Finding Charm at the LHCb

An Analysis of Trigger Selections on $K_s \to \pi\pi$ and $D^0 \to K\pi$ Decays
MPhys Report

Robert Hartley
Department of Physics and Astronomy, University of Manchester
(Project performed in collaboration with Elisabeth Peak)
(Dated: January 7, 2024)

This report investigates selections that will be made on the software High-Level Trigger 1 that will be implemented at the LHCb as part of the Allen upgrade. The report focuses on selections on the $K_s^0 \to \pi^\pm\pi^\mp$ and $D^0 \to K^\pm\pi^\mp$ decays. The report uses a data set from LHCb-Run 3 and a Monte Carlo simulated data set to evaluate cuts to select both decays. For the K_s decay a selection was found that results in a 92.8% signal fraction. For the D^0 decay, a set of the most relaxed and most strict selections was found. Within this range, one iteration of p_T >500MeV/c, p >4000MeV/c, track χ^2_{IP} >3, DIRA>0.995, D^0_{IP} <0.15mm resulted in a signal efficiency of 13.6% and an event rate of 2.93%.

1. INTRODUCTION

1.1. LHC and LHCb

The Large Hadron Collider (LHC) is the world's largest and most powerful particle collider. It is a 26.7km circumference, two-ring, hadron collider built 100m underground at the European Organization for Nuclear Research (CERN) near Geneva [1]. Among the array of experiments conducted at the LHC, the LHCb detector focuses on studying the 'beauty' and 'charm' particle decays in proton-proton collisions. The LHCb has produced over 600 papers on topics such as Charge-Parity (CP) violation, electroweak physics and rare b and c quark decays [2].

The LHCb has a different geometry from the other 3 main detectors along the LHC beam as it is a single-arm spectrometer with a forward angular coverage from approximately 15 mrad to 300 (250) mrad in the bending (non-bending) plane [3]. Whilst only covering 4% of the solid angle, the LHCb includes ≈ 40% of $b\bar{b}$ production cross-section [4]. As it only has detectors in the forward direction, the LHCb can afford more precise tracking detectors than the other three main detectors, allowing the collaboration to push the frontier of precision physics. The main detectors used in reconstructing particle tracks are described below.

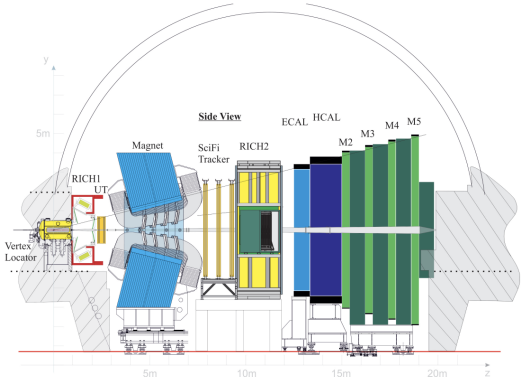


FIG. 1: LHCb Upgrade I Schematic [5]

1.2. LHCb Detectors

1.2.1. Tracking Detectors

VErtex LOcator (VELO) - The first detector in the chain is the VErtex LOcator (VELO). Through tracking charged particles, the VELO measures the location of primary vertices (proton-proton interaction), secondary vertices (decay vertices) and the distances between them. The precision of this information is vital for event selection as it causes consequential uncertainties in other calculated variables. The VELO is a pixilated hybrid silicon detector arranged in a diamond pattern around the beam axis. To achieve optimal resolution for the primary and secondary vertex locations, the VELO needs to be as close to the beam as possible. However, during the beam injection and before reaching the maximum beam energy, the two proton beams are separated and therefore slightly dispersed. Positioning the VELO too close to the beam axis during injection could cause significant damage to the detector. Therefore, the VELO detector is retractable and sits away from the beam at a distance of 3cm during injection and tuning. The arms move in and enclose around the beam at a distance of 5.11mm once the beam is stable. This configuration allows the VELO to achieve a primary vertex spatial resolution of a few microns [6].
Upstream Tracker (UT) - This is a silicon microstrip detector that minimises the occurrence of 'fake' or 'ghost' tracks. Particles are required to leave hits in the VELO, UT and SciFi detectors to contribute to a real particle track. Measuring the momentum in both VELO and UT before the magnet, further reduces the search window in SciFi, reducing event reconstruction times.
Magnet - The magnet applies a magnetic field with bending power 4 Tm [7] which curves the charged particle tracks to allow calculation of momentum which is vital for analysis. **SciFi Trackers** - After the magnet bends the charged particles, a series of scintillating fibres provides another area to detect charged particles. These are designed to cover the larger spatial range caused by being around 10m away from the primary vertex [8].
The hits left in all three of these detectors can then be reconstructed into tracks with the required variables needed to perform selections.

1.2.2. The particle identification system

The particle identification system at the LHCb consists of the RICH1, RICH2, hadronic and electromagnetic calorimeters and the muon chambers. However, at the level of the trigger, this information is not yet available and therefore is not as relevant in this report.

1.3. The Trigger

LHCb generates around 40 million *pp* collisions per second, producing an overwhelming amount of data that cannot be stored offline for further analysis. Therefore, LHCb incorporates a trigger system responsible for selecting and saving specific 'interesting' events. Previously, this involved a level-0 (L0) hardware trigger that relied on crude measurements from the muon chambers and electromagnetic calorimeters. The events that passed this trigger were then passed to a software High-Level Trigger (HLT) for further selection.
However, LHCb is currently in the Phase-I upgrade which focuses on increasing the luminosity, upgrading tracking detectors, particle identification and integrating a fully software trigger [9]. Post-Phase I upgrades, the LHC will supply an instantaneous luminosity of L=2×10^{33} cm^{-2}s^{-1} to LHCb[9]. At this luminosity, various channels in the L0 hardware trigger would saturate due to the 1 MHz output rate limit [10]. This, and the constantly more complex processes being analysed are motivations for the fully software trigger. The software HLT will allow for "unprecedented flexibility" and efficiency for trigger selections [11].
The upgraded trigger is now broken down into two software high-level triggers; the HLT1 and the HLT2. The HLT1 has a similar function to the legacy L0 hardware trigger and is required to reduce the output rate to 1 MHz whilst maximising the amount of specific decays kept. The HLT2 then applies further restrictions to further refine specific selections. This report concentrates on the strategic design and optimization of selection criteria and cuts of the HLT1. It will focus on the selection process for the $K_s \to \pi\pi$ decay briefly before moving on to the $D^0 \to K\pi$ decay.

2. THEORY

2.1. Theory of particle decays

There is a lot of jargon in the particle physics field that can have many interpretations across different projects. Therefore for this report, the following section will define the meaning of several keywords used.

2.1.1. Keywords

Event - An event is classified as each time the two proton beams cross, which is a bunch crossing. An event can have, in general, any number of proton-proton collisions, however, due to the cross-section of interaction, they average around 1-3 per bunch crossing.
Track - In each event, charged particles from the proton-proton collision will leave behind hits in the tracking detectors of the VELO, UT and SciFi, these will be reconstructed to form tracks of particles.
Decay - In the context of this report, as it only considers two-body decays, a decay is defined as a pair of two tracks. These are not guaranteed to be actual decays (as they could be two random tracks), however, this report will classify each pair of particles that are considered to be a 'decay'. Some of these will be from the specified decay and shall be called *true decays*, others could be from the combinatoric background or other similar decays, these will be called *false decays*.
Event Rate - This is the ratio of the total amount of events that are saved against the total number of events in the sample. The LHCb has a bunch-crossing rate of 40 MHz [12], to be able to store the data offline, this needs to be reduced to 1 MHz. Gaps in the beam reduce the bunch-crossing rate to 31.6 MHz [13], therefore, selections need an event rate of under 3%.
Signal purity - Eq. 2 details the signal purity and this is used when there is a way to distinguish between background and signal. This is used in the D^0 selections when there is access to the true identification of a track through simulated data. The change in signal purity is used to evaluate the selections that are applied.
Signal Fraction - In the K_s selections, the nature of each decay is not known and therefore the signal fraction is used to estimate the signal purity of the distribution. It is formulated in Eq. 1, where the mass region is defined as (498 MeV/c^2 $\pm$ 11 MeV/c^2).
Signal Efficiency - Eq. 3 shows the formula for signal efficiency which is another important ratio. It indicates that of the total 'true' decays of interest in a sample, how many are kept by the selection applied. The two most important numbers for the selections are signal efficiency and event rate. The event rate must be below 3% and the signal efficiency must be as high as possible. The other fractions are used to give insight into how effective each cut is.

