DEPARTMENT OF PHYSICS
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

A Feasibility Study for Electron-Muon Correlations at ALICE

Author:
Nina Nathanson

Supervised By:
Dr. Thomas Dietel
Co-supervised By:
Dr. Zinhle Buthelezi
Dr. Siegfried Förtsch

August 21, 2024

Abstract

This analysis aims to assess the feasibility of using e-μ pairs as an alternative dilepton probe for heavy quark studies in proton-proton collisions. The e-μ channel would provide a much cleaner/low background signal than other currently-utilized probing methods, but it is statistics-limited, a restriction which is of increased concern at ALICE due to the detector's geometry. With the use of Pythia simulations, the viability of undertaking an analysis of the channel using ALICE p-p data from LHC Run 3 is considered. The analysis finds that it is feasible to undertake a data study of the e-μ pairs and presents an estimate of the expected signal in a year of Run 3 data.

Contents

•

1 Introduction

1.1 Content Overview

The semileptonic decay of heavy flavor hadrons (i.e., when a hadron containing a charm or bottom quark decays to a combination of lepton and hadron products) will sometimes produce correlated electron-muon (e-μ) pairs. The pairs are oppositely charged (or "signed") and, at sufficiently high transverse momenta, reflect the azimuthal direction of their parent particles.

In this way e-μ pairs are a potential probe of heavy flavor quark pairs, both for independent heavy flavor research and eventually in studies of the Quark Gluon Plasma (QGP) [1]. The e-μ signal has fewer contributing background sources than other probes of its type, so it is a desirable alternative to current heavy flavor measurement methods [2]. However, the decay channel is restrictively rare and as such, no investigation at ALICE utilized the pairs during previous data-taking periods.

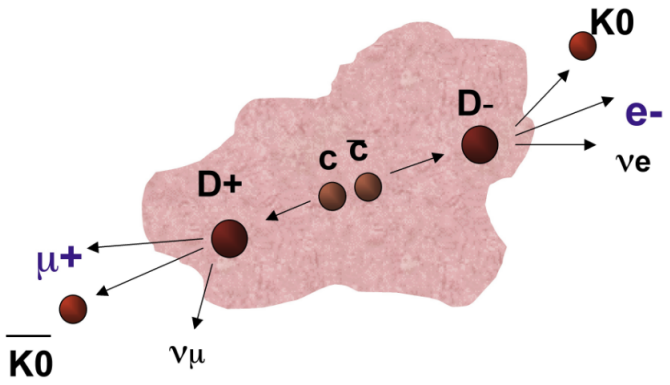


Figure 1: An illustration depicting the production of an oppositely-signed electron-muon pair through the semileptonic decay of a heavy quark pair [2]

With the increased quantity of data that will be recorded by the ALICE detector during LHC Run 3, more of the pairs will be observed, potentially lifting this restriction and making the e-μ channel a viable heavy quark probe. In this thesis, the feasibility of such an investigation is assessed using simulated 13.6 TeV proton-proton collision data to mirror the capabilities of the ALICE detector as of LHC Run 3. To provide context, the theoretical concepts behind the study are first introduced, along with a brief discussion of the ALICE detector hardware that would be used in a real data study of the e-μ pairs. Finally, the available ALICE analysis software tools are reviewed and the steps required to undertake such a study are presented and tested.

1.2 The ALICE Collaboration

The Large Hadron Collider (LHC) is located on the border of France and Switzerland at CERN, the European Organization for Nuclear Research. It has a radius of 27km, making it the largest circular particle collider globally. It is currently in its third run of data taking, which will continue to the end of 2025. Upgrades and continued operations of the LHC have been planned until 2041.

A Large Ion Collider Experiment (ALICE) is one of the four large experiments located on the main ring of the LHC [3]. The ALICE collaboration was created in 1993, and began taking data during the LHC's first run in 2009. Over the course of three decades, the collaboration has grown and is now made up of over 2000 scientists, spanning 170 institutions in 40 countries across the world [3].

ALICE has a unique mission statement among the LHC experiments. The experiment is designed to study a state of matter called the Quark Gluon Plasma (QGP), often described as a "hot, dense soup" of quarks and gluons. In the first 10^{-6} seconds after the Big Bang, after the immediate phase of inflation, the universe existed as a QGP [4] and in high-energy heavy ion collisions, such as those detected at ALICE, it is possible to simulate these conditions and temporarily produce QGP in the lab.

2 Theoretical Background

2.1 The Standard Model

The leading theory of modern particle physics is known as the Standard Model. This theory encapsulates all matter and forces in nature as being described by the interactions between a collection of "elementary particles", which are represented in Figure 2 [6]. Though there are physical phenomena that are not explainable using the Standard Model, it has so far been extremely successful at describing and predicting observations in experimental data.

The particles in the Standard Model can be split into two categories - bosons and fermions. Fermions make up the matter of the universe, while bosons mediate the fundamental forces which dictate how the fermions interact. Each force has one or more associated "gauge boson(s)", and the exchange of these gauge bosons between fermions cause the forces seen in nature. The strong force is mediated by the gluon, the weak force is mediated by the W and Z bosons and electromagnetism is mediated by the photon. Gravity, the fourth and final fundamental force, is the notable exception. It behaves differently to the other forces (for example, it is the only force that is purely attractive) and no gauge boson for gravity is included in the Standard Model, except on a hypothetical basis.

The twelve matter particles (fermions) can be further categorized into six quarks and six leptons. Every matter particle has a corresponding antimatter equivalent with oppositely signed electric charge and, in the case of quarks, opposite color charge (more on this in section 2.2). The fermions are arranged into three "generations", each containing two quarks and two leptons. Corresponding particles across successive generations are identical in all but mass, with increasing particle mass in every generation [6].

The type of quark or lepton is referred to as its "flavor". It is possible for particle flavor to

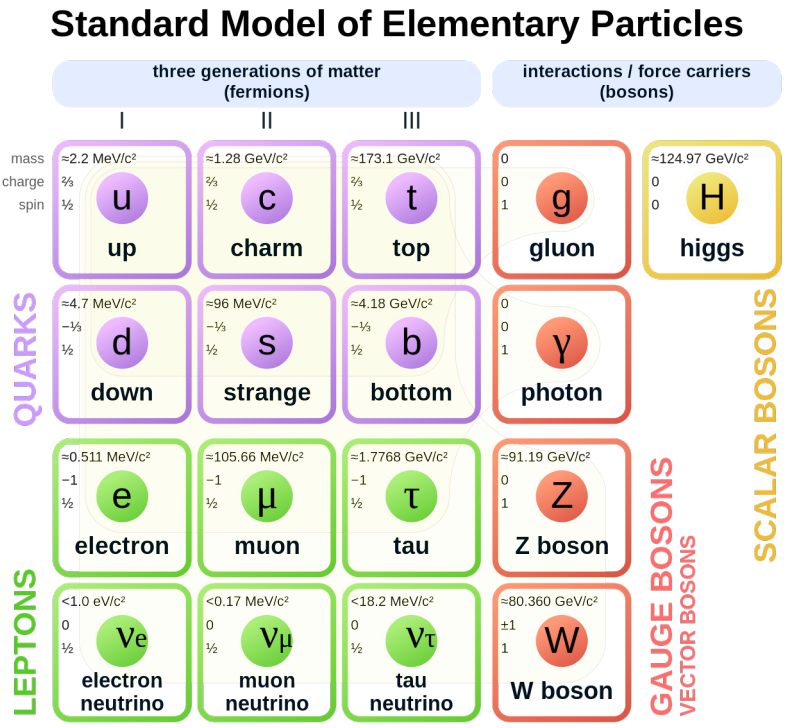


Figure 2: The Standard Model of particle physics, depicting the 17 fundamental particles along with their mass, electric charge and spin [5]

change during weak interactions, though there are certain rules regulating what transitions are possible. Flavor change occurs via the exchange of the W boson, which has an electric charge of ± 1. Since electric charge is a conserved quantity, the changed flavor of the particle involved in the interaction cannot have the same electric charge as it started out with. Specifically in the case of quarks this means that the top row of quarks in Figure 2 can transform into the second row of quarks (and vice-versa) but it is not possible to observe, for example, the transformation of a charm quark into an up quark.

2.2 Quantum Chromodynamics

The field of physics dedicated to studying the strong force is known as Quantum Chromodynamics (QCD). QCD describes the interactions between the six quarks and the strong force gauge boson, the gluon. While quarks experience all four of the fundamental forces, they are unique in that they are the only fermions to experience the strong force. Therefore, the theoretical and experimental study of quarks is crucial to investigations of QCD.

In nature, quarks exist in groups that are tightly bound together by the strong force. These bound states are called hadrons, and are categorized based on the number of valence quarks that exist in the group. The most common hadron categories are mesons (with two valence quarks) and baryons (with three valence quarks), though larger groupings (e.g. tetraquark states) have been observed in experiment. The type (or "flavor") of the valence quarks defines the type of

hadron. Also present within the hadronic bound state is a "sea of virtual gluons" that exist around the valence quarks [6]. Through the process of gluon splitting these gluons can produce virtual quark-antiquark pairs. These "sea quarks" are unstable and will typically then annihilate to form new sea gluons. In this way, the valence quarks are surrounded by a cloud of constantly changing virtual quarks and gluons.

2.2.1 Color Charges

The prefix **chromo-** in "quantum chromodynamics" is derived from the type of charge that is carried by strongly interacting particles, called "color charge". In reality, these charges have very little to do with the human understanding of visual color, rather each of the possible states of QCD charge are named after a color, which gives rise to the name. Color charge is a conserved quantity, just as electric charge is. However while electric charge has only one possible state, color charge has three - red, green and blue (r, g and b). The color charge carried by any given quark does not define its flavor.

Similarly to how a particle carrying electric charge can be either positively or negatively charged, each of the three color charge states have an equivalent, opposite anticolor charge. Antimatter quarks (antiquarks) carry the "negatives" of the three color states - they are either charged antired, antigreen or antiblue ($\bar{r}$, $\bar{g}$ and $\bar{b}$) [6]. Hadronic bound states of quarks are color-neutral, or "colorless", and themselves do not strongly interact with other hadrons.

Another major distinction when comparing QCD and quantum electrodynamics (QED) is the behavior of their gauge bosons. Photons do not carry electric charge, and therefore do not interact with electromagnetic fields. However, every gluon carries two charges (a color and an anticolor), allowing them to interact with the color field. This fundamental difference is important to keep in mind, and it leads to an effect called "anti-screening" [4]. When viewing a quark from a large distance, the cloud of virtual gluons that exists around the original quark augment the strength of the quark's color charge. This effect is named to highlight the contrast with "screening" in QED, where a particle's electric charge is felt more weakly at a distance due to the cloud of electrically-neutral photons which dampen the perceived strength of the charge.

2.2.2 Asymptotic Freedom

Asymptotic freedom is a property of QCD that is characterized by a significant change in the QCD coupling strength (α_S) at different energy scales and distances. In Figure 3, a comparison is made between the QCD coupling strength and the much weaker coupling strength of the electrodynamic interaction. While the coupling strength in QED has a minor change over the range of energy, temperature and distance on the x-axis, the slope of the α curve is much less extreme than that of α_S. This demonstrates the degree to which α_S is dependent on these factors. It can also be seen that in QED, coupling is strongest at short distances, whereas the opposite is true in QCD - this is a result of the screening/anti-screening effect discussed in the previous section.

The consequence of this asymptotic freedom is that QCD particles at low temperatures/energies, or particles which are far apart, are incredibly strongly coupled, leading to an effect known as color confinement (see section 2.2.4). Conversely, in order for quarks to act as free particles, they must exist under the conditions required for a weak α_S.

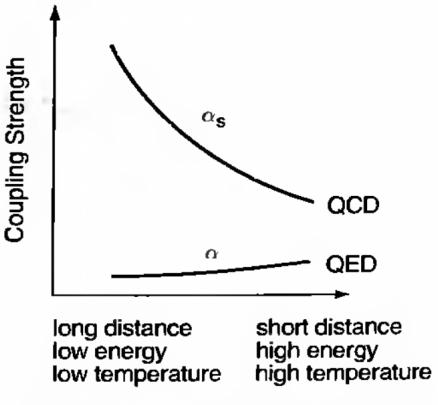


Figure 3: Graph demonstrating and comparing the change in the QED (α) and QCD (α_S) coupling strengths at high vs low temperatures, momenta and distances [4]

2.2.3 Hard vs Soft Processes

With the variable coupling strength associated with asymptotic freedom in QCD now established, a comment can also be made on how the QCD cross section is described. In QED, perturbation theory is used to calculate the behavior of the system across a range of momenta. Perturbation theory is a useful mathematical tool that involves representing the solution to a complex system as a series of corrections to a simpler system's solution [7], but it can only be successfully applied to QCD processes at a weak coupling strength. When a stronger coupling applies, the perturbative solution diverges and is therefore deemed a bad descriptor for the system [6].

To handle this, the QCD cross section is split into two parts using what is known as the factorization theorem. The non-perturbative cross section covers "soft" QCD processes at a stronger coupling strength and the perturbative cross section covers "hard" processes at weaker coupling strengths [8]. Soft processes are more abundant in high-energy particle collisions due to the less demanding conditions at which a strong coupling strength is achieved (i.e. lower energies and temperatures).

The divide between the hard and soft QCD cross sections informs how the entire QCD cross section is modeled. When simulating QCD processes, this separation must be taken into consideration. This will become very relevant in undertaking the simulations in section 5, so it is being introduced and highlighted here in preparation for later discussions.

2.2.4 Color Confinement and Hadronization

As has already been briefly introduced, one of the main consequences of asymptotic freedom is color confinement. At the beginning of this section it was mentioned that quarks generally only exist in the form of hadrons. This can now be understood, in that except in some extreme cases (e.g. neutron stars), the high energies and temperatures that are required to reduce the QCD coupling constant do not exist in nature [4]. For this reason, quarks can only be observed in bound states without a color charge, which provides a significant challenge for experimental

studies of QCD. Although color confinement makes direct measurements of color-charged particles impossible, indirect methods can be employed.

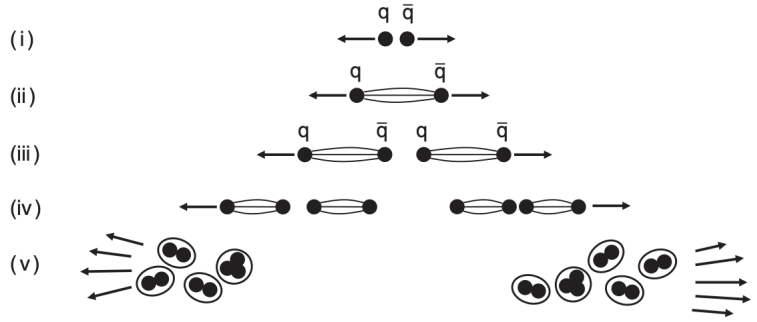


Figure 4: The quark hadronization process [6]

Figure 4 illustrates the formation of hadrons and the effects of color confinement. As a quark-antiquark ($q\bar{q}$) pair move away from each other, potential energy builds up in the color field between them. This eventually reaches a point where it is not only possible, but "energetically favorable" [6] to produce a new quark of opposite color charge for each of the originals. This is known as quark fragmentation [8]. This results in two quark pairs, each with slightly lower energy than the original, which then repeat this process of quark fragmentation continuously. After a few fragmentations, the quarks reach a low enough energy that color confinement takes hold and hadronization begins.

In particle collisions, this hadronization process results in a jet of energetic particles radiating outwards from the quark's origin point, in the general direction of its movement. Since hadrons in the jet carry a signature of the quarks that formed them, the decay products of the hadronization can be used as a probe for the original pair.

2.2.5 Quark Gluon Plasma

Though asymptotic freedom leads to color confinement at the point of strongest coupling, the converse of the effect is that quark deconfinement is possible above a critical point where conditions result in a sufficiently weak coupling strength, as seen in the QCD phase diagram of Figure 5. Past this boundary, color charged particles are able to exist freely in the QGP, as was introduced during the description of the ALICE detector in section 1.2. The extreme conditions of the QGP are a direct consequence of asymptotic freedom.

After the deconfinement phase transition has occurred, the particles within the QGP are still experiencing the effects of the strong force. Since they continue to interact strongly with the other quarks and gluons within the plasma, it is expected that quarks moving through the QGP will experience a drag force thereby losing energy and depositing it into the medium around them [1]. This results in a signal quenching effect: i.e., it is expected that quark signals from collisions with more available energy will be reduced (or "quenched") and not enhanced in comparison to signals from collisions with less available energy, entirely due to the drag experienced in the QGP [10].

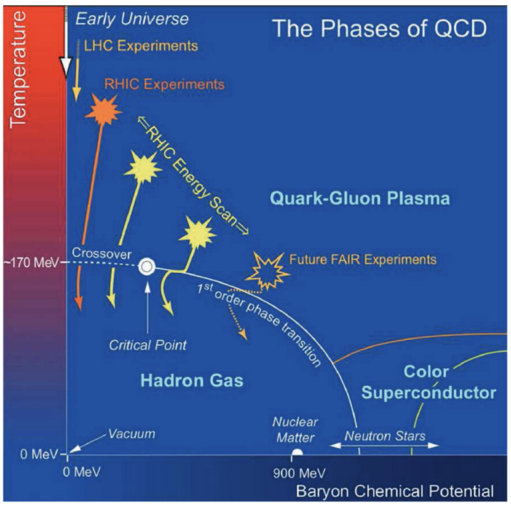


Figure 5: The QCD phase diagram, with the first-order phase transition between hadron gas and the quark-gluon plasma indicated by a solid white line [9]

To measure this in experiment, signals from different collision systems are used. Generally, QGP is predicted in heavy ion (A-A) collision systems but not in proton-proton (p-p) or proton-ion (p-A) systems. The differences between the signals in these two collision systems can be used both to identify the presence of QGP, and to quantify how it is affecting the particles moving through it.

2.2.6 QGP Evolution

The first stage of the QGP's evolution is thermalization. Soon after the collision, the system reaches a state of thermal equilibrium at which point it can be effectively described by relativistic hydrodynamics [4]. At this point, the QGP is anisotropic/elliptic in shape and so pressure gradients begin to form at between different points in the medium, inducing flow.

During this phase of relative stability, it has been found that the QGP fits the expectations of a "perfect fluid" better than any other observed substance [10]. However, the system is constantly expanding and eventually reaches a "freeze-out" point where this hydrodynamic equilibrium cannot be maintained. As the quarks and gluons lose energy and move apart in the expansion, α_S increases and color confinement again becomes relevant as quarks within the plasma begin to undergo hadronization. The hadrons that are formed consist of the quarks that have interacted with and subsequently lost energy in the QGP. Therefore, they carry some signature of the medium and can be used as a probing mechanism as desired.

2.3 Heavy Flavor Quarks

The six quark flavors can be grouped by their mass, and quark studies are often split into three distinct fields based on this categorization. The up, down and strange quarks are generally defined as the light flavors, the charm and bottom quarks are the heavy flavors and the top quark, which is orders of magnitude more massive than any other quark, is in its own league. This analysis is specifically focused on the study of heavy flavor quarks, which are sometimes

denoted with a capital Q (or $\bar{Q}$ for antiquarks) to differentiate them. This convention will be used going forward.

Heavy flavor quarks are valued as probes for QGP studies since they lose energy while flowing through the medium, yet survive past the point of freeze-out in the form of hadrons that contain heavy quarks (such as D mesons). Therefore they are prime candidates for which to observe the signal quenching effect discussed in section 2.2.5. However, there is also a lot of independent value in studying heavy quarks past their use in QGP research. Their higher mass implies that they can only be produced during hard processes, so those with sufficiently high energy - in this way they are an ideal test for the perturbative sector of the QCD cross section [11].

2.3.1 History and Context

In the mid-20th century, only three quarks had been discovered experimentally: the light flavor quarks, up, down and strange. The Standard Model had not yet been defined, and the pattern of the three generations of matter as they are understood today had not yet emerged. Still, physicists of the time speculated about the existence of a fourth quark to reflect a symmetry with the four leptons that had so-far been discovered [2]. Sheldon Glashow, one of the predictors of the charm quark, described coining the name due to how charmed he and his peers felt at the idea of such a symmetry [12].

Some time after these initial speculations Glashow and two contemporaries published a paper putting forward a theory called the GIM mechanism, which predicted the charm quark more formally. The GIM mechanism solidified the importance of a fourth quark in making electroweak theory compatible with the latest experimental observations. To understand the role played by the charm quark, one must refer back to the laws of flavor changing via exchange of the W boson as first introduced in section 2.1. If there existed only three quarks, electroweak theory, which otherwise succeeded at describing experimental data, would require that strange to down quark decay processes (and vice versa) were physically possible [11]. This is a phenomenon known as "flavor changing neutral currents", and the important take away is that under this model quarks must be able to change flavor without changing their electric charge, breaking charge conservation.

However, charge conservation was well supported by experiment, and the changing of lepton flavor without the changing of electric charge was not observed. To rectify this disconnect between theory and experiment, the GIM mechanism predicted an alternative involving a fourth quark. Electroweak theory would then no longer require flavor changing neutral currents to function, thus solving the problem [11]. In the context of the GIM mechanism, Glashow is quoted as describing the evolution of the quark's name, as it had now become more like an arcane charm - a "magical device to avert evil" that would have otherwise been brought about by the failure of the three quark model [12].

The GIM mechanism's prediction proved accurate and the charm quark was finally discovered in experiment in 1974. The second heavy flavor quark, the bottom quark, was predicted and then experimentally substantiated soon after.

2.3.2 Production and Decays

There are three mechanisms through which heavy flavor quarks can be produced: pair creation, flavor excitation and gluon splitting [13].

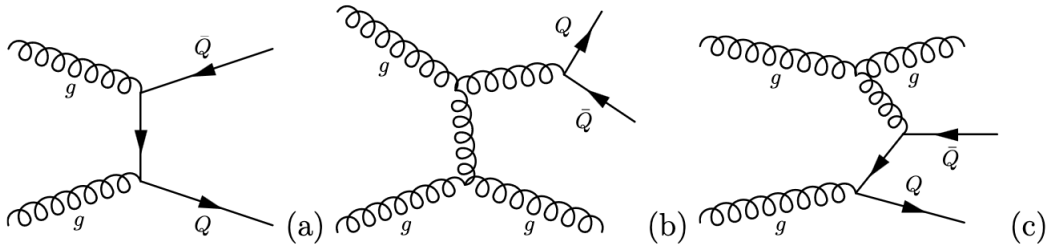


Figure 6: An example of a Feynmann diagram for each of the heavy quark production processes: pair creation (a), gluon splitting (b) and flavor excitation (c) [14]

Pair creation is the leading-order process, and it dominates at lower center of mass energies [2]. This involves the production of $Q\bar{Q}$ pairs either from two gluons as displayed in Figure 6, or from another $Q\bar{Q}$ pair. In this process, the quark and antiquark are created "back to back" in azimuth, which means that they move apart from each other in opposite directions. This means we expect the azimuthal separation (denoted $\Delta\phi$) between the two quarks to be equal to π. Azimuthal angle is a major kinematic variable in the collider coordinate system, which will be discussed in greater depth in section 3.1.

The next process to be discussed, gluon splitting, is the dominant process at higher center of mass energies [2]. As suggested by the name, it involves the production of a $Q\bar{Q}$ pair from a single parent gluon. Gluon splitting was mentioned earlier, at the beginning of section 2.2, as it is also the fundamental mechanism responsible for the production of sea quarks from sea gluons within hadrons. Lastly, flavor excitation involves the further evolution of an already-existing quark. In particle collisions, this often occurs when a quark in one of the incoming beam protons interacts with a sea quark from the other [13].

