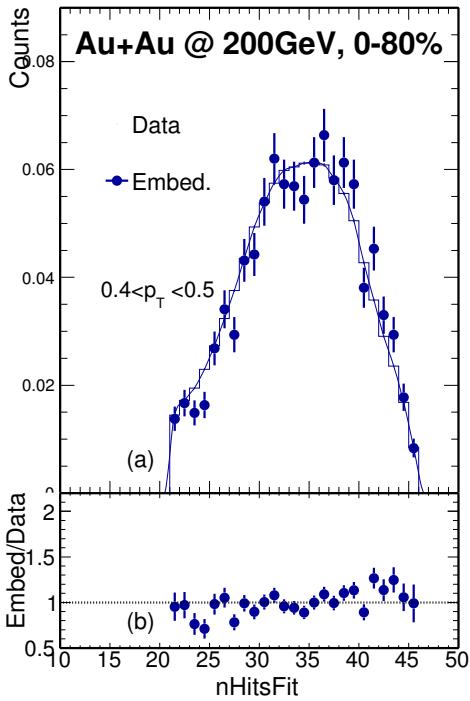


Figure 93: nHits comparison between real data and embedding for 0-80%.

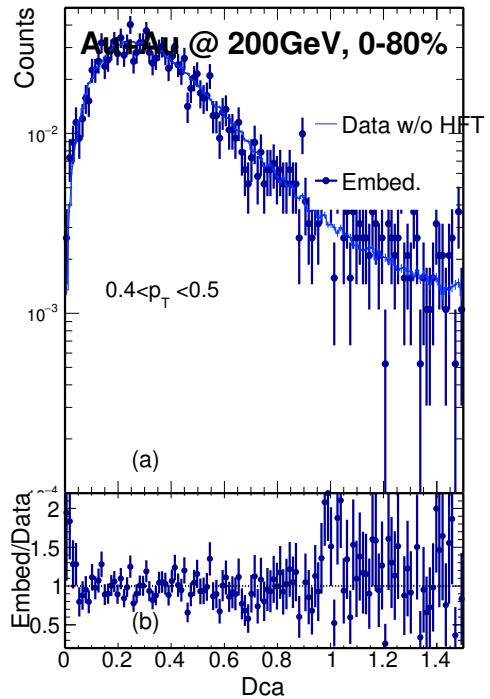


Figure 94: Dca comparison between real data and embedding for 0-80%.

The next source is by varying the topological cuts and daughter  $p_t$  cuts. The standard TMVA cuts, the 50% efficiency and 150% efficiency cuts are calculated, and also the daughter  $p_T$  cuts are

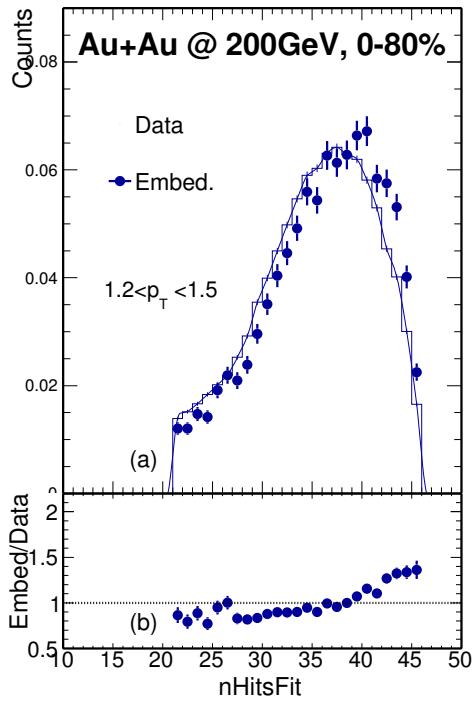


Figure 95: nHits comparison between real data and embedding for 0-80%.

