Analysis of $K_S^0 \to \pi^+\pi^-e^+e^-$ decays using data from Run 3 at the LHCb Experiment

Experimental Physics (EP) Department, LHCb Experiment,
European Organization for Nuclear Research (CERN)

CERN Summer Student Programme 2024

Danna Alondra Herrera Rodríguez
Institute of Technology of Costa Rica (ITCR)

Supervisors:

- Sergio Arguedas
- Titus Mombaecher

August 2024

1 Abstract

This report presents an analysis of $K_S^0 \to \pi^+\pi^-e^+e^-$ decays using data from Run 3 at the LHCb experiment. The performance of the analysis is evaluated through a comparison between real data and Monte Carlo (MC) simulations and employs an offline selection process as well as an alternative approach using a Boosted Decision Tree (BDT) algorithm. The analysis conducted in this project did not reveal any signal peak.

2 Introduction

The study of particle physics has led to important insights into the fundamental forces and subatomic particles that constitute matter. Among the most intriguing topics in this field is the analysis of kaons, since the discovery of CP violation was first made through the analysis of $K_L^0 \to \pi\pi$ decays.

Although the $K_S^0 \to \pi^+\pi^-e^+e^-$ decay is a rare process, it is frequent enough to be observed by LHCb and to be thoroughly studied. Due to this, this decay serves as an excellent control channel, providing a well-understood reference point for studying very rare Kaon decays where electrons are present in the final state, such as the $K_S \to e^+e^-e^+e^-$ decay that has never been observed. [1, 2].

The LHCb experiment is primarily designed to investigate the decays and interactions of particles containing b- or c-quarks, which are essential for studying phenomena like CP violation and rare decays within the framework of the Standard Model. However, due to the enormous number of particles produced in high-energy proton-proton collisions at the LHC, the LHCb experiment also observes an abundance of particles containing s-quarks. This has enabled LHCb to achieve remarkable results in the study of rare strange decays, including those involving K_S^0 mesons [1].

3 The LHCb Detector

The Large Hadron Collider beauty (LHCb) experiment is a specialized detector at the Large Hadron Collider (LHC) designed to study heavy-flavor physics, focusing primarily on the behavior of beauty and charm hadrons. Its main objective is to search for indirect signs of new physics beyond the Standard Model by investigating CP violation and rare decays of these particles [3].

The LHCb detector is designed to capture the large volume of B meson produced. For this to be achieved, precision in vertex and momentum resolution is essential to get the required accuracy in studying rapidly systems which is necessary to minimize background noise. The detector's layout is a single-arm spectrometer with forward angular coverage, designed to capture the majority of b and anti-b hadrons produced at high energies, which typically move in the same forward or backward direction. It is constituted of a Magnet used for determining the momentum of particle by bending their path, as well as a Vertex Locator to track the particle interaction points [3, 4].

Also, the particle trajectory reconstruction is enabled by the Tracking System featuring the Upstream Tracker (UT) and the Scintillating Fibre Tracker (SciFi), while the Ring Imaging Cherenkov Counters (RICH1 and RICH2) use specific radiators to achieve separation of pions and kaons over a wide momentum range; the energy of the particles is measured by the Calorimeter System, and the muons are identified by the Muon Detector System [5]. These configurations can be observed in Figure 1.

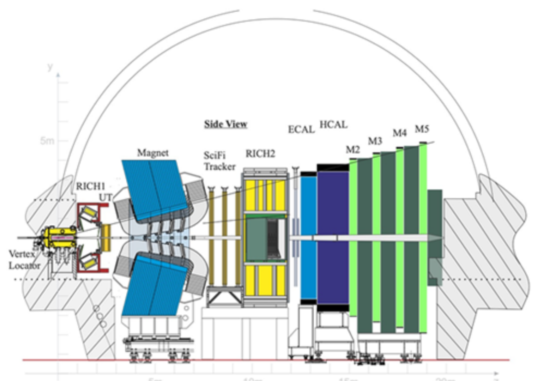


Figure 1: LHCb detector [5]

4 Offline selection

The offline selection includes cuts on various variables, which are crucial for reducing background noise and ensuring that the events are indeed K_S^0 decays. By carefully refining the selection criteria, the analysis maximizes the signal-to-noise ratio, enabling the extraction of meaningful results.