This analysis is specifically focused on the semileptonic decays of "open" heavy-flavor hadrons, whereby a hadron containing a single heavy flavor quark decays into a mixture of leptons and hadrons [15]. For charm and bottom quarks respectively, these open hadrons are typically some variation of D (charm) or B (bottom) meson.

2.4 Electron-Muon Correlations

With a basic explanation of QCD and heavy quark studies now complete, an introduction can be made to the e-μ decay channel - its benefits, drawbacks, and the practicality of experimental studies.

The e-μ correlation method allows for a measurement to be made of heavy flavor quark production in high energy collisions. Unlike heavy quark studies which make use of single-particle decay

products, utilizing multi-particle probes can provide insight on the origin of the heavy quarks being measured [15] and e-μ pairs in particular are at an advantage. Of the three heavy quark production mechanisms outlined in section 2.3.2, only the leading order process of pair creation results in the production of a back-to-back quark-antiquark pair. This azimuthal separation is reflected in the decay products of the parent quark pair, especially when looking at final state particles with high momenta (explored further in section 2.4.2).

By looking for electrons and muons with this same azimuthal separation, it is possible to measure heavy quark pairs specifically produced through pair creation [2]. This ability to identify the mechanism behind the production of the measured heavy quarks is beneficial, as it allows for comparisons to be made with theoretical models of the pair creation cross section.

Additionally, the measurement has the potential to contribute to studies of the QGP at ALICE in the long term. In heavy ion collisions, it is expected that measurements of heavy quarks production via this method would exhibit the quenching effect described in section 2.2.5. The way to quantify this is to use the "quenching factor", denoted I_{AA}:

$$I_{AA} = \Sigma_{AA}/\Sigma_{pp} \tag{1}$$

where

$$\Sigma = \int_{\phi_{min}}^{\phi_{max}} d(\Delta\phi) \left[\frac{1}{N_{trig}} \frac{dN_{assoc}}{d\Delta\phi} \right] \tag{2}$$

The study of e-μ pairs in p-p collisions would provide the baseline measurement of Σ_{pp} required for this calculation, which has so far only been performed using simulations in previous investigations [1].

Finally, the e-μ correlation channel could contribute to QGP searches in p-p collisions, an ongoing area of research. In p-p collisions with higher outputs of charged particles (known as "high multiplicity" collisions), a comparison could be made of the e-μ signal in Run 3 data with simulated predictions that use current theoretical models of quark fragmentation. If heavy quark production in these collisions exhibits patterns of suppression reminiscent to those seen in heavy ion collisions, it would contribute to the effort to search for QGP production in smaller collision systems.

2.4.1 Comparison with Other Probes

The main advantage of using the e-μ channel is that it is an extremely clean process. The e-μ sample is uncontaminated by many of the backgrounds that must be accounted for in other probes. Though other probes (such as jet probes, or lepton-hadron correlations) do exist, the focus here will be on comparing e-μ correlations with other dilepton probes. The dilepton probes typically utilized in heavy flavor studies at ALICE are same species probes, so either electron-electron (e-e) or muon-muon (μ-μ).

The biggest drawback of using same species correlations, which e-μ correlations neatly avoid, is the large amounts of background signal that specifically produce false pairs that interfere with their measurement. The contributors to this same-species background can be broken down into two: those processes that also result in an e-μ background, and those that don't. The fact that this second category of background exists in itself is clearly an advantage, as it implies that the

background of the e-μ signal is reduced in general. Included in this category are leptons that come from resonance decays and Drell Yan production [2].

However even those processes that exist in the first category, and do produce a background signal for the e-μ measurement, can be conveniently addressed. These processes (e.g. light-flavor meson decays, photon conversions and dalitz decays [15]) are equally likely to produce like-signed e-μ pairs [2]. Therefore, the desired e-μ signal exists in the discrepancy between the number of like- and opposite-signed pairs recorded. By doing a "like-signed subtraction" from the opposite-sign signal, a large portion of this background signal can be removed.

Of course, no signal will ever be fully and perfectly background subtracted, but the cleaner nature of the e-μ signal still sets it apart. There will also always be a certain amount of background and error associated with the detector being used to take measurements and particle identification from detector signals. This will be discussed more in the context of ALICE throughout section 3 and section 4.

The major disadvantage associated with e-μ studies is that e-μ pairs are produced very rarely, which means any study that is undertaken will be statistics-limited in a way that e-e and μ-μ studies are not [2]. In previous runs at ALICE, the effect of this has been so great that the e-μ channel was not investigated, and the other dilepton probes were prioritized. This is the motivation for the feasibility study at the basis of this analysis - to see if it is viable to analyze e-μ probes in ALICE Run 3 data, or if the statistics limitation would still be too great.

2.4.2 Experimental Studies

The practicalities of studying e-μ pairs can now be discussed. First, it is necessary to cut out particles with lower momenta, which are more likely to come from background processes. This is not done equally for both electrons and muons - instead, one lepton flavor is chosen as the "trigger" particle and the other as the "associated" particle in the pair. A high momentum cut is first made on the trigger lepton, which initially reduces the sample of events being analyzed. Then, a slightly lower momentum cut is made on the associated lepton. The choice of trigger/associated lepton and the magnitude of the momentum cuts will be carefully considered in section 5.6.

In addition to cutting out background, momentum cuts are also important to choose the leptons that most closely reflect their parents' azimuthal direction and therefore make a more accurate measurement of the initial quarks that are being probed. Heavy quarks are by definition produced as high energy particles, so it is likely that when a decay lepton is recorded at high momentum, it is because the lepton has retained a large fraction of its parents energy. Since momentum has a directional component, this in turn means that higher momentum leptons will by propelled in an azimuthal direction closer to their parents. They are also less likely to be influenced by scattering effects.

So much emphasis is placed on the azimuthal angle of the quarks and their lepton children because the azimuthal correlation $\Delta\phi$ is the main measurement of the pairs. For every e-μ pair that is detected, $\Delta\phi$ is recorded and then stored in a histogram, through which it would be possible to visualize trends in the correlation. In general, these azimuthal correlation histograms will start to form two peaks - one centered on 0 (the "near-side" peak) and a second centered on π (the "away-side" peak). In the e-μ case, the expectation is that only an away-side peak

will form, so pairs in the near-side peak are most likely background.

During previous investigations at PHENIX (the Pioneering High Energy Nuclear Interaction eXperiment), the e-μ channel was analyzed in proton-proton and deuteron-gold collisions [2]. The azimuthal peak of the d-Au collisions was observed to be significantly quenched, more so than was expected (see Figure 7). The cause of this is still unclear, and further investigation into the pairs would provide more insight [2].

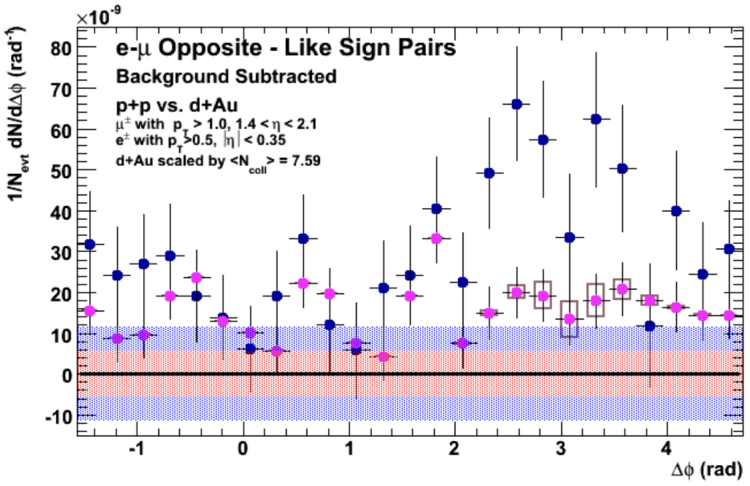


Figure 7: The azimuthal correlation between e-μ pairs in a previous study at PHENIX, for both the p-p and d-Au collision systems [2]

No further investigation has been undertaken to catalog e-μ correlations in heavy ion collision data, where there would be a strong expectation of QGP.

3 Hardware Setup

3.1 Kinematic Coordinates

In particle physics experiments, a unique coordinate system (referred to throughout this analysis as the "kinematic coordinates" is used to define the location of particles based on their relation to the beam line and the interaction point (IP) where the collision occurred. A brief overview will now be provided along with the visual representation in Figure 8.

The primary coordinates are transverse momentum (p_T), azimuthal angle (ϕ) and pseudorapidity (η). These are extremely convenient to work with, as they provide important information about the properties of any given particle without being influenced by the aftereffects of the collision. For instance, transverse momentum (the component of a particle's momentum in the xy plane) is inherently perpendicular to the direction of the colliding beams along the z-axis [6] and it therefore excludes the large remnant of the beam's momentum that would have obscured information about the particle itself. Azimuthal angle is also measured perpendicular to the beam line, in the same way that it would be in cylindrical coordinates.

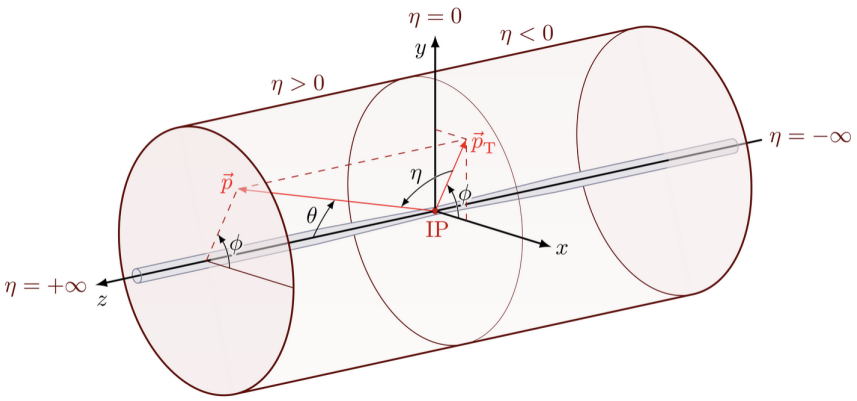


Figure 8: Visualization of the detector (p_T, η, ϕ) coordinate system, with comparisons made to the cylindrical (ϕ, θ, z) and cartesian (x, y, z) coordinate systems [16]

3.1.1 Note on Pseudorapidity

The final kinematic coordinate, pseudorapidity, is less intuitive than the others and so will now be addressed in more detail. Pseudorapidity is related to the concept of rapidity (y), which is in turn related to velocity and momentum. Rapidity has the advantage of being Lorentz invariant, meaning that it doesn't change based on the reference frame of an observer under special relativity [17]. It is defined in equation 3:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{3}$$

There is a feature of this rapidity definition that is a complication in experiment - it is inherently reliant on energy, and therefore on particle mass, which requires knowledge of what type of particle is being recorded. In experiment, it is not possible to identify the particles at the data capture level. Instead, this can only be done much later on during analysis (as will be discussed in section 3.3). To remove this dependence, an approximation can be used that is entirely dependent on polar angle (θ, the angular direction of movement in the z-plane). This is called the pseudorapidity,

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \tag{4}$$

In Figure 9, the corresponding value of η is shown for a range of polar angles. Due to this relationship with polar angle, the range of particles that a detector is able to record is often quoted in pseudorapidity regions. At ALICE, the detector is split into two main η regions - a central barrel which covers $|\eta| \leq 0.9$ and a forward region which covers $-4.0 \leq \eta \leq -2.5$.

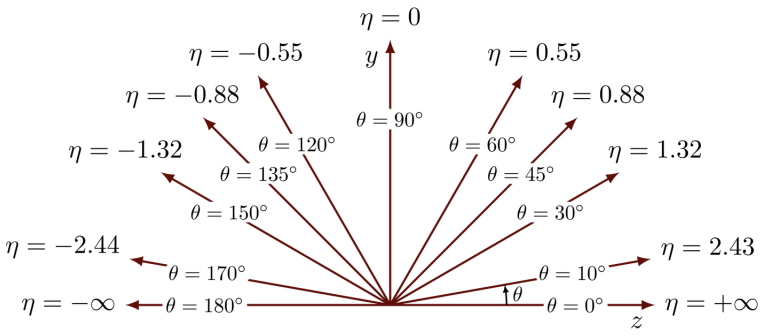


Figure 9: The relationship between polar angle (θ) and pseudorapidity (η) [16]

3.2 The ALICE Detector

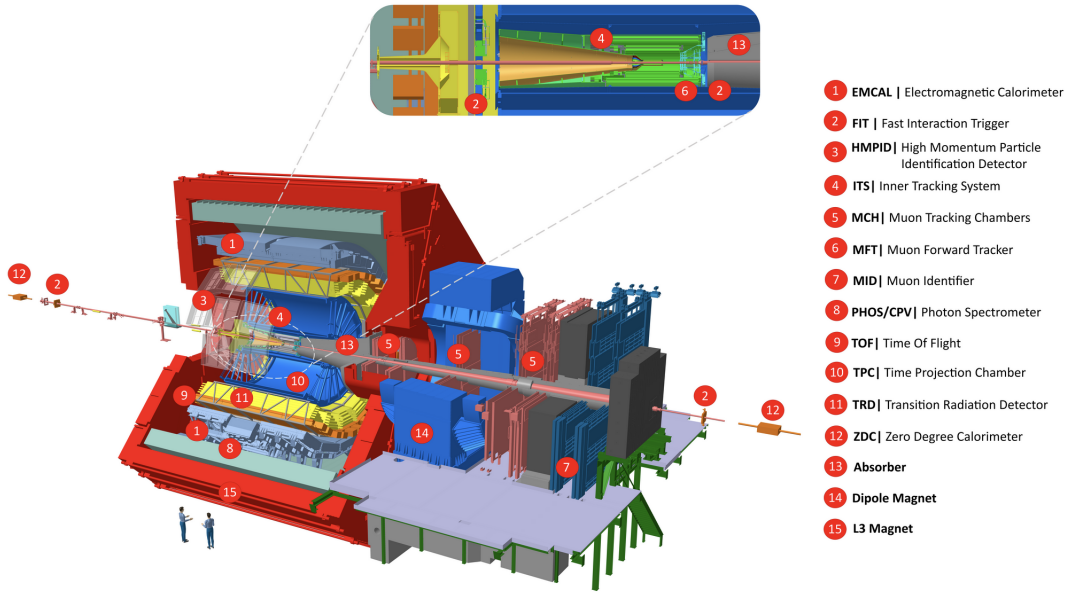


Figure 10: An overview diagram of the ALICE collaboration detector as of LHC Run 3 [3]

In Figure 10, the various subdetectors at ALICE as of Run 3 are displayed. The original iteration of the detector operated during Runs 1 and 2 before undergoing major hardware upgrades. The inner tracking system (ITS) was completely replaced. The time projection chamber (TPC), the transition radiation detector (TRD), the time of flight chamber (TOF) and the muon spectrometer [18] subdetectors underwent significant electronic upgrades. Additionally, the diameter of the beam pipe was reduced and a muon forward tracker (MFT) was added to supplement the tracking capabilities of the muon spectrometer. These hardware changes, along with software upgrades that will be described in section 4 will improve ALICE's capabilities for both electron and muon detection.

The detector is split into two main sections based on η coverage (the "central barrel" and "forward" regions). The main difference in purpose of the two sections is the particles that they are designed to observe. The forward region specifically handles muon detection, while all other particles are picked up in the central barrel. This layout affects the inherent properties of the e-μ pairs which are detectable using the ALICE detector. Unlike with the same-species dilepton probes introduced in section 2.4, where both leptons in the pair would have to have similar pseudorapidities to be detected, the ALICE detector geometry defines that the electrons and muons which make up the pair signal must instead exist in two different regions.

Though this η separation further reduces the number of e-μ pairs that are available for analysis, it also provides access to a different region of physics. The total momentum of any hadron is distributed between its constituent quarks and gluons (known as "partons"). There is a correlation between the fraction ("x") of the hadron's momentum carried by any given parton and the pseudorapidity of the particles produced in its decay. In this way, particle correlations at midrapidity can be used to probe partons in the high-x region, while forward rapidity correlations provide a probe of the low-x region [2]. By utilizing η-separated leptons, the e-μ signal will provide a probe for a region between these two extremes that is currently not accessible at ALICE via the other dilepton channels.

3.3 Central Barrel Detectors

In this analysis, the central barrel is relevant for its use in electron detection. However, due to the many other types of particles that the central barrel detects, the major challenge that will be faced is particle identification. It is crucial to determine which of the many particle tracks belong to electrons, but doing so requires several different subdetectors with the central barrel and will always come with some degree of uncertainty. The primary method of particle identification is to measure how much energy is lost by a particle, and therefore deposited in the detector material, as it moves through various detector layers. This quantity of energy loss is referred to as $\frac{dE}{dx}$. In this section, the most relevant detectors involved in electron PID will be discussed, with the knowledge that there are many more components to the overall ALICE detector.

A crucial element in the central barrel is the solenoid magnet which surrounds all of the central subdetectors. The magnet was repurposed from the L3 experiment, which previously occupied the location that now houses the ALICE detector. It weighs 7800 tonnes and generates a magnetic field of 0.5 T [19]. As particles are produced in proton or heavy ion collisions, they are exposed to this field. The curvature of their path through the detector reflects the manner in which they interacted with the field, which allows for conclusions to be made about their charge and momentum.

3.3.1 Inner Tracking System

The first detector encountered by particles outgoing from a collision is the inner tracking system (ITS). This is the main detector used in reconstructing the location of the IP of each event. It is also important in the identification of secondary verticies, which are the points of particle decays after the collision has occurred [19]. In the Run 3 upgrade, its resolution capabilities in this area improved by a factor of 3 [18]. Also expanded in the upgrade was its particle tracking abilities. It can now track electrons at lower momenta then before, with a minimum trackable

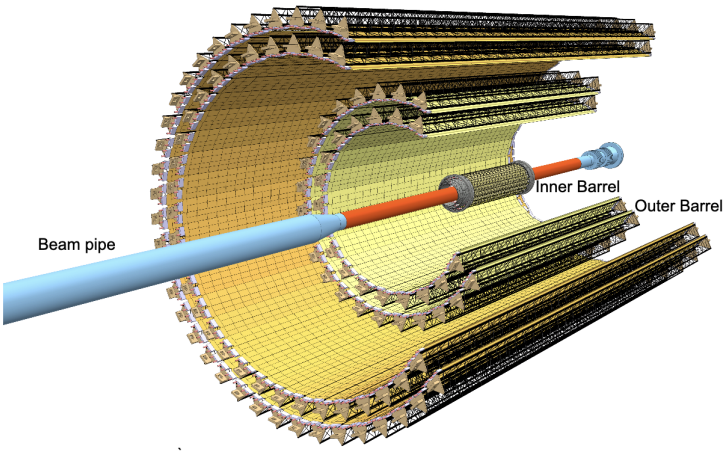


Figure 11: Diagram of the ALICE inner tracking system, highlighting the inner and outer barrels around the beam pipe [18]

p_T of 0.05 GeV [18].

The ITS is constructed of seven distinct silicon layers grouped into two cylindrical barrels - the inner barrel (3 layers) is closest to the beam pipe and is surrounded by the outer barrel (4 layers). Each of these layers is made out of long triangular prisms (called "staves") arranged into concentric cylinders. Before Run 3, the ITS had only six layers. In order moving outwards from the beam pipe, there were two of each of the following: silicon pixel detectors (SPDs), silicon drift detectors (SDDs) and silicone strip detectors (SSDs) [19]. With the second iteration of the ITS, all seven layers are now specialized SPDs that specifically use "monolithic" pixel chips, where all the circuitry is included on one piece of silicon [20].

3.3.2 Time Projection Chamber

Surrounding the ITS is the time projection chamber (TPC), the largest subdetector at ALICE. The TPC is a gas-based tracking detector that plays an important role in particle identification. The ALICE TPC has a minimum tracking efficiency of 98%, and can perform particle identification for $p_T \leq 20$ GeV.

As shown in Figure 12, the TPC is made up of a central cylindrical "cage" chamber. A potential difference is produced within the cage between the central electrode and the endplates. As a charged particle moves through the gas-filled cage, it ionizes its surroundings and produces ionization electrons, which then drift along the lines of potential difference to the endplates [20]. This will happen throughout the particle's movement, meaning a series of ionization electrons will be produced corresponding to locations on the original path (or "track") of the charged particle through the detector.

As the ionization electrons reach the endplates they are read as data points. Information about their position on the end plates and the time that they reached the plates can be used to reconstruct the particle's path. At ALICE, the TPC end plates are split into 18 sectors. In Run 3, the readout chambers on the end plates were upgraded to gas electron multipliers (GEM

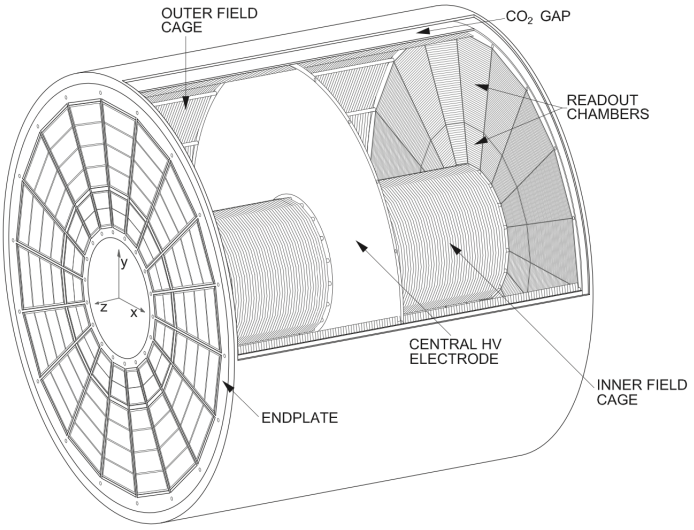


Figure 12: Diagram of the ALICE time projection chamber [21]

chambers) which are better equipped to handle high luminosities [21].

3.3.3 Transition Radiation Detector

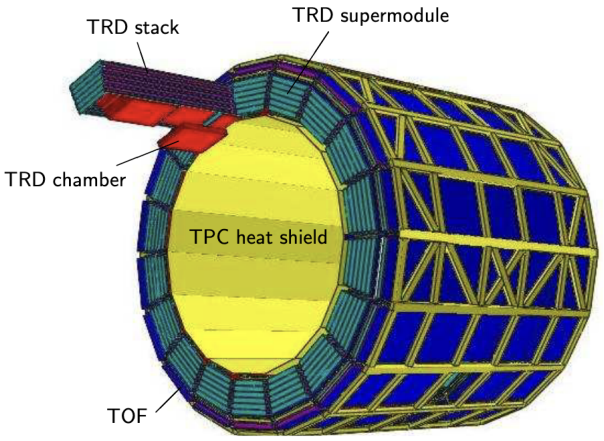


Figure 13: Diagram of the ALICE transition radiation detector, showing the chamber/stack/supermodule structure [19]

The transition radiation detector (TRD) lies beyond the TPC and supplements the particle positioning capabilities of the ITS and TPC. The main design purpose of the TRD is high-momentum electron identification, with a focus on differentiating electrons from pions, that contribute the largest background signal for electrons ("pion rejection") [19]. Due to pion contamination in the sample, especially at higher momenta, the TRD is especially helpful for the analysis of semileptonic decay products [22], making it of great interest to this analysis.

The full TRD is constructed of 540 individual chambers, arranged in stacks of 6 layers each. As illustrated in the diagram of Figure 13, 5 adjacent stacks is called a "supermodule", and 18 supermodules arranged in an approximate cylinder make up the TRD. Hits in each layer are called tracklets and can be examined in order to reconstruct the path of particle tracks through the detector.

