JOINT INSTITUTE FOR NUCLEAR RESEARCH
Veksler and Baldin laboratory of High Energy Physics

FINAL REPORT ON THE
START PROGRAMME

Track analysis on the framework of the MPDroot

Supervisor:
Alexey Aparin

Student:
Rodrigo Guzmán Castro, Mexico
Nuclear Sciences Institute,
National Autonomous University
of Mexico

Participation period:
July 3rd – August 25th,
Summer session, 2023

Dubna, 2023

Abstract

A systematic study of the behavior of the reconstructed tracks under several conditions has been done in this work. This study only considers some effects presented in the reconstruction of the tracks, thus, calculations on the efficiencies of track reconstruction has been the main focus. Different Monte Carlo produced data has been employed for this analysis, labeled as request 25, 30 and 31, which consist on Bismuth+Bismuth collisions at 9.2 GeV produced at the JINR with UrQMD, PHSD and UrQMD generators respectively. Important differences have been found for the case of the request 25, and some attention should be put when making use of this data.

Contents

Acknowledges

Chapter 1

Introduction

1.1 Motivation

The Multi-Purpose Detector (MPD) is one of the 3 main experiments at NICA facility. Its construction is planned as a 2 stages project; in the first stage, all the necessary subsystems will be implemented to enable initial measurements of particle momentum, tracking, centrality, and collective flow studies. This stage focuses on establishing the core functionality of the detector and conducting fundamental research. The second stage of construction will involve the incorporation of additional detectors to further enhance the capabilities of the MPD. These additional detectors will enable the detection of muons originating from external sources, expanding the range of measurements that can be performed. The MPD aims to broaden the scientific scope of the phenomena occurring in the baryon-rich region of the QCD phase diagram.

Once the MPD is under operation and the correct calibration is completed, basic analysis on the spectra of the data can be performed. Typically, a Monte Carlo embedding technique is employed to calculate the efficiency and contamination due to the identification process. This process consists on the blending of Monte Carlo (MC) and raw data, a full simulation of the reconstructed process is done to the MC data and the mixture is treated as real data. In order to perform the corrections, the mixed events should reproduce the characteristics of the real data, and then efficiencies on the track reconstruction can be performed by an association procedure [1, 2].

Given that the detector is still under construction, only analysis over the MC generated data can be performed. The generated data is treated as real data, then detector efficiencies can be estimated by simulating the passing of data through the full detector configuration. For this work, no identification process is considered (or more precisely, particle identification is 100% accurate), this should be considered as the most ideal case, as if the detectors considered for this analysis were as perfect as the algorithms for the reconstruction process itself could be.

4

1.2   MPD layout configuration

The first stage of the MPD consists of several components shown in Figure 1.1. The structure of the full apparatus reminds that of a **матрёшка**, and a brief description is given below, following Refs. [3, 4]:

- Time Projection Chamber (TPC): The central barrel of the MPD houses the TPC, a cylindrical structure composed of 12 modules. The TPC is 340 cm long with inner and outer radii of 27 cm and 140 cm, respectively. It is divided into two halves by a central electrode, creating a uniform electrical field of 140 V/cm along the axis. Charged particles passing through this region ionize the gas mixture of $90\%Ar+10\%CH_4$, and the ionization charge is collected by the Read-out chambers (ROC), which consist of Multi-Wire Proportional Chambers (MWPC).

- Time Of Flight system (TOF): Surrounding the TPC, the TOF consists of 14 plate sectors, each formed by two modules. The TOF has a total length of 590 cm and employs the Multigap Resistive Plate Chambers (MRPC) technology (280 in total). The gas mixture used in the TOF is 90% $C_2H_2F_4$, 5% SF_6 and 5% i-C_4H_{10}.

- Electromagnetic calorimeter: The MPD features an electromagnetic calorimeter consisting of 50 isolated half-sectors forming a cylindrical shell with a length of 624 cm. It has inner and outer radii of approximately 168 cm and 230 cm, respectively. Each half-sector contains 48 calorimeters: (8 modules of different types in the longitudinal direction) × (6 modules in the transverse direction). Each module consists of of 16 towers of 40×40 mm^2 transverse cross-section with a lead-scintillator sandwich that contains 210 tiles of Pb interleaved with 210 tiles of plastic scintillator.

- Forward Hadron Calorimeters: Includes two identical Forward Hadron Calorimeters placed at 3.2 m upstream and downstream from the center of the MPD. Each calorimeter has 44 modules, and each module consists of 42 lead-scintillator sandwiches enclosed in a stainless steel box. Scintillator tiles are covered with a white reflector to enhance light collection, and compact SiPMs are coupled to the optical connectors at the rear side of the module.

- Fast Forward Detector (FFD): The FFD consists of two modular arrays with 20 Cherenkov modules. Each module comprises a lead converter, a quartz radiator, PMTs, and board circuitry. The FFD has an acceptance range of $2.7 < |\eta| < 4.1$, corresponding to a polar angle range of $1.9° < |\theta| < 7.3°$.

After the collision, a multitude of particles is emitted at all directions and interactions with the detectors occur. Each component of the MPD plays a crucial role in the identification process. The FFD is responsible for triggering A+A collisions and providing the start time required for the Time Of Flight (TOF) measurement. The track reconstruction is based on the drift time and $R - \varphi$ cylindrical coordinate measurement