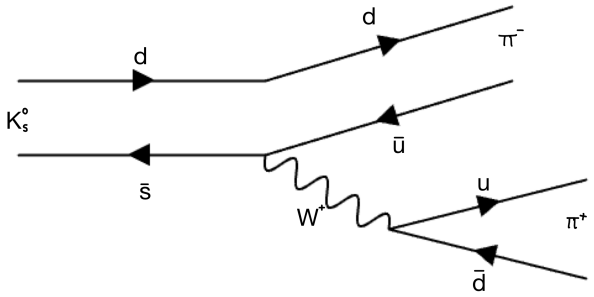


FIG. 2: Feynman diagram of the $K_s \to \pi\pi$ decay

$$\text{Signal Fraction} = \frac{\text{Decays inside mass window}}{\text{Total Decays}} \quad (1)$$

$$\text{Signal Purity} = \frac{\text{True } D^0 \text{ decays}}{\text{Total Decays}} \quad (2)$$

$$\text{Signal Efficiency} = \frac{\text{True decays kept}}{\text{Total True Decays In Sample}} \quad (3)$$

2.1.2. Particle Decays

This report focuses on two specific particle decays and the selection process required for each one. First,

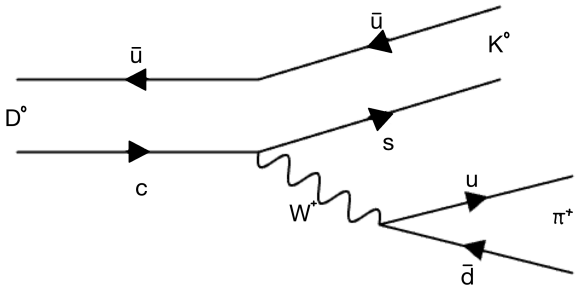


FIG. 3: Feynman diagram of the $D^0 \to K\pi$ decay

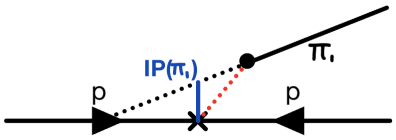


FIG. 4: Diagram showing the impact parameter of a track. Two protons (p), collide at the cross, and a particle is created (red line). π_1 is a daughter track from this particle and its impact parameter is shown to be IP(π_1)

the $K_s^0 \to \pi^\pm\pi^\mp$ decay, shown in Fig. 2 was studied. The invariant mass of the pair of tracks that survive the selections should peak at the mass of the K_s, which is 497.6 MeV/c^2 [14]. The K_s is a commonly produced particle and therefore has a very strong signal [15]. This makes the decay easier to study as changes in signal fraction and event rate can be seen visually. For these reasons, it was chosen as a way to gain familiarity with the trigger and the selection process.

The other decay studied is $D^0 \to K^\pm\pi^\mp$. The motivation behind studying this decay is that the D^0 meson is one meson that exhibits Charge-Parity (CP) violation [16]. CP violation is of great interest to the LHCb experiment due to its implications in understanding the asymmetry between matter and antimatter. The phenomenon involves the violation of the charge and parity symmetries in certain particle interactions, leading to observable differences in particles and their antiparticles. By investigating CP violation, LHCb aims to unveil discrepancies in the behaviour of matter and antimatter, which could potentially explain why the universe is predominantly made up of matter while antimatter is notably scarce. Understanding the mechanisms behind CP violation could provide crucial insights into fundamental physics principles, the Standard Model and one of modern cosmology's most profound mysteries. Despite this, it is unlikely that CP violation will be found in this decay. $D^0 \to K\pi$ is Cabbibo-favoured and therefore any smaller contributions (from $D^0\bar{D}^0$ mixing) are relatively suppressed. However, the $D^0 \to K\pi$ decay will be used to gain an understanding that can be applied to multi-body decays. The D^0 is a heavier particle than the K_s, having a rest mass of 1864.8 MeV/c^2 [14]. It is also much shorter lived, having a proper lifetime of 4.10×10^{-13}s, compared to the K_s lifetime of 8.95×10^{-11}s. The D^0 meson also has a much weaker signal with around one charm quark being produced every five events [17]. These reasons make it a more challenging decay to select.

2.2. How the HLT1 algorithm will work

The HLT1, as previously discussed, operates entirely on software, meaning that it can reconstruct each event. At the selection stage, track reconstruction has been completed, granting the algorithm access to track variables like impact parameter and 4-momentum. However, data from the particle identification system, used to further identify particles, is not yet available at this level of the trigger. Initially, the algorithm reads in all of the tracks from an event, it will then iterate through the tracks and discard any that don't individually meet the conditions outlined below. Tracks meeting these criteria are then evaluated as pairs against further cuts. If at least one pair of tracks from an event survives all of these selections, the whole event is saved and contributes to the event rate. In this report, every pair of tracks that survives the selections is saved and its invariant mass, 4-momentum and other properties are used in the graphs in this report. The selections that will be applied are detailed below.

2.3. Theoretical reasons behind cuts

There are two main types of background signals in the data, the combinatoric background and actual decays that are similar to the specified one. Combinatoric background originates when you calculate the invariant mass from two random tracks and there is a probability of it being close to the specified mother particle's mass. Actual decays on the other hand are more likely to get through these initial selections, however, they will peak at a different invariant mass. The main thing to do in the initial selections is to reduce the combinatoric background as much as possible, this would overwhelm most decay signatures due to the sheer number of combinations of tracks. Theoretical reasoning can be applied to the introduction of several loose cuts aimed at reducing the number of background decays that pass through the trigger.

2.3.1. Momentum

Particle momentum is distributed exponentially, with daughter particles from massive particle decays having a higher average momentum than commonly produced slow-moving pions [18]. Therefore, implementing a minimum momentum cut will significantly reduce the number of tracks needing analysis in the second phase of the trigger, effectively removing a larger amount of background compared to true decays.

2.3.2. Transverse momentum

The transverse momentum of a particle (p_T) is its momentum in the plane perpendicular to the beam axis. It is easier to think about instead of selecting interesting events, to think about not selecting background data. The majority of background or uninteresting decay will happen close to the beam axis with low transverse momentum. The more interesting decays will require a larger exchange of momentum and therefore will be more likely to have a higher transverse momentum.

2.3.3. Impact Parameter

The impact parameter of a particle refers to the distance between a track and primary vertex in the transverse direction as shown in Fig. 4 [19]. A large

impact parameter suggests that the daughter tracks do not originate from the primary vertex, indicating that their mother particle has travelled some distance before decaying. This reduces the abundance of tracks that are created in the pp collisions, which are not of interest. The impact parameter chi-squared, χ^2_{IP}, is proportional to $\chi^2_{\text{IP}} \propto [\text{IP}/\sigma(\text{IP})]^2$ and is a more useful quantity for the selections. By setting a high χ^2_{IP}, the cut requires the ratio between the impact parameter distance and its error to be high, ensuring that only tracks that are sure to have been displaced are kept.

