Master Thesis

Czech Technical University in Prague

F4 Faculty of Nuclear Sciences and Physical Engineering
Department of Physics

Reconstruction of strange hadrons in collisions of nuclei at RHIC

Bc. Jakub Kubát

ZADÁNÍ DIPLOMOVÉ PRÁCE

Katedra: fyziky   *Akademický rok:* 2019/2020

Student: Bc. Jakub Kubát

Studijní program: Aplikace přírodních věd

Obor: Experimentální jaderná a částicová fyzika

Název práce: Rekonstrukce podivných hadronů v jádro-jaderných srážkách na
(česky)   urychlovači RHIC

Název práce: Reconstruction of strange hadrons in collisions of nuclei at RHIC
(anglicky)

Pokyny pro vypracování:

1) Ultra-relativistické srážky jader a kvarkgluonové plasma

2) Experiment STAR na urychlovači RHIC

3) Měření produkce podivných hadronů ve srážkách zlato-zlato na experimentu STAR při nízkých těžišťových energiích.

a. seznámení se s fyzikálním programem nízkých energií na RHIC, tzv. "Beam Energy Scan (BES) II",

b. rekonstrukce a podivných barynů s použitím systému KFParticle Finder z dat nabraných během programu BES II,

c. Optimalizace selekčních kriterií pomocí algoritmů strojového učení,

d. Studium efektivity rekonstrukce.

4) Shrnutí a závěr

Doporučená literatura:

[1] K.Yagi et al.: Quark-Gluon Plasma, Cambridge University Press 2005

[2] I. Kisel: Event Topology Reconstruction in the CBM Experiment, J.Phys.Conf.Ser. 1070 (2018) , 012015

[3] T.O. Ablyazimov, et al.: Kalman filter-based fast track reconstruction for charged particles in a Compressed Baryonic Matter experiment using parallel computing on a multicore server at the Laboratory of Information Technologies, Joint Institute for Nuclear Research, Phys.Part.Nucl.Lett. 12 (2015), 423-427

[4] S. Shi: Strangeness in STAR experiment at RHIC,J.Phys.Conf.Ser. 779 (2017), 012008

Jméno a pracoviště vedoucího diplomové práce:

RNDr. Petr Chaloupka, Ph.D., Katedra fyziky, Fakulta jaderná a fyzikálně inženýrská ČVUT v Praze
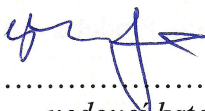
Datum zadání diplomové práce: 25.10.2019

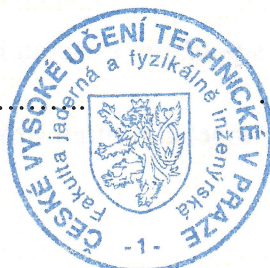*Termín odevzdání diplomové práce:*   04.05.2020

Doba platnosti zadání je dva roky od data zadání.

....
garant oboru

....
vedoucí katedry

....
děkan

V Praze dne   25.10.2019

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Petr Chaloupka for guidance through the analysis presented in this work, for his time, advices, and shared experience. Also, I would like to show gratitude to Doc. Jaroslav Bielčík, Dr. Pavol Federič, Dr. Yuri Fisyak, Ing. Michal Kocan and Dr. Maksym Zyzak for valuable conversations throughout my Master's studies. Last but not least, I am grateful to my friends and family for all their support.

Declaration

I declare that this thesis has been composed solely by myself and that I have cited all sources I have used in the bibliography.

I have no serious reason against the use of this school work in the sense of § 60 of Act No. 121/2000 Sb., On Copyright, on Rights Related to Copyright and on Amendments to Certain Acts (Copyright Act).

Prague, August 12, 2020

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně a použil jsem pouze podklady uvedené v přiloženém seznamu.

Nemám závažný důvod proti použití tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

V Praze, 12. srpna 2020

Abstract

KF Particle Finder (KFPF) is a C++ package developed for fast and effective reconstruction of short-lived particles. It is expected to help with the joint effort of FIAS and STAR's High Level Trigger (HLT) group and Tracking Focus Group (TFG) to develop a fast data processing system which could speed up physics analyses of data collected during second phase of Beam Energy Scan program at STAR and prepare software for future CBM experiment. The KFPF package was used in this thesis for analysis of BES-II data from Au+Au collisions at various energies. Strange baryons were reconstructed using KFPF in express production of BES-II data to demonstrate qualities of both KFPF and STAR's express production.

Keywords: RHIC, STAR, quark-gluon plasma, Beam Energy Scan, KF Particle Finder, High-Level Trigger, particle reconstruction

Supervisor: RNDr. Petr Chaloupka, PhD.
Fakulta jaderná a fyzikálně inženýrská, Břehová 7,
115 19 Praha 1

Abstrakt

KF Particle Finder (KFPF) je C++ balíček vyvinutý pro rychlou a efektivní rekonstrukci částic s krátkou dobou života. Jeho účelem je pomoci při společné snaze skupiny FIAS a skupin High Level Trigger (HLT) a Tracking Focus Group (TFG) na STAR vyvinout systém rychlého zpracování dat, který by mohl urychlit fyzikální analýzy dat shromážděných během druhé fáze programu Beam Energy Scan na STAR a připravit software pro budoucí experiment CBM. Balíček KFPF byl v této práci použit pro analýzu dat z BES-II ze srážek Au+Au při různých energiích. Podivné baryony byly rekonstruovány pomocí KFPF v expresní produkci dat z BES-II pro demonstraci kvalit KFPF a expresní produkce dat na STAR.

Klíčová slova: RHIC, STAR, kvark-gluonové plazma, Beam Energy Scan, KF Particle Finder, High-Level Trigger, rekonstrukce částic

Překlad názvu: Rekonstrukce podivných hadronů v jádro-jaderných srážkách na urychlovači RHIC

Contents

Chapter 1

Introduction

With modern particle accelerators such as Large Hadron Collider (LHC) in CERN, Geneva, or Relativistic Heavy Ion Collider (RHIC) in Brookhaven National Laboratory, New York, physicists are able to collide nuclei of heavier elements such as lead or gold at relativistic energies in order to simulate conditions that were present in the very beginning of the Universe. Under extreme temperatures and energy densities protons and neutrons melt into partons that form them. Exotic state of matter is created which we call quark-gluon plasma (QGP).

The Beam Energy Scan program at STAR is dedicated to the study of nuclear matter phase diagram and phase transition from hadron gas to QGP. In order to map out the nuclear matter phase diagram, physicists need to measure heavy ion collisions at various collision energies. That is why the latest development in collider technology was focused not only on further increasing the collision energy, as it was in the past, but also on improving luminosity at lower energies to compensate for small cross sections, i.e. probabilities, of researched phenomena. Naturally, experimental techniques of detection and data processing had to evolve as well.

The Compressed Baryonic Matter (CBM) experiment at future Facility for Antiproton and Ion Research (FAIR) in GSI, Darmstadt is another planned experiment meant to investigate the QCD phase diagram [1]. The data rate promised to be delivered at FAIR is so high that new solutions for online reconstruction of particles had to be found. This motivated FIAS group to develop the KF Particle Finder package which employs Kalman Filter method for the purpose of reconstruction of short-lived particles [2, 3]. Besides its speed, KF Particle Finder brings also whole new approach to reconstruction of particles. The FIAS group have joined forces with groups behind High Level Trigger and tracking at STAR in order to develop fast data processing system which would speed up the work on data from second phase of Beam Energy program at STAR. This would also help to prepare base for software of future CBM experiment.

The first goal of this thesis is to test properties of KF Particle Finder, additionally in combination with Boosted Decision Trees classifier, in reconstruction of strange particles and look for possible improvements in comparison to conventional analysis approach. The second goal is to explore possibilities of express production of data done by High Level Trigger group and Tracking Focus Group at STAR by employing KF Particle Finder in analysis of expressly produces data from Beam Energy Scan program. The final ambition is to estimate efficiencies of reconstruction with KF Particle Finder using Monte Carlo simulation based on online calibrations done by High Level Trigger and thus investigate qualities of such simulation.

In Chapter 2 brief summary of Quantum Chromodynamics and quark-gluon plasma is given and RHIC Beam Energy Scan is introduced as well. Chapter 3 describes experimental setup of Relativistic Heavy Ion Collider (RHIC) and Solenoidal Tracker at RHIC (STAR) is described. Chapter 4 deals with KF Particle and KF Particle Finder packages. In Chapter 5 we have a look at Toolkit for Multivariate Analysis with focus on Boosted Decision Trees and Chapter 6 present authors analysis of data from collisions of gold nuclei at BES energies with KF Particle Finder and TMVA. In Chapter 7 the work finishes with a summary of what has been presented.

Chapter 2

Quark-Gluon Plasma and QCD Phase Diagram

2.1 Brief Summary of Quantum Chromodynamics

Between 1906 and 1916 Hans Geiger and Ernest Marsden have carried out a series of famous experiments which have led Ernest Rutherford to formulate his model of atom [4]. They have basically bombarded thin gold foil with α particles and contrary to their expectations they have found out that some of the α particles have scattered at large angles. This led Rutherford to the conclusion that atoms are made of very small positively charged nucleus in the center with negatively charged electrons orbiting around.

These observations started a long journey of nuclear and particle physicists to study the structure of matter by colliding particles at increasingly higher energies. Since than we have found out that while electrons are elementary particles, it is not the case with protons and neutrons inside the nucleus. Subnuclear structure of protons was probed in 1968 with deep inelastic scattering experiments at the Stanford Linear Accelerator Complex. Deep inelastic scattering hinted at nucleons being made up from point-like constituents which were called partons at that time but later they were identified with *up* and *down* quarks whose existence was predicted independently by Murray Gell-Mann [5] and George Zweig [6] in 1964. Although every day matter we encounter consist from already mentioned light quarks named *up* and *down*, there are actually three generations of quarks which gives us a total of six quark flavours.

Quarks interact with each other through strong interaction which is mediated by gluons. Similarly to how electrons carry an electric charge, we assign a color charge to every strongly-interacting object. In contrast to electromagnetism, where the mediators (photons) are electrically neutral, gluons are charged with color too. Under standard circumstances quarks and gluons are confined in color neutral particles called hadrons which proton and neutron are the most stable representatives of. Therefore, quarks or gluons are never observed directly. This phenomenon is called color confinement. We

3

now know of two groups of hadrons, baryons consist of three valence quarks with all possible color charges, i.e. red, green and blue, while mesons are formed by pair of quark and anti-quark. Our current understanding of strong interactions is encapsulated in theory of Quantum Chromodynamics (QCD) which together with the theory of Electroweak interactions form a Standard Model of elementary particles.

The Quantum Chromodynamics predicts that the state of deconfined quarks and gluons can actually exist. Under extremely high temperatures and energy densities quarks are expected to become asymptotically free and able to move inside a hot and dense medium quasi-freely on distances larger than the size of hadrons. This medium is called quark-gluon plasma (QGP) and has been intensely studied in heavy ion collisions at Relativistic Heavy Ion Collider [7] and Large Hadron Collider [8]. In order to establish QGP as a phase of a QCD matter it is important to study phase transition between gas of hadrons and the partonic degrees of freedom. Current understanding of QCD matter phase diagram is that there is a smooth crossover between hadron gas and QGP at low baryon densities and high temperature, while at higher densities the transition is of first order which means that there should also be a critical point in the diagram where the transition switches its nature.

2.2 Experimental Evidence for QGP

2.2.1 Jet Quenching

In particle physics jets are collimated flows of hadrons which stem from fragmentation of partons which were scattered in high-energy collisions [9]. As these highly virtual partons fly from the collision point, they produce new increasingly less energetic partons and form a collimated shower. Eventually, partons in the shower reach non-perturbative, low energy scale and produce hadrons which we detect as jets. Conservation of momentum requires compensation of jet's momentum at least by another jet in the opposite direction.

In the presence of quark-gluon plasma, partons are expected to interact with the medium through parton scattering and gluon emission. This leads to energy, momentum loss of partons and consequently to less energetic jets. Hypothetically, if the creation of parton pair happened at the edge of region with strongly interacting medium, one parton would loose substantially more energy by traversing it. This has been experimentally observed by examining azimuthal di-hadron correlations in $p + p$,$d + $ Au, and Au $+$ Au collisions as measured by STAR [10]. In Fig. 2.1 these di-hadron correlations are presented. One can see that in $d + $ Au and $p + p$ collisions jets are compensated by their away side counterparts, however, in Au $+$ Au collisions the away side peaks are suppressed which is argued to be due to the presence of QGP.
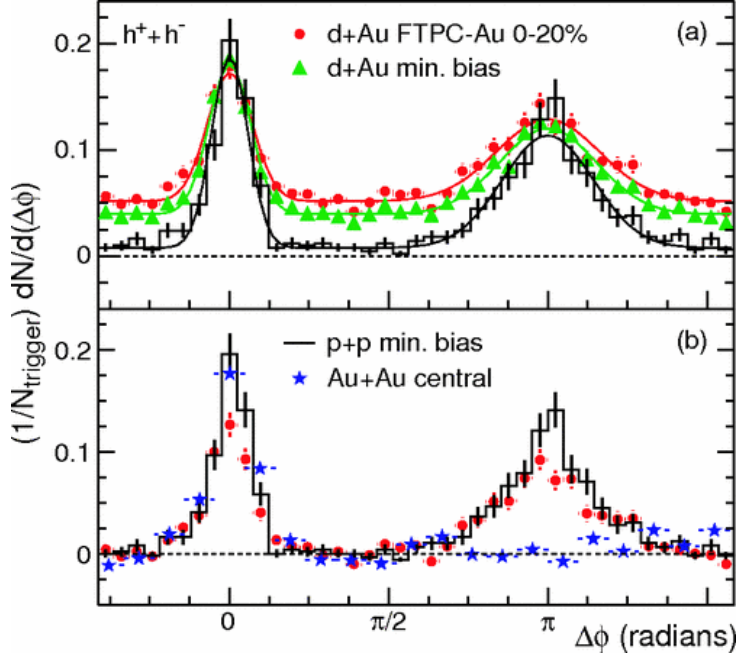
Figure 2.1: Azimuthal di-hadron correlations in $p+p$ (a) and $d+$Au, Au$+$Au (b) collisions as measured by STAR. Taken from [10].

2.2.2 Anisotropic Flow

The overlap region of two colliding nuclei in non-central collision is of an elliptic shape as is illustrated in Fig. 2.2. This spatial anisotropy in xy plane can be described by eccentricity

$$\epsilon = \frac{\langle y^2 - x^2 \rangle}{\langle y^2 + x^2 \rangle}, \tag{2.1}$$

where the average is calculated over some distribution which characterizes participant nucleons, e.g. energy density. If the medium created within the overlap region reaches thermal equilibrium, the initial state geometrical anisotropy converts into pressure gradients which lead to flow. Anisotropies in the collective flow manifest in the final-state particle momentum distributions related to the reaction plane which is defined by beam line z and the impact parameter $\vec{b}$. Typically, we describe anisotropies by the Fourier expansion of momentum distributions

$$E\frac{\mathrm{d}^3 N}{\mathrm{d}^3 \mathbf{p}} = \frac{1}{2\pi} \frac{\mathrm{d}^2 N}{p_{\mathrm{T}} \mathrm{d} p_{\mathrm{T}} \mathrm{d} y} \left(1 + 2 \sum_{n=1}^{\infty} v_n \cos\left[n\left(\varphi - \Psi_{\mathrm{RP}} \right) \right] \right), \tag{2.2}$$

where E is the energy of the particle, p the momentum, p_{T} the transverse momentum, φ the azimuthal angle, y the rapidity, and Ψ_{RP} the reaction plane angle. The second Fourier coefficient v_2 is referred to as elliptic flow and is considered to be a proof for existence of QGP as we would expect particle distributions to be spherical without the creation of deconfined phase [11]. Besides v_2, higher harmonics have been also measured. The v_3 coefficient

is called triangular flow and is supposed to be connected to event-by-event fluctuations in the participant-nucleon collision points [12].
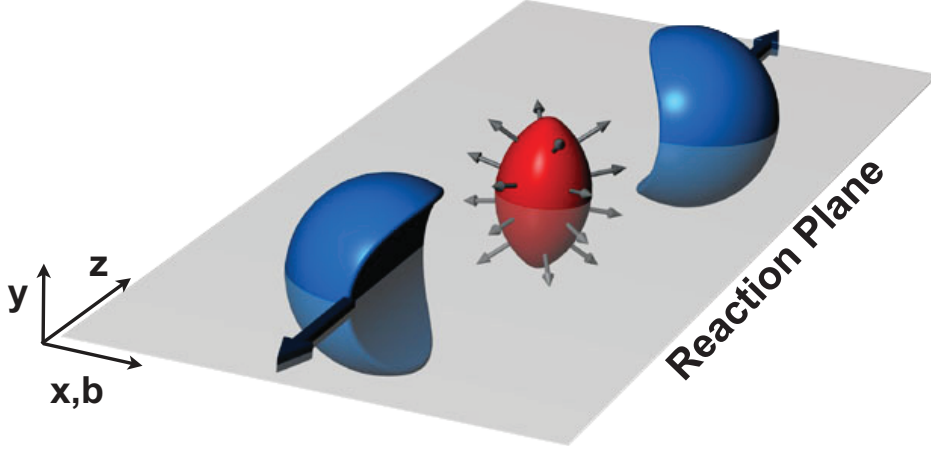
**Figure 2.2:** Illustration of non-central heavy ion collison and definition of reaction plane. Taken from [11].

2.2.3 Strangeness Production

Amongst the most useful tools to study dynamics of heavy ion collisions is also strange particle production. Back in 1982, it was proposed that enhancement in strange particle production in A+A collisions with respect to p+A interactions signifies the phase transition from hadronic gas to quark-gluon plasma [13]. It is argued that in quark-gluon plasma strangeness can be easily produced via pair production of strange and anti-strange quark pairs through gluon fusion processes like $gg \rightarrow s\bar{s}$ and thus in later hadronization stage strange hadrons can be formed in abundance via recombination or coalescence of these quarks which would be very unexpected without onset of deconfinement [14, 15].

Moreover, multi-strange baryons, which consist of more than one strange quark, are expected to have small hadronic cross section, therefore, considering also their large mass, they should be less prone to hadronic rescattering in later stages of heavy ion collisions. Employing models based on hydrodynamics it has been demonstarted that thermal freeze-out of multi-strange baryons occurs close to chemical freeze-out, i.e the composition of strange hadron is set shortly before produced hadrons fly away from the collison point without any further interactions. This would mean that they are indeed less affected by later stages of heavy ion collisions and their transverse flow reflects the flow of partons in the initial stages. Consequently, elliptic flow of multi-strange baryons could tell us about collectivity within QGP [16].

In the Fig. 2.3 measurement of v_2 in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV

by STAR is presented. It can be seen that the flow of multi-strange hadrons $\Omega(sss)$ and $\phi(s\bar{s})$ scale similarly to proton and π-mesons which consist of lighter u and d quarks. If we consider the above argument about strangeness not being influenced by hadron scattering, we can say that elliptic flow of all particles is indeed built up mostly in the partonic stage. The consistency in v_2 of particles is demonstrated nicely by Fig. 2.4 where the magnitude of flow is scaled by number of constituent quarks. The latest STAR results show that open charm D^0 mesons are subject to flow comparably to lighter particles and further establish the argument of flow build up within QGP [17].
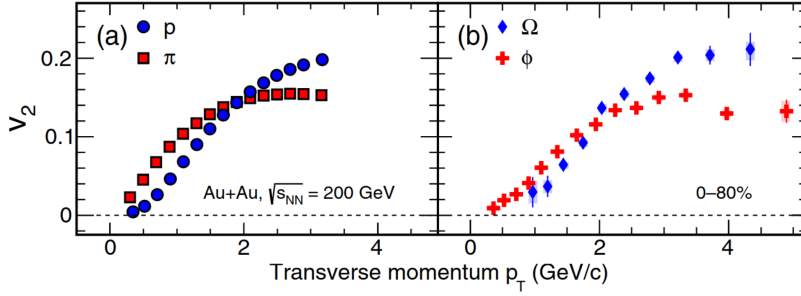
Figure 2.3: Elliptic flow coefficient v_2 of hadrons measured at STAR in $\sqrt{s_{NN}} = 200$ GeV Au+Au collisions. Taken from [18].
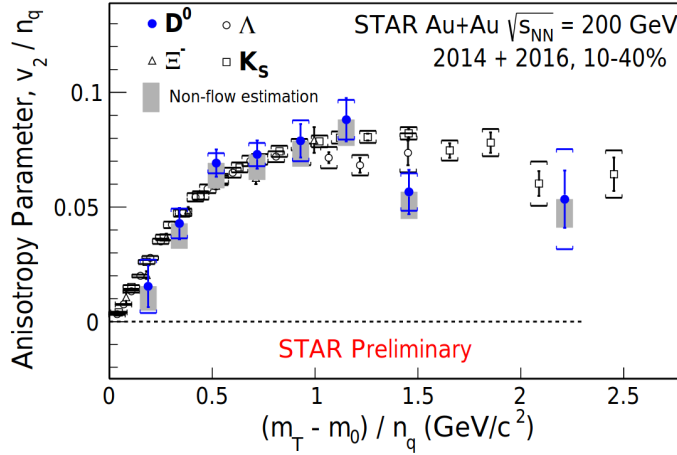

Figure 2.4: v_2/n_q as a function of $(m_T - m_0/n_q)$ for D_0 and $\bar{D}_0$ mesons combined in $10 - 40\%$ central Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV along with K_S^0, Λ and Ξ. Taken from [17].

2.3   QCD Phase Diagram and Beam Energy Scan

The main idea of Beam Energy Scan program is to investigate the phase diagram of strongly interacting matter. The current understanding of nuclear matter is illustrated by phase diagram of strongly interacting matter

in Fig. 2.5. The diagram depicts relation between temperature and the so called baryon chemical potential μ_B which quantifies balance between matter and anti-matter. Finite temperature lattice QCD calculations suggest that the transition from hadron gas to the state of deconfined quarks and gluons takes form of a crossover at vanishing baryon chemical potential and temperature around $T_c = 154 \pm 9$ MeV [19]. That means the transition is not accompanied by any discontinuities. On the other hand, based on QCD calculations, it is believed that at lower temperature and higher chemical potential the transition is of first order [20]. If correct, this would mean that there has to be a critical point at which the first order phase transition changes to crossover. Finding the critical point would be a smoking gun for existence of the phase transition and give us a permanent mark in the QCD diagram.
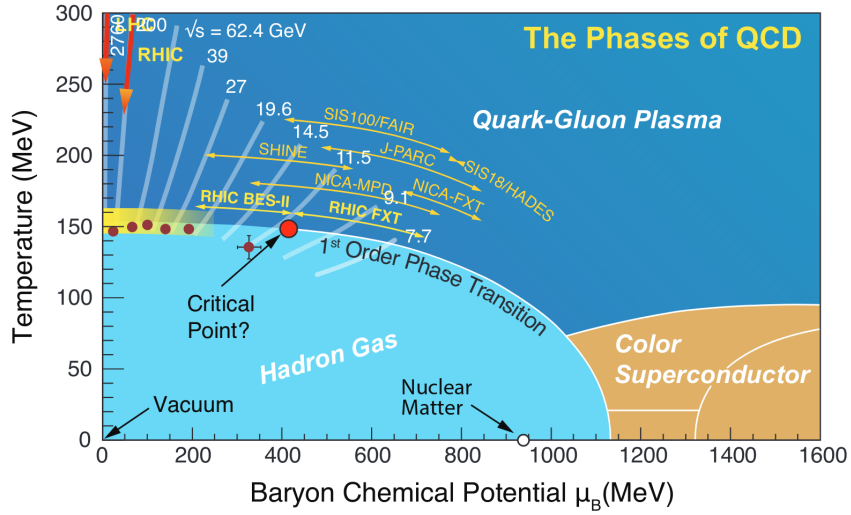

Figure 2.5: Illustration of phase diagram of nuclear matter. Taken from [21].