As indicated by its name, the TRD works on the principle of transition radiation. When a charged particle passes (or "transitions") from one material to another, it has a chance to emit transition radiation in the form of a photon. Conveniently, pions do not emit this transition radiation, so it is easy to identify electrons that are recorded as having emitted such a photon. Since this radiation is probabilistic (not guaranteed to occur with every transition), to increase the likelihood of photon emission in each TRD chamber, particles first encounter and pass through a "radiator" made up of alternating materials. This increases the transitions of each particle, therefore increasing the emission likelihood [22].

3.4 Muon Detectors

Muons have some unique features that result in them being extremely convenient to identify and track through detectors. When charged particles move materials of higher density, the amount of energy they lose through ionizing their surroundings increases. Being heavier than electrons, muons are able to travel ("penetrate") much farther through dense material without losing all of their available energy. In addition, muons do not interact with the material they are moving through via any mechanism other than ionization, which is not true of other charged particles [6]. This property can be exploited in an experimental setup to achieve specific muon detection.

3.4.1 Muon Spectrometer

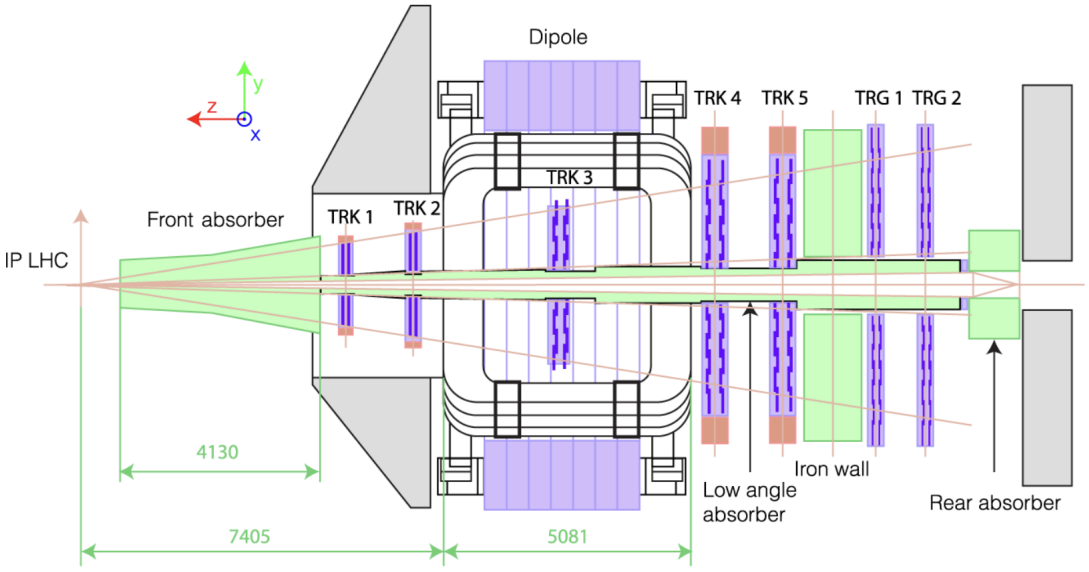


Figure 14: Diagram of the ALICE muon spectrometer [23]

The muon spectrometer at ALICE is specifically designed to take advantage of the penetrating power of muons. Before particles can reach the main body of the spectrometer they encounter the front absorber, which consists of a mixture of carbon and concrete. The absorber aims to prevent other particles, specifically hadrons and photons that might also survive until the forward region, from reaching the muon tracking chambers and contaminating the muon sample. As an additional measure to guard against radiation from the incoming collision beams, the entire length of the beam pipe that the spectrometer covers is surrounded by a beam absorber of tungsten, lead and stainless steel [19].

Past the front absorber are the five muon tracking chambers (MCH). Two chambers are placed on either side of the central dipole magnet, with the fifth existing within the magnet. The positioning of the MCH allows for the bend of the muon's trajectory due to the magnetic field to be analyzed. In Runs 1 and 2, the MCH was the only method of muon tracking available, but during Run 3 it has been supplemented with the addition of the muon forward tracker [24].

Finally, at the end of the muon spectrometer that is farthest away from the interaction point, particles passing through the detector encounter the muon filter and subsequently, the muon trigger system. The filter is made of iron and therefore has a higher density than the front absorber, which provides a final layer of protection against non-muons. This ensures that only high-energy muons are able to reach the trigger system, which identifies which events have a significant muon signal [23].

3.4.2 Muon Forward Tracker

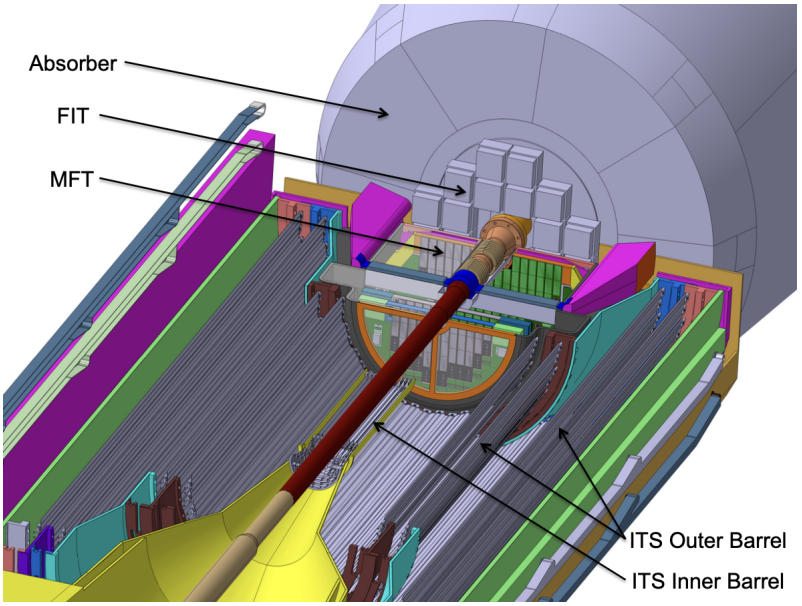


Figure 15: Diagram of the ALICE muon forward tracker's position in relation to the inner tracking system and muon spectrometer [24]

As previously introduced, the muon forward tracker (MFT), is a new installation of the ALICE detector for Run 3. It is a monolithic silicone pixel chip detector (similar in design to the new

ITS staves) that is placed before the muon absorber. As a result of this positioning, the MFT is able to carry out precise tracking of charged particles within the observable range of the muon spectrometer and in this way is used to improve the capabilities of muon tracking at ALICE. Though it cannot differentiate between muon tracks and those from other charged particles in the way that the spectrometer is able to, MFT tracks can be matched with the signals from the MCH in order to reconstruct the full path of muons (before and after interaction with the front absorber).

4 Run 3 Data Analysis Setup

Though it is not within the scope of this project to undertake a full analysis using Run 3 data, the steps that would need to be taken in order to produce such an analysis were investigated and tested. In this section, a description of how to go about a real data e-μ study is presented and initial tests are performed to confirm that a the study would be possible within the existing ALICE code base.

In addition, this section presents the setup of the ALICE software framework (O2) locally at UCT for the use of students within the UCT-ALICE group. The setup was developed over the course of this project. The process of doing so, along with the practicalities of using the setup, will be discussed. The structure of the software framework is also reviewed.

4.1 The O2 Software Framework

As of the LHC's third run, the ALICE experiment has transitioned to using a ROOT-based software framework called O2. The two O's stand for Online and Offline, as the framework is designed to handle ALICE's software requirements at every step in the experimental process. This means that not only is O2 used for handling data readout during the online period when collisions are occurring in the detector, it is also responsible for the offline processing of data and analysis of results [25].

O2 relies on manual, user-driven updates due to the constantly evolving code requirements of the experiment. There are over 2000 active members of ALICE, many of whom are actively using the software for their own analysis and code development. To avoid conflicts, especially those that would result in the loss of experimental work and data, each user is responsible for maintaining their O2 version and ensuring that any changes inherited from the main branch do not create issues in their own work. Users are then able to dynamically develop code for their own purposes, rebuilding and updating the software regularly.

4.1.1 Setup on the UCT HPC

O2 is supported in a limited capacity on several platforms, however installation can be challenging and demanding on local machines. The hardware recommendations are out of scope for the average student at UCT - a large amount of storage and memory is required to use it reliably. To mitigate these computing challenges while undertaking this report and in order to make O2 more accessible for use by future students in the UCT ALICE group, a period of time was dedicated to setting up a process by which it can be run on the UCT high processing cluster (HPC) [26].

It was decided that the optimal approach was through the containerization software Singularity [27]. Containers in this context are light-weight virtual operating systems that can be used dynamically after configuration. While there are many widely-used containerization platforms, the primary example being Docker, Singularity is generally the prime choice when working on an HPC. Docker executes commands through the Docker daemon, which requires root privileges to function [28]. This is not automatically an issue but in an HPC environment with many users, authorizing the Docker daemon to function can pose a serious security risk as it provides users the ability to obtain unrestricted access to the system [29]. Singularity is designed to be run by an unprivileged user, and therefore provides a convenient alternative that does not suffer the same risk.

However, there is a complication to using Singularity. Containers by default exist in read-only mode, and while theoretically it is possible to enter a container in a way that allows write-ability (using the "sandbox" mode), it requires `sudo` access, which nullifies the previously explained advantages for unprivileged users. While it is possible to use Singularity to simulate root permissions (using the flag `--fakeroot` when entering the container environment), this can be disabled by the HPC administrators for security reasons, as is the case at the UCT HPC. Therefore, the only way to use Singularity on the HPC is to first build the container on a local machine with full permissions, install the needed software (either by using a Singularity definition file or sandbox mode), and finally load a static version of the container onto the HPC in the form of a `.sif` file.

This presents a direct challenge for O2 usage. Considering the previously-described dynamic nature of the framework, one cannot simply load O2 onto a container - not only would this restrict the development of new analysis code in the UCT group, every time a user wished to update their version of O2 (often an unavoidable necessity due to incompatibility of old code with the newest data formats) it would require a new Singularity container to be built and O2 to be re-installed. Beyond this installing O2 locally, even within the virtual environment of a Singularity container, does not address some of the hardware issues originally encountered. In fact, this process would still require a large amount of storage space and processing power which the average student may not have access to.

The proposed solution was to create a custom container locally that would function as a virtual CentOS 7 machine (the primary supported platform for O2) and only setup `aliBuild` (the installation tool for ALICE) and it's prerequisites, following the steps in the relevant documentation [30]. The created file was moved to the `\scratch` folder on the HPC and an attempt was made to directly install O2 from a shell within the container.

Unfortunately, despite a promising start to the installation, nothing about this setup would ever work easily during the period of time that this procedure was being developed. The core of the issue was that during the installation process, O2 tries to make hard links between data and files in separate directories. In a Linux filesystem, a hard link points directly to the data being stored while its counterpoint (a soft link) is more indirect, pointing to another file which then in turn points to the data in storage [31]. The file system used on the HPC, `BeeGFS`, does not support hard links between different directories.

Many attempts were made to understand and find a work around for this obstacle. After much experimentation, a solution was found in the form of Singularity's persistent overlay feature. If

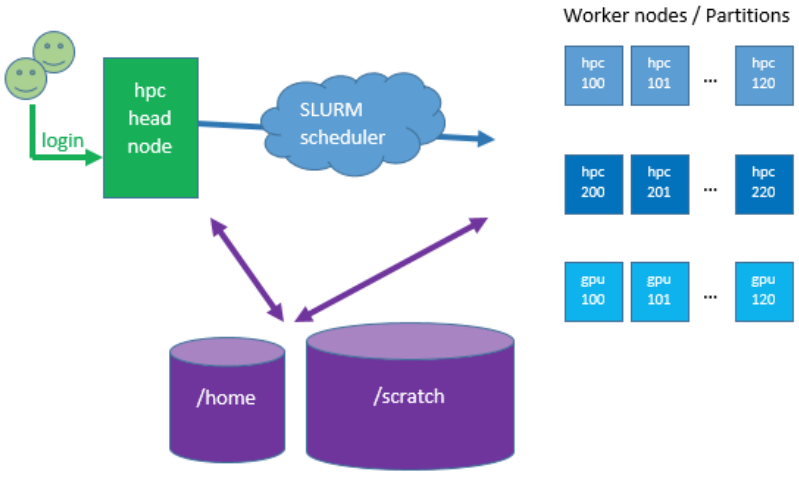


Figure 16: An illustration of the UCT HPC's node structure [26]

an external file system is available, it is possible to mount it to the Singularity container upon activation and use it instead of the default. In Linux, creating an empty file system is trivial with the `mkfs` command. Using the available tool, an `ext3` file system was created and after testing it was determined that when using it as an overlay, the container could be successfully used as if it was not a read-only object.

From within the Singularity container (now equipped with the empty `ext3` file system), it was then possible to continue with the installation of O2. The final piece in the process was explicitly directing `aliBuild` to use a folder in the new file system's `\home` directory as the "work directory" for O2. This folder is not visible to the user from outside of the container, but is easily accessible within it and can be used to access the O2 environment. In this way, the hard link limitations of `BeeGFS` were bypassed and O2 was successfully installed.

4.1.2 Using O2 on the HPC

Detailed instructions for setting up an O2 build on the HPC, as well as the relevant Singularity container file, can be found on a linked github repository [32]. A brief description of how to access and run O2 once these instructions have been completed will now follow.

The HPC is a complex multi-user system, with specific rules that must be respected when using it. A layout of the HPC architecture is displayed in Figure 16 below. Upon logging in, a user is directed to the "head node", the primary Linux server of the cluster. Any processing job run on the HPC must be run on one of the worker nodes, so it is crucial to never run any Singularity or O2 commands without first moving onto one of these secondary servers. As shown in the figure, this is achieved through the use of SLURM (Simple Linux Utility for Resource Management), a scheduler that manages the optimal allocation of nodes based on job requests [33].

When running O2, users should either submit resource requests through the `sbatch` command, which allows for the submission of prewritten bash scripts, or create a bash terminal on a worker node using the `srun` command. This second method is more interactive, and therefore often

more convenient when working on code development in O2. In either case, it is important to request resources (such as cores and time needed) based on estimated usage. Once resources have been assigned and the worker node has been accessed, users can enter the Singularity container and from there, the O2 environment.

The separation of the working directory as described previously allows for the user to view and edit their analysis code upon logging in to the HPC, without needing to move to a worker node to avoid inducing load on the head node. However, building and re-building O2 is only possible when the work directory is accessible, and therefore users must be within the Singularity environment to compile their code and test any changes made. Code compilation can be achieved using the O2 `ninja` tool [25].

It should be noted that when working with O2 on the HPC, the chosen storage location is `\scratch`. The main benefit of using `\scratch` is the space associated with it. In a user's `\home` folder one can only access 20 GB of storage space, while on `\scratch` one has 100 GB by default (and this can be increased on request). Considering the storage requirements of O2, this is extremely desirable. As the `\scratch` folder is not backed up and is intended to be treated as temporary storage, the regular use of Git is highly recommended to ensure work is not lost (a general good practice).

4.1.3 O2 Data Structure

When trying to understand the way that data is structured in O2, it is important to look at the wider context of the transition from Run 2 to Run 3. O2 which was introduced as a successor for AliPhysics (the software framework utilized by ALICE throughout Runs 1 and 2) and includes a complete restructuring of how data is stored. The design decisions that were made are a direct consequence of the increased requirements associated with the higher luminosities of Run 3. In previous years of operation, data was stored on a "hierarchical" event-by-event basis [34], which is not possible in Run 3 due to a key change in how data is captured. Instead of triggering based on individual events, ALICE now triggers for entire bunch crossings of particles (BCs). This is known as "continuous data taking". Since events are no longer the default unit of classification, data is now grouped and stored in "timeframes" of 10 ms.

Timeframe data is processed and stored as a series of tables in Analysis Object Data (AOD) ROOT files. These tables can be treated similarly to a relational database and many of the relational operations associated with typical database handling have been reproduced for use by analyzers in O2. Keys (known as indexes in O2) form the links between objects in different tables and are used to relate particle tracks back to the event that they came from. These events in turn have indexes that relate them to a specific bunch crossing, etc.

Figure 17 details the most important of the table relationships, with arrows representing the indexes. The outlined boxes represent groups of tables that contain information about the same objects and can be joined together in different ways. To understand how this applies practically, one can look at the `Tracks` and `FwdTracks` tables. They both represent particle tracks that have been recorded in the detector. Each track in these tables has an index linking it to an event in the `Collisions` table that, during reconstruction, it was classified as originating from.

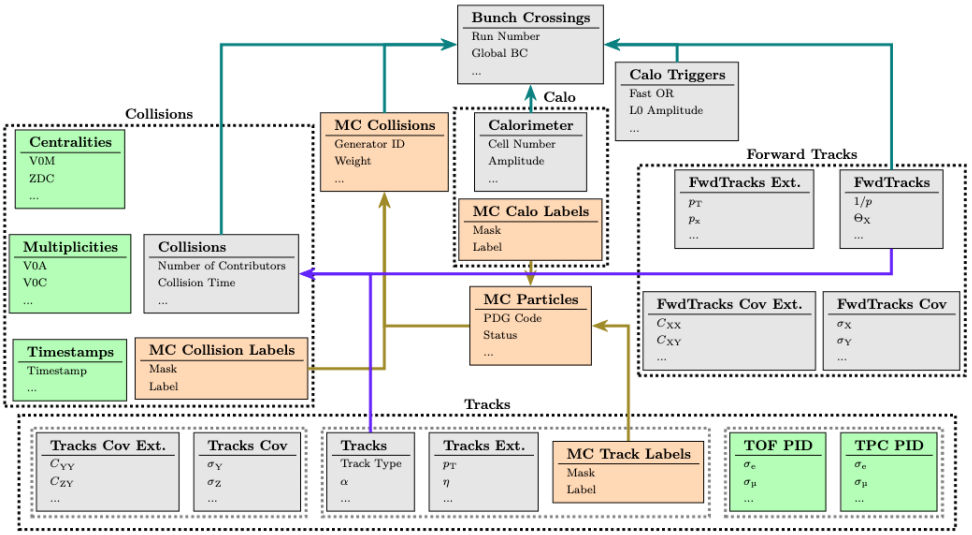


Figure 17: An overview of the major table data structures in O2 and the links between different table groups [34]

However, as shown by their grouping in the figure, they are stored fundamentally as separate types of objects. `Tracks` stores track information from the central barrel detectors, while `FwdTracks` does the same for the forward arm/muon detectors and as such each has their own requirements. The other tables in each group contain additional information that may or may not be necessary, such as particle identification from the TOF and TPC detectors for central barrel tracks. In this way every analysis task can choose what tables to use and not suffer the loss of efficiency that would come from processing unnecessary information.

It should be noted that in Figure 17 there are some tables and links that have been omitted - for example, there are tables of "ambiguous" tracks and forward tracks. These are tracks that met the detector requirements to be classified as a particle instead of background noise, but could not be cleanly reconstructed to belong to one specific collision event. Since data is no longer stored on an event-by-event basis, there is an inherent uncertainty in classifying particles to events. Even tracks not stored in the ambiguous tables are affected by this. A process known as event mixing is typically applied to account for reconstruction ambiguity when it comes to particle correlations. It involves intentionally matching pairs of particles that have been classified as coming from different events, using the resulting false signal as a type of background subtraction.

Other relational concepts that can be used when writing an analysis task in O2 include partitioning, which can be used to split a table into smaller components based on a given criteria (such as particle charge or position in the detector) and filtering, the primary method of applying cuts in an analysis task. Filtering a table follows the same logic as a relational selection operator. A user can define multiple filters based on different table parameters and then instruct O2 to apply them all at the processing step of their code.

O2's analysis framework is complex and multifaceted, and there are many tools that have not been outlined here due to a lack of relevance to the analysis tasks referenced in this report. For

further information, the O2 documentation [25] can be referred to along with the latest edition of the O2 analysis tutorials, organized on a semi-regular basis by the members of the ALICE experiment.

4.1.4 The DQ Analysis Framework

The ALICE experiment is divided into several Physics Working Groups (PWGs), each with their own area of focus. One of these groups is PWG-DQ, focusing on dilepton and quarkonia analysis. The DQ group has developed its own specialization of the O2 analysis framework to accommodate the specific needs of dilepton analyses. Electron-muon correlations are a dilepton probe, so this is the optimal environment to investigate the pairs. There is already some code infrastructure available to facilitate the analysis of the e-μ channel incorporated into the DQ framework.

The DQ framework introduces a series of new "reduced" tables into the data model. These tables do not replace their O2 equivalents, instead acting as a follow-on step in data processing before the actual analysis is performed. Broad cuts are chosen depending on the type of analysis to be done and are provided as a configuration input to the `TableMaker` task. The list of available cuts, which are specialized for DQ analysis topics, can be found in the DQ `CutsLibrary` [35].

From here, the original O2 tables are "skimmed" such that only events and tracks that pass the inputted cuts exist in the dataset. Due to the smaller size of the now-reduced tables, they are much more manageable in an analysis context. This means that a user undertaking an analysis has the ability to re-run their desired analysis task in a way that is more time and resource efficient than it would have been when using the original tables, allowing experimentation for tuning cuts, etc.

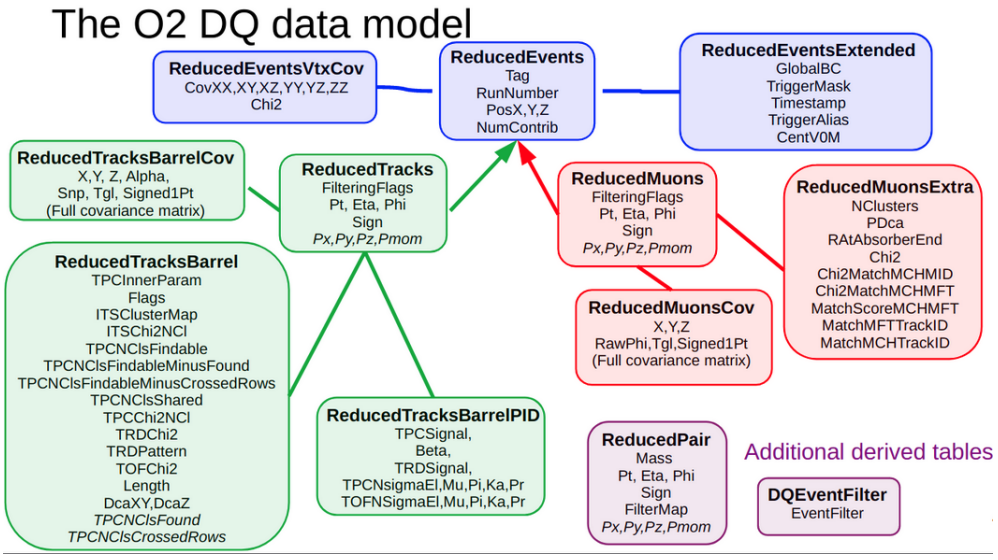


Figure 18: The reduced table structure of the DQ framework within O2 [35]

4.2 Analysis Task Requirements

As mentioned, there is already a framework in which to undertake an e-μ analysis in the DQ code base. Ideally, the `TableReader` task [35] would be used with the correct configuration, and e-μ specific cuts would be developed in line with DQ conventions and added to the current library. In this section, the basic components of how to process reduced DQ tables in a task such as `TableReader` are reviewed and the specifics that would need to apply in the e-μ case are discussed.

Before the analysis itself is undertaken, it is necessary to run through all elements of the reduced tables and classify which entries pass the chosen cut requirements. To accommodate this, the analysis task is separated into functions called `structs`, which are executed in a specific order as specified by the user. There are three selections that are relevant in the e-μ case - event selection, barrel track selection and muon track selection.

Each of the selection `structs` in the DQ framework return a vector that corresponds to the length of the table being cut (either events, barrel tracks or muons) and can be joined to these tables in a new "selected" data type. As the `struct` executes, it assigns a numerical value to each entry in the table based on whether it passes each selection criteria (a positive value is assigned to those that do). In this way, the success or failure of a single element to pass multiple desired cuts can be assessed with a single check during analysis, simplifying the process.

