Postgraduate Studies Program in Physics
Physics Department, School of Sciences
University of Ioannina, Greece

Master's Thesis

Measurement of the Ratio $R_{K_S^0}(\mu) = BR(B^0 \to \mu\mu K_S^0)/BR(B^0 \to J/\psi(\mu\mu)K_S^0)$ with the CMS experiment at CERN

Spyridon Nessis
AM: 826
Supervisor: Costas Fountas

Ioannina November 6, 2024

Abstract

This thesis presents a measurement of specific branching functions for the decay of the B^0 meson. The blinded measurement of the relative branching ratio $R_{K_S^0}(\mu)$ of the $B^0 \to \mu\mu K_S^0$ Flavor Changing Neutral Current (FCNC) decay with respect to the tree-level resonant decay $B^0 \to J/\psi(\mu\mu)K_S^0$, as well as the tree-level decay $B^0 \to \psi(2S)(\mu\mu)K_S^0$ relative to the tree-level decay $B^0 \to J/\psi(\mu\mu)K_S^0$ is also measured. FCNC processes are forbidden at tree-level in the Standard Model with branching functions at the order of magnitude of 10^{-7} compared to resonant tree-level processes at the order of magnitude of 10^{-4}. FCNC decays are an important aspect in searching for New Physics beyond the Standard Model whose effects can affect the value of the $R_{K_S^0}(\mu)$ observable that is measured in this thesis. The data provided for the analysis of this thesis have been recorded by the CMS experiment at CERN LHC during Run 3 at 2022 with center of mass energy $\sqrt{s} = 13.4$TeV and nominal integrated luminosity of $\mathcal{L} = 39.7 fb^{-1}$. This analysis is part of a bigger ongoing project which aims to probe Lepton Flavor Universality by studying $b \to sX$ transitions at the CMS experiment at CERN LHC.

Contents

1 The Large Hadron Collider

1.1 Overview

The Large Hadron Collider (LHC)[1], built by the European Organization for Nuclear Research (CERN)[2], is a two-pipe ring 27km long superconducting collider designed to study proton-proton event collisions at extremely high luminosity at the center of mass energy of 14TeV, operating at a very low temperature of 1.9K assisted by superfluid helium coolant. Its main purpose is the exploration of electroweak symmetry breaking via the Higgs mechanism and the search of new physics past the Standard Model, e.g. Supersymmetry. CERN was founded in 1954 and is currently comprised of 24 member states and is receiving contributions from other nations around the world in its research programmes.

Construction of the LHC was authorized by the CERN Council in 1994 and between 1996 and 1998 the four detectors were approved for construction, ALICE, LHCb, ATLAS and CMS. The LHC's operation is not perpetual but includes long scheduled stops (shutdowns) for the machine's maintenance and upgrades. The first lengthy operation of data collection, Run 1, started in November 2009 and lasted until 2012, where afterwards the first shutdown took place. Run 2 lasted from 2015 to 2018 followed by another shutdown. The LHC operated again in 2022 and is expected to stop at 2024. These three runs constitute Phase 1 of the LHC, where a shutdown will commence between 2024 and 2027 which will mark the beginning for the first Run of Phase 2. Phase 2 of the LHC is planned to have a major upgrade to the beam's luminosity, and is called the *High Luminosity LHC* (HL-LHC) upgrade.

The two rings of the LHC are responsible for transporting the two beams throughout the machine, conveniently named Ring 1 and Ring 2[3]. The beams in the two rings, Ring 1 and Ring 2 are named Beam 1 and Beam 2 and circulate clockwise and counterclockwise respectively. The two rings are spanned by pipes at ultrahigh vacuum. Particles travelling inside the beam pipes are controlled by superconducting electromagnets, operating at -271.3°C[4] using liquid helium coolant. Different types of magnets are used to guide the beams of particles throughout the LHC. Specifically, the 1232 dipole magnets 15m long each are tasked with bending the beam to follow the circular track of the Rings and the 392 quadrupole magnets, each 5-7m long are tasked with focusing the beams. Throughout the two circular rings of the LHC, there are four Interaction Points (IPs) where the beam collisions take place. Around these IPs, the aforementioned particle detectors operate. ALICE and LHCb are referred to as low luminosity experiments, whereas ATLAS and CMS are referred to as high luminosity experiments.

1.2 Accelerator Complex

This thesis is within the context of the CMS experiment and therefore a small brief history and description of a proton's journey throughout the accelerator complex until the CMS detector is provided. In order for the particles to reach the high energies required for the experiment, particles are injected in a series of accelerators before they are inserted into the two rings of the LHC. Before their final injection into the LHC rings, the protons journey through four accelerators.

The first accelerator that prepares the two final beams in the LHC is the 86m long Linear Accelerator 4 (Linac4)[5] that boosts negative hydrogen (H^-) ions to 160MeV. Specifically, there are four stages of acceleration throughout Linac4. First, particles from

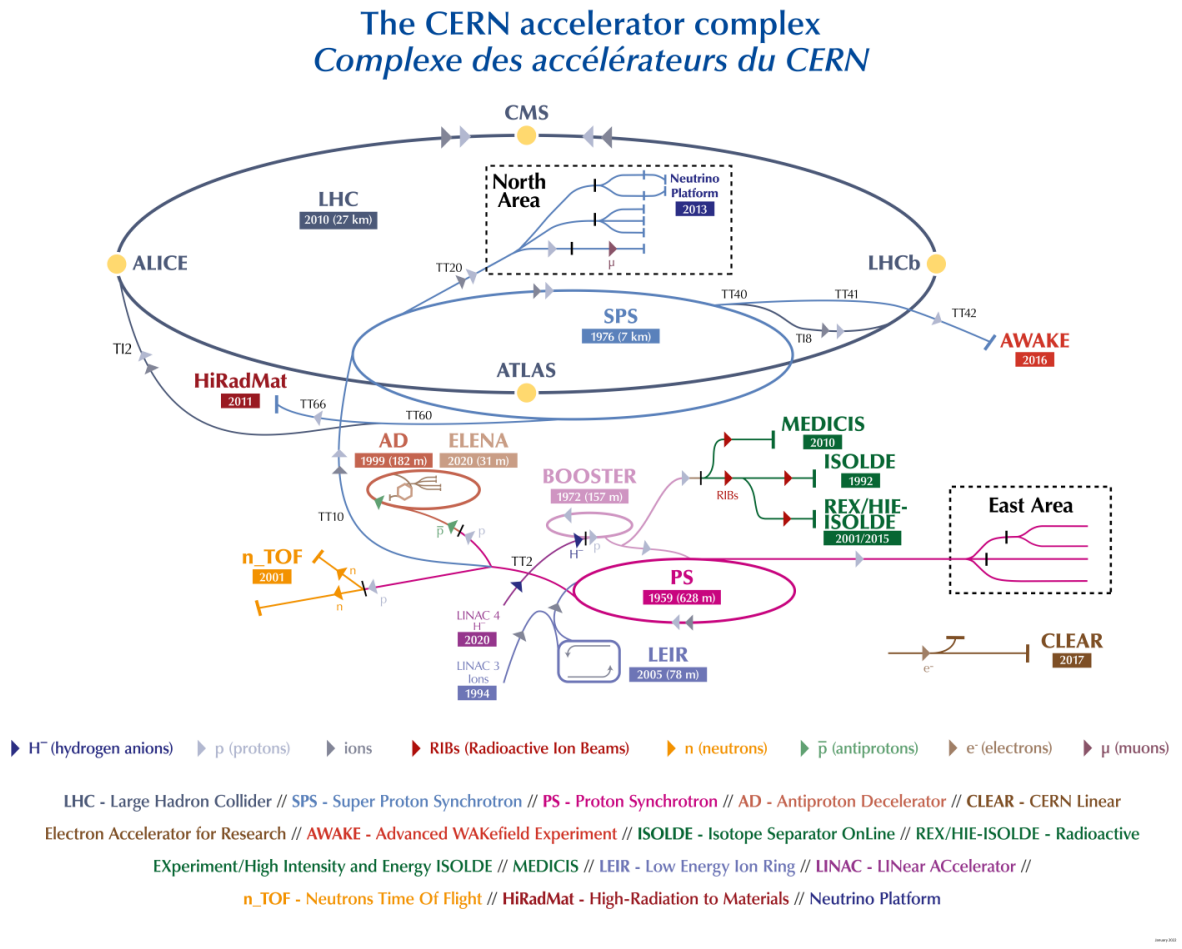


Figure 1.1: Graphic of the CERN Accelerator Complex. Each accelerator may serve multiples purposes and experiments.

the H^- source are accelerated to 3MeV by a RF quadrupole, then to 50MeV via drift tube linear accelerators, followed by coupled-cavity drift tube linear accelerators to 100MeV and finally Pi-mode structures accelerate the ions to 160MeV. Before the beam of H^- enters the next accelerator, the Proton Synchrotron Booster, the electrons are stripped away from the proton nucleus in order to ensure better control of the beam. The Linac4 accelerator replaced its previous iteration, the Linac2 in 2020, which had a source of hydrogen gas instead of hydrogen ions and a maximum acceleration of 50MeV for the accelerated protons. Linac4 is a key element in the increase of beam luminosity during the next LHC upgrades.

The second accelerator tasked with the preparation of the beams is the Proton Synchrotron Booster (PSB)[6]. Constituted from four superimposed synchrotron rings, it accelerates Protons to 2GeV and prepares the particles for injection into the next accelerator, the Proton Synchrotron (PS)[7]. The PS has a circumference of 628m and accelerates protons up to 26GeV. A number of 277 conventional room-temperature electromagnets are used by the machine, including 100 dipoles to bend the beams around its rings. The final accelerator in the accelerator complex prior the main LHC rings is the Super Proton Synchrotron (SPS). It is a machine[8] that resembles its predecessor, the Proton Synchrotron. Constituted by 1317 conventional room-temperature electromagnets and 744 dipoles that bend the ring, it can accelerate protons to energies up to 450GeV. Then the beams are injected into the LHC two ring pipes and are accelerated to 7TeV.

The radiofrequency (RF) cavities of the LHC are discrete chambers that house oscil-

lating electromagnetic fields that are used to accelerate charged particles[9]. The LHC has 16 RF cavities, eight for each ring. The electric field inside a cavity oscillates in tandem with the incoming charged beam in order to accelerate the beam. Once the beam reaches the required energy, protons that are ideally focused will no longer accelerate or decelerate. Protons with energies slightly off the desired energy value eventually stabilize to the desired value at earlier or later times. At this stage, the protons are grouped into bunches, forming a discrete but well-focused beam. The beam is characterized by 2808 bunches of protons, with each bunch composed of about $1.64 \cdot 10^{11}$ protons, separated by a time interval of 25ns per bunch. This time interval is referred to as the bunch crossing time and describes the time interval between two consecutive bunches in a proton beam. The beam is characterized by its beam parameters, which will change after the Phase Two upgrade. The maximum beam energy is reached in around 20 minutes of operation where the bunches have passed through the RF cavities more than 10 million times.

1.3 LHC Beam Parameters

The LHC was designed to discover the Higgs particle which at the time was expected to be lighter than 1 - 2TeV. The LHC uses two colliding beams each with a 6.7TeV proton energy. This results to a proton-proton cross section of about 100mb. The beam particles are chosen to be protons due to their composite structure and their reduced energy loss due to radiation. The radiation loss of a relativistic particle is given by Lienard's formula[10] in SI units:

$$P = \frac{\mu_0 c q^2 \gamma^6}{6\pi} \left(\left| \dot{\vec{\beta}} \right|^2 - \left| \vec{\beta} \times \dot{\vec{\beta}} \right|^2 \right) \tag{1.1}$$

where γ is the Lorentz factor and $\beta = v/c$. The power radiated perpendicular to the direction of acceleration is equal to:

$$P = \frac{\mu_0 c q^2}{6\pi} \frac{\gamma^2}{m^2 c^2} \left(\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} \right)^2$$

For a particle performing circular motion of radius R, the magnitude of the force that is acted upon it is equal to:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\gamma \beta^2 m c^2}{R}$$

which yields the synchrotron radiation loss:

$$P = \frac{\mu_0 c^5 q^2}{6\pi} \frac{E^4}{R^2} \frac{1}{m^4}$$

where E is the energy of the particle. Lienard's formula for relativistic particles in circular motion establishes that the radiative power is proportional to $1/m^4$. This means that more power is radiated for lighter particles compared to heavier particles with the same energy, charge and trajectory radius. Massive particles cannot exceed the speed of light c, therefore a light particle must have increased radiative power loss, dissipating more energy compared to a heavier particle. The proton beam is mainly characterized by its instantaneous luminosity $\mathcal{L}$ (in SI units $m^{-2}s^{-1}$), which is in turn defined by the beam parameters as shown in table 1.2. Luminosity describes the strength of the beam and determines the number of events per unit time. The chance of recording rare processes is increased as the beam luminosity strengthens. Phase 2 of the LHC regards an upgrade to the beam's luminosity for this very reason. The instantaneous luminosity is given by:

$$\mathcal{L} = \frac{\gamma f n_b N_p^2}{4\pi \epsilon_n \beta^*} F$$

Parameter Name	Symbol	Value
Center of Mass Energy	$\sqrt{s}$	$14TeV$
Instantaneous Luminosity	$\mathcal{L}$	$10^{34}cm^{-2}s^{-1}$
Lorenz factor	γ	7461
Revolution frequency	f	$11245Hz$
Bunches per beam	n_b	2808
Protons per bunch	N_p	$1.15 \cdot 10^{11}$
Normalized transverse emittance	ϵ_n	$3.75\mu m$
Betatron value at IP	β^*	$0.55m$
Reduction Factor	F	0.81
Bunch crossing time	τ_b	$25ns$

Figure 1.2: Some of the proton-proton beam parameters of the LHC beam[11][12].

The number of events per second dN/dt generated in the LHC collisions in the IPs is related to the instantaneous luminosity $\mathcal{L}$ and is given by:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \mathcal{L}\sigma$$

where σ is the cross section of the event, which is nominally equal to as 100mb. The high luminosity creates a problem. Among the interesting hard collisions and reactions, there's always the presence of soft, low energy collisions. This presence of the soft collisions is called a pileup (PU) and is responsible for data contamination and further introduction of measurement uncertainty. Furthermore, pileup makes the identification and reconstruction of various physics objects (like trajectories and physical quantities) tedious. In the LHC, the beams collide every $\Delta t = 25ns$, which for the aforementioned nominal instantaneous luminosity and cross section, the number of collisions ΔN is nominally equal to:

$$\frac{\Delta N}{\Delta t} = \mathcal{L}\sigma \Rightarrow \Delta N = 25 \text{ events every 25ns}$$

The total amount of data collected is associated with the integrated luminosity:

$$L = \int \mathcal{L}\, dt$$

with associated units $\text{fb}^{-1} = 10^{-39}\text{cm}^{-2}$. The luminosity of the LHC is not constant over time. During the operation of the machine in a Run, the luminosity slowly increases to its maximum value. This is done in order to carefully monitor the experiment and identify any signs of damage or component misconfigurement in the machine. The highest value of luminosity achieved in CMS during Run 3 in 2022 was 39.7fb^{-1}. During Run 3 in 2024, the LHC machine has provided CMS with a record-breaking luminosity of 88.9fb^{-1}. The general discussion of the LHC is now over and discussion on one of the four main detectors, the CMS, will now commence.