Monte Carlo (MC) simulations play a critical role in understanding complex particle interactions and experimental setups. In the context of the LHCb experiment, MC simulations are employed to model the behavior of particles, simulate their decays, and compare the resulting data with real experimental observations. The goal of these simulations is to improve the accuracy of the analysis, understand potential sources of background, and optimize the selection criteria for identifying signal events.

By applying selection criteria to both simulated signal events and real data, the efficiency of their cuts can be determined. Signal efficiency is the fraction of true signal events that pass the selection criteria; by optimizing these criteria using MC simulations, the analysis can be tuned to maximize signal efficiency while minimizing background contamination.

5 Project Overview

As a starting point, the analysis was based on the data from Run 2 of the LHCb experiment, conducted in 2016. During this period, 100 events of K_S^0 decays were observed with a total integrated luminosity of 3 inverse femtobarns (fb^{-1}) [1]. Based on these previous results, it was estimated that the pretechnical stop dataset used in this analysis, with a total luminosity of 1.85 fb^{-1}, should contain around 61 K_S^0 decay events, and for the posttechnical stop, with a luminosity of 1.5 fb^{-1}, should be around 50 events.

In order to model the K_S^0 decays, MC simulations were used to generate a sample providing a controlled environment in which to test the analysis methods. The comparison with the real data was made by simulating the decay of K_S^0 $\rightarrow \pi^+\pi^-e^+e^-$ using this tool. The histograms for the mass of the K_S^0 resulting from the simulation and the data before any offline selection, are observed in Figure 2 and Figure 3, respectively.

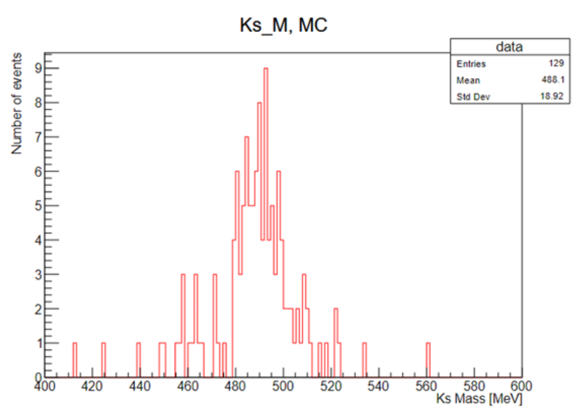


Figure 2: Histogram of the K_S^0 mass variable resulting from the simulation.

The data were collected in two primary configurations: magnet-up and magnet-down. These configurations refer to the orientation of the magnetic field within the LHCb detector, which affects the trajectories of charged particles. By analyzing data from both configurations, it can be ensured that the results are robust and not dependent on a specific magnetic field orientation.

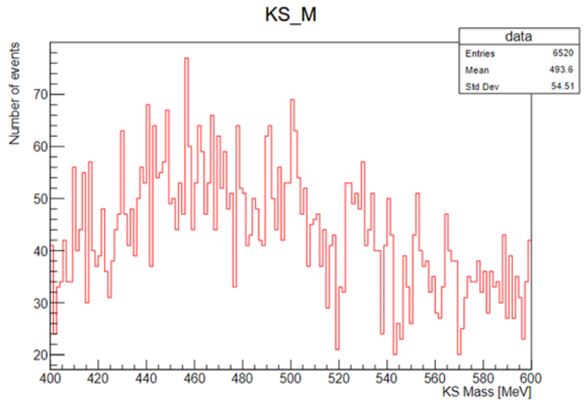


Figure 3: Preliminary distribution of the K_S^0 mass without any selection cuts, using a single ROOT file from the pretechnical stop dataset.

5.1 Analysis Methods and Results

5.1.1 Selection of Variables for Analysis

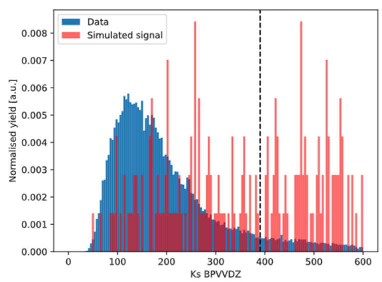
The choice of variables for analysis is a critical step in distinguishing between signal and background events. The goal is to identify variables where the distribution of the signal differs significantly from that of the background, as can be observed in Figures 4, 5 and 6.

For this analysis, several key variables were considered including the distance along the z-axis between the primary vertex (where the initial collision occurred) and the decay vertex (where the K_S^0 decayed). The lifetime of the K_S^0 meson is a crucial variable as well. By selecting events with lifetimes within a certain range, the purity of the sample can be enhanced, focusing on events that are more likely to correspond to K_S^0.