Event cuts for proton-proton events are generally only made on the z-axis position of the collision and the performance of the detectors while the run was taking place. Barrel track cuts in the e-μ case would need to include particle identification requirements to choose an electron sample. This is also where the associated p_T cut could be performed, as well as general quality cuts on the barrel tracks themselves, including track matching information between the various barrel subdetectors. Similarly, the muon cuts would need to incorporate a trigger p_T cut and muon track quality/matching.

During the sample analysis, a filter which requires reduced tables to have a positive selection flag can be created and applied to the reduced tables joined to their selection vectors. The surviving barrel and muon tracks of all events that passed the cut requirements are then processed. In order to analyze e-μ pairs, the task then needs propagate over combinations of barrel and muon tracks in the filtered events.

An additional step that is required in this type of data analysis is event mixing, a process which is also facilitated in the DQ framework. Event mixing is a method which aims to removes ambiguous tracks and background by deliberately incorrectly pairing tracks from different events. This would result in a purposefully uncorrelated background measurement which could then be subtracted from the e-μ spectrum to assist in removing noise from the signal.

4.3 Testing the Analysis Task

In order to gauge whether an e-μ analysis can be implemented and executed with relative ease within the existing ALICE software, a simplified adaptation of the `TableReader` analysis task was written and locally tested on a small dataset skimmed using `TableMaker` from a single run in 2022. The original dataset contained 4350056 p-p collisions. A basic event cut was applied to se-

lect collisions with a primary vertex within 10cm of the central interaction point along the z-axis.

During data skimming, initial cuts were made on the pseudorapidity of the barrel and forward tracks to ensure that they lay within the valid range of the ALICE detector. An additional "matching" cut was made on the forward region data to ensure that selected tracks had been found in both the muon chambers and the muon identification region.

Due to the lack of e-μ specific cuts in the `CutsLibrary`, local cuts were created within the task according to DQ convention. The barrel and muon selections as described in section 4.2 were implemented and tested, though event selection and mixing were not included in the analysis task. For both barrel and forward tracks, the local cut was designed to allow a user to adjust track p_T in a separate configuration file. The barrel track cut additionally contained the option for election identification cuts to be made using the TPC, though due to the size of the dataset this option was not used in the production of the figures presented below.

The test version of the analysis task (available at the code repository for this project [36]) was indeed able to run successfully, though due to the small size of the dataset it must be emphasized that no analysis was attempted or would be possible using the obtained output. It is expected that the test dataset is composed entirely of background, with no signal visible. However in order demonstrate the initial functionality, Figures 19, 20 and 21 are presented.

In Figure 19, the transverse momenta of potential pair partners are displayed. All tracks recorded in the histogram come from events in which signals were picked up in both the central barrel and muon detectors, therefore having the possibility of coming from a shared parent quark pair. The azimuthal correlation histograms for these tracks are displayed in figure 20. In the top subplot, all correlations are considered with no regard for the electric charge of the constituent leptons. This is then categorized into two more subplots, displaying the like- and opposite-signed pairs independently.

Though no cuts were applied to the tracks in Figures 19 and 20, it is possible to impose restrictions on the transverse momenta of the barrel and muon tracks within the `emuConfig.json` configuration file (refer to [36]). The azimuthal correlation histograms of the pairs with such a momentum cut imposed are shown in Figure 21. There are also cuts related to particle identification available to be configured.

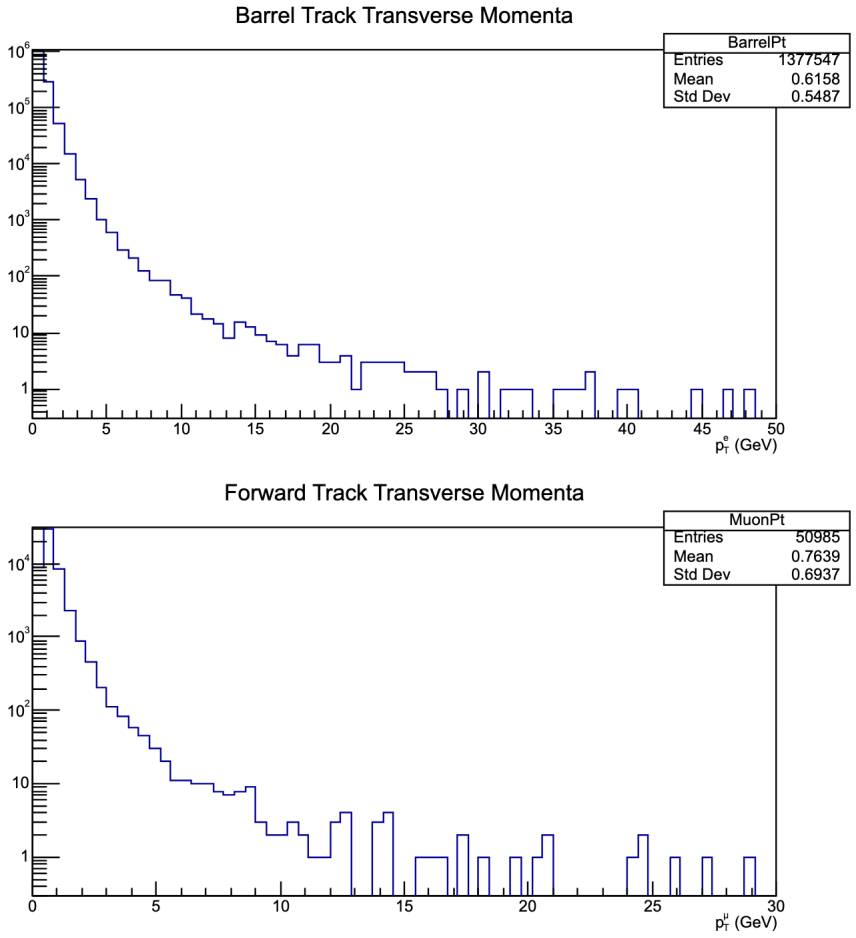


Figure 19: Transverse momenta of reduced barrel and muon detector tracks in a single 2022 run

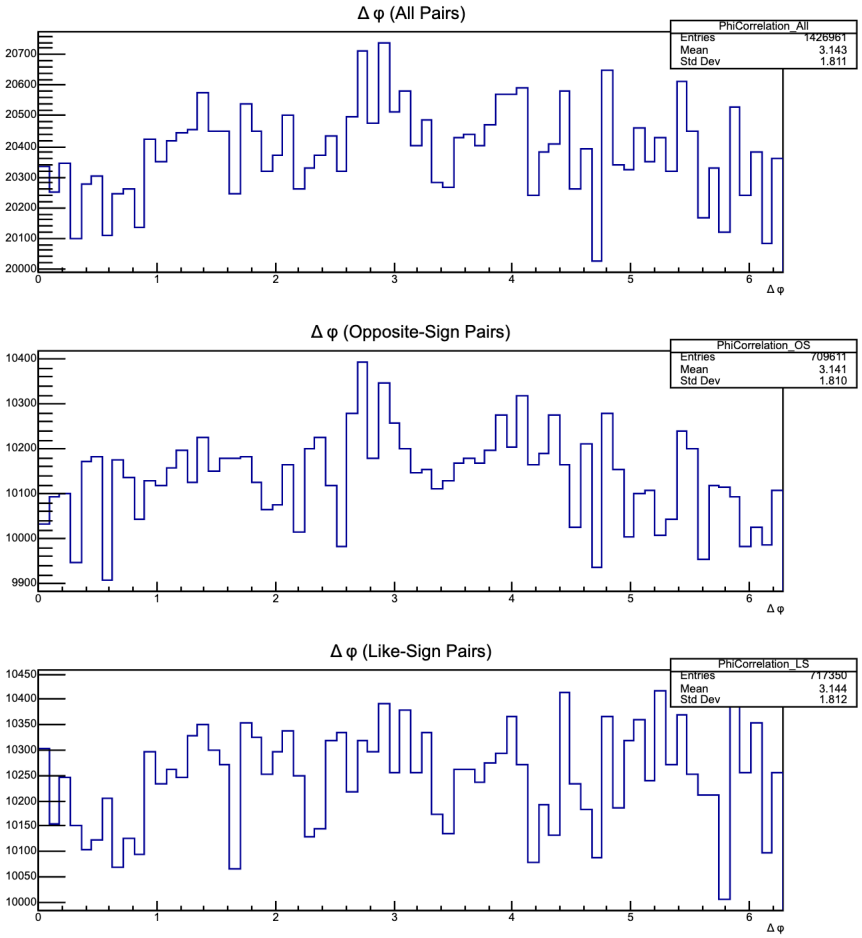


Figure 20: Azimuthal correlation between reduced barrel and muon detector tracks in a single 2022 run

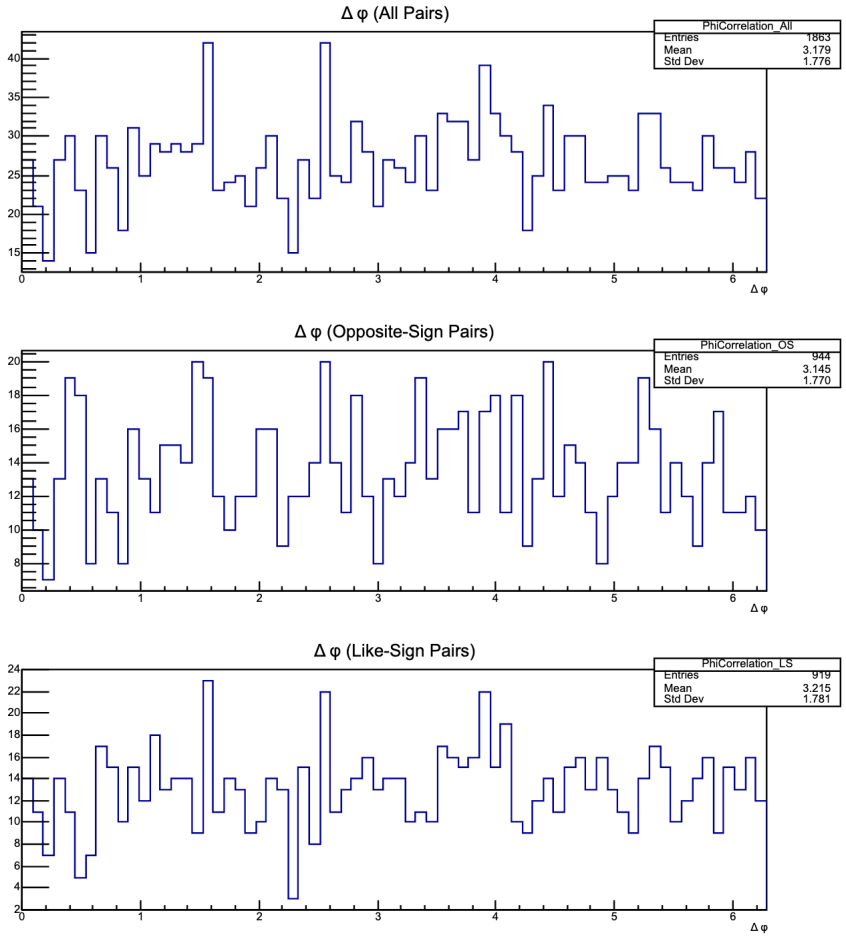


Figure 21: Azimuthal correlation between reduced barrel and muon detector tracks in a single 2022 run, with an imposed cut requirement of $p_T^e \geq 1$ GeV and $p_T^\mu \geq 3$ GeV

5 Simulations

As discussed in section 2.4, a major limiting factor of using the e-μ channel is the low rate of pair production compared to other dilepton heavy quark probes. While the LHC's third run will provide the opportunity to the study a larger sample of data than in previous experimental runs, the kinematic acceptance range of the ALICE detector (as quoted in section 3.2) and the p_T cuts required to increase sample purity further reduce the available statistics. Before undertaking an e-μ correlation study using Run 3 data, it is necessary to understand to what extent these limitations will further reduce the sample of data. How many heavy quark pairs are produced that can be detected in the form of an e-μ pair, considering the detector restrictions? Gauging this will allow for an estimation of the quantity of real data that would need to be analyzed to see a viable e-μ signal and from there a statement can be made about the feasibility of the method.

To achieve this, simulated events must be used. In high-energy particle physics, there is an inherent restriction on what can be directly measured in a detector due to the time and distance

scales of particle decays. Therefore, the data output of such experiments comes in the form of the surviving decay products of the particles created during any given collision. Detailed particle identification and reconstruction must then be undertaken to determine, within some degree of uncertainty, the type of particle that the detector recorded, where it originated from within the machine and what parent particles may have originally existed. Simulated events do not have this restriction - since it is possible to track the "particles" through every stage of the decay process, a full breakdown of every particle's properties and history can be extracted.

Particle collision simulations are based on the current understanding of the Standard Model and are therefore not infallible. They can only provide information to a degree of accuracy which reflects the success of theoretical models in describing underlying physical processes. However, with an understanding of the degree of success that these theoretical models achieve in describing experimental results, an expectation can be set for what may be seen in data. Specifically, for this study, the goal is to gain an understanding of

(i) how many p-p collisions produce heavy quark pairs that decay into oppositely-signed e-μ pairs,

(ii) how often these e-μ pairs fit the criteria of $\eta_e \leq |0.9|$ and $-4 \leq \eta_\mu \leq -2.5$ necessitated by the ALICE detector geometry and

(iii) how much this sample is reduced by cuts on the p_T of the trigger and associated final-state particles.

These restrictions will be used to build a reduced cross section of quark pairs that have a viable chance of being measured by the ALICE detector in the form of e-μ pairs.

5.1 Pythia

The simulations run throughout this analysis use Pythia, a full event generation software originally developed by researchers from Lund University [37]. The usefulness and importance of using simulations in high-energy particle physics has been discussed, but it is also necessary to understand how these simulations are produced and in this way provide context for the reliability, benefits and limitations of any analysis that makes use of them.

5.1.1 Random Sampling

Pythia uses Monte Carlo methods, a group of computational techniques characterized by the random sampling of theoretical probability distributions. Mathematical descriptions of the standard model provide the current best estimate of the distributions that describe the likelihood of a given physical processes occurring during a particle collision. Knowledge of these probabilities, coupled with the inherent randomness of Monte Carlo methodologies, allows for a computational picture to be constructed that is reminiscent of what is observed in real collisions. This approach of harnessing random number generation is at the core of Pythia's event generation.

However, the complexity of high energy particle collisions cannot be overstated. In order to succeed at reflecting real physical effects, it is necessary to use a large parameter space with many variables incorporated. This is where classic Monte Carlo methods run into one of their major limitations: the "curse of dimensionality". When dealing with multidimensional systems

such as these, the scaling of the problem and its associated processing requirements becomes exponentially large [38].

Because of this effect, independent sampling of the distributions becomes unwieldy and impractical. To remedy this, Pythia employs a method known as Markov Chain Monte Carlo (MCMC) [39]. This method removes the requirement for statistical independence when performing random generation - instead, each new parameter is dependent on the previous one, creating a "chain of dependence" known as a Markov chain [40].

The final component to be mentioned are the various "tunes" that can be set in Pythia. While the MCMC method takes into account theoretical predictions of nature, these distributions alone are not a complete descriptor for all interactions within a collision. To account for this Pythia uses a pre-set parameter space, called a tune, to adjust its output and best reflect real-world observations [39]. There are several of these tunes available, each one taking into consideration the event data from a different experiment. The simulations run for this analysis utilized the default Pythia tune (`Monash 2013`).

5.1.2 Simulated Events

To understand Pythia's inner workings, it is useful to break down the basics of what occurs when an event is generated. For the purposes of this review, a basic overview of the major components is provided, but it should be noted that many other complex interactions occur throughout the process.

Contrary to what might be expected, Pythia's first step is not to generate the first process that would happen linearly in time. Instead, the software begins with what is known as the "hardest process". "Hardness" in this context is complex and linked to the definition of hard and soft processes as described in section 2.2.3. It is also a measure of the degree to which particles and the subsequent processes they are involved in are mathematically virtual according to the Heisenberg uncertainty principle [15, 6].

For the purposes of this analysis and in order to quantify what sets apart the hardest process as it is discussed here, it can generally be thought of as the most energetic subprocess that occurs as a result of interactions between the quarks within the two colliding protons ("partons"). The transverse momentum of this hardest process is denoted $\hat{p}_T$. The energy involved in this process is variable and dependent on what fraction of each proton's momentum (p_z) is carried by the interacting partons that are involved.

Pythia events are classified according to the nature of the hardest process (stored by the software as the event's "code"), and users are able to specify which category of events are desired for any given simulation [39] based on this classification. Further details on the options that are available, as well as the choice made for this study, will be discussed in section 5.2.1.

After this hardest process is generated and the event is defined, Pythia moves on to simulating processes that exist at increasingly lower hardness scales. This is achieved through what is known as a parton shower [15]. In Pythia, parton shower processes can be split into two distinct classifications - those that occurred before the hardest process on a scale of linear time (known as the initial state radiation or ISR), and those that occurred afterwards (known as the final

state radiation or FSR) [39].

ISR processes generally come from the internal interactions of each colliding proton. Then as previously discussed, the interactions between partons in the two separate protons result in the hardest process, which has already been determined as the basis of the simulated event. It should be noted that during this stage of parton interaction, it is possible for the other partons to interact in ways that are not the event's hardest process - these are known as multi-parton interactions (or MPIs) and while they are not accounted for in depth during this study, they are a major complicating factor in Pythia and come with their own set of user input options. Finally, the hardest process and other hard scatterings lead to further particle production, which characterizes FSR interactions [39].

Once all partonic interactions are defined for an event, Pythia must begin to simulate hadronization (as outlined in section 2.2.4). Pythia uses the Lund String Model to describe the hadronization process. [39, 8]. The first hadrons that form often do not exist in a stable state and so may decay into a range of child particles. Pythia continues to simulate these decay processes until all remaining particles exist in their "final state". These final state particles are analogous to the particles that, in real events, reach the detector and therefore have a chance of being recorded in an experiment.

5.1.3 Particle Information

With this understanding of how Pythia approaches event generation, the output provided to the user can be discussed. The main channel of information in high-energy physics simulations is called an event listing (or record). In Table 1, a selection of rows taken from a full Pythia event listing are displayed.

no.	**id**	**name**	**status**	**mothers**		**daughters**		**colors**		p_x	p_y	p_z	**e**	**m**
0	90	(system)	-11	0	0	0	0	0	0	0	0	0	13700	13700
1	2212	(p+)	-12	0	0	93	0	0	0	0	0	6850	6850	0.938
2	2212	(p+)	-12	0	0	94	0	0	0	0	0	-6850	6850	0.938
							...							
5	4	(c)	-23	3	4	9	9	101	0	4.409	3.936	-89.004	89.212	1.500
6	-4	(cbar)	-23	3	4	10	10	0	103	-4.409	-3.936	-60.131	60.440	1.500
							...							
311	321	K+	91	148	0	0	0	0	0	-1.519	-1.579	-24.719	24.821	0.494
312	-211	pi-	91	148	0	0	0	0	0	-0.026	-0.002	-0.348	0.375	0.140

Table 1: Sample of the Pythia event listing from one p-p event used in this study

There are several parameters of note to discuss in Table 1. The first three columns are straightforward, they contain the particle's position in the event listing and its identification (both in the form of a PDG classification code or "id", [41] and written out by name). The fourth column, "status", is how Pythia classifies the particle's production mechanism, and a detailed breakdown of what the various status codes indicate is available in the online Pythia documentation [37].

These status codes can be interpreted by breaking them into two parts, the numerical value and the sign (positive or negative). The numerical value indicates the manner in which the particle was produced while the sign tells the user whether it is in its final state (positive means that it is, negative means it decays further). In this way the status code of the particle can be used to extract an overview of the particle's history. For example, the charm and anti-charm quark displayed in the table (no. 5 and 6) both have the numerical status `23`, which indicates that they are the outgoing particles of the event's hardest process. In contrast the last two row contains a final state pion, with the specific numerical value of `91` indicating it was produced in a "normal decay" as opposed to a resonance or oscillation.

With the nature of the production process covered by the status code, the next column ("mothers") displays the location in the event listing of the specific parent particle(s) which underwent that process. The "daughters" column follows on from this, and provides the event listing location of the immediate products of a particle's decay where applicable. It is possible to use this information to trace the lifetime of any particle in the event, working back through its production history until the original protons are encountered or forward until a final state particle is found.

In the outputted event listing, Pythia displays a maximum of two mothers or daughters in each of these columns, but in actuality there are sometimes more than two in existence. The full list can be obtained in the form of a vector. A useful feature provided by Pythia is the user's ability to also return a complete vector of daughters produced by a particle past its immediate decay products (so including the products of subsequent decays). This method was extremely useful throughout this analysis, as it facilitated the identification of heavy quarks that decayed into final-state electrons and muons. However, since recursion is used to achieve this, it should be noted that the vector will likely contain repeats of particles. To avoid double counting in this analysis and obtain a list populated by unique particles, the vector (which contains a list of positions in the event listing) would be sorted in ascending numerical order, with the duplicates being removed using the `C++ erase` method.

The "color" column contains information about the color and anti-color charge of the particles, and as such only quarks and gluons will be listed as non-zero. The next four columns, "p_x", "p_y", "p_z" and "e" are the components of the particle's 4-vector momentum, both in physical terms and computationally. Pythia has a dedicated class, `Vec4`, for the handling of 4 vectors. Finally, the mass of the particle is provided.

Other particle properties, such as the important kinematic coordinates (azimuthal angle and pseudorapidity, as discussed in section 3.1) are not outputted in the default listing but can still be accessed. The charge of the particle can also be returned. The full list of methods available to access particle properties can be found in Pythia's online documentation [37].

5.1.4 Event Information

In addition to individual particle properties, Pythia also provides information about the event in general. Some already-discussed examples of this include the value of an event's $\hat{p}_T$ and its code classification based on the hardest generated process. When preparing a Pythia simulation, it is possible for a user to set a $\hat{p}_T$ range for the events being generated, defining bounds for the

minimum and maximum hardness that the simulation can exist within.

Another important piece of event information is the generated cross section, `sigmaGen` (denoted σ_{gen} in future equations) and its associated uncertainty `sigmaErr` (σ_{err}). This is a measure of how likely the generated event is to have occurred out of all possible outcomes in a proton-proton interaction. In this way it allows for comparison with observations that would be made in real data. This is crucial when simulating specific reactions/processes in Pythia, as will be done in this analysis. Scaling each event by an individual cross section and uncertainty is unwieldy, so once the requested number of events have been generated Pythia provides a total value of $\sigma_{gen} \pm \sigma_{err}$ that can be applied to all events within the defined parameters.

5.2 Simulation Setup

5.2.1 Process Selection

As mentioned in section 5.1.2, Pythia relies on user input when making decisions about the nature of a simulated event. An important variable input is the classification of the hardest process, so which code is to be generated. The codes `121 - 124` have been reserved for heavy flavor quark processes. To decrease the time needed to run the relevant simulations and increase the likelihood that any single simulated event will result in a relevant e-μ pair being produced, the flags `HardQCD:hardccbar` and `HardQCD:hardbbbar` were used throughout this analysis.

However, these processes can not be turned on without consideration. In Pythia, processes are classified as either hard or soft in alignment with the factorization of the QCD cross section as discussed in section 2.2.3. This analysis is looking at hard QCD processes and as such, at low $\hat{p}_T$ values the cross section of the processes diverges and the perturbative solution no longer holds. Therefore, a user cannot turn on such processes and generate simulated events across the entire range of $\hat{p}_T$ without consideration.