When establishing signatures of presence of QGP in heavy ion collisions, STAR investigated mainly collisions of gold nuclei at center of mass energy $\sqrt{s_{NN}} = 200$ GeV and found no evidence of first order transition. At this energy or at even higher energies at LHC, the baryon chemical potential is very low since matter and anti-matter are created equally. So, in order to map out the QCD phase diagram, one has to go lower with collision energy. This is why STAR already collected data at energies from $\sqrt{s_{NN}} = 200$ to 7.7 GeV during first phase of Beam Energy Scan in 2010 and 2011 [22] although with very low statistics. The second phase of BES introduces upgrades to both STAR (iTPC, eToF, EPD) and RHIC (electron cooling) and will add statistics from collisions at $\sqrt{s_{NN}} = 7.7, 9.1, 11.5, 14.6$ and 19.6 GeV while the linked fixed target program will supply data from collisions at $\sqrt{s_{NN}} = 3.0, 3.2, 3.5, 3.9$ and 4.5 GeV [23]. The goals of BES include searching for the predicted first order phase transition and the critical point, investigate expected turn-off of QGP signatures and look for evidence of chiral symmetry restoration.

As for the turn-off of QGP signatures, it can be clearly observed by looking at R_{CP} factor of charged particles. The R_{CP} factor represents ratio of yield of particles produced in central heavy ion collisions to yield in peripheral collisions related to number of participating partons. In the Fig. 2.6 one can see that at higher collision energies charged high-p_{T} particles are suppressed in central collisions with respect to peripheral ones which suggests a presence of medium which is opaque for particles with color charge. On the other hand, with decreasing collision energy we observe enhancement of charged high-p_{T} particles in central collisions and therefore believe that the quark-gluon plasma is not created.
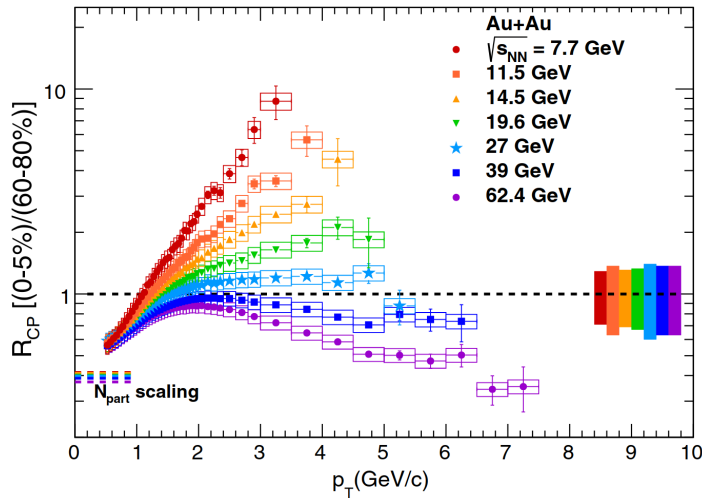
Figure 2.6: Charged hadron R_{CP} for RHIC BES energies as measured by STAR. Taken from [24].

One of the approaches to search for QCD critical point comes from sensitivity of fluctuations of conserved quantities such as baryon number, charge, and strangeness to the phase transition. Non-monotonic variations of fluctuations in these quantities with changing beam energy are considered to be one of the distincitve signatures of the QCD critical point. Experimentally, one can measure moments of the event-by-event net-particle distributions (particle multiplicity minus antiparticle multiplicity), such as net-proton, net-kaon and net-charge multiplicity distributions in heavy-ion collisions. The moments are mean (M), variance (σ^2), skewness (S), and kurtosis (κ) and their products $S\sigma$, $\kappa\sigma^2$ and the ratio σ^2/M are usually constructed [25]. Fig. 2.7 shows STAR's measurements for net-kaon multiplicity distributions from 0-5% most central and 70-80% peripheral Au+Au collisions at various energies. Although there are large uncertainties, a hint of non-monotonic behaviour of $\kappa\sigma^2$ for central collisions can be seen in the region of $7.7-19.6$ GeV. It is the target of BES-II to provide more statistics for these collisions which will shed more light on the results.

9
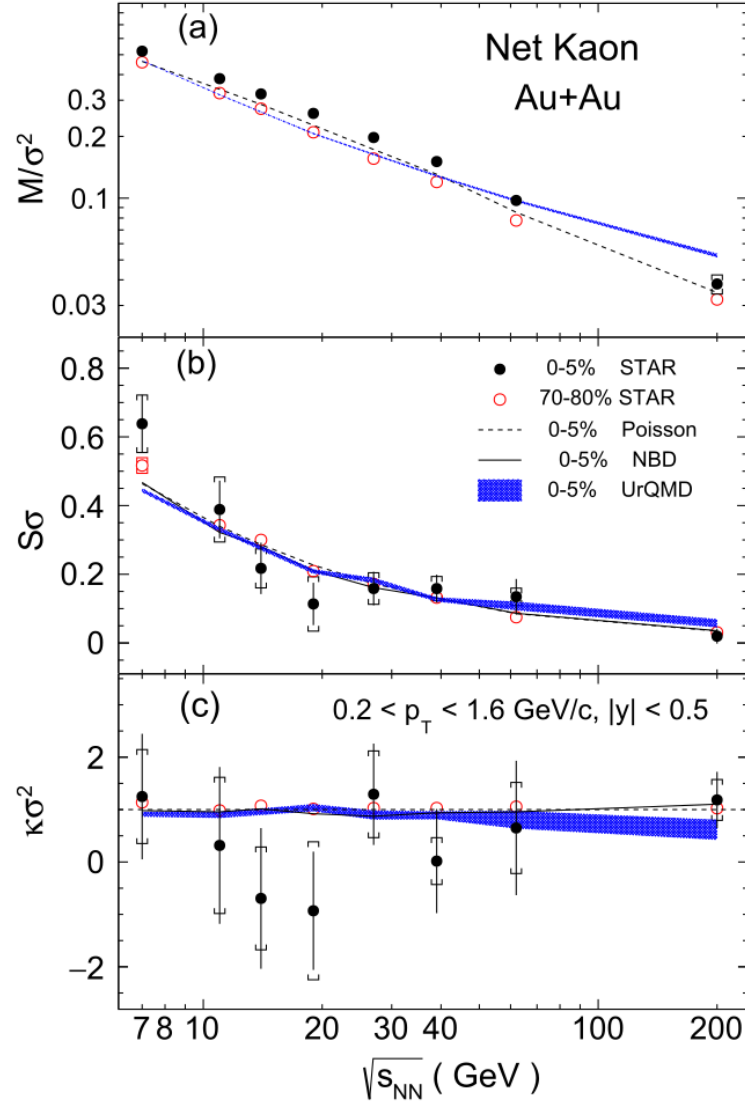
Figure 2.7: Collision energy dependence of the values of M/σ^2, $S\sigma$, $\kappa\sigma^2$ for net-kaon multiplicity distributions from 0–5% most central and 70–80% peripheral collisions in Au+Au collisions at $\sqrt{s_{NN}} = 7.7$, 11.5, 14.5, 19.6, 27, 39, 62.4 and 200 GeV. Taken from [25].

Chapter 3

Experimental Setup

3.1 Relativistic Heavy Ion Collider

Since the establishment of quark model in 1960s, the existence of medium with partonic degrees of freedom was debated. This concept was further developed by advancement of Quantum Chromodynamics, the theory of strong interaction. The possibility of exploring a new state of matter called Quark-gluon plasma became a common interest to the nuclear and particle physics community and the idea to create it by colliding heavy ions at relativistic energies was born [26].

The vision for construction of Relativistic Heavy Ion Collider (RHIC) in the United States was formulated in the long-range plan for nuclear physics in 1983 [27] by the Nuclear Science Advisory Committee (NSAC), an advisory body to the Department of Energy (DOE). The conceptual design for RHIC was submitted to DOE in 1984 by Brookhaven National Laboratory (BNL). The plan was to build two-ring superconducting hadron collider in the tunnel of Colliding Beam Accelerator (CBA) experiment and utilize the existing Alternating Gradient Synchrotron (AGS) as injector in order to meet cost expectations of NSAC [26]. The construction of the accelerator began in 1991 and was completed in 1999 [28].

The collider ring is 3.8 km in diameter and consists of two hexagonal concentric rings named "Blue" for clockwise and "Yellow" for counter-clockwise beams. The rings intersect at six points from which four have had experiments built around them. The four experiments at RHIC are BRAHMS, PHENIX, PHOBOS, and STAR. However, currently only STAR is still active because PHOBOS and BRAHMS were decommissioned in 2005 and 2006 respectively and PHENIX is undergoing modernization into new experiment called sPHENIX [29]. The whole accelerator complex can be seen in Fig. 3.1.

RHIC is able to operate with both ion and proton beams. The top energy for ion beam is 100 GeV/u (where u stands for atomic mass unit) and for proton beam it is 250 GeV. Before the beam is accelerated to its desired energy in the RHIC itself, it is pre-accelerated in a chain of other

machines. In the past, the injection chain for RHIC consisted of three different accelerators, namely the Tandem Van de Graaff, the Booster, and the AGS [30]. Nowadays, the Tandem Van de Graaff is replaced with Electron Beam Ion Source (EBIS) which is followed by Radio Frequency Quadrupole accelerator (RFQ) and a short linear accelerator. The new EBIS setup requires shorter transfer line to the Booster and is able to produce Au^{32+} beams at 2 MeV/u. EBIS is also able to deliver uranium beams [31, 32]. In the Booster ions are accelerated to 95 MeV/u and at the exit further stripped before acceleration in AGS to the RHIC injection energy of 10.8 GeV/u. Gold ions are fully stripped to a charge state of +79 at the exit from the AGS [28].

The flexibility of RHIC and the ability to cover wide range of beam energies is vital to the study of phase diagram of strongly interacting matter. At RHIC it is possible to operate beam with energy as low as 3.85 GeV/u which is lower than the injection energy. In order to control such a beam electron cooling upgrade was commisioned for the ongoing second phase of Beam Energy Scan program. The upgarde aims to combat intrabeam scattering and is helping significantly with reaching desired luminosity and statistics [33].
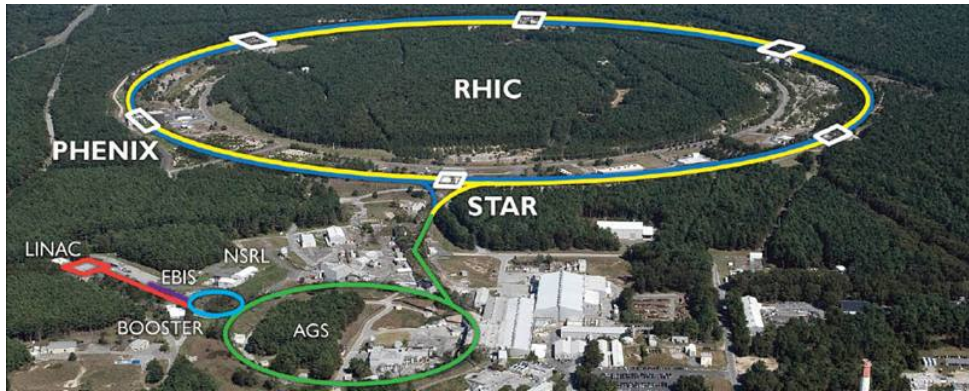

Figure 3.1: RHIC accelerator complex in Brookhaven National Laboratory. Taken from [34].

3.2 Solenoidal Tracker at RHIC

The experiment STAR (Solenoidal Tracker at RHIC) was constructed for the purpose of study of heavy ion collisions and quark-gluon plasma. For this reason it was designed to record high multiplicity events and cover full azimuth within pseudo-rapidity region $|\eta| < 1$. The detection system consists of Time Projection Chamber inside a solenoidal magnet to enable tracking, momentum analysis, and particle identification via dE/dx. A Time of Flight detector for particle identification at higher momenta surrounds TPC. Outside of magnet there is also Barrel Electro-Magnetic Calorimeter for triggering and distinguishing electrons from hadrons [35].

Throughout the years STAR subsystems were upgraded, new ones like Muon Telescope Detector (MTD) were added and some were already removed. In Fig. 3.2 reader can see a diagram of STAR and its subdetectors from year 2016. Note that there is Heavy Flavor Tracker (HFT) present, which was added to STAR in 2014 and remained in use until 2016 [36]. Most recent upgrades of STAR include inner Time Projection Chamber (iTPC) and endcap Time of Flight (eToF) which are now already completely built into tracking and time of flight measurement. Another planned upgrade denoted simply as STAR forward upgrade is meant to combine tracking system with electromagnetic and hadronic calorimeters in order to cover region $2.5 < \eta < 4.5$ in pseudo-rapidity [37]. Below two components of STAR will be described in more detail, namely Time Projection Chamber and Time of Flight detector.
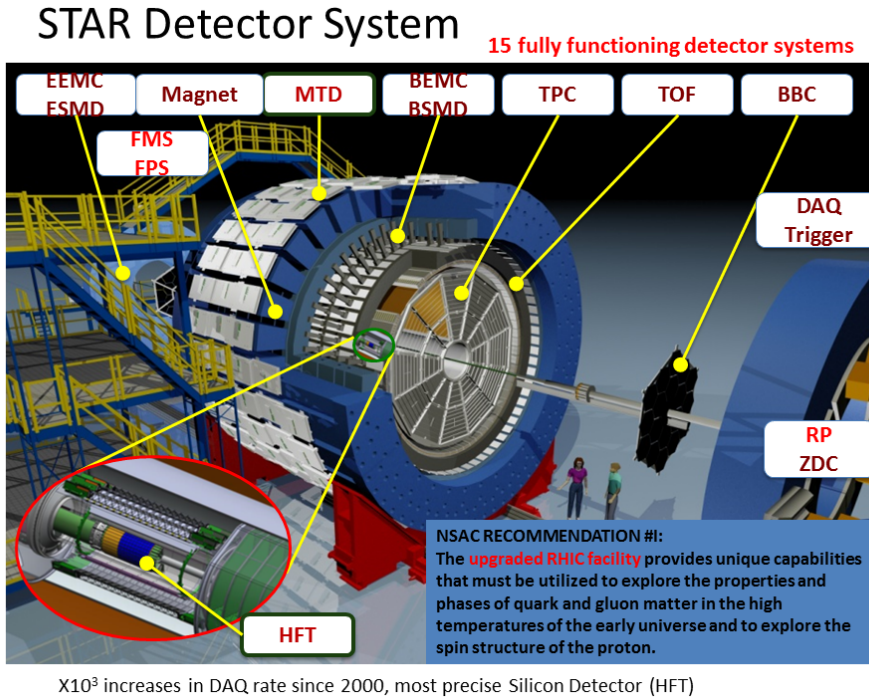

Figure 3.2: STAR detector configuration from 2016. Taken from [38].

3.2.1 Time Projection Chamber

The Time Projection Chamber is a main STAR tool used for tracking of particles. It is positioned inside magnetic field of a solenoidal magnet with induction of $B = 0.5$ T. The chamber itself is 4.2 m long and 4 m in diameter. It is filled with P10 gas (10% methane, 90% argon). The gas gets ionized by charged particles passing through volume of the chamber and releases electrons which drift up to 2.1 m to the readout end caps thanks to the present uniform electric field of ≈ 135 V/cm. The readout system consists of Multi-Wire Proportional Chambers with readout pads [39].

13

Besides recording of tracks, TPC also measures their momenta and allows to determine the energy loss through ionization which helps to identify charged particles. Acceptance of the TPC covers $|\eta| < 1.8$ and it is possible to measure particle momenta in range from 100 MeV/c to 30 GeV/c [39].

The inner Time Projection Chamber (iTPC) upgrade was introduced to STAR in 2019 for the purposes of Beam Energy Scan II and its physics goals, i.e. study of phase diagram of nuclear matter. The idea of the upgrade is to increase the segmentation on the inner padplane and renew the inner sector wire chambers. This should lead to improvement of tracking at small angles relative to beam line and hence expansion of acceptance of the chamber to $|\eta| \leq 1.5$ [40]. The momentum and dE/dx resolution as well as acceptance at low momenta should also benefit from this update as is discussed further in the analysis section of this work.
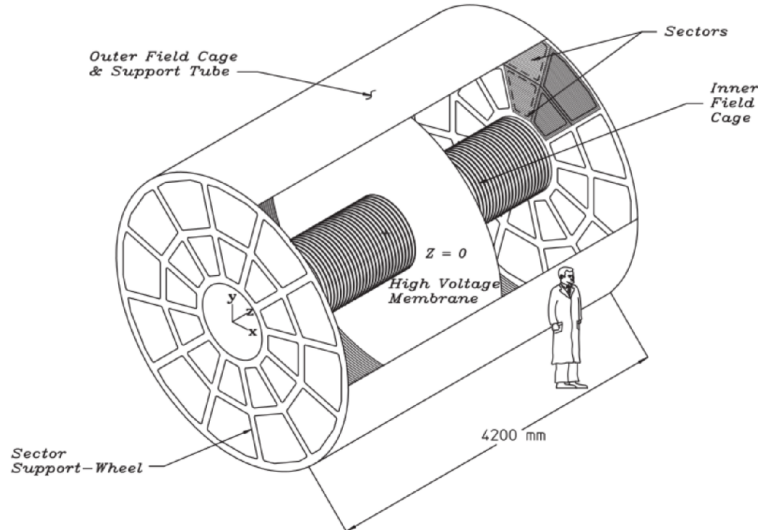
Figure 3.3: Schematic depiction of STAR TPC. Taken from [39].

■ 3.2.2 Time of Flight detector

The Time of Flight detector is based on Multi-gap Resistive Plate Chamber Technology (MRPC) [41]. It was built in order to improve particle identification for particles with momenta up to 3 GeV/c. MRPC modules are arranged in 120 segments around a cylinder which covers $|\eta| < 0.9$ in pseudo-rapidity. They take the time when particle passes through the detector, while Vertex Position Detector (VPD) sets start time of the flight. Endcap Time of Flight (eToF) was similarly to iTPC installed at STAR for in 2019 in order to further enhance particle identification. It incorporates technology that will be used at future project CBM in FAIR facility which has been adapted for STAR.

14

▉ 3.3 **Fixed Target Experiment Setup**

As mentioned in previous chapter fixed target program is meant to complement the collection of data from collisions with low center of mass energies and thus high baryon chemical potential. It allows to go as low as $\sqrt{s_{NN}} = 3.0$ GeV with collision energy which corresponds to $\mu_B = 721$ MeV [23]. When running in fixed target mode there is only one beam circulating in RHIC which is subsequently deflected by RHIC operators to hit a fixed gold target placed inside beam pipe 200 cm from the center of TPC. The energy of the beam is actually the same as for one of the BES-II settings but the center of mass energy in fixed target mode is lower. So for example the 14.6 GeV collisions were measured with two beams having 7.3 GeV each and the same beam energy was used for fixed target mode to get 3.9 GeV collisions.

Before the launch of fully-fledged fixed target program there had to be several tests done [42]. First, as a proof of principle, analysis of collisions between ions in beam halo and aluminium nuclei in the vacuum pipe was carried out. The plot on the left side of Fig. 3.4 shows vertices of these collisions. After that in 2014, a thin gold foil target was installed in the beam pipe on the west edge of the TPC about 210 cm from the center. The target was, however, illuminated only with gold beam halo again. In Fig. 3.5 one can see detector setup from Run14 during which the fixed target test was done. On the right side of Fig. 3.4 there is a photo of the gold foil target and its support structure. Fig. 3.6 shows vertices of collisions between gold beam halo and gold nuclei in target.
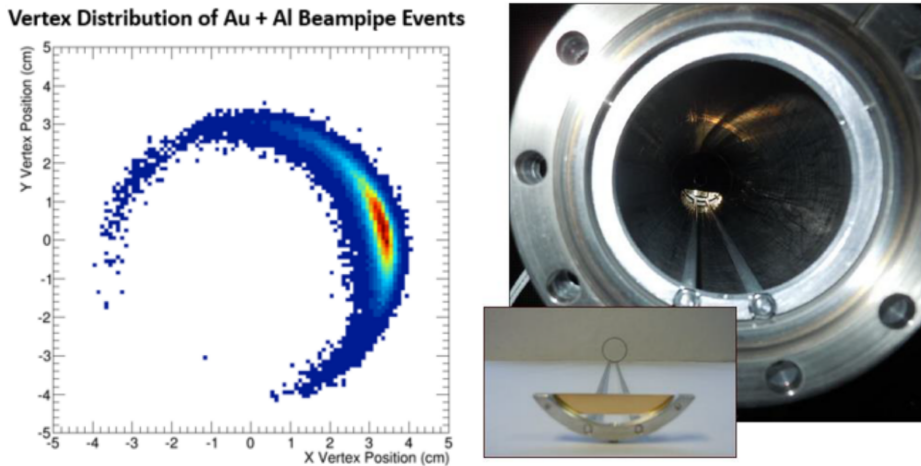
Figure 3.4: Vertices of collisions between gold beam halo and aluminium nuclei in beam pipe (left), photograph of the gold foil target and its support structure (right). Taken from [42].
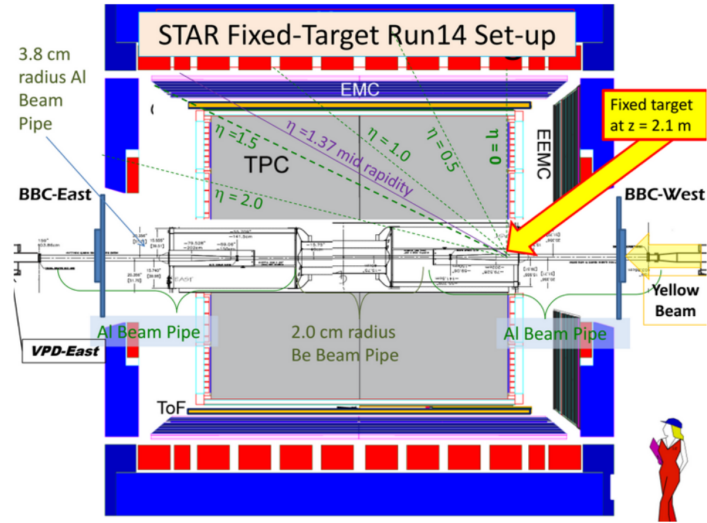
15

Figure 3.5: Diagram of the fixed target test run detector setup. Taken from [42].
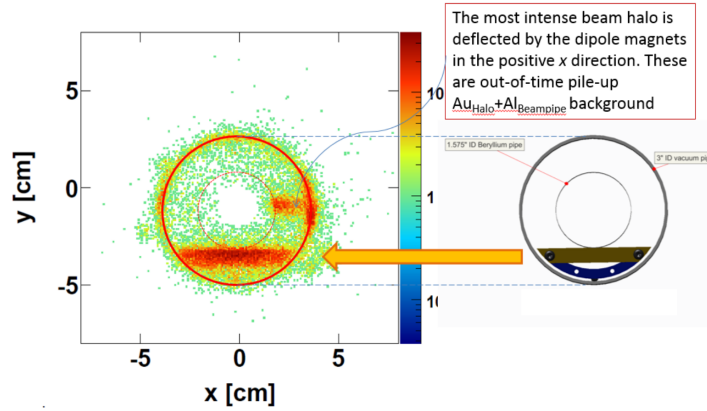

Figure 3.6: Vertices of collisions between gold beam halo and gold nuclei in fixed target. Taken from [42].