1.4 The Compact Muon Solenoid Detector

The Compact Muon Solenoid (CMS) detector[13][14] operated by the CMS Collaboration has been commissioned at IP5. It is a general-purpose detector and is used for studying the Standard Model as well as for providing additional searches for physics beyond the Standard Model, such as signatures that predict dark matter particle candidates and

extra dimensions. The detector is built around the usefulness of the muon, an elementary lepton. Due to their higher mass than the electrons, are highly penetrating particles and therefore can be detected easily. Muons originate from decays of heavy particles as well as other exotic particles and therefore can be used as signatures of new physics. Due to the construction of the detector, muon identification is excellent. In this section an overview of the detector's geometry and its main components will be given.

Figures 1.3 and 1.4 show a visual description of the detector. The CMS detector is of hermetic cylindrical geometry and is 21.6m long, with a maximum diameter of 14.6m. Surrounding the IP, in the heart of the detector, lies the Pixel Detector, a primary component of the Silicon Tracker. These detectors provide the appropriate precision for particle detection and aid the measurement of the position of secondary vertices and the impact parameter b of charged particle tracks, two important quantities for analysis. Extending outwards, immediately follows the compact Electromagnetic Calorimeter (ECAL) along with the Preshower Detector (shown in figure 1.3), which surrounds the silicon tracker, tasked with the detection of electrons and photons, spanning a thickness greater than 25 radiations lengths. The Hadronic Calorimeter (HCAL) engulfs the ECAL and is solely responsible for the detection and halt of hadrons. These two calorimeters are encased within the solenoid magnet, which generates a high magnetic field strength. Lastly, the Muon Detectors are part of the Muon System which is tasked with the detection of muons, the experiment's primary point of interest. The Very Forward Calorimeter is an additional detector that succeeds the Muon chambers and detects particles for very large η.

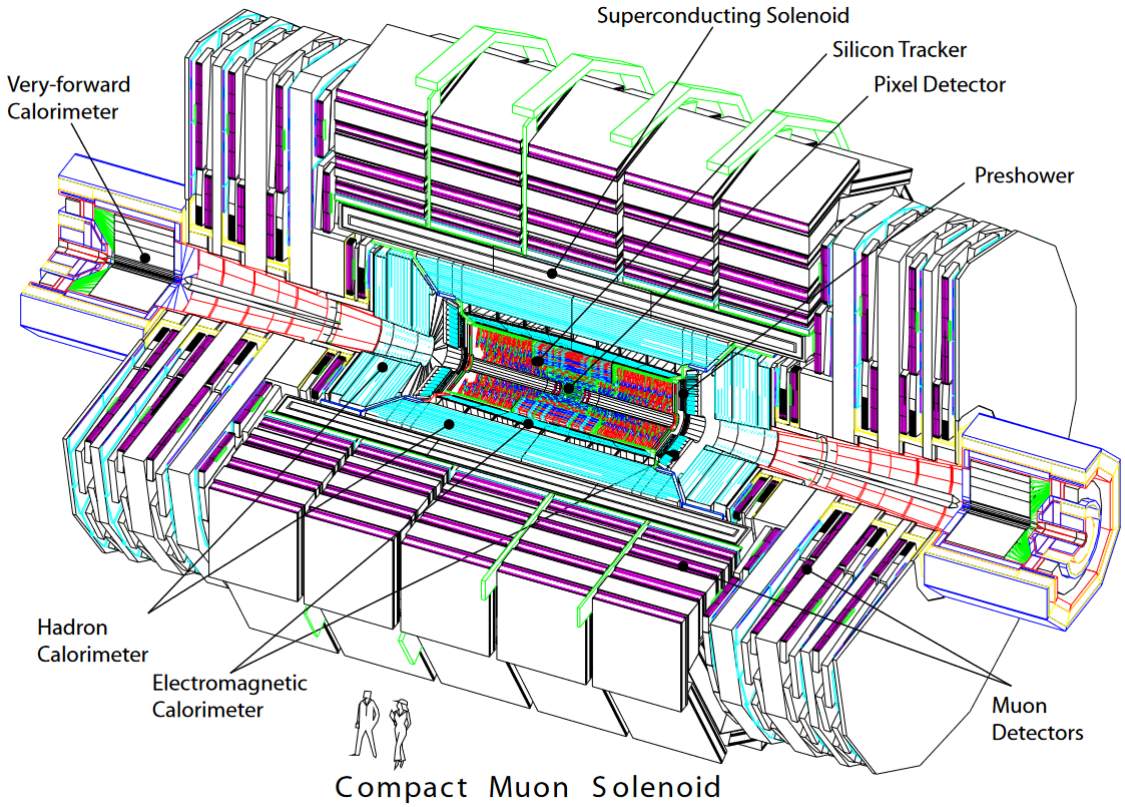


Figure 1.3: Mechanical Geometry of the CMS detector including key parts in perspective view.

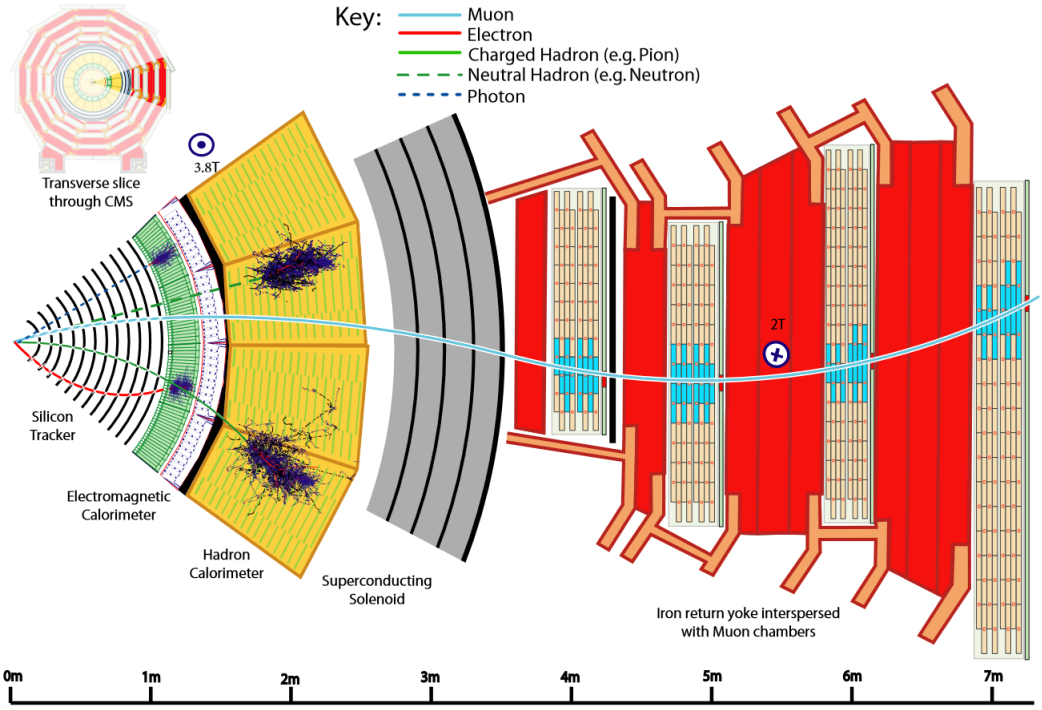


Figure 1.4: Cross-sectional radial view of the CMS detector showcasing each major subsystem.

1.4.1 Superconducting Solenoid Magnet

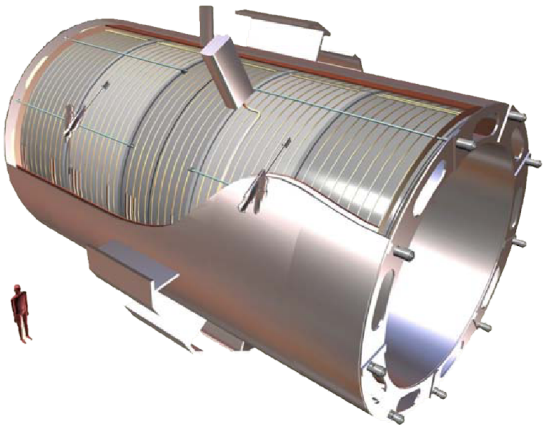
CMS uses a 4T magnetic field which results to the momentum resolution of $\Delta p/p$ of 1% at momenta of $200 GeV/c$. The superconducting solenoid magnet is comprised of Niobium-Titanium (NbTi) superconducting strands, a high purity Aluminum (Al) stabilizer and an aluminum alloy reinforcement[15]. The superconducting stands (SC strands) have a critical current of 55kA at 4.2K and 5T, and a rated current of 19.14kA. The high purity Al stabilizer surrounds a Rutherford-type cable comprised of 32 SC strands. The choice of Al to serve as the stabilizer is due to the low compacting ratio to ensure small critical current degradation. The Al alloy AA 6082 mechanically reinforces the structure into a cold mass[14] (figure 1.5a) weighing 220t and radiation length $3.9X_0$ which provides a uniform magnetic field strength of 3.8T and stored energy per unit cold mass 11.6kJ/kg within the solenoid.

The CMS solenoid is powered by a bipolar thyristor power converter at 520kW with the aid of passive L-C filters, covering a range of voltages between -23V to +26V and a nominal electrical current of 19.1kA. The high value of the stored energy per unit of cold mass results in high hoop strain ϵ of 0.15%. This magnetic field strength is achieved by a four layer winding of Rutherford-type cable which corresponds to a total magnetomotive force of 41.7MA-turns and constitutes the cold mass of the solenoid magnet. The solenoid has a diameter of 6.3m and a length of 12.5m and weighs 220 tons, parallel to the beam pipe.

A 10000 ton iron yoke comprised of two endcaps and a barrel weighing 2000t and 6000t each is installed to support the detector outside the Hadronic Calorimeter and to return the field back in the opposite direction according to Maxwell's equation $\vec{\nabla} \cdot \vec{B} = 0$. The yoke is the frame of the magnet and is responsible for the support of the entire apparatus of the CMS and includes the coil and the cryostat. Each endcap is composed of three endcap disks and the barrel is composed of five barrel wheels (Fig. 1.5b). This choice of layered structure allows the insertion of muon chambers which are supported by the return yoke.

The solenoid also houses other mechanical geometries such as a vacuum system for the insulation of the cryostat and transportation of the liquid helium cooling, a grounding circuit, a cryogenic plant which prepares the coolant and a quench detection system, responsible to combat the quench back effect. The quench back effect describes induced surface eddy currents that form at the trigger of a current fast discharge which can heat up the coil above the superconducting critical temperature, removing the superconducting properties of the magnet for an extended period of time. Electrical protection circuits have been installed in order to prevent local overheating near the superconducting wirings.

(a) Cross section of the cold mass with the details of the 4-layer winding with reinforced conductor.



(b) 3D Model of the five wheels of the cold mass barrel inside the iron return yoke.

Figure 1.5: Two components of the Superconducting Solenoid Magnet.

1.4.2 Inner tracking system

The inner tracking system[14][16] covers the interaction point with a diameter of 2.5m and longitudinal length of 5.8m and is primarily responsible for the precise measurement of charged particle trajectories and the reconstruction of secondary vertices, as well as the primary proton-proton interaction point of each event. The high magnetic field strength of the solenoid and the high luminosity of the LHC impose stringent constraints on detector electronics. Furthermore, due to the high flux of particles per bunch crossing, the tracking system must be able of resolving and identifying events and trajectories with high resolution. Additionally, the design of the tracker must use as little material as possible, so that particles cannot further damage the tracking system. All these requirements led to the choice of silicon detector technology for the tracker. The tracker is composed of a pixel detector with four barrel layers and a silicon strip tracker with ten barrel layers. Two endcaps complete each of the two subsystems, providing a pseudorapidity range of $|\eta| < 2.5$. The active detector area is 200m^2.

The Pixel Detector The pixel detector provides high accuracy for particle path tracking and path reconstruction and is the part of the tracking system closest to the Interaction Point of the detector, covering 1.06m^2 for a total of 66 million pixels. A pixel's composition includes a common silicon detector imbued with readout electronics, resulting in a very thin layout. Charged particles passing through the silicon pixels are identified by an electric signal which is created by electron-hole pairs created in the pixel's material resulting into currents which are processed by the electronics to measure the position that each particle crosses the silicon detector. The presence of the silicon pixels throughout

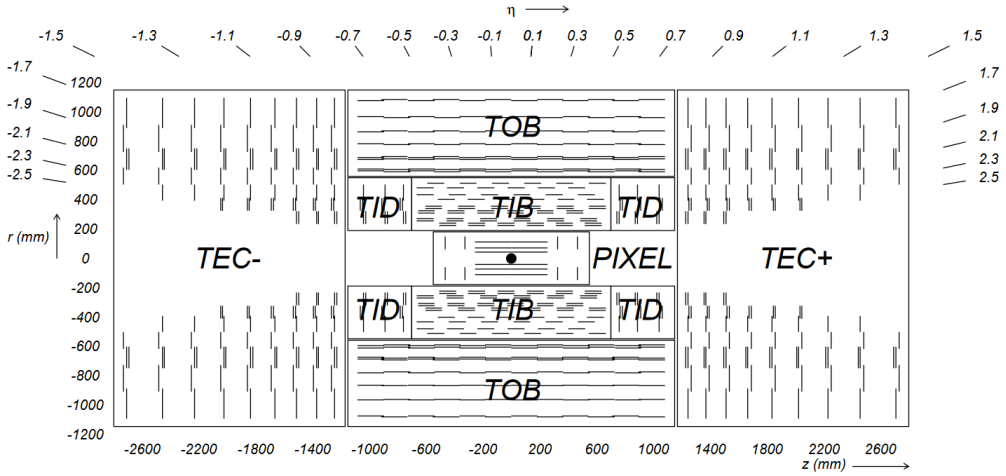


Figure 1.6: Schematic cross section through the inner tracking system. Each line represents a detector module.

the tracker provide precise information for particle trajectory reconstruction. In figure 1.6, at radii 4.4, 7.3 and 10.2cm, three cylindrical layers of hybrid pixel modules containing 48 million pixels spanning a total area of 0.78m^2 surround the IP, complemented by two disks of pixel modules on each side, at z=±34.5cm and z=±46.5cm, containing 18 million pixels spanning a total area of 0.28m^2.