2.3.4. Second Stage - Combining tracks

These three cuts can be applied at the first stage of the selection process on the individual tracks, ensuring that in the second part of the selection process, there are considerably fewer tracks to analyse. This is important due to the time complexity of the two stages. The first stage has a linear time complexity of $O(N)$. The second stage finds every pair in its array of tracks and therefore, has a time complexity of $O(N_1^2)$, where N is the number of tracks in a given event and N_1 are the number of tracks that pass the first stage of the selection process. Reducing N_1 is therefore essential for the trigger to be able to function at the speed of the LHC. This report will now focus on the selections applicable to a pair of tracks. The most common-sense cuts to apply on a pair of tracks is to make sure that the tracks are near each other, and have opposite charges. To ensure the tracks are near each other a distance of closest approach (DOCA) [19] of 0.2mm is applied. This is the standard at the LHCb [15]. The HLT1 then does a quick selection that discards any pairs of tracks that have the same charge, as these would not have come from a neutral particle. Next, the HLT1 ensures the secondary vertex position is well-defined with a $\chi^2/\text{NDOF} < 10$ cut where NDOF is the number of degrees of freedom the position of the secondary vertex has.

2.3.5. Direction Angle

Another common sense cut that is implemented utilizes the direction angle (DIRA) of the reconstructed mother particle [19]. This angle is shown in Fig. 5, where θ is the angle between the direction of the reconstructed momentum and the direction between the measured primary and secondary vertices. The DIRA is then the cosine of this angle.
If the two tracks that are paired together have come from the same two-body decay, the resulting direction angle should be very small. As DIRA is the cosine of this, it will be close to one. This reasoning allows for a very good selection to be applied that removes the majority of multi-body (>2) decays and combinatoric background decays. Furthermore, negative DIRA values indicate that the secondary vertex is behind the primary vertex, and the mother particle was travelling in the opposite direction to both its daughter particles. This is either due to the tracks being assigned the wrong primary vertex, or them not belonging to a decay at all. Implementing this cut is crucial for this reason.

2.3.6. Mother Particle Impact Parameter

In the D^0 selections, a cut is introduced on the impact parameter of the mother particle, ensuring that the mother particle originates from the primary vertex (PV) or shortly thereafter. Therefore, the selections primarily focus on prompt D^0 mesons, which are particles that have been created at the primary vertex [20]. It's worth noting that this selection criterion is dependent on what type of D^0 are of interest to the experiment. Nonetheless, for this report, the priority lies in evaluating the purity, efficiency, and event rate of the selections. An impact parameter chi-squared cut is not applied here as it is not yet available in the LHCb upgrade at the level of the high-level trigger. [15]. This selection is not applied in the K_s selections, as kaons are more easily produced from secondary and tertiary particles due to their lighter mass.

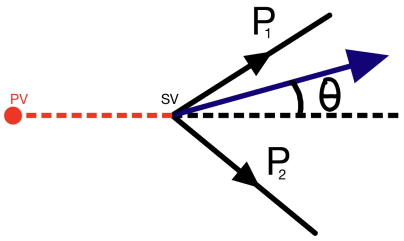


FIG. 5: Diagram showing the direction angle of a reconstructed particle. The blue arrow represents the sum of momenta of the daughter particles P_1, P_2. DIRA is then equal to $\cos\theta$

3. $K_s \to \pi\pi$ SELECTION ANALYSIS

3.1. Data Sets

In this project, two data sets were used. One data set contains data collected during Run 3 and is used in the K_s selections. This data set had 4,414,005 events and is equivalent to around 0.1 seconds of data taking.
For the D^0 selections, the signal is much weaker and therefore an additional data set was required. A Monte Carlo simulated data set was generated that guaranteed at least one $D^0 \to K\pi$ decay to be produced per event. The simulated data enabled access to the true identification of tracks and their respective mother particles. This new information enhanced the accuracy of the D^0 signal purity and efficiency calculations. There are 85421 events in this data set. The D^0 selections were applied to the Monte Carlo data set to determine the signal purity and efficiency, and then to the real data set to determine the event rate.

3.2. $K_s \to \pi\pi$ Selection Process

At the start of the project, the selections applied were focused on getting the most pure signal possible. It was only later on during the study of D^0 decays that the importance of event rates and selecting as many true decays were highlighted. The metrics signal fraction, event rate, and the number of 'true' K_s decays selected are used to quantify each selection. The number of 'true' K_s decays selected is equivalent to the signal efficiency described in Eq. 3. The signal fraction as described in Eq. 1 is an estimate of the signal purity and was the main metric evaluated during this part of the project.

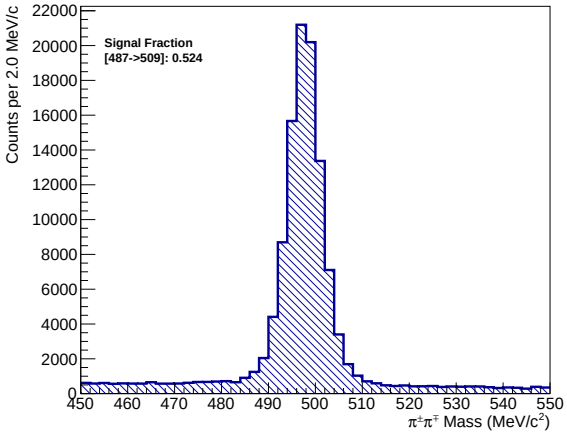


FIG. 6: Plot showing the $\pi^{\pm}\pi^{\mp}$ invariant mass distribution with cuts from Table. I

Cuts applied	Signal Fraction	Event Rate	Total 'true' K_s decays
$p_T > 200$ MeV/c $p > 1500$ MeV/c MIN $\chi^2_{\text{IP}} > 30$ DIRA> 0.8	0.524	2.02%	100047

TABLE I: Inital Selections for the $K_s \to \pi\pi$ decay. Applied to the real data.

3.2.1. Inital cuts

The initial cuts that were chosen are laid out in Table. I. The transverse momentum cut p_T was chosen to be 200 MeV/c, slightly less than half the K_s mass. The momentum cut of $p > 1500$ MeV/c is justified by the minimum energy required for a charged particle to make it through the bending magnet is 1.5 GeV/c [21]. In reality, the data for this run has no tracks with a momentum of less than 3000 MeV/c, as seen in Fig. 7a. The DIRA cut was chosen to be 0.8 initially to further justify and investigate the theoretical reasons laid out in Section. 2.3. These initial selections resulted in the invariant mass distribution shown in Fig. 6 and a signal fraction of 52.4% Next, an analysis of how individual cuts can be improved to increase the signal fraction was performed. To this a high ratio of background to signal needs to be removed in each cut.

3.2.2. Momentum Selections

Fig. 7a reveals that there are no decays registered below 3000MeV/c as discussed earlier. The cut can then easily be increased to 3000 MeV/c without removing any K_s decays that were in the sample already. Fig. 7b shows the distribution of transverse momentum against invariant mass. The main region of the signal is around $p_T = 350 - 400$MeV/c. Increasing the p_T cut from 200 to 350 will remove comparatively more background than the signal, increasing the signal fraction. Increasing this further still increases the signal fraction, but at the cost of signal efficiency.

3.2.3. Direction Angle - DIRA

It is clear that in Fig. 7d, a cut on DIRA will improve the selection greatly, due to the extreme contrast between signal and background regions. Table. II shows the effect of increasing the DIRA cut. The value of 0.9999993 is selected because Fig. 7d shows that a cut here would remove the majority of combinatoric background, whilst keeping the main concentration of signal. This is reflected in Table. II with signal fraction increasing to 80.38%.