3.4 High Level Trigger and Tracking at STAR

Since modern particle physics experiments measure events at high rates and observe collisions which usually produce large amount of particles, they have to deal with immense amount of data. It becomes impossible to store this amount of data due to the real-world limitations in data storage capacity and rates. This is why trigger systems are essential parts of the experimental setup in high-energy physics. These systems are able to swiftly decide whether events are worth keeping or not. These decisions are initially based on simple criteria on a level of hardware, but there are also higher level triggers which take events from lower levels as an input and can take more time to take a better decision. Such a high level trigger can include simple physics analysis of events with fast software. This is why there is continuous effort being

16

made to develop fast algorithms for data analysis.

Many high-energy physics experiments pursue development of a uniform system of data processing which would remove the traditional separation between the online data processing in real time and the offline analysis of the stored data. For future experiments like CBM, where they will work with interaction rates of up to 10 MHz and consequently have to deal with data flow of up to 1 TB/s, this becomes a necessity. At STAR it is High Level Trigger (HLT) group and also Tracking Focus Group (TFG) who have development of fast online and offline algorithms among their agenda. Since 2011 they cooperate with Frankfurt Institute of Advanced Studies (FIAS) on development of software for online and offline reconstruction of events based on the Cellular Automaton (CA) and the Kalman Filter (KF) approaches. The FIAS group helped to implement their KF Particle package into STAR's HLT in order to speed up STAR's data processing ability and also test the software which is supposed to make a base for the CBM physics program.

Because STAR collaboration was concerned about the constrained computing resources and disk storage, the ultimate goal of HLT, TFG and FIAS groups for BES-II program is to provide offline quality data practically online and in this way publish the first BES-II results as soon as possible. The speed up is to be reached among other techniques by exploiting computational power of HLT's cluster by parallelizing and porting the KF Particle Finder to Intel Xeon Phi cards. More on KF Particle based software is presented in Chapter 4. Express production of BES-II data is analysed in this thesis. Besides triggering, HLT benefits BES-II program also by online 3D primary vertex reconstruction, real time beam position monitoring in RHIC low energy runs and providing live feedback to Collider-Accelerator Department for accelerator monitoring and performance tuning.

Based on the calibrations made online by HLT, Tracking Focus Group is able to produce a Monte Carlo simulation with full detector effects. It is a simulation of a pure particle signal and currently it is not embedded into data, however, it can be used to get the idea about efficiencies of reconstruction algorithms. In this thesis an attempt to calculate efficiency of Λ baryon reconstruction with KF Particle Finder using this online calibrated MC simulation is presented. Understanding the efficiencies is a crucial part of any physics analysis and having a decent Monte Carlo simulation available so early would benefit the pursue of almost online analysis of data substantially.

Chapter 4

Reconstruction of Particles with KF Particle Framework

The Beam Energy Scan at RHIC is not the only program aiming for exploration of the phase diagram of strongly interacting matter. Worldwide efforts are being put into research of this topic from both theoretical and experimental physicists. The Compressed Baryonic Matter (CBM) at future Facility for Anti-Proton and Ion Research (FAIR, Darmstadt,Germany) will be one of the experiments devoted to the region of large net-baryon densities in the QCD phase diagram [43]. CBM's main focus will be measurement of very rare probes. For this purpose it will have to work with interaction rates of up to 10 MHz and consequently deal with data flow of up to 1 TB/s. That is why CBM is planing to fully reconstruct events online and make event selection with advanced trigger which will require information from several detector sub-systems. The majority of trigger signatures will be complex and will incorporate short-lived particles [44].

For the purpose of reconstruction of full event topology so called First Level Event Selection (FLES) package was developed. It includes several modules for track finding, track fitting, particle reconstruction and physics selection. The reconstruction of short-lived particles is done with KF Particle Finder which is based on the KF Particle package. KF Particle Finder combines already found tracks of charged particles which are decay products of particles in reconstruction. It also selects particle candidates from these random combinations. KF Particle Finder uses intrinsically local and parallel algorithms as well as the rest of the FLES package and thus can run online on dedicated many-core CPU/GPU cluster [45]. Both KF Particle and KF Particle Finder are further described below.

4.1 KF Particle

KF Particle is a C++ package developed for the purpose of reconstructing particles together with their parameters [2]. The estimation of parameters is based on Kalman filter (KF) algorithm, although the standard Kalman

filter approach is modified. In the package particles are parametrized by state vector

$$\mathbf{r} = (x, y, z, p_x, p_y, p_z, E, s)^{\mathrm{T}}, \qquad (4.1)$$

where (x, y, z) are coordinates of the particle, (p_x, p_y, p_z) are momentum components and $s = l/p$ is a distance between decay points normalized on the momentum. This parametrization makes the algorithm independent on geometry of the detector system. Together with state vector Kalman filter outputs also particle's covariance matrix, which contains parameter uncertainties. This allows for calculation of χ^2 criteria which estimate the quality of the reconstruction. These criteria can be later used for combinatorial background rejection.

In comparison to Kalman filter method, conventional reconstruction packages search for point of closest approach of daughter tracks, extrapolate parameters of daughter tracks to this point and simply sum up obtained momenta and energies to calculate parameters of decaying particle. This approach, however, usually neglects parameter uncertainties and hence does not allow for calculation of statistical criteria which can be used for background rejection. Instead, these methods usually rely on cuts on absolute values of topological variables.

The design of KF Particle package allows to treat mother and daughter particles in the same way and daughter particles are added to the mother particle independently from each other. This allows for convenient reconstruction of decay chains. The package is also implemented in single precision and is parallelized so it can run on machines with SIMD (Single Instruction, Multiple Data) architecture [3]. This is essential for the computing speed CBM is aiming for.

Moreover, the KF Particle package provides, besides the reconstruction of mother particle, user with functionalities which can be exploited to further simplify physics analyses, such as:

- access to the physical parameters of the particle (mass, momentum, decay length, lifetime, rapidity, etc.) together with their uncertainties,
- transport of the particle to an arbitrary point, to the decay and production points, to another particle, to a vertex, on the certain distance,
- calculation of a distance from a particle to a point, to another particle, to a vertex,
- calculation of a deviation in terms of χ^2 from a point, from another particle, from a vertex;
- calculation of the angle between two particles;
- setting constraints on the particle state vector on mass of the particle and on the production point of the particle [3].

◼ **4.2 KF Particle Finder**

The aim of KF Particle Finder is to utilize mathematics defined in KF Particle and offer an algorithm which finds, reconstructs and selects short-lived particles. The resulting software package which implements the algorithm is based on SIMD instructions and the algorithm itself is parallelized between cores of the CPU. This enables to efficiently exploit resources of state-of-the-art CPUs [3]. The speed of the algorithm will be essential for it to be used as a part of First Level Selection Package at CBM since the ultimate goal there is to reconstruct and select events online. It was already demonstrated that the whole FLES package with inclusion of KF Particle Finder scales on the many-core CPU systems with respect to the number of cores as can be seen in Fig. 4.1 [44].
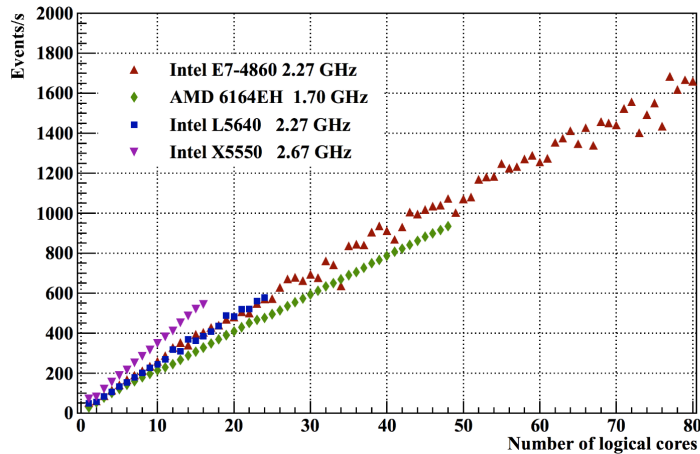
**Figure 4.1:** Scalability of the FLES package on many-core servers. Taken from [44].

The fact that KF Particle treats daughter and mother particles in the same way was used conveniently in KF Particle Finder. The reconstruction algorithm is able to use reconstructed particles as daughters and thus in a recursive way reconstruct decay chains in one session. Currently, there are over 100 decays implemented in KF Particle Finder as is illustrated in Fig. 4.2.

The reconstruction of short-lived particles begins with providing tracks of charged particles detected in the experiment as an input. The input has to also include the correlation matrix with information about track uncertainties. At first, KF Particle Finder classifies tracks into primary and secondary, i.e. tracks that either do or do not come from primary vertex - the position where the collision took place. For this purpose χ^2_{prim} criterion is calculated

$$\chi^2_{\mathrm{prim}} = \Delta \vec{r}^T \left(C_{track} + C_{PV} \right)^{-1} \Delta \vec{r}, \tag{4.2}$$

where $\Delta \vec{r}$ is difference between the track and the primary vertex position and C_{track} and C_{PV} are covariance matrices of the track and primary vertex

respectively. The criterion is basically distance between the track and primary vertex normalized on the total error. Under the assumption of particle parameters following Gaussian distribution and χ^2_{prim} following χ^2 distribution, the criterion represents probability of the trajectory intersecting primary vertex within uncertainties, i.e. classifying as primary. So for example if the criterion value $\chi^2_{\text{prim}} = 18.6$ there is an probability of 0.01% that the track is primary. As a matter of fact this is the value that is used in the KF Particle Finder code by default to divide tracks into primary and secondary.
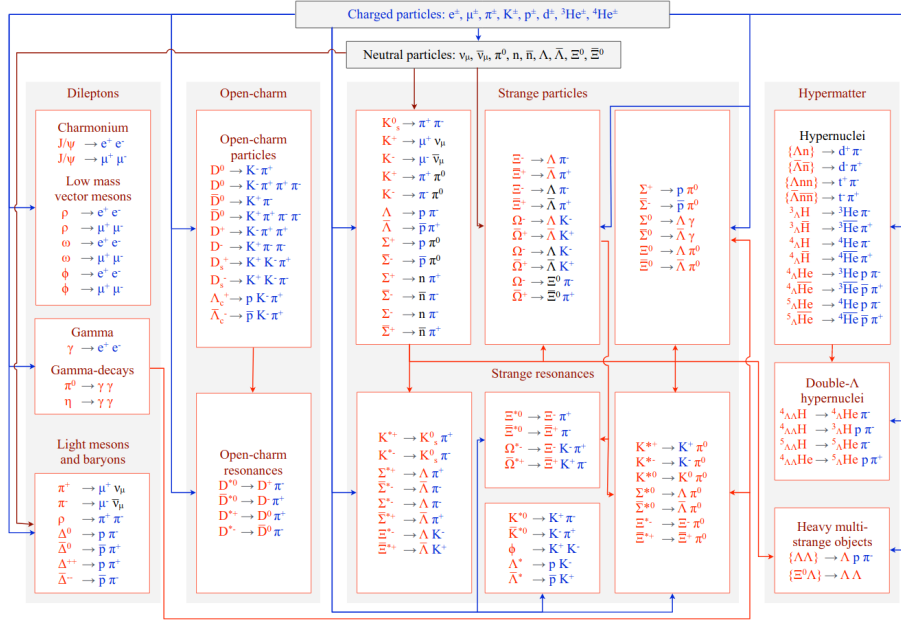
Figure 4.2: Block diagram of the KF Particle Finder package. Taken from [46].

After all tracks are divided, short-lived mother particles are reconstructed using KF Particle. In principle, candidate particles are constructed simply by making all possible combinations of corresponding daughters. In this way, however, a combinatorial background is created which needs to be suppressed somehow for most of the decays. For this purpose KF Particle Finder enforces cuts on statistical criteria that relate to the topology of the decay in question. Some of these are based on χ^2 values calculated within KF Particle using the covariance matrix. The crucial criteria are following:

- χ^2_{fit}/NDF criterion characterizes probability of daughter trajectories intersecting within their uncertainties,
- χ^2_{topo}/NDF characterizes whether the mother particle comes from the primary vertex region and therefore it is used to divide mother particles into primary and secondary,
- $l/\Delta l$ is a distance from the primary vertex to the decay point (decay length) normalized on its error [3].

In order to test abilities of FK Particle Finder the framework was im-

plemented into High Level Trigger at STAR which contributed to further development of the package by testing it in established environment. Furthermore, the package promises to be beneficial for offline analyses as well. There have been studies of various decays done with KF Particle Finder presented within STAR collaboration, which suggest that the statistical approach to topological cuts is more efficient and this work aspires to add to this argument. Particularly, the improvement in low p_T region is expected since the statistical approach should be better when working with low p_T tracks which have increased curvature in the magnetic filed of the detector and come with greater uncertainties. An example of previous work done with KFPF which was presented at STAR Analysis Meeting is shown in Fig. 4.3. It shows an improvement in signal significance of D-mesons for whole p_T range.
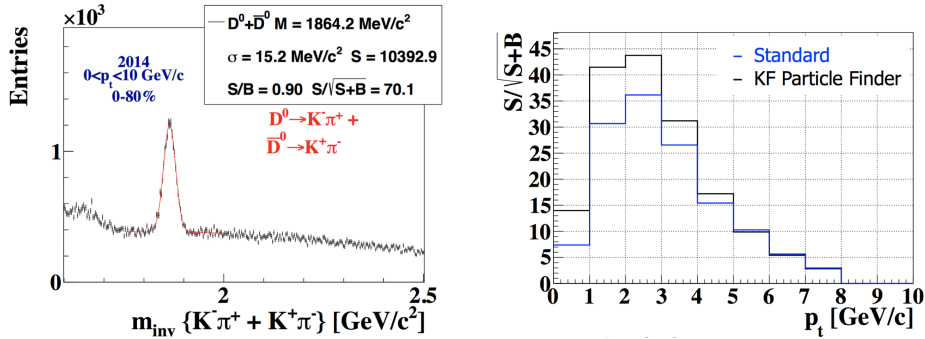
Figure 4.3: Reconstruction of D-mesons with KF Particle Finder presented at STAR Analysis Meeting shows improvement in signal significance with KF Particle Finder.

At STAR the KF Particle Finder framework was customized to follow the standard structure of analysis code as well as accept STAR internal data formats as input (picoDst, MuDst). The core class of the STAR's version of the package is the `StKFParticleAnalysisMaker` class, which is a standard STAR maker with `Init`, `Make` and `Finish` methods. In this class the reconstruction of particles is initiated and it is where user is able to go through all reconstructed particles and work with them, e.g. store their parameters into ROOT ntuples.

The `StKFPartcileInterface` class controls processing of events. In the method `ProcessEvent` one can setup their own event-selection cuts, PID or track quality cuts on track parameters. The class contains also methods for controlling particle selection cuts. These can be called once the instance of `StKFParticleInterface` is created in the analysis code. This way one can apply cuts not only on criteria such as χ^2_{fit} or $l/\Delta l$ but also on maximum distance between particles and so forth.

Chapter 5

Toolkit for Multivariate Analysis and Boosted Decision Trees

5.1 Toolkit for Multivariate Analysis

Machine learning and techniques of multivariate analysis become increasingly more popular in both science and industry. This applies to high-energy physics (HEP) as well, since there is a common task to identify rare signal events in immense background and for that multivariate methods can be utilized. Typically, variables relevant to the physics problem in question are selected and a machine learning model is trained for classification or regression using samples of signal and background events. Moreover there is also an interest to exploit machine learning in order to shorten execution time of computationally expensive parts of event simulation, pattern recognition, calibration or other tasks relevant for HEP in general [47].

Toolkit for Multivariate Analysis (TMVA) contains a large selection of multivariate classification and regression algorithms. It is integrated into C++ based analysis framework ROOT and offers user interfaces for training, testing, performance evaluation and application of all available classifiers. The toolkit is designed specifically for HEP, however in principle it is not restricted to it. TMVA also comes with a graphical user interface (GUI) for displaying various control plots. Among the available methods are:

- Rectangular cut optimisation
- Projective likelihood estimation
- Multi-dimensional likelihood estimation
- Linear and nonlinear discriminant analysis
- Artificial neural networks
- Support vector machine
- Boosted/bagged decision trees
- Predictive learning via rule ensembles
- A generic boost classifier for boosting of any of the above classifiers [48, 49].

The analysis employing TMVA usually consists of two phases. The first phase includes the training of requested multivariate methods as well as testing and evaluation. This is handled by the TMVA `Factory` object, which should be created in the beginning of the program. The input data are passed to the `Factory` in the form of ROOT trees, i.e. objects derived from class `TTree`. Samples for both training and testing can be contained in the same tree and the division can be carried out within the `Factory`. The results of training are stored in a xml "weight" file whose structure is specific for each classifier. The output of testing and evaluation can be saved in a root file and subsequently viewed with TMVA's GUI.

The second phase aims to apply trained results on data set with unclassified samples. For this the `Reader` class is designated. The data set described by the same set of variables as in the training phase must be given to the `Reader`. All events from the data set are evaluated in a loop in which the response value of multivariate method in question is calculated for each event based on the results from the training contained in the weight file. In the analysis of the physics problem one than usually makes cut on this value in order to classify signal and background events with corresponding efficiency [49].

■ 5.2 Boosted Decision Trees

Together with neural networks Boosted Decision Trees (BDT) belong to the most popular machine learning classifiers in high-energy physics [47]. For example, they were employed in analysis of Higgs boson at CMS in CERN [50]. This classifier was also used in the analysis presented in this thesis.

Let us first start with growing a single decision tree. A decision tree consists of nodes. The se nodes represent a division of training data into two subsets based on a value of particular variable. This division aims to separate signal from background as much as possible. The separation can be quantified by various separation criteria, e.g. Gini Index or statistical significance. Starting from the root node, each node is further divided until a stopping condition is met, i.e. when a desired separation of signal and background is reached or in some cases when the maximum depth of the tree is reached. This whole process of forming a decision tree is illustrated by Fig. 5.1. End-nodes of a decision tree are called leaf nodes and these are classified as background or signal depending on the majority of training events that end up in the node.

The boosting of a decision tree means growing a large set of trees - a forest. Such a forest typically consists of a large number of shallow trees since the boosting in general is based on combining a large amount of weak learners (classifiers) into effective single on e. The final classifier is given by weighted average of individual decision trees. In the case of TMVA the output of BDT
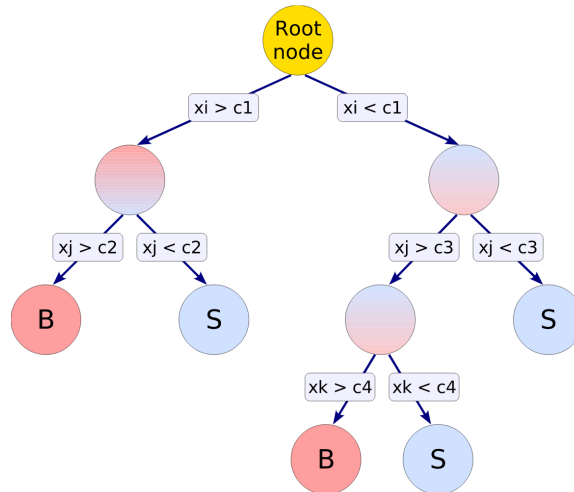
Figure 5.1: Schematic view of a decision tree. Taken from [49].

is called BDT response value. It ranges from -1 to +1 and characterizes how is the particular event "background-like" or "signal-like". When compared to a single decision tree, boosting is better in performance and it is also able to stabilize the decision response with respect to fluctuations in the training sample. The re are several ways to introduce boosting to decision trees. The approach used in the analysis of this work is called Adaptive Boosting (AdaBoost). In the process of growing a forest with adaptive boosting, events that were misclassified during the training of a decision tree are given a higher event weight in the training of the following tree [49].

As for any other method in the Toolkit for Multivariate analysis there are various useful control plots based on evaluation of trained classifier on testing sample. The power of classifier is illustrated nicely with the so called ROC curve (Receiver Operating Characteristic). It demonstrates how the increase of background rejection leads to loss in signal efficiency. An example from the analysis presented further in this work is shown in Fig. 5.2. The curve follows expected trend which expresses the fact that it is not possible to train a classifier which is able to reject increasingly more background without sacrificing signal, i.e. loosing signal efficiency. In naive ideal case scenario the curve would form a rectangle by touching the top right corner of the plot. In practice the closer the curve gets to the top right corner the better the classifier has been trained.

In Fig. 5.3 one can see an example of plot of correlation matrices for signal and background samples which in this case display linear correlation coefficients of training variables used in the analysis part of this thesis. In one's analysis it might come out beneficial to check these plots since the study of intertwined variables is the essence of multivariate analysis.

Another example of an important plot from this theis's analysis part is
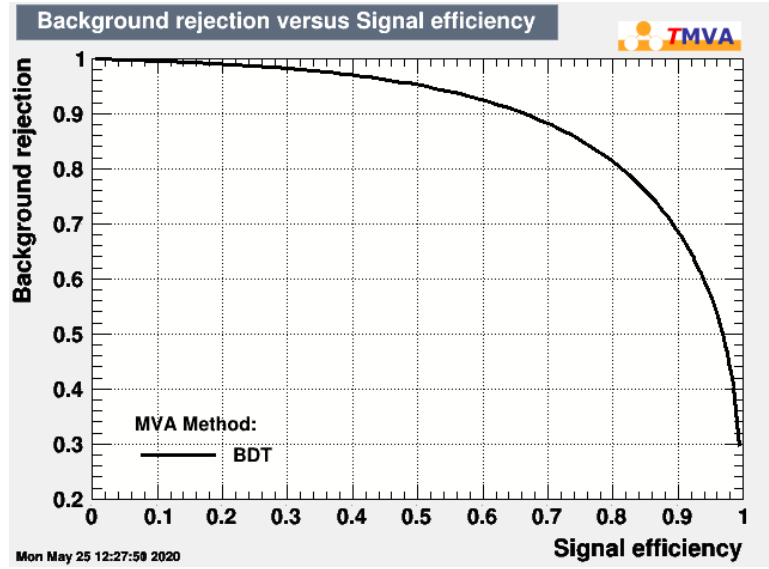
Figure 5.2: *THIS THESIS:* Receiver Operating Characteristic (ROC) curve demonstrating classifying power of Boosted Decision Trees trained for analysis presented in this thesis. It shows that it is impossible to reject all background without sacrificing signal.

displayed in Fig. 5.4. It shows distributions of BDT response value for training and testing data sets superimposed. This plot can be checked to see whether the classifier was overtrained. Decision trees are said to be especially prone to overtraining [51] which means that the classifier learns various statistical fluctuations in the signal sample and then looks for these when applied on un-classified data. This would manifest in different distributions of response value for training and testing sample and thus can be checked by inspection of superimposed distributions.