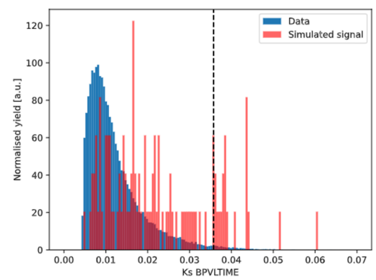
The cosine of the angle between Momentum and Flight Direction measures the alignment between the particle's momentum and the direction it traveled before decaying. A high cosine value indicates that the momentum and flight direction are closely aligned, which is expected for K_S^0 decays. The distance between the primary and decay vertices in the transverse plane (perpendicular to the beam axis) is also studied, as well as other angles between the particles. The list of variables can be seen in Table 1.

5.1.2 Cut Optimization Process

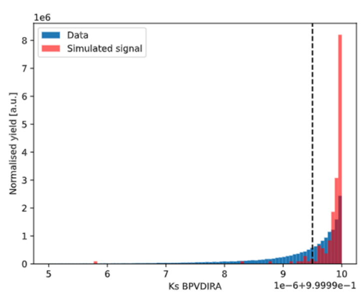
Once the relevant variables were selected, the next step was to optimize the cuts applied to these variables. The cuts are criteria used to filter the data, selecting events that are likely to correspond to the signal while rejecting background



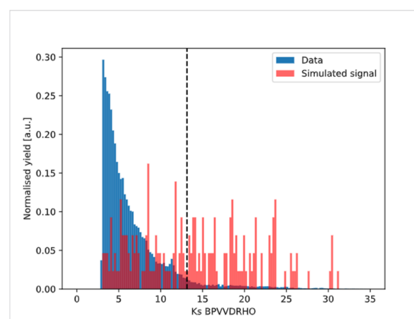
(a) BPVVDZ.



(b) BPVLTIME.

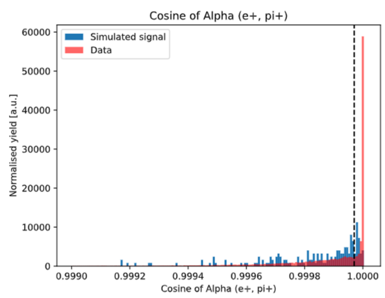


(c) BPVDIRA.

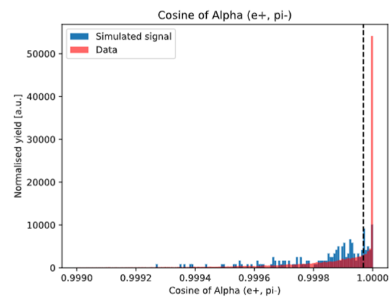


(d) BPVVDRHO.

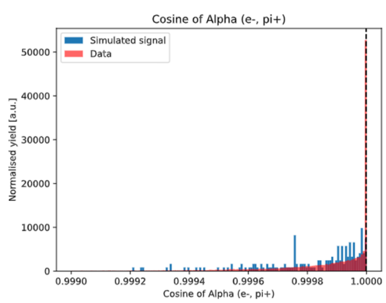
Figure 4: Histograms of BPVVDZ, BPVLTIME, BPVDIRA, and BPVVDRHO variables for the simulation and the data, resulting from 400 ROOT files for the Magnet Down configuration and 400 ROOT files for the Magnet Up configuration pretechnical stop.



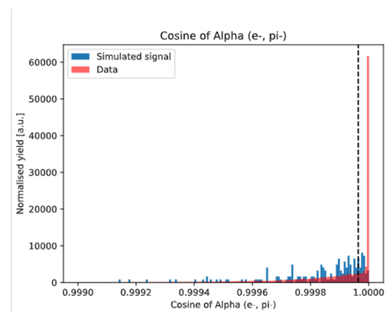
(a) Cosine of the angle between e^+ and π^+.



(b) Cosine of the angle between e^+ and π^-.



(c) Cosine of the angle between e^- and π^+.



(d) Cosine of the angle between e^- and π^-.

Figure 5: Histograms of the cosine of angles between e and π for the simulation and the data, resulting from 400 ROOT files for the Magnet Down configuration and 400 ROOT files for the Magnet Up configuration pretechnical stop.