There are a few ways to deal with this complication. It is possible to set Pythia to simulate different process types at different hardness scales, so one option is to turn on the `softQCD` flag below some threshold value of $\hat{p}_T$. Once that threshold has been reached, the flag can be turned off and the desired `hardQCD` process can be switched on.

However, since this study is focused on evaluating hard QCD processes, the soft QCD region of the cross section holds far less interest. In addition, since the quark pairs being probed come from the hardest process, the transverse momentum of each quark in the pair is equivalent to the value of $\hat{p}_T$. Quarks with higher energies are more likely to produce high p_T leptons, which consequently increases the chance that these leptons will pass the high p_T cuts that make up the third requirement for an "observable" pair as introduced at the beginning of this section. This means that harder events are more desirable in this study, so the decision was made to simply choose a minimum value of $\hat{p}_T$ above 0 and not generate events in the soft QCD region.

To make this choice of minimum $\hat{p}_T$, a comparison was made between the soft and hard QCD cross sections, following the procedure laid out in previous investigations [15], now at 13.6 TeV.

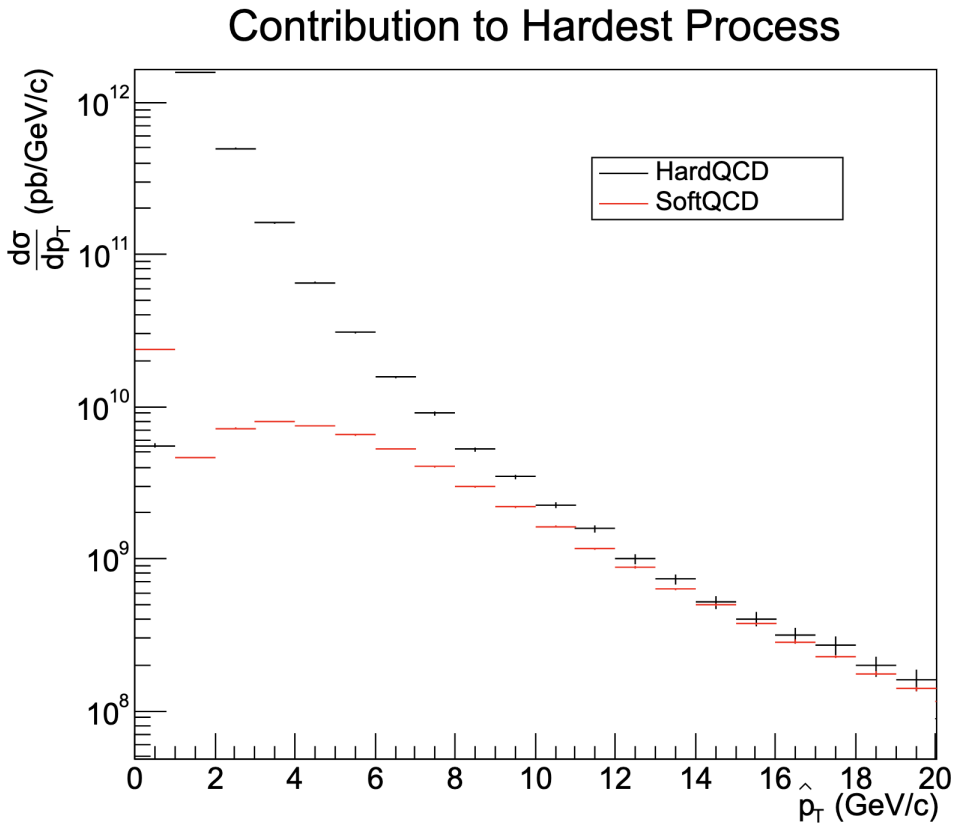


Figure 22: A comparison of the Pythia cross sections for hard and soft QCD processes at a collision energy of 13.6 TeV

The point of intersection between the two cross sections can be taken as the scale at which soft QCD no longer dominates, and it is reasonable to generate exclusively hard QCD processes. In Figure 22, this point is observed to be at 14 GeV.

5.2.2 Hard Process Binning

In addition to setting a minimum $\hat{p}_T$, it is also possible to set a maximum. Using these two parameters (in Pythia called `PhaseSpace:pTHatMin` and `PhaseSpace:pTHatMax`), a user can define a hardness "bin" within which Pythia is required to simulate events. This bin could span the entire desired range of $\hat{p}_T$, but the range can also be divided into several $\hat{p}_T$ bins of various sizes.

Approaching the simulation in this way is useful, as it allows for a cross section to be constructed that has comparable statistics in each bin. Very hard events are relatively rare so when Pythia is allowed to simulate over a wide range of $\hat{p}_T$ values, the likelihood of a higher $\hat{p}_T$ event being generated is low, causing limited statistics at harder scales. An additional issue is the reduction in statistics of quarks with lower values of $\hat{p}_T$ once the high lepton p_T cuts are imposed. These problems are well handled by the binning as described.

In order to reasonably compare the different bins, certain transformations must be applied after

the simulation has completed. Pythia is now simulating the same number of events in each bin, but the cross section of the processes at various hardness scales is quite different. This means that if the data points from different bins are simply combined in a histogram, a significant disconnect will be observed and no reasonable analysis can be performed. It is therefore necessary to scale each bin by its luminosity, which takes into account the number of events as well as the cross section as provided by Pythia's `sigmaGen`. In general,

$$N_{events} = L \times \sigma \tag{5}$$

Since the Pythia cross section is provided in units of millibarn, it should be noted that appropriate scaling has been done done to represent the cross section in units of inverse picobarn (pb^{-1}). The luminosity of each bin can be expressed as follows:

$$L_{bin} = \frac{N_{events}}{\sigma_{gen}} \tag{6}$$

The bins are then normalized using this value of L_{bin}. In terms of the `ROOT` code, this is achieved by multiplying the bin contents with a scaling factor s, where

$$s = \frac{1}{L_{bin}} = \frac{\sigma_{gen}}{N_{events}} \tag{7}$$

5.2.3 Cross Section Uncertainty Contributions

Just as Pythia provides a value for the cross section of each $\hat{p}_T$ bin, so too does it store the associated uncertainties as discussed in section 5.1.4. When the various histogram bins are normalized by their luminosity, `ROOT` automatically recalculates the errors of the histogram according to the scale but does not consider this additional uncertainty from Pythia. As such, an effort was undertaken to manually propagate the σ_{gen} uncertainties into the histogram bins to determine their impact on the overall histogram errors.

When scaling a value, the uncertainty must be scaled proportionally. In this way, the uncertainty of the scaling factor s in equation 7 becomes

$$\delta\left(s\right) = \delta\left(\frac{1}{L_{bin}}\right) = \left(\frac{1}{N_{events}}\right)\sigma_{err} \tag{8}$$

If this is to be taken into account, the automatic error recalculation of each histogram bin would need to be propagated with the scaling factor error as follows:

$$\delta\left(H_{bin} \times s\right) = \left(H_{bin} \times s\right)\sqrt{\left(\frac{\delta H_{bin}}{H_{bin}}\right)^2 + \left(\frac{\delta s}{s}\right)^2} \tag{9}$$

where H_{bin} is the number of histogram entries in any given bin, and the scaling factor is dependent on the specific value of L_{bin}.

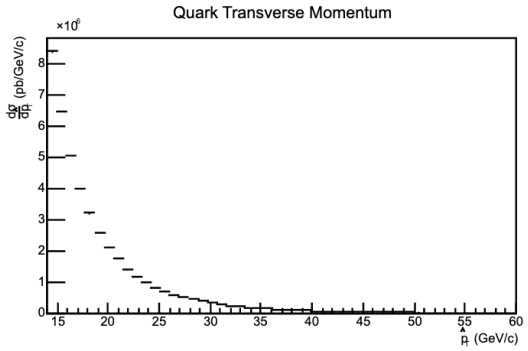
Therefore, to determine if the uncertainty σ_{err} will have a significant impact on the histogram's overall errors, it is necessary to compare the two ratio values, $\left(\frac{\delta H_{bin}}{H_{bin}}\right)^2$ and $\left(\frac{\delta s}{s}\right)^2$. Due to the large number of events being simulated in each bin (on the order of magnitude of 10^9), it was found that the ratio $\left(\frac{\delta H_{bin}}{H_{bin}}\right)^2$ was the significant contributor to this uncertainty. In comparison,

the contribution from $\left(\frac{\delta s}{s}\right)^2$ was negligible. It was therefore decided that fully propagating the errors from the Pythia cross section was unnecessary.

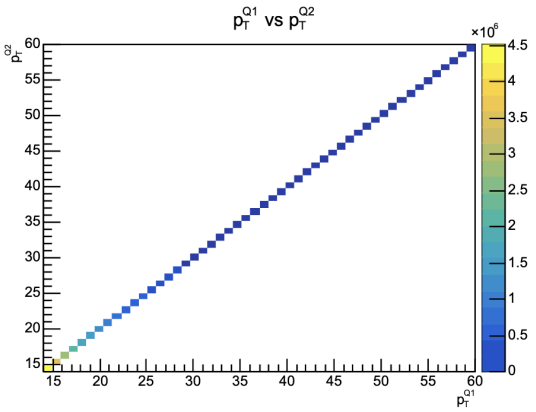
5.3 Quark Kinematics

5.3.1 Quark Pairs

In this section, the relevant kinematic coordinates (p_T, ϕ and η) of the heavy quark pairs are analyzed. In each case, a 1D distribution is displayed first to demonstrate the behavior of the quarks individually (with both quarks in the pair being counted as distinct particles). Accompanying these are 2D representations comparing the properties of the pairs, in order to examine the relationships between the quark pair partners and identify any correlations between the two particles.



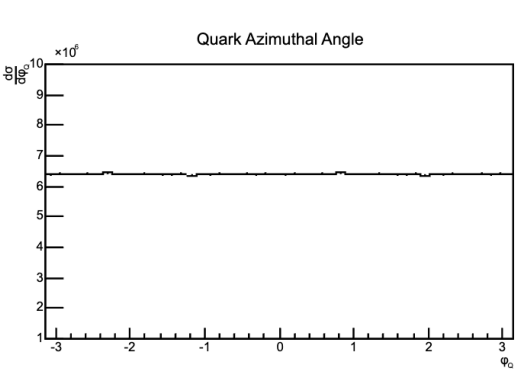
(a)



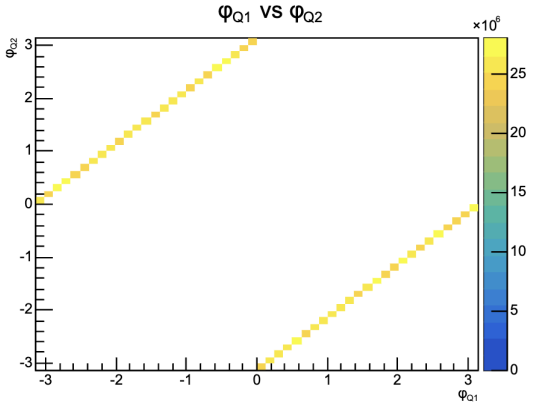
(b)

Figure 23: Histograms showing the transverse momenta of quarks created during Pythia's hardest generated process

The Pythia convention of denoting the transverse momentum of the hardest process as $\hat{p}_T$ has been continued here, though from now on it will also be used to signify the transverse momentum of the heavy quarks (as these are equivalent by definition). The distribution in subfigure 23a decreases exponentially (note the logarithmic scaling), indicating that the majority of quarks are produced at lower energy scales, which fits with expectation. This decrease is reflected in the changing color grade of the z-axis in subfigure 23b. Both quarks in each pair are produced at the same energy, hence the clean correlation between the x and y axis of the 2D distribution.



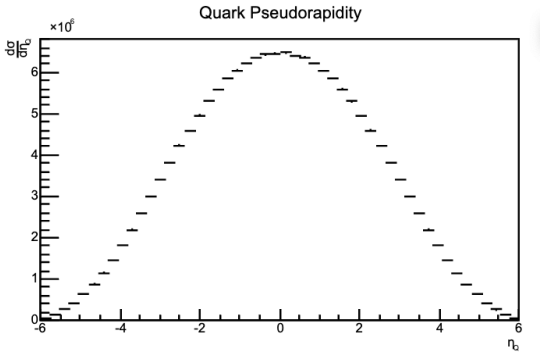
(a)



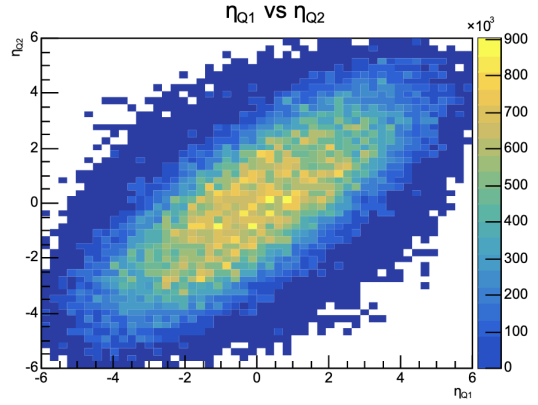
(b)

Figure 24: Histograms showing the azimuthal angles of quarks created during Pythia's hardest generated process

In Figure 24, the azimuthal angles (ϕ) of the quarks are examined. In subfigure 24a, it is confirmed that ϕ_Q is distributed isotropically across a range of $[-\pi, \pi]$, i.e. the quarks are not more likely to be emitted at any specific value of ϕ_Q. Subfigure 24b is important to take note of as it is where the back-to-back nature of the decay is first represented. This behavior will be reflected in the away-side peak of the e-μ azimuthal correlation distribution, and is therefore key in any analysis that will be undertaken of the pairs.



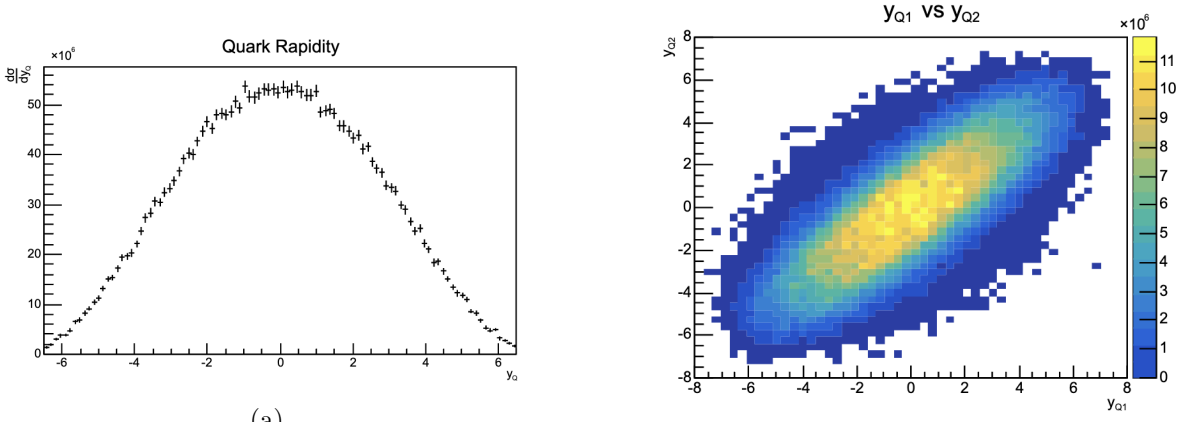
(a)



(b)

Figure 25: Histograms showing the pseudorapidities of quarks created during Pythia's hardest generated process

The pseudorapidity distributions in Figure 25 are the least straightforward of the three coordinates. Though a correlation can still be observed in the 2D distribution (25b), it is not 1:1 as it was with the other coordinates - instead, there is a notable spread. This is a function of the quarks' rapidity (y), from which η is derived (as was first introduced in section 3.1.1). In Figure 26, the rapidity is plotted explicitly. The 2D distribution in subfigure 26b is similarly spread.



(a) (b)

Figure 26: Histograms showing the rapidities of quarks created during Pythia's hardest generated process

When discussing the hardest process of a simulated event in section 5.1.2, it was discussed that the partons within each original colliding proton carry different fractions of the proton's total momentum in the beam direction, z. The distribution of a proton's momentum between its partons is not a set constant, but instead is probabilistic and can be described using a "parton distribution function" (PDF). This implies that the partons involved in the production of the heavy quark pairs have a range of momenta in the direction of the beam axis, which therefore causes a spread in the p_z of the pairs themselves. Equation 3 established that the rapidity of a particle is reliant on p_z, so this spread carries over into the comparison of the pairs' rapidities and pseudorapidities.

5.3.2 Quark to Lepton Decays

Here the relationship between the kinematic variables of the parent quark pairs and their electron and muon decay products is cataloged. Before the e-μ pairs can be used as a probe of their parent particles, analyzing these kinematic relationships will indicate in what ways the e-μ pairs reflect the properties of the quarks and how well they do so (i.e. how closely related they are). In figures 27 and 28 the transverse momentum, azimuthal angle and pseudorapidity are presented. There are no major differences in the quarks' relationship with electrons vs with muons, so the two will be discussed together.

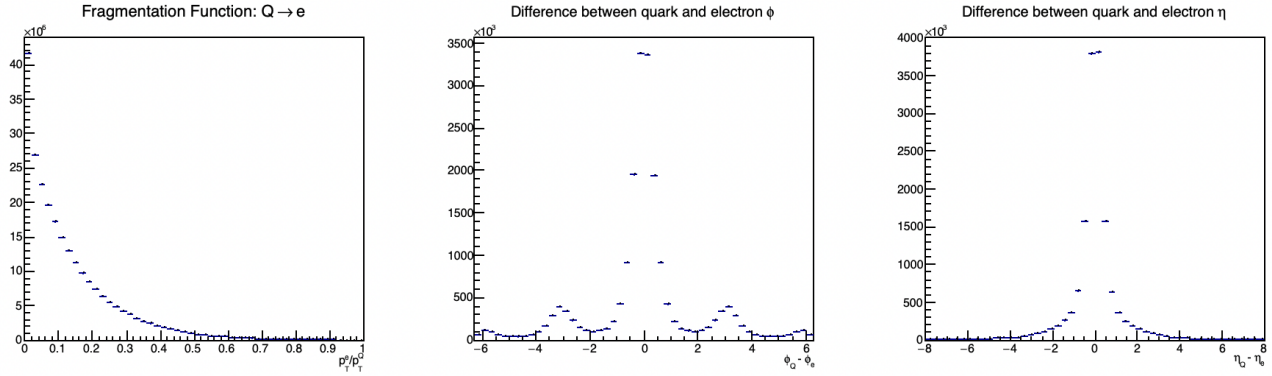


Figure 27: 1D histograms illustrating the relationship between kinematic variables of heavy quarks and electrons that are produced in their semileptonic decays

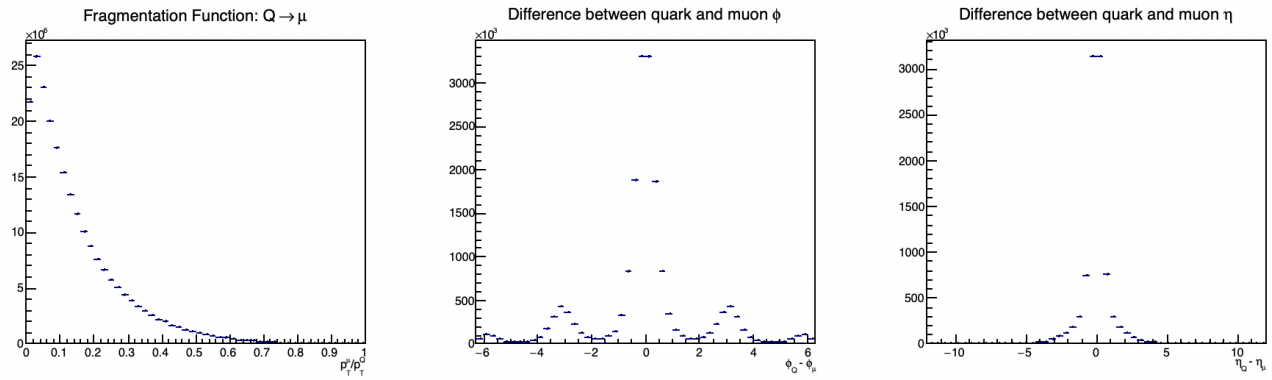


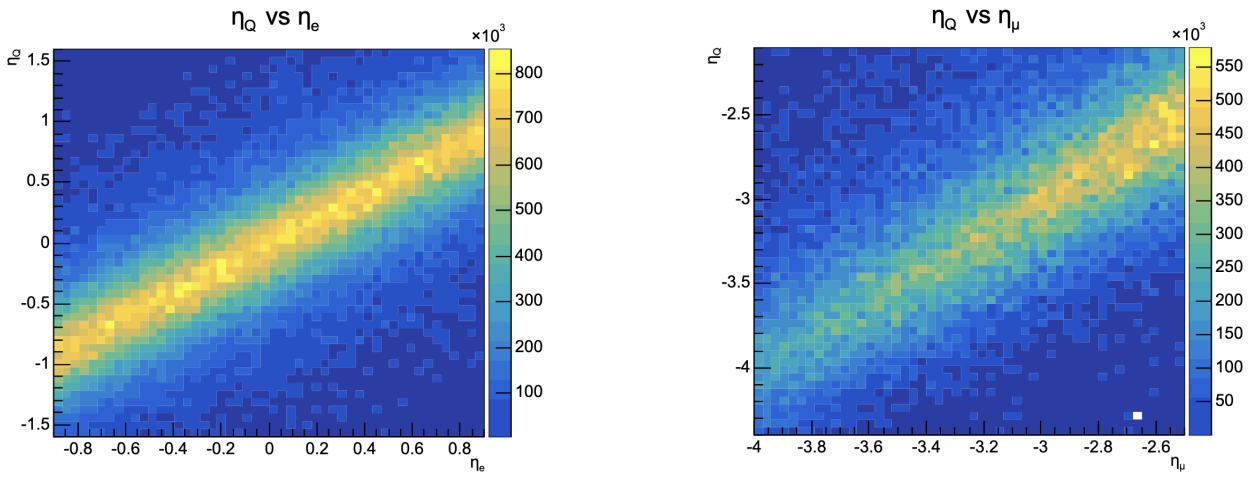
Figure 28: 1D histograms illustrating the relationship between kinematic variables of heavy quarks and muons that are produced in their semileptonic decays

First, the transverse momenta of the two particles are compared. This is done by dividing the p_T of the child lepton with that of its parent quark, in order to visualize the fragment of quark momentum that is carried by the final state decay product. The histogram does not have any entries past a value of $p_T^l/p_T^Q = 1$, indicating that a decay lepton will not have a larger value of p_T than its parent particle. As the quarks decay, their momenta will be shared between the particles created in those processes (i.e. momentum is conserved), leading to decay products with lower momenta. The histogram fits into this expectation. As the quark to lepton ratio approaches 0, the slope of the $Q \to e$ histogram is slightly steeper than that of the $Q \to \mu$ histogram. This implies that muons are more likely to carry a larger fragment of a parent quark's momentum than electrons are.

The second and third comparisons are in the form of correlation histograms, with the lepton's kinematic property being subtracted from that of its quark parent. The azimuthal correlation graphs show the effect described in section 2.4.2. the majority of the quarks' decay leptons reflect their azimuthal angles, creating a strong near-side peak at $\phi_l - \phi_Q = 0$. When combined with the back-to-back correlation of the quark pairs (as shown in 24), this is a precursor

for the away-side peak that is expected in the azimuthal correlation distributions of the e-μ pairs.

Finally, though there is some degree of spread, it can be seen that there is a close correlation between the pseudorapidities of the parent quarks and their decay leptons. This will be exploited in section 5.4. Though there is no great difference in the general pseudorapidity distributions for decays to electrons vs muons, there is a significant difference in the detector regions these particles are recorded in at ALICE, as discussed in section 3. To understand how this impacts the sample, a 2D histogram is plotted in Figure 29, focused on the central barrel and forward regions.