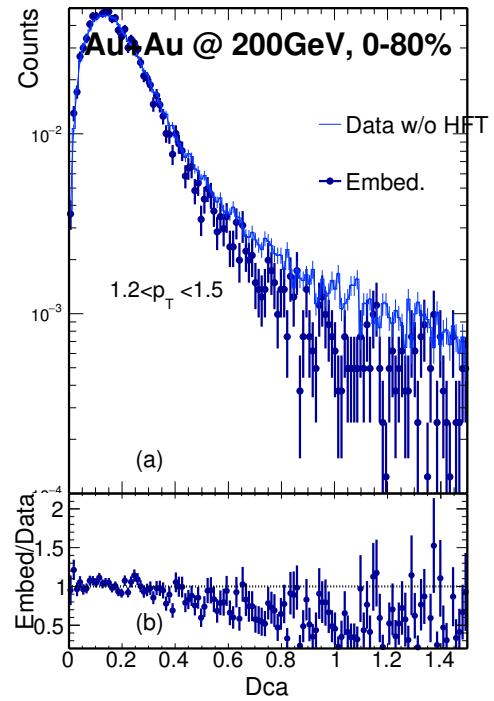


Figure 96: Dca comparison between real data and embedding for 0-80%.

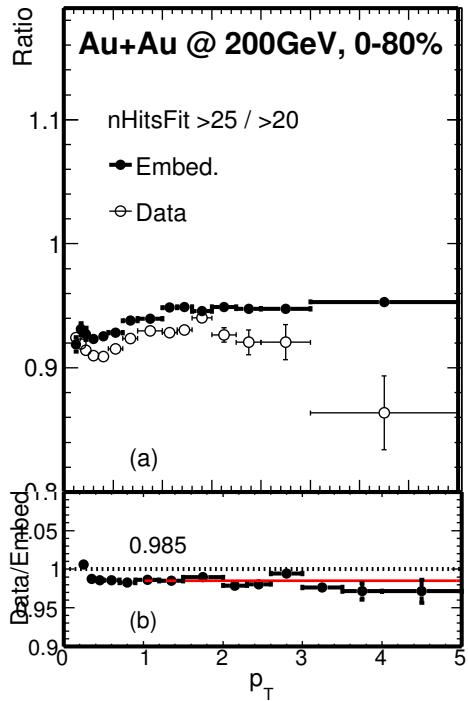


Figure 97: Varying nHit(left) and Dca(right) comparison between real data and embedding for 0-80%.

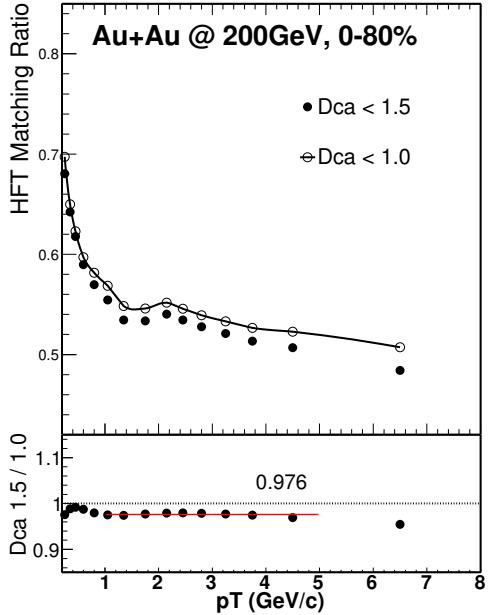


Figure 98: HFT Matching ratio comparison with the Dca cut varying from 1.5 to 1.0 for 0-80%

checked for 300 MeV and 500 MeV beside of the default value 600 MeV. The difference between the corrected yield are quoted as the systematic source. Also there is a source coming from the double counting and vertex resolution contribution as we discussed in the previous sections.

Note here, for the  $R_{CP}$  calculations since there are correlated uncertainties, so similar as TPC part, for the daughter  $p_T$  scan and topological cuts scan uncertainties, we calculate the  $R_{CP}$  for each individual setup or cuts, then quota the difference of  $R_{CP}$  for each cuts as systematic errors. For the others, such as yield extraction, double counting and vertex resolutions, those uncertainties was quadratical added up for  $R_{CP}$ .

Fig. 99 to Fig. 102 shows the  $D^0$  spectrum different sources contribution in various centralities. As we see, the systematic uncertainties is quite small in the most of the  $p_T$  range except some of the  $p_T$  ranges due to the limited statistics and large contribution from yield extraction.