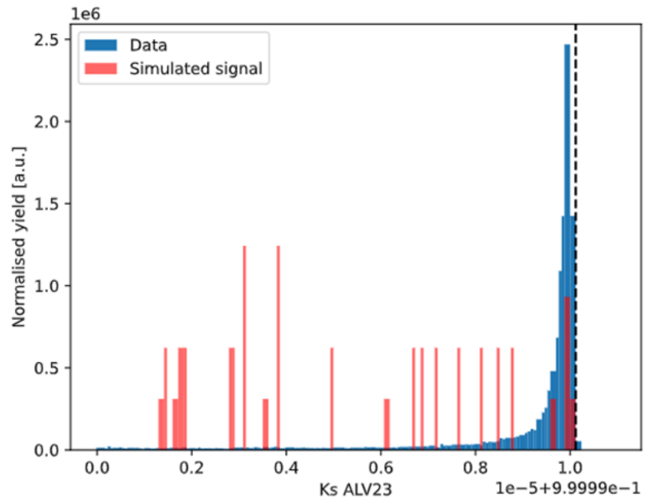


Figure 6: Histogram of the ALV23 variable for the simulation and the data, resulting from 400 ROOT files for the Magnet Down configuration and 400 ROOT files for the Magnet Up configuration pretechnical stop.

events.

The optimization process was guided by a figure of merit (FoM), which is a mathematical function that quantifies the effectiveness of the cuts and can be seen in equation 1. The FoM typically balances the need to maximize the signal efficiency (the fraction of signal events that pass the cuts) while minimizing the background efficiency (the fraction of background events that pass the cuts). An optimizer script was developed to automate the cut optimization process, ensuring that the cuts were applied in the most effective way possible. The resulting optimized cut values can be found in Table 1.

$$\text{FoM} = \frac{\epsilon_{\text{sig}}^2}{\epsilon_{\text{bkg}}} \tag{1}$$

where:

- ϵ_{sig} represents the efficiency of the signal.
- ϵ_{bkg} represents the efficiency of the background.

After applying the optimized cuts to the datasets, the resulting distributions were analyzed to identify the kaon mass peak. The mass peak is a critical feature in the data, corresponding to the mass of the K_S^0 particle (approximately 493.677 MeV). In an ideal scenario, this peak should be clearly visible in the

Variable	Value
BPVVDZ (z-axis distance)	> 390.3905
K_S^0 Lifetime	> 0.03573
K_S^0 BPVDIRA	> 0.9999997
K_S^0 BPVVDRHO	> 13.1359
Cosine of the Angle Between Electrons and Pions	< 0.99997
K_S^0 ALV23	< 1

Table 1: Optimized Cut Values for Event Selection

mass distribution, indicating that the selected events correspond to K_S^0 decays.

However, in this analysis, the kaon mass peak was not clearly observed in the real data. Specifically, the plot generated after applying all the cuts did not show a significant peak around the expected mass of K_S^0, as can be observed in Figure 7.

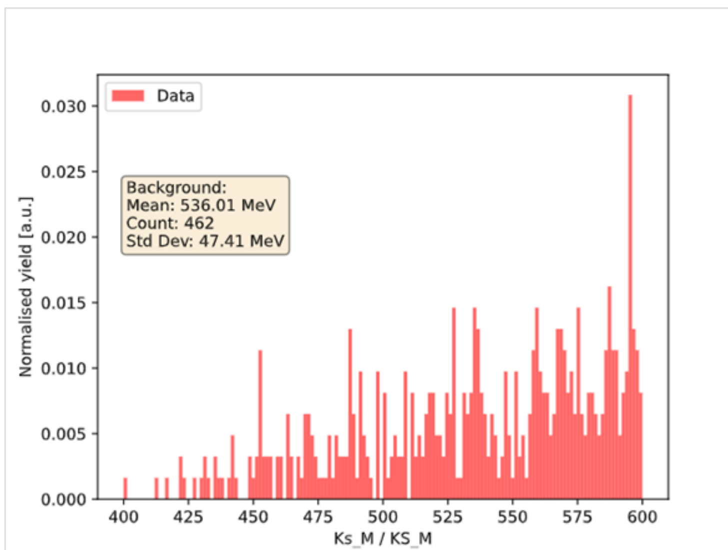


Figure 7: Resulting histogram of the K_S^0 mass variable with cuts applied, resulting from 400 ROOT files for the Magnet Down configuration and 400 ROOT files for the Magnet Up configuration.

5.1.3 Current Status: An alternative approach using BDT

The former outcome led to the consideration of an alternative approach to improve the selection process and enhance the detection of the signal.