of the primary ionization clusters created by the charged particles passing through the TPC. The minimal two-track resolution is ~ 1 cm, then a measurement of the specific ionization energy loss dE/dx can be applied on a track-by-track basis. Momentum reconstruction relies on the curvature of the reconstructed tracks, enabling the identification of charged particles with transverse momenta (p_T) greater than 50 MeV/c. The TPC alone can discriminate between charged pions and kaons up to momenta of about 0.7 GeV/c, and between kaons and protons up to approximately 1.1 GeV/c. The TOF detector provides both time and coordinate measurements with an accuracy of around 80 ps and 0.5 cm, respectively. It employs a matching procedure to associate TPC tracks with hits in the TOF detector. The matching involves extrapolating the TPC track to the TOF surface and finding the nearest TOF hit within a predetermined window (matching window). The size of the matching window is determined to balance the TOF's intrinsic performance (time and coordinate resolutions) with the overall occupancy in heavy-ion collisions. The expected reconstruction efficiency is approximately 80% for particles with momenta up to 1.7 GeV/c. The primary role of the ECal is to measure the spatial position and total deposited energy of electromagnetic cascades induced by electrons and photons produced in heavy-ion collisions.
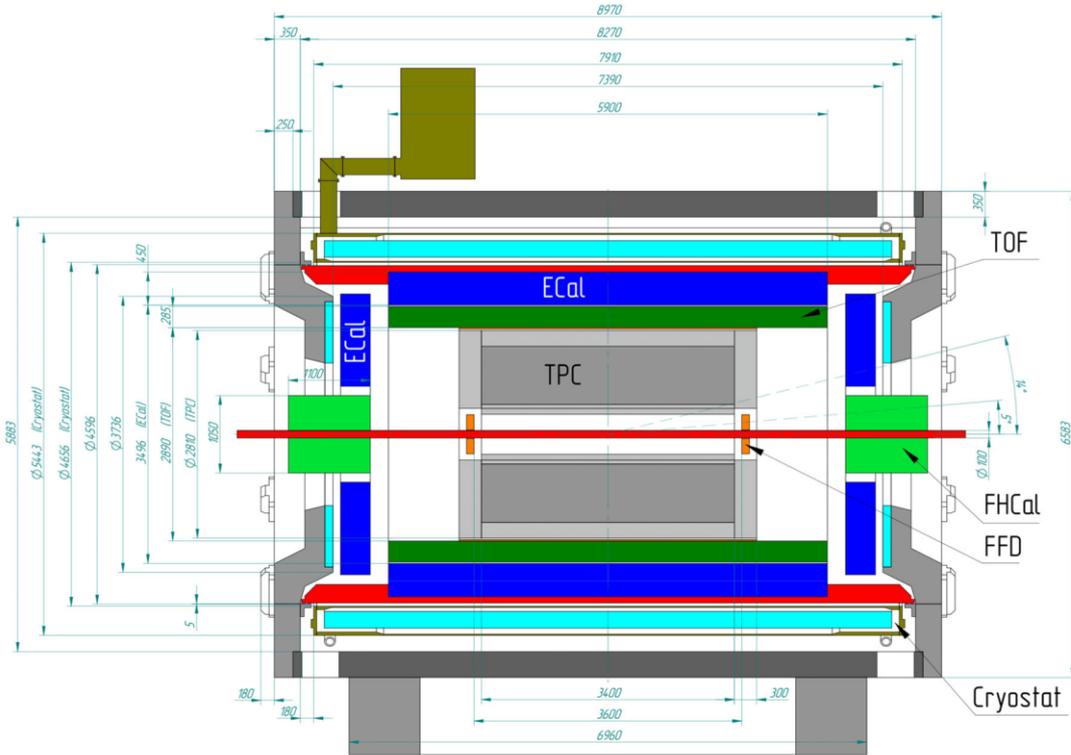

Figure 1.1: Cross section view of the MPD detector, including: time projection chamber, time of flight system, electromagnetic calorimeter, forward hadron calorimeters and fast forward detector. Figure taken from Ref. [3]

1.3 Software environments

The Monte Carlo generated events used for this analysis consist on three different sets of data produced at the JINR cluster for different purposes, but each data set consists on Bi+Bi collisions at 9.2 GeV. The request 25 is described for general purpose studies in the MPDforum, and the MC production was done with the Ultra relativistic Quantum Molecular Dynamics (UrQMD) event generator. On the other hand, request 30 is a sample for femtoscopy studies, produced by the Parton-Hadron-String Dynamics (PHSD) event generator. Request 31 is done for femtoscopy-purpose, and generated by UrQMD. The geometry of the detector as well as the full simulation of the interaction of the passing particles through the detector is done with GEANT4. And the reconstruction and analysis of the data is done within the MPDroot enviroment [5].

The MPDroot root developers version 23.03.23 was used, and so the recent wagon system, but given the type of analysis no other analysis train was needed. If in the future is necessary to include additional wagons such as centrality or particle identification for a more complete analysis of this kind, the compatibility is almost immediate.

1.4 Purpose of the study

A typical collision at which the MPD will be operating $\sim$(4-11) GeV can produce thousands of particles, out of which the reconstruction algorithm should be able to create a full mapping of the tracks left by the passing particles through the detectors, and then correctly match information obtained from different detectors which corresponds to that singular track. This process is by itself not perfect, and not every particle of interest will leave a track or the necessary information to make an adequate reconstruction. The main goal of this work is to summarize some of the characteristics of those reconstructed tracks, for example, how are the tracks affected when imposing different cuts on the variables needed to make the reconstruction process. Different effects on the tracks will be described and quantified. Also efficiencies when comparing to the MC are given per particle species. The idea is to acquire some insights on what we could expect when the analysis with the real data is done, and so the corrections and systematic errors are properly taken into account.