## 6 Validation with Ks Spectra Measurement ... To Be Added

To verify the data-driven fast-simulation procedures, we calculated the efficiency corrected spectra for  $K_s$  in 10-20% centralities using the data-driven methods. The result shows very good agreement with the published result as in Fig. 103. The lower panel shows the double ratio comparison between run 14 and the published one.

## 7 Results ... To Be Added

## 8 Re-analysis of Run10/11 data ... To Be Added

### 8.1 Long and Yifei's re-analysis ... To Be Added

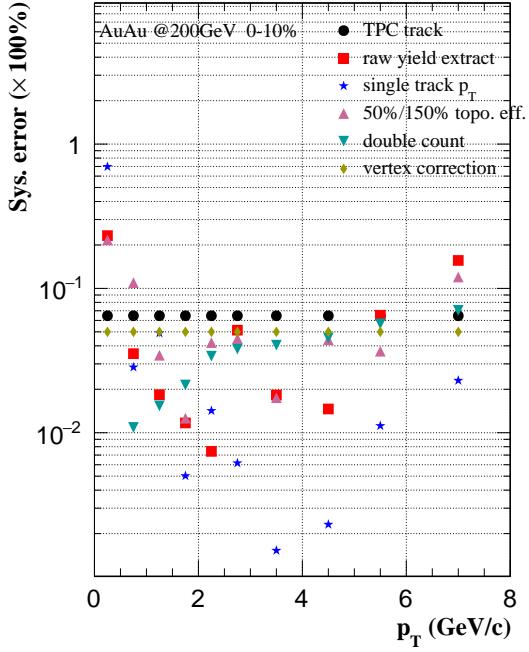


Figure 99: Systematic uncertainties from different sources for 0-10% spectra.

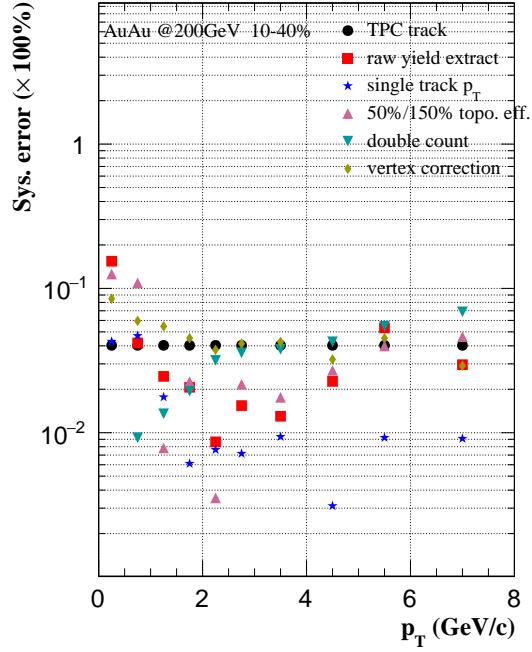


Figure 100: Systematic uncertainties from different sources for 10-40% spectra.

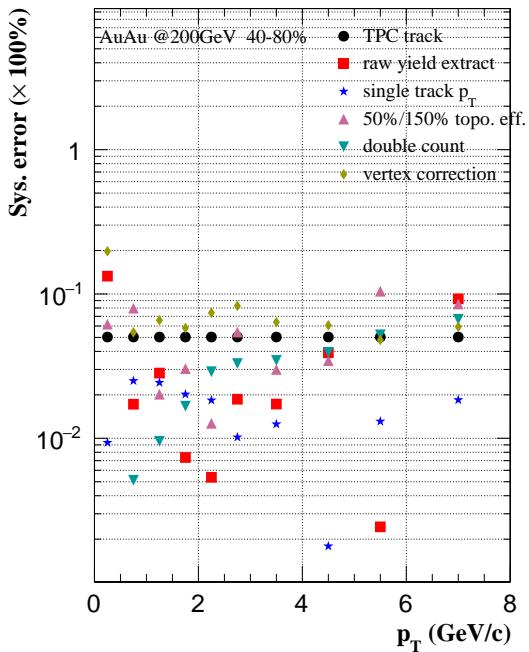


Figure 101: Systematic uncertainties from different sources for 40-80% spectra.

