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

# FINAL REPORT ON THE START PROGRAMME

## *Introduction to programming language and data analysis*

**Supervisor:**
MSc. Luong Ba Vinh
**Student:**
Nguyen Viet Quy, Viet Nam
University of Science VNUHCM
**Participation period:**
February 16 – March 29
Winter Session 2025

Dubna, 2025

## Abstract

During my START programme, study of the basics of C++ programming language, ROOT framework, and data analysis were carried out. Using C++ and ROOT, I have studied to add event, track, and particle identification (PID) selections to the experimental data of heavy-ion collisions collected at the STAR experiment. Furthermore, quality assurrance (QA) histograms were obtained to aid the data analysis procedure.

# CONTENTS

# CHAPTER 1. OVERVIEW OF PROGRAMMING LANGUAGE

## 1.1. C++ programming language

C++ is a powerful, high-performance programming language that offers several advantages for various types of software development. Here are some key benefits of using C++:

The first part is advanced programming with C++, including C++ programming with pointers and arrays,structured data types and file input and output operations.

The second part is object-oriented programming with C++, including definitions and operations :

C++ is a cross-platform language that can be used to create high-performance applications which was developed by Bjarne Stroustrup, as an extension to the C language. It gives programmers a high level of control over system resources and memory which can be found in today's operating systems, Graphical User Interfaces, and embedded systems. It is an object-oriented programming language which gives a clear structure to programs and allows code to be reused, lowering development costs which is portable and can be used to develop applications that can be adapted to multiple platforms.

We use C++ programming language to add event, track, and particle identification (PID) selection:

- Event selection:
+ Radial collision vertex cut: Vr < 2 cm
+ z-position of collision vertex cut: |Vz| < 70 cm
- Track selection:
+ Number of hits in the Time Projection Chamber (TPC) used for track recon-struction: $N_{hits}^{fit} \geq 15$
+ Number of TPC hits over possible hits : $N_{hits}^{fit}/N_{hits}^{poss} > 0.52$
+ Primary track (track coming from the collision vertex)

+ Distance of Closest Approach (DCA): |DCA| < 3 cm
+ Transverse momentum of the primary track: $0.15 < P_T^{prim} < 1.5$ GeV/c
+ Pseudorapidity of the primary track: $|\eta^{prim}| < 1$

- Kaon identification:

+ For $0.15 < p <$ 1.5 GeV and $|n\sigma_K| < 3$ (TPC) and $0.2 < m^2 < 0.32$ GeV$^2$/c$^4$ (TOF)

## 1.2. ROOT programming language

ROOT is a data analysis framework and programming language primarily used in the field of high-energy physics and astrophysics. Developed by CERN, it provides a powerful set of tools for data manipulation, visualization, and analysis. ROOT combines the capabilities of a programming language (with features similar to C++) with advanced libraries for graphical representation and statistical analysis.

- Data Handling: ROOT can store complex data structures in a compact format (ROOT files), making it efficient for large datasets commonly found in scientific research.
- Histogramming: It has extensive features for creating and manipulating histograms and other types of data visualizations.
- Statistical Analysis: Built-in functions for statistical analysis are critical for interpreting experimental results.
- Object-Oriented: The language is object-oriented, which facilitates code reuse and modular programming.
- Graphs and Plots: Offers powerful graphing capabilities to visualize results effectively.
- Interpreted Language: ROOT scripts can be run in an interpreted manner using its C++-based scripting language
- Extensive Documentation and Community: ROOT is well-documented, and it has an active user community, providing users with support and resources.

  We use ROOT programming language to add quality assurance (QA) histograms:

1. 2D distribution of Vx versus Vy (before and after event selection)
2. Vz (before and after event selection)
3. $N_{hits}^{fit}$( before and after track selectrion)
4. |DCA| (before and after track selection)
5. $P_T^{prim}$(before and after track selection)
6. $\eta^{prim}$(before and after track selection)
7. 2D $P_T^{prim}$versus $\eta^{prim}$ (before and after track selection
8. dE/dx $P_T^{prim}/q$versus (before and after PID selection), where q is the track's charge
9. nσ(K) versus $P_T^{prim}/q$ (before and after PID selection)
10. $m^2$ versus $P_T^{prim}/q$ (before and after PID selection)

# CHAPTER 2. INTRODUCTION TO DATA ANALYSIS OF HEAVY-ION COLLISIONS

## 2.1. Introduction of Quark-Gluon Plasma

The idea of building a fundamental theory of strong interaction - Quantum Chromodynamics (QCD) came simultaneously with the idea that when the temperature increases above a certain critical T ~ 100 MeV or baryon density above ~ 10 nuclear densities, ordinary hadronic matter should transform into a new phase state – ideal fluid of free quarks and gluons shown in figure 2.1 [1]. Moreover, ideas came from different areas. The stability of neutron stars requires a gentler equation than can be obtained from hadronic matter. One was put forward in the works of Ivanenko and Kurdelaidze, where it was predicted that matter in a similar state should be in massive stars for their stability [1]. Itoch then showed the hydrostatic stability of such stars. Baym and Chin later explored the possibility of stars existing as huge bags of free quarks and gluons. On the other hand, the idea of a gas of free quarks and gluons as a form of matter at very high temperatures is naturally obtained from the asymptotic freedom of quarks in QCD.

The task of studying the phase transition in QCD turned out to be very exciting, and over the next decades, a whole line of research was formed in high-level physics, so today the program of every large accelerator or collider capable of accelerating heavy and light ions includes an item on the study of quark-gluon plasma [1]. Interest is drawn to the fact that this is the only way to study in the laboratory the sector of the Standard Model of QCD phase diagram, phase transitions, and thermalization of fundamental fields [1]. And given the very rich dynamic content of the theory, such studies become extremely exciting. The next phase transition in the Standard Model, associated with the restoration of electroweak symmetry, is expected at temperatures of the order of 100 GeV and its experimental study in the laboratory is hardly possible in the near future [1]. The study of QCD turns out to be closely related to fundamental problems in various areas of modern physics.

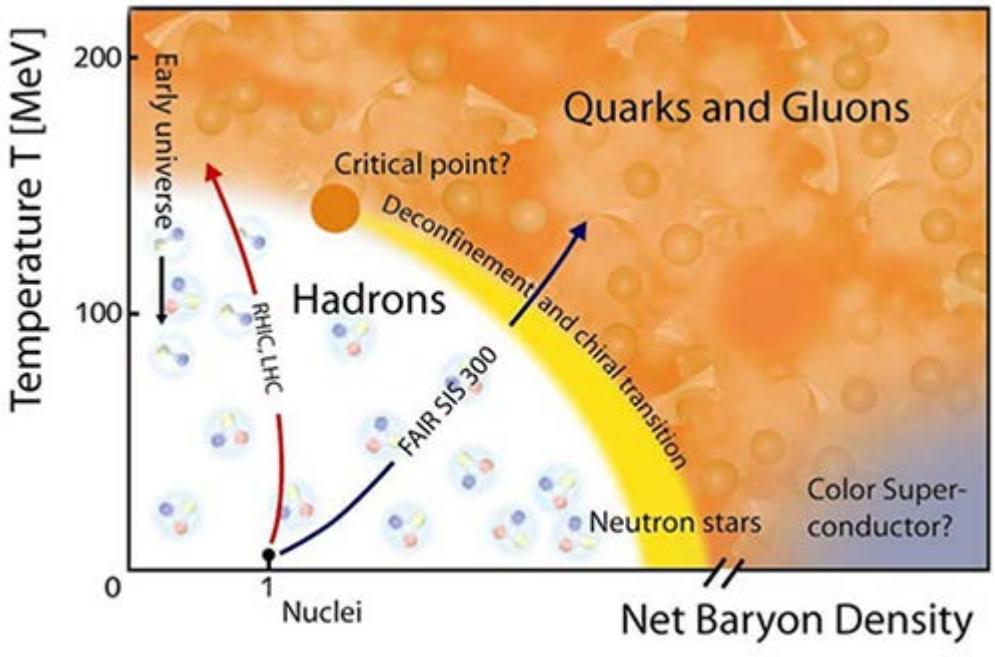


Figure 2.1. Phase diagram of hadronic matter with transition lines from hadronic gas to quark-gluon plasma and the critical point [2].

## 2.2. Experimental Setup

The Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Laboratory is the first machine in the world, designed to accelerate and collide the heavy ion beams at relativistic speed to explore the matter at early universe. RHIC is a versatile accelerator, capable of operating with both polarized protons and a variety of ion species over a broad range of energies [3].

The polarized proton-proton collisions at RHIC is a part of spin physics program with the aim of studying the spin structure of the nucleon. The RHIC construction was completed at BNL in 1999 and the first commissioning runs with Au+Au collisions took place in 2000 [3]

The STAR experiment at RHIC is to study the formation and properties of quark gluon plasma which is believed to exist at very high energy density generated by heavy ion collisions. Because of the complexity of the system produced in collisions the STAR detector system consists of several types of detectors which are functioning to measure different types of particles. With these detectors working

together experiment data of heavy ion collisions is collected for scientific analysis [4]. A schematic view of the STAR experimental setup is shown in Figure 2.2.