Chapter 2

Analysis of the data

2.1 Monte Carlo Association procedure

In order to study the efficiencies we need a proper way to "track" the tracks. To this end an association is made between the generated MC tracks and the reconstructed tracks. In the framework of the MPDroot we can extract the information of all the reconstructed tracks from MpdGlobalTracks. Whereas the information of the MC is stored in the MCtracks. Now the process consists of obtaining the ID of the reconstructed tracks, which is just a label put during the reconstruction process to the track. Then you can associate each reconstructed track to the real track (the MC track). The advantage is that you can then obtain information of the reconstructed tracks coming directly from the MC, which is the real information to this end. This means that every variable you measure will be 100% accurate, starting with the identity of the particle (its pdg code) corresponding to a particular track. A simple example of the code is as follows:

```
1  fTMCTracks = event.fMCTrack; // MC tracks of the event
2  Int_t nmctracks=fTMCTracks->GetEntriesFast(); // No. of MC tracks
3
4  fTDstEvent = event.fMPDEvent;
5  fTMpdGlobalTracks = event.fMPDEvent->GetGlobalTracks(); // Global rec.
       tracks of the event
6  TClonesArray *MpdGlobalTracks = (TClonesArray*)fTDstEvent->
       GetGlobalTracks();
7  Int_t ntracks=fTMpdGlobalTracks->GetEntriesFast(); // No. of rec. tracks
8
9  for (Int_t i = 0; i < nmctracks; i++){  // loop over MC
10   MpdMCTrack *MCtrack = (MpdMCTrack*) fTMCTracks->UncheckedAt(i);
11   Double_t ptmc = MCtrack->GetPt();
12   Int_t pdgmc = MCtrack->GetPdgCode();
13   }
14
15  for (Int_t i = 0; i < ntracks; i++){  // loop reconstructed tracks
16   MpdTrack *track = (MpdTrack*)fTMpdGlobalTracks->UncheckedAt(i);
17   Int_t ID = track->GetID();
18   MpdMCTrack *MCtrack = (MpdMCTrack*) fTMCTracks->UncheckedAt(ID);
19   Double_t ptmc = MCtrack->GetPt();
```

```
20    Int_t pdgmc = MCtrack ->GetPdgCode ();
21    }
```

The macro consist on mainly two loops: which runs over the MC and reconstructed tracks. All the information corresponding to the MC should then be filled within the MC loop and the Associated Monte Carlo (AMC) is filled inside the reconstructed tracks loop. A comparison between the MC and reconstruction then tell us how is our original data being transferred from the raw MC to the reconstruction in an idealistic scenario of our detector's measurements.

2.2 Repeated tracks

One particular effect (which in reality is a conglomerate of different effects) is that of the repeated tracks. Repeated tracks (RT) are defined in this work as **tracks with the same ID for the same event**. The effect is divided in two types for this analysis. One corresponding to the Splitted Tracks (ST) and the other to the Ghost Tracks (GT). To distinguish properly between this two effects, it is necessary to mention that the TPC consist of 52 layers in the radial direction in which a hit (point of a track) can occur. This means that a given track can be (at most) reconstructed with 52 hits. Ghost tracks are considered as those repeated tracks which adding up their number of hits surpass this limit of 52 hits. Lets say there are N repeated tracks $RT_1, RT_2, \ldots, RT_N$ with $nh_1, nh_2, \ldots, nh_N$ number of hits respectively. If the condition $nh_1 + nh_2 + \cdots + nh_N > 52$ is met, those tracks (minus one) are labeled as ghost tracks. This means that there are some points shared between ghost tracks, which should not be happening in the case of splitted tracks. Just as a clarification, even if this is true, this only works as a lower bound for the real number of ghost tracks, because there can exist repeated tracks sharing some points and still having less than 52 hit points. The discussion is clearer if we look at the distribution of number of hits per number of tracks in figure 2.1a.

It is worth to mention that if we have M number of repeated tracks, there will only be one considered as the "real" track while the rest will be either splitted or ghost tracks. All the numbers presented in this work with the label "repeated tracks" are not taking into account the "real track". The selection criteria for this one track is that it should be the track with the most number of hits. This criteria even if somewhat arbitrary, makes sense if we consider that a track is better reconstructed whenever more points are used. A distribution of the number of hits per repeated track is shown in figure 2.1b, which shows that a considerable amount of this type of tracks can be avoided if we make a cut over the minimal number of hits per track.

2.3 Primary and secondary tracks

As stated previously, there are three main stages for the posterior analysis of the tracks, consisting on the simulation of the collisions via the MC event generator, then the simulation of the interaction of the particles generated with the detectors and finally
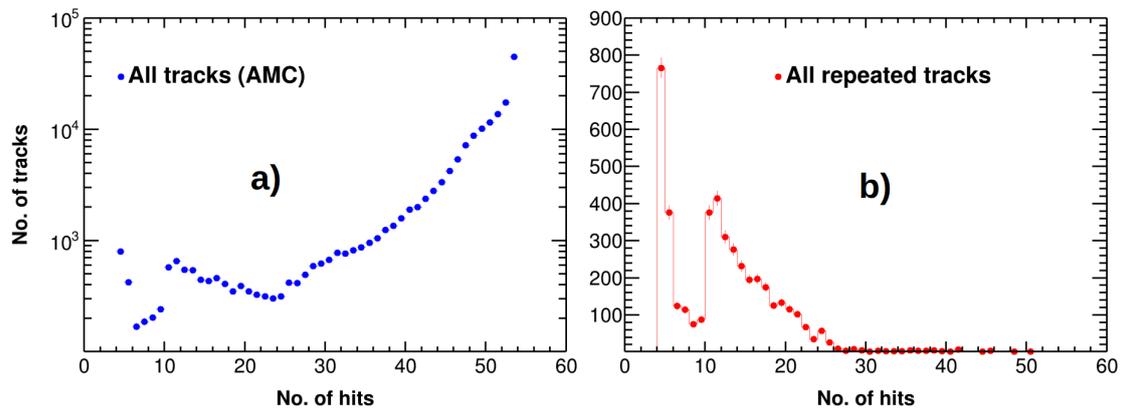
Figure 2.1: Number of hits distribution for Bi+Bi Collisions at 9.2 GeV. a) Distribution for all the Associated Monte Carlo tracks. b) Distribution only of the repeated tracks. These results corresponds to the request 25, but no differences appear for requests 30 or 31.

the reconstruction of the tracks. All of the particles produced until the end of the MC simulation will be considered, as for the reconstruction process corresponds, as **primary particles**. The rest of the particles, either coming from decays or interacting with the walls of the detectors or any other physical process will be considered as **secondary particles**. This is as for the reconstruction algorithm corresponds, but the physical reality considers the primary particles as those produced by the first interaction of two of the original nucleons of any of the two nuclei. Those process occur so quickly that in practice it would be an impossible task to measure and track them all, and so a good approximation is that considered by the reconstruction algorithm.