The decision on value of the cut on BDT response value in one's analysis can be based on resulting value of signal significance $S/\sqrt{S+B}$, where S stand for number of signal events and B for background. This means that the optimal cut value depends on the actual number of signal and background events in the particular data set which is a priori unknown. The dependency of signal significance for any number of signal and background entries can be plotted together with cut efficiencies within TMVA as is demonstrated in Fig. 5.5. For the purpose of determination of cut value, however, one needs to estimate the true count of signal and background. The estimation can be made by means discussed in Chapter 6 about the analysis part of this work.
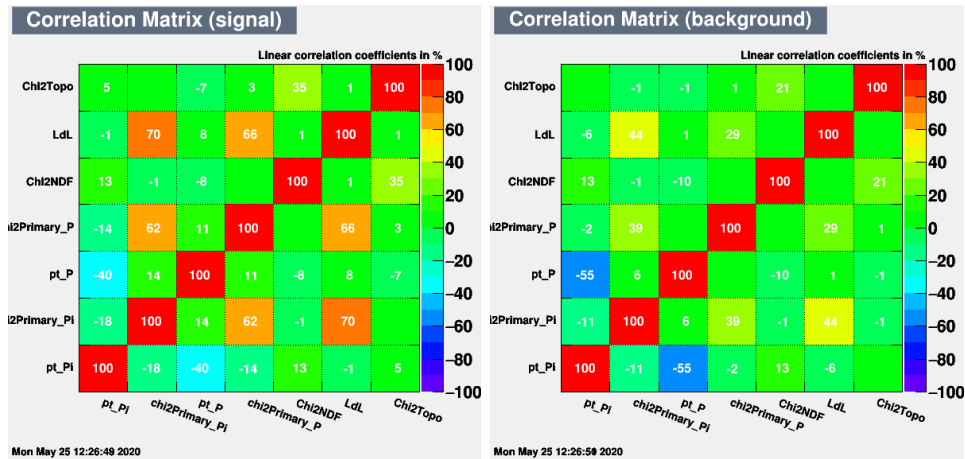
Figure 5.3: *THIS THESIS:* Linear correlation coefficients of training variables in a matrix for signal (left) and background (right) samples obtained within TMVA.
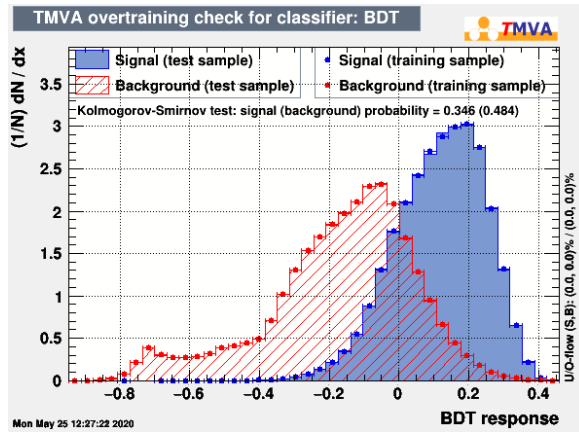

Figure 5.4: *THIS THESIS:* Superimposed distributions of BDT response value for training and testing samples.
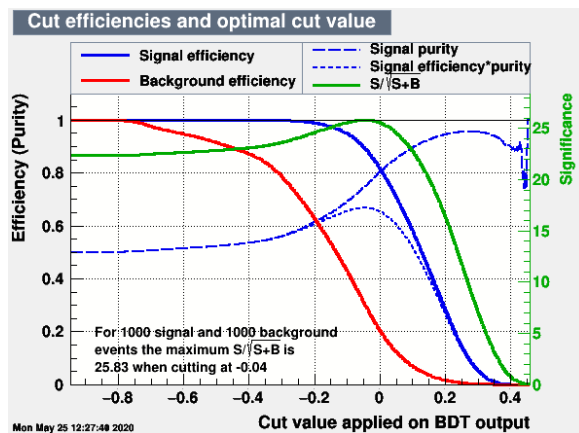

Figure 5.5: *THIS THESIS:* Cut efficiencies and optimal cut on BDT value as calculated by TMVA while assuming 1000 signal and 1000 background events.

29

Chapter 6

Strange hadrons reconstruction with KF Particle Finder

In this chapter author's analysis of data from Au+Au collisions measured by STAR during second phase of Beam Energy Scan program will be presented. This includes results from Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV, $\sqrt{s_{NN}} = 14.6$ GeV and fixed target collisions at $\sqrt{s_{NN}} = 3.9$ GeV. Please note that although results obtained by analysis of STAR data are presented, all figures marked with "*THIS THESIS*" represent purely author's personal work and have not been approved by STAR collaboration for public presentation yet.

6.1 Λ in Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV

6.1.1 Dataset and event selection

The data from Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV were collected by experiment STAR during 2018. Although author's work on data from these collisions begun with express production made by HLT+TFG, the official dataset with full calibrations was produced since then and thus results from analysis of official full production are presented bellow. That includes raw yields of Λ baryon obtained with KF Particle Finder and also with KFPF in combination with Boosted Decision Trees.

   The production code-named P19ib of picoDst data format was done at STAR using library SL19b. Total number of minimum bias events sampled in this analysis is 610M. From these minimum bias events 370M were selected for further analysis. The event selection was based on the position of primary vertex of each collision. It was required that the primary vertex is less than 70 cm away from the center of the Time Projection Chamber, i.e. $|V_z| < 70$ cm. Other criteria for event selection demanded at least 10% of tracks to classify as primary and required precise reconstruction of primary vertex in transverse plane, specifically $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} \leq 0.45$ cm. Fig. 6.1 shows the distribution of z coordinate of primary vertex before event selec-

tion on the left, the plot on the right demonstrates distribution after event selection.
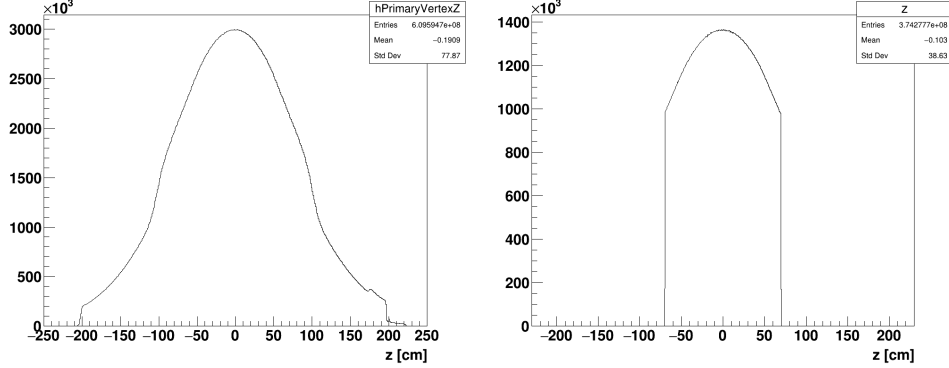
Figure 6.1: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Distribution of z coordinate of primary vertex in minimum bias events (left) from Au+Au at $\sqrt{s_{NN}} = 27$ GeV dataset. Distribution of z coordinate of primary vertex after event selection (right).

◼ 6.1.2   Analysis with default KFPF cuts

At first, in order to test the functionality of KF Particle Finder with custom changes in event-processing part of the code, the analysis of Λ was done with default particle cuts in KFPF. This means that cuts on statistical criteria defined by KFPF were set by constructors of relevant classes of the package as implemented by original developers. These values are explicitly listed in Tab. 6.1.

Cut value	Cut description
$\chi^2_{geom} < 10$	χ^2 of the track to the second daughter track
$l/dl > 5$	decay length normalized on the error
$l > 5$ cm	decay length
$\chi^2_{prim} > 18.6$	χ^2 of the track to primary vertex
$\chi^2_{topo} < 5$	χ^2 of the mother particle to primary vertex
$d_{max} < 1$ cm	maximum distance between daughter particles

Table 6.1: Default values of KF Particle Finder cuts.

The invariant mass spectra were extracted for all Λ particle candidates satisfying the criteria mentioned above in several transverse momentum bins with changing bin width in range from 0.1 to 6.0 GeV/c. There is no signal in transverse momentum bin from 0.0 to 0.1 GeV/c. This is caused by kinematics of the decay and the momentum acceptance of TPC. In the rest frame of Λ, its decay daughters, proton and pion, each carry momentum of around 101 MeV/c [52]. For stationary Λ daughter tracks are out of acceptance of TPC since it is able to detect only pions with momentum $p_\pi > 150$ MeV/c

[39]. For non-stationary Λ with low transverse momentum the situation gets more complicated, since one needs to boost mother particle into lab frame in order to calculate $p_{\rm T}$ of daughters, but the logic stays the same. The $p_{\rm T}$ coverage should be improved with iTPC upgrade [40].

The distribution of mass was fitted in fixed range $M_\Lambda = 1.096$ - 1.138 GeV/c^2 with double Gaussian sitting on a polynomial background. There was no physics motivation behind the employment of double Gaussian to describe the signal other than inspiration from other analyses done at STAR and reasonably successful fit of pure Λ signal simulation generated with STAR MC, i.e. Monte Carlo simulation of pure signal with detector effects done with GEANT and calibrations based on online measurements done with High Level Trigger. The polynomial background was of third order for $p_{\rm T} < 0.4$ GeV/c and of second order for $p_{\rm T} > 0.4$ GeV/c.

The broader of the two Gaussians was selected to define the 3σ mass window in which the sum of signal and background was calculated as a sum of bin content in this region. The background was estimated by integrating the polynomial function. The resulting raw signal yield was than obtained by subtracting the background counts from the total sum of bin content defined above.

In order to evaluate the effectiveness of the analysis with KFPF or to compare it with other methods, the significance of the signal was calculated as it is usually done in particle physics assuming Poisson distribution of data

$$\alpha = \frac{S}{\sqrt{S+B}}. \tag{6.1}$$

For the purpose of comparison with previous analysis done at STAR, the significance was also recalculated for the total number of events after selection equal to 1M. Employing the assumption of signal and background counts being proportional to the number of events, the projected significance can be calculated as follows

$$\alpha_1 = \sqrt{\frac{N_1}{N_0}}\alpha_0. \tag{6.2}$$

As an example, plots from the lowest two transverse momentum bins, from the mid-$p_{\rm T}$ bin and from the high-$p_{\rm T}$ bin are shown in Fig. 6.2. Note that in the first bin the signal was not successfully extracted, however there was an indication of a peak visible which further strengthened the motivation for employing machine learning techniques with ambition to reach lower $p_{\rm T}$. Fig. 6.3 presents raw yield divided by bin width and significance dependence on transverse momentum. Both quantities are recalculated per 1M events post selection.
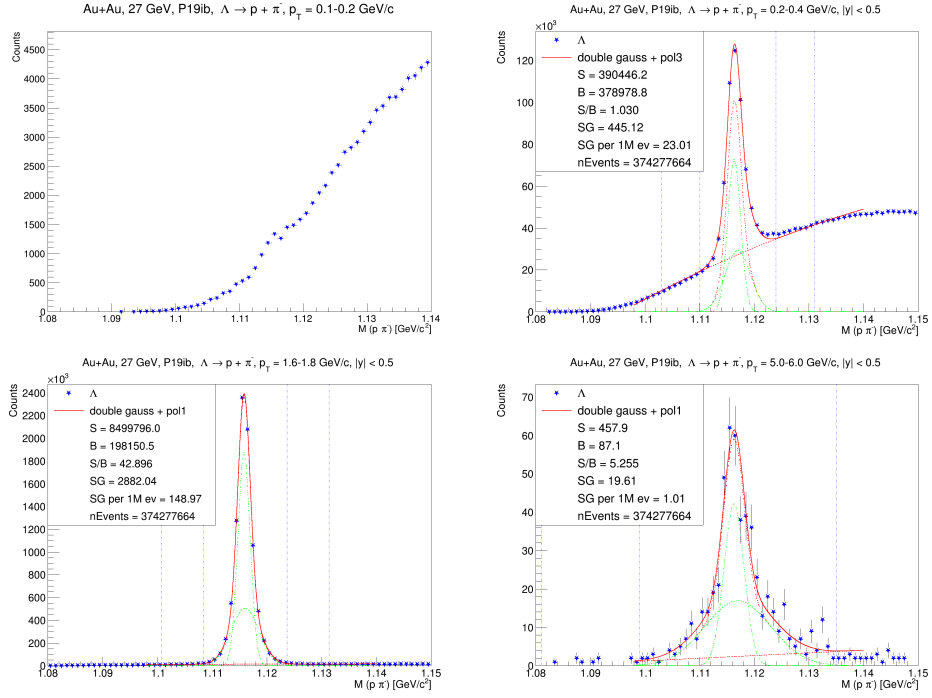
Figure 6.2: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Invariant mass spectra of Λ candidates in various transverse momentum bins.
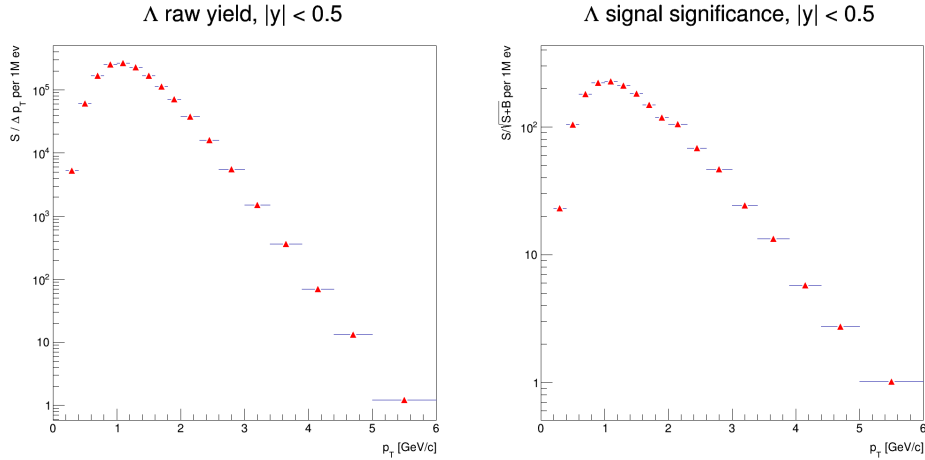
Figure 6.3: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Uncorrected Λ yield (left) and signal significance (right) dependence on transverse momentum. Both quantities are recalculated per 1M events post selection.

6.1.3 BDT training

The TMVA package offers broad selection of machine learning methods. For this particular analysis, Boosted Decison Trees were employed. The training was done using default BDT settings within TMVA, i.e. adaptive boosting, number of trees $N_{tree} = 850$ with maximum depth equal to 3, number of cuts

34

$N_{cuts} = 20$ and Gini Index for separation in nodes. Gini Index is defined by $p \cdot (1 - p)$, where p is a purity of a sample (fraction of signal events) [49]. For more detailed description of the method see Chapter 5.

To train the BDT, user has to supply a sample of signal and background. For both ntuples (data structure derived from ROOT's TTree class) of training variables were passed to TMVA. These ntuples come from analysis of particular data with KFPF. In order to obtain signal sample simulation of pure Λ signal was generated with HLT's MC simulation based on online calibrations and than processed with KFPF using looser cuts in order to give BDT some space to operate in. There were 20 Λ particles in each event generated with thermal distribution of transverse momentum. Background sample was obtained from the data. Approximately 1/5 of full statistics was processed with KFPF using the same looser cuts on KFPF variables. From obtained Λ candidates those corresponding to sideband region were selected to represent background. The width of the mass window that was cut from the background sample was decided after investigating width of the simulated signal peak. The list of looser cuts on KFPF criteria can be found in Tab. 6.2.

Cut value	Cut description
$\chi^2_{geom} < 14$	χ^2 of the track to the second daughter track
$l/dl > 3$	decay length normalized on the error
$l > 1$ cm	decay length
$\chi^2_{prim} > 3$	χ^2 of the track to primary vertex
$\chi^2_{topo} < 5$	χ^2 of the mother particle to primary vertex
$d_{max} < 1$ cm	maximum distance between daughter particles

Table 6.2: Selection criteria used in KFPF for obtaining samples for the training of Boosted Decision Trees.

Since the training was completed in each transverse momentum bin separately, following plots serve as an example from bin with $p_{\mathrm{T}} = 0.2 - 0.4$ GeV/c. The variables the BDT were trained on are presented in Tab. 6.3. Due to the energy loss of daughter particles not being described well in the MC simulation and time of flight information was not included at all, PID variables were not used for training. Rather topological variables and transverse momenta of daughters were utilized. Distributions of these quantities for both signal and background are plotted in Fig. 6.4. In Fig. 6.5 one can observe linear correlation coefficients of training variables for signal and background. The Fig. 6.6 contains the ROC curve which describes the relation between background rejection and signal efficiency (left plot). As was explained in Chapter 5 the curve shows that it is impossible to train a classifier which is able to reject increasingly more background without decreasing signal efficiency. The possible overtraining of BDT can be checked by consulting the plot of superimposed BDT value distributions for training and testing sample. These should match if the overtraining did not occur. The example

35

from p_T bin in question is presented in Fig. 6.7.

Variable code-name	Description
pt_P	p_T of daughter proton
chi2Primary_P	χ^2 of daughter proton track to PV
pt_Pi	p_T of daughter pion
chi2Primary_Pi	χ^2 of daughter pion track to PV
Chi2NDF	χ^2 of daughter track to another
LdL	normalized decay length
Chi2Topo	χ^2 of mother particle to PV

Table 6.3: Training variables used for classification with BDT.
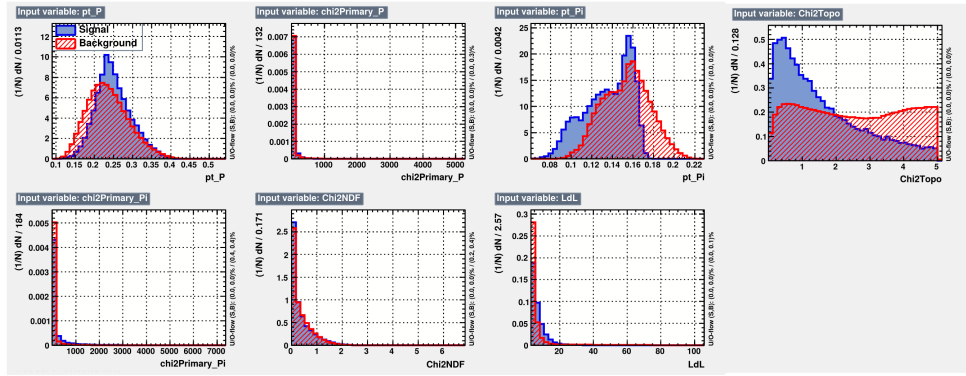
Figure 6.4: *THIS THESIS:* Distributions of training variables from training sample.
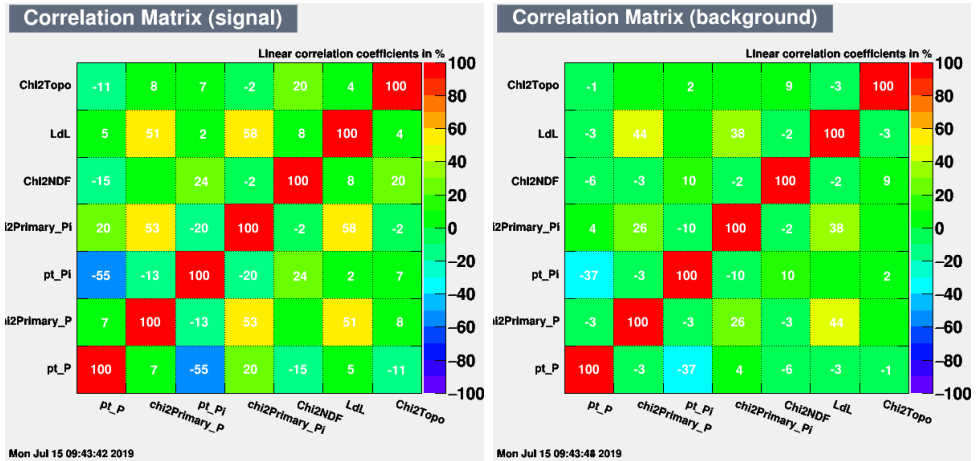
Figure 6.5: *THIS THESIS:* Linear correlation coefficients of variables used in training of BDT for signal (left) and background (right).

Based on the training TMVA is able to estimate the dependency of signal significance on the BDT value cut. However, the optimal cut on BDT response
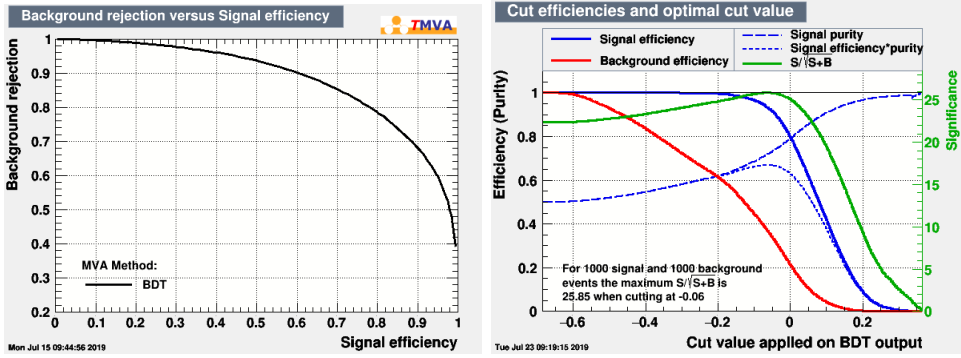
Figure 6.6: *THIS THESIS:* ROC curve characterizing the relation between background rejection and signal efficiency of BDT (left), cut efficiencies and optimal BDT response value cut as calculated by TMVA (right).
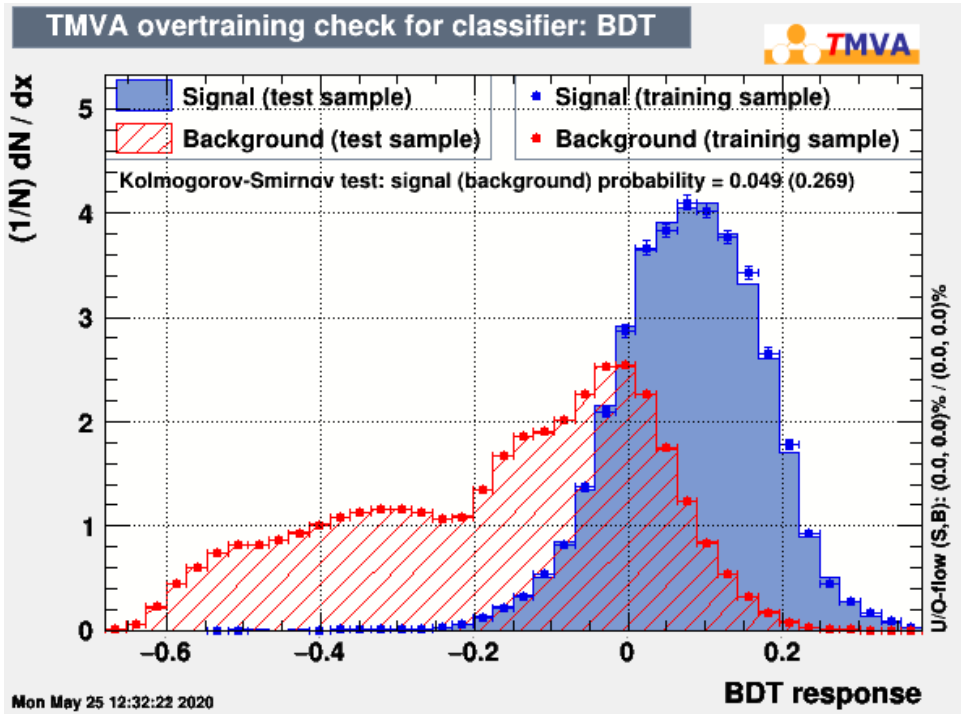
Figure 6.7: *THIS THESIS:* Superimposed distributions of BDT response value for training and testing sample used to check for overtraining. Plot comes from BDT training for $p_{\mathrm{T}} = 0.2 - 0.4$ GeV/c.

value (maximizing signal significance) depends on actual number of signal and background events. In the plot on the right in Fig. 6.6 the optimal cut value is determined with assumption of number of signal and background events being equal to 1000 (by default). Since the ratio of signal to background events supplied to TMVA is usually artificial, one has a option to estimate the number of signal and background events in the data, for example in the same way the raw signal and background were extracted in the previous section. The other option is to apply different values of BDT

cut on the data directly, investigate the dependency of significance on the cut value and find the value which maximizes the signal significance.