Cuts applied	SF	Event Rate	True Decays
Table. I cuts + DIRA> 0.9966	58.87	1.816	98589
Table. I cuts + DIRA> 0.999	66.03	1.595	97054
Table. I cuts + DIRA> 0.9999993	80.38	1.064	77480
Table. I cuts + DIRA> 0.9999999	79.08	0.648	46226

TABLE II: Table showing the effect of changing the DIRA selection on signal fraction (SF), event rate and the number of decays inside signal region

3.3. Track impact parameter chi-squared (χ^2_{IP})

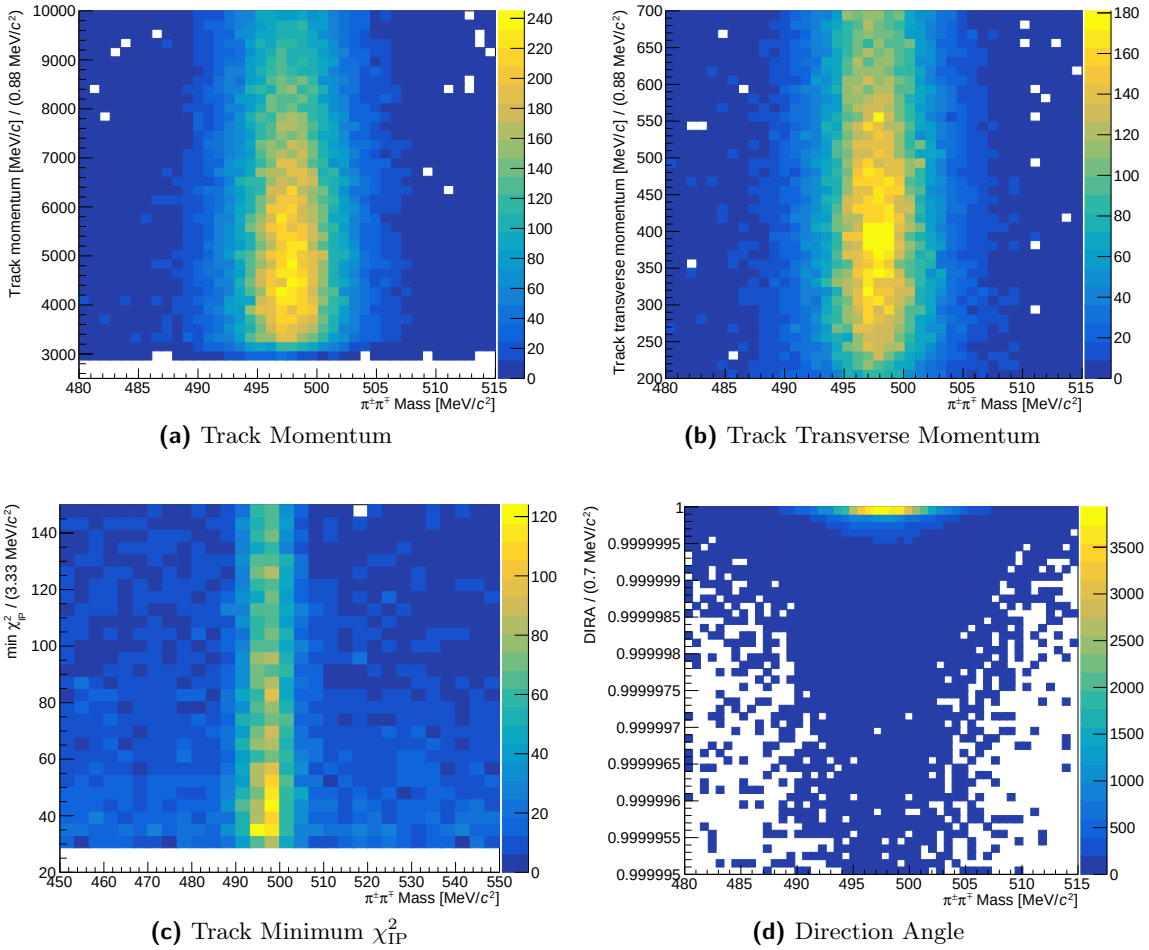
The χ^2_{IP} distribution, shown in Fig. 7c, has less distinction between signal and background. In each horizontal slice of the distribution (between 30-40 for example) there is a smaller ratio of signal-background. This means that the χ^2_{IP} are good cuts to remove a large number of background events. Due to the exponential distribution of impact parameters for true K_s decays, increasing the impact parameter cut here only takes away a relatively small amount of signal.

3.4. K_s Selection Results

This part of the project involved a lot of trial and error and iteration. One of the most pure distributions that were found were with selections $p_T >$500MeV/c, $p >$3400 MeV/c, $\chi^2_{\text{IP}} >$100 and DIRA < 0.9999993. This resulted in a signal fraction of 92.8%, as shown in Fig. 8. Despite being a very pure distribution, it is not a good selection, as the event rate is down to 0.352%, or 141KHz. This is more of a focus when selections were made for the D^0 decay.

4. $D^0 \to K\pi$ SELECTION ANALYSIS

A similar method to the K_s selection was applied to the D^0 selections. However, certain differences in the kinematics of the two decays had to be accounted for. The D^0 meson is a much shorter-lived particle than the K_s, having a lifetime almost 200 times as short. This means that the impact parameter chi-squared of the daughter tracks will be much smaller. Secondly, due to the much larger mass of the D^0 meson, it will have a larger average track transverse momentum and larger average track momentum. In this section, each cut will be analysed separately and a minimum and maximum value of the cut will be stated. These will indicate the region where an optimal selection can be found. This will leave one relaxed and one strict array of selections, in which there will be an ensemble that uses a combination of cuts inside the range given that optimises both signal efficiency and event rate.



(a) Track Momentum

(b) Track Transverse Momentum

(c) Track Minimum χ^2_{IP}

(d) Direction Angle

FIG. 7: 2D histograms of various properties against the $\pi^{\pm}\pi^{\mp}$ invariant mass. Existing cuts on the data from Table. I.

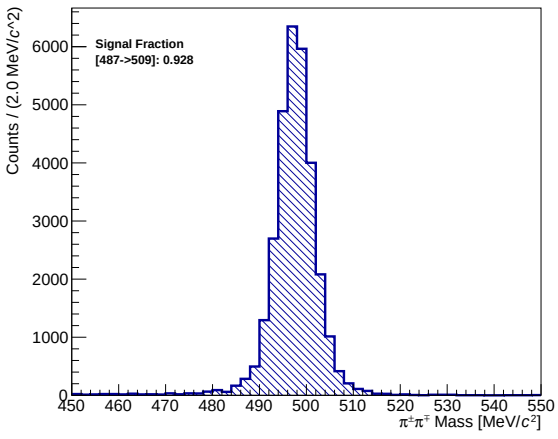


FIG. 8: Invariant Mass Distribution for best K_s selections: p_T >500MeV/c, p >3400 MeV/c, χ^2_{IP} >100 and DIRA < 0.9999993

4.1. Mass combinations

The $D^0 \to K\pi$ decay is different from the $K^0_s \to \pi\pi$ decay in more than just kinematics. The D^0 meson decays into distinguishable particles, the kaon and pion, whereas the K_s decays into two indistinguishable pions. This is significant as the invariant mass is dependent on which mass is assigned to which track, and considerable time was spent evaluating how to deal with this.

Saving the wrong combination of invariant mass could skew the data, causing true D^0 decays to fall outside the mass window and hence look like background. This aspect was a larger problem before there was access to particle identification in the Monte Carlo data set, nonetheless, it is something important to consider. There was evidence that the higher momentum particle would be selected first in the detector and therefore may be more likely to be the first particle in the algorithm [15]. Following this it was shown that due to the relativistic boost, a higher mass particle is likely to have a higher momentum in a two-body decay at this energy range.