In the following sections, as they are closely related to particle identification, the two detectors called time projection chamber (TPC) and time of flight detector (TOF) will be mainly discussed [4]. The STAR detector consists of several subsystems in a large solenoidal analyzing magnet. With its axis aligned along the beam direction, the time projection chamber (TPC) is the main tracking device for charged particles, covering a pseudorapidity range $|\eta| \leq 1.8$ and providing complete azimuthal coverage [5]. The entire TPC is located inside a solenoidal magnet, and data are taken at the maximum magnetic field $|Bz| = 0.5$, where the z axis is parallel to the beam direction . Radial-drift TPCs (FTPCs) are also installed to extend particle tracking into the forward and backward regions ($2.5 < |\eta| < 4.0$) [5]. Surrounding the TPC is the central trigger barrel (CTB) , which is a scintillator counter array whose analog signal is sensitive to the total charged particle multiplicity with coverage $|\eta| \leq 1.0$. A pair of beam-beam counters (BBCs) at $3.3 < \eta < 5.0$ and a pair of zero degree calorimeters (ZDCs) at $\theta < 2$ mrad are located on either side of the collision region along the beam line, and are used to provide event triggers for data taking [5].

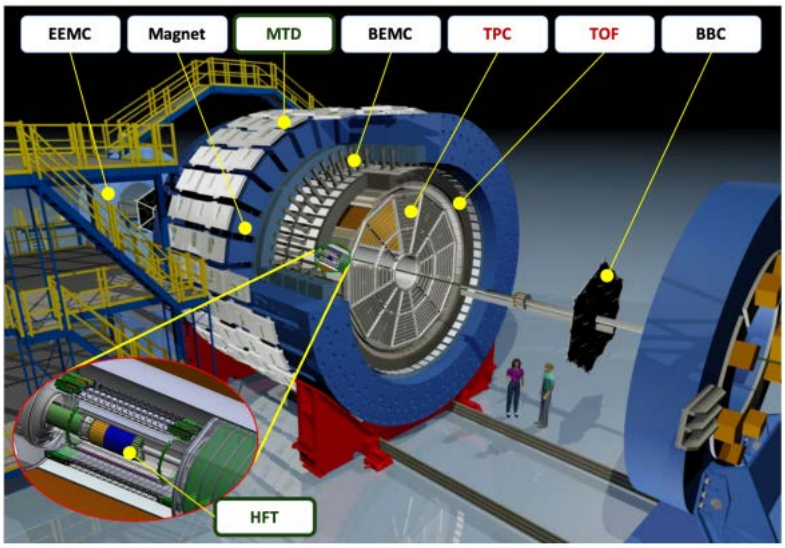


Figure 2.2. An overview of the Solenoidal Tracker At RHIC (STAR) detector. Major subsystems are Magnet, TPC, TOF, BEMC/EEMC, MTD, HFT, HLT. STAR features over eighteen detector subsystems in total [4].

## 2.3. Data set for Au+Au collisions at 27 GeV

The Beam Energy Scan (BES) program at RHIC aims to explore the QCD phase diagram, including the search for the evidence of the 1st order phase transition from hadronic matter to Quark-Gluon Plasma (QGP) and the location of the QCD critical point. One of the features previously observed in the study of QGP is the effect of suppression of particle production with high transverse momenta (>2 GeV/c) at energies $\sqrt{s_{NN}}$ = 27 GeV which was deduced from the charged-particle nuclear modification factor measured using the data from Beam Energy Scan Program Phase I (BES-I) of STAR experiment. In 2018, STAR has collected over 500 million events from Au+Au collisions at $\sqrt{s_{NN}}$ = 27 GeV as a part of the STAR BES-II program, which is about a factor of 10 higher than BES-I $\sqrt{s_{NN}}$ = 27 GeV data size.

## 2.4. Selection criteria

### 2.4.1. Event selection

The longitudinal z position of the interaction point is determined on-line by the measured time difference of the two ZDCs's signals. A cut on the z position of the interaction point is applied on-line for all data sets in order to maximize the amount of useful data for physics analysis, since events with primary vertices far away from the center of the TPC have a significant nonuniform acceptance. In the off-line data analysis further cuts are applied on the z position of the reconstructed primary vertex ($V_Z$), to ensure nearly uniform detector acceptance [5].

Radial collision vertex cut: Vr < 2 cm

z-position of collision vertex cut: |Vz| < 70 cm

### 2.4.2. Track selection

Several quality cuts were applied to ensure selection of good tracks. During the TPC track reconstruction, a charged track was extrapolated back to the beam line by using the reconstructed helix parameters. If the distance of closest approach (DCA) of the track to the event vertex was less than 3 cm and the track had at least ten hit points in the TPC, the reconstructed track was labeled as a primary track. The helix parameters for primary tracks were then refitted by requiring that the helix pass through the primary vertex location. This procedure improved the momentum

resolution of tracks [5]. The DCA selection for primary tracks was 3 cm. The differences in the results were small and were included in the estimate of systematic uncertainties [5]. Tracks with transverse momentum less than 0.15 GeV/c were not used, as their combined acceptance and efficiency becomes very small. Each track was required to have at least 15 hit points out of 45 used in the fitting of the tracks helix parameters. The ratio of the number of space points used in the track reconstruction to the maximum possible number of hit points was required to be greater than 55% to avoid split tracks where a real track is reconstructed in two or more segments. A pseudo-rapidity cut |η| < 1.0 was applied to select tracks that are well within the TPC acceptance [5].

1. Number of hits in the Time Projection Chamber (TPC) used for track recon - struction: $N_{hits}^{fit} \geq 15$
2. Number of TPC hits over possible hits : $N_{hits}^{fit}/N_{hits}^{poss} > 0.52$
3. Primary track (track coming from the collision vertex)
4. Distance of Closest Approach (DCA): |DCA| < 3 cm
5. Transverse momentum of the primary track: $0.15 < P_T^{prim} < 1.5$ GeV/c
6. Pseudorapidity of the primary track: $|\eta^{prim}| < 1$

### 2.4.3. Kaon selection

Particle identification (PID) was achieved by correlating the ionization energy loss (dE/dx) of charged particles in the TPC gas with their measured momentum.The measurement of mean dE/dx was achieved by averaging the measured dE/dx samples along the track after truncating the top 30%. The measured dE/dx versus momentum curve is reasonably well described by the Bethe-Bloch function smeared with the detector's resolution [5]. Note that from the dE/dx measurement, kaons cannot be clearly separated from pions above p ~ 0.6 GeV/c and from protons/antiprotons above p ~ 1.1 GeV/c [5].

The selection criteria to identify kaons are: $0.15 < p < 1.5$ GeV and $|n\sigma_K| < 3$ (TPC) and $0.2 < m^2 < 0.32$ GeV$^2$/c$^4$ (TOF).

# CHAPTER 3.RESULTS OF INTRODUCTION TO DATA ANALYSIS

## 3.1. Event selection

### 3.1.1. 2D distribution of Vx versus Vy (before and after event selection)

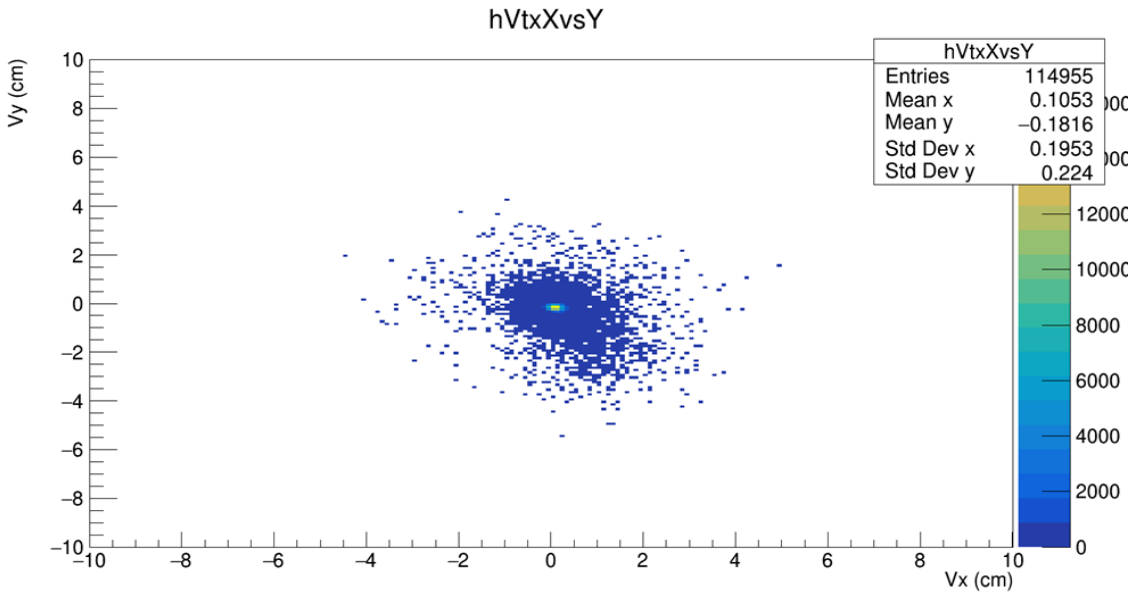


Figure 3.1. 2D distribution of Vx versus Vy (before event selection)