This makes it so easy to distinguish between primary and secondary tracks when information of the MC is provided. In the MPDroot a primary selection can be done by means of the motherID of the particle, which is just a label to distinguish if the particle comes from the raw MC data or no. Nevertheless in a experimental situation the problem becomes more complicated, and so a new variable is needed for the selection of the primary tracks. To do a proper selection, first we have to know the position of the primary vertex, this is the position where the collision occurs. The position of this vertex is determined by an extrapolation of all the tracks produced by a single collision event. Then we can individually extrapolate each particle's track to the nearest position to this primary vertex, this is called the Distance of Closest Approach (DCA). A DCA cut for each track should, in principle, make a reasonable job at distinguishing primary from secondary tracks, this will be important for example when doing centrality studies, because the particles of interest are the primaries, and thus an adequate primary vertex and track reconstruction is fundamental. Normally, a DCA is expressed as a radial distance (typically in a distance of the order $\sim$1cm) between the nearest point of the track to the primary vertex.

2.4 χ^2 and vertex cuts

There are two other variables considered for the analysis in this work. The first one is the χ^2, which for convenience of this analysis has been modified by dividing each tracks' χ^2 over its number of hits (χ^2/Nofh). This quantity is to represent the quality of the reconstruction of the track, then a good χ^2/Nofh should be less than a small given number. But the overall effect of a cut over the number of hits and this variable is expected to be similar because they are closely related. The shape of this distribution is shown in figure 2.2a.

The vertex has already been described in the previous section, but the main reason on why there is a necessity in making a cut on its position is to avoid bad events. In this case, bad events is referred to two main sources of problems. The first is produced along the beam direction (in the Z vertex), if the collision occurs at a considerable distance from the interaction point of the detector, a reduced multiplicity will be observed as a consequence of the loss of some particles because the collision is occurring at the edge of the detectors. As for the transverse (XY vertex) direction, there could be events occurring near the beam pipe, or directly collisions with the walls, which of course are not of interest. Typically the MC event generators are built in such a way that the collision occurs at coordinates of the vertex (0,0,0). Thus, the smearing of the vertex is given artificially when the simulation through the detectors is done. In the case of the three request under analysis, the smearing of the XY vertex is so small that is negligible. Whereas for the Z vertex the smearing is considerable, figure 2.2b shows the distribution of this vertex. The peaks at ± 150 cm most likely represents interactions with either the edges of the TPC or with the FFD.
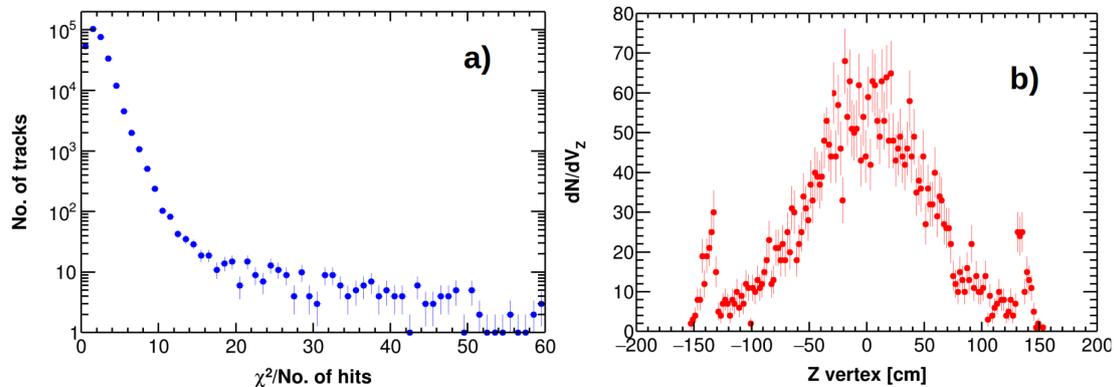

Figure 2.2: Distributions for Bi+Bi Collisions at 9.2 GeV (request 31). a)χ^2/Nofh for all the AMC tracks. b) Z vertex for 4000 events. Distribution for χ^2/Nofh presents no difference for other requests. As for the Z vertex, request 25 presents the same shape, whereas request 30 presents smearing up to ± 50 cm.

Chapter 3

Results and discussion

A compilation of the most important results are presented in this chapter, the tables presented can serve as a guide of the effect you would expect when doing a determinate selection cut. A more detailed and systematic study can be done in order to obtain the most optimal value for a given variable depending on the type of study required. All results are presented with two basic kinematic selection criteria; $p_T > 1$ GeV/c and $|y| > 0.5$. Additional selection is always indicated.

3.1 Results for the repeated tracks

Given the definitions on the previous chapter, a summarize of the most relevant results for the repeated tracks is given in Tables 3.1-3.4.

From Table 3.1 it is shown that the efficiencies and contamination is more or less in the same proportion for each request. The number in parenthesis from the third column indicates the efficiency of the reconstructed tracks compared to the MC tracks, it means that about 80% of the tracks are lost if no selection over primary or secondary tracks is done. Out of the total number of reconstructed tracks, the fourth column indicates that about 3% are repeated tracks. Whereas at least 15% of those repeated tracks are ghost tracks.

Request	ToT MC	ToT AMC	ToT RT	ToT GT
25	1873881	319659 ($\sim 17\%$)	9617 ($\sim 3\%$)	1514 ($\sim 15\%$)
30	2255585	508396 ($\sim 23\%$)	16596 ($\sim 3\%$)	2494 ($\sim 15\%$)
31	1984770	365130 ($\sim 18\%$)	11735 ($\sim 3\%$)	1775 ($\sim 15\%$)

Table 3.1: Table shows the number of repeated tracks in a sample without any selection of either primary or secondary tracks. Columns represent, from left to right: request No., Total number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from the Repeated tracks, and for the Ghost tracks. The number in parenthesis represents the percentage of the sample compared to the one on its left side.