4.1.1. Two Body Decay Kinematics

In the rest frame of a two-body decay, each daughter particle will have equal and opposite momenta due to the conservation of momentum. This can be represented as a 4-vector. In the case of the $D^0 \to K\pi$ decay, the resulting 4-momenta are

$$\tilde{P}_K = (\frac{1}{2}M, \mathbf{p}), \quad \tilde{P}_\pi = (\frac{1}{2}M, -\mathbf{p}). \tag{4}$$

These momenta allow the energies of the two parties to be calculated as;

$$E_K = \frac{M^2 + m_K^2 - m_\pi^2}{2M^2}, E_\pi = \frac{M^2 + m_\pi^2 - m_K^2}{2M^2}. \tag{5}$$

When the particles are boosted into the lab frame they undergo a Lorentz transformation and become

$$p'_{K_x} = \gamma(p_{K_x} + vE_K) \tag{6}$$

$$p'_{\pi_x} = \gamma(p_{\pi_x} + vE_\pi). \tag{7}$$

As the energy of the K_s in the rest frame of the D^0 will always be bigger than the energy of the pion, the kaon is more likely to have more momentum in the lab frame [22]. An investigation was performed into whether the order of selection was related to the lab-frame momentum. This was shown to be incorrect, the lab-frame momentum is not related to the order of selection and therefore it was decided that for each selection each combination of mass will be saved. As the algorithm uses a nested for-loop, as shown in the pseudocode below, this can be done by assuming that *Particle 1* is the kaon. This works as the nested for-loop considers both permutations of particles so it is guaranteed to save the correct one. By saving both it also levels out any bias inherent in the system.

```
for Particle 1 in tracks:
    for Particle 2 in tracks:
        Apply Selections
        if Pass Selections:
            Save Event
```

4.2. $D^0 \to K\pi$ selection process

During this process, cuts will be justified and analysed individually, with a maximum and minimum range set. The cuts will be analysed on the Monte Carlo data, as there is less background and access to the true particle identification from the simulation. This enables the signal purity and signal efficiency to be known so cuts can be more quantitatively analysed. Once a set of minimum cuts and a set of maximum cuts are defined these will be run on the real data set to determine the event rate. One iteration within the ranges set on each cut will be shown to have an event rate lower than 3% and a high signal efficiency.

4.2.1. Initial Cuts

Due to the differences in kinematics discussed above, the initial selection has been changed to be p_T >300, p >3000, χ^2_{IP} >3. These initial selections, even in the Monte Carlo data set, leave an overwhelming amount of background, shown in Fig. 9. The selections kept 18752 true D^0 decays. As there were 85421 events in the Monte Carlo data and there is at least one $D^0 \to K\pi$ decay in each event this corresponds to a maximum signal efficiency of 21.95%. These minimal cuts have already reduced the number of true D^0 decays by 78.05%. This is mainly because the D^0 is governed by the exponential law of decay and despite having a longer lifetime than most background, the majority of D^0s created, still decay very close to the primary vertex, and thus its daughter tracks have much smaller χ^2_{IP}. Due to the amount of background, these cuts were increased to p_T >400, p >4000, χ^2_{IP} >4 for analysis.

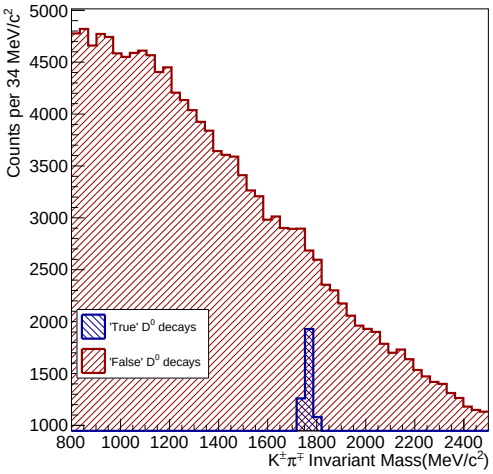


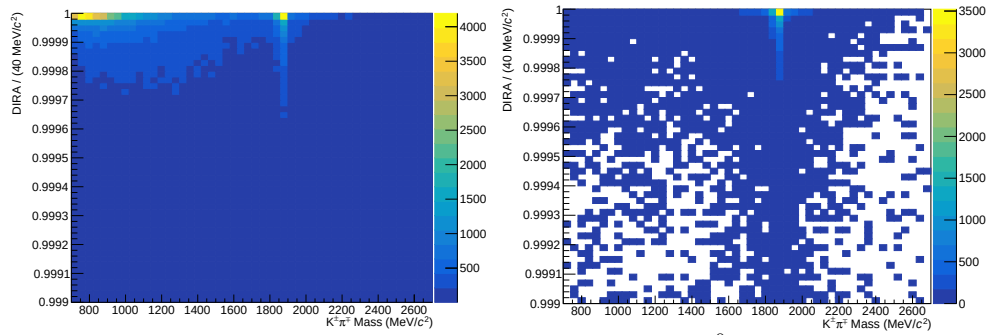
FIG. 9: Initial Cuts of p_T >300, p >3000 and χ^2_{IP} >3, applied to the Monte Carlo Data.

4.2.2. Direction Angle - DIRA

The first selection applied was DIRA, as in the K_s data it was one of the most effective cuts. Making the direction angle positive, as discussed previously in Section. 2.3, removes a large percentage of combinatoric background whilst keeping the majority of signal. Starting with the refined initial selections of p_T >400MeV/c, p >4000MeV/c, track χ^2_{IP} >4, the effect of increasing the DIRA cut can be studied. Table. III shows the purity of decays remaining after the cut, the purity of the decays that have been removed since the previous cut and the signal efficiency. Applying the cut of DIRA> 0.95 reduces the total decays selected by 44.1%. Of these decays removed, only 1.15% are actual D^0 decays. This is a very successful cut and will be the minimum DIRA cut required. If this minimum was lowered it would not be enough to reduce the output rate to 1 MHz. Fig. 10b shows that this cut can be pushed much higher to further isolate the signal. The maximum DIRA cut will be set at 0.9997, as beyond this point 14.81% of the decays being removed are signal. This is a large sacrifice, and results in a signal efficiency of 8.85%. The number of decays kept between the DIRA cut of 0.95 and 0.9997 decreases by 61.2%.

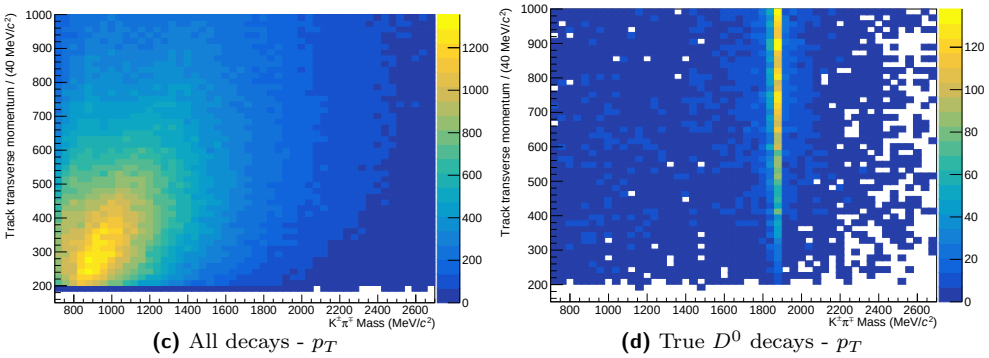
4.2.3. Transverse Momentum

Due to the larger mass of the D^0 compared to the K_s, its decay products are expected to have higher transverse momentum. This can be seen in the difference between Fig. 10c and Fig. 10d, where Fig. 10c includes K_s decays as discussed in Section 4.2.8. A minimum p_T cut of 400 MeV/c efficiently removes background whilst retaining a significant proportion of true D^0 decays. The minimum value of p_T will therefore be set to 400MeV/c. To analyse where the maximum p_T will be set Table. IV is used. This again shows the purity of decays remaining, removed and signal efficiency for increasing p_T cut, with initial selections of (p_T > 300MeV/c, p > 4000MeV/c and track impact parameter chi-squared >4). By 800MeV/c, the purity of decays removed has increased to 10.3% and the signal efficiency decreased to 10.45%. The maximum p_T will be set here as in-