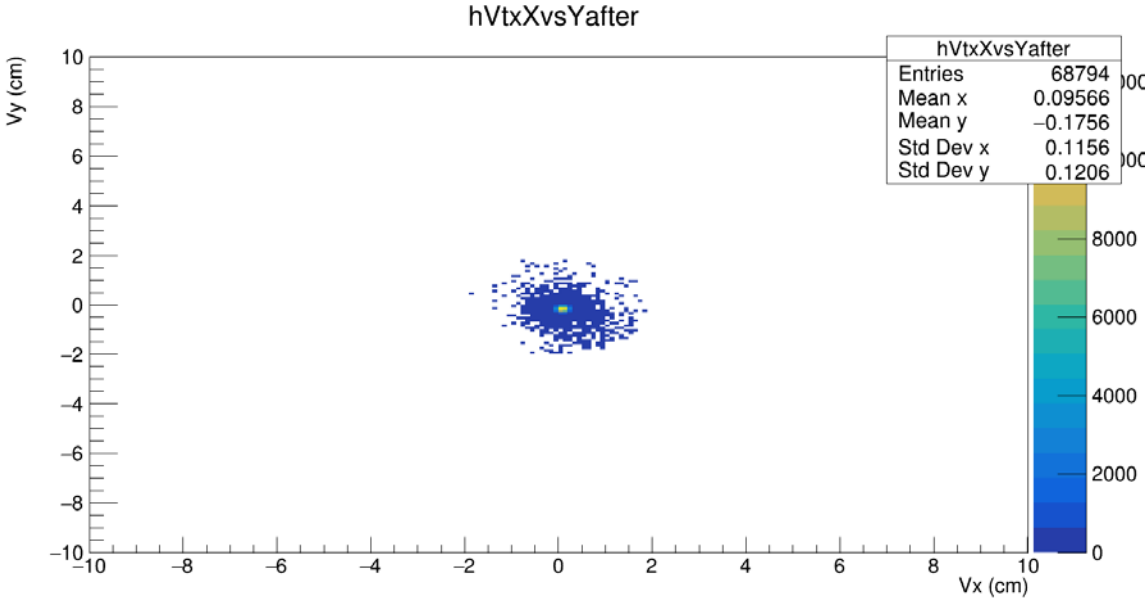


Figure 3.2. 2D distribution of Vx versus Vy (after event selection)

In Figure 3.1 and 3.2 shows a 2D histogram representing the distribution of primary vertices in the XY-plane before and after add event selection: Radial collision

vertex cut: Vr < 2 cm, specifically in $Vx$ and $Vy$ centimeters. The color gradient indicates the density of events, with yellow representing areas of higher density and blue representing areas of lower density. The concentration of points around the center indicates the typical region where primary vertices are reconstructed in this experiment. Purpose of the radial collision vertex cut limits the radial distance of the primary vertex position to ensure event quality by excluding collisions that occur near the beam pipe, thereby improving data reliability and helps to minimize background contributions from beam-gas or beam-beam pipe collisions, which can distort the experimental results. During low-energy runs at RHIC, where maintaining a clean event sample is crucial for accurate physics analyses, the STAR experiment can more effectively select primary particles, enhancing the overall event selection efficiency and the detected tracks are more likely linked to the collision event rather than spurious backgrounds, providing cleaner data for subsequent analysis.

### 3.1.2. Vz (before and after event selection)

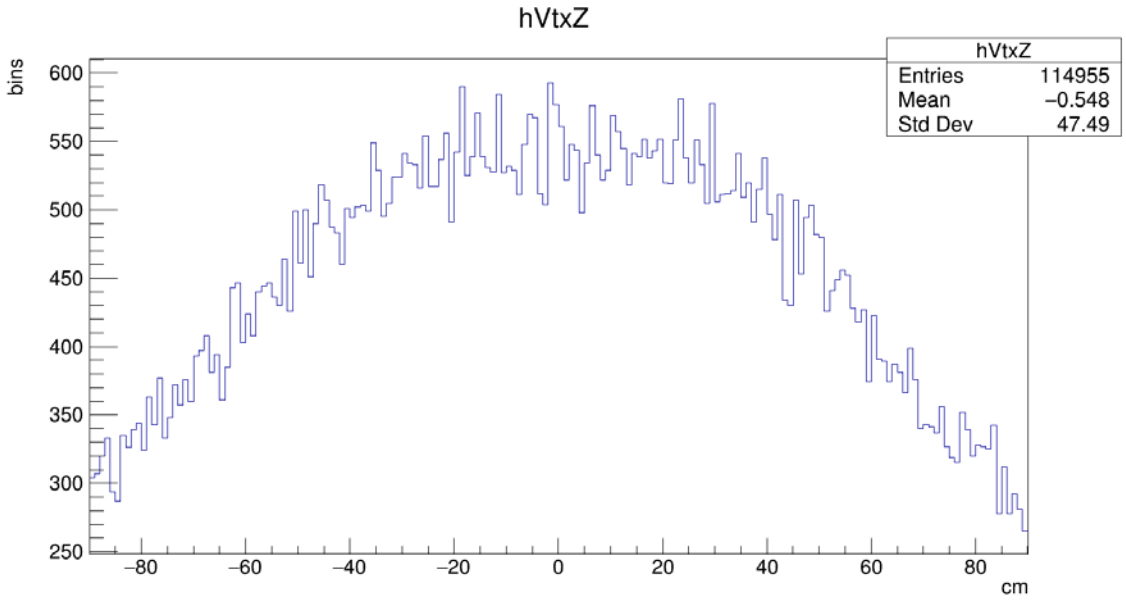


Figure 3.3. Vz before event selection

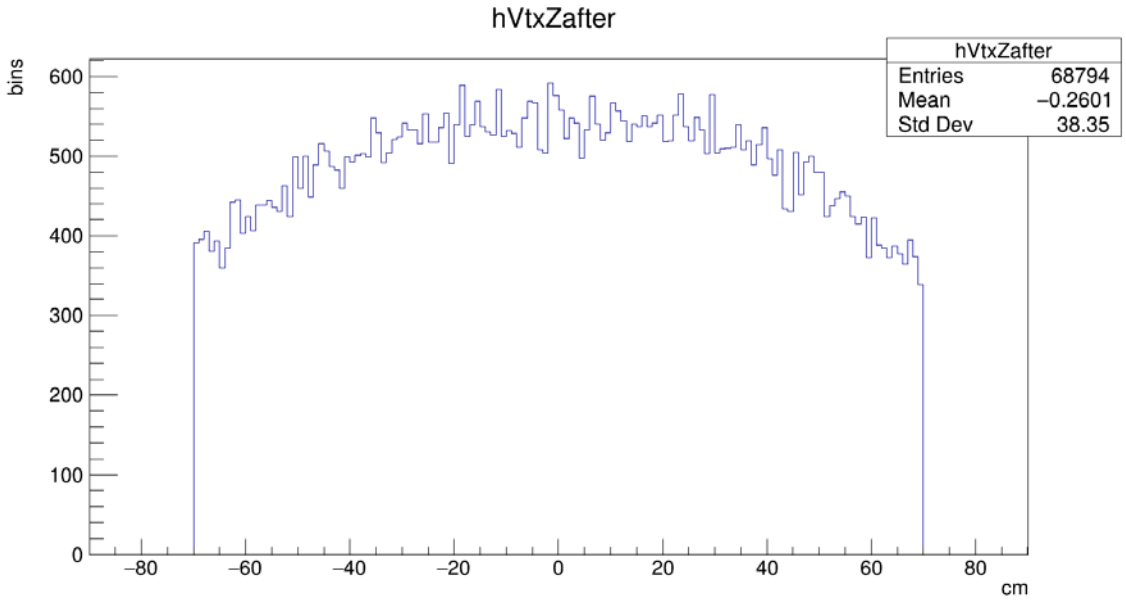


Figure 3.4, Vz after event selection

In Figure 3.3 and 3.4 shows the position of the primary vertex z before and after selection: z-position of collision vertex cut: |Vz| < 70 cm. In the Cartesian coordinate system, there are x, y, and z coordinates of the vertex. The z-coordinate is a longitudinal component that is responsible for the position of the primary vertex along. Purpose of the z-position of collision vertex cut ensures that the primary vertex of the collision is within 70 cm of the center of the Time Projection Chamber (TPC), enhancing the quality and reliability of the reconstructed events by limiting the z-position, the experiment reduces the likelihood of including events that originate from beam-gas interactions, thereby improving data purity. Also helps maintain uniform detector acceptance across varying collision scenarios, which is crucial for accurate physics measurements. It is designed to optimize the detection of particles by ensuring they fall within the effective operational range of the detector systems