From Table 3.2, where a selection of primary tracks is done with the motherID of
the tracks (meaning that we can be sure that all tracks are in fact primary tracks), we
can observe that the reconstruction efficiency is good, because only about 6% of the
primary tracks have been lost in the process. The total number of repeated tracks stays
about the same as the previous case, and the ghost track's lower bound is only reduced
to ~13%.

Request	ToT MC	ToT AMC	ToT RT	ToT GT
25	166730	156659 ($\sim 94\%$)	4459 ($\sim 3\%$)	604 ($\sim 13\%$)
30	420065	396702 ($\sim 94\%$)	11678 ($\sim 3\%$)	1489 ($\sim 13\%$)
31	307355	289530 ($\sim 94\%$)	8487 ($\sim 3\%$)	1063 ($\sim 13\%$)

Table 3.2: Table shows the number of repeated tracks in a sample of only primary tracks
via the motherID of the track. Columns represent, from left to right: request No., Total
number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from
the Repeated tracks, and for the Ghost tracks. The number in parenthesis represents
the percentage of the sample compared to the one on its left side.

In Table 3.3, a DCA cut is implemented to make the selection of primary tracks.
The results appear rather strange for the case of the request 25, because the efficiency
of reconstructed tracks is about 134%. This means that there should be a source of
contamination, but if we look into the numbers of repeated tracks we can observe that
they represent about 1% of the sample. Then we cannot explain this contamination
in terms of the repeated tracks. Most likely the source of contamination should come
from the secondary tracks, and this situation will be explored in the next section. The
efficiencies and contamination for request 30 and 31 are in a good correspondence with
each other. In fact, even when the DCA cut is reducing the reconstruction efficiency to
about 87%, the contamination due to repeated tracks is also reduced up to 1% and the
lower bound for ghost tracks is reduced even more.

Request	ToT MC	ToT AMC	ToT RT	ToT GT
25	166730	228243 ($\sim 137\%$)	2836 ($\sim 1\%$)	5 ($\sim 0.2\%$)
30	420065	368689 ($\sim 88\%$)	4920 ($\sim 1\%$)	12 ($\sim 0.2\%$)
31	307355	264723 ($\sim 86\%$)	3605 ($\sim 1\%$)	9 ($\sim 0.2\%$)

Table 3.3: Table shows the number of repeated tracks in a sample of only primary
tracks via a DCA< 1 cm. Columns represent, from left to right: request No., Total
number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from
the Repeated tracks, and for the Ghost tracks. The number in parenthesis represents
the percentage of the sample compared to the one on its left side.

In table 3.4 is shown that a cut over the number of hits (additional to the DCA
cut) is in fact reducing the number repeated tracks and ghost tracks, but the DCA cut

by itself makes a better job in reducing this number for some reason. Probably the explanation is because the repeated tracks usually have a reduced number of hits, so when making the extrapolation of those tracks to find its DCA, this number turns out to be big because is far apart from the primary vertex, and so a cut over the DCA is taking most of the RT off.

Request	ToT MC	ToT AMC	ToT RT	ToT GT
25	166730	224092 ($\sim 134\%$)	2078 ($\sim 1\%$)	1 ($\sim 0\%$)
30	420065	361956 ($\sim 86\%$)	3576 ($\sim 1\%$)	1 ($\sim 0\%$)
31	307355	259754 ($\sim 85\%$)	2683 ($\sim 1\%$)	2 ($\sim 0.1\%$)

Table 3.4: Table shows the number of repeated tracks in a sample of only primary tracks via a DCA< 1 cm $+$ a cut over the number of hits > 30. Columns represent, from left to right: request No., Total number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from the Repeated tracks, and for the Ghost tracks. The number in parenthesis represents the percentage of the sample compared to the one on its left side.

3.2 Results for the primary and secondary tracks

Tables 3.5-3.7 summarize the results for the primary and secondary tracks, and together with the previous tables give a better understanding of the whole analysis.

From Table 3.5 it is observed that without making any selection of the primary or secondary tracks, about half of the tracks come from primary tracks and half from secondary tracks in request 25. A considerable difference is observed for the request 30 and 31, which presents similar results to each other. About 80% of the tracks consist of primary tracks and 20% of secondary tracks.

Request	ToT MC	ToT AMC	ToT PrimT	ToT SecT
25	1873881	319659 ($\sim 17\%$)	156659 ($\sim 49\%$)	163000 ($\sim 51\%$)
30	2255585	508396 ($\sim 23\%$)	396702 ($\sim 78\%$)	111694 ($\sim 22\%$)
31	1984770	365130 ($\sim 18\%$)	289530 ($\sim 79\%$)	75600 ($\sim 21\%$)

Table 3.5: Table shows the number of primary and secondary tracks with only basic cuts. Columns represent, from left to right: request No., Total number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from the Primary tracks, and from Secondary tracks. The number in parenthesis for columns 4-5 represents the percentage of the sample compared to the total number of AMC tracks.

In Table 3.6 a cut over the DCA< 1 cm is applied to the tracks. The result for the request 25 shows that the DCA cut is not reducing considerably the contamination

due to secondary tracks, they represent up to ~40% of the tracks. This means that the 137% efficiency in track reconstruction of primary tracks has quite a big error. As for the request 30 and 31 corresponds, they are in a good agreement with each other. Primary tracks represents about 90% of the reconstructed tracks, while there is a contamination of 10% of secondary tracks.