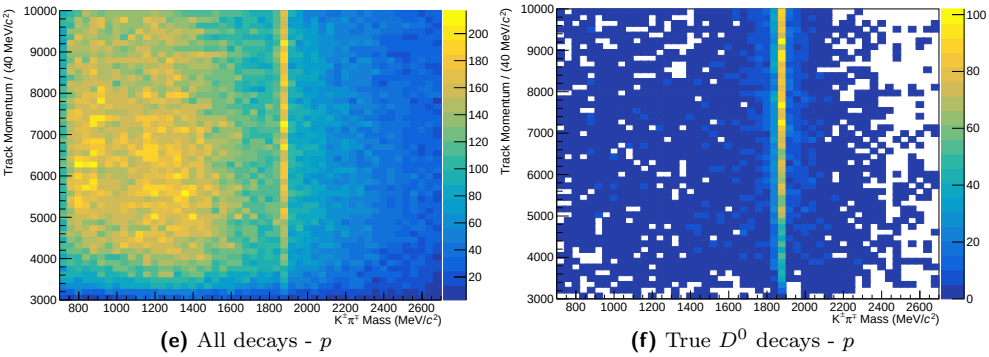
(a) All decays - DIRA

(b) True D^0 decays - DIRA



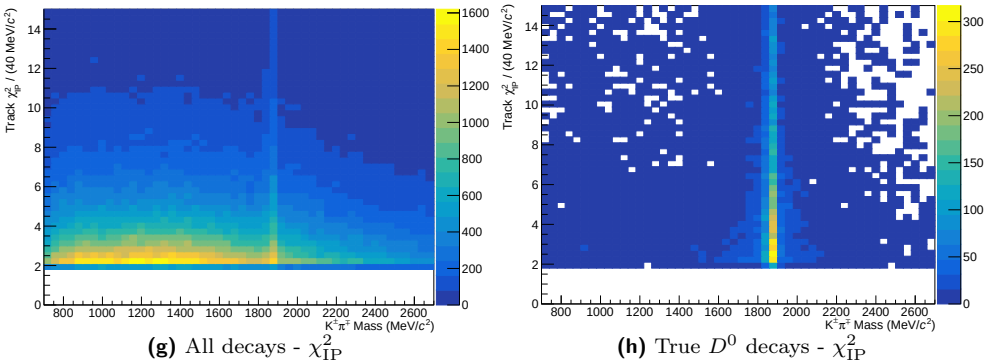
(c) All decays - p_T

(d) True D^0 decays - p_T



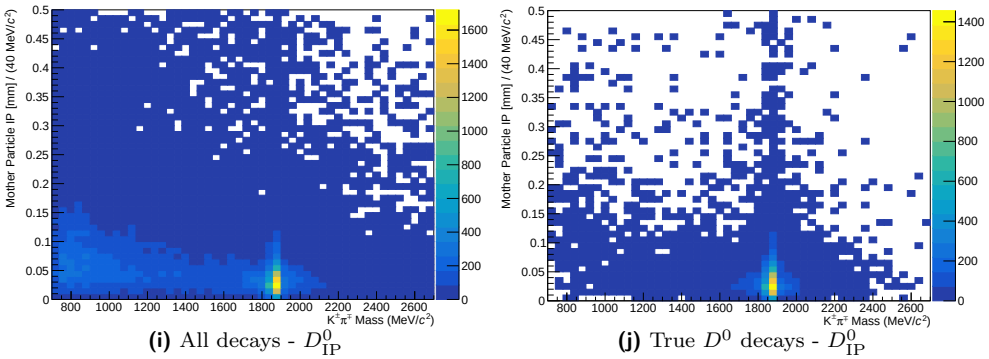
(e) All decays - p

(f) True D^0 decays - p



(g) All decays - $\chi^2_{\rm IP}$

(h) True D^0 decays - $\chi^2_{\rm IP}$



(i) All decays - $D^0_{\rm IP}$

(j) True D^0 decays - $D^0_{\rm IP}$

FIG. 10: 2D histograms of the given variable against the pair of tracks invariant mass. Selections of $p_T > 200\text{MeV}/c$, $p > 3000\text{MeV}/c$, track $\chi^2_{\rm IP} > 2$ applied to the Monte Carlo data set. All of the decays are plotted on the left. Only true D^0 decays are plotted on the right.

creasing the cut beyond this will remove more signal with diminishing returns on the distribution.

4.2.4. Momentum

The momentum of the tracks, depicted in Fig. 10e, have more dispersion than their track transverse momentum counterpart. This means that each selection will remove a higher percentage of the signal and will therefore be a less effective cut. Using Fig. 10f a minimum cut of 4000 MeV/c can be set, however, this already removes a sample of data that has a purity of 7.61%. Increasing the cut to 6000 MeV/c removes data with a purity of 8.50% and results in a signal efficiency of 13.34%. It is clear from this that the momentum is not as an effective cut as DIRA and p_T and therefore the maximum will be set at 6000 MeV/c.

4.2.5. Track Impact Parameter Chi-Squared

Due to the nature of exponential decay, most D^0 mesons decay shortly after travelling from the primary vertex, resulting in small impact parameters of the daughter tracks. However, background tracks also tend to have a small impact parameter, as most particles originate from near the primary vertex. Therefore, cuts on the minimum impact parameter chi-squared are indiscriminate and can remove large samples of both signal and background. An impact parameter chi-squared cut of 2 is the minimum that can be allowed without having an immense amount of data, even in the Monte Carlo data set. Table. V shows the change in purity and efficiency from varying the χ^2_{IP} cut. An increase from 2 to 3 reduces the total amount of decays kept by 46.51%, where 1.967% of those removed are actual D^0 decays. While this adjustment is drastic, it is required for the selections to be run on the real date set. The maximum cut will be set at 6, as it results in 15.44% of the total decays being kept.

4.2.6. Mother Partice Impact Parameter

Finally, a selection on the impact parameter of the mother particle can impose the criterion that the mother particle had to be created close to the primary vertex. This cut could reduce the number of tracks that contribute to the combinatoric background, by removing tracks that have been produced further away from the primary vertex. Fig. 10i and Fig. 10j show that a cut of D^0_{IP} <0.2mm could remove a large proportion of background data. Table. VI shows the results of these cuts. The minimally restrictive cut will be set at 0.2mm, as it increases purity well whilst only removing 21.93% of the total decays. The maximumly restrictive cut will be set at 0.1mm, as beyond this massively reduces the signal efficiency as expected from Fig. 10j.

4.2.7. Other cuts

Throughout the project, a few other cuts were considered. My MPhys partner worked on calculating the decay distance and lifetime of each mother particle as discussed in her report [23]. This revealed some interesting information that hardly any true D^0 decays had a negative $(PV_z - SV_z)$, where this is the difference between the z-coordinate of the primary vertex and secondary vertex. This is as expected due to the LHCb being a forward direction detector and daughter particles released from backward travelling decays are unlikely to be detected. However, this cut was not applied because a positive DIRA already enforces this. She also did work on a cut on the lifetime of the mother particle. This would have been used to focus on longer-lived $D^0 \to K\pi$ decays. These are potentially more interesting as they would have more time to undergo $D^0\bar{D}^0$ mixing, and potentially exhibit CP violation. However, this was also not considered as it is another cut that would exponentially remove true D^0 decays and made the signal purity worse in some cases.