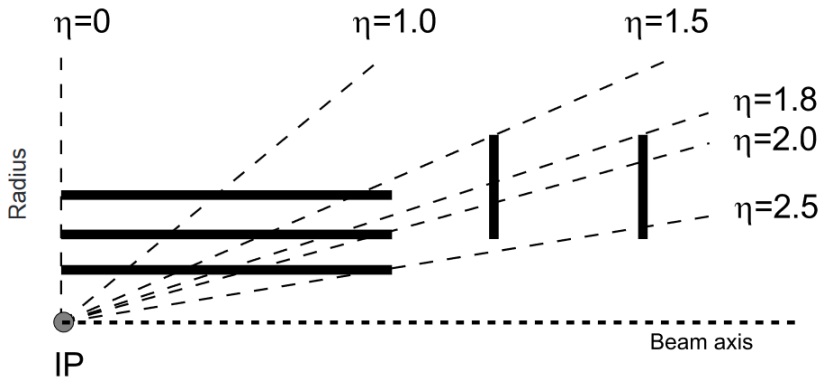


Figure 1.7: Coverage of the Pixel Detector in η. The three cylindrical layers and the two disks are shown.

The Silicon Strip Tracker Following the pixel detector, ten layers of n-doped silicon strip trackers are installed in the experimental apparatus in the barrel region. At this point, the outermost radius reaches about 20-116cm. The strip trackers are placed in four different regions as shown in the schematic of Fig. 1.6, namely, the tracker inner barrel (TIB), the tracker disks (TID), the tracker outer barrel (TOD) and the tracker endcaps (TEC). These four distinct areas packed with the strip detectors enclose the pixel system and the interaction point. The silicon strip tracker has a total of 9.3 million strips and 198 m^2 of silicon area.

The TIB and TID regions have a maximum radius of 55cm and are composed by 4 barrel layers and three supplemental disks respectively. Due to the 4 barrel layers, they are able to provide up to 4 measurements in radius r and azimuthial angle ϕ (r-ϕ) using silicon micro-strip sensors parallel to the beam axis. The strip pitch in he TIB region is different between the first and last two layers, leading to a single point resolution of 35μm and 23μm respectively. The TOB region extends from 55cm to 116cm and up to ±118cm

on the beam axis and consists of 6 barrel layers providing up to 6 r-ϕ measurements with single point resolution of 35μm and 53μm for the first two innermost layers and the four outermost layers respectively. The two TEC regions enclose the aforementioned geometry with two endcaps with $z \in [124, 282]$cm and $r \in [22.5, 113.5]$cm. Each TEC is composed of 9 disks, providing up to 9 measurements of ϕ.

The first two layers and rings of TIB, TID and TOB and three disks in the TEC regions carry a second micro-strip detector module to measure the second coordinate, which is z for the barrel regions and r in the disks. In TIB and TOB, the single point resolution varies between 230-530μm. This layout of the tracker ensures about 9 hits within $|\eta| < 2.4$ with about 4 of them being two dimensional measurements due to the second coordinate measurement. The tracker sensors are cooled to protect them from radiation damage and other harmful effects.

1.4.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL)[17][11] is a homogenous, hermetic and fine-grained calorimeter, composed of lead tungstate $PbWO_4$ scintillation crystals in both of its barrel part and the two endcaps. The ECAL encloses the silicon tracker extending outwards. The high mass density (8.28 g/cm^3), radiation hardness, short radiation length[1] X_0 (0.89cm) and the Moliere radius[2] of the scintillation crystals R_M (2.19cm), constitute the crystals as a perfect fit for the CMS detector, making the ECAL so compact that it can fit inside the superconducting solenoid. Furthermore, the scintillation decay time is in the same order of magnitude as the bunch crossing time of 25ns (40MHz), with 80% of the photons emitted within 25ns. The energy resolution of the barrel region's supermodules for energies below about $500GeV$, is given by the relation:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12.4\%}{E}\right)^2 + (0.26\%)^2$$

The relation is due to fitted incident electron data from 3×3 arrays centered on a EB crystal with beam hodoscope cuts of 4×4 mm^2[18]. The first term is the stochastic term, where photostatistics, event-to-event fluctuations and measurements in the preshower detector contribute. As for the second term represents the contributions of noise from the electronics and digitization, as well as particle pileups. The last term represents the non-uniformity of different detector elements.

The Barrel Region The composition in the barrel region (EB) covers the pseudorapidity range $|\eta| < 1.479$ and is characterized by the 36 crystal supermodules, with each supermodule containing about 1700 crystals bringing the crystal total to 61200. This categorization of modules allows stability, weight control and enables flexible handling in cases of various moderation services, like cooling and maintenance. Each crystal has a front face cross section of about 22×22mm^2, a rear face cross section of 26×26mm^2 and a radiation length of 25.8X$_0$ (which corresponds to a length of 230mm). The granularity in ϕ, η is 360-fold and 2×85-fold respectively. These crystals are purposefully misaligned by 3° with respect to the line segment between the cross sectional face center and IP.

The Endcap Region The endcaps (EE) cover the region of pseudorapidity between the EB region to $|\eta| < 3.0$ and are placed 315cm from the IP along the beam axis. Each

[1] defined as the mean length at which the energy of an electron is reduced to a factor 1/e of its energy
[2] defined as the radius of a cylinder containing on average 90% of the shower's energy deposition

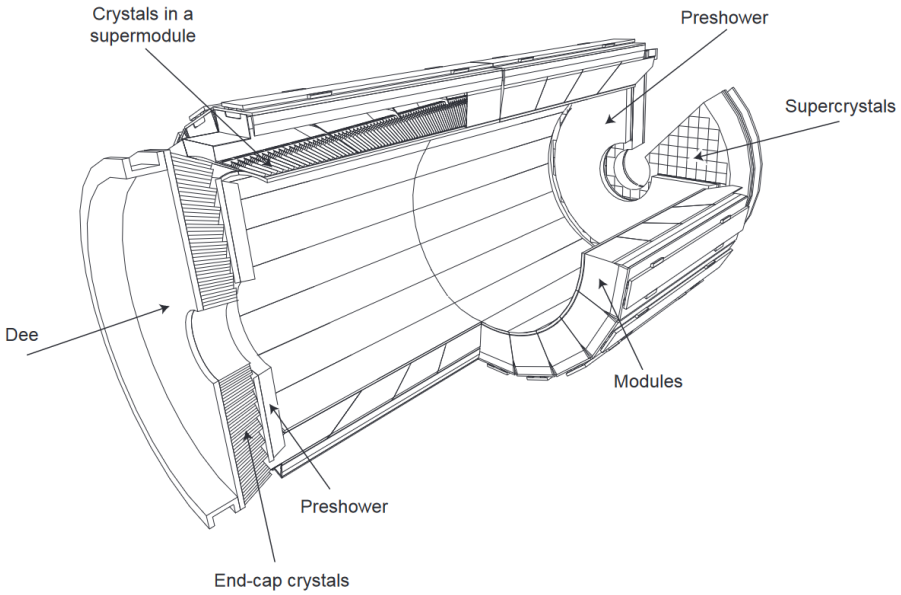


Figure 1.8: Detailed schematic of the ECAL. The three primary regions (not explicitly denoted in this figure) are the barrel region (EB), endcap region (EE) and the preshower detector (ES). Components across these regions are shown in the schematic.

of the two endcaps is divided into two halves, the Dees, in which the supercrystals, an arrangement of 5×5 crystals and have a front face cross section of 28.6×28.6mm^2, radiation length 24.7X_0 (length 220mm) and are placed in a rectangular grid (x-y) as opposed to a radial grid (η-ϕ) as constructed in the EB. Each Dee is comprised of 3,662 crystals, summing to a total of 14,648 crystals across all four Dees. The two regions combined bring the total amount of crystals to 75,848. The ECAL's outer radius reaches $177cm$ from the central interaction point at this stage.

The Preshower Detector Lastly, the preshower detector (ES), preceding the endcaps inward, aims to identify light neutral pions in the endcaps of the ECAL in the pseudorapidity range $1.653 < |\eta| < 2.6$ as well as aiding the identification of electrons against minimum ionizing particles. Since the most common decay mode of a π^0 is to $\gamma\gamma$ (98.823 ± 0.034 %)[19], this detector aims to distinguish whether the incoming photons came from the neutral pions or some other event. With a thickness of $20cm$, this sampling calorimeter is characterized by a layer of lead radiator and a layer of silicon sensors. The first layer is responsible for causing electromagnetic showers from incident photons or electrons throughout the ES. The latter layer is composed of silicon strip sensors that measure the energy deposited along with transverse shower profiles. Each silicon sensor has an active area of 61×61mm^2, thickness of 320μm and is divided into 32 strips of 1.9mm pitch each. A normal incident minimum ionizing particle will deposit ~3.6fC in the sensor.

The $PbWO_4$ crystals The $PbWO_4$ crystals emit blue-green light (420nm) when scintillated and its intensity varies with temperature at a rate of -1.9%/°C at the operational temperature 18°C, the $PbWO_4$ crystals are maintained at a constant temperature of 18° to high precision (within ±0.05°C), with minimal fluctuation from that value. Due to the conditions of the experiment, the crystals are subject to radiation damage. Ionizing radiation produces absorption bands in the lattice, causing wavelength-dependent losses of light transmission due to those absorption bands. This damage is investigated via monitoring the optical transparency through injected laser light.

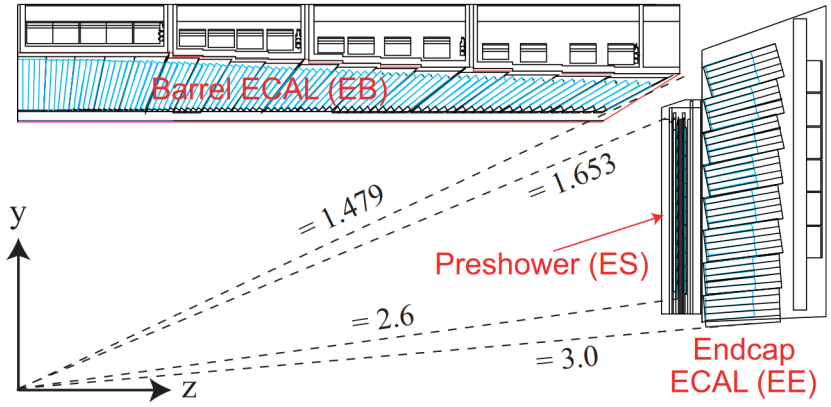


Figure 1.9: Transverse geometry of the ECAL, including the pseudorapidity ranges for the three regions (EB, EE, ES). A quarter of the transverse cross sectional view is shown.

The Photodetectors The presence of photodetectors with intrinsic gain which operate within a magnetic field are necessary since the scintillators previously described output a low light yield signal of about 30γ/MeV. Therefore, appropriate photodetectors must be used depending on the geometry of the experiment in order to amplify, and eventually digitize the signal. The EB and EE regions are equipped with photodetectors but of different types. A pair of avalanche photodiodes (APDs, figure Fig. 1.10, left) is attached on each crystal of a supermodule in the EB, whereas in the crystals in the EE, vacuum phototriodes (VPTs, figure Fig. 1.10, right) are used instead and a single one is attached at the back of each crystal as opposed to a pair of APDs in the EB. The APDs have an active area of 5×5mm^2, quantum efficiency of 75 ± 2% at 430nm and a mean gain of 50 at the operating temperature of 18°. The choice of the single gain stage VPTs is due to the high radiation close to the endcaps, where APDs can not operate and were specifically designed for the CMS experiment and its harsh operating conditions. Each VPT is 25mm in diameter, has a mean quantum efficiency 22% at 430nm and a mean gain of 10.2 at 0T. The magnitude of the magnetic field of the CMS at 4T negatively affects the response of the VPT along the EE region. For a VPT with its axis 15° with respect to the $\vec{B}$ direction, the mean response is 94.5% of the zero magnetic field.

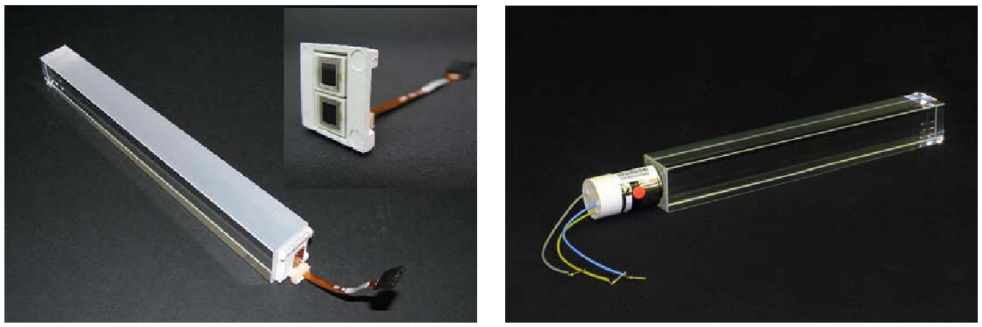


Figure 1.10: Left, Hamamatsu type S8148 reverse structure APD for the EB region and type PMT188 from JNC NRI Electron VPT for the EE region. The socket for the APD is also shown on the left side of the figure.

The ECAL is subjected to a cooling system that extracts the dissipated heat by the electronics. In the barrel of the calorimeter, each supermodule is directly supplied with water at the aforementioned temperature through a thermal screen between the crystals and the outer part of the silicon tracker. What's more, a thick layer of insulating foam is placed between the crystals and the electronics which carry the signals to prevent

overheating from high power consumption.

1.4.4 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is a sampling calorimeter and can identify and measure the energy and direction of particle jets and provide data on missing transverse energy in events, in conjunction with the ECAL. The HCAL, akin to the ECAL, is also separated into specialized geometrical regions[11], namely the Hadron Barrel (HB), Hadron Endcaps (HE), Hadron Outer (HO) and Hadron Forward (HF). All but one of the HCAL regions (the HO) are placed inside the Superconducting solenoid. The mechanical design of the HCAL is determined by the need to absorb hadrons between within the limited region inside the Superconducting Solenoid. For this reason, the HCAL is designed to have an absorber thickness of 5.82 interaction lengths[3] at $\theta = 90°$ and increased to 10.6 interaction lengths at $|\eta| = 1.3$[21]. The increase in the interaction length leads to less interactions between particles and the material and therefore the reduction of the tails of the energy resolution function[11]. The active material is Kuraray SCSN81 plastic scintillator, chosen for its long-term stability and moderate radiation hardness. There are two types of absorber material used in the HCAL. The first absorber material is a flat brass alloy (known as C26000/cartridge brass[22], nominally 70%Cu and 30&Zn with mass density 8.53g/cm^3) due to its short interaction length[4] and its non-magnetic property. The second material is stainless steel and is primarily used for structural support.