In this alternative method, a Boosted Decision Tree (BDT), a machine learning tool, was implemented. BDTs are particularly effective in distinguishing between signal and background events by combining multiple variables into a single discriminator. The machine learning algorithm was fed with both simulated signal events and background events to learn the most effective way to differentiate them using the same variables used for the former method.

The decision tree algorithm iteratively splits the dataset based on these input variables, creating a model that predicts the probability of an event being a signal rather than background. Once the BDT was trained and validated, it was applied to the data to re-evaluate the selection criteria.

This machine learning approach allows for a potentially more effective selection process than manual cut optimization, as the BDT can capture complex correlations between variables that are not easily visible through traditional methods. The implementation of the BDT aimed to improve the sensitivity of the analysis and to find the K_S^0 mass peak in the data, however, the resulting plots for the K_S^0 mass using this approach still show no peak as it can be seen in Figure 8.

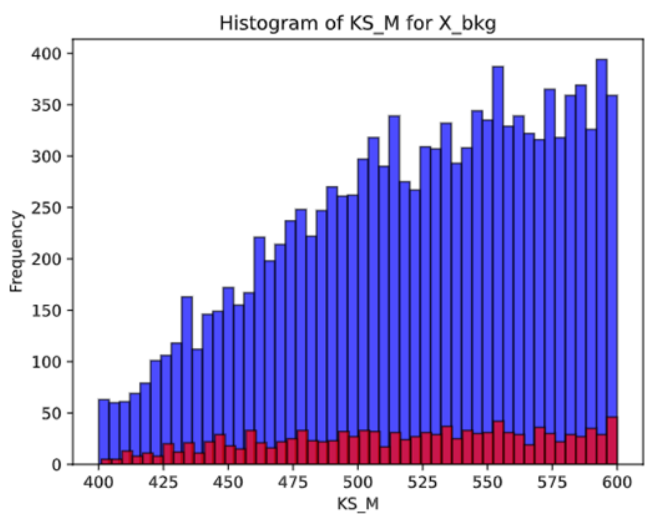


Figure 8: Resulting histogram of the K_S^0 mass variable using BDT, using 929 ROOT files from the Magnet Down configuration post technical stop.

5.2 Future Work

Although the implementation of the BDT was used as an alternative tool for the analysis, future work should focus on reproducing the selection criteria used in the 2016 study, as well as investigating any differences found in data-simulation analysis, for example, by studying the $K_S^0 \to \pi\pi$ decays to ensure that no selection is being applied that might suppress the signal. Also, on refining and optimizing the BDT tool to further increase its effectiveness in distinguishing between signal and background events, possibly expanding the set of input variables used by the BDT. Additionally, incorporating additional variables that capture more subtle aspects of the K_S^0 decay and its background could allow the BDT to make more accurate distinctions. Finally, exploring different machine learning models or even ensemble methods that combine multiple models could lead to improved results.

5.3 Conclusions

The expected K_S^0 mass peak was not initially observed in the real data. This outcome prompted the exploration of alternative methods, leading to the implementation of a Boosted Decision Tree (BDT) machine learning tool. There is significant potential to refine and enhance the BDT tool further, by optimizing the BDT's parameters, exploring additional variables, and even combining the BDT with other machine learning models to increase the likelihood of accurately identifying the K_S^0 signal. This could yield more definitive results and contribute valuable insights to the ongoing study of K_S^0 meson decays.

While the current analysis did not achieve all its goals, it laid a strong foundation for future investigations and provided experience in using advanced data analysis techniques. Continued development and refinement of these methods hold promise for future successes in the field of rare decay studies.

References

[1] Pere Gironella Gironell. Testing the Standard Model with Radiative Beauty and Rare Strange Decays at LHCb. *CERN-THESIS*, 2023. `https://cds.cern.ch/record/2861742/?ln=es`.

[2] A. A. Alves Junior et al. Prospects for measurements with strange hadrons at LHCb. 2019. `https://arxiv.org/abs/1808.03477`.

[3] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 2008. `http://dx.doi.org/10.1088/1748-0221/3/08/S08005`.

[4] Michal Mazurek, Gloria Corti, and Dominik Müller. New simulation software technologies at the LHCb experiment at CERN. 2021. `https://arxiv.org/abs/2112.04789`.

[5] The LHCb Collaboration. The LHCb Upgrade I. *Journal of Instrumentation*, 19(05), 2024. `http://dx.doi.org/10.1088/1748-0221/19/05/P05065`.