4.2.8. False mass peak

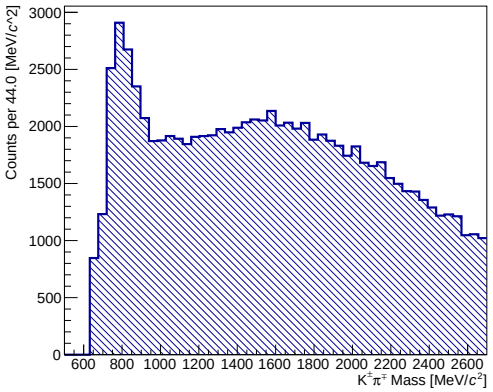


FIG. 11: Shows a mass peak at 800MeV/c in the

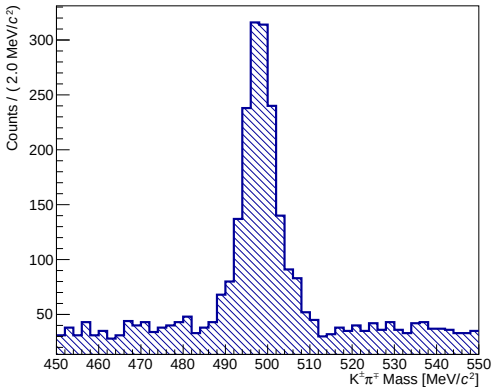


FIG. 12: The Monte Carlo data when the daughter tracks are assumed to be two pions. Shows a peak at the K_s mass

In both the Monte Carlo and real data a second mas peak was observed as illustrated in Fig. 11. To pass the selections the second peak has to have a similar decay signature as the $D^0 \to K\pi$ and have a lighter mother particle. The second mass peak is located at around 800MeV/c^2 so it was hypothesised that these were $K_s \to \pi\pi$ decays. These decays were peaking at a higher mass than normal due to the algorithm assuming that the decay products were $K\pi$, instead of $\pi\pi$. When this assumption was changed to $\pi\pi$, the peak was confirmed to be at the K_s mass, as shown in Fig. 12. These decays were contributing to the event rate and reducing the number of D^0 decays that could be kept. A crude selection was implemented that discarded any permutation that had an invariant mass of below 1200 MeV/c. This limit is high enough to remove any trace of the K_s mass peak. The mass cut also needs to be low enough that a big enough range is given for the distribution to be fitted.

DIRA Selection	Purity of Decays Remaining (%)	Purity of Decays Removed (%)	Signal Efficiency (%)	Total Decays Kept (%)
No DIRA	6.906	N/A	16.76	100
0	10.40	1.075	16.43	64.32
0.95	11.45	1.15	15.53	55.89
0.99	12.90	1.895	15.19	48.50
0.999	18.03	3.590	13.69	31.28
0.9995	20.24	8.007	12.59	25.64
0.9997	21.93	10.99	11.54	21.68
0.9999	25.67	14.81	8.850	14.21

TABLE III: Changes in the signal purity of decays remaining, decays removed and signal efficiency due to an increase in the DIRA cut. Data had selections of $p_T > 400\text{MeV}/c$, $p > 4000\text{MeV}/c$, track $\chi^2_{\text{IP}} > 4$.

p_T (MeV/c) Selection	Purity of Decays Remaining (%)	Purity of Decays Removed (%)	Signal Efficiency (%)	Total Decays Kept (%)
300	9.651	N/A	16.19	100
400	11.45	2.057	15.53	80.86
500	13.62	3.402	14.55	63.71
600	15.87	5.313	13.34	50.10
700	18.08	7.738	11.95	39.40
800	20.26	10.30	10.45	30.76
900	22.58	11.93	9.115	24.07

TABLE IV: Shows the change in the quality of the signal as different transverse momentum selections are applied on the Monte Carlo Data. Initial cuts were: $p > 4000\text{MeV}/c$, track $\chi^2_{\text{IP}} > 4$, DIRA > 0.95

Track χ^2_{IP} Selection	Purity of Decays Remaining (%)	Purity of Decays Removed (%)	Signal Efficiency (%)	Total Decays Kept (%)
2	5.210	N/A	22.11	100
3	8.030	1.967	18.23	53.49
4	11.45	2.955	15.53	31.96
5	15.06	4.325	13.55	21.21
6	18.42	6.061	12.07	15.44
7	21.45	7.842	10.93	12.00

TABLE V: Shows the change in the quality of the signal as different impact parameter chi-squared selections are applied on the Monte Carlo Data. Initial cuts were: $p_T > 400\text{MeV}/c$, $p > 4000\text{MeV}/c$, track $\chi^2_{\text{IP}} > 2$, DIRA > 0.95

Mother Particle IP Selection (mm)	Purity of Decays Remaining (%)	Purity of Decays Removed (%)	Signal Efficiency (%)	Total Decays Kept (%)
No Cut	11.45	N/A	15.53	100
< 0.3	12.54	3.946	14.84	87.27
< 0.2	13.49	5.028	14.22	78.07
< 0.1	19.05	5.020	12.09	46.78
< 0.05	30.52	11.07	7.947	19.20

TABLE VI: Shows the change in the quality of signal with different D^0_{IP} selections are applied on the Monte Carlo data. Inital cuts were $p_T > 400\text{MeV}/c$, $p > 4000\text{MeV}/c$, track $\chi^2_{\text{IP}} > 4$, DIRA > 0.95

5. CONCLUSION

It is clear from the above data that cuts on the direction angle, transverse momentum and impact parameter of the mother particles are the most effective for selecting the D^0 decay. They have the lowest signal purity of decays removed and therefore can optimise both signal efficiency and event rate. Using the results above, one iteration of cuts was found to have a high signal efficiency whilst being below the event rate of 3%. Cuts at $p_T > 500\text{MeV}/c$, $p > 4000\text{MeV}/c$, track χ^2_{IP} >3, DIRA>0.995 and D^0_{IP} <0.15mm were chosen. This resulted in a signal efficiency of 13.62% with an event rate of 2.933%, as shown in Table. VII and Fig. 14. This report has explored the variables that the LHCb detects and how they can be used within a software high-level trigger to select events that include the $K_s \to \pi\pi$ and $D^0 \to K\pi$ decays. The report has concluded a range in which an optimal event rate and signal efficiency of D^0 mesons can be found, detailed in Table. VII. The relaxed selections have a signal purity on the Monte Carlo data set of 11.52%, a signal efficiency of 15.46% and a real event rate of 7.27%. This results in an output rate of 2.91 MHz when the beam is running at 40 MHz. The strictest selections have a signal purity of 68.0%, however with only a signal efficiency of 5.64% and event rate of 0.117% or 46.8 KHz. These distributions are shown in Fig. 13. Whilst these selections are not optimised they allow a range to be set in which an optimization can be found. The graphs in Fig. 10 and tables in Section. 4.2.1 give insight into which cuts are more effective. This report has laid the foundations for the importance of a high-level trigger, which cuts are more effective to apply, and provides insight into how to optimise signal efficiency below the output rate limit. The results summarised in this report can be used to further optimise the D^0 signal or to expand onto more complex decays.

In the one iteration given, to get the most amount of $D^0 \to K\pi$ decays below the event rate limit, very relaxed track χ^2_{IP} and p were required, at 3 and $4000\text{MeV}/c$ respectively. This leaves the majority of the cuts to be done in the second stage of the algorithm. Although this results in a large signal purity and efficiency, this could have other problems, due to the time complexity $0(N_1^2)$ of the second stage. If the first stage of the algorithm is not selective enough, there is a risk that the HLT1 system does not decide quickly enough and misses events in the LHCb. This is another thing to consider when going further. For example, if three-body decays are considered, this would the time complexity to $O(N_1^3)$, increasing the magnitude of this issue.

There are a lot of things to continue on this project. One obvious place is to optimise the event rate and signal efficiency further. This is non-trivial as there are at least five different variables to consider and therefore machine learning techniques may be implemented. Another direction is to consider selections on more complex processes, for example, multi-body charm decays. This will be more useful for studying at the frontier of particle physics, and allow more constraints to be applied to the Standard Model. The time efficiency of selections can also be looked into further, as highlighted above.