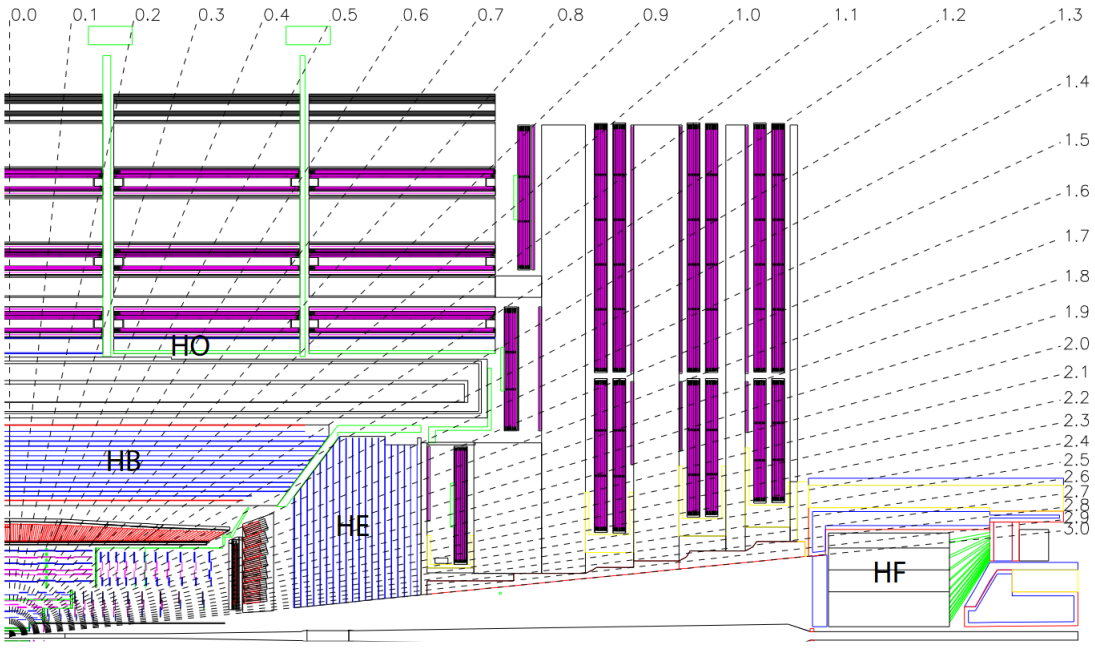


Figure 1.11: Longitudinal cross-sectional view of the HCAL and its four components. The range in η is shown. The purple regions are the Muon Chambers.

The Hadron Barrel The Hadron Barrel (HB) region consists of two equal face-cut halves covering the pseudorapidity region $|\eta|$ <1.4 and extends from 177.7cm to 287.65cm outwards. Each half barrel is composed of 18 identical 20° brass alloy absorber plate wedges in ϕ. The combination of the two barrel halves results in 18×20°=360◦ in ϕ. Seventeen layers (labeled as Layers 0-16), encompass the plastic scintillator active material

[3] defined as the mean free path between two inelastic interactions[20]

[4] the absorber must have a short interaction length in order to stop particles.

and are interspersed between the absorber material plates with variable thickness per barrel half. The Scintillator Layer 0 follows directly after the ECAL and has a scintillator thickness of 9mm and is separated by a 61mm stainless steel absorber, Layers 1-8 are 3.7mm thick, separated by 50.5mm thick brass absorber, Layers 9-14 are 3.7mm thick, separated by 56.5mm thick brass absorber, Layer 15 is 3.7mm thick, separated by a 75mm stainless steel absorber, followed by the last layer, Layer 16 which is 9mm thick. The innermost and outermost absorber layers are made of stainless steel and not brass for structural strength. Furthermore, the increased thickness of the scintillator Layers 0 and 16 is to actively sample low energy showering particles from the support material between the ECAL and HCAL[11] and correct for late developing showers leaking out from the HB[22]. Lastly, the active material of Layer 0 is Bicron BC408 instead of the default choice Kuraray SCSN81[22]. Scintillator tiles are made from 10mm active material plates. Wavelength Shifting Fibres (WLS) are intertwined inside the tile and lead the light captured from the scintillator active material to hybrid photodiodes (HPD) with a gain of ≈2000 to prepare the signal for the readout electronics. All tiles in a ϕ sector are grouped together and form a mechanical unit called a tray. Trays can be replaced without disassembly of the absorber of an HCAL region for maintenance. Trays are inserted in the absorber geometries of each region, except for the HF. The energy resolution of the ECAL+HCAL is expressed by the following relation[23]:

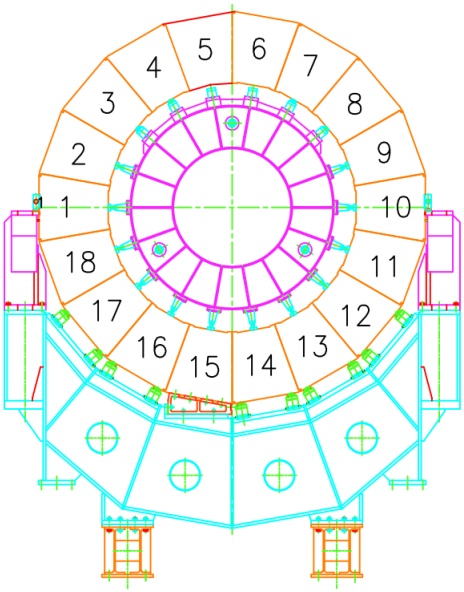


Figure 1.12: Face cross-sectional view of the HCAL, showcasing the 18 wedges.

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2$$

where the terms have been described in the section regarding the ECAL's energy resolution. The electronic noise is negligible and is not taken into account. The relation for energy resolution comes from the combined efforts of the ECAL and HCAL and not using the HCAL individually, as the hadronization process can start before the radial geometrical coverage of the HCAL. The above relation describes the energy resolution for pions after non-linearity corrections (the stochastic term is reduced from 110.7% to 84.7%)[23].

The Hadron Endcaps The Hadron Endcaps (HE), composed of the same brass absorber as of the two HB regions[22], cover the pseudorapidity range 1.3< $|\eta|$ <3.0 and must withstand tough conditions to operate, such as high particle counting rates and high radiation tolerance and are attached to the muon endcap yoke. At the front face of the HE, two ECAL components, the EE and ES are attached. The HE is composed of 79mm thick brass plate disks with 9mm gaps in between them to house plastic scintillators trays of the same material as in the HB region. The WLS fibres inside the scintillator tiles shift the wavelength of the captured light and optical cables transfer signals from the trays to multipixel hybrid photodiodes (HPDs) in order to digitize the signal for the readout electronics. The choice of HPDs as photodetectors is due to their low sensitivity to magnetic fields and large dynamical range. The granularity of the HB calorimeters is variable, $\Delta\eta \times \Delta\phi$ = 0.087×0.087 for $|\eta|$ <1.6 and $\Delta\eta \times \Delta\phi \approx$0.17×0.17 for 1.6< $|\eta|$ <3.0.

The Hadron Outer Detector The Hadron Outer (HO) detector comprises of scintillators and its purpose is to detect charged particles that leak out of the HCAL. For this reason, it is placed outside of the Superconducting Solenoid and inside the Barrel Muon System, covering the pseudorapidity range $|\eta| < 1.26$. The HO is divided into 5 sections along the beam axis, called rings with labels -2, -1, 0, 1, 2 at positions in z -5.342, -2.686m, 0m, +2.686m, +5.342m respectively[22]. Each ring has a single layer of scintillators at a radial distance of 409.7cm and a beam axis length of 2.5m, except for Ring 0, which has an additional scintillator layer at a radial distance 385.0cm. These two layers are separated by an iron absorber with an approximate thickness of 18cm. The 10mm thick scintillators sample the energy from penetrating hadron showers leaking through the rear of the calorimeters and increase the effective thickness of the hadron calorimetry to over 10 interaction lengths.

The Hadron Forward Detector The Hadron Forward (HF) calorimeters cover the pseudorapidity range of 3.0< $|\eta|$ <5. The faces of the HF detectors are located at ±11.2m from the IP and the depth of the steel absorbers is 1.65m (about 10 interaction lengths). Fibres are inserted inside the absorber. Two sets of fibres are placed within this detector one set covering the full depth, and the other set covering half the depth, starting 22cm from the face of the HF. These two sets, which are readout separately, is to distinguish showers between electrons and photons (first set, labeled L for long) and hadrons (second set, labeled S for short). The former particle showers deposit a large part of their energy at the first 22cm of the detector, where the L set is responsible for their measurement, whereas the latter produce nearly equal signals throughout the absorber material depth, and therefore the S set is used. Both HF regions are also divided to 36 total 20° wedges. The task of the two HF detectors is to measure energetic forward jet profile and increase the hermeticity of the missing transverse energy measurement. The active material of this calorimeter is composed of steel/quartz fibre. It is very radiation-hard and therefore can survive in the extreme conditions at high 4.5< $|eta|$ <5. A 10cm thick lead plate, located in front of HF around the detector reduces exposure to radiation from the absorber. The signal is collected from Cerenkov light emitted in the fibres and then processed by photomultipliers. The diameter of the fibres is 0.6mm and they are placed in a square grid 5mm apart, parallel to the beam line.

1.4.5 Muon System

The engineering of the CMS experiment has been specifically designed so that the Muon System will measure muons with extreme precision through their trajectory curvature. While the ECAL can measure photons and electrons, which are very lightweight particles, muons are the most penetrating particles and are measured in this detector which succeeds the HCAL after the solenoid. Furthermore, over the high background rate at the nominal operational luminosity of the LHC, muon detection is a powerful tool for the recognition of signatures of interesting processes, such as $H^0 \rightarrow Z^0Z^0 \rightarrow l_1^+l_1^-l_2^+l_2^-$ or Supersymmetry (SUSY) signatures. Three types of gaseous particle detectors for the identification of muons are used, namely Drift Tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). More information on the geometry and schematics of the components can also be found in [11] and [24].

The Barrel Region The barrel detector geometry is composed of four muon stations[11] that extend outwards of the magnet solenoid, built-in within the iron return yoke. The return magnetic field of the yoke provides for an additional momentum measurement for the muons. The drift tubes (DTs), exclusive to this geometrical region of the Muon System, and the RPCs, are responsible for the measurement due to the uniform magnetic field and low muon rates. Electrons originating from excitations of the atoms of the Ar/CO_2 gas mixture inside the tube start an avalanche. The position of the incident muon can be calculated using the drift velocity and the time that electrons take to reach the anode wire[25]. The curvature of the path of the muons is measured by the hits on the four station detectors, thanks to the return $\vec{B}$, where the momentum of a muon is inferred. The pseudorapidity range in this sector is up to $|\eta| < 1.2$.

The four muon stations of 250 chambers are divided into five wheels (labeled as YB+2, YB+1, YB+0, YB-1, YB-2) across the beam axis z, similar to the rings of the Hadron Outer (HO) system. Each wheel, in turn, is divided into 12 sectors, covering a 30° ϕ angle. All chambers are 40cm shorter along the beam direction in wheels YB+1 and YB-1 as these wheels host the chimneys for the magnet cryogenic lines. In each wheel, the two innermost stations, labeled MB1 and MB2 are constructed of a DT chamber placed between 2 RPCs. For the two outermost stations, MB3 and MB4, a DT chamber is coupled to a layer made of one, two or four RPCs, depending on the wheel and station. See figure 1.13 for a schematic.

The Endcap Region The Endcap Region is exposed to high particle flux and a non-uniform magnetic field. 468 Cathode strip chambers (CSCs)[11], exclusive to this geometrical region of the Muon System due to the high radiation, non-uniform magnetic field and high muon rates are used in conjunction with RPCs in this detector which covers a pseudorapidity range of $0.9 < |\eta| < 2.4$. A CSC is constituted by an array of anode wires perpendicularly crossed against an array of cathode strips within a gas mixture of $Ar/CO_2/CF_4$ (nominally 30%/50%/20%). Muons passing through the CSC detach electrons from the gas which in turn are attracted towards the positively charged anode wires. The positive charged ions of the gas move towards the cathode strips[26]. An RPC is made from two parallel plates of high resistivity material, the anode and the cathode, separated by a a thin volume of gas mixture of $C_2H_2F_4$/i-C_4H_10 (nominally 95.5%/4.5%). A travelling muon ejects electrons from the gaseous material. The electrons start an avalanche of electrons, which in turn give hits to the external metallic strips behind the resistive plates. The identification of electrons forms a pattern and the momentum of the

muon is inferred. The RPCs have excellent time resolution of 1ns[27]. Like the Barrel Region, four stations (labeled ME1, ME2, ME3, ME4 in order of increasing distance in the beam axis) of chambers are mounted on the disks which enclose the CMS magnet. Each disk is divided into 2 concentric rings around the beam axis, with the exception of chambers ME1 which is divided into 3 rings.

The Gas Electron Multiplier (GEM) chamber[28] is a new addition installed in Run 2 in the Muon System. The role of the new GEM detector in the first muon station in the endcap region is to maintain and improve the forward muon triggering and reconstruction in the region $1.6 < |\eta| < 2.2$ in the face of high luminosity[29]. GEMs are gaseous detectors, using a gas mixture of Ar/CO_2. A 50μm thick insulating polymer with microscopic holes is etched in a regular hexagonal pattern and is surrounded by copper conductors. The GEM chambers consist of foils along with the gas mixture. A voltage is applied at the copper conductors, generating sharp electric fields in the holes. Travelling muons eject electrons from the gas which are then drifted to the foils and multiplied at the holes. The induced electron avalanche is read out by electronics.