## 3.2. Track selection

### 3.2.1. $N_{hits}^{fit}$ (before and after event selection)

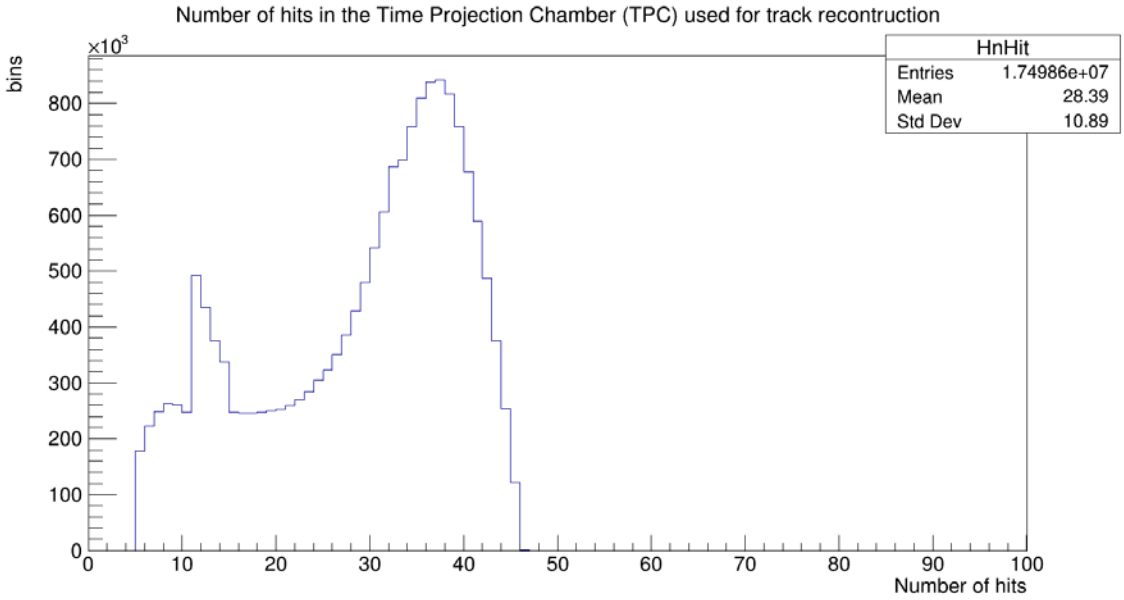


Figure 3.5. $N_{hits}^{fit}$ before selection

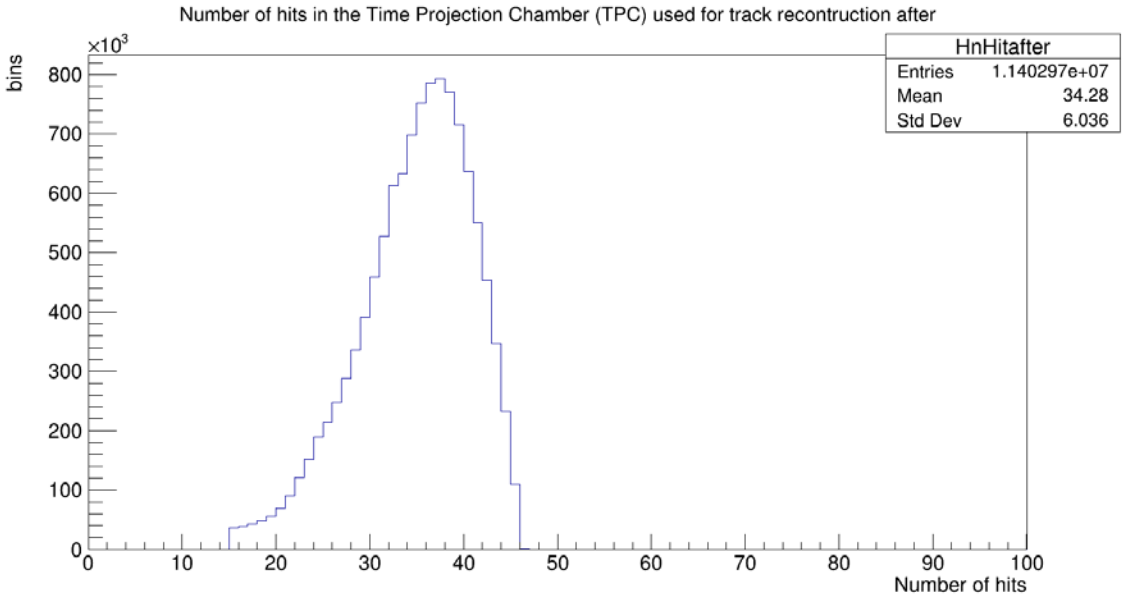


Figure 3.6. $N_{hits}^{fit}$ after selection

In Figure 3.5 and 3.6 display histograms representing the distribution of the number of hits produced by a particle's track before and after add track selection: Number of hits in the Time Projection Chamber (TPC) used for track reconstruction: $N_{hits}^{fit} \geq 15$. The x-axis indicates the number of hits, while the y-axis shows the count

of occurrences. Purpose of the number of hits in the Time Projection Chamber (TPC) used for track reconstruction with $N_{hits}^{fit} \geq 15$ is requiring a minimum of 15 hits ensures that the track reconstruction is based on sufficient data points, leading to a more accurate and reliable determination of particle trajectories, helps eliminate poorly reconstructed tracks, which could arise from random noise or low-quality hits, thereby enhancing overall event quality. Number of hits reduces the potential for track splitting, where a single particle may be incorrectly reconstructed as multiple tracks, aiding in accurate particle identification and tracks with at least 15 hit points allow for better momentum resolution and energy-loss measurements (dE/dx), which are critical for particle identification and subsequent analyses