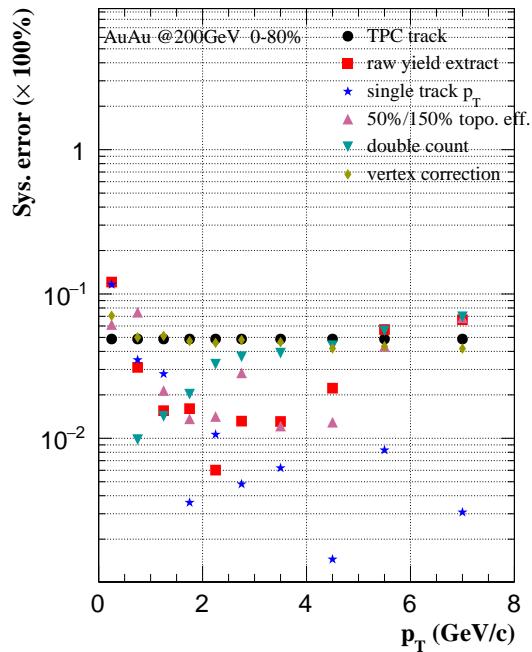


Figure 102: Systematic uncertainties from different sources for 0-80% spectra.

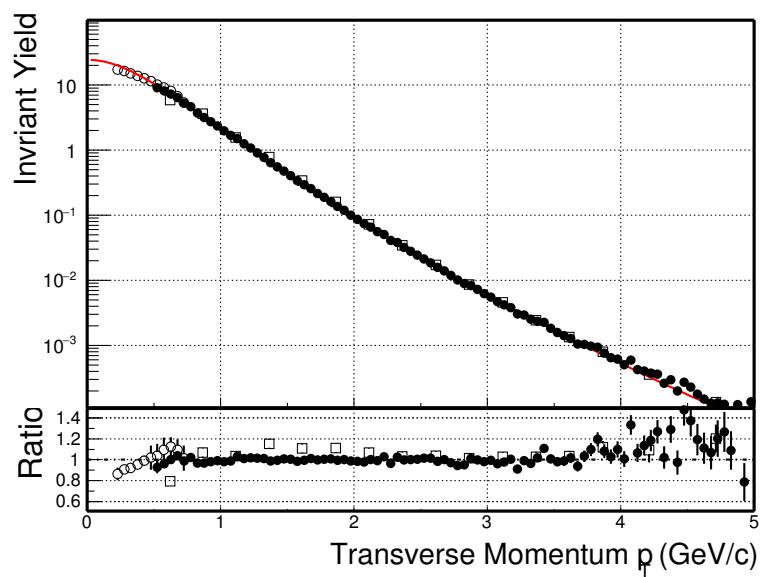


Figure 103:  $K_s$  spectra.