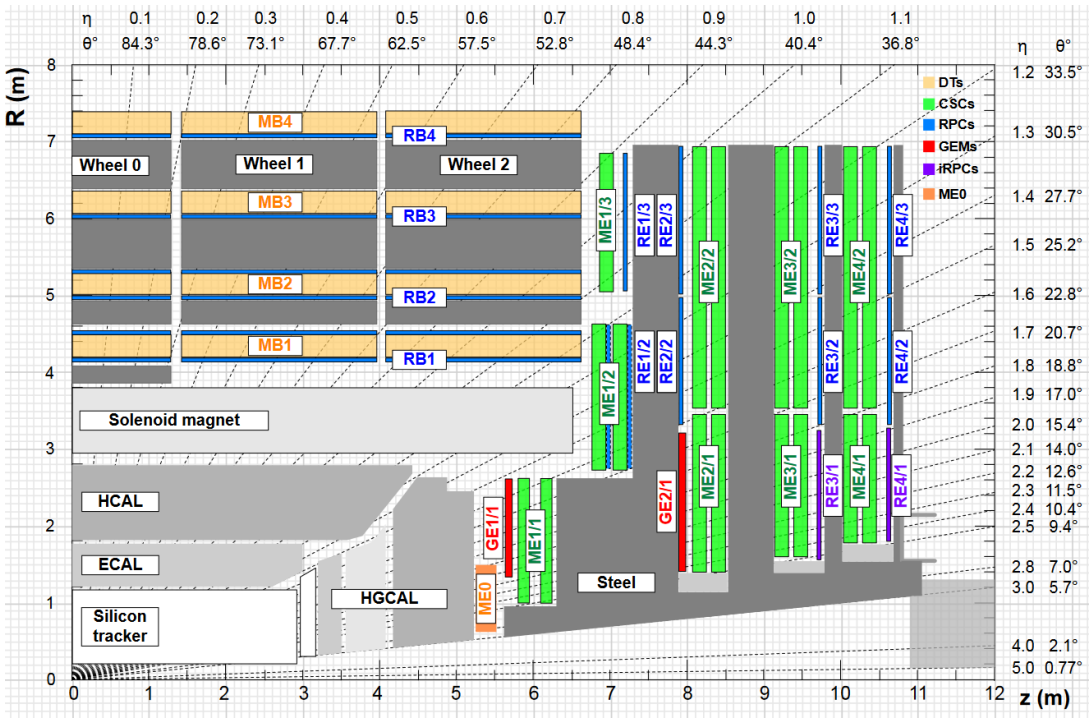


Figure 1.13: Longitudinal cross sectional view of the Muon System. Each region is color coded. In the endcap region, the components that are expressed in the form X/Y denote the disk muon station number (X) and the ring muon station number (Y). The chambers GEM1/1, GEM2/1, ME0 and the iRPCs (improved RPCs) are upcoming LHC Phase-2 upgrades.[30]

Momentum Measurement The Muon System is optimized for measurement of 100GeV p_T muons. The momentum measurement accuracy using the Muon System is limited by several factors[24].

- Multiple scattering in the calorimeters and in the thick steel plates separating the muon stations
- The intrinsic resolution of the detectors
- Extreme energy loss, e.g. hard photon Bremsstrahlung

- Extra detector hits generated by muon radiation, δ-rays and various background types
- Chamber and component misalignment
- $\vec{B}$ field uncertainty

Low p_T muons have greater track curvature due to the magnetic field which improves the momentum estimate. However, the mean multiple scattering angle (a non-Gaussian effect) is inversely proportional to momentum, increasing the deflection of the trajectory for small p_T as the particle passes through absorber material[24]. The momentum error is said to be multiple scattering limited if this effect dominates. Muon tracks at 10GeV are used to explore this effect. By increasing the track momentum above this effect, the error is labeled measurement limited and increases linearly with p_T. High momentum muons lose energy via secondary radiation more often, skewing the p_T measurement towards lower p_T. This radiation creates background effects along with the muon measurement, worsening the measurement. This secondary radiation effect has larger errors than the multiple scattering effect and position measurement in the chambers. The momentum measurement can be studied via the momentum p_T residual

$$\frac{\Delta p_T}{p_T} = \frac{1/p_T^{\text{meas}} - 1/p_T^{\text{gen}}}{1/p_T^{\text{gen}}}$$

and fitting a Gaussian to this event distribution. The width of the fitted Gaussian distribution is the residual error estimate[24].

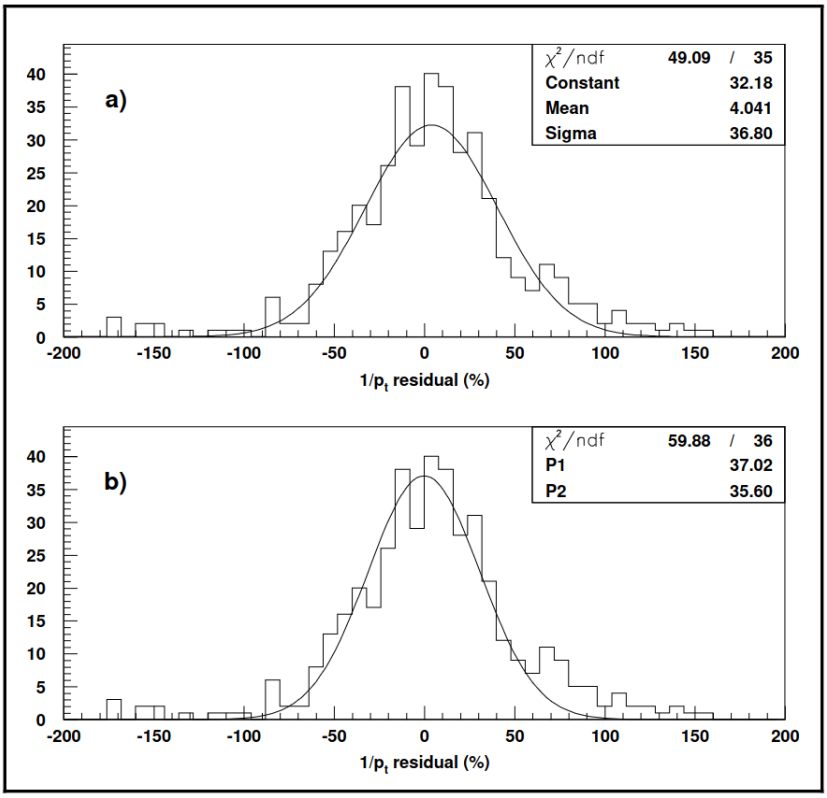


Figure 1.14: $1/p_T$ *residual distributions for 1TeV muons. a) The distribution is fitted by an unconstrained Gaussian. b) The distribution is fitted by a constrained Gaussian, where its mean parameter is zero. The presence of events on the right of the fitted constrained Gaussian tail denotes the catastrophic energy loss due to underestimating the value of a* p_T *track[24].*

1.4.6 Trigger System

The trigger system is tasked to reduce the high number of data intake before offline storage[11]. A large amount of the proton beam collision event data is of mundane interest to the experiment and since data storage speed and capacity is technologically limited, a mechanism that is responsible for the rejection of data online is imperative to the experiment. This event selection process, the CMS Trigger System[31], is divided into two main steps, the Level-1 (L1T) Trigger which reduces the event rate from 40MHz to about 100kHz, and the High-Level Trigger (HLT) which reduces the event rate further down to about 1kHz during Run 2.

The L1 Trigger The L1 Trigger system is based on FPGA technology which allows to reprogram its algorithms according to the needs of the experiment. The data from the two main calorimeters (ECAL, HCAL) and the Muon System are fed into this system. In order to process and decide how to reject incoming data, low resolution objects that describe physical quantities and particles called L1 candidates, such as photons, electrons, muons and jets satisfying requirements and thresholds are formed. Typical thresholds are the E_T and p_T quantities. The L1 Trigger also constructs global sums of E_T and E_T^{miss} (referred to as Missing Energy Transverse - MET). The L1 Trigger limits the data output rate according to the upper limit imposed by the CMS readout electronics at about 100kHz, before the data is passed into the next level in the Trigger system. The hardware system collects the information from the calorimeters and the muon detectors (figure 1.15) and decides whether to roughly accept or reject track candidates or events using quality factor criteria.

The Calorimeter Trigger consists of two Layers (figure 1.15, right). Inputs to Layer 1 are Trigger Primitives (TPs) from the ECAL and HCAL detectors, which are the measured energy deposits, which are calibrated and sorted by Layer 1. Layer 2 reconstructs and calibrates physics objects such as electrons, tau leptons, jets and energy sums from the calibrated TPs from its previous Layer as input. A demultiplexer (DeMux) board reformats the data of the event in such a way that the Global Trigger can process them.

The Muon Trigger architecture is shown in the left of figure1.15. The TPs consist of coordinates, timing and quality information from the Muon System detector. The TwinMux layer merges and refines barrel muon detector TPs into superprimitives, which are assigned a quality factor and sends them as input to the Barrel Muon Track Finder (BMTF). The TwinMux also sends the original TPs into the Overlap Muon Track Finder (OMTF). The CPPF[31] (figure 1.15, left) clusters RPC hits in the endcap detector and computes the θ and ϕ coordinates and passes its output to the Endcap Muon Track Finder (EMTF). The BMTF, EMTF and OMTF take processed inputs from components in the barrel detector (RPCs, DTs), the endcap detector (RPCs, CSCs) and from all components respectively. Each track finder builds muon track candidates, assign a quality factor to the track, measure the charge and p_T of each candidate muon. The Global Muon Trigger (μGMT) receives input from the aforementioned track finders. It sorts the candidate muons, removes possible duplicates and corrects the spatial coordinates of each muon candidate by extrapolation. It provides a track quality to the Global Trigger (μGT) for specific trigger paths for data analyses and takes the processed information from the Calorimeter Trigger and the Muon Trigger.

The High Level Trigger Unlike the L1 Trigger, the HLT is a software system primarily composed of algorithms running on commercial processors. These algorithms run on an

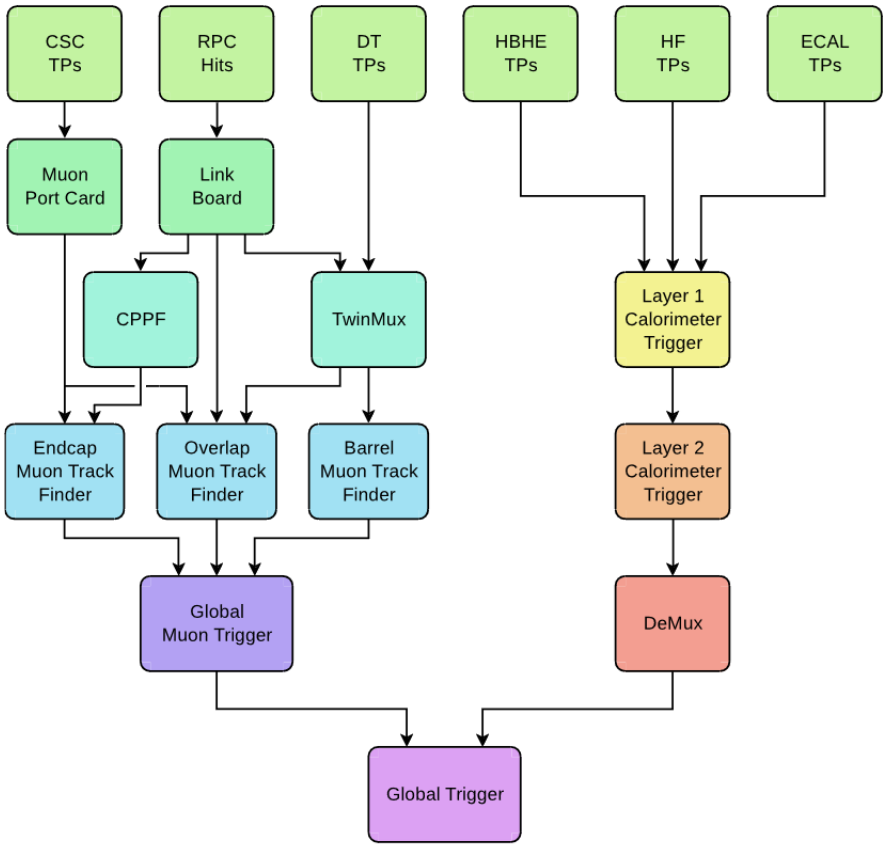


Figure 1.15: Diagram of the upgraded CMS L1 Trigger system during Run 2[31]. The abbreviations correspond to: Trigger Primitive (TP), Cathode Strip Chamber (CSC), Drift Tube (DT), Resistive Plate Chambers (RPC), Concentration Preprocessing & Fan-out (CPPF), Hadron Barrel (HB), Hadron Endcap (HE), Hadron Forward (HF), Demultiplexing card (DeMux).

event filter farm, located on a single machine running the Scientific Linux OS which is composed of at least 13000 CPU cores[32]. The data output rate from the previous level is reduced to just about $1kHz$ for offline storage. The modular structure of trigger paths, which are sequences of reconstruction and filtering blocks of increasing complexity, aim to optimize the processing time. If an event does not pass the filter, subsequent filters are not checked and the event along with its physics objects is rejected. A Physics object can be reconstructed with different methods, sometimes with two methods. The Particle Flow (PF)[33] algorithm is used to reconstruct hadronic τ, jets and MET. Jets, MET, electrons and photons as well as muons are reconstructed from clusters of ECAL and HCAL energy deposits. Other methods of reconstruction or identification include a neural network based classifier (Deep Combined Secondary Vertex) for b tagged jets[34]. Disks contain the processed data after the reconstruction in their final format and can be distributed to various groups of scientists for analysis. Data Scouting is a method where a small summary of the reconstructed event quantities is saved, reducing the event size, which allows recording events with a higher rate, allowing loosened filters. Another method that increases the amount of data throughout a Run is Data Parking, where data is ”parked” (saved) on a disk where physics objects are not reconstructed. The reconstruction can be performed whenever the LHC is on a Shutdown period. The HLT algorithm software is similar to the software used for offline analysis.

2 The Standard Model

2.1 Overview

The Standard Model of Particle Physics is the currently accepted model that describes the fundamental constituents of the Universe[35] and their interactions excluding Gravity. It is the result of the accumulation of years of theoretical and experimental work and marks the baseline for current research in High Energy Physics. According to the Standard Model, there are 17 elementary particles responsible for making up the more composite systems, ranging from the commonplace atoms and molecules to extremely short-lived baryons and mesons. The particles are divided into two main categories, fermions and bosons, according by their spin.

Fermions are particles that have half odd integer spin in multiples of $\hbar$. All 12 elementary fermions have a spin of $\hbar/2$ and are further separated into two more categories, quarks and leptons. These twelve particles constitute all of matter, as combinations of quarks and leptons make up atoms and molecules, which are present in organic and inorganic matter. What's more, leptons and quarks are also divided into three generations each, showcased in table 2.1.

Particle	Mass (MeV)	Charge (e)	Particle	Mass (MeV)	Charge (e)
Leptons					
Electron e	0.511	-1	e-neutrino ν_e	$< 1.1 \cdot 10^{-6}$	0
Muon μ	105.658	-1	μ-neutrino ν_μ	< 0.19	0
Tauon τ	1776.86	-1	τ-neutrino ν_τ	< 18.2	0
Quarks					
Up u	2.16	$2/3$	Down d	4.67	$-1/3$
Charm c	$1.27 \cdot 10^3$	$2/3$	Strange s	93	$-1/3$
Top t	$172.76 \cdot 10^3$	$2/3$	Bottom b	$4.18 \cdot 10^3$	$-1/3$

Table 2.1: Fermions with mass and electric charge. Each row represents each of the three generations[19].