(a) (b)

Figure 29: 2D histogram illustrating the relationship in pseudorapidity between parent heavy quarks and their decay leptons, focused on the central barrel (a) and forward (b) regions of the ALICE detector

The first noticeable difference in the two regions is that there is a higher density of electrons in the central barrel then there are muons in the forward region. This is because quark production is more abundant close to a value of $\eta_Q = 0$, as seen in Figure 25. In addition to this abundance of electrons, the value of η_Q (and therefore η_μ) is centered around 0. This results in a symmetry in the central barrel that doesn't exist in the forward region. Instead, muon production in the forward region is skewed towards the $\eta = -2.5$ edge.

5.4 Approximation Method

It is possible to determine whether a simulated event produces a true e-μ pair from the semileptonic decay of a quark pair, as will be presented in section 5.5 below. It is also possible to examine the semileptonic decays of individual quarks into either electrons or muons, within the restrictions of the required cuts on η and p_T. These three factors will all contribute to the final reduced quark pair cross section, but there is a disconnect - none of the factors provide information about the likelihood that a pair is created where both the electron and muon exist within the cut restrictions.

The most straightforward method to determine this would be to apply the electron and muon kinematic cuts to the leptons from those events that have been identified as producing true pairs (tag `6` in section 5.5). However, this is not as simple to achieve. As will be demonstrated, the cuts imposed by the ALICE detector geometry and the accuracy of the analysis are extremely restrictive so when they are applied to the true semileptonic e-μ sample, only around $\sim 1.9\%$ of the sample remains. Even in a simulation with tens of millions of events, the reduction is too great to work with. After much consideration, it was decided that there was no reasonable way to simulate a sufficient number of events, requiring an alternate solution be put forward.

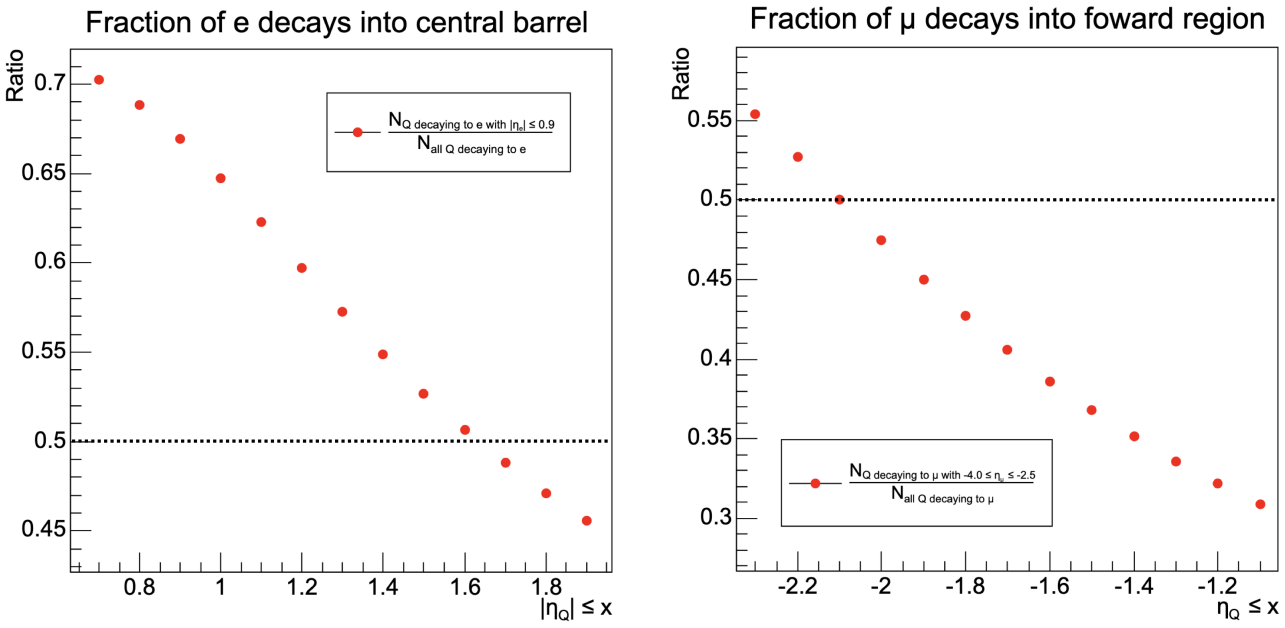
In section 5.3.2, a close relationship between the pseudorapidity of the parent quarks and decay leptons was identified. Due to the chosen Pythia configuration, each generated event is guaranteed to produce a $Q\bar{Q}$ pair, so there is a significantly larger sample of heavy quark pairs to choose from. Therefore, instead of directly searching for the probability that an e-μ pair is produced within the bounds of the detector's η regions, it is possible to approximate that probability by examining the η of the parent quarks. Additionally, since there is no restriction on which of the two parent quark decays to which lepton, both permutations ($Q \to e/\bar{Q} \to \mu$ and $Q \to \mu/\bar{Q} \to e$) are allowed in this model, further increasing the statistics.

To implement this approximation method, a choice of η bounds must be made for the parent quarks. To ensure that the quarks well represent electrons decaying into the central barrel and muons decaying into the forward region, these bounds must be slightly wider than those of the actual detector. The details of how this choice was made are presented in section 5.4.1. Once the bounds are well defined, electrons and muons coming specifically from individual quarks within those bounds can be considered to achieve the desired reduced cross section.

5.4.1 Choice of Quark Pseudorapidity Boundaries

The strategy of choosing the parent quark η bounds was to find the regions in which least half of all electrons or muons produced through semileptonic decays decay into the appropriate detector region. For a range of potential η_Q boundaries, a ratio was calculated that compared the total number of quarks that decayed into an electron or muon with the number of those electrons or muons that decayed "correctly" in η. In Figure 30, this ratio is plotted for potential η_Q ranges, starting slightly within the bounds of the detector regions and increasing in increments of 0.1 on both edges.

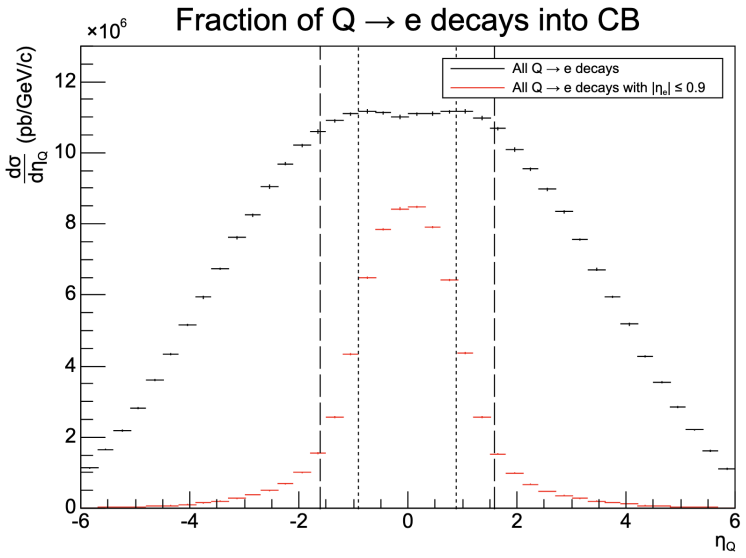
From subfigure 30a, the cutoff past which most decay electrons do not reach the central barrel is the region $|\eta_Q| \leq 1.6$, so this was chosen as the electron parent quark boundary. Since the forward arm is not symmetric in pseudorapidity in the same way as the central barrel, the x-axis of subfigure 30b displays only the upper η bound, with the lower bound corresponding to the same increment away from -4.0 as the upper bound is from -2.5. The chosen boundary for muon parent quarks is therefore $-4.4 \leq \eta_Q \leq -2.1$.



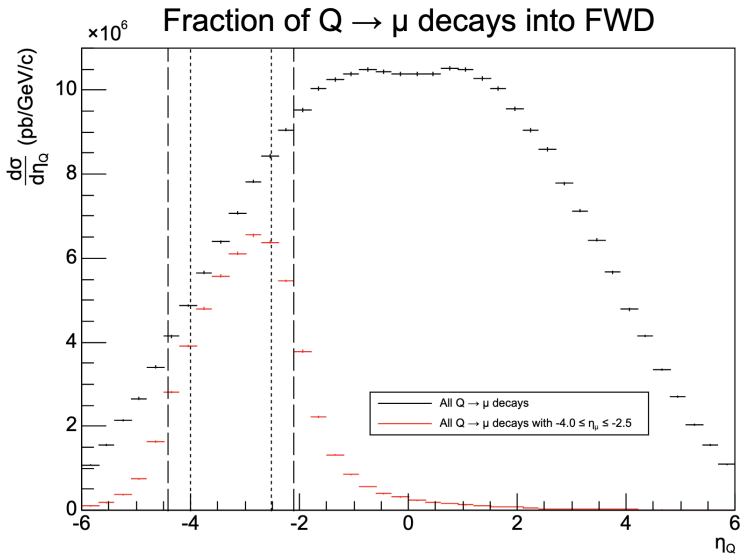
(a) (b)

Figure 30: The ratio of number of leptons produced through semileptonic decay within the boundaries of the ALICE detector vs all leptons produced through semileptonic decay plotted for various parent quark boundaries

To further illustrate the value of the ratios in Figure 30, the pseudorapidity distributions for heavy quarks which produce electrons and muons through semileptonic decay are displayed in Figure 31. The ratios plotted in Figure 30 represent the integral of the black histogram divided by the integral of the red histogram within various bounded ranges of the x-axis.



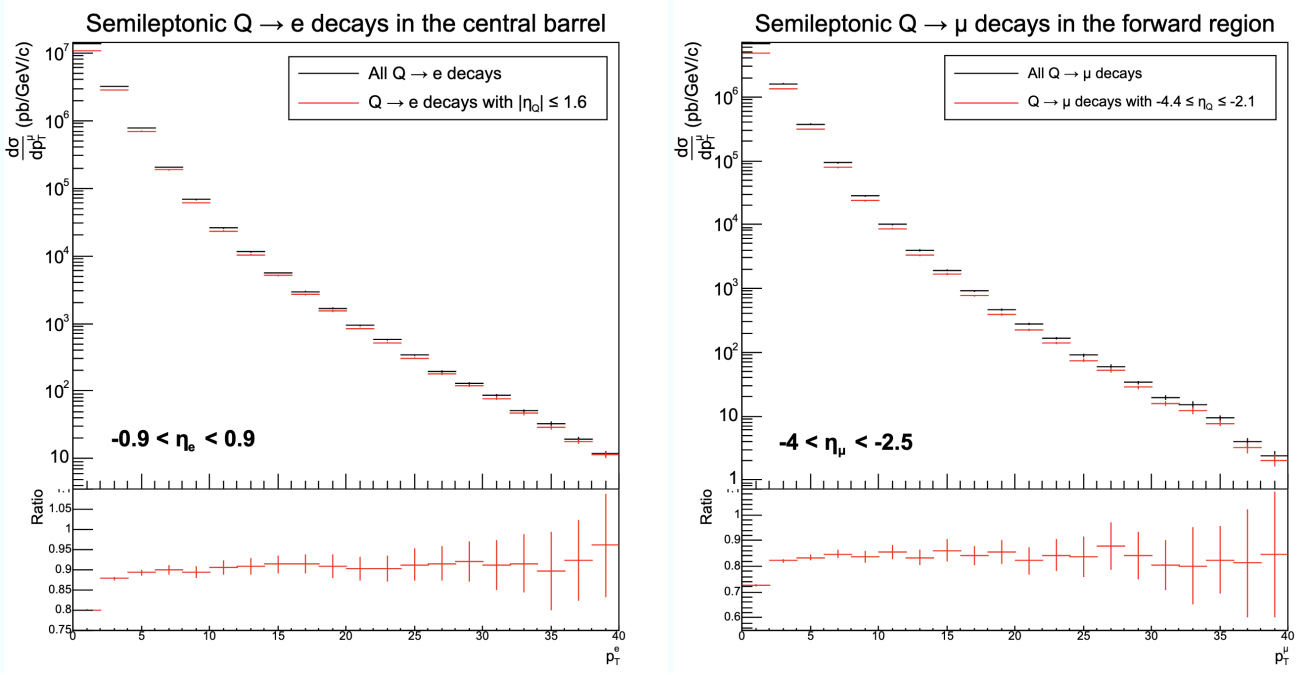
(a)



(b)

Figure 31: Histograms comparing the pseudorapidities of all heavy quarks which decay into either electrons or muons with those that decay into leptons which exist within the ALICE detector's range. The chosen η bounds for the parent quarks are demarcated by vertical dashed lines, with dotted lines indicating the central barrel and forward regions respectively

To confirm that these boundaries will deliver a quark sample that represents a reasonable approximation for their decay products, the final state leptons must also be evaluated. Of all the decay electrons in the central barrel and decay muons in the forward arm, the number that come from quarks within the chosen bounds is examined.



(a) (b)

Figure 32: A comparison of the number of leptons in the two detector regions that come from quarks within the chosen η_Q bounds of the approximation method

In Figure 32, a large proportion of the electrons and muons that are produced within the η bounds of the ALICE detector are observed as coming from parent quarks within the chosen bounds of $|\eta_Q^e| \leq 1.6$ and $-4.4 \leq \eta_Q^\mu \leq -2.1$ across the full range of p_T^e and p_T^μ. For this reason, the bounds are accepted. Now that the decision has been made, it is possible to determine the likelihood that a generated $Q\bar{Q}$ pair meets the η restrictions.

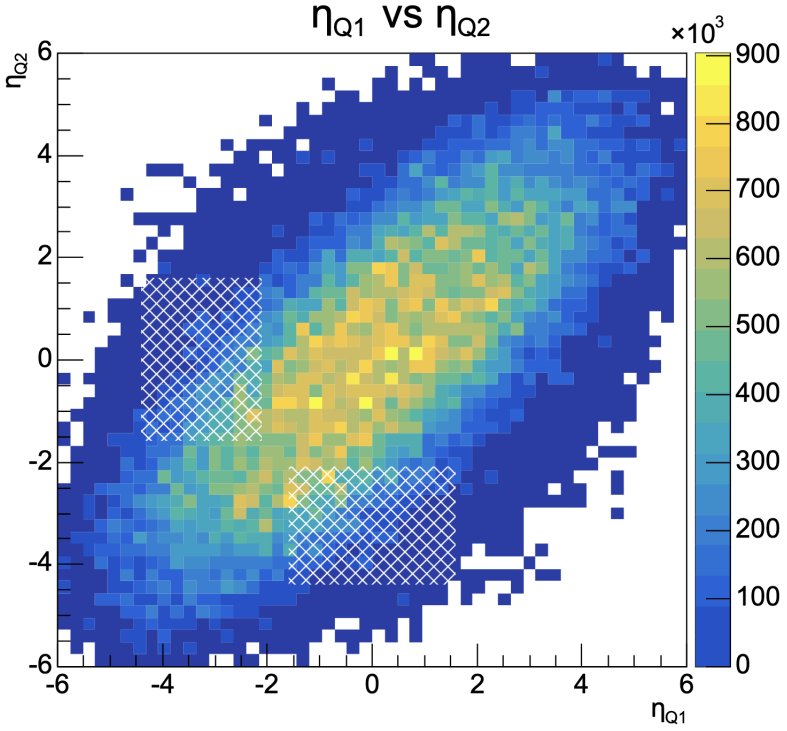


Figure 33: An illustration of the regions in which $Q\bar{Q}$ pairs match the bounds chosen for the approximation method

In Figure 33, the 2D representation of η_Q vs $\eta_{\bar{Q}}$ is again shown, now with the chosen η regions marked by a black hash pattern. About 9.3% of the quark pairs exist within the bounds.

5.5 Event Categorization

In order to obtain an overview of the production of electrons and muons from quark pairs on the event level, each simulated collision was categorized with a numerical tag as follows:

Tag	Configuration
0	No electrons or muons produced from $Q\bar{Q}$ decay
1	Only muon(s) produced from $Q\bar{Q}$ decay
2	Only electron(s) produced from $Q\bar{Q}$ decay
3	Electron and muon produced from $Q\bar{Q}$ decay
4	Back-to-back $Q\bar{Q}$ decay into electron(s) and muon(s)
5	Back-to-back $Q\bar{Q}$ decay into oppositely signed electron(s) and muon(s)
6	Semileptonic $Q\bar{Q}$ decay into oppositely signed electron(s) and muon(s)

Table 2: A breakdown of the different tags used for event categorization in the analysis code

Each tag is indicative of the decay behavior of the original quark pair in the hardest process. Given the nature of the e-μ pairs this study is investigating, events labeled with the tag **6** are

desirable. In these events, one of the original quarks in the pair decays semileptonically into an electron while the other decays semileptonically into a muon. The two leptons are of opposite sign to each other while matching the sign of their original parent quark.

In Figure 34, the simulated events are displayed according to this decay behavior. Each curve is associated with one of the numerical tags in Table 2, but all events that meet the conditions of the tag (as described in the legend) are included in each curve. This means that events are sometimes recorded multiple times in different curves - for instance, a tag **6** event would be included in every curve. The purpose of this is to understand how each additional restriction on the quark decay mode impacts the reduction in the quark pair cross section. The number of $Q\bar{Q}$ pairs which decay to a "true" e-μ are reduced by about two order of magnitudes compared to the total $Q\bar{Q}$ sample.

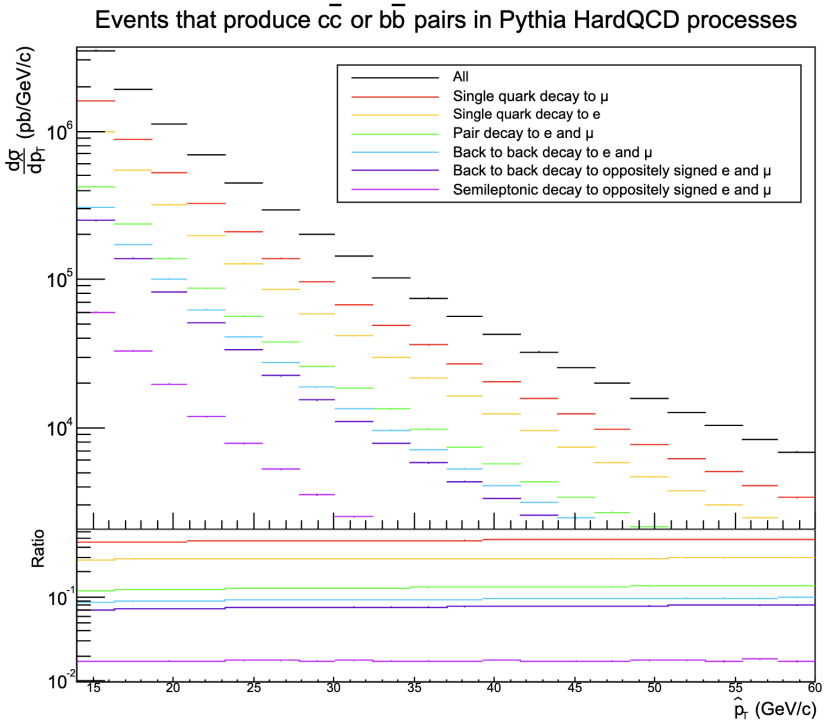


Figure 34: An overview of all simulated heavy quark events, classified based on the decay of the $Q\bar{Q}$ pair into an e-μ pair

5.6 Electron-Muon Pairs

Before exploring the impact of the kinematic cuts that will be made when selecting electrons and muons, a closer look must be taken at the e-μ pairs themselves. The choice of trigger particle and subsequent cuts on p_T^{μ} and p_T^{e} are discussed.

5.6.1 Choice of Trigger Particle

When undertaking a dilepton study, one of the particles is generally labeled the "trigger" particle for any given event, with the second taking the role of the "associated" particle. In any given event, the presence of a trigger and associated particle must be confirmed as a requirement for

its inclusion in the analysis. The two particles cannot be searched for simultaneously, so an order of priority is established. The presence of a trigger particle is confirmed first and its existence is used as a requirement to then search for the associated particle. In this way, events with no particles that pass the conditions of the trigger are rejected immediately, making the analysis more efficient and reducing unnecessary processing time.

Since the goal is to minimize the amount of analysis on events that will end up being rejected, it follows that the requirements for the trigger should be the more restrictive of the two. It is for this reason that higher p_T cuts are usually applied to the trigger than to the associated. In the case of e-μ correlations, one lepton flavor will be chosen to act as the trigger. This adds an additional element to the choice, since the difference in electron and muon production is also relevant.

As seen in Figure 29, there is a higher density of quarks decaying to electrons in the central barrel than there are quarks decaying to muons in the forward region. Therefore the presence of a detectable muon in an event is a more restrictive condition than the presence of detectable electron. It is for this reason that in this analysis a muon trigger was chosen, with electrons acting as the associated particles.

The second motivation for this choice of muon trigger has to do with particle identification. As discussed in section 3, electron identification is a multi-detector process and will always result in a sample with some degree of contamination from other types of particles. However the muon arm is specifically designed to filter out other particles, so there is a high confidence that a signal being detected comes from a muon. By triggering on muons, it ensures that the events which are not immediately rejected have a higher likelihood of including at least one of the desired pair partners.

5.6.2 Note on Transverse Momentum Cuts

Throughout the remainder of this analysis two options for trigger and associated p_T cut will be considered. The first is a lower/less restrictive cut, of $p_T^\mu \geq 3$ GeV and $p_T^e \geq 1$ GeV. The second is a higher/more restrictive cut, of $p_T^\mu \geq 5$ GeV and $p_T^e \geq 3$ GeV. These two cuts were chosen based on previous reports, which suggested that the minimum successful cut for a trigger particle in this scenario would be 3 GeV [1]. It is expected that the higher cut will result in a more accurate e-μ sample but will greatly reduce the statistics. This tradeoff will be considered when making final statements about the feasibility of the e-μ method.

5.6.3 Kinematic Analysis

The relationships between the different kinematic variables of the final state e-μ pairs are presented here, first with no cuts applied and then introducing cuts to catalog the effects on the pairs which remain. In Figure 35, a pseudorapidity distribution is shown which is reminiscent of the relationship in Figure 25b. Considering the strong correlation between η_Q and η_e/η_μ, this aligns with expectation. The transverse momenta of each lepton in the pair are distributed fairly evenly, with both electrons and muons being produced in higher quantities at lower values of p_T. Finally the azimuthal distribution shows that the majority of e-μ pairs have an azimuthal difference of π, which would translate to the expected away-side peak in a 1D distribution and demonstrates the back-to-back behavior of the pairs.

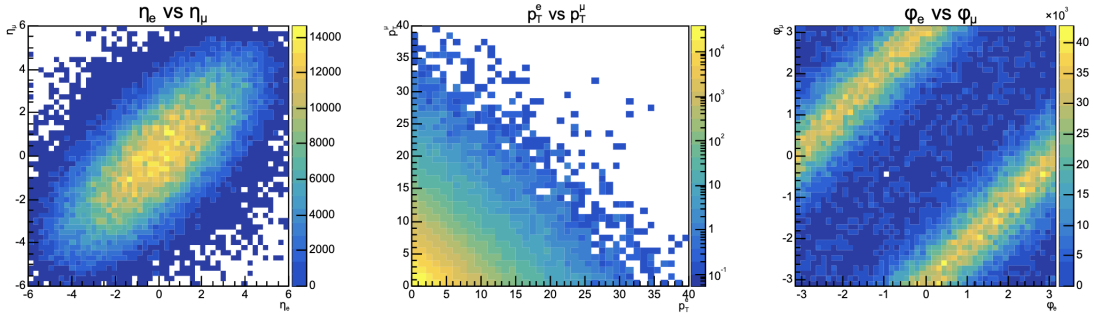


Figure 35: Relationships between the kinematic variables of true e-μ pairs, with no cuts applied

Next the effect of the low p_T cut on these distributions is considered, first in isolation (Figure 36) and then applied in addition to the detector region η cut (Figure 37).