### 3.2.2. |DCA| (before and after track selection)

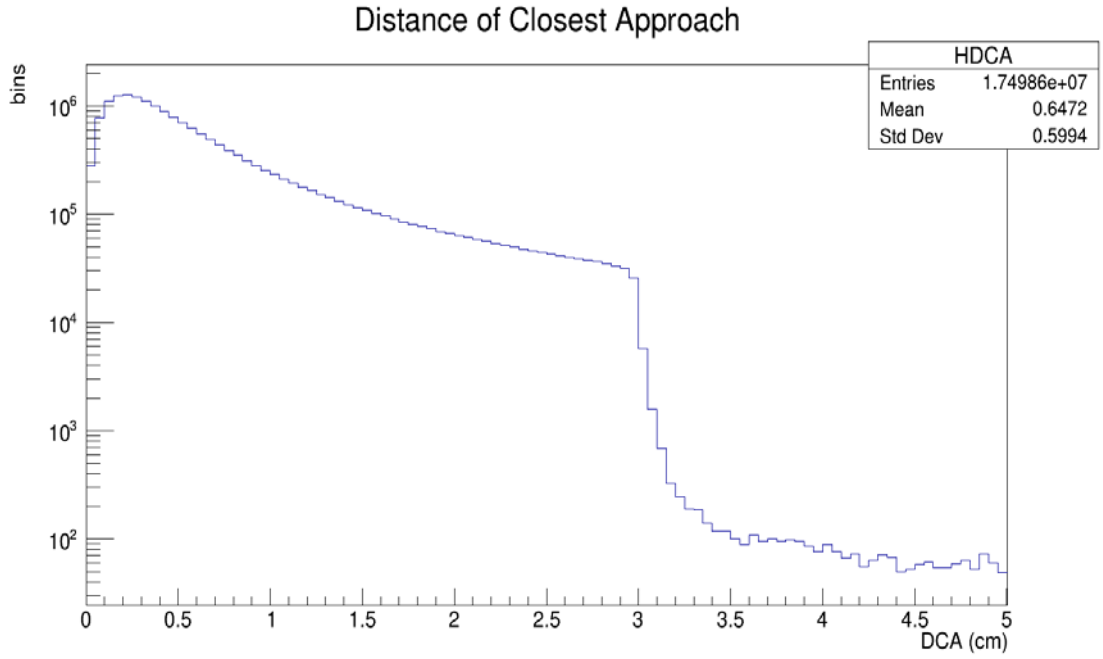


Figure 3.7: |DCA| before selection

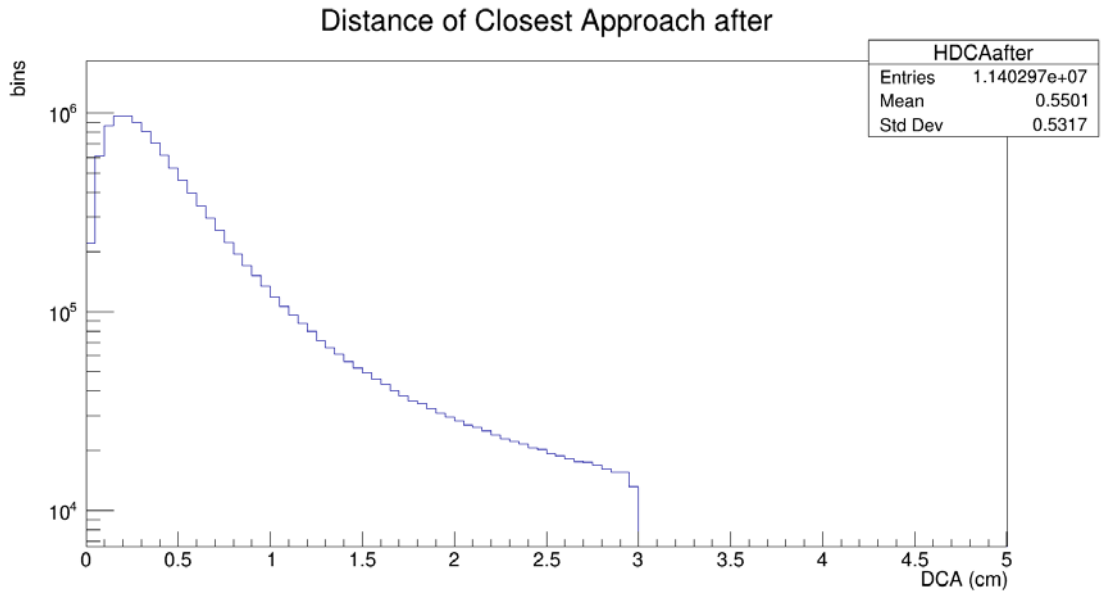


Figure 3.8: |DCA| after selection

In Figure 3.7 and 3.8 DCA (Distance of Closest Approach) is the shortest distance between particle's trajectory and the primary vertex. Distance of Closest Approach (DCA): |DCA| < 3 cm. Purpose of the Distance of Closest Approach (DCA): |DCA| < 3 cm is set to suppress the influence of secondary tracks that may arise from particle decays or interactions within the detector material and the purity of the collected data, as it reduces the background noise in track reconstruction, facilitating more accurate measurements in high-energy collisions.

### 3.2.3. $P_T^{prim}$ (before and after selection)

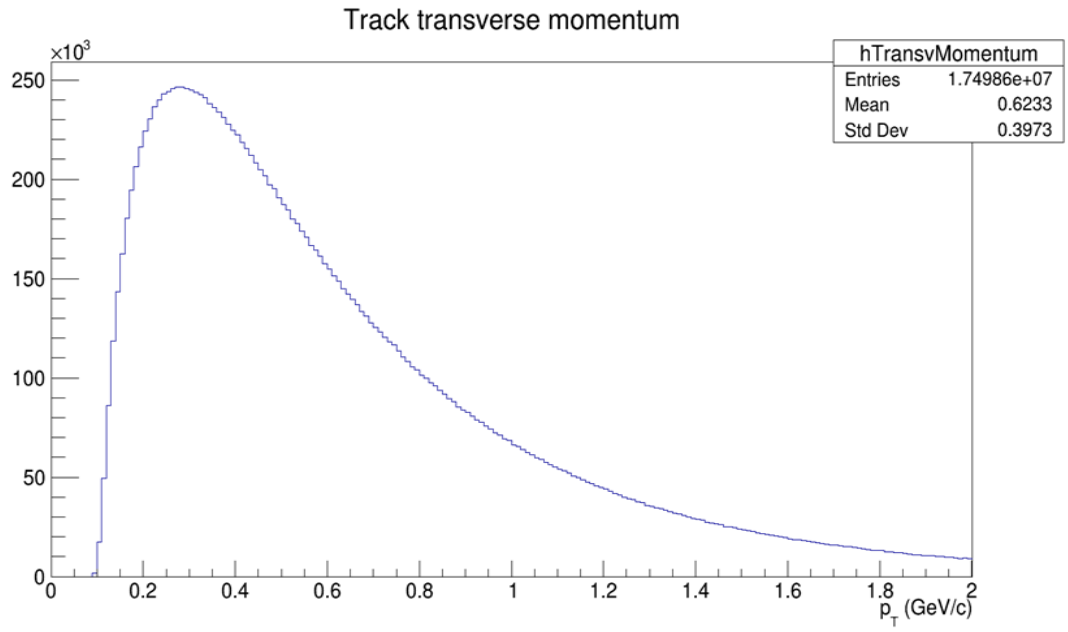


Figure 3.9. $P_T^{prim}$ before selection

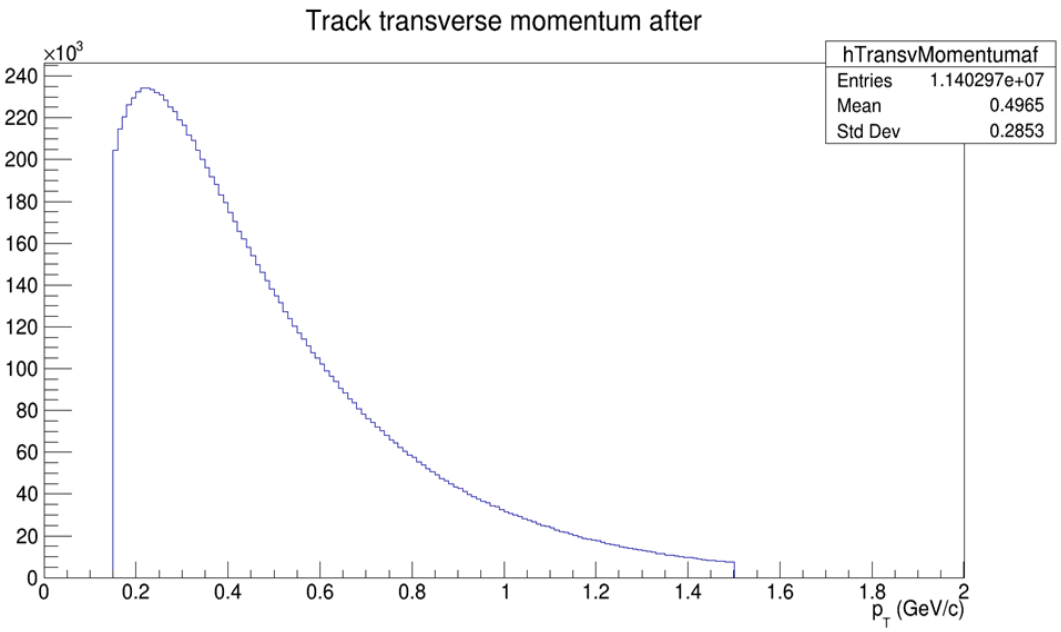


Figure 3.10. $P_T^{prim}$ after selection

In Figure 3.9 and 3.10 shows a one-dimensional distribution of particles over $pt$ before and after track selection: Transverse momentum of the primary track: $0.15 < P_T^{prim} < 1.5$ GeV/c The x-axis indicates the transverse momentum of the primary track , while the y-axis shows the count of occurrences