Request	ToT MC	ToT AMC	ToT PrimT	ToT SecT
25	166730	228243 ($\sim 137\%$)	133681 ($\sim 59\%$)	94562 ($\sim 41\%$)
30	420065	368689 ($\sim 88\%$)	336032 ($\sim 91\%$)	32657 ($\sim 9\%$)
31	307355	264723 ($\sim 86\%$)	246643 ($\sim 93\%$)	18080 ($\sim 7\%$)

Table 3.6: Table shows the number of primary and secondary tracks with a DCA< 1 cm. Columns represent, from left to right: request No., Total number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from the Primary tracks, and from Secondary tracks. The number in parenthesis for columns 4-5 represents the percentage of the sample compared to the total number of AMC tracks.

Finally, Table 3.7 shows even a more restrictive cut over the DCA (< 0.5 cm). The effect is a reduction overall in the reconstruction efficiency for all request about 20-30%. In the case of request 25, the percentage of primary and secondary tracks stays about the same. But for the case of the request 30 and 31, a little improvement is observed, less contamination from secondary particles on the condition of loosing about 20% of the tracks.

Request	ToT MC	ToT AMC	ToT PrimT	ToT SecT
25	166730	180337 ($\sim 108\%$)	107108 ($\sim 59\%$)	73229 ($\sim 41\%$)
30	420065	287815 ($\sim 68\%$)	269884 ($\sim 94\%$)	17931 ($\sim 6\%$)
31	307355	207325 ($\sim 67\%$)	197168 ($\sim 95\%$)	10157 ($\sim 5\%$)

Table 3.7: Table shows the number of primary and secondary tracks with a DCA< 0.5 cm. Columns represent, from left to right: request No., Total number of tracks (ToT) from the Monte Carlo, from the Associated Monte Carlo, from the Primary tracks, and from Secondary tracks. The number in parenthesis for columns 4-5 represents the percentage of the sample compared to the total number of AMC tracks.

3.3 Efficiency and contamination per particle species

Some results for the 6 particle of interest in this work ($\pi^{\pm}, K^{\pm}, p^{\pm}$) is shown in this section. The effects under consideration for this analysis are those studied in the 2 previous sections.

In the Table 3.8 is shown how many tracks corresponding to a particular particle

species are there in the total reconstructed tracks which are primary tracks (motherID criteria). As it can be observed, the overall efficiency stays about the same for each request, being the highest for $\pi^\pm, p^\pm$ ($\sim 95\%$) and the lowest for $K^\pm$ ($\sim 80\%$). The relative abundances of each particle species is rather similar for each request, having only a subtle difference in the case of p^+. Most of the primary particles produced in the collision are π^+ ($\sim 70 - 80\%$), the rest $\sim 15 - 25\%$ are $p^\pm$ and the least contribution comes from the $K^\pm$ ($\sim 7\%$). The relative abundances remain more or less similar even when applying several different cuts.

Part.	ToT rq25	ToT rq30	ToT rq31	% in ToT AMC
π^+	51938 ($\sim 95\%$)	143870 ($\sim 96\%$)	107635 ($\sim 95\%$)	33/36/37
K^+	6658 ($\sim 82\%$)	25115 ($\sim 82\%$)	13257 ($\sim 82\%$)	4/6/5
p^+	37061 ($\sim 95\%$)	96 ($\sim 96\%$)	45139 ($\sim 96\%$)	24/14/16
π^-	56493 ($\sim 95\%$)	167336 ($\sim 96\%$)	115793 ($\sim 95\%$)	36/40/40
K^-	3964 ($\sim 81\%$)	13458 ($\sim 82\%$)	7146 ($\sim 80\%$)	3/3/2
p^-	500 ($\sim 96\%$)	1042 ($\sim 95\%$)	560 ($\sim 94\%$)	0.3/0.3/0.2

Table 3.8: Table shows information per particle species only for primary tracks. Columns represent, from left to right: particle type, total number of reconstructed tracks from request 25, total number of reconstructed tracks from request 30, total number of reconstructed tracks from request 31, and the percentage of tracks of that particle specie for request 25 / request 30 / request 31. The number in parenthesis for columns 1-3 represents reconstruction efficiency.

Table 3.9 is showing how are the effects of the two previous sections propagated per particle species, with the basic cuts + a DCA>1 cm cut. As it can be observed from the column corresponding to the request 25, most of the contamination comes from secondary particles, either from $\pi^\pm$ or $K^\pm$, whereas the contamination from $p^\pm$ even if considerable, is not the main source. For the case of the requests 30 and 31 the results presents similarities overall. The contamination coming from secondary particles is quite reduced for the case of $K^\pm$, whereas the contamination from $\pi^\pm$ is greatly reduced compared to the request 25. The main source of contamination from secondary tracks is due to the $p^\pm$ and is in correspondence to that of the request 25. The contamination due to the repeated tracks is very low and equal for all requests, at most about 1% per particle specie.

3.4 Results for cuts on the Z vertex and χ^2/Nofh

To complete the analysis, data for a cut over the Z vertex is shown in Table 3.10, whereas data for a cut in the χ^2/Nofh is shown in Table 3.11. Cuts for the Z vertex are selected according to the figures 3.1-3.3. It is interesting to observe the difference between the request 25 and 31, even if the two come from UrQMD simulations, the secondary pions are behaving differently. Cuts where based on the π^+ distribution for two main reasons.