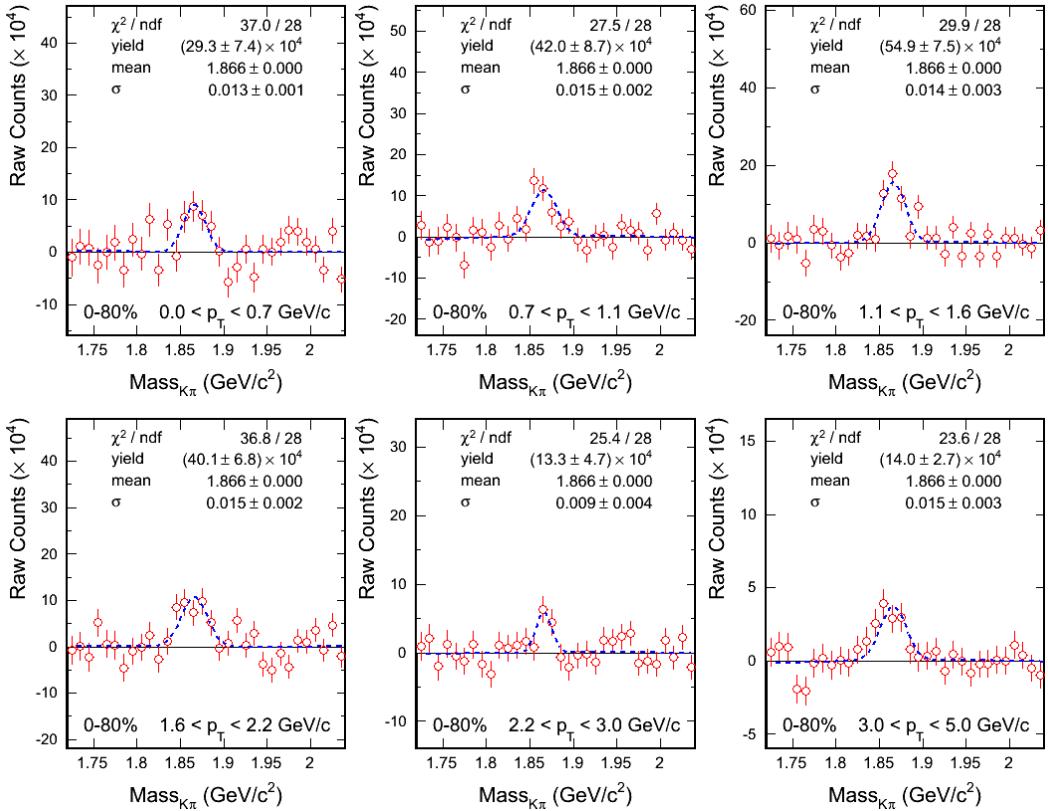


Figure 104:  $p_T$  dependence of  $D^0$  signals in 0-80% from partial Run11 data.

Since at  $p_T < 2$  GeV/c a discrepancy was found between new data from Run14 with HFT and published data from Run10+Run11. A re-analysis on Run11 data was performed to check if anything was incorrect. The analysis cuts are the same as in previous analysis note:

<https://drupal.star.bnl.gov/STAR/system/>

files/dzeroAuAu\_updated.pdf, also listed in Xiaolong 's analysis, see next section. The only difference is varying DCA cut with  $< 1$  cm or  $< 2$  cm.

The raw signals as a function of  $p_T$  from part of Run11 data with 85% of full statistics were extracted as Fig.104,

The raw signals were compared with Xiaolong 's independent analysis and found to be consistent, shown in Fig.105. The slight lower yield in Xiaolong 's analysis is due to the tight DCA cut  $< 1$  cm, while the DCA cut applied here is  $< 2$  cm.

In early days, there was no vertex detector, we had to use random combination of decay products to reconstruct  $D^0$  meson in hadronic channel, which suffered from huge combinatorial background. The way to improve the significance was to enrich the kaon and pion PID probability and enhance the statistics as much as possible at the same time. Thus we developed a hybrid method: at low  $p_T$  we used TOF PID to ensure the purity of kaons and pions, at high  $p_T$  , if there was an available TOF matching, we applied TOF PID, otherwise we used TPC PID to enhance the PID efficiency. But an issue was found in the code, that the condition to reject those tracks without matching to TOF at low  $p_T$  was different from what we used to calculate the efficiency. In data analysis, we reject those tracks with no TOF matching, while in calculating efficiency we reject those tracks with no TOF matching and those with no valid  $\beta$  information in TOF. This introduces about 15% difference at low  $p_T$  but does not affect high  $p_T$  much.

The other issue is that we applied an additional  $DCA_{XY}$  cut efficiency in previous analysis. This cut was used in TOF matching algorithm (in TOFMatchMaker) to require tracks with a distance of closest approach to the beam line in xy plane within 1 cm. This efficiency was studied in early days and found there was about up to 20% of tracks at low  $p_T$  rejected with this cut, see Fig.107. However, when we studied TOF matching efficiency, we used a data driven method and this efficiency should be already included. Thus we double counting this efficiency.