At first, the second approach was selected for this analysis and the BDT were applied on a fraction of 3% of the full statistics to scan the significance. In this way the optimal cut value was found which was later fixed and applied on the whole dataset. However, as one can observe in the plot of the significance scan in Fig. 6.8 on the left, this approach might be prone to fluctuations in the data and thus the first approach was adopted in the end. The blue line in the Fig. 6.8 (left) highlights where the final cut was made using the TMVA calculation based on estimation of signal and background events from data. The right hand side plot in Fig. 6.8 shows dependency of signal to background ratio on BDT response value. It demonstrates well that with further increase of BDT value threshold after reaching maximum significance one can still notably improve the purity of the signal. Final values of BDT cuts are listed in Tab. 6.4 .

p_{T} [GeV/c]	0.1-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0	1.0-1.2
BDT cut	-0.0690	-0.0471	-0.1130	-0.1216	-0.1426	-0.1732
p_{T} [GeV/c]	1.2-1.4	1.4-1.6	1.6-1.8	1.8-2.0	2.0-2.3	2.3-2.6
BDT cut	-0.1841	-0.1877	-0.2062	-0.1913	-0.1934	-0.1838
p_{T} [GeV/c]	2.6-3.0	3.0-3.4	3.4-3.9	3.9-4.4	4.4-5.0	5.0-6.0
BDT cut	-0.1697	-0.1970	-0.1602	-0.0565	0.0079	0.1519

Table 6.4: Final BDT cuts used in this work. Values are based on TMVA's estimation of dependence of significance on BDT threshold. Number of signal and background candidates needed for this estimation are obtained by fitting the invariant mass histogram in analysis without BDT.
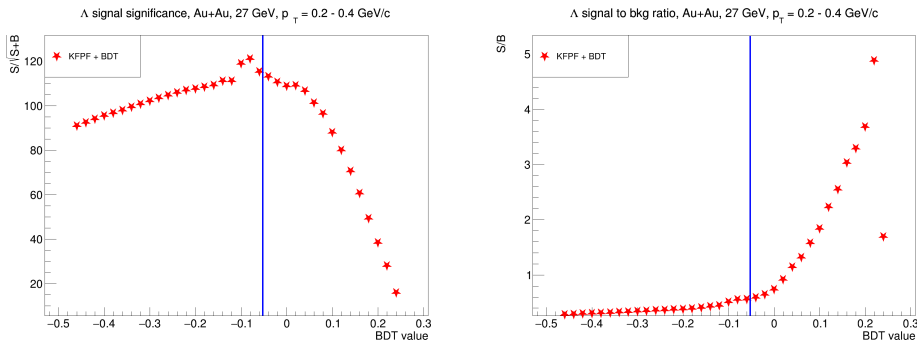
Figure 6.8: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Dependence of signal significance on BDT cut (left), signal to background ratio as a function of BDT cut (right) for transverse momentum bin $p_{\mathrm{T}} = 0.2-0.4$ GeV/c. Blue lines highlight were the final cut was made after obtaining the cut value from TMVA as described in the text.

6.1.4 BDT application

With the cut values found by means explored in previous section BDT were applied directly in KF Particle Finder to process the whole dataset. The invariant mass histograms were fitted in the same manner as before and also the raw yields and signal significance were calculated in the same way. By employing BDT it was now possible to reconstruct signal of Λ in region $p_{\mathrm{T}} = 0.1 - 0.2$ GeV/c as can be seen in Fig. 6.9. In Fig. 6.10 significance obtained with default KFPF cuts is compared to the one achieved with BDT. In general, there is an enhancement of up to 50% at low transverse momentum.
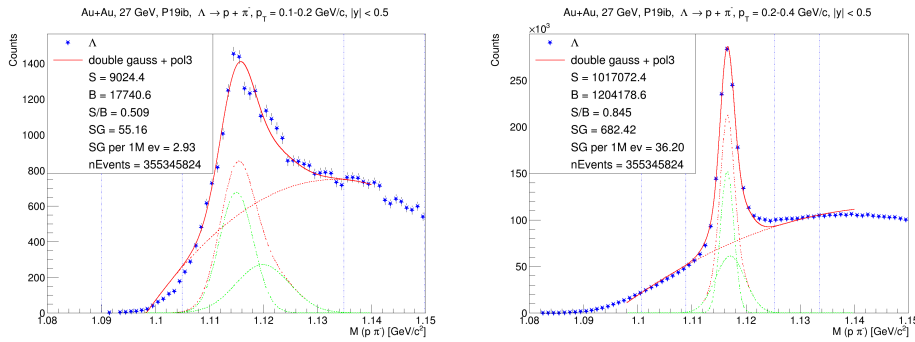
Figure 6.9: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Invariant mass of Λ candidates with $p_{\mathrm{T}} = 0.1 - 0.2$ GeV/c and $p_{\mathrm{T}} = 0.2 - 0.4$ GeV/c obtained with KFPF and BDT.

6.1.5 Comparison with conventional analysis

In this section results obtained with KFPF and BDT will be compared to conventional analysis which employs topological cuts to reject combinatorial background. For this purpose study of production of strange particles in Au+Au collisions at 27 GeV done by STAR during phase one of Beam Energy Scan was selected [53]. In order to estimate the signal significance obtained with conventional means described in the paper yield of Λ particles was divided by statistical error for each p_T bin. Note that no centrality division of collisions was done in analysis presented in this thesis and the significance in the paper was calculated by using yields from most central collisions only (0-5%). For this reason, the comparison has to be taken with caution. However, the crucial conclusion is that conventional analysis does not reach under $p_{\mathrm{T}} = 0.4$ GeV/c and so it seems probable that even without BDT KF Particle Finder enhances significance in low transverse momentum region. The resulting comparison is displayed in the Fig. 6.11.
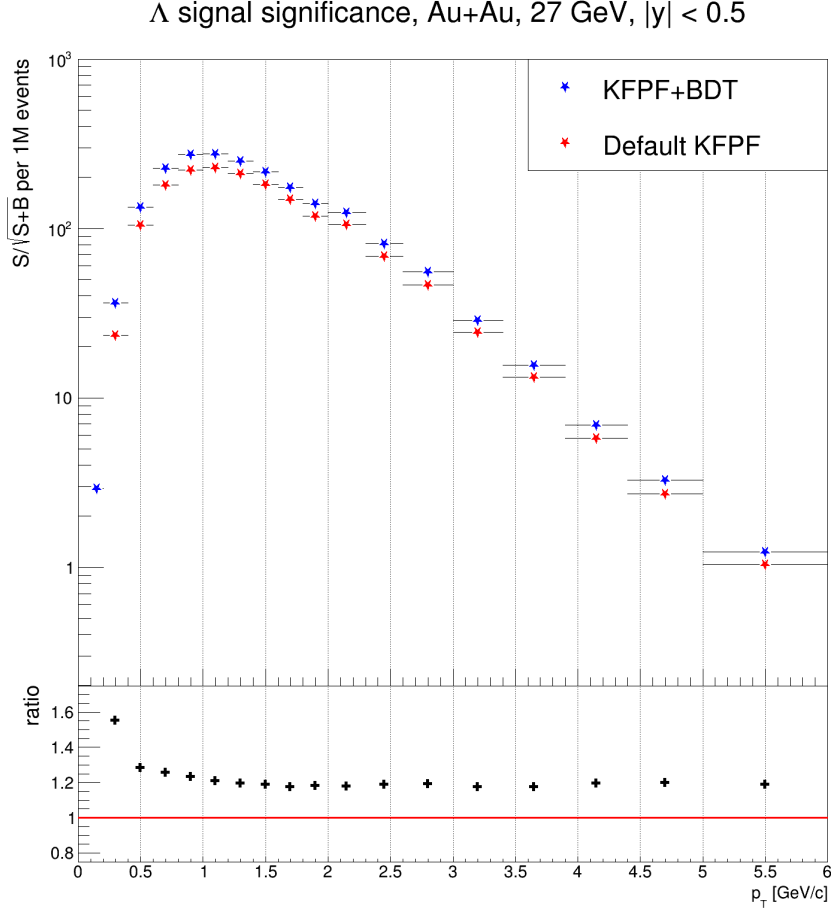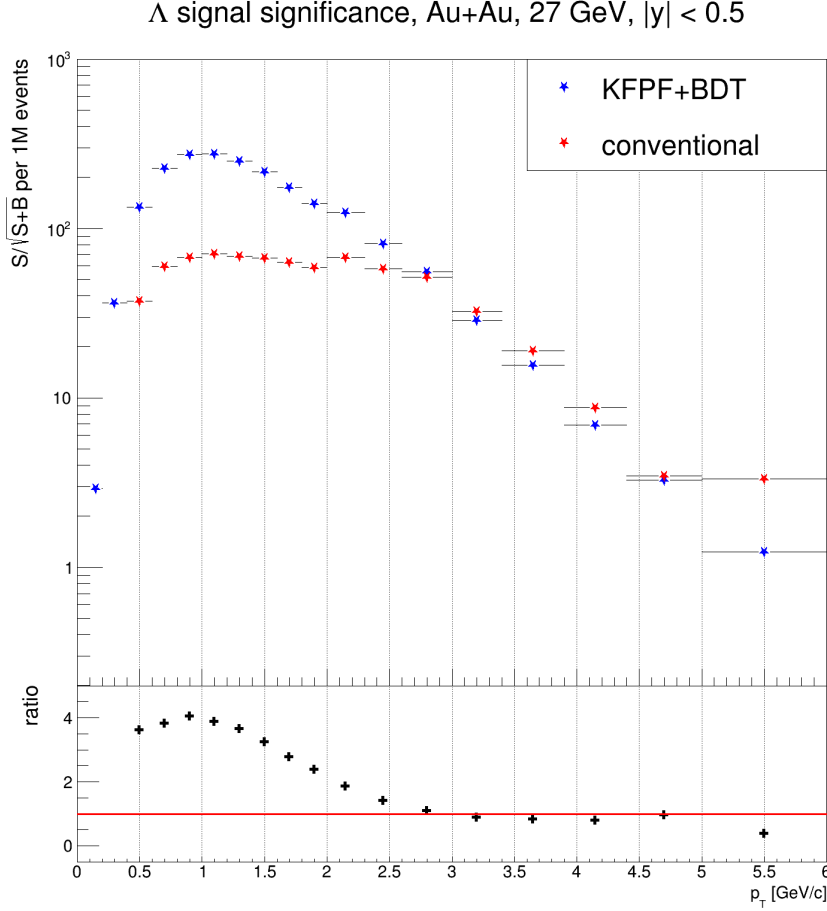
Figure 6.10: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Comparison of significance obtained with application of BDT to the significance achieved with default KFPF cuts.

## 6.2 Efficiencies of Λ reconstruction in 27 GeV Au+Au collisions

The possibility of obtaining fast physics results by means of analysis of express production data at STAR with KF Particle Finder is promising. However, in order to reach any meaningful conclusions we must understand the efficiency and systematic uncertainties of such analyses. That is why Monte Carlo simulation of studied decays is needed including detector effects. For this purpose the working group responsible for express production at STAR currently has a pure signal MC simulation at its disposal. The detector effects are handled by GEANT and all calibrations are based on online analyses of measured collisions during data taking. Therefore, the simulation is available almost instantly. In this section an attempt to calculate efficiencies and corrected spectra of strange baryons reconstructed with KF Particle Finder using simulation based on online calibrations will be presented and possible shortcomings

Figure 6.11: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV: Comparison of significance obtained with application of BDT to the significance achieved with conventional analysis.

will be discussed. Note that complete official embedding of MC simulation into data was not available at STAR during writing of this work.

The efficiency of Λ reconstruction in Au+Au collisions at 27 GeV was studied using express production based MC simulation. Two different samples of pure Λ signal were generated. The first one was generated with flat transverse momentum distribution and 20 Λ particles per event. In the second one there are 20 Λ baryons in each event as well and they follow thermal transverse momentum distribution. Primary vertices of simulated collision were generated following distribution based on real data and fixed uncertainty was assigned to them as discussed later on in this section. For both samples particles were generated within rapidity range $|y| < 1.6$. The p_{T} and y distributions of MC particles are shown in Fig. 6.12 and Fig. 6.13.

Since KF Particle Finder exploits information about uncertainties it is desirable to simulate properly also errors of tracks and vertices. Specifically
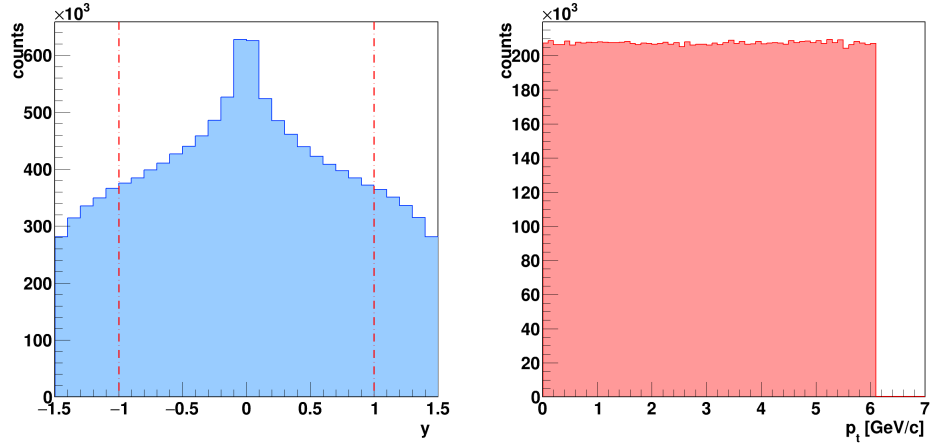
Figure 6.12: *THIS THESIS:* Distributions of rapidity y and transverse momentum p_T for MC simulation sample with flat p_T.
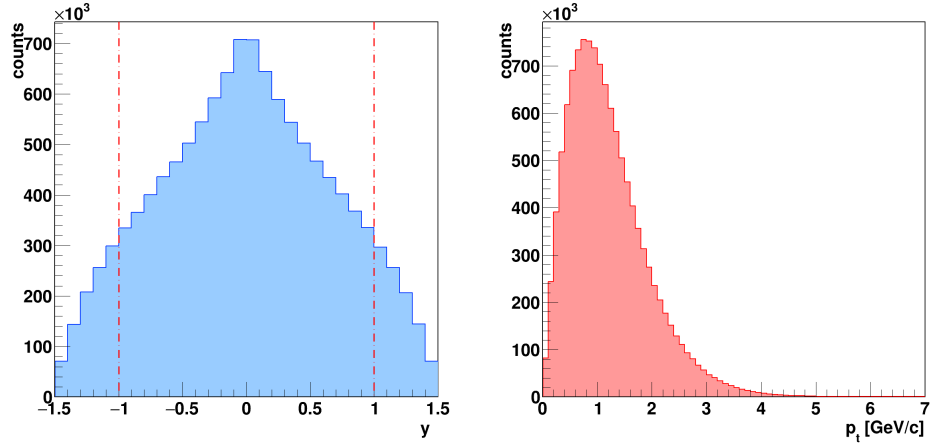
Figure 6.13: *THIS THESIS:* Distributions of rapidity y and transverse momentum p_T for MC simulation sample with thermal p_T.

the handling of primary vertex within MC should be discussed as it simulates only signal particles and therefore the primary vertex is not obtained as it would be in data, i.e. from tracks. Although the position of primary vertex in MC is not fitted from tracks, it follows distributions obtained from data and has fixed uncertainty assigned to it ($\delta r = 200 \ \mu$m).

In order to investigate how well the MC simulation actually describes data in terms of statistical criteria and other variables defined within KF Particle Finder, distributions of these variables were compared for Monte Carlo and for signal candidates in real data within particular p_T bin. Since in the data primary vertex resolution depends on multiplicity of the collision and thus its centrality it was desirable to attempt a rough division of collisions into centrality classes before comparing to the Monte Carlo. The division was

42

based on cuts on reference multiplicity (refMult) of each event. Cut values were inspired by those used for isobar collisions at 200 GeV since it was argued within the STAR collaboration that the refMult distribution of this system is conveniently similar to the Au+Au collisions at 27 GeV.

The distribution of reference multiplicity (refMult) can be seen in Fig. 6.14. Distributions for both MB events and events after selection cuts described previously are shown. Red lines in the plot demonstrate cut values that define centrality classes. For the purpose of comparison of MC simulation to data events were only divided into "central" (0-5%), for which $19 < \text{refMult} \leq 67$, and "peripheral" (50-80%), for which $267 < \text{refMult}$.
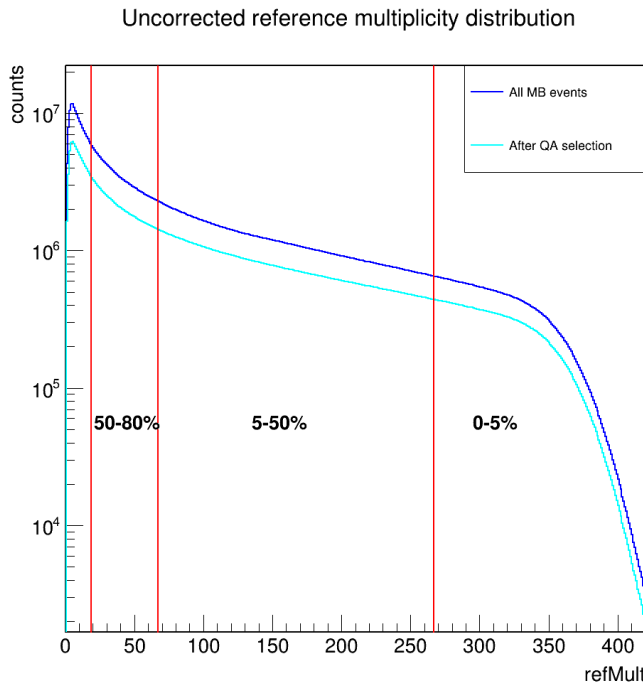
Figure 6.14: *THIS THESIS:* Uncorrected reference multiplicity distribution for 27 GeV Au+Au collisions. Red lines highlight cut values which define centrality classes used in this work.

The signal distributions of studied quantities in data were obtained by subsequent procedure. First, rapidity cut was done on Λ candidates so that $|y| < 1.0$. Second, distribution for Λ candidates in 3σ region under the signal peak was extracted. Third, distribution for candidates in the sideband region was extracted, this distribution was meant to represent background. Next, the signal distribution was obtained by subtracting background distribution from the distribution which represented all candidates under the peak while simultaneously scaling both distributions with background and sum of background and yield respectively. Finally, the MC distribution was scaled so its integral is equal to the one of signal distribution coming from data.

Example of a resulting comparison for $p_\mathrm{T} = 1.8 - 2.0$ GeV/c and centrality 50-80 % is shown in Fig. 6.15 and Fig. 6.16. Distributions for most central collision (0-5%) and the same p_T range are presented in Fig. 6.17 and Fig. 6.18.