Bosons are characterized by their integer spin in multiples of $\hbar$. All spin-1 elementary bosons are mediators of interactions, responsible for describing the force of their respective interaction. The Strong, Weak and Electromagnetic Interactions are described this way by the Standard Model. The photon γ is the mediator of the Electromagnetic Interaction, the gluons g are the mediator of the Strong Interaction and the $W^{\pm}, Z$ bosons are the mediators of the Weak Interaction. The Higgs Boson H^0 is the particle that grants mass to other particles through Spontaneous Symmetry Breaking. As of yet, Gravity has not been successfully incorporated in the Standard Model. The hypothetical boson responsible for mediating information about Gravity is the graviton. Gravitational forces between two elementary and composite particles are negligible.

Composite particles are created when two or more elementary particles form a bound state. A bound state of two or more particles is a stable system with definite energy. Hadrons are bound states of two or more quarks and are separated into two categories, mesons and baryons. Mesons are bound systems containing two quarks whereas baryons are bound systems containing an odd number of quarks. Atoms are bound states between electrons, neutrons n (ddu) and protons p^+ (duu), which in turn are baryonic bound states of up and down quarks. Molecules are bound states of two or more atoms. Composite

Particle	Mass (MeV)	Charge (e)	Spin ($\hbar$)
Gauge Bosons			
Photon γ	0	0	1
Gluon g	0	0	1
W	$80.37 \cdot 10^3$	± 1	1
Z	$91.188 \cdot 10^3$	0	1
Higgs H	$125.25 \cdot 10^3$	0	0

Table 2.2: Gauge bosons with mass, electric charge and spin. [19].

particles make up ordinary matter and are the object of study of numerous fields of Science.

What's more, much like every other field in Physics, a special system of units is chosen as preference in High Energy Physics. The International System (SI) of Units and the CGS (Centimetre-Gram-Second) are not very applicable due to the order of magnitude of physical quantities (energies, masses, wavelengths, time constants, etc). The Natural System is therefore attained for High Energy Physics. The Natural System does not simply define the scale of the order of magnitude for the quantities of interest, but it also changes their respective dimensions for faster theoretical calculations. Dimension setting is performed via:

$$\hbar = c = \epsilon_0 = \mu_0 = 1$$

The Natural System of Units forces length and time dimensions to be inversely proportional to energy and mass dimensions. This process of dimension setting is reversible through the Buckingham Pi Theorem and it is used when experimental data is required to be expressed and showcased. Throughout this thesis from this point on, the Natural System is applied, unless stated otherwise. Typical atomic processes such as electron ionization take place at energies of about 1eV (=1.6×10^{-19}J), nuclear processes such as alpha or beta decays have energies of about several MeV, quark related processes have energies in the GeV regime.

2.2 Theoretical Framework

This section regards a short review of the theoretical mathematical framework of Particle Physics of the Standard Model. It includes the Lagrangian Formalism, necessary formalism in Symmetries and Group Theory, as well as description for various mathematical technicalities such as Lorentz four-vectors and Dirac spinors to complement the underlying framework.

2.2.1 Lagrangian Formulation

The Lagrangian formulation is the foundation of the Standard Model. In order to produce equations that describe the allowed interactions within a theory, ones needs to start from a Classical Field Theory (CFT) and then introduce the first quantization into a Quantum Field Theory (QFT). Special Relativity is inherently included, as opposed to simple Classical Newtonian due to the nature and energy range of phenomena that need to be properly described. Both CFTs and QFTs use the concept of the field $\phi(x^\mu)$, a functional that assigns a value in each spacetime[5] point $x^\mu = (t, \vec{x})$ instead of the concept

[5] $\mu \in \{0,1,2,3\}$. For $\mu = 0$, $x^0 = t$ and for $\mu = i \in \{1,2,3\}$ x^i is the ith component of the $\vec{x}$ vector. Typically, Greek letters denote all four possible components of spacetime, whereas Latin letters denote the spatial

of the particle itself. QFTs have the advantage of resolving several inconsistencies such as negative energies in the Klein-Gordon equation and the Causality violation. The spin of the particle is attributed to this functional, and depending on its value, a different kind of mathematical object describes it (e.g. scalar, spinor). To start, the minimization of the Action functional $\mathcal{S}$ yields the equations that describe the correct possible interactions as predicted by the Standard Model. The starting point is the functional itself[36]:

$$\mathcal{S} = \int L\, dt$$

where L is the Lagrange function. The functional typically used is the Lagrangian density $\mathcal{L}$ and not the Lagrange function as L denoted above. The Lagrangian density is given by:

$$L = \int \mathcal{L}\left(\phi(x^\mu), \partial_\mu \phi(x^\mu)\right)\, d^3x$$

where $\partial_\mu = (\partial/\partial t, \vec{\nabla})$. The Action is then written as:

$$\mathcal{S} = \int \mathcal{L}\left(\phi(x^\mu), \partial_\mu \phi(x^\mu)\right)\, d^4x$$

It is common practice to drop the Lorentz four-vector index μ since it prevents unnecessary clutter while reading and writing equations. The field is then simply written as $\phi(x)$ and not $\phi(x^\mu)$ where the reader is always aware that x denotes a Lorentz four-vector. When the need for a three-dimensional vector arises, $\vec{x}$ is used instead. All four-vector notation is used throughout Particle Physics from formulating Lagrangians to calculating cross sections and decay rates of particles. What's more, the presence of Lorentz Invariant quantities are realized via the notion of $x_\mu x^\mu$ or $p_\mu p^\mu$, an inner product operation between a four-vector and its dual four-vector. These Lorentz invariants are scalar quantities that are independent of reference frame choices. In order to extract the equations of motion for a given Lagrangian density to find the solutions of the fields $\phi(x)$, minimization of $\mathcal{S}$ is performed ($\delta\mathcal{S} = 0$) with respect to $\phi(x)$ and $\partial_\mu\phi(x)$. The result attained is the well known Euler-Lagrange equations from Classical Mechanics, now upgraded for a Classical Field Theory:

$$\frac{\partial \mathcal{L}\left(\phi(x), \partial_\mu\phi(x)\right)}{\partial \phi(x)} = \partial_\mu \frac{\partial \mathcal{L}\left(\phi(x), \partial_\mu\phi(x)\right)}{\partial_\mu \phi(x)} \tag{2.1}$$

An alternative to the Lagrangian formalism is the Hamiltonian formalism. As a final step before introducing the first quantization, the conjugate momentum $\pi(x)$ to the field $\phi(x)$ is defined:

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial\left(\partial_0 \phi(x)\right)} \tag{2.2}$$

and then the Hamiltonian H as well as the Hamiltonian density $\mathcal{H}$ is acquired, via a Legendre transformation:

$$H = \int \pi(x)\dot{\phi}(x) - \mathcal{L}\, d^3x = \int \mathcal{H}\, d^3x$$

It is from this point on that the quantization is required to promote the Field Theory from a Classical formulation into a Quantum one. This is done by upgrading the Poisson Brackets into commutators, and the two canonical variables $\phi(x), \pi(x)$ into operators:

$$[\phi(x), \pi(y)] = i\delta(x-y), \quad [\phi(x), \phi(y)] = [\pi(x), \pi(y)] = 0$$

component of a four-vector. This notation is followed throughout this thesis, unless stated otherwise.

where the commutators above are expressed in the Heisenberg picture. With the formalism established, what remains is the addition of mathematical constructs that allow operations to take place. This means that in order to make use of the formalism, one needs to supply the required rules and technicalities that will eventually become the interactions with the particles that the Standard Model describes.

2.2.2 Symmetries

The mathematical forms that the Symmetries that the Lagrangians take are described with Group Theory. What's more, there is no preceding step; all Lagrangian terms are entered by hand provided that they satisfy the given Symmetries. This means that the construction of a Lagrangian term by term is not a definite process as it is the equivalent of the theorist's experimental apparatus. The primary tool for a Lagrangian's construction are Symmetries. Symmetries are an integral part of the theoretical formulation and it's the physicist's job to be fluent in that language, as they set the groundwork of a theory and how that theory reflects and interacts in the world that it is tested on. As for Group Theory, the language where Symmetries are based on, require a method that can express those symmetries. This is done through the *group generators*, which are responsible for generating the elements of a group, a set of elements combined with a binary operation which follow specific axioms. Symmetries are also understood as topological defects in recent developments, however for the mathematical formulation of the Standard Model in this thesis, the descriptions given here are satisfactory. Furthermore, a Lie Group is group which is also a differentiable manifold, meaning that the group itself has continuous elements.

Definition 2.1 *For a Lie Group $\mathcal{G}$, the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$ in the matrix representation form is given by the elements X: $e^{tX} \in \mathcal{G}\ \forall\, t \in \mathbb{R}$, together with a binary operation $[\bullet, \bullet]$, called the Lie bracket.*

Group generators obey specific algebras, where in the scope of the Standard Model they are given within a commutator or anticommutator relation. Lie groups allow for the establishment of fields once strong conditions are required, for example Local Gauge Invariance, which will be explained later. The Standard Model makes use of three simple continuous Lie groups in its entirety to formulate the elementary interactions: $U(1)$, $SU(2)$ and $SU(3)$. The group $SU(n)$ has n^2-1 generators. This is because of the relationship $U(n) = SU(n) \times U(1)$. where $U(n)$ has n^2 generators, and the condition that $\det[U] = +1$ for the group elements of $SU(n)$ elements eliminates one generator. The Lorentz Lie group (and by extension, the Poincare Group) is used to connect coordinate system transformations and translations within the Standard Model. Despite continuous symmetries, discrete groups are also present within the theory in order to explain more symmetries, such as parity, time and charge conjugation.

Noether's Theorem Noether's theorem states that for a differentiable symmetry in the Lagrangian there is a corresponding conserved charge or current. In the case of internal symmetries, which are important for interacting field theories, the invariance of the Lagrangian $\delta\mathcal{L} = 0$ under an infinitesimal transformation of a field Φ_i itself, $\Phi_i' = \Phi_i + \delta\Phi_i$ is written as:

$$\delta\mathcal{L} = \mathcal{L}(\Phi^i, \partial_\mu\Phi^i) - \mathcal{L}(\Phi^i + \delta\Phi^i, \partial_\mu\left(\Phi^i + \delta\Phi^i\right))$$

Since the transformation is infinitesimal, the Taylor expansion of the above expression to first order is enough to determine the conserved current:

$$\delta\mathcal{L} = -\frac{\partial\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial\Phi^i}\delta\Phi^i - \frac{\partial\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial(\partial_\mu\Phi^i)}\partial_\mu\delta\Phi^i = 0$$

The Euler-Lagrange equation for fields (eq. 2.1) can be used to replace the first term of the above expression:

$$\delta\mathcal{L} = -\partial_\mu\left(\partial\frac{\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial(\partial_\mu\Phi^i)}\right)\delta\Phi^i - \frac{\partial\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial(\partial_\mu\Phi^i)}\partial_\mu\delta\Phi^i = 0$$

Finally, by using the product rule:

$$\partial_\mu\left(\frac{\partial\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial(\partial_\mu\Phi^i)}\delta\Phi^i\right) = 0$$

Therefore, if the Lagrangian is invariant under the field transformation $\Phi_i' = \Phi_i + \delta\Phi_i$, the following quantity J^μ, called Noether current which fullfills the continuity equation:

$$J^\mu = \frac{\mathcal{L}(\Phi^i, \partial_\mu\Phi^i)}{\partial(\partial_\mu\Phi^i)}\delta\Phi^i, \quad \partial_\mu J^\mu = 0$$

Furthermore, it can be shown that the following equation is also true by using the continuity equation:

$$\partial_t\int J^0\, d^3x = 0$$

which implies that an associated quantity is also conserved in time. Specifically, invariance under displacements of the field itself leads to the conservation of conjugate momentum density $J^0 = \pi(x)$ (eq. 2.2). Other applications of Noether's Theorem on non-internal symmetry cases include that the physical momentum, angular momentum and energy of a particle are invariant under spatial, rotational[6] and temporal transformations. The link between the physical momentum and spatial translations is given by the transformation $\Phi(x') = \Phi(x + \epsilon)$.

Group $U(1)$ The trivial group $U(1)$ describes the set of all elements of the field $\mathbb{C}$ with modulus equal to 1. Each transformation describes rotations in the complex plane $\mathbb{C}$. The label of the group stands for *Unitary*. The trivial $U(1)$ group is the only connected Abelian Lie group which implies that particles described by $U(1)$ do not interact with themselves. What's more, it has no finite set of generators and is not described in a matrix representation.

Group $SU(2)$ The group $SU(2)$, abbreviated from *Special Unitary*, describes the set of 2×2 unitary complex Hermitian ($U^\dagger U$, $[U] \in SU(2)$) matrices in matrix representation form $[U]$. The label *Special* indicates that all group elements in matrix representation $\det[U] = 1$. If the element $U \in SU(2)$ is referred to a matrix, it is correct to state it as $[U]$ although this notation is dropped throughout discussion in the Standard Model of Physics. The three generators for this group are:

$$J_i = \frac{1}{2}\sigma_i$$

[6]Rotations must be two-dimensional or above so that the group associated with the transformations can be continuous (and therefore a Lie group).

where σ_i, $i = \{1, 2, 3\}$ are the three Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

following the Lie algebra:

$$[J_i, J_j] = i\epsilon_{ijk} J_k$$

The traceless property of the Pauli matrices enforces them as the generators of the group $SU(2)$ up to a multiplicative factor $1/2$. The group elements all obey the relation:

$$U(\theta_i) = e^{-i\theta_i \sigma_i / 2}$$

where $\theta_i \sigma_i$ is a dot product shorthand for:

$$\theta_i \sigma_i = \theta_1 \sigma_1 + \theta_2 \sigma_2 + \theta_3 \sigma_3 = \vec{\theta} \cdot \vec{\sigma}$$

these types of dot product shorthands will have the indices of the two objects in the same origin in order to differentiate between Lorentz four-vector dot products, or be expressed as dot products in vector notation.