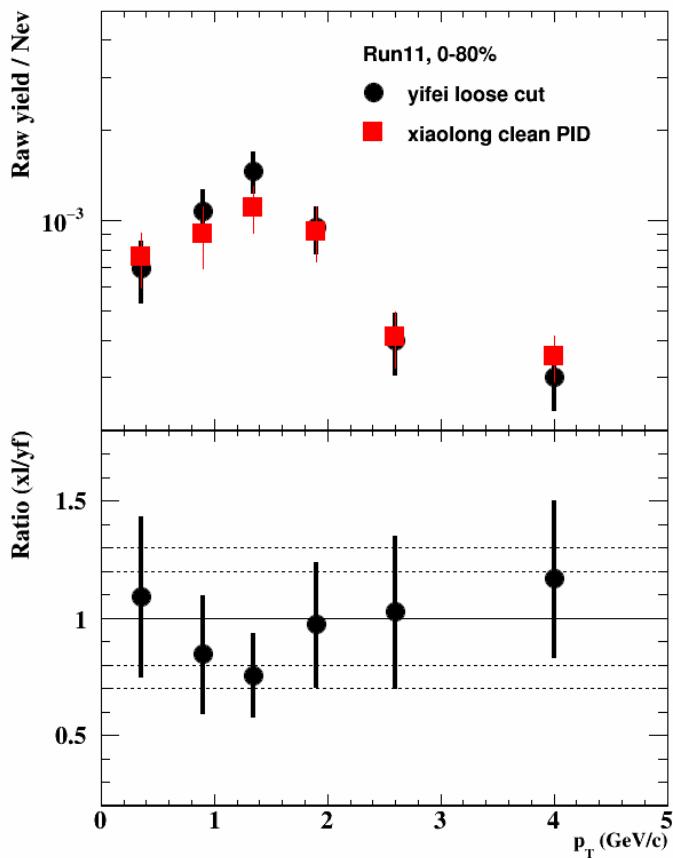


Figure 105: Raw signals comparison with Xiaolong 's.

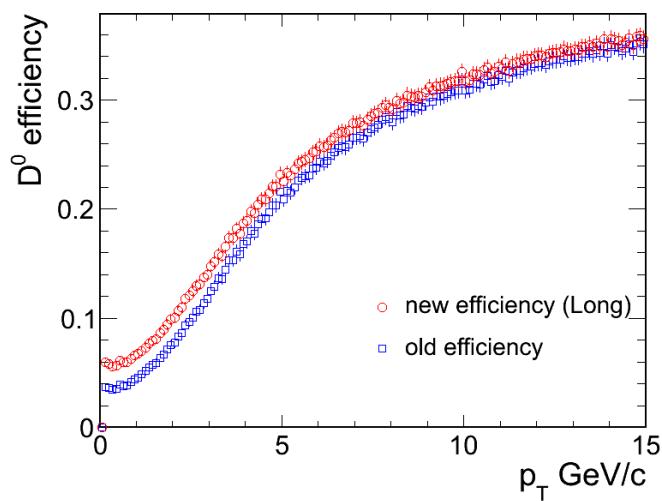


Figure 106: New efficiency checked from Long compared with old one.

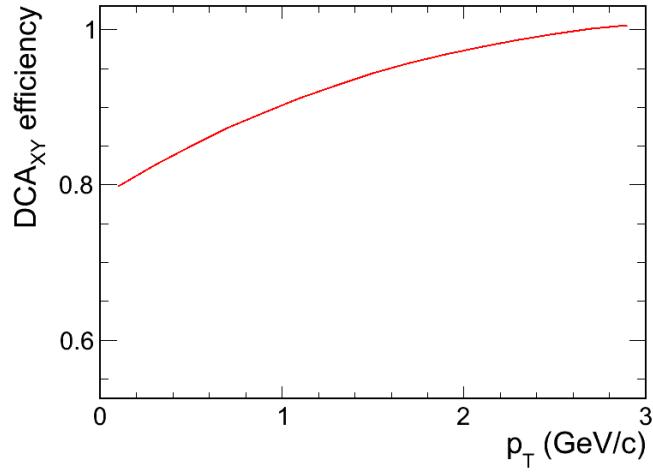


Figure 107:  $DCA_{XY}$  efficiency.