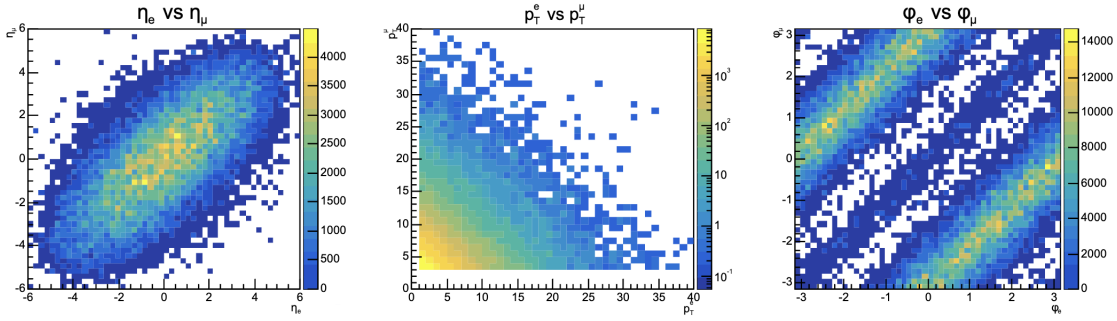


Figure 36: Relationships between the kinematic variables of true e-μ pairs, with the low p_T cut applied ($p_T^e \geq 1$ GeV, $p_T^\mu \geq 3$ GeV)

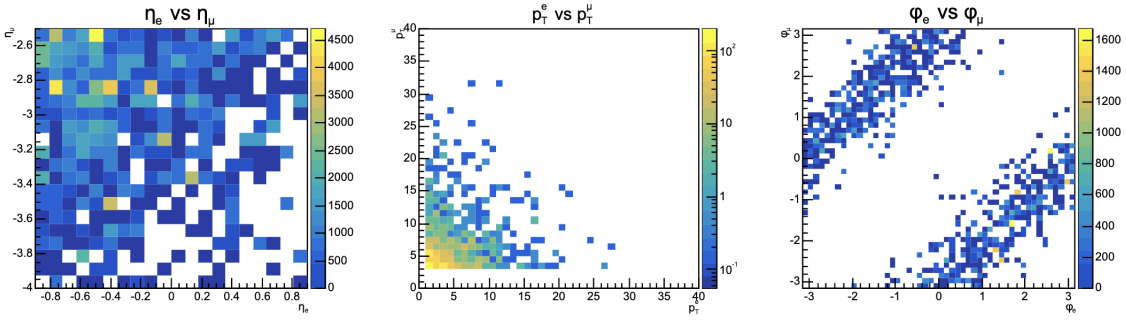


Figure 37: Relationships between the kinematic variables of true e-μ pairs, with the detector geometry ($|\eta_e| \leq 0.9$, $-4.0 \leq \eta_\mu \leq -2.5$) and low p_T cuts applied ($p_T^e \geq 1$ GeV, $p_T^\mu \geq 3$ GeV)

When the p_T cut is applied in isolation in Figure 36, there is a reduction in statistics which excludes the majority of e-μ pairs that exist outside of the dominant away-side peak of the 2D azimuthal distribution. There is a minor secondary near-side peak which becomes visible in this plot, but is removed with the addition of an η cut reflecting the range of the ALICE detector in

Figure 37.

To investigate this, the relationship between $\Delta\eta$ and $\Delta\phi$ is shown in Figure 38. There is a visible relationship between the two quantities and it can be seen that all pairs within this secondary peak have both the electron and muon within the same η region ($\Delta\eta \lesssim 1$). Considering the separation between the central barrel and forward regions, the minimum possible separation in η which would be detectable at ALICE is $\Delta\eta = 1.6$, so this secondary peak would not expected in a data analysis.

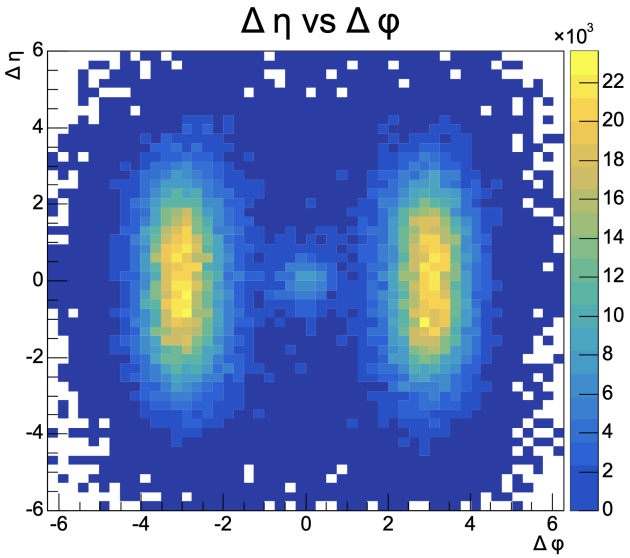


Figure 38: The relationship between $\Delta\eta$ ($\eta_\mu - \eta_e$) and $\Delta\phi$ ($\phi_\mu - \phi_e$) of the e-μ pair signal

The same consideration is made for the high p_T cut in figures 39 and 40. A similar effect is seen in the azimuthal correlation as was the case for the low p_T cut, but to an increased degree. The secondary near-side peak in Figure 39 is less pronounced and when the η cut is added the spread around $\Delta\phi = \pi$ is narrower than before, indicating a distribution that is more sharply peaked.

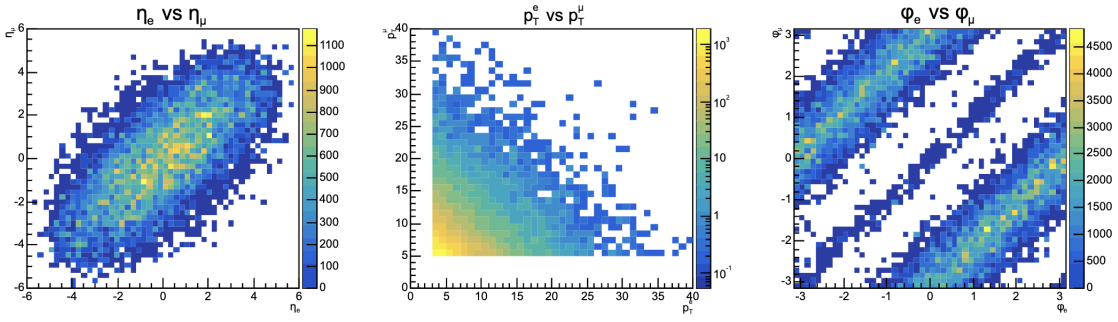


Figure 39: Relationships between the kinematic variables of true e-μ pairs, with the high p_T cut applied ($p_T^e \geq 3$ GeV, $p_T^\mu \geq 5$ GeV)

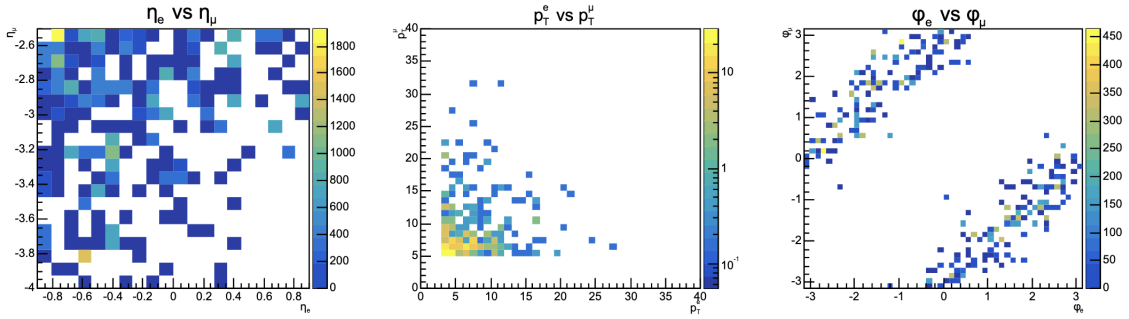


Figure 40: Relationships between the kinematic variables of true e-μ pairs, with the detector geometry ($|\eta_e| \leq 0.9$, $-4.0 \leq \eta_\mu \leq -2.5$) and high p_T cuts applied ($p_T^e \geq 3$ GeV, $p_T^\mu \geq 5$ GeV)

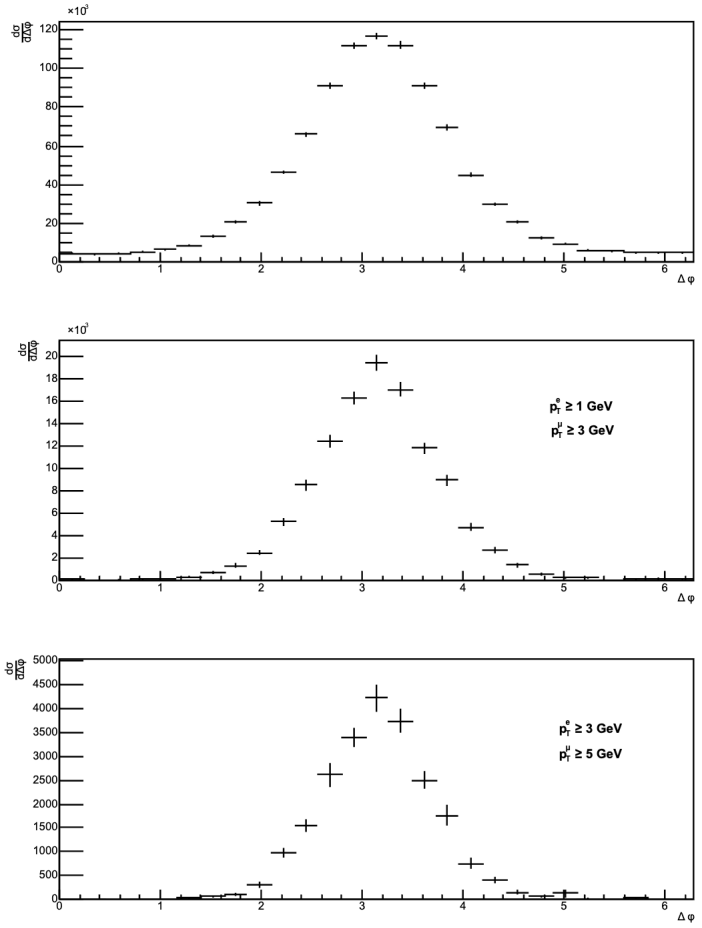


Figure 41: The 1D azimuthal correlation histograms for true e-μ pairs with a separation in pseudorapidity of $|\Delta\eta| \geq 1.6$. The pairs are first displayed before any p_T cuts are made to the trigger muon or associated electron and then after a low and high p_T cut has been applied

Figure 41 examines the impact that the two described p_T cuts have on the 1D azimuthal correlation histograms of the e-μ pairs. A generalized cut of $|\Delta\eta| \geq 1.6$ (where $\Delta\eta = (\eta_\mu - \eta_e)$) is applied to ensure the pairs have a degree of separation in η which is greater than or equal to the minimum separation that is possible considering the bounds of the ALICE detector.

The main observed effect of the p_T cuts is on the shape of the histogram. As the cut becomes more restrictive, the slope of the graph becomes steeper and more sharply peaked around $x = \pi$. This aligns with the expectation from Figures 37 and 40 and indicates an approach towards the behavior of the parent quark pairs with their direct azimuthal separation of $\Delta\phi = \pi$.

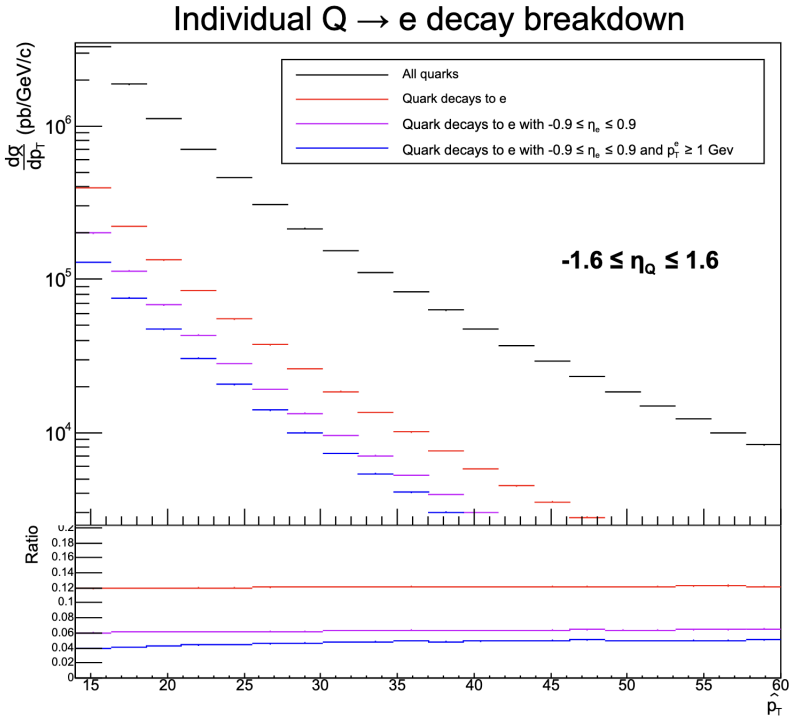
5.7 Impact of Kinematic Cuts

The final element in determining the observable $Q\bar{Q}$ cross section is the reduction to the individual $Q \to l$ decay samples caused by cuts on lepton kinematics. To abide by the approximation method laid out in section 5.4, only leptons decaying from parent quarks in the chosen η_Q ranges will be examined. In each case ($Q \to e$ and $Q \to \mu$), the total heavy quark sample within the appropriate η region will be compared with the sample after three cuts are applied in succession.

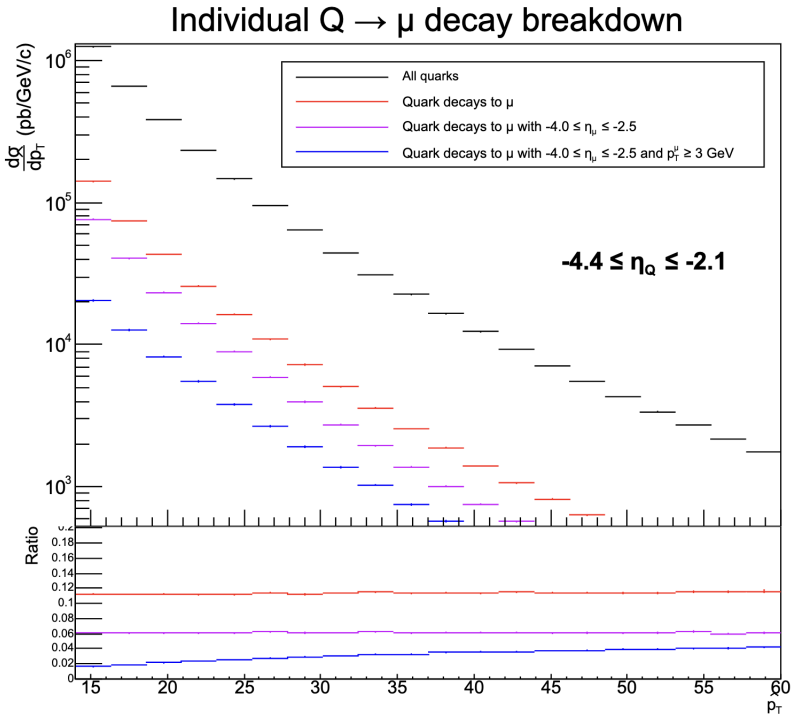
First, a cut is made based on whether the parent quark underwent a semileptonic decay to the correct lepton. Second, a lepton η cut is added to separate those quarks which decay into leptons in the correct detector regions. Third, the p_T cut is applied - the low cut in Figure 42 and the high cut in 43. In the case of both p_T cuts, a ratio is then calculated of the quarks that are observable in the form of their decay lepton (passing all relevant cuts, shown in blue in the figures) divided by the total heavy quark signal (shown in black).

For both the electron and muon decay channels, the reduction from cuts on semileptonic decay and detector pseudorapidity are fairly consistent. Approximately 20% of the quarks undergo semileptonic decay and produce a final state lepton, Of these, about half then remain after cuts are made to η_e or η_μ. This fits in with the ratios used to choose the quark η range in Figure 30.

The signal reduction from the cuts to p_T^e and p_T^μ is highly variable across the range of $\hat{p}_T$ bins, being far greater in the low momentum bins. This is increasingly more pronounced as the cuts to lepton p_T are increased, culminating in the effect observed in Figure 43b where the value of the cross section in the first $\hat{p}_T$ bin actually falls below the value in the second bin. The conclusion that can be drawn from this is that as one looks at a more energetic lepton sample, the particles are increasingly less likely to come from lower energy parent quarks, which aligns with expectation.

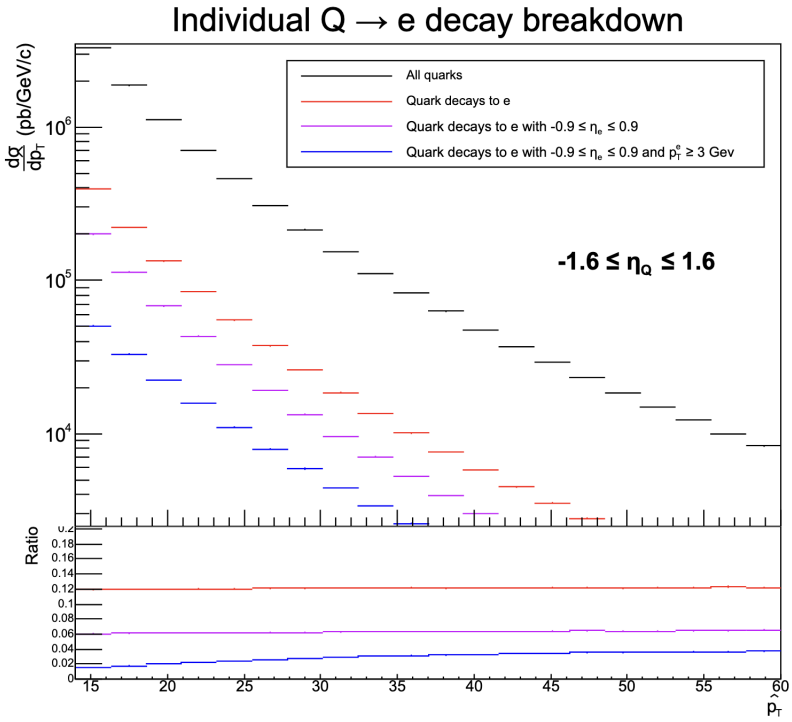


(a)

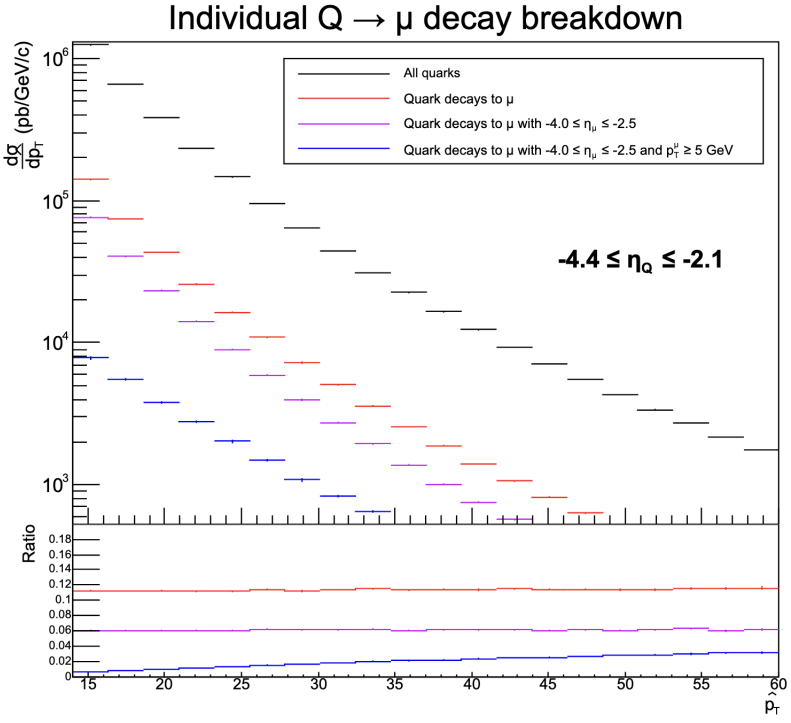


(b)

Figure 42: Histograms showing the impact of various kinematic cuts on the single $Q \to e$ (a) and $Q \to \mu$ decay signals. The low trigger/associated p_T cut option is applied ($p_T^\mu \geq 3$ GeV, $p_T^e \geq 1$ GeV)



(a)



(b)

Figure 43: Histograms showing the impact of various kinematic cuts on the single $Q \to e$ (a) and $Q \to \mu$ decay signals. The high trigger/associated p_T cut option is applied ($p_T^\mu \geq 5$ GeV, $p_T^e \geq 3$ GeV)

5.8 Reduced Cross Section

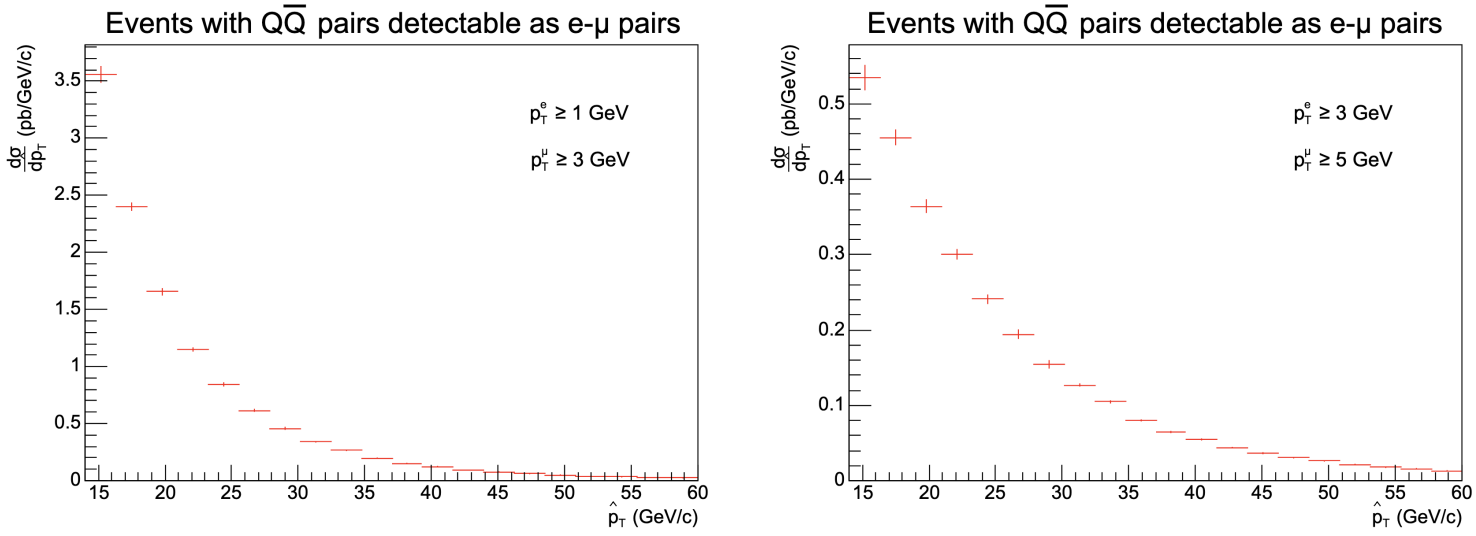
It is now possible to calculate the final reduced cross section of all $Q\bar{Q}$ pairs which are potentially observable as e-μ pairs within the cut restrictions. To review, the contributing factors are:

(i) The amount of quark pairs which exist within the chosen η region bounds of the approximation method (section 5.4);

(ii) The proportion of $Q\bar{Q}$ which undergo back-to-back semileptonic decays into an oppositely signed e-μ pair (section 5.5); and

(iii) The proportion of quarks within these η bounds that decay to central barrel electrons which pass the chosen associated p_T cuts and forward muons which pass the chosen trigger p_T cuts respectively (section 5.7).