[1] LHC Facts and Figures (Jan 3rd 2024), accessed at `https://home.web.cern.ch/resources/faqs/facts-and-figures-about-lhc`.

[2] CERN Document Server, Accessed at `https://cds.cern.ch/collection/LHCb?ln=enonDec292023`.

[3] LHCb Collaboration, *The LHCb Detector at the LHC* , Journal of Instrumentation : LHCb-DP-2008-001 (3rd August 2008).

[4] R. Jacobsson, *Performance of the LHCb Detector During the LHC Proton Runs 2010 - 2012*, Tech. Rep. (CERN, Geneva, 2013) LHCb-PROC-2013-030.

[5] LHCb Collaboration, *The LHCb upgrade I*, Tech. Rep. (2023) LHCb-DP-2022-002.

[6] LHCb Collaboration, *LHCb VELO Upgrade Technical Design Report*, Tech. Rep. (2013) LHCb-TDR-013.

[7] LHCb Collaboration, *LHCb magnet: Technical Design Report*, Technical design report. (CERN, Geneva, 2000) LHCb-TDR-1.

[8] P. Hopchev, *SciFi: A large Scintillating Fibre Tracker for LHCb* (2017), arXiv:1710.08325.

[9] Upgrade Schedule Status (17 April 2019), accessed `https://lhcb.web.cern.ch/lhcb_page/lhcb_documents/progress/Source/RRB/April_2019/Written_Upgrade_Status.pdf` (4 Jan 2024).

[10] C. Fitzpatrick and V. V. Gligorov, *Anatomy of an upgrade event in the upgrade era, and implications for the LHCb trigger*, Tech. Rep. (CERN, Geneva, 2014) LHCb-PUB-2014-027.

[11] C. Agapopoulou, *Commissioning LHCb's GPU high level trigger* (2023), LHCb-PROC-2022-002.

[12] LHCb Collaboration, *The 40 MHz trigger-less DAQ for the LHCb Upgrade* (2016), Nucl. Instrum. Methods Phys. Res., A **824**, 280-283.

[13] Collisions at the LHC, accessed at `https://lhc-machine-outreach.web.cern.ch/collisions.htm`, Jan 2 2023.

[14] R.L. Workman *et al.* (Particle Data Group), Progress of Theoretical and Experimental Physics **2020**, 083C01 (2020).

[15] Conor Fitzpatrick and Timothy Evans, Personal Communication.

[16] LHCb Collaboration, *Observation of CP Violation in Charm Decays* (May 2019, month=may), Physical Review Letters **122** 21, American Physical Society.

[17] LHCb Collaboration, *Measurements of prompt charm production cross-sections in pp collisions at* $\sqrt{s} = 13$ *TeV* (2016), JHEP, **03**, 159, (2016), LHCB-PAPER-2015-041, arXiv:1510.01707.

[18] M. R. M. Jacob, *High energy collisions production processes at high energy theory and experiment*, Tech. Rep. (CERN, 1972) CERN-TH-1570.

[19] LHCb Glossary (4 Jan 2024), Accessed at `https://lhcb.github.io/glossary/glossary/D.html`.

[20] The ALICE Collaboration, *The ALICE definition of primary particles* (2017), ALICE-PUBLIC-2017-005.

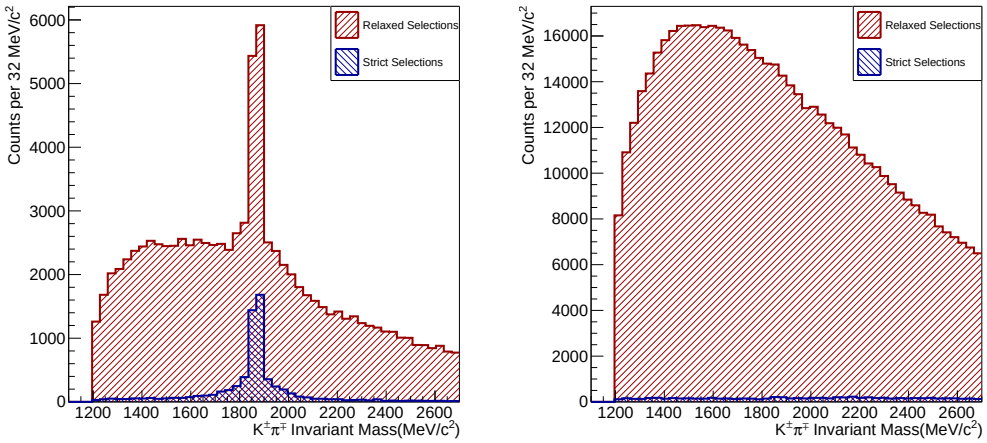
[21] International Journal of Modern Physics A **30**, 1530022 (2015), *LHCb detector performance*, https://doi.org/10.1142/S0217751X15300227.

[22] Hok-Chuen Cheng, *Two body decay kinematics* (May 1 2015), Accessed at `https://indico.cern.ch/event/391122/contributions/928962/attachments/782786/1073126/twoBodyDecay.pdf`.

[23] Elizabeth Peak, *Finding charm at the LHCb; Analysing LHCb trigger selections* MPhys Project (Jan 2024).

Cuts applied	Type of Selection	Signal Purity (%)	Real Event Rate (%)	Signal Efficiency (%)
$p_T > 400\text{MeV}/c$ $p > 4000\text{MeV}/c$ Track $\chi^2_{\text{IP}} > 3$ DIRA > 0.95 D^0 IP < 0.2mm	Most Relaxed	11.52	7.274	15.46
$p_T > 800\text{MeV}/c$ $p > 6000\text{MeV}/c$ Track $\chi^2_{\text{IP}} > 6$ DIRA > 0.9997 D^0 IP < 0.1mm	Most Strict	68.04	0.1173	5.638
$p_T > 500\text{MeV}/c$ $p > 4000\text{MeV}/c$ Track $\chi^2_{\text{IP}} > 3$ DIRA > 0.995 D^0 IP < 0.15mm	One Combination	18.24	2.933	13.62

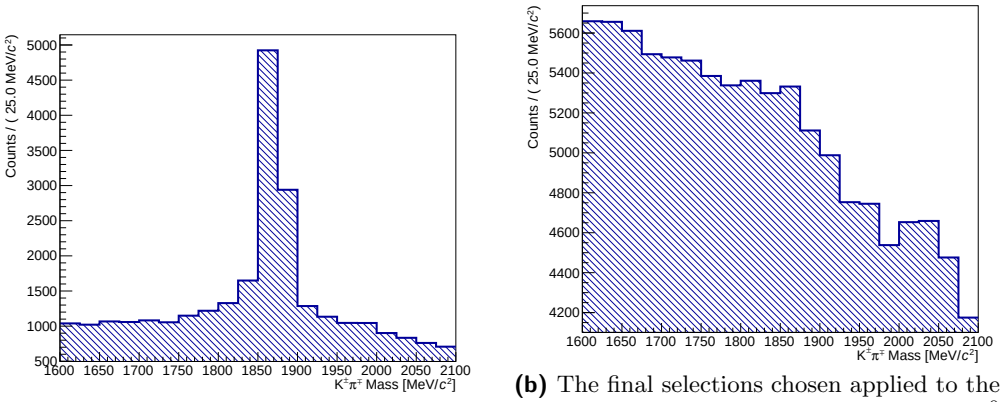
TABLE VII: Table showing the results from the D^0 selection analysis.



(a) Most Relaxed and Most Strict Selections on the Monte Carlo Data

(b) Most Relaxed and Most Strict Selections on the Real Data

FIG. 13: Figures showing the final invariant mass distributions of the arrays of most strict and most relaxed selections



(a) The final selections chosen applied to the Monte Carlo Data

(b) The final selections chosen applied to the Real Data. A slight peak is seen near the D^0 mass.

FIG. 14: The final selections that were chosen to be below the 3% event rate are applied on both data sets.