### 3.2.4. $\eta^{prim}$ (before and after selection)

Pseudorapidity of the primary track: $|\eta^{prim}| < 1$

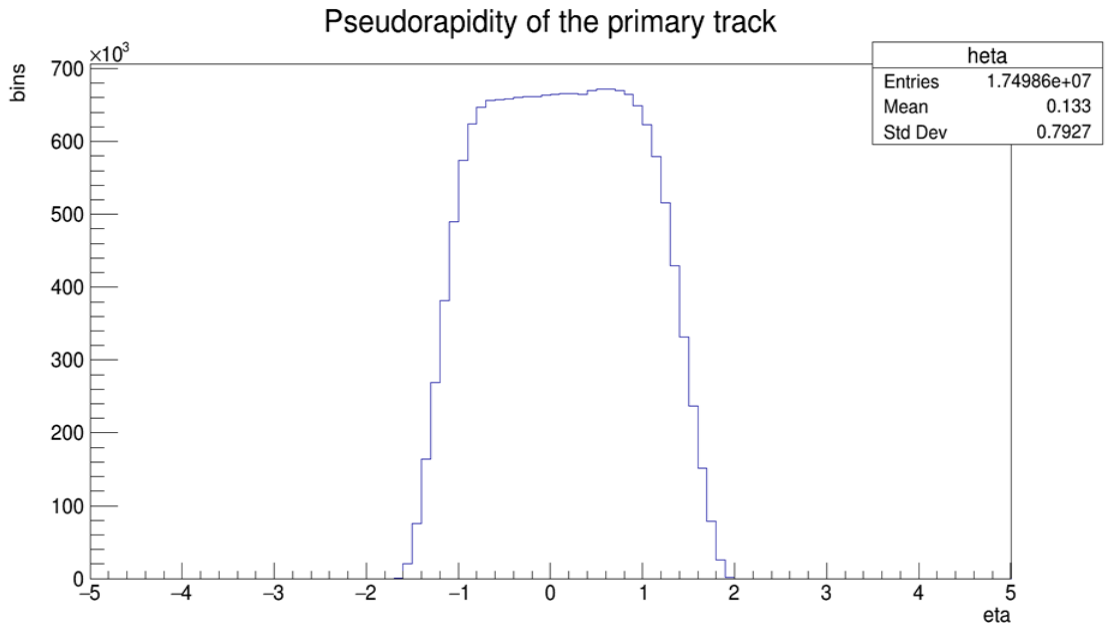


Figure 3.11. Pseudorapidity before selection

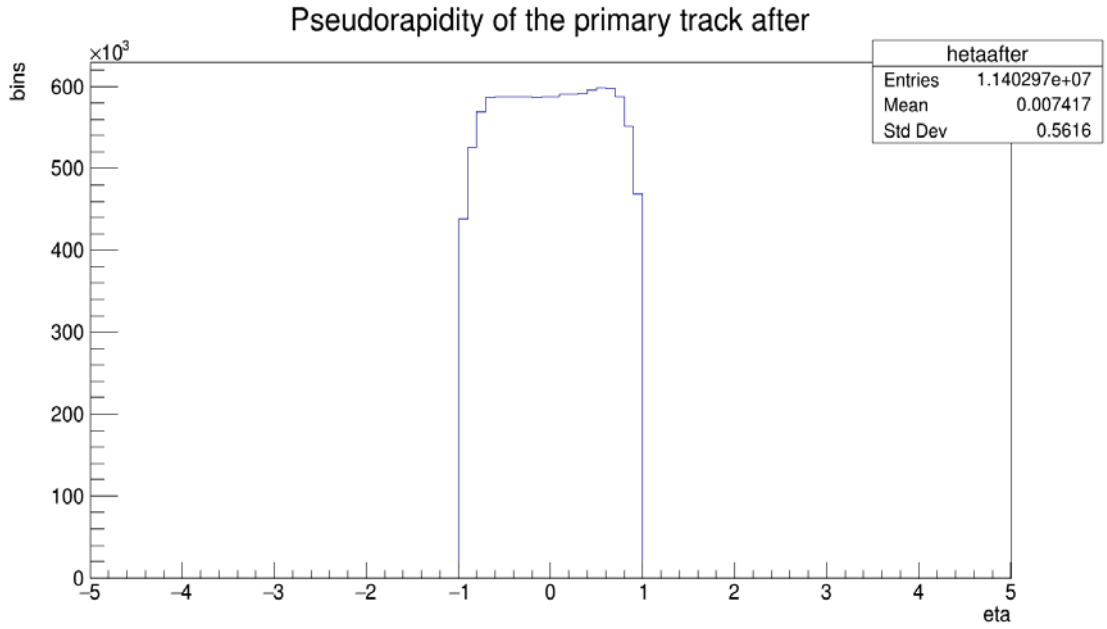


Figure 3.12. Pseudorapidity after selection

In Figure 3.11 and 3.12 shows a one-dimensional distribution of pseudorapidity of the primary track before and after track selection: pseudorapidity of the primary track: $|\eta^{prim}| < 1$. The x-axis indicates pseudorapidity of the primary

track, while the y-axis shows the count of occurrences. Purpose of the pseudorapidity of the primary track: $|\eta^{prim}| < 1$ is ensures that tracks fall within the effective acceptance range of the Time Projection Chamber (TPC) and other detectors, maximizing data collection efficiency by constraining primary tracks to this range, the experiment seeks to achieve uniform detector performance across different pseudorapidity regions, allowing for better statistical analysis and comparisons. It helps minimize biases that could arise from effects related to tracks originating outside the central rapidity region, which can complicate interpretations of the data. This cut facilitates more accurate characterizations of phenomena like particle production and flow patterns, critical for studies of quark-gluon plasma and other nuclear matter properties

### 3.2.5. 2D $P_T^{prim}$ versus $\eta^{prim}$ (before and after track selection)

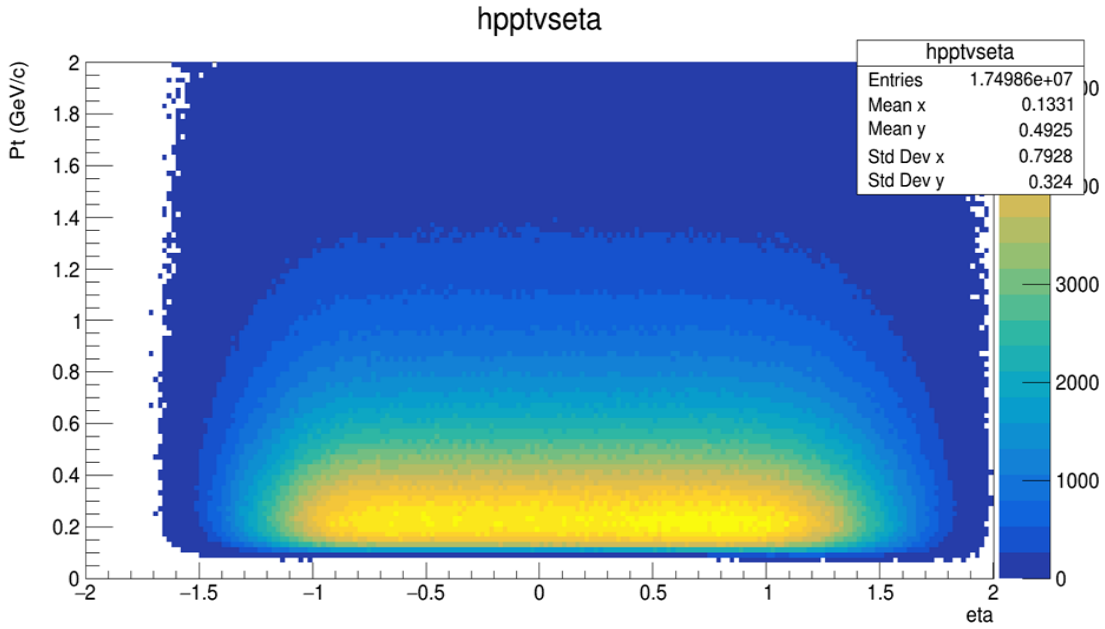


Figure 3.13. $P_T^{prim}$ versus $\eta^{prim}$ before track selection

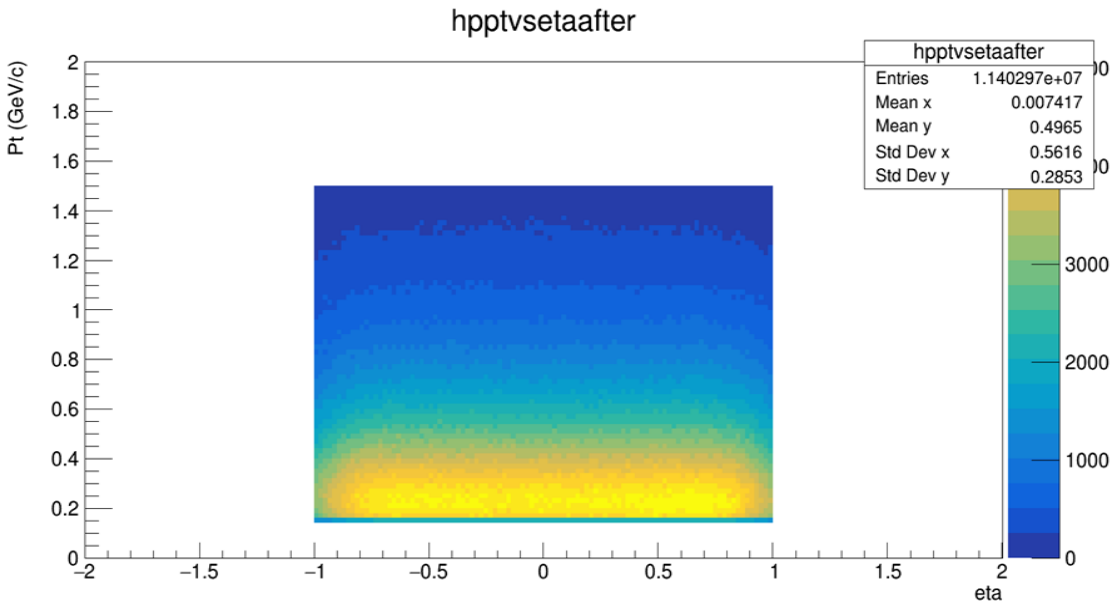


Figure 3.13. $P_T^{prim}$ versus $\eta^{prim}$ after track selection

In Figure 3.12 and 3.13 show a two-dimensional distribution of particles over η and $pt$ before and after add track selection : Pseudorapidity of the primary track: $|\eta^{prim}| < 1$ and $0.15 < P_T^{prim} < 1.5$ GeV/c. The x-axis indicates pseudorapidity of the primary track , while the y-axis shows the transverse momentum of the primary track.