The Fig. 6.15 demonstrates that energy loss through ionization is not described precisely. More importantly, however, the topological variables are generally described well as can be seen in Fig. 6.16, although there are some discrepancies. The χ^2_prim variables for protons and pions, which characterize the probability of particles coming from primary vertex, do not match for low values, which is the region where the optimal cut value is searched for. This may be connected to the manner in which the primary vertex error is handled in the Monte Carlo and could be potential problem for calculation of efficiency of KF Particle Finder with this simulation. Nevertheless, the calculation of efficiencies was carried out in a way described further. By comparing Fig. 6.16 to Fig. 6.18 one can find that χ^2_prim for simulated pions matches real data better in peripheral collisions. This could be due to the fact that in central collisions, i.e. high multiplicity events, the primary vertex resolution is actually better than the MC simulation assumes by assigning the vertex a fixed uncertainty.
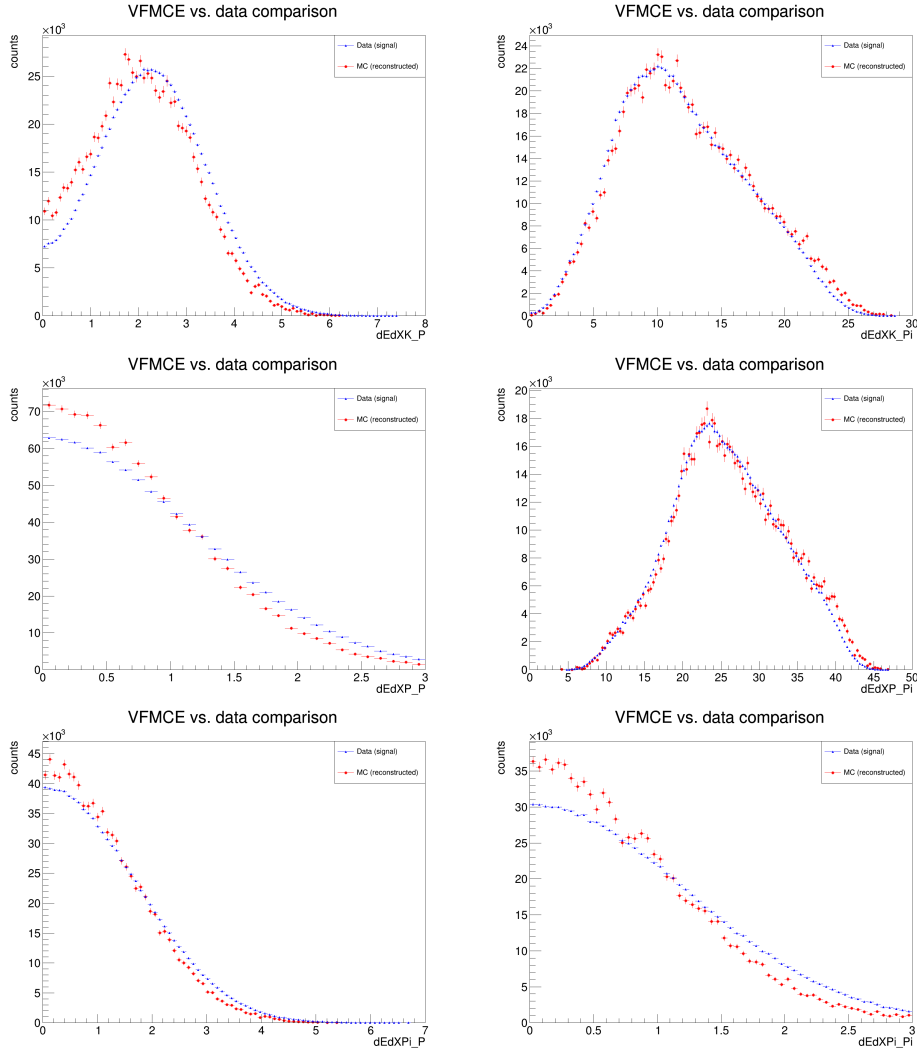
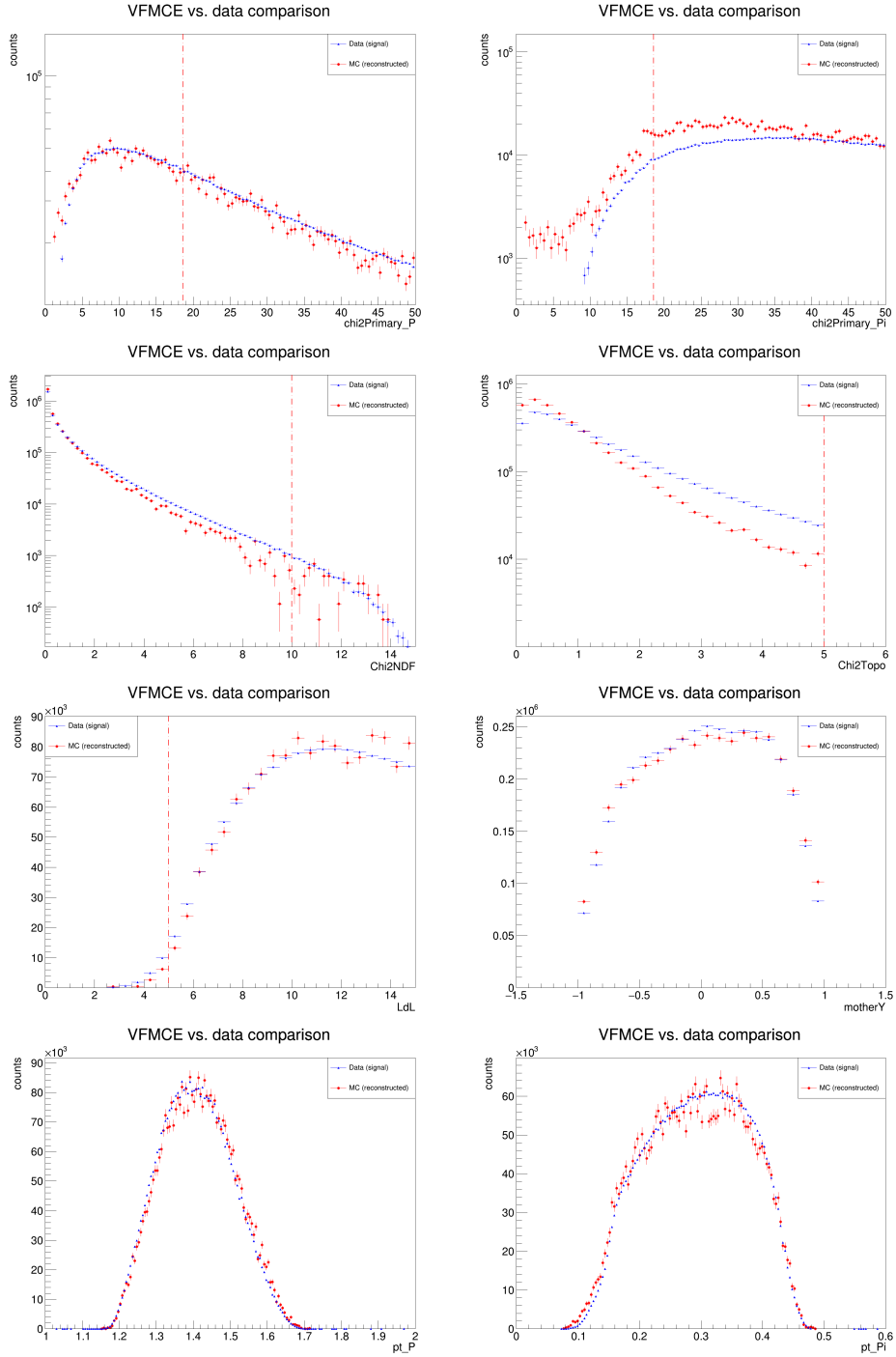**Figure 6.15:** *THIS THESIS:* Comparison of distributions of KF Particle Finder's energy loss variables for data and MC simulation, peripheral collisions (50-80%), $p_{\mathrm{T}} = 1.8 - 2.0$ GeV/c. The dEdXKP variable represents how far were particles identified as protons from energy loss hypothesis for kaons in terms of Gaussian sigma, all other variables are defined analogically.
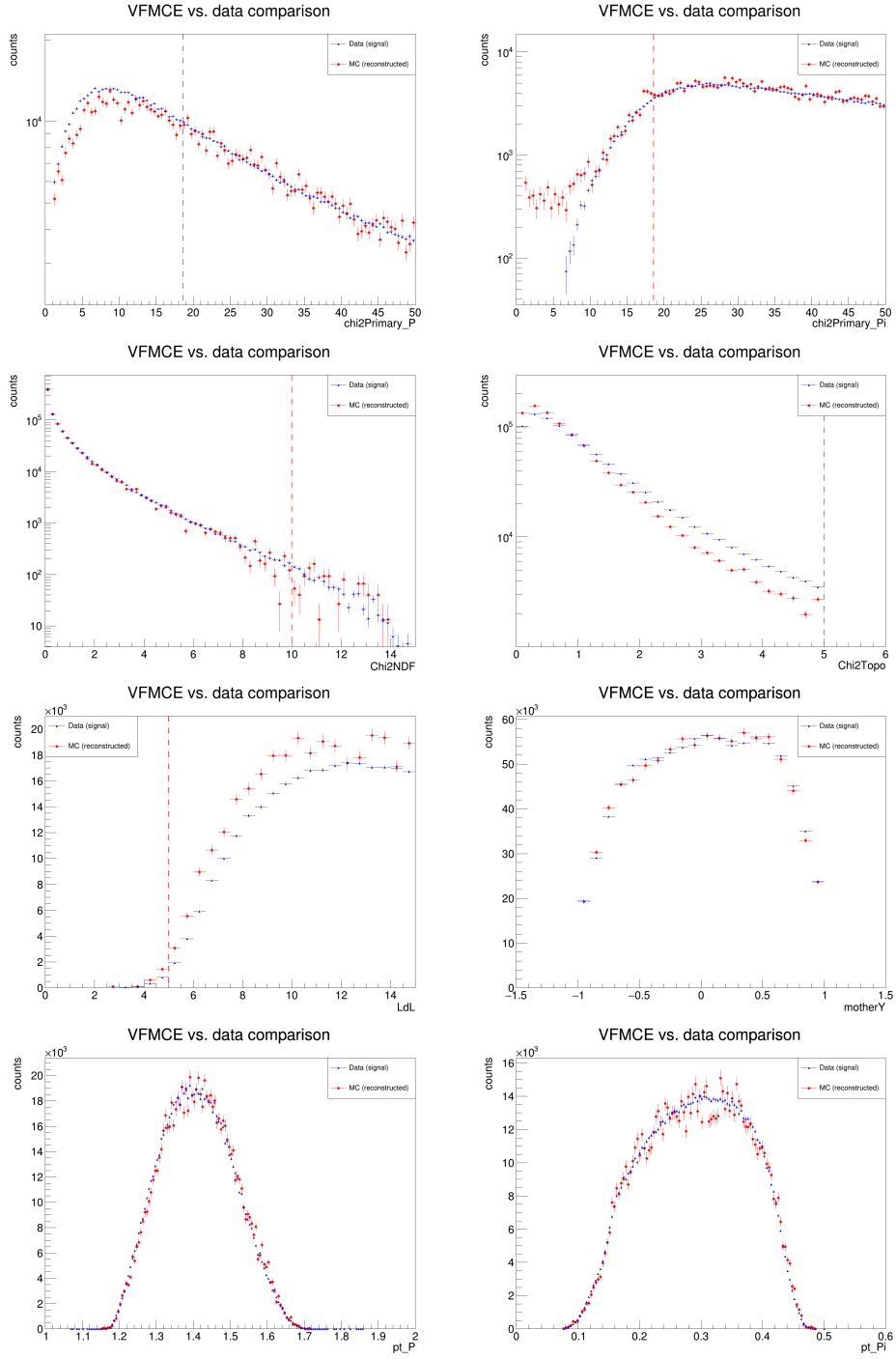
Figure 6.16: *THIS THESIS:* Comparison of distributions of KF Particle Finder's variables for data and MC simulation, peripheral collisions (50-80%), $p_T = 1.8 - 2.0$ GeV/c. Red lines highlight where the cut is made within KFPF by default.
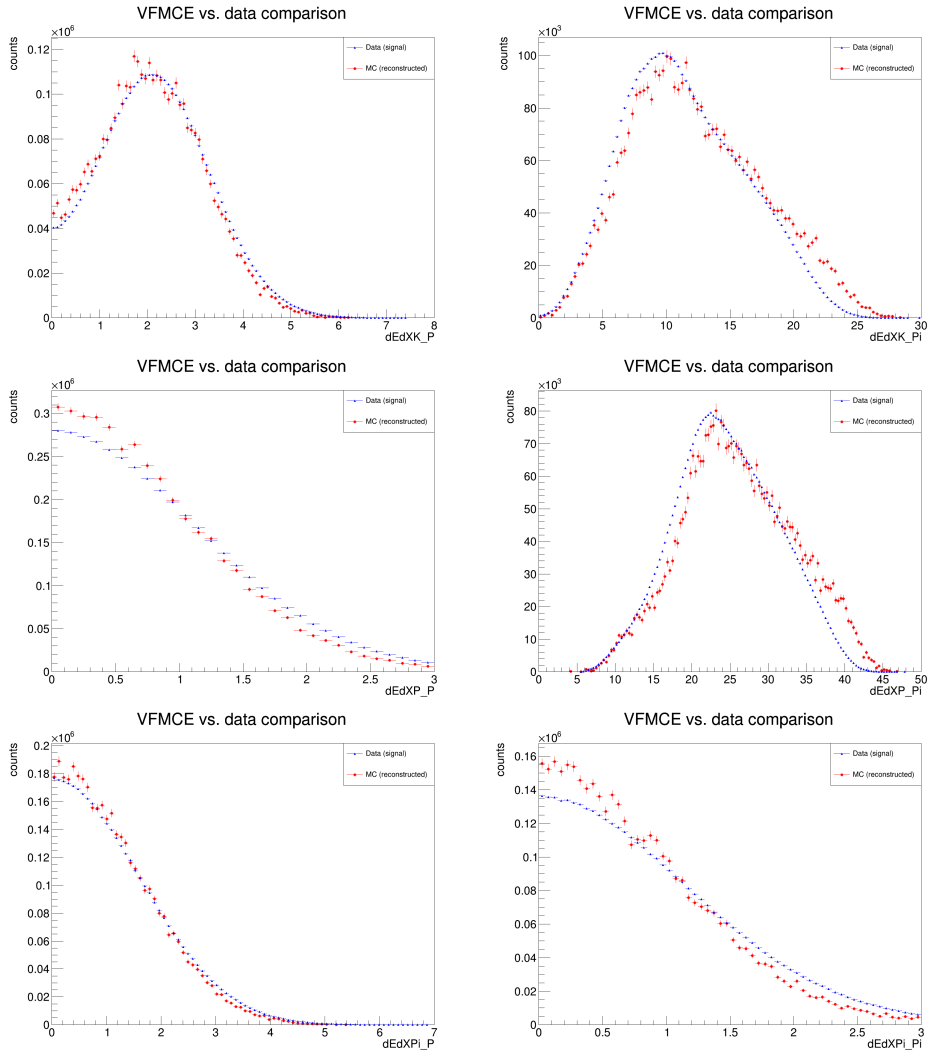
Figure 6.17: *THIS THESIS:* Comparison of distributions of KF Particle Finder's energy loss variables for data and MC simulation, central collisions (0-5%), $p_{\mathrm{T}} = 1.8 - 2.0$ GeV/c.

Figure 6.18: *THIS THESIS:* Comparison of distributions of KF Particle Finder's energy loss variables for data and MC simulation, central collisions (0-5%), $p_\mathrm{T} = 1.8 - 2.0$ GeV/c. Red lines highlight where the cut is made within KFPF by default.

In order to obtain efficiencies of reconstruction, the simulation of Λ with flat and thermal spectrum was processed with both KF Particle Finder and with KF Particle Finder combined with Boosted Decision Trees as it would be done with real data. There are means of handling the MC particles implemented in KF Particle Finder and therefore one is able to produce histograms for true MC particles and for reconstructed candidates which have a match to MC particle separately. Before the calculation of efficiencies cut on rapidity of both Monte Carlo and reconstructed particles was imposed in order to calculate the efficiencies in well defined acceptance - Λ particles with $|y| < 0.5$ were selected. This value was motivated by attempt to compare final spectra with STAR's paper on strangeness [53]. Note that the cut changed the p_T spectra of both MC samples as is demonstrated in Fig. 6.19. The efficiency of reconstruction was calculated for each p_T bin by dividing the yield of matched reconstructed particles by number of MC particles in particular bin. The yield of reconstructed MC particles was extracted analogically to the case of real data. However, the invariant mass spectrum was fitted only with double Gaussian as no background was expected in the simulation of pure signal. The resulting efficiencies of reconstruction are plotted in Fig. 6.20.
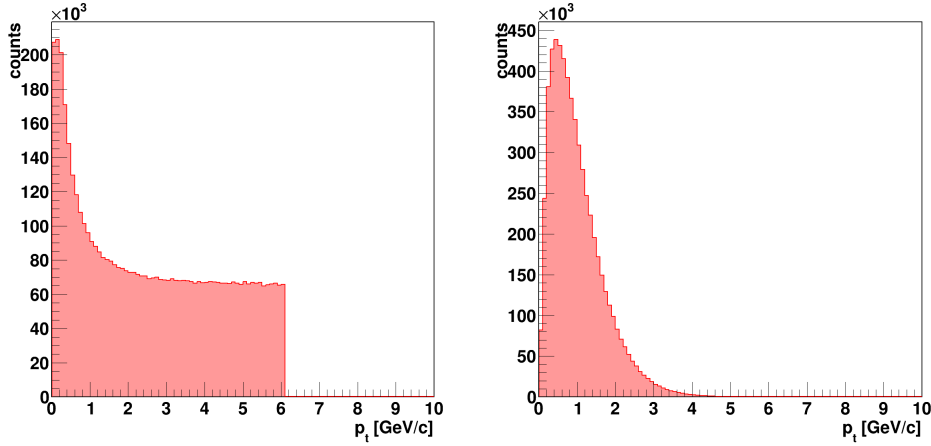
Figure 6.19: *THIS THESIS:* Distributions of transverse momentum p_T for MC simulation sample with thermal and flat p_T after cut on rapidity of Λ particles $|y| < 0.5$.

With the efficiencies obtained as described above, raw yields of Λ extracted with KFPF and with KFPF in combination with BDT were corrected. After that invariant spectra were calculated for the purpose of comparison with the paper defined as

$$\frac{\mathrm{d}^2 N}{2\pi N_{evt} p_T \mathrm{d}p_T \mathrm{d}y},\qquad(6.3)$$

where N_{evt} is a number of events after event selection, p_T is a center of particular p_T bin and $\mathrm{d}p_T$, $\mathrm{d}y$ are transverse momentum and rapidity

49

Λ reconstruction efficiency
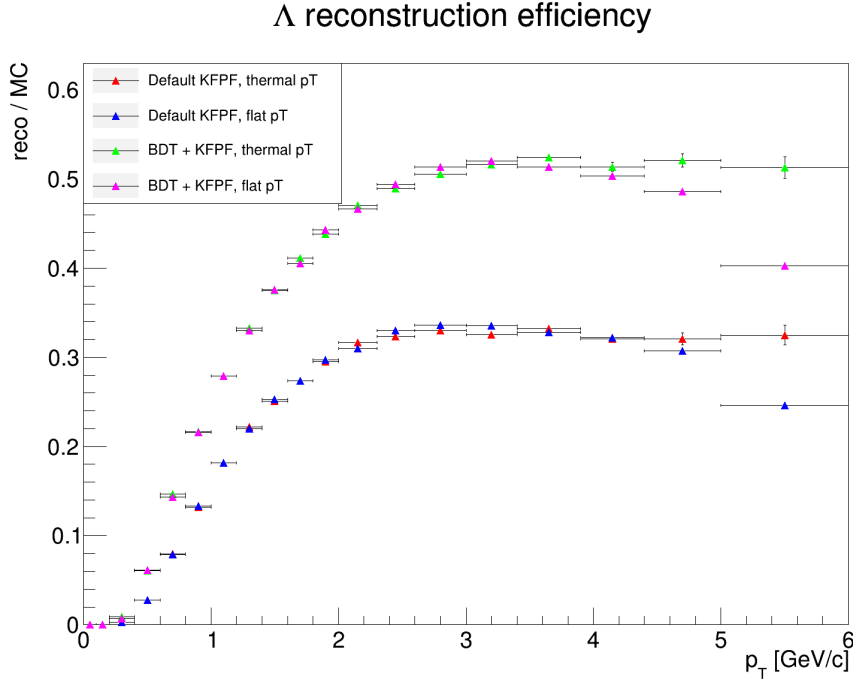
**Figure 6.20:** *THIS THESIS:* Efficiencies of Λ reconstruction with KF Particle Finder and with KF Particle Finder combined with BDT based on MC simulations generated with flat and thermal p_T distributions.

bin widths respectively. The p_T value in 6.3 should be the mean p_T within the particular bin. In this analysis, however, center of bin was used as a approximation while assuming that bin widths are narrow enough. The resulting invariant spectra are displayed in Fig. 6.21. As can be seen the spectra of Λ do not match well at low p_T and the difference is too significant to accept it as a systematic uncertainty which emerged by using two different p_T distributions for efficiency calculation. Therefore, further study of both reconstruction efficiency and the MC simulation is necessary. Note that the point for $p_T = 0.1 - 0.2$ GeV/c is there only for efficiencies based on MC simulation with thermal distribution as there was not enough statistics to reconstruct Λ also from MC with flat p_T. Also the systematic error of this point was not computed and since the fit of invariant mass for this p_T bin was rather crude it could be considerably large.

For the purpose of comparison spectra in paper [53] were summed over all centrality classes as no centrality study was done for 27 GeV Au+Au collisions in this thesis. More specifically the result was given by weighted average over all centrality classes where for each class the weights corresponded to fractions of total event count. Finally, before the comparison a cut on refMult was imposed on results from this thesis in order to roughly select collisions within 0-80% centrality range. This had to be done since the least central collisions were discarded in the paper as well and by doing so one reduces

the number of events significantly while keeping the majority of Λ candidates and thus shifts the invariant spectra substantially. The final comparison is shown in Fig. 6.22. The ratio presented in the plot is calculated for spectra from paper and spectra obtained with KFPF+BDT and corrected with efficiency based on MC simulation with thermal p_{T} distribution. One can observe that with increasing p_{T} spectra increasingly deviate. This could be connected to the primary vertex issue discussed above as well.
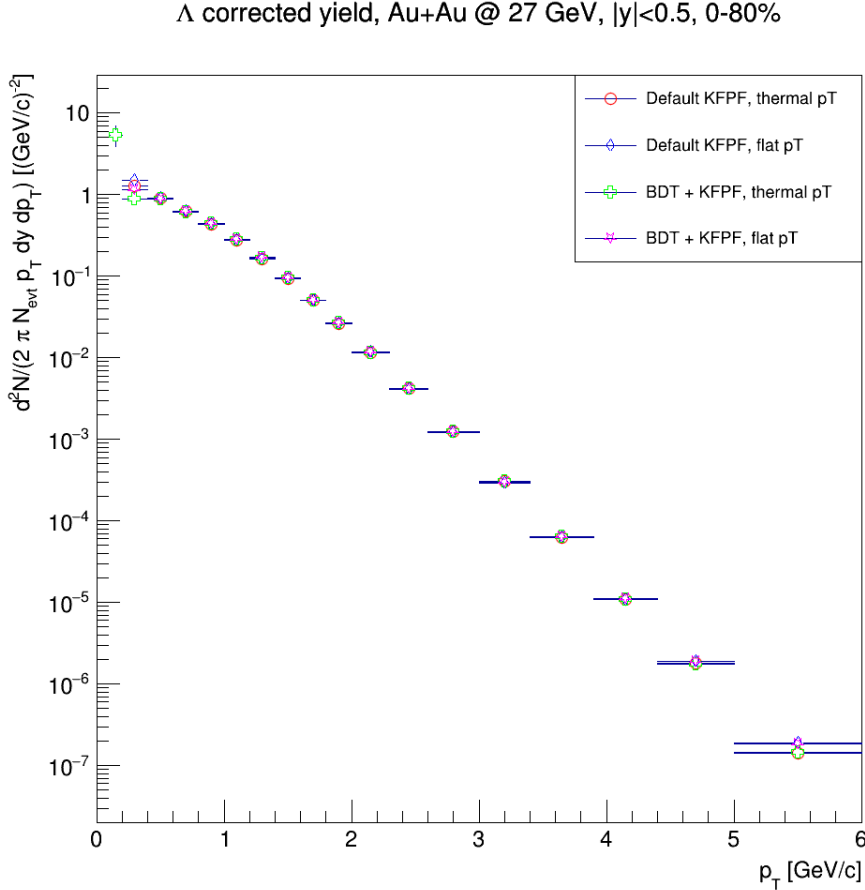
Figure 6.21: *THIS THESIS:* Corrected invariant spectra of Λ in $\sqrt{s_{NN}} = 27$ GeV Au+Au collisions obtained with KF Particled Finder and KF Particle Finder combined with BDT with corrections based on MC simulations generated with various p_{T} distributions.
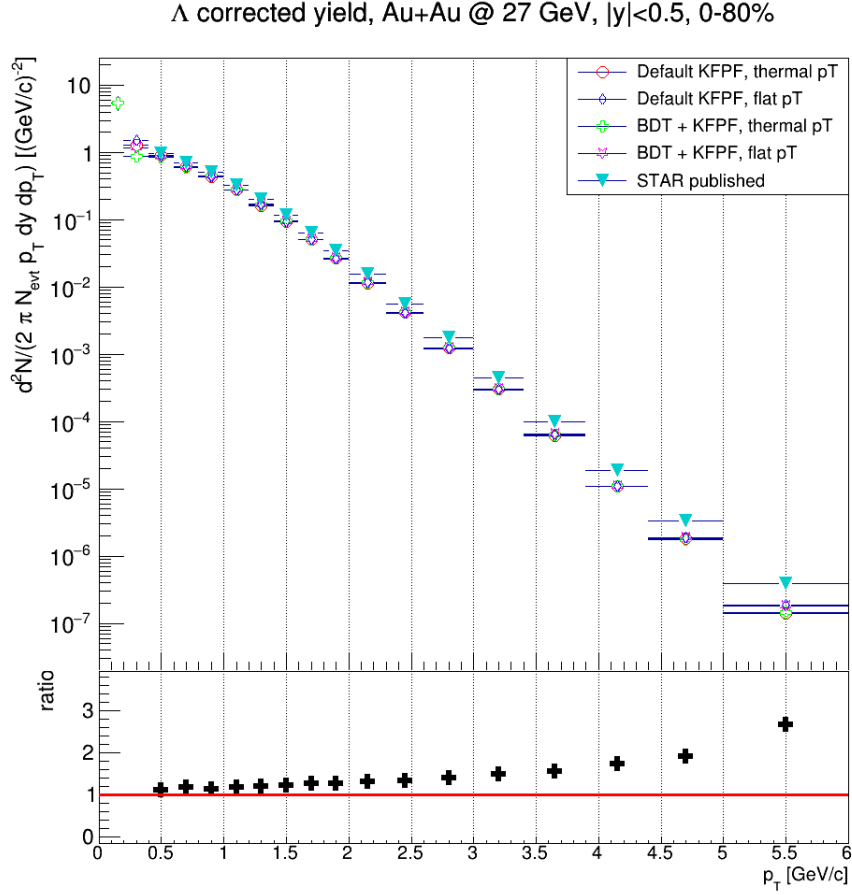
Figure 6.22: *THIS THESIS:* Comparison of corrected invariant spectra of Λ in $\sqrt{s_{NN}} = 27$ GeV Au+Au collisions from STAR's paper on strange hadrons production [53] with spectra obtained with KF Particled Finder and KF Particle Fineder combined with BDT and corrections based on MC simulation. The ratio is calculated for spectra from paper and spectra obtained with KFPF+BDT and corrected with efficiency based on MC simulation with thermal p_T distribution.