Where relevant, the reduced signal after the applied cuts was divided by the total $Q\bar{Q}$ signal and this ratio was recorded and saved for each $\hat{p}_T$ bin. These effects can now all be combined and applied to the original $Q\bar{Q}$ production cross section to achieve a reduced cross section of all the $Q\bar{Q}$ pairs which would have the possibility of being observed by the ALICE detector in the form of an e-μ pair. The reduced cross sections in the case of both the high and low p_T cuts are displayed in Figure 44.



(a) (b)

Figure 44: Reduced cross section per $\hat{p}_T$ bin of $Q\bar{Q}$ pairs which are potentially observable as e-μ pairs within the cut restrictions of the ALICE detector pseudorapidity regions, with low (a) and high (b) cut options for trigger and associated lepton p_T

In order to assess the feasibility of this method, it is now necessary to use these cross sections to calculate an estimate of the number of potentially observable pairs expected to be produced over the course of Run 3. Recall equation 5, where the relationship between cross section (σ) and luminosity was defined:

$$N_{events} = L \times \sigma$$

The final piece of information needed to calculate N_{events} is the integrated luminosity of ALICE. This is measured using a method known as a van der Meer scan, which was most recently performed at ALICE in September 2023 [42]. It was found that in 2023, 9.7 pb^{-1} of data was recorded. Combined with the readings from 2022 (19.3 pb^{-1}), this indicates that for all of Run 3 thus far, ALICE has recorded 29 pb^{-1} of data, as shown in Figure 45.

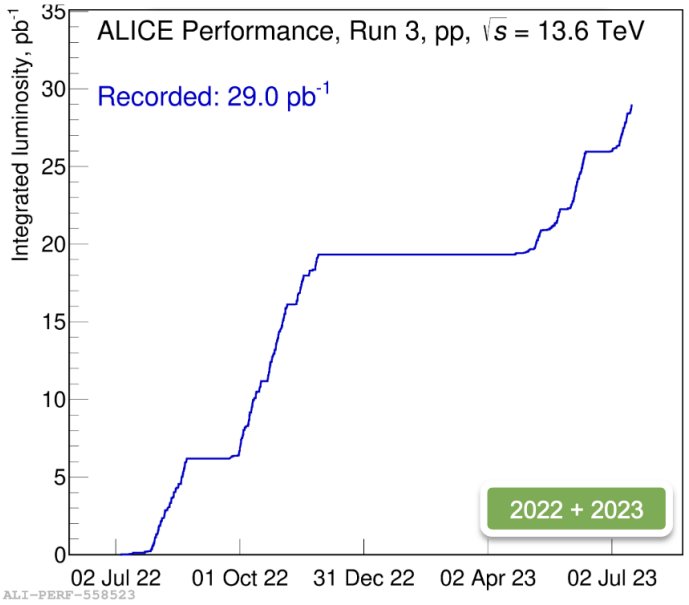
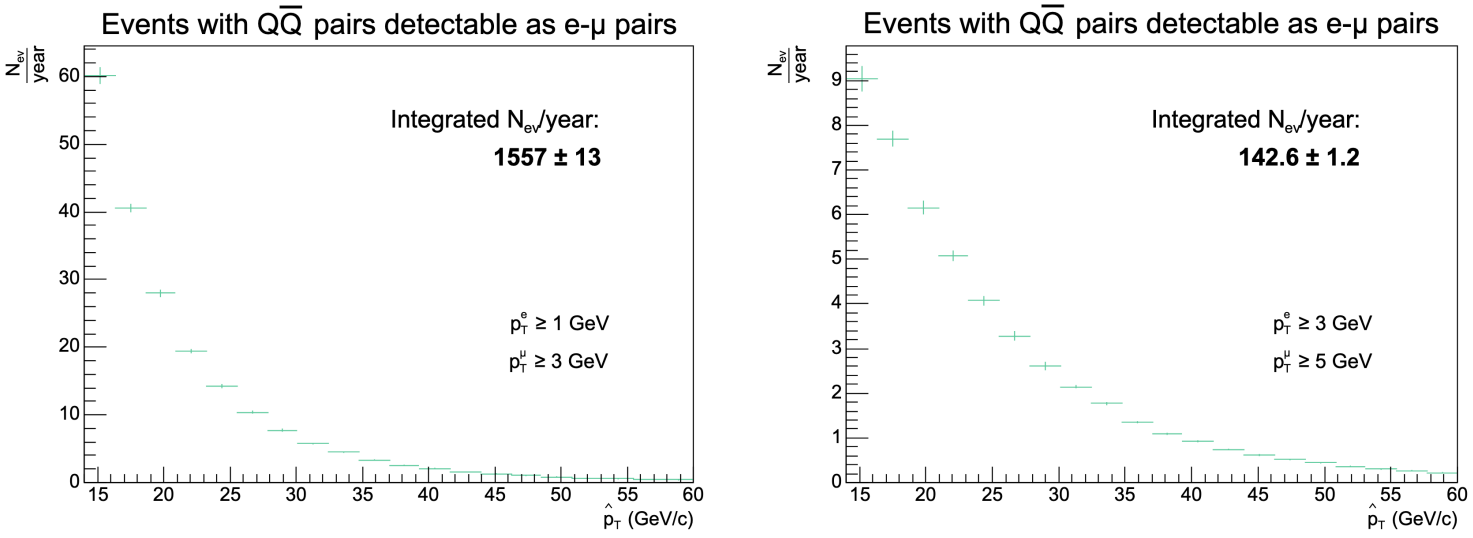


Figure 45: The integrated luminosity of the ALICE detector in 2022 and 2023 [42]

In 2023, the LHC experienced an unscheduled temporary shutdown to allow for repairs after a vacuum leak [42]. As a result of this, less data was recorded than was originally planned. It was decided that in order to obtain an estimate of the expected e-μ signal over one year of Run 3 data, the average value of integrated luminosity over the two years would be used (16.9 pb^{-1}). This can then be multiplied by the reduced cross section per $\hat{p}_T$ bin from Figure 44. The resulting histogram can be integrated over the entire range.

In Figure 46, the estimate of the true e-μ sample observable at ALICE in one year of Run 3 data is shown for both the low and high lepton p_T cuts. Since Run 3 is scheduled to last for two more years, until the end of 2025, these can be multiplied by 4 to obtain an estimate for the size of the e-μ sample throughout Run 3. The results are summarized in Table 3.



(a) (b)

Figure 46: Luminosity per $\hat{p}_T$ bin of $Q\bar{Q}$ pairs which are potentially observable as e-μ pairs within the cut restrictions of the ALICE detector pseudorapidity regions, with low (a) and high (b) cut options for trigger and associated lepton p_T

	Estimate of observable pairs in 1 year of Run 3 data	**Estimate of observable pairs in 4 years of Run 3 data**
Low p_T cut	1557 ± 13	6228 ± 52
High p_T cut	142.6 ± 1.2	570.4 ± 4.8

Table 3: Number of $Q\bar{Q}$ pairs potentially observable as e-μ pairs during Run 3 at ALICE

6 Discussion and Conclusions

6.1 Summary

In this report, the e-μ channel was discussed in depth. The pairs are a desirable probe for heavy flavor quarks, the study of which allows for testing the perturbative QCD cross section. Heavy quarks can also provide insight on interactions within the QGP, since they are able to survive past the point of freeze-out in the form of hadrons. Correlated e-μ pairs are a desirable alternative to other heavy quark probes as they have fewer contributing background sources.

However, the semileptonic decay process which creates the e-μ pairs is quite rare, meaning that analysis is restricted by the limited statistics that can be recorded. At ALICE, the design of the detector additionally requires that the electron and the muon be produced in the central and forward pseudorapidity regions respectively, which increases the restriction. Finally, pairs in the sample must survive past high p_T cuts for background reduction and in order to best reflect the

parent heavy quark pairs.

In previous LHC runs, these restrictions have made it impossible to use the pairs as a method of probing heavy quarks at ALICE. The increase in amount of recorded data (integrated luminosity) and center of mass energy during Run 3 may lift this restriction. To assess whether the method is now feasible, 13.6 TeV proton-proton events were simulated using Pythia and the $Q\bar{Q} \rightarrow e\text{-}\mu$ decay channel was examined in depth. Accompanying the simulated study, an assessment was made of the software that would be required to undertake an analysis of the pairs with Run 3 data. The O2 analysis framework was set up on the UCT HPC and an initial version of a data analysis task was produced and tested on a small dataset.

Over the course of the simulated study, the kinematic properties of the parent heavy quark pairs and their semileptonic decay products were presented and compared. Due to statistics limitations in the simulation, the close correlation between quark and lepton pseudorapidity was taken advantage of to develop an approximation for the likelihood of a produced $e\text{-}\mu$ pair decaying into the correct respective η regions. The likelihood that a heavy quark pair would decay semileptonically into an $e\text{-}\mu$ pair was also determined. The role of the muon as a trigger particle, with the electron acting as the associate, was established and the azimuthal correlation between simulated $e\text{-}\mu$ pairs was presented. The effect of p_T cuts at two levels of restriction was investigated.

Finally, the various factors contributing to the reduction of the original parent quark signal were combined to produce a reduced cross section representing the amount of heavy quark pairs which could be observed with the ALICE detector in the form of correlated $e\text{-}\mu$ pairs. This was combined with the integrated luminosity of the ALICE detector as recorded thus far in Run 3 to obtain a numerical value for the number of pairs that could be expected within a Run 3 signal.

6.2 Feasibility

With an approximation of the number of observable pairs calculated in section 5, it is now necessary to assess whether this expected signal would be significant enough for the pairs to provide a successful heavy quark pair measurement. To make this determination, two points of comparison will be provided. First, the integrated luminosity of Run 2 will be considered to provide context for how much of an improvement Run 3 has brought. Then, the number of pairs used during the previous data study performed at PHENIX [2] will be presented to facilitate a discussion of whether a comparable sample could be analyzed over the course of Run 3.

6.2.1 Comparison to Run 2

A direct comparison to the $e\text{-}\mu$ signal that would have been possible to observe in previous LHC runs is difficult on two accounts. To start, the center-of-mass energy of the proton collision beams during Run 2 was 13 TeV, rather than the 13.6 TeV of Run 3 (and used for the simulations in this report). In addition, as discussed in Section 4, the Run 3 ALICE triggering system was adjusted to accommodate the continuous data taking mode.

Despite these discrepancies it is possible to use the reduced cross sections of section 5.8 and the integrated luminosity of ALICE from previous runs to obtain an upper limit for the amount of pairs that would have been observable. Due to the nature of the pairs and the high p_T cuts as discussed throughout the analysis, it is expected that the lower energy and event-by-event

trigger system would result in fewer e-μ pairs produced that pass all cut thresholds. This implies that signal calculated with the simulated 13.6 TeV data would be an overestimate (the "true" signal for Run 2 would lie below it).

Throughout the three years of Run 2 proton-proton collisions at 13 TeV (2016, 2017 and 2018), a total integrated luminosity of $31.93 \pm 0.43 nb^{-1}$ was recorded [43]. Since $1 pb^{-1}$ is equal to $1000 nb^{-1}$, this can be converted to $0.03193 \pm 0.00043 nb^{-1}$. This means that the data recorded over the entirety of Run 2 is equivalent to

$$\frac{0.03193}{16.9} = 0.19\%$$

of the amount of data to be collected on average per year in Run 3.

When this percentage is multiplied by the pair estimates over a single year in Table 3, the following upper limit estimates are obtained.

	Estimate of observable pairs in 1 year of Run 3 data	**Upper limit estimate of observable pairs in 3 years of Run 2 data**
Low p_T cut	1557 ± 13	2.958 ± 0.025
High p_T cut	142.6 ± 1.2	0.2709 ± 0.0023

The numbers quoted do not enable a definitive statement about the feasibility of the method within the Run 3 estimate. However, the lack of feasibility during Run 2, as well as the subsequent improvement in the Run 3 sample, are both clear. For both levels of p_T cut restriction, the estimated number of pairs increases by a factor of 10^3 in Run 3, which is consistent with the increase in integrated luminosity.

6.2.2 Comparison to Past Studies

In the previously performed PHENIX data analysis cited throughout this study [2], the exact number of raw e-μ pairs identified during the proton-proton run are quoted. A total of 7613 opposite-signed pairs and 5788 like-signed pairs survived after kinematic cuts had been applied. This suggests that after the like-signed subtraction, approximately 1825 raw pairs remained.

Event mixing, background subtraction and detector efficiency were then considered in the PHENIX study. The signal to background ratio for the PHENIX study is not explicit, so no final figure can be quoted for comparison. However, from the plotted background contributions, it can be estimated that the signal was reduced by at least 15%. This implies an upper limit of approximately 1552 pairs. This can be used for comparison with the calculated ALICE Run 3 expectations as presented.

The estimates in Table 3 represent the number of heavy quark pairs which produce an e-μ pair that passes all cut requirements. Therefore they also represent the number of true e-μ pairs produced from semileptonic heavy quark decays which have the chance of being recorded by the ALICE detector (i.e. the number that would be expected after a like signed subtraction, event mixing and background subtraction are performed).

For the most restrictive p_T cut ($p_T^\mu \geq 5$ GeV and $p_T^e \geq 3$ GeV) over the entire course of Run 3 (all four years), this estimate lies well below the raw pair count in the PHENIX study. In order to construct an azimuthal correlation plot at this level of cut restriction, the full Run 3 dataset would need to be used, at which point a sufficient signal would be guaranteed. This would require the use of significant computing resources in order to run through the entire data sample, which would have to be requested from the appropriate ALICE board. Considering the small expected sample and the fact that the impact of detector effects on the pairs is not yet known, there are not sufficient grounds on which to do this. In addition, no such analysis could be performed until after 2025 once all Run 3 data has been collected and processed.

On the other hand, the less restrictive p_T cut ($p_T^\mu \geq 3$ GeV and $p_T^e \geq 1$ GeV) is estimated to produce a comparable e-μ pair sample to the PHENIX estimate in a single year of Run 3 data. This implies that within the two years of data already collected during Run 3, there is the potential to perform a comparable analysis with slightly higher statistics, with a further increase by the end of Run 3 in 2025.

6.2.3 Outlook

Considering the bounds of the simulated analysis, the conclusion of this thesis is that a study of correlated e-μ pairs in Run 3 proton-proton data is feasible, at the least restrictive level of cut restrictions ($p_T^\mu \geq 3$ GeV and $p_T^e \geq 1$ GeV). It is expected that a comparable e-μ signal to past investigations should already be present in the Run 3 proton-proton data collected as of 2024, with increased feasibility if the study is undertaken at the end of the Run 3 data taking period. In addition, the software infrastructure required to perform such a study would be achievable within the currently existing framework.

6.3 Further Investigations

The clear next step in this process is undertaking an study that utilizes Run 3 data. This would require the completion of a relevant analysis task such as the one tested throughout this thesis and the development of event-level cuts for data reduction to be integrated into the already-existing O2-DQ framework. Finally, due to the amount of data that must be examined in even a single year of Run 3, it would be necessary to obtain access to the computing resources required to perform the analysis.

Looking past the immediate future, a successful p-p study has the potential to lead in to an analysis of e-μ pairs in Run 3 heavy ion collision data. A similar simulated feasibility work could be undertaken to determine the expectations of this scenario, followed by a comparison of the two collision systems if successful.

Acknowledgments

I would like to sincerely thank my supervisor Thomas Dietel and my co-supervisors, Zinhle Buthelezi and Siegfried Förtsch, for their assistance, advice and support throughout this analysis. Thanks is also extended to my fellow UCT-ALICE postgraduate students, Stephan Potgieter and Salmaan Barday, who always had good ideas and were willing to talk through issues as they came up.

I also wish to recognize and thank the friends and family members who engaged with this thesis, helped to identify the many grammatical/stylistic errors contained within and reminded me to "muon" when I got stuck in my own head. Finally thank you to Pippin, my wonderful and fluffy canine friend, for keeping me in good company (and good spirits) during the final months of writing.

References

[1] Yukinao Akamatsu, Tetsuo Hatsuda, and Tetsufumi Hirano. Electron-muon Correlation as a New Probe to Strongly Interacting Quark-Gluon Plasma. *Physical Review C*, 2009. doi: 10.1103/physrevc.80.031901.

[2] Tatia Englemore. *Electron-Muon Correlations in Proton+Proton and Deuteron+Gold Collisions at PHENIX*. PhD thesis, Columbia University, 2011.

[3] Alice experiment. `https://alice.cern/`, 2024.

[4] Kohsuke Yagi, Tetsuo Hatsuda, and Yasuo Miake. *Quark-Gluon Plasma: From Big Bang to Little Bang*. Cambridge University Press, 2005.

[5] Cush. `https://upload.wikimedia.org/wikipedia/commons/0/00/Standard_Model_of_Elementary_Particles.svg`, 2019.

[6] Mark Thompson. *Modern Particle Physics*. Cambridge University Press, 2013.

[7] David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Education Limited, 2014.

[8] Peter Skands. Introduction to QCD. In *Searching for New Physics at Small and Large Scales*. World Scientific, 2013. doi: 10.1142/9789814525220_0008. URL `http://dx.doi.org/10.1142/9789814525220_0008`.

[9] L.Udaya Kumar and Declan Keane. Experimental studies of the quantum chromodynamics phase diagram at the STAR experiment. *Pramana*, 84:773–786, 2015. doi: 10.1007/s12043-015-0969-9.

[10] W.A. Zajc. The Fluid Nature of Quark-Gluon Plasma. *Nuclear Physics A*, 805(1–4), 2008. ISSN 0375-9474. doi: 10.1016/j.nuclphysa.2008.02.285. URL `http://dx.doi.org/10.1016/j.nuclphysa.2008.02.285`.

[11] Luciano Maiani. The GIM Mechanism: Origin, Predictions and Recent Uses, 2013.

[12] Sheldon L. Glashow. The Hunting of the Atoms and Their Nuclei. `https://www.nytimes.com/1976/07/18/archives/the-hunting-of-the-atoms-and-their-nuclei-protons-and-neutrons.html`.

[13] E. Norrbin and T. Sjöstrand. Production and Hadronization of Heavy Quarks, 2000.

[14] D Souza and N H Brook. Analysis of Charm Pair Production at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 43(1), 2015. ISSN 1361-6471. doi: 10.1088/0954-3899/43/1/015001. URL `http://dx.doi.org/10.1088/0954-3899/43/1/015001`.

[15] Katharina Garner. Separation of Heavy-Flavour Production Mechanisms via Two-Particle Angular Correlations in pp Collisions at s = 2.76 TeV. Master's thesis, Westfälische Wilhelms-Universität Münster, 2017.

[16] Izaak Neutelings. CMS coordinate system, Psuedorapidity. `https://tikz.net/author/izaak/`.

[17] John Collins. Light-cone Variables, Rapidity and All That, 1997.

[18] The ALICE Collaboration. Technical Design Report for the Upgrade of the ALICE Inner Tracking System. Technical report, CERN, 2013.

[19] The ALICE Collaboration. The ALICE Experiment at the CERN LHC, 2008.

[20] Rudolf Frühwirth and Are Strandlie. *Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors*. Springer Charm, 2021. doi: 10.1007/978-3-030-65771-0.

[21] The ALICE Collaboration. Technical Design Report for the Upgrade of the ALICE Time Projection Chamber. Technical report, CERN, 2014.

[22] The ALICE Collaboration. Technical Design Report of the Transition Radiation Detector. Technical report, CERN, 2001.

[23] The ALICE Collaboration. Technical Design Report of the Dimuon Forward Spectrometer. Technical report, CERN, 1999.

[24] The ALICE Collaboration. Technical Design Report for the Muon Forward Tracker. Technical report, CERN, 2015.

[25] The ALICE Collaboration. O2 Analysis Framework Documentation. `https://aliceo2group.github.io/analysis-framework/docs/`.

[26] University of Cape Town, Timothy Carr, and Andrew Lewis. UCT HPC Facility. `https://ucthpc.uct.ac.za/index.php/hpc-cluster/`, 2023. URL `https://doi.org/10.5281/zenodo.10021613`.

[27] Gregory M. Kurtzer, Dave Trudgian, Michael Bauer, Ian Kaneshiro, David Godlove, Yannick Cote, Carlos Eduardo Arango Gutierrez, Geoffroy Vallee, Justin Cook, Jason Stover, Adam Hughes, Brian P Bockelman, Marcelo Magallon, Jacob Chappell, Mike Frisch, Daniele Tamino, Carl Madison, Sasha Yakovtseva, Amanda Duffy, Satrajit Ghosh, Tru Huynh, Mike Gray, and Yaroslav Halchenko. sylabs/singularity: SingularityCE 4.0.2. `https://docs.sylabs.io/guides/latest/user-guide/`, 2023. URL `https://doi.org/10.5281/zenodo.10143865`.

[28] Docker Documentation. `https://docs.docker.com/`.

[29] Lorenz Gerber. Containerization for HPC in the Cloud: Docker vs Singularity - A Comparative Performance Benchmark. `https://api.semanticscholar.org/CorpusID:204087303`, 2018.

[30] Installation via Alibuild. `https://alice-doc.github.io/alice-analysis-tutorial/building/custom.html`.

[31] Korbin Brown. Linux Hard Link vs. Soft Link. `https://linuxconfig.org/linux-hard-link-vs-soft-link`, 2023.

[32] Nina Nathanson. O2 on the HPC. `https://github.com/NNathanson/o2hpc`, 2023.

[33] SLURM Workload Manager. `https://slurm.schedmd.com/overview.html`.

[34] Alkin, Anton, Eulisse, Giulio, Grosse-Oetringhaus, Jan Fiete, Hristov, Peter, and Kabus, Maja. ALICE Run 3 Analysis Framework. *EPJ Web Conf.*, 2021. doi: "10.1051/epjconf/202125103063". URL "`https://doi.org/10.1051/epjconf/202125103063`".

[35] O2 DQ group. Dileptons-Quarkonia (DQ) analysis framework. https://aliceo2group.github.io/analysis-framework/docs/advanced-specifics/pwgdq.html.

[36] Nina Nathanson. E-Mu Pythia. https://github.com/NinaFN/E-Mu-Pythia, 2024.

[37] Christian Bierlich, Smita Chakraborty, Naomi Cooke, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Utheim, and Rob Verheyen. Documentation - PYTHIA 8.3. https://pythia.org/documentation/.

[38] Frances Y. Kuo and Ian H. Sloan. Lifting the Curse of Dimensionality. 2005. URL https://api.semanticscholar.org/CorpusID:15991040.

[39] Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Utheim, and Rob Verheyen. A comprehensive guide to the physics and usage of PYTHIA 8.3, 2022.

[40] Joshua S. Speagle. A Conceptual Introduction to Markov Chain Monte Carlo Methods, 2020.

[41] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022. ISSN 2050-3911. doi: 10.1093/ptep/ptac097. URL https://doi.org/10.1093/ptep/ptac097.

[42] Sarah Porteboeuf and Andrea Ferrero. Report from Run Coordination. ALICE Week Presentation, 2023.

[43] The ALICE Collaboration. ALICE 2016-2017-2018 luminosity determination for pp collisions at 13 TeV, 2021.