### 3.3. Kaon identification

#### 3.3.1. dE/dx versus $P_{tot}^{prim}$ /q (before and after PID selection), where q is the track's charge

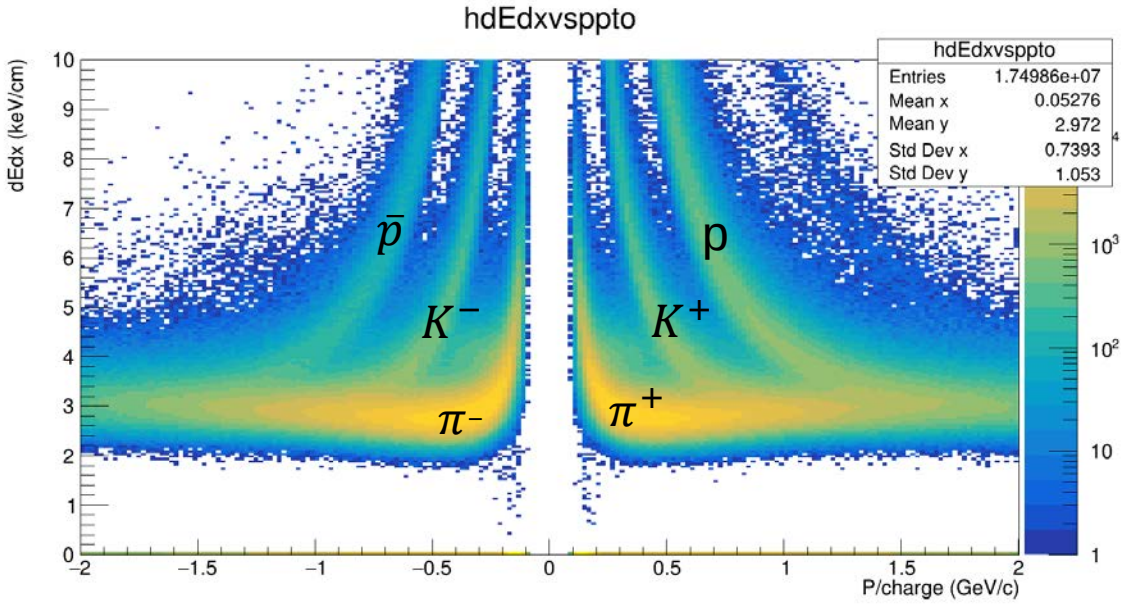


Figure 3.14. dE/dx versus $P_{tot}^{prim}$ /q before PID

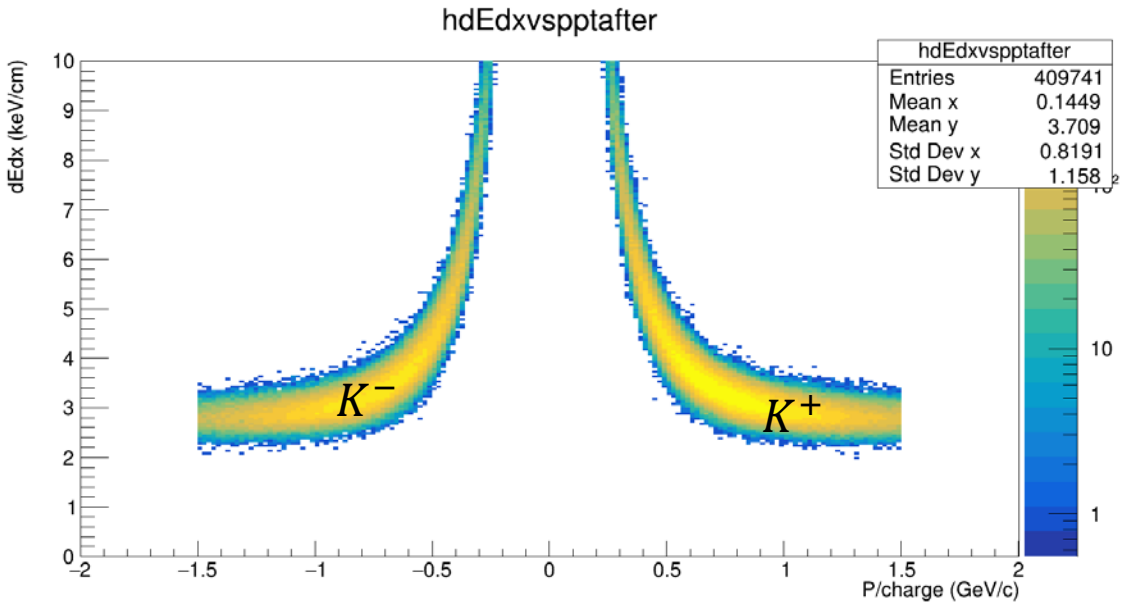


Figure 3.15. dE/dx versus $P_{tot}^{prim}$ /q after PID

In Figure 3.14 and 3.15 shows a distribution of ionization energy loss (dE/dx) versus momentum (p/charge) before and after add kaon identification: For $0.15 < p < 1.5$ GeV and $|n\sigma_K| < 3$ (TPC) and $0.2 < m^2 < 0.32$ GeV$^2$/c$^4$ (TOF). Before we

add kaon identification, different particles (electrons, pions, kaons, protons) are seen. After adding kaon identification, only kaons are visible, which demostrates the correctness of the particle identification procedure.

### 3.3.2. nσ(K) versus $P_{tot}^{prim}$ /q (before and after PID selection)

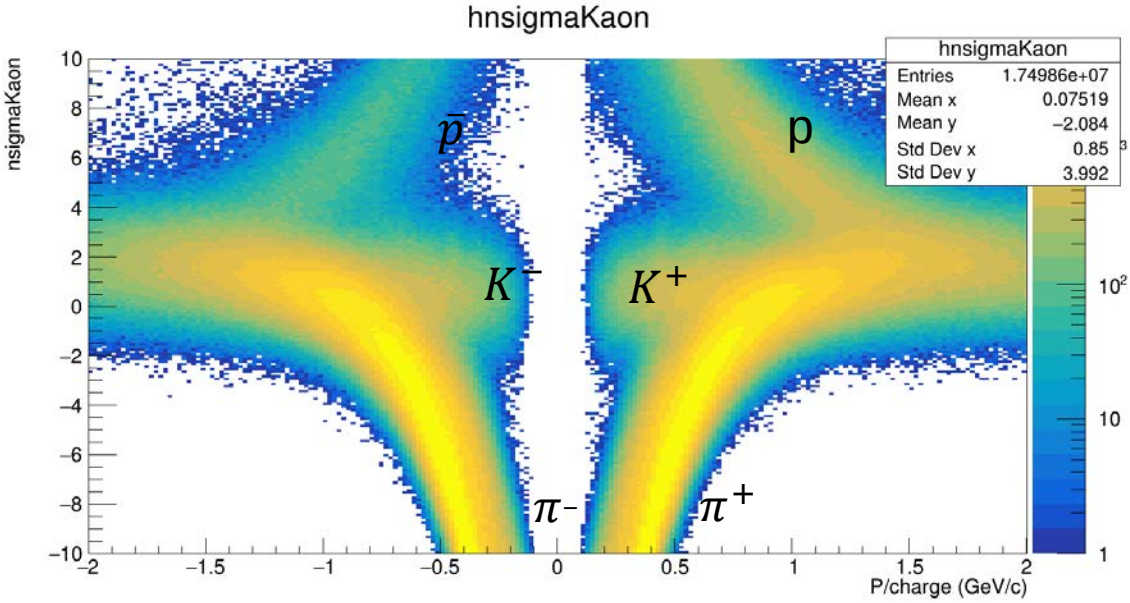


Figure 3.16: nσ(K) versus $P_{tot}^{prim}$ /q before PID selection

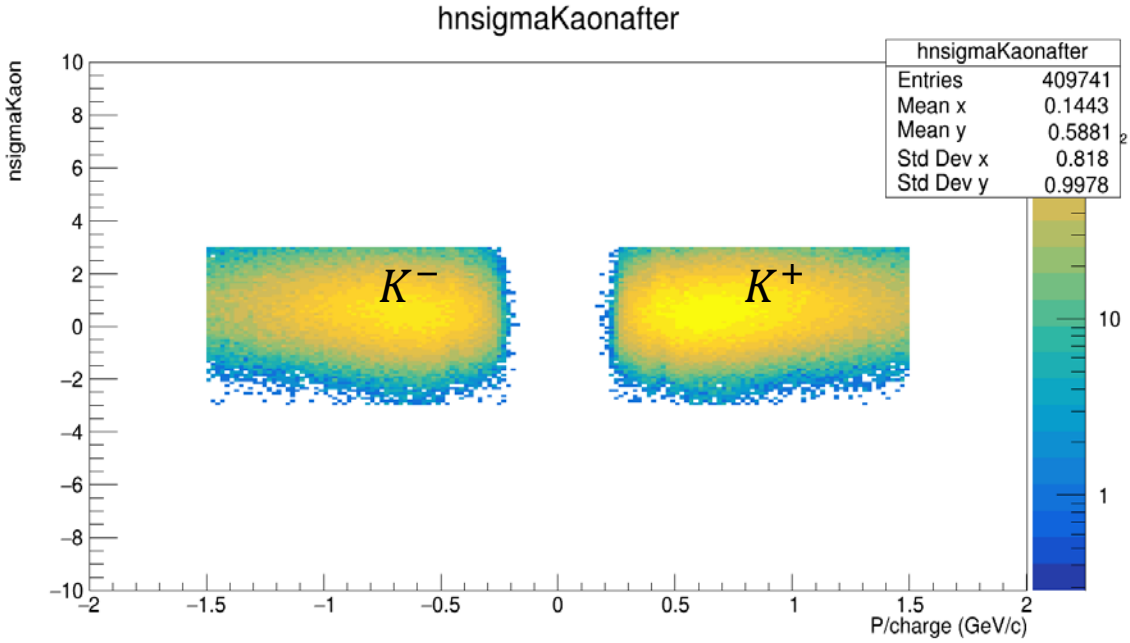


Figure 3.17: nσ(K) versus $P_{tot}^{prim}$ /q after PID selection

In Figure 3.16 and 3.17 shows a distribution of nσ(K) versus momentum (p/charge ) before and after add Kaon identification: For $0.15 < p < 1.5$ GeV and $|n\sigma_K| < 3$ (TPC) and $0.2 < m^2 < 0.32$ GeV$^2$/c$^4$ (TOF). Before we add kaon identification, different particles (electrons, pions, kaons, protons) are seen. After adding kaon identification, only kaons are visible, which demostrates the correctness of the particle identification procedure.