6.3 Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV

The data from Au+Au collisions at center of mass energy $\sqrt{s_{NN}} = 14.6$ GeV were taken at STAR experiment during 2019 which as a part of second phase of the Beam Energy Scan program. In the time of writing this work, the data were available only through the express production of BES-II data and thus this part of thesis is dedicated to research of viability of express production data for fast physics analyses. Note that iTPC upgrade was implemented at STAR prior to the 14.6 GeV run. Therefore, an improvement in signal extraction was expected, especially for low p_T since the inner sectors of Time Projection Chamber were completely renewed (as illustrated by Fig. 6.23) and the p_T tracking was enhanced.
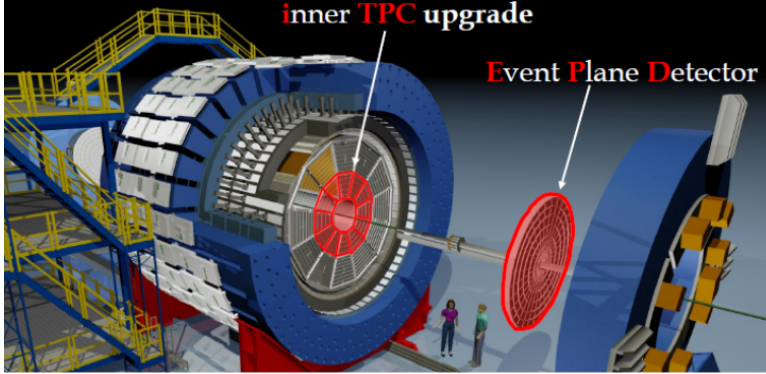
Figure 6.23: The inner Time Projection Chamber upgrade completely renews inner sector and improves low p_T tracking. Taken from [54].

6.3.1 Dataset and event selection

The express production of the data was done with TFG19 library by TFG (Tracking Focus Group) at STAR. For this dataset results from Λ baryon are presented similarly to the previous one. Besides that, results from reconstruction of antibaryon $\bar{\Lambda}$ and also result from analysis of cascading decays of multistrange baryons Ω^-, Ξ^- and their antiparticles are shown.

The event selection was executed in the same way as in the 27 GeV case, i.e. selected events had at least 10% tracks classified as primary and satisfied conditions $|V_z| < 70$ cm and $\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2} \leq 0.45$ cm. The total number of sampled minimum bias events counted 210M, from those 194M were selected for particle reconstruction.

6.3.2 Analysis with default KFPF cuts

In the same way as with the 27 GeV data default KFPF cuts were applied for the 14.6 GeV collisions. The cut values for reconstruction of cascades are listed in Tab. 6.5. The rest of the reconstruction, including extraction of signal and calculation of significance, was done analogically to the analysis of 27 GeV. For extracting the signal of multistrange baryons single Gaussian was used and the background was fitted with polynomial of first order.

As an example for Λ baryon, there are invariant mass histograms from two lowest transverse momentum bins plotted in the Fig. 6.24. The left plot demonstrates that it is possible to reach practically zero transverse momentum even without inclusion of any machine learning methods. Reaching lower p_T is important for any future improvements of Λ total cross section. Fig. 6.30 presents the resulting uncorrected yields for all transverse momentum bins together with signal significance per 1M events.

For other particles examples of invariant mass spectra in various transverse momentum bins as well as raw yields and signal significance are presented throughout Fig. 6.24 to Fig. 6.35.

Cut value	Cut description
$\chi^2_{geom} < 6$	χ^2 of the track to the second daughter track
$l/dl > 10$	decay length normalized on the error
$l > 5$ cm	decay length
$\chi^2_{prim} > 18.6$	χ^2 of the track to primary vertex
$\chi^2_{topo} < 5$	χ^2 of the mother particle to primary vertex
$d_{max} < 1$ cm	maximum distance between daughter particles

Table 6.5: *THIS THESIS:* Selection criteria used in KFPF for selection of reconstructed cascades Ω and Ξ.
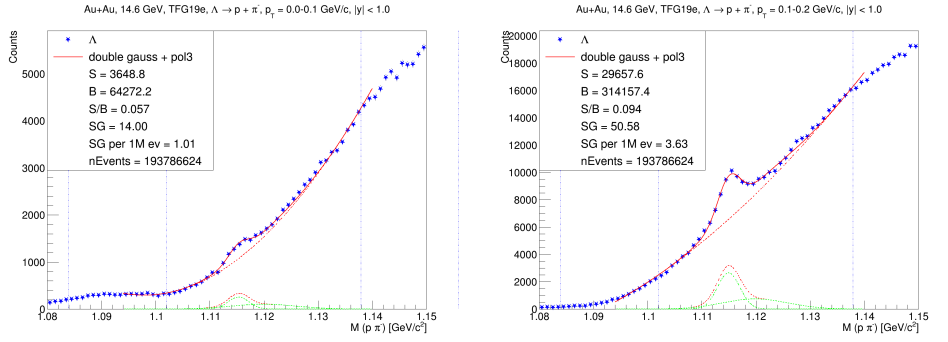
Figure 6.24: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for Λ with $p_T = 0.0-0.1$ GeV/c (left) and $p_T = 0.1 - 0.2$ GeV/c (right).
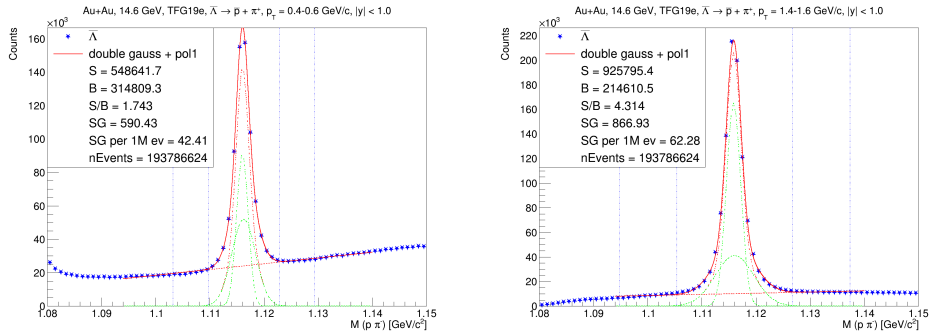
Figure 6.25: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for $\bar{\Lambda}$ with $p_T = 0.4-0.6$ GeV/c (left) and $p_T = 1.4 - 1.6$ GeV/c (right).
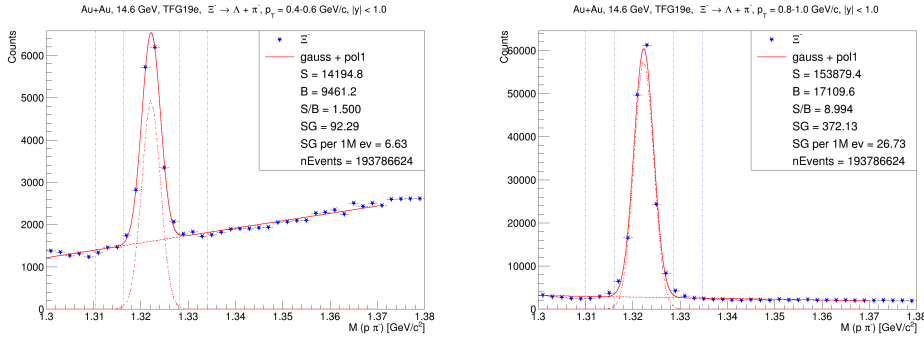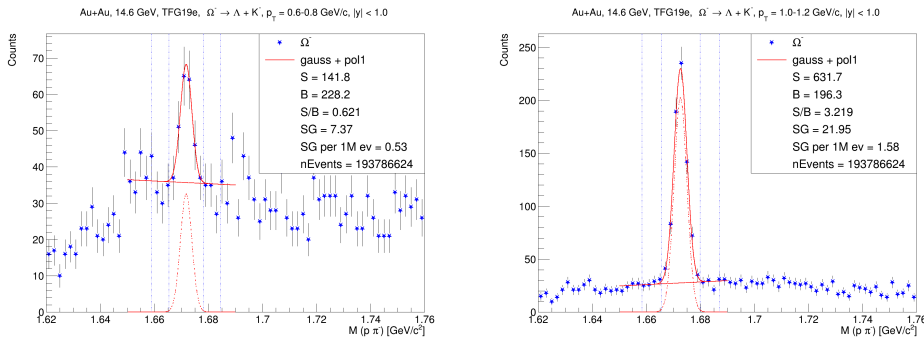
Figure 6.26: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for Ξ^- with $p_T = 0.4 - 0.6$ GeV/c (left) and $p_T = 0.8 - 1.0$ GeV/c (right).
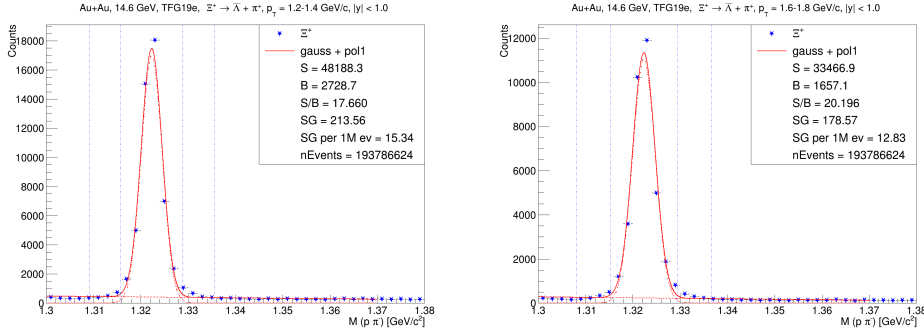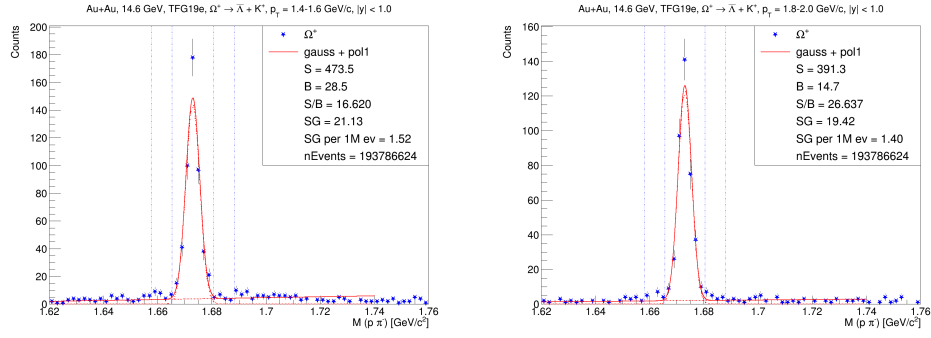

Figure 6.27: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for Ξ^+ with $p_T = 1.2 - 1.4$ GeV/c (left) and $p_T = 1.6 - 1.8$ GeV/c (right).
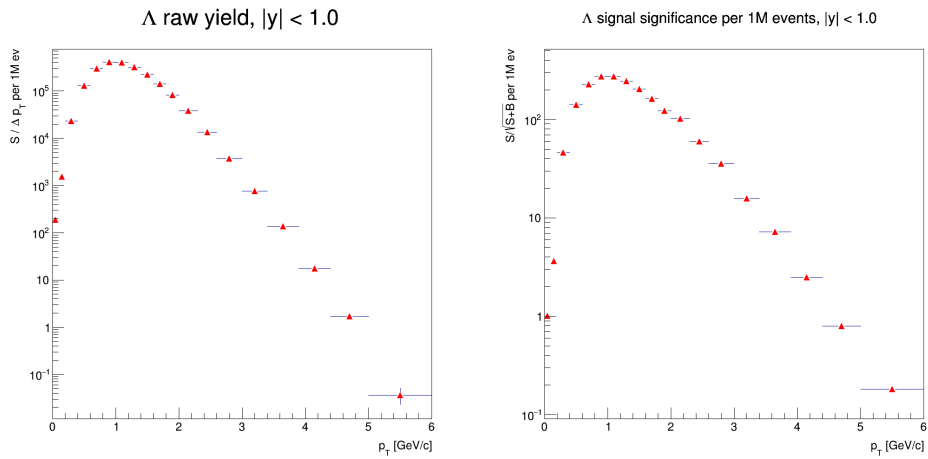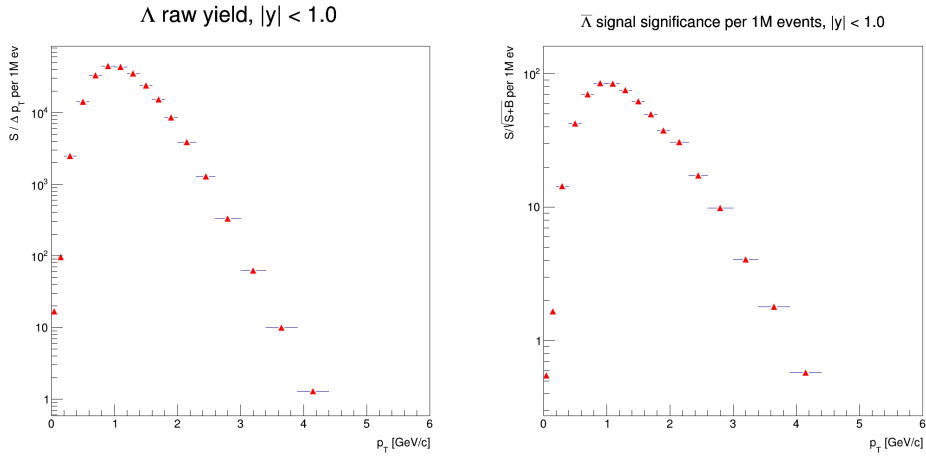

Figure 6.28: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for Ω^- with $p_T = 0.6 - 0.8$ GeV/c (left) and $p_T = 1.0 - 1.2$ GeV/c (right).

55

Figure 6.29: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Invariant mass histograms for Particle with $p_{\mathrm{T}} = 1.4 - 1.6$ GeV/c (left) and $p_{\mathrm{T}} = 1.8 - 2.0$ GeV/c (right).
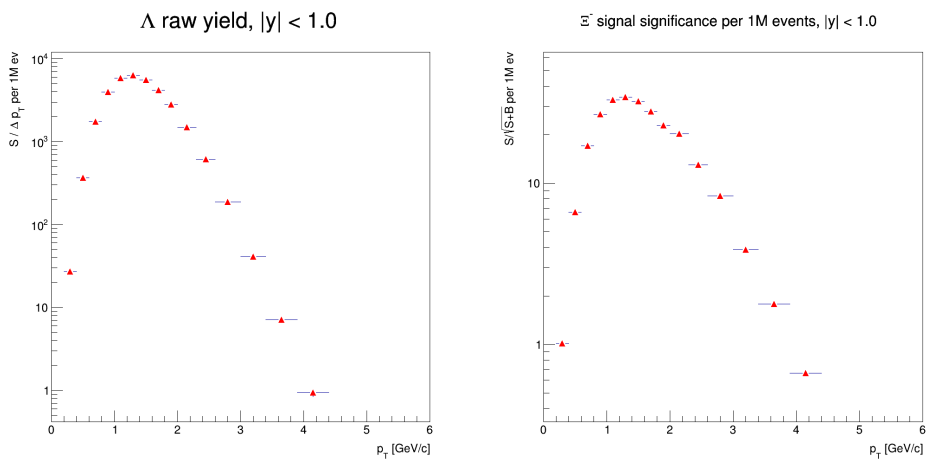

Figure 6.30: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected Λ yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.
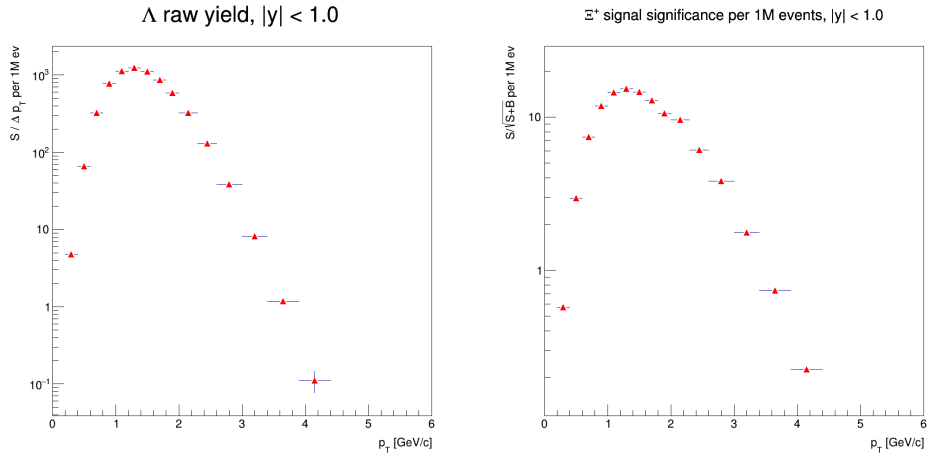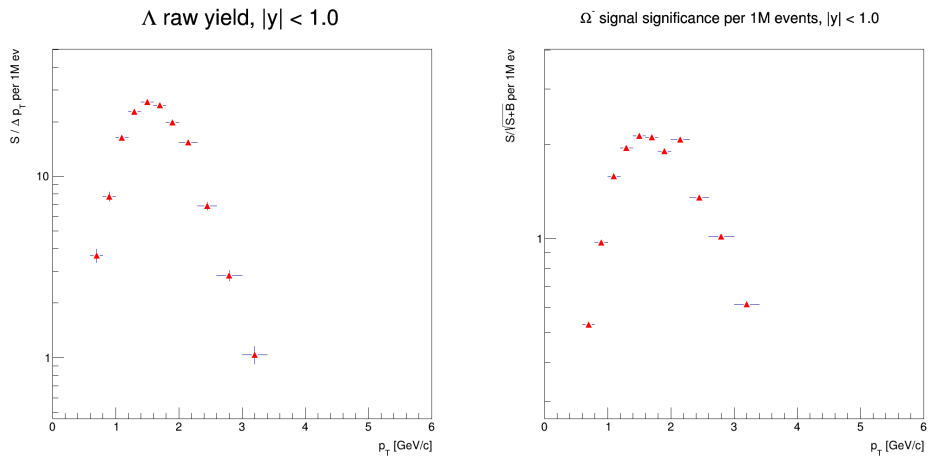
Figure 6.31: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected $\bar{\Lambda}$ yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.
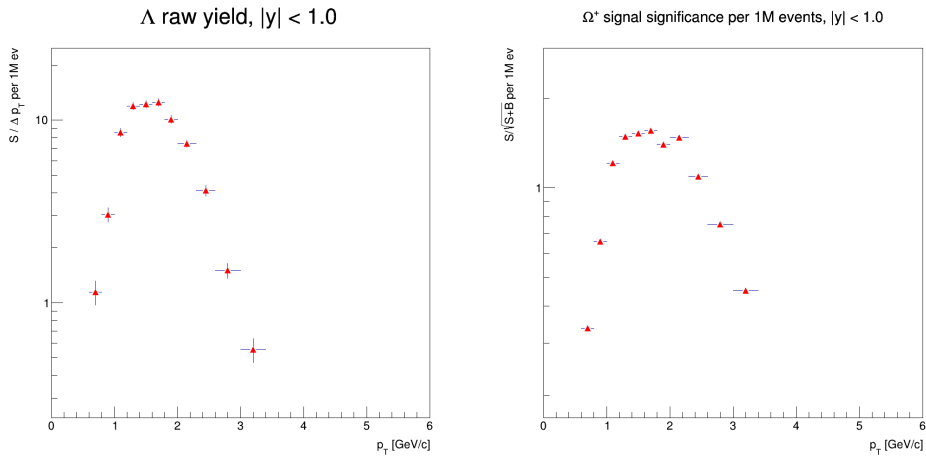

Figure 6.32: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected Ξ^- yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.

Figure 6.33: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected $\Xi+$ yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.

Figure 6.34: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected Ω^- yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.
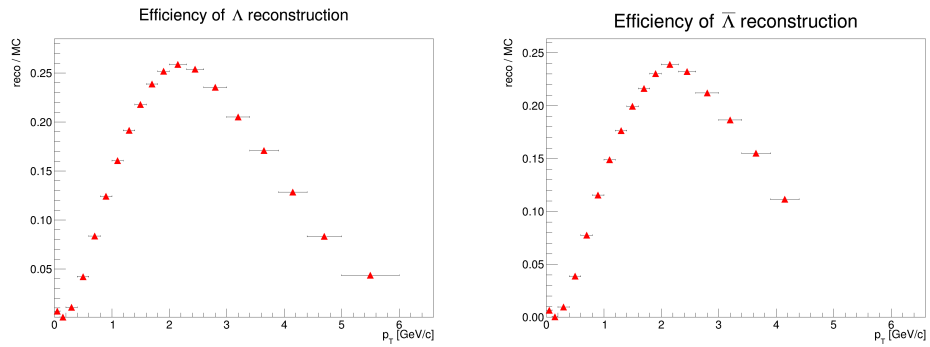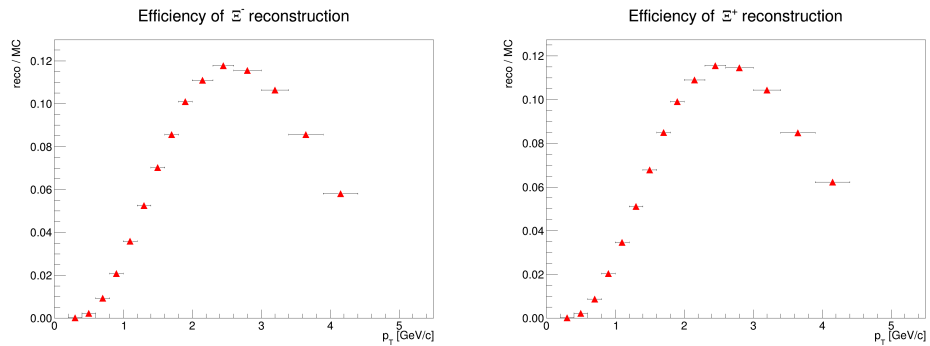
Figure 6.35: *THIS THESIS:* Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV: Dependence of uncorrected Ω^+ yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated per 1M events post selection.