Group $SU(3)$ The $SU(3)$ group describes the set of 3×3 unitary matrices of $\det\{U\} = 1$. The group elements are generated similarly to the $SU(2)$ group:

$$U(\theta_i) = e^{-i\theta_\alpha \lambda_\alpha / 2}$$

In $SU(3)$ there are eight generators T_α $\alpha \in \{l\}_1^8$:

$$T_\alpha = \lambda_\alpha / 2$$

where λ_α are the Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

These eight generators follow the Lie bracket algebra:

$$[T_\alpha, T_b] = i f_{\alpha bc} T_c$$

where $f_{\alpha bc}$ are the structure constants of the group. The structure constants are antisymmetric under the interchange of any pair of indices and equal to zero for any repetition of the same value of two or more indices (e.g. f_{113}). Moreover:

$$f_{123} = 1, \qquad f_{458} = f_{678} = \frac{\sqrt{3}}{2}, \qquad f_{147} = f_{165} = f_{246} = f_{257} = f_{345} = f_{376} = \frac{1}{2}$$

Lorentz Group $SO(1,3)$ The $SO(1,3)$ group is the set of all transformations that preserve the inner product of Minkowski space $x^\mu x_\mu = x^\mu g_{\nu\mu} x^\nu$. The label $(1,3)$ refers to the structure of the Minknowski metric diagonal in the matrix representation form[7]:

$$[g] = diag(1,-1,-1,-1)$$

The elements of the Lorentz Group are the Lorentz Transformations of Special Relativity. Particularly, the transformations are *special* (proper, $\det\Lambda = +1$) and have $\Lambda^0_0 = +1$ (orthochronous). The six group generators J_i, K_i $i \in \{l\}^3_1$ of the Lorentz group obey the following relations:

$$[J_i, J_j] = i\epsilon_{ijk}J_k, \quad [J_i, K_j] = i\epsilon_{ijk}K_k, \quad [K_i, K_j] = -i\epsilon_{ijk}J_k,$$

where J_i, K_i are regarded as the generators of rotations and boosts respectively. The rotation generator is Hermitian $J_i^\dagger = J_i$ and the boost generator is Anti-Hermitian $K_i^\dagger = -K_i$. The explicit form of the generators is given by:

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk}, \qquad K_i = M_{0i}$$

where $M^{\mu\nu}$:

$$(M^{\mu\nu})_{\alpha\beta} = g^\mu_\alpha g^\nu_\beta - g^\mu_\beta g^\nu_\alpha, \quad M^{\mu\nu} = [(M^{\mu\nu})^\alpha_\beta]$$

The group elements are expressed via:

$$\Lambda(\vec{\theta},\vec{\phi}) = e^{i\vec{J}\cdot\vec{\theta}+i\vec{K}\cdot\vec{\phi}}$$

Using the generators J_i, K_i, a new set of six generators $N_i^\pm$ equivalent to the former can be defined,

$$N_i^\pm = \frac{1}{2}(J_i \pm iK_i)$$

showcasing that the Algebra of $SO(1,3)$ described above consists of two copies of the $SU(2)$ algebra:

$$[N_i^+, N_j^+] = i\epsilon_{ijk}N_k^+, \quad [N_i^-, N_j^-] = i\epsilon_{ijk}N_k^-, \quad [N_i^+, N_j^-] = 0$$

Poincare Group $ISO(1,3)$ The Poincare Group is a semidirect product $SO(1,3) \rtimes \mathbb{R}^4$ between the Lorentz group and translations in Minkowski space respectively. A Poincare group element (Λ, α) is equivalent to the transformation $x \to \Lambda x + \alpha$. The generators of the Poincare group include the generators of the Lorentz group plus four additional generators P_μ. These generators in position space take the familiar form $P^\mu = i\partial^\mu$. The Poincare group algebra is described by:

$$[J_i, J_j] = i\epsilon_{ijk}J_k, \quad [J_i, K_j] = i\epsilon_{ijk}K_k, \quad [K_i, K_j] = -i\epsilon_{ijk}J_k, \quad (= \text{Lorentz algebra})$$

$$[J_i, P_j] = i\epsilon_{ijk}P_k, \quad [J_i, P_0] = 0, \quad [K_i, P_j] = -i\delta_{ij}P_0, \quad [K_i, P_0] = -iP_i$$

Using the definition of $M_{\mu\nu}$ described above, the Poincare Algebra simplifies to:

$$[P_\mu, P_\nu] = 0, \quad [M_{\mu\nu}, P_\rho] = i\left(g_{\mu\rho}P_\nu - g_{\nu\rho}P_\mu\right)$$

[7] The group $O(4)$ would use the identity metric $diag(1,1,1,1)$.

2.3 The Dirac Equation

This section is appropriate in introducing the behavior of fermions inside the Standard Model, as all of them have spin-$1/2$. In addition, the existence of antiparticles is predicted by the Dirac Equation. The free Dirac Lagrangian showcases the existence of spin-$1/2$ particles and implies the existence of anti-particles. In order to fully interpret the free Dirac Lagrangian, one needs to supply some additional mathematical framework. A Clifford algebra[8] obeying the anti-commutation relation, in matrix representation:

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$$

where γ^μ are the well-known Dirac Gamma matrices, are crucial in order to arrive into a proper expression of an equation that is not second order in x for spin-$1/2$ particles. They are given here for reference in the Weyl (or Chiral) representation, along with γ_5, a matrix that will be referred to later:

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \gamma_5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma^\mu = \{\hat{1}, \sigma^i\}, \quad \bar{\sigma}^\mu = \{\hat{1}, -\sigma^i\} \tag{2.3}$$

Naturally, this implies the presence of the Gamma matrices throughout the mathematical background of the Electromagnetic Interaction that is being formulated, or any other formulation that by extension requires spin-$1/2$ particles. The Dirac Lagrangian is expressed as:

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu\partial_\mu - m\right)\psi$$

where $\psi = \psi(x)$ is the the Dirac Field, a four component vector referred to as a Dirac spinor, not to be confused with a Lorentz four-vector. The Dirac field is Lorenz invariant only as billinear forms $\bar{\psi}\psi$ where $\bar{\psi} = \psi^\dagger\gamma^0$. Applying the Euler-Lagrange equations (2.1), the Dirac equation is acquired:

$$\left(i\gamma^\mu\partial_\mu - m\right)\psi = 0 \tag{2.4}$$

2.4 Electromagnetic Interaction

It is worth nothing that the Dirac Lagrangian is globally gauge invariant. A global gauge transformation for the field ψ is simply described as:

$$\psi(x) \to e^{iga}\psi(x)$$

where a, g are constants that do not depend on x. In this case, a is referred to as a phase constant and g as the coupling strength of the interaction, whose magnitude meaningful when at least an additional interaction is added to the Lagrangian which is a measure of relative interaction strength. According to Noether's theorem, there must be a conserved current with this internal symmetry. The infinitesimal transformation $\psi' \to \psi + \delta\psi$ for the global gauge transformation $\psi' = e^{iga}\psi$ expanded to first order yields:

$$\psi' = \psi + iga\psi$$

this leads to a conserved current:

$$J^\mu = -ag\bar{\psi}\gamma^0\psi$$

[8]Fermion algebras are described with anticommutators, whereas boson algebras are described with commutators.

for the continuity equation:

$$\partial_\mu J^\mu = 0$$

The zeroth component $\mu = 0$ of the conserved charge corresponds to the electric charge density ρ and integrating this quantity in $\mathbb{R}^3$ yields the conserved Noether charge:

$$Q = \int \rho \, d^3x = Q = \int j^0 \, d^3x = \int -ga\bar{\psi}\gamma^0\psi \, d^3x = -ga \int \bar{\psi}\gamma^0\psi \, d^3x$$

This conserved charge is the electrical charge. Therefore, the global $U(1)$ symmetry leads to the conservation of the electric charge. However, one needs to demand Local Gauge Invariance, which prevents the group's parameters from being constant in spacetime x by allowing the a phase to have a dependence on x. The following transformation denotes Local Gauge Invariance:

$$\psi(x) \to e^{iga(x)}\psi(x)$$

This transformation does not guarantee that a given Lagrangian will be invariant under Local Gauge Invariance. In order to circumvent this issue, the introduction of the covariant derivative D_μ is introduced:

$$D_\mu = \partial_\mu - igB_\mu$$

This is equivalent to treating the factor $e^{ia(x)}$ as a constant when taking the covariant derivative D_μ. However, this comes with a cost of the introduction of a new quantity, B_μ which transforms as:

$$B_\mu \to B_\mu - \partial_\mu a(x)$$

The field B_μ, which is a Lorentz four-vector, is eventually given the role of the photon field, and the Dirac Equation describes charged particles of spin-$1/2$. The process above fully incorporates Local Gauge Invariance. Finally, before writing out the full form of the Lagrangian, a Lorentz invariant kinetic term for the new field B_μ is required in order for the new field to propagate, which means that it can transport information throughout spacetime. This term takes the form of $F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu}$ is the well known Electromagnetic tensor:

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

Finally, the Lagrangian that describes the Electromagnetic Interaction, or better yet, Quantum Electrodynamics (QED) under Local Gauge Invariance is expressed through the Lagrangian density:

$$\mathcal{L} = \bar{\psi}\left(i\gamma^\mu D_\mu - m\right)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

It's worth noting here, that should one add a photon mass term by hand, the Local Gauge Invariance would be violated. What's more, a mass term also introduces unrenormalizable divergences when calculating Feynman diagrams. This is an adequate description of QED, provided that no other interaction is present in the theory. This formulation is eventually re-formulated together with the Weak Interaction, which will be discussed later.

2.5 Quantum Chromodynamics

The formulation of the Strong Interaction, Quantum Chromodynamics (QCD) is similar to that of Quantum Electrodynamics, apart from a few changes due to the non-Abelian property of $SU(3)$. The $SU(3)$ group takes the role of facilitating the framework for QCD. Three color fields are required so that the theory is in line with experimental data, hence

the choice of the group. To impose Local Gauge Invariance, the following transformation is applied[9]:

$$Q(x) \to e^{iga_\alpha(x)T_\alpha}Q(x), \quad Q = (q_1\ q_2\ q_3)^T$$

One needs to start with a free Dirac field Lagrangian density for a triplet field Q:

$$\mathcal{L} = \bar{Q}\left(i\gamma^\mu\partial_\mu - m\right)Q$$

For the justification of the three color fields, eight traceless Hermitian generators T_α are required from the symmetry group $SU(3)$. The imposition of Local Gauge Invariance is not as trivial as in QED due to the fact that the group $SU(3)$ is non-Abelian. Despite this, the process is the same, where a covariant derivative is once again defined:

$$D_\mu = \partial_\mu - ig\mathcal{G}_\mu, \quad \mathcal{G}_\mu = T^\alpha G^\alpha_\mu$$

and the new fields transform as:

$$G^\alpha_\mu \to G^\alpha_\mu - \frac{1}{g}\partial_\mu a^\alpha(x) + G^b_\mu f^{\alpha bc}a^c(x)$$

where g is the coupling strength of the interaction. Once again, a kinetic term for the new color fields must be included in the form of $G^\alpha_{\mu\nu}G^{\mu\nu}_\alpha$. It is remarked that since $SU(3)$ is non-Abelian, the eight aforementioned strength tensors $G^\alpha_{\mu\nu}$ also take a different form:

$$G^\alpha_{\mu\nu} = \partial_\mu G^\alpha_\nu - \partial_\nu G^\alpha_\mu - g f_{\alpha bc}G^b_\mu G^c_\nu$$

Color is an exclusive property of the Strong Interaction and all the particles that can interact with gluons. According to Noether's theorem, the conserved charge is color. The final Lagrangian reads:

$$\mathcal{L} = \bar{Q}\left(i\gamma^\mu D_\mu - m\right)Q - \frac{1}{4}G^\alpha_{\mu\nu}G^{\mu\nu}_\alpha$$

A mass term for the mediators, which are the eight gluons, is prohibited by Local Gauge Invariance. The mediators are once again massless and this result is determined experimentally and the interactions between the quarks and gluons can be read from the Lagrangian density. Gluon interactions carrying different color charge are allowed and enabled by the fact that $SU(3)$ is non-Abelian. This is reflected to the new form of the field strength tensor compared to QED.

2.6 Electroweak Interaction

Following the formulation of the $U(1)$ symmetry, one asks how to repeat the process for the $SU(2)$ symmetry. As previously discussed, $SU(2)$ has three generators, and a global gauge transformation summing on the three generators for a doublet field ψ would be:

$$\psi' = e^{iga_i\frac{\sigma_i}{2}}\psi, \quad \psi = \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix}$$

The conserved charge of the Global Gauge Invariant $SU(2)$ is the isospin. The same process is done to upgrade into a local gauge symmetry by letting $a \to a(x)$. In the case of the $U(1)$ symmetry, where the change from global to local was performed, a new spin 1

[9]Once again, a distinct factor of g is present to denote the coupling strength of the interaction.

field was introduced. In this case, three new spin 1 fields need to be introduced with the same method. The covariant derivative in this case is:

$$D^\mu = \partial^\mu - ig\frac{\sigma^j}{2}W_j^\mu, j \in \{1,2,3\}$$

The locally $SU(2)$ invariant Lagrangian is:

$$\mathcal{L} = i\bar{\psi}\gamma_\mu D^\mu\psi - \frac{1}{2}\,\mathrm{Tr}\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right)$$

where:

$$\mathcal{W}^{\mu\nu} = \partial^\mu\mathcal{W}^\nu - \partial^\nu\mathcal{W}^\mu - ig\left[\mathcal{W}^\mu, \mathcal{W}^\nu\right], \quad \mathcal{W}^\mu = W_i^\mu\frac{\sigma^i}{2}$$

The fields $\mathcal{W}^\mu$ need to transform as:

$$\mathcal{W}^\mu \to \mathcal{U}(x)\mathcal{W}^\mu\mathcal{U}^{-1}(x) + \frac{i}{g}\left(\partial^\mu\mathcal{U}(x)\right)\mathcal{U}^{-1}(x), \quad \mathcal{U}(x) = e^{ia_i\frac{\sigma_i}{2}}$$

However, this is not the case with the $SU(2)$ symmetry as its gauge bosons, the three spin 1 fields, must be massive, as experiments have shown[37][38]. The inclusion of a mass term for the gauge bosons breaks Local Gauge Invariance, therefore the aforementioned Lagrangian cannot be used to describe interactions with three massive bosons as mediators. It wouldn't be ideal to show the formulation of a pure $SU(2)$ theory, but instead the unification between the Weak and Electromagnetic Interactions into the Electroweak Interaction. In order to do so, it is required to give a description of Glashow-Weinberg-Salam's Electroweak model and the more accurate form of the Weak Interaction as is present in the Standard Model will be mentioned in this section (section 2.7). However, the notion of chirality will be explained in this section. Since the experimental evidence shows that three mediator bosons of the Weak Interaction exist, the $SU(2)$ symmetry can be picked as a base to describe this interaction. A four component spinor ψ can be written as:

$$\psi = \begin{pmatrix}\psi_L\\ \psi_R\end{pmatrix}$$

What's more, the Hermitian Parity operator $\hat{P}$ is a transformation[35] that reverses the spatial and time component of a Lorentz four-vector. Three component vectors are classified in two categories, *vector* and *axial vector*. Under the Parity transformation, vector quantities, such as position and momentum obey $\vec{v} \to -\vec{v}$ and axial vector quantities, such as angular momentum obey $\vec{v} \to \vec{v}$. The most general Lorentz-invariant form for the interaction between a fermion and a spin-1 boson is a linear combination of bilinear covariants. This general form is required in the case of the Weak Interaction as it has been proven to violate Parity[39]. The interaction current j^μ is proportional to:

$$j^\mu \propto \bar{\psi}\left(g_V\gamma^\mu + g_A\gamma^\mu\gamma^5\right)\psi = g_V j_V^\mu + g_A j_A^\mu$$

where j_V^μ and j_A^μ are the vector and axial vector currents respectively. The absence of the proper factor is irrelevant to the discussion, as only the labels of j_V, j_A are required for the description of the Parity transformation. The Weak Interaction violates Parity *maximally*, satisfying when $|g_V| = |g_A|$. Experimental measurements show that $g_V = -g_A$, leading to:

$$j^\mu \propto \bar{\psi}\gamma^\mu\left(1-\gamma^5\right)\psi \tag{2.5}$$

which is known as the vector minus axial vector (V-A) interaction. Furthermore, it is seen that the ψ_L and ψ_R of components ψ yield the eigenvalues $-1, +1$ on the matrix γ^5 respectively. It is convenient to define the projection operators $\hat{P}_L, \hat{P}_R$:

$$\hat{P}_L = \frac{1-\gamma^5}{2}, \quad \hat{P}_R = \frac{1+\gamma^5}{2}$$

so that, when acting on the spinor ψ:

$$\hat{P}_L\psi = \begin{pmatrix} \psi_L \\ 0 \end{pmatrix}, \quad \hat{P}_R\psi = \begin{pmatrix} 0 \\ \psi_R \end{pmatrix}$$

or more conveniently, $\hat{P}_L\psi = \psi_L$ and $\hat{P}_R\psi = \psi_R$. The Weak Interaction couples only to "left-handed" particles and "right-handed" antiparticles. This notion of "left" and "right" is attributed to chirality, a distinction that only the Weak Interaction is able to interpret, similar to the color charge of the Strong Interaction. Therefore, from the expression in 2.5, the current interaction includes the projection operator P_L. Hence, only left-handed chiral particle states and right-handed chiral antiparticle sates participate in the Weak Interaction.