Part.	ToT rq25	ToT rq30	ToT rq31	~RP(%)	~SecT(%)
π^+	84735 ($\sim 155\%$)	131027 ($\sim 87\%$)	96884 ($\sim 86\%$)	1/1/1	48/7/6
K^+	9733 ($\sim 120\%$)	21004 ($\sim 69\%$)	11102 ($\sim 69\%$)	1/1/1	41/0.1/0.1
p^+	37339 ($\sim 95\%$)	58117 ($\sim 101\%$)	44766 ($\sim 95\%$)	1/1/1	14/19/13
π^-	90595 ($\sim 152\%$)	147729 ($\sim 88\%$)	105407 ($\sim 87\%$)	1/1/1	47/8/6
K^-	5317 ($\sim 109\%$)	9341 ($\sim 69\%$)	5975 ($\sim 67\%$)	1/1/1	37/0.3/0.2
p^-	524 ($\sim 101\%$)	1471 ($\sim 134\%$)	589 ($\sim 99\%$)	1/1/1	17/37/19

Table 3.9: Table shows information per particle species with a DCA< 1 cm. Columns represent, from left to right: particle type, total number of reconstructed tracks from request 25, total number of reconstructed tracks from request 30, total number of reconstructed tracks from request 31, contamination due to repeated tracks for request 25 / request 30 / request 31, and contamination due to secondary tracks for request 25 / request 30 / request 31. The number in parenthesis for columns 1-3 represents reconstruction efficiency.

First, they represent most of the contamination coming from secondary tracks in the sample for request 25, and secondly, they represent most of the tracks.

As seen in figure 3.1, both the primary and secondary pions are denser in the low DCA region, and only some extra secondary pions are appearing when collisions at a position closer to the edge of the TPC is occurring. This explains why a DCA cut is not working for the selection of primary particles, but it does not explain why this is happening. A possible explanation will be given on the final section of this chapter.
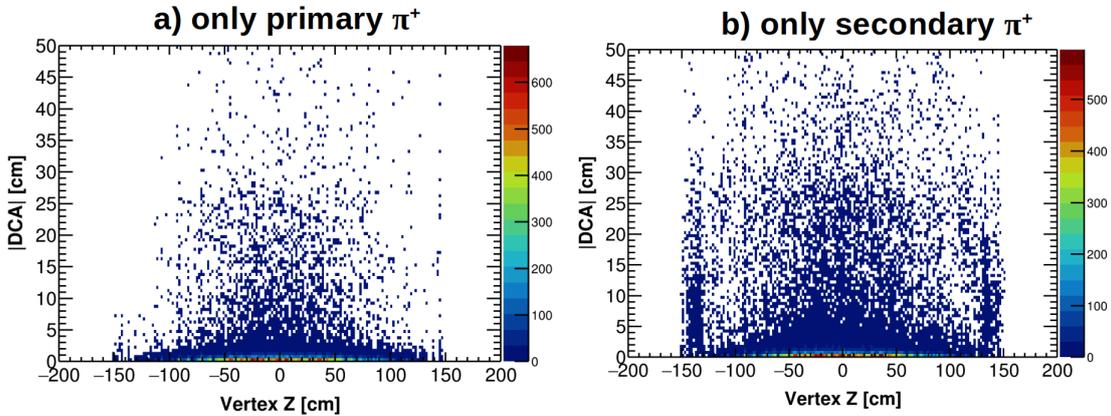

Figure 3.1: Z vertex vs DCA of π^+ from the request 25 with basic cuts.

In figure 3.2 appears a different situation, the smearing of the Z vertex is not as big as for the other requests. The distribution for primary and secondary particles looks similar in the Z vertex direction, but different in the DCA. Basically, there is a considerable amount of π^+ having a high (> 1) DCA, which in fact are taken out by the DCA cut

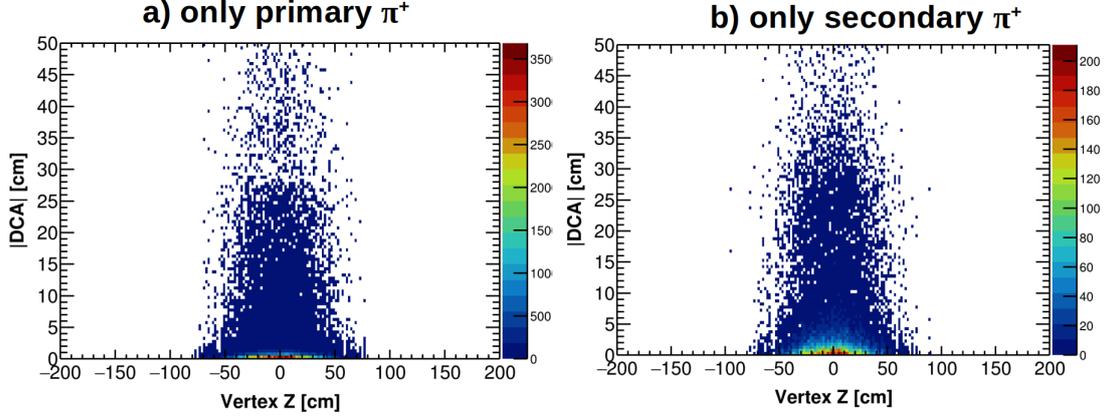and then reducing the contamination from secondary tracks for this request.

Figure 3.2: Z vertex vs DCA of π^+ from the request 30 with basic cuts.

The situation shown in figure 3.3 looks rather similar to that of the request 25 if we only look for the Z vertex direction. But if we only look at the DCA direction, the situation looks more similar to the request 30. Then a DCA cut will be a good condition to select primary tracks, which is one of the biggest difference compared to the reconstructed data from the request 25.
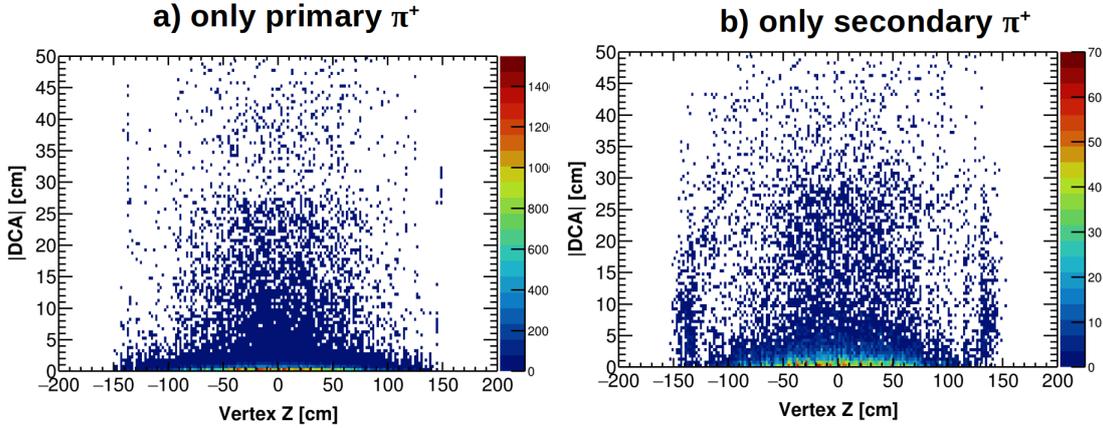

Figure 3.3: Z vertex vs DCA of π^+ from the request 31 with basic cuts.