### 3.3.3. $m^2$ versus $P_{tot}^{prim}$ /q (before and after PID selection)

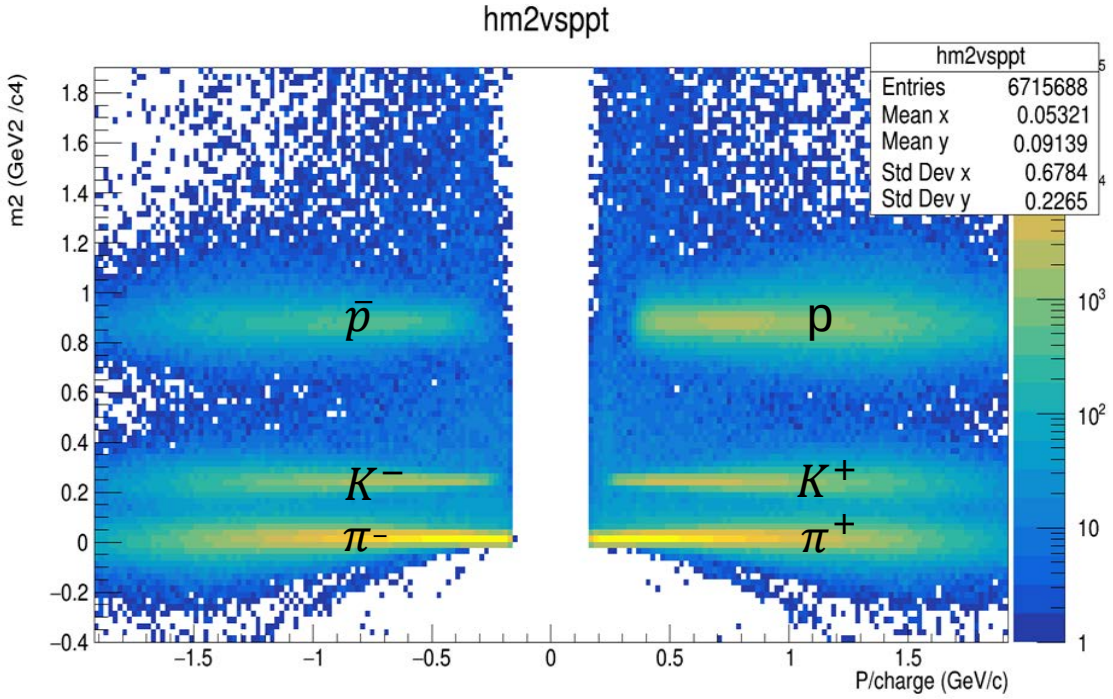


Figure 3.18: $m^2$ versus $P_{tot}^{prim}$ /q before PID selection

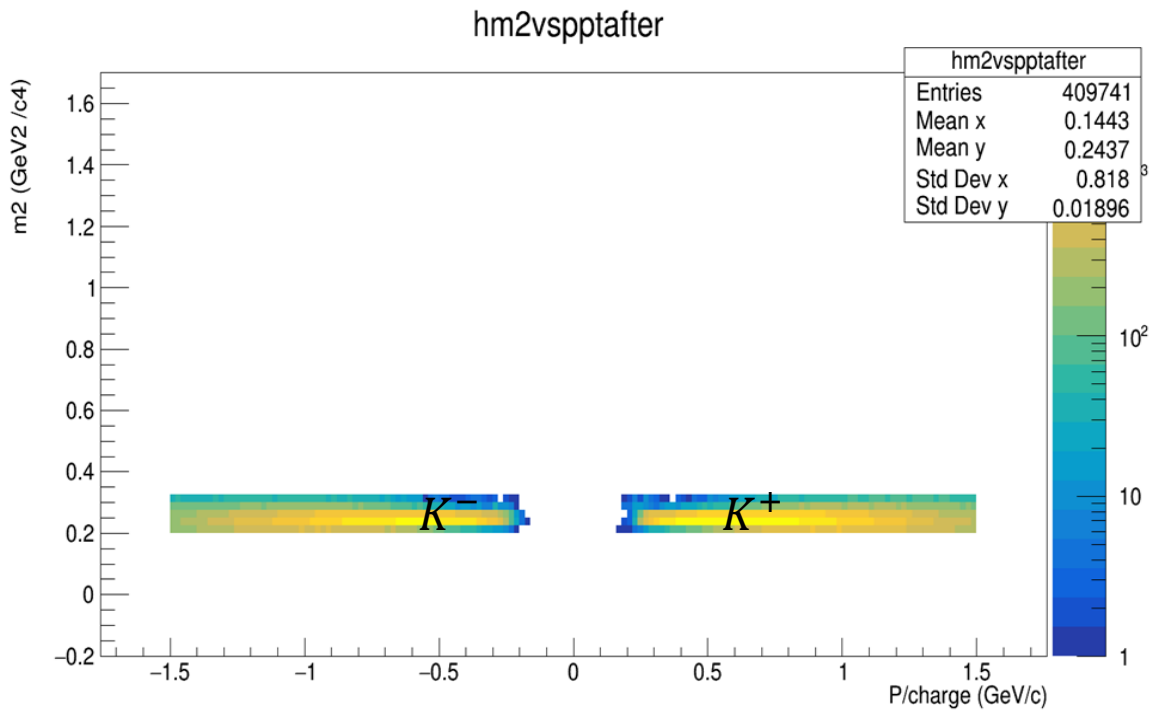


Figure 3.19: $m^2$ versus $P_{tot}^{prim}$ /q after PID selection

In Figure 3.18 and 3.19 give information about distribution of $m^2$ and $p/\text{charge}$ with before and after add Kaon identification: For $0.15 < p < 1.5$ GeV and $|n\sigma_K| < 3$ (TPC) and $0.2 < m^2 < 0.32$ GeV$^2$/c$^4$ (TOF). In this analysis particle tracks are reconstructed using global tracks, which are tracks reconstructed using both TPC and TOF signals. Particles are also identified by using only the TPCs signal. TOF measure the particle velocity β relative to the speed of light in vacuum, $\beta = v/c$ which allowing the mass square $m^2$ of the particle to be determined:

$$m^2 = p^2(1/\beta^2 - 1)$$

Stop time - start time = τ (time of flight) is associated with reconstructed tracks in the TPC by track extrapolation to the TOF. The TPC provides the momentum p, and total length L, so we can calculate inverse velocity [4]. Before we add kaon identification, different particles (electrons, pions, kaons, protons) are seen. After adding kaon identification, only kaons are visible, which demostrates the correctness of the particle identification procedure.

# CONCLUSIONS

I study the basic skills and abilities of working with the CERN ROOT framework and C++ programming language. Using it, we obtained and examined the main characteristics of heavy-ion collisions, such as distributions of the primary vertex position along the longitudinal (beam) direction (Vz), histogram of the position of vertex X and vertex Y, and also the dE/dx and $m^2$ of tracks, necessary for the particle identification.

# ACKNOWLEDGEMENTS

I would like to express my gratitude to JINR for providing the opportunity to visit the START program and paying for all the necessary needs for this period and travel. I am also grateful to my supervisor Luong Ba Vinh for his help and organization of my activities during this period. And, of course, I say a special thank you to my partners from VBLHEP Artem, Sonya and everyone else who helped me master all the tools in solving the problems I needed and helped me compile the programs I needed.

# LIST OF REFERENCES

[1] Maxim Ganin, *Analysis of particle production in heavy ion collision.* JOINT INSTITUTE FOR NUCLEAR RESEARCH Veksler and Baldin laboratory of High Energy Physics, 2024

[2] https://physics.stackexchange.com/questions/662655/how-exactly-does-the-quark-gluon-plasma-prevent-protons-from-forming

[3] CHITRASEN JENA, *PARTICLE PRODUCTION AND ELLIPTIC FLOW OF LIGHT NUCLEI IN RELATIVISTIC HEAVY ION COLLISIONS AT RHIC, 2012*

[4] Yu Zhang, *Measurements of Higher Moments of Proton Multiplicity Distributions in Fixed-Target Au+Au Collisions by the STAR Experiment at RHIC* , 2022

[5] B. I. Abelev and M. M. Aggarwal and Z. Ahammed,..., J. X. Zuo, *Measurements of φ meson production in relativistic heavy-ion collisions at RHIC*, 2009

[6] Arkhipkin and Jerome Lauret, *STAR Online Meta-Data Collection Framework: Integration with the Pre-existing Controls Infrastructure*, 2017. DOI: iopscience.iop.org/article/10.1088/1742-6596/898/3/032023

[7] https://www.star.bnl.gov/public/tof/publications/TOF_20040524.pdf