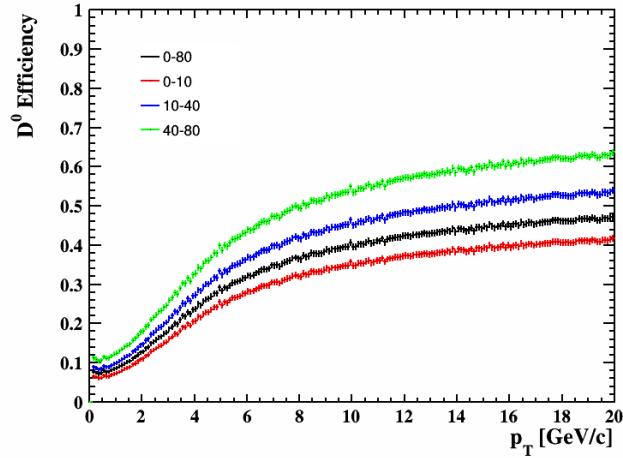


Figure 108: The reconstruction efficiency at each centrality bin

After corrected above two main issues, the corrected reconstruction efficiency can be found at Fig.108

The final  $D^0 R_{AA}$  spectra will be corrected by efficiency ratio between the published reconstruction efficiency and the corrected reconstruction efficiency. The corrected  $D^0 R_{AA}$  is shown in Fig.109

*Need the  $R_{AA}$  comparison plots between my  $R_{AA}$  and Run 14  $R_{AA}$ (from Xiaolong)*

More material can be found at here :

[https://drupal.star.bnl.gov/STAR/system/files/D0\\_Eff\\_discussion\\_2016\\_11\\_7.pdf](https://drupal.star.bnl.gov/STAR/system/files/D0_Eff_discussion_2016_11_7.pdf)  
[https://drupal.star.bnl.gov/STAR/system/files/D0\\_Eff\\_discussion\\_2017\\_05\\_12.pdf](https://drupal.star.bnl.gov/STAR/system/files/D0_Eff_discussion_2017_05_12.pdf)  
[https://drupal.star.bnl.gov/STAR/system/files/D0\\_BNL170516\\_0.pdf](https://drupal.star.bnl.gov/STAR/system/files/D0_BNL170516_0.pdf)

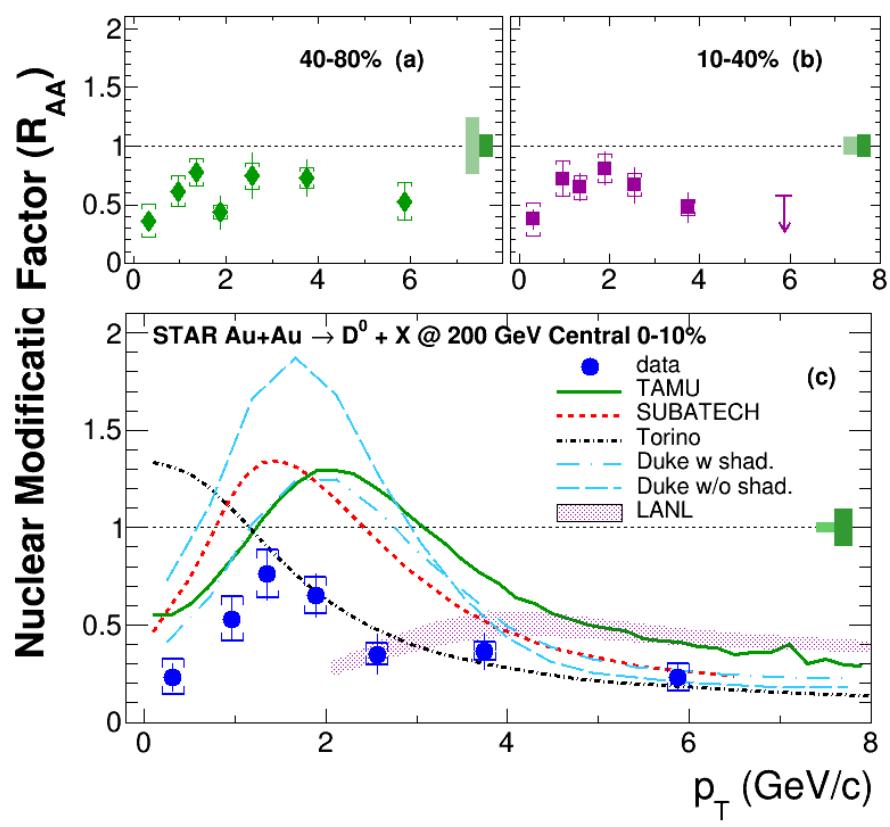


Figure 109: The Corrected  $D^0$   $R_{AA}$  in each centrality class, and compared with several model calculation