In the Table 3.10 we can observe that the results are comparable to those of the Table 3.6. Differences do exist but they are very small and mainly given by the amount of events considered for the analysis. If we were to observe a considerable difference, a more restrictive cut over the Z vertex should be applied, but as we reduce the cut, more events will be taken off, so a very restrictive cut should only be applied when a big sample is analyzed.

Table 3.11 shows the efficiencies and contamination with a cut over the χ^2/Number

Request	$V_z <$ [cm]	Rec eff. (%)	RT cont. (%)	SecT cont. (%)
25	50	137	1	41
30	30	87	1	9
31	50	85	1	7

Table 3.10: Table shows different efficiencies for a selection of basic cuts + Z vertex cut + DCA< 1 cm. Columns represent, from left to right: request No., the value of the Z vertex cut, reconstruction efficiency, contamination from repeated tracks and contamination from secondary tracks.

of hits < 4. This Table is also in a good agreement with the previous one. Then a more restrictive cut should be applied if one is to observe some differences. But one should be careful to not take out too many tracks in this case and thus reduce the reconstruction efficiency greatly.

Request	Rec eff. (%)	RT cont. (%)	SecT cont. (%)
25	136	1	41
30	87	1	9
31	85	1	7

Table 3.11: Table shows different efficiencies for a selection of basic cuts + χ^2/Nofh< 4 + DCA< 1 cm. Columns represent, from left to right: request No., reconstruction efficiency, contamination from repeated tracks and contamination from secondary tracks.

3.5   On the secondary particles

One last detail which could explain the situation occurring with the impossibility to properly select primary tracks from the request 25 with a DCA cut is explored briefly in this section.

In Table 3.12 is shown the total amount of secondary particles produced in the analyzed data. Also is shown the number of secondary tracks which are produced at the same time at which the collision occurs. This is because for the reconstruction process, the collision occurs at times = 0. This means that most likely, the secondary particles produced in the request 25 come from heavy resonances which are decaying intermediately (for any practical purposes). This makes sense when one is to look for the inputfile provided to UrQMD, because usually the model force a decaying of all unstable particles when the simulation ends [6]. Now for the case of the request 25, as indicated in MPDforum, some resonances where considered as stable throughout the whole simulation process, by providing a restriction to the inputfile. Finally, when the simulation through the detectors is done by the GEANT software, those resonances are

again "activated", and so they decay quickly. All those particles are then considered by the reconstruction process as secondary tracks and then a cut over the DCA will not be able to distinguish properly with the definition of primary/secondary tracks considered by the reconstruction algorithm. Further analysis in this direction is needed, but this could also explain the similarities between the results obtained for the protons case in the three requests.

Request	No. SecT	SecT T_0
25	163000 ($\sim$51%)	93969 ($\sim$58%)
30	111694 ($\sim$22%)	8289 ($\sim$7%)
31	75600 ($\sim$21%)	5110 ($\sim$7%)

Table 3.12: Table shows the amount of secondary tracks with basic selection cuts. The last column indicates the number of secondary tracks which are produced at time = 0, meaning that they are produced when the initial collision occurs. Numbers in parenthesis indicates the percentage of secondary tracks with respect of the total number of tracks (left), and the percentage of secondary tracks produced at T=0 with respect to the total number of secondary tracks (right).

Chapter 4

Conclusions

The current state of the analysis shows the overall behavior of the reconstructed tracks under several conditions. The reconstruction efficiency has been quantified for the ideal case in which the detectors are as accurate as they can be. The effect of the repeated tracks represents about 1% of the reconstructed tracks when only an ordinary kinematic selection + DCA cuts are applied. So even if low, some considerations should be done when calculating the systematic errors in the processing of the real data.

On the other hand, if the analysis in consideration requires a clear distinction between primary and secondary tracks via the DCA, some special attention should be put in order to avoid large contamination. The analysis has shown that for the case of the requests 30 and 31, the DCA cut introduces a low contamination when selecting only primary tracks. As for the request 25, the DCA approach cannot really distinguish properly between primary and secondary tracks.

This same analysis can be applied to any MC produced data, and further analysis to calculate detectors' efficiency with a proper PID wagon would make a more complete analysis of the MPD's characteristics.

Bibliography

[1] B. I. Abelev *et al.* [STAR Collaboration], Systematic measurements of identified particle spectra in pp, d+Au, and Au+Au collisions at the STAR detector. *Phys. Rev. C* **79** (2009) 034909.

[2] L. Adamczyk *et al.* [STAR Collaboration], Bulk Properties of the Medium Produced in Relativistic Heavy-Ion Collisions from the Beam Energy Scan Program. *Phys. Rev. C* **96** (2017) 044904.

[3] MPD technical design report. http://mpd.jinr.ru/doc/mpd-tdr/

[4] V. Abgaryan *et al.* [MPD Collaboration], Status and initial physics performance studies of the MPD experiment at NICA. *Eur. Phys. J. A* **58** (2022) 140.

[5] MPDroot. http://mpdroot.jinr.ru/

[6] UrQMD user manual. https://itp.uni-frankfurt.de/~bleicher/userguide.pdf