6.4 Efficiencies of strange baryons reconstruction in 14.6 GeV Au+Au collisions

The efficiencies of reconstruction of strange particles in the 14.6 GeV dataset were obtained similarly to the 27 GeV case. This time there were 20 particles generated per event and the transverse momentum followed thermal distribution. Resulting efficiencies for all species of strange baryons are presented in Fig. **??**. There are several phenomena present in the behaviour of efficiencies we do not yet understand. Although the efficiencies of reconstruction of Ξ baryons behave adequately, there is yet unexplained drop of efficiency for both Λ baryons for $p_{\mathrm{T}} = 0.1 - 0.2$ GeV/c. In the case of Ω cascades there is a rapid drop around $p_{\mathrm{T}} = 2$ GeV/c and subsequent rise back to higher efficiency. The plot in Fig. 6.39 shows that it will be necessary to divide efficiency calculations into rapidity bins as well. For that, however, more statistics of simulation will have to be generated.
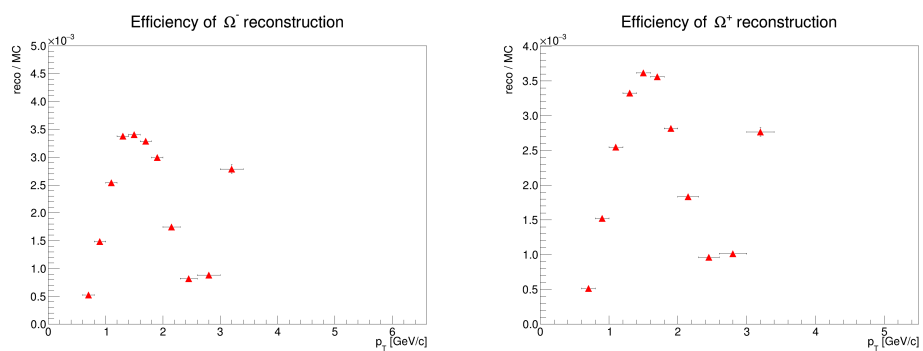

Figure 6.36: *THIS THESIS:* Efficiencies of strange hadrons Λ, $\bar{\Lambda}$ reconstruction in 14.6 GeV Au+Au collisions.

Figure 6.37: *THIS THESIS:* Efficiencies of strange hadrons Ξ reconstruction in 14.6 GeV Au+Au collisions.

60

Figure 6.38: *THIS THESIS:* Efficiencies of strange hadrons Ω reconstruction in 14.6 GeV Au+Au collisions.
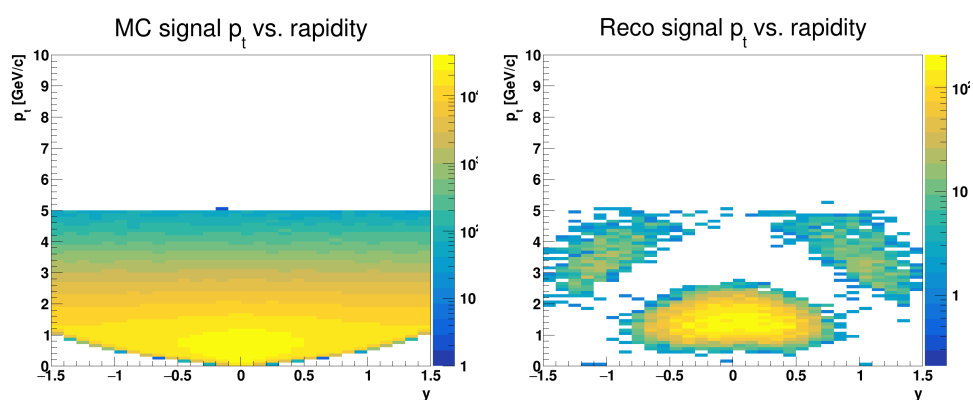
Figure 6.39: *THIS THESIS:* Transverse momentum p_T vs. rapidity y distribution for MC signal (left) of Ω and reconstructed Ω baryons (right).

6.5 Fixed target collisions at $\sqrt{s_{NN}} = 3.9$ GeV

6.5.1 Dataset and event selection

Fixed target data from collisions with center of mass energy $\sqrt{s_{NN}} = 3.9$ GeV come also from the express production of 2019 measurements. In order to collect this data accelerator was basically in the same settings as for 14.6 GeV, but there was only one beam at energy 7.3 GeV which was deflected to hit the fixed gold target and produce a collision at 3.9 GeV. The analysis of this dataset with KFPF is an important test of its geometrical independence. The Fig. 6.40 shows visualization of a fixed target collision with its unique geometry in collider experiment. The total number of sampled minimum bias events in this analysis counted 2.8M and all of these were used for reconstruction of Λ baryons.
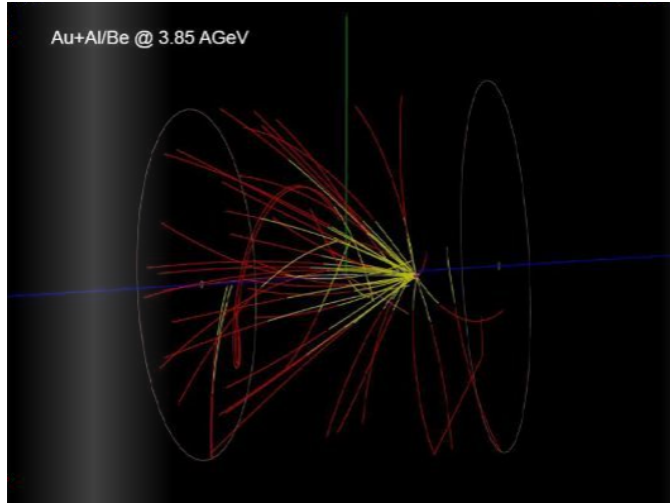

Figure 6.40: Visualization of fixed target collision at STAR. Taken from [55]

6.5.2 Analysis with default KFPF cuts

For the reconstruction of Λ particles the default KFPF cuts were employed again. This time soft ToF PID mode was turned on in KFPF which means that all hypothesis in 3σ window around predicted time of flight were accepted when identifying daughter particles. The calculation of yield and significance follows the same steps as before, although this time around the dependence on rapidity was explored as well since the CMS of the colliding nuclei is boosted in the laboratory frame. As an example plot from two transverse momentum bins and two plots from rapidity bins are plotted in Fig. 6.41 and Fig 6.42 respectively. Resulting yields and signal significance for all bins are presented in Fig. 6.43 and Fig. 6.44.
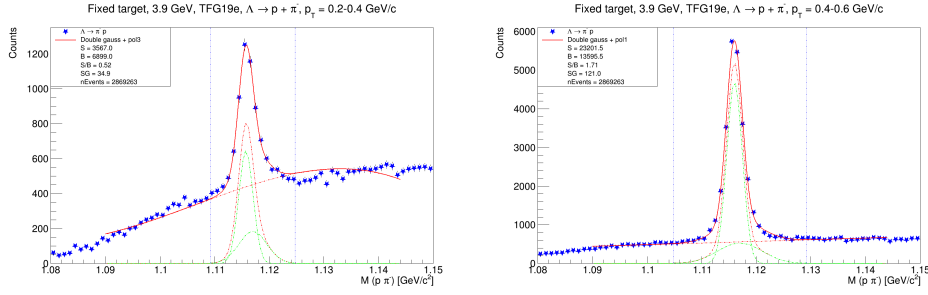
Figure 6.41: *THIS THESIS:* FXT collisions at $\sqrt{s_{NN}} = 3.9$ GeV: Invariant mass histograms for $p_\mathrm{T} = 0.2 - 0.4$ GeV/c (left) and $p_\mathrm{T} = 0.4 - 0.6$GeV/c (right).
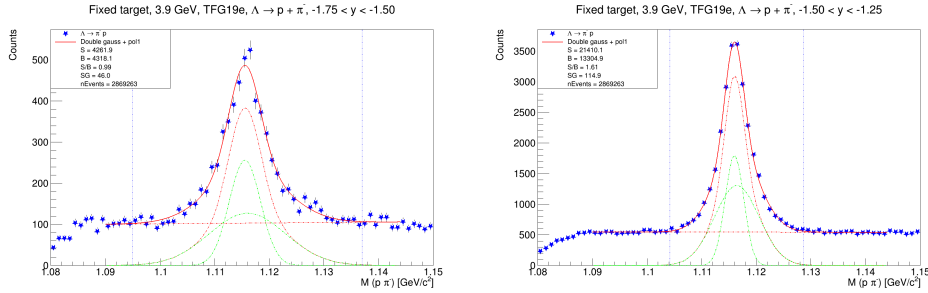
Figure 6.42: *THIS THESIS:* FXT collisions at $\sqrt{s_{NN}} = 3.9$ GeV: Invariant mass histograms for $-1.75 < y < -1.50$ (left) and $-1.50 < y < -1.25$ (right).
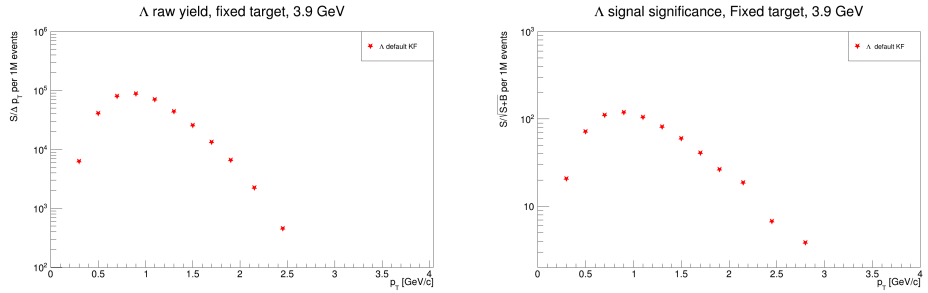
Figure 6.43: *THIS THESIS:* FXT collisions at $\sqrt{s_{NN}} = 3.9$ GeV: Dependence of uncorrected Λ yield (left) and signal significance (right) on transverse momentum. Both quantities are recalculated assuming 1M events post selection.
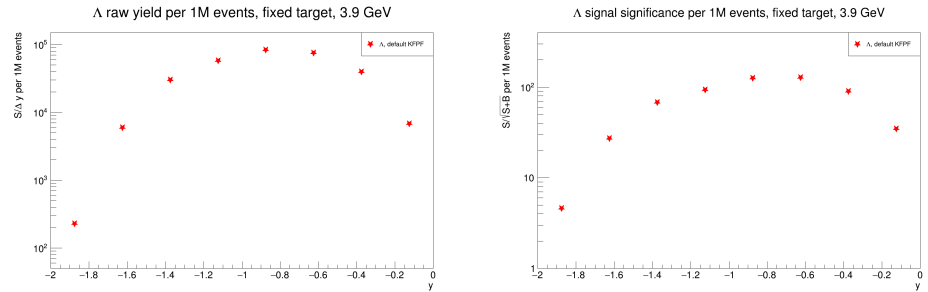
Figure 6.44: *THIS THESIS:* FXT collisions at $\sqrt{s_{NN}} = 3.9$ GeV: Dependence of uncorrected Λ yield (left) and signal significance (right) on rapidity. Both quantities are recalculated assuming 1M events post selection.

63

Chapter 7

Conclusion

The aim of this work was to test properties of KF Particle Finder (KFPF) software by reconstructing strange hadrons in various datasets from second phase of Beam Energy Scan program at RHIC. KFPF is a novel unified approach to particle reconstruction developed by FIAS group and implemented with their help at STAR. Besides investigating the qualities of KF Particle Finder, viability of express production of BES-II data for physics analyses was to be tested. The express production is based on online calibrations done by High Level Trigger (HLT) group and Tracking Focus Group (TFG) at STAR which I have participated in while working on this thesis. Monte Carlo simulations based on online calibrations were tested in this work and used to calculate efficiencies of strange hadrons reconstruction with KFPF. The ultimate goal of HLT and TFG together with FIAS group is to develop a fast data processing system which will be able to speed up physics analyses at STAR and subsequently prepare a base for future CBM experiment in GSI, Darmstadt.

First, strange baryons Λ were successfully reconstructed in Au+Au collisions at $\sqrt{s_{NN}} = 27$ GeV. It was shown that analysis with KF Particle based on cuts on probabilities rather than on absolute values of variables describing decay topology provides advantageous alternative to conventional approach. The impact of KF Particle can be seen especially in low p_T region, where, with the addition of Boosted Decision Trees, signal was reconstructed all the way down to $p_T = 0.1 - 0.2$ GeV/c. The statistical approach is more effective in this region because low p_T daughter tracks are more curved in the magnetic field of the detector and come with greater uncertainties. These findings are important for studies of strangeness production since in previously STAR published results the production at low p_T was only inferred from model fits to data.

As a next task, Monte Carlo simulation based on online calibrations done by HLT was used to calculate efficiencies of Λ reconstruction with KF Particle Finder and KF Particle Finder in combination with Boosted Decision Trees. The invariant yield of Λ baryons in $\sqrt{s_{NN}} = 27$ GeV collisions was computed and compared with results presented with STAR's previous pa-

per on strangeness. Good agreement was found at low p_T, however, with increasing p_T the spectra start to deviate. The discrepancies observed could be argued to stem from the way the primary vertex definition is handled within the simulation. Hence, there is still a need to understand better the Monte Carlo simulation as well as systematic uncertainties which would be the topic for continuation of this work. Still, the ability to see physics results so fast after data taking by analysing express production and using Monte Carlo simulation based on online calibrations to calculate efficiencies is intriguing and could help STAR collaboration to publish result from BES-II in shorter time than it is used to.

Next, the Au+Au collisions at $\sqrt{s_{NN}} = 14.6$ GeV were also analysed. The data came from express production of the dataset and a range of strange baryons including multi-strange hadrons were reconstructed to demonstrate viability of data produced with online based calibrations for physics analyses. It was interesting to see that even without BDT the signal of Λ was reconstructed almost to zero transverse momentum. This shows the benefits of inner Time Projection Chamber upgrade which improves low p_{T} tracking.

Finally, expressly produced data from fixed target experiment were analysed as well in order to demonstrate the geometrical independency of KF Particle package and show how convenient it is to switch between different detector setups once the analysis code is established. This could help to speed up physics analyses in STAR collaboration considerably.

Bibliography

[1] P. Senger, "The CBM experiment at FAIR," *J. Phys. Conf. Ser.*, vol. 50, pp. 357–360, 2006.

[2] S. Gorbunov, "On-line reconstruction algorithms for the CBM and ALICE experiments," Ph.D. dissertation, Johann Wolfgang Goethe-Universität, 2013.

[3] M. Zyzak, "Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR," Ph.D. dissertation, Johann Wolfgang Goethe-Universität, 2015.

[4] H. Geiger and E. Marsden, "Lxi. the laws of deflexion of a particles through large angles," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 25, no. 148, pp. 604–623, 1913.

[5] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys. Lett.*, vol. 8, pp. 214–215, 1964.

[6] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking. Version 1," 1964.

[7] M. J. Tannenbaum, "Highlights from BNL-RHIC," *Subnucl. Ser.*, vol. 49, pp. 295–342, 2013.

[8] T. Nayak and B. Sinha, "Search and study of Quark Gluon Plasma at the CERN-LHC," pp. 131–144, 2009.

[9] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets: an introduction to jet substructure and boosted-object phenomenology.* Springer, 2019, vol. 958.

[10] J. Adams *et al.*, "Evidence from d + Au measurements for final state suppression of high-p_T hadrons in Au+Au collisions at RHIC," *Phys. Rev. Lett.*, vol. 91, p. 072304, 2003.

[11] R. Snellings, "Elliptic Flow: A Brief Review," *New J. Phys.*, vol. 13, p. 055008, 2011.

[12] B. Alver and G. Roland, "Collision geometry fluctuations and triangular flow in heavy-ion collisions," *Phys. Rev. C*, vol. 81, p. 054905, 2010, [Erratum: Phys.Rev.C 82, 039903 (2010)].

[13] J. Rafelski and B. Müller, "Strangeness production in the quark-gluon plasma," *Phys. Rev. Lett.*, vol. 48, pp. 1066–1069, Apr 1982. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.48.1066

[14] J. Rafelski and J. Letessier, "Strangeness and quark gluon plasma," *J. Phys.*, vol. G30, pp. S1–S28, 2004.

[15] L. Adamczyk *et al.*, "Probing parton dynamics of QCD matter with Ω and ϕ production," *Phys. Rev.*, vol. C93, no. 2, p. 021903, 2016.

[16] J. Adams *et al.*, "Multi-strange baryon elliptic flow in Au + Au collisions at $\sqrt{s_{NN}} = 200$ GeV," *Phys. Rev. Lett.*, vol. 95, p. 122301, 2005.

[17] S. Singha, "Measurements of directed and elliptic flow for D^0 and $\overline{D^0}$ mesons using the STAR detector at RHIC," *Nucl. Phys. A*, vol. 982, pp. 671–674, 2019.

[18] L. Adamczyk *et al.*, "Centrality and transverse momentum dependence of elliptic flow of multistrange hadrons and ϕ meson in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV," *Phys. Rev. Lett.*, vol. 116, no. 6, p. 062301, 2016.

[19] A. Bazavov *et al.*, "Chiral and deconfinement aspects of the qcd transition," *Phys. Rev. D*, vol. 85, p. 054503, Mar 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.85.054503

[20] S. Ejiri, "Canonical partition function and finite density phase transition in lattice QCD," *Phys. Rev.*, vol. D78, p. 074507, 2008.

[21] D. Tlusty, "The RHIC Beam Energy Scan Phase II: Physics and Upgrades," October 2018, https://arxiv.org/pdf/1810.04767.pdf.

[22] G. Odyniec, "The RHIC Beam Energy Scan program in STAR and what's next ..." *J. Phys. Conf. Ser.*, vol. 455, p. 012037, 2013.

[23] C. Yang, "The STAR beam energy scan phase II physics and upgrades," *Nucl. Phys.*, vol. A967, pp. 800–803, 2017.

[24] L. Adamczyk *et al.*, "Beam Energy Dependence of Jet-Quenching Effects in Au+Au Collisions at $\sqrt{s_{\mathrm{NN}}} = 7.7$, 11.5, 14.5, 19.6, 27, 39, and 62.4 GeV," *Phys. Rev. Lett.*, vol. 121, no. 3, p. 032301, 2018.

[25] L. Adamczyk *et al.*, "Collision Energy Dependence of Moments of Net-Kaon Multiplicity Distributions at RHIC," *Phys. Lett. B*, vol. 785, pp. 551–560, 2018.

[26] Brookhaven National Laboratory. Upton, "Conceptual design of the Relativistic Heavy Ion Collider (RHIC)," 1986.

[27] DOE/NSF Nuclear Science Advisory Committee, "A Long Range Plan for Nuclear Science: A Report," 1983.

[28] S. Ozaki and T. Roser, "Relativistic Heavy Ion Collider, its construction and upgrade," *PTEP*, vol. 2015, no. 3, p. 03A102, 2015.

[29] A. Adare *et al.*, "An Upgrade Proposal from the PHENIX Collaboration," 2015.

[30] M. Harrison, T. Ludlam, and S. Ozaki, "RHIC project overview," *Nucl. Instrum. Meth. A*, vol. 499, pp. 235–244, 2003.

[31] J. G. Alessi, E. Beebe, O. Gould, A. E. Kponou, R. Lockey, A. Pikin, D. Raparia, J. Ritter, and L. Snydstrup, "High Performance EBIS for RHIC," *Conf. Proc. C*, vol. 070625, p. 3782, 2007.

[32] J. Alessi *et al.*, "The Brookhaven National Laboratory electron beam ion source for RHIC," *Rev. Sci. Instrum.*, vol. 81, p. 02A509, 2010.

[33] C. Liu *et al.*, "Improving the Luminosity for Beam Energy Scan II at RHIC," in *10th International Particle Accelerator Conference*, 2019, p. MOPMP044.

[34] U.S. Department of Energy, "Office of Science," 2020. [Online]. Available: https://science.osti.gov/np/Facilities/User-Facilities/RHIC

[35] M. Beddo *et al.*, "STAR: Conceptual design report for the Solenoidal Tracker at RHIC," STAR Collaboration, Tech. Rep., 1992.

[36] H. Qui, "STAR heavy flavor tracker," *Nuclear Physics A*, vol. 931, no. 1, pp. 1141–1146, 2014.

[37] Y. Wang, "STAR upgrade program and future physics," *J. Phys. Conf. Ser.*, vol. 535, pp. 12–22, 2014.

[38] "STAR Detector System." [Online]. Available: https://drupal.star.bnl.gov/STAR/future

[39] M. Anderson, J. Berkovitz, W. Bettes *et al.*, "The STAR time projection chamber," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 499, no. 2-3, pp. 659–678, 2003.

[40] S. Collaboration, "Technical Design Report for the iTPC Upgrade," STAR Collaboration, Tech. Rep., 2015, [Online]. Available: https://www.star.bnl.gov/public/tpc/tpc.html.

[41] P. Fachini, Z. Xu *et al.*, "Proposal for a Large Area Time of Flight System for STAR: STAR-TOF," STAR Collaboration, Tech. Rep., 2004, [Online]. Available: https://www.star.bnl.gov/public/tof/.

69

[42] K. C. Meehan, "The fixed-target experiment at STAR," *J. Phys. Conf. Ser.*, vol. 742, no. 1, p. 012022, 2016.

[43] T. Ablyazimov *et al.*, "Challenges in QCD matter physics –The scientific programme of the Compressed Baryonic Matter experiment at FAIR," *Eur. Phys. J. A*, vol. 53, no. 3, p. 60, 2017.

[44] I. Kisel, I. Kulakov, and M. Zyzak, "Standalone first level event selection package for the CBM experiment," in *18th Real-Time Conference*, 2012.

[45] V. Akishina, I. Kisel, I. Kulakov, and M. Zyzak, "FLES - First Level Event Selection Package for the CBM Experiment," in *GPU Computing in High-Energy Physics*, 2015, pp. 23–29.

[46] I. Kisel, "Event Topology Reconstruction in the CBM Experiment," *J. Phys. Conf. Ser.*, vol. 1070, no. 1, p. 012015, 2018.

[47] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper," *J. Phys. Conf. Ser.*, vol. 1085, no. 2, p. 022008, 2018.

[48] K. Albertsson, "TMVA Summary," 2018. [Online]. Available: https://drupal.star.bnl.gov/STAR/future

[49] A. Hocker *et al.*, "TMVA - Toolkit for Multivariate Data Analysis," 3 2007.

[50] S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett.*, vol. B716, pp. 30–61, 2012.

[51] C. M. Bishop, *Pattern recognition and machine learning.* Springer, 2006.

[52] M. Tanabashi *et al.*, "Review of Particle Physics," *Phys. Rev.*, vol. D98, no. 3, p. 030001, 2018.

[53] J. Adam *et al.*, "Strange hadron production in Au+Au collisions at $\sqrt{s_{\mathrm{NN}}}$ = 7.7, 11.5, 19.6, 27, and 39 GeV," 2019.

[54] F. Videbae, "Overview of iTPC," 2016. [Online]. Available: https://indico.bnl.gov/event/2357/contributions/5579/attachments/5063/6073/Project_Overview_v5.pdf

[55] D. Cebra, "STAR as a Fixed Target Experiment?" 2011. [Online]. Available: https://www.yumpu.com/en/document/read/24556481/star-as-a-fixed-target-experiment