2.6.1 Cabibbo-Kobayashi-Maskawa Matrix

The unitary Cabibbo-Kobayashi-Maskawa Matrix (CKM) V_{CKM} is a component of the Lagrangian that regards transitions from up-type quarks to down-type quarks and vice versa in the Weak Interaction. The mass eigenstates of the three generations of quarks are not equivalent in the Weak Interaction unlike the lepton mass eigenstates [40]. This matrix arises from the Yukawa interactions with the Higgs condensate (see section 2.7.1.It can be shown that for charged current interactions using the $W^{\pm}$ boson, terms such as $u_L^{\dagger}(i\bar{\sigma}^{\mu})V_{\text{CKM}}d_L$ appear, where:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \quad d_L = \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \quad u_L = \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix}$$

Ideally, to prevent this differentiation between the mass eigenstates in the Weak Interaction, $V_{\text{CKM}} = I$. The CKM Matrix has four independent components that can be measured experimentally. The CKM matrix can be parametrized with four independent parameters, one of which can be denoted as:

$$V_{\text{CKM}} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}$$

where $s_{ij} = \sin\theta_{ij}$, $c_{ij} = \cos\theta_{ij}$ for $\{i, j\} \in \{k\}_1^3$, $i \neq j$ and δ is the phase responsible for all CP-violating phenomena in flavor changing processes. The three Euler anglers θ_{ij} along with the phase δ form the set of four independent parameters of the above representation of the CKM matrix. The values of these four parameters have been determined[19]:

$$s_{12} = 0.22500 \pm 0.00067 \qquad s_{13} = 0.00369 \pm 0.00011$$
$$s_{23} = 0.04182 \pm 0.00085 \qquad \delta = 1.444 \pm 0.027$$

Let U_{CKM} be defined through $U_{\text{CKM}\,ij} = \left|V_{\text{CKM}\,ij}\right|$, which expresses the moduli of the nine entries according to the aforementioned values of the four parameters of the CKM

matrix[10][19]. Then:

$$U_{\text{CKM}} = \begin{pmatrix} 0.97401 \pm 0.00011 & 0.22650 \pm 0.00048 & 0.00361 \pm 0.00011 \\ 0.22636 \pm 0.00048 & 0.97320 \pm 0.00011 & 0.04053 \pm 0.00083 \\ 0.00854 \pm 0.00023 & 0.03978 \pm 0.00082 & 0.999172 \pm 0.000035 \end{pmatrix}$$

The V_{CKM} matrix is an important probe of New Physics.

2.7 Spontaneous Symmetry Breaking and the Higgs Mechanism

The mass of the spin 1 mediators of the Electromagnetic and Strong interactions is zero, as demanded by Local Gauge Invariance. However, in the case of the Weak Interaction the three mediators are massive, as determined by experiments[37][38]. Since the addition of a mass term is forbidden, a different process must be responsible for the acquisition of mass for the mediators. This process is realized thanks to two concepts: Spontaneous Symmetry Breaking and the Higgs Mechanism. The former introduces a new spin 0 particle potential under a local Gauge Invariant $SU(2) \times U(1)$ symmetry, and the latter breaks this symmetry by choosing a specific minimum for the Higgs Potential. This process allows three of the four particles of the $SU(2) \times U(1)$ symmetry to acquire mass thanks to the form of the Higgs Potential, namely its parameters μ, λ and the choice of its minimum. Specifically, in Spontaneous $SU(2) \times U(1)$ Symmetry Breaking, Local Gauge Invariance yields three gauge fields W_μ^i from $SU(2)$ and one B_μ from $U(1)$. Out of the complex $SU(2) \times U(1)$ doublet, three fields become Goldstone bosons, whereas the remaining one, the Higgs boson, gains mass after the symmetry is spontaneously broken, and the symmetry becomes hidden from the choice of the ground state. What remains is to relocate those three Goldstone bosons as degrees of freedom into the three fields W_μ^i that arose through Local Gauge Invariance and its infinitesimal transformations on the fields ϕ. By substituting the new pertrubative about the minimum field, the three W_μ^i fields acquire mass and longitudinal polarization by eliminating the three massless Goldstone bosons. A massive scalar boson remains which is the Higgs Boson. The process of the elimination of the Goldstone bosons and the mass acquisition of the W_μ^i fields is referred to as the Higgs Mechanism and allows the avoidance of massless particles. The Lagrangian for a free spin zero complex field ϕ is globally $U(1)$ invariant:

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi^\dagger \partial^\mu \phi - \mu^2 \phi^\dagger \phi\right)$$

By introducing a self-interaction quadratic term for the ϕ field, $-\lambda\left(\phi^\dagger\phi\right)^2$, gauging this Lagrangian and introducing 4 spin-1 fields, the following Lagrangian is obtained:

$$\mathcal{L} = \left(\left(\partial_\mu + ig'\mathcal{W}_\mu + igB_\mu\right)\Phi^\dagger\right)\left(\left(\partial^\mu - ig'\mathcal{W}^\mu - igB^\mu\right)\Phi\right) + \mu^2\Phi^\dagger\Phi - \lambda\left(\Phi^\dagger\Phi\right)^2$$

where $\Phi = (\phi_1, \phi_2)^T$ The Higgs potential is defined as $V(\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda\left(\Phi^\dagger\Phi\right)^2$. Let ϕ_1, ϕ_2 be the two complex component fields so that $\Phi = (\phi_1\ \phi_2)^T$. A scalar field with a negative mass term $\sim -m^2\phi^2$ has a minimum at $\phi = 0$[41]. For a positive mass term this isn't the case. It can be seen from the Higgs potential that it is separable additively $V(\Phi) = V(\phi_1) + V(\phi_2)$. The minima of the potential $V(\Phi)$ can be calculated through $\partial V(\Phi)/\partial\phi_1$[11]. This leads to the condition:

$$|\phi|_{1\,\text{min}} = \sqrt{\frac{\mu^2}{2\lambda}} \Leftrightarrow \phi_{1\,\text{min}} = e^{i\theta}\sqrt{\frac{\mu^2}{2\lambda}} \quad \theta \in \mathbb{R}$$

[10] where the larger error is kept from the reference.

[11] since the potential is separable in the two complex field components, one component, let ϕ_1 can be chosen. Let, for simplicity, $\phi_2 = 0$

These minima $\phi_{1\,\min}$ represent a circular curve in the $\mathbb{C}$ plane with radius $\sqrt{\mu^2/2\lambda}$. The process of Symmetry Breaking picks one of the possible minima, therefore, let $\Phi_{\min}$:

$$\Phi_{\min} = \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{2\lambda}} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad v = \sqrt{\frac{\mu^2}{\lambda}}$$

In such scenarios, calculations are performed by expanding around the minimum as no exact solutions are available. It is to be remarked that expanding about a stable minimum is the proper technique in Perturbation Theory. This is done by picking a vacuum for the theory by minimizing the Lagrangian with respect to that field and performing perturbative calculations around that new vacuum, without affecting the conservation of currents so that Noether's Theorem isn't violated. If the calculation is performed on the pre-transformed Lagrangian, the calculation would not converge. Since this transformation of fields does not change the rules and interactions, the pre-transformation and post-transformation Lagrangians are identical. By this field transformation to a petrurbative minimum, the mass term of the Lagrangian is revealed at the cost of manually choosing one of the many possible minima which gives the name of the mechanism, Spontaneous Symmetry Breaking. It will now be demonstrated how the choice of an appropriate gauge eliminates three of the four real valued field components due to the $SU(2)\times U(1)$ symmetry. From the two complex components ϕ_1, ϕ_2, let, around the minimum:

$$\begin{aligned}\Phi_{\min,\,\text{expand}} = \begin{pmatrix} \phi_{1\text{expand}} \\ \frac{v}{\sqrt{2}} + \phi_{2\text{expand}} \end{pmatrix} = \begin{pmatrix} \phi_{1r} + i\phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + i\phi_{2c} \end{pmatrix} &= e^{i\theta_i\sigma_i/2} e^{i\beta I/2} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} 1 + \frac{i}{2}\theta_3 + i\frac{\beta}{2} & i\frac{1}{2}\theta_1 + \frac{1}{2}\theta_2 \\ i\frac{1}{2}\theta_1 - \frac{1}{2}\theta_2 & 1 - i\frac{1}{2}\theta_3 + i\frac{\beta}{2} \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}\end{aligned}$$

where $\theta_1 = \theta_2 = 0$ and $\theta_3 = \beta$. Out of the four parameters under a $SU(2) \times U(1)$ transformation, the vacuum is invariant if there exists one massless gauge boson. The other three are eliminated from this process and instead gain mass by using the expanded vacuum. To make the masses apparent, the covariant derivative D_μ is written for the $SU(2) \times U(1)$ symmetry acting on the chosen minimum (where Y is the hypercharge):

$$D_\mu \Phi_{\min} = \partial_\mu - ig\frac{\tau^\alpha}{2} W_\mu^\alpha - ig' B_\mu Y \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}$$

The masses of the remaining three mediators come from the kinetic term $\left|D_\mu \Phi_{\min}\right|^2$. The evaluation of the expression yields:

$$\frac{1}{2}\left(\frac{gv}{2}\right)^2 \left(\left(W_\mu^1\right)^2 + \left(W_\mu^2\right)^2\right) = \left(\frac{gv}{2}\right)^2 W_\mu^+ W^{\mu-} = m_W^2 W_\mu^+ W^{\mu-}, \quad W^\pm = \frac{1}{\sqrt{2}}\left(W_\mu^1 \pm W_\mu^2\right)$$

and

$$\frac{v^2}{8}\left(-gW_\mu^3 + g'B_\mu\right)^2 = \frac{(g^2 + g'^2)v^2}{4}\left(Z^\mu\right)^2 + 0 \cdot \left(A^\mu\right)^2$$

where Z_μ and A_μ are the orthogonal fields:

$$Z_\mu = \frac{g}{\sqrt{g^2+g'^2}} W_\mu^3 - \frac{g'}{\sqrt{g^2+g'^2}} B_\mu, \qquad A_\mu = \frac{g'}{\sqrt{g^2+g'^2}} A_\mu^3 + \frac{g}{\sqrt{g^2+g'^2}} B_\mu$$

It has become evident that in place of the $SU(2)\times U(1)$ symmetry, the four spin 1 mediator fields mix into a new linear combination of the aforementioned mediators. During this

process, given the choice of the minimum, which was necessary for the evaluation of the expressions, three of these fields acquire mass. All left chiral fermions are grouped into doublets, each containing the two members of each generation, for a total of six doublets, three for the leptons and three for the quarks. T^3 follows the conventional isospin choice of $+1/2$ for the first entry of the doublet, and $-1/2$ for the second.

$$\begin{pmatrix} e_i \\ \nu_{e_i} \end{pmatrix}_L, \qquad \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L$$

The right chiral fermions are grouped into singlets, since they are unaffected by the Weak Interaction and all have $T^3 = 0$:

$$\left(e_i\right)_R, \quad \left(u_i\right)_R, \quad \left(d_i\right)_R$$

Only the first generation was shown in the cases regarding the chiral charge above. The following relation between electric charge Q, the third weak isospin component T^3 and hypercharge Y is introduced:

$$Q = T^3 + \frac{1}{2}Y$$

The Hypercharge Y describes the current interactions given the values of the charges and weak isospins of the particles participating in the interaction. This description of Electroweak theory is known as the *Glashow-Weinberg-Salam model*.

2.7.1 Fermion Mass Acquisition

In the Glashow-Weinberg-Salam model, the masses in the Lagrangian are of the form $-m_f\left(f_R^\dagger f_L + f_L^\dagger f_R\right)$, linking the two chiral components of a fermion field. However, the inclusion of such terms without the process of Symmetry breaking is forbidden due to the gauge group. The correct, general choice is the expression[12]:

$$-y_e^{ij} {L_a}^\dagger i\phi_\alpha e_R^j + y_d^{ij} Q_\alpha^{\dagger i}\phi_\alpha d_R^j - y_u^{ij} Q_\alpha^{\dagger i}\epsilon_{\alpha b}\phi_b^* u_R^j + \text{h.c.}$$

where the complex 3×3 matrices are called the Yukawa matrices. The L, Q doublets are the left chiral lepton and quark doublets, e, d, u are the right chiral lepton and quark doublets ϕ is the Higgs field doublet and $\epsilon_{\alpha\beta}$ is the two-dimensional antisymmetric symbol with $\epsilon_{12} = -\epsilon_{21} = -1$ and 0 otherwise. It is important to note that Gauge Invariance requires that the gauge couplings of the fermions of the three generations are *absolutely identical*. This leads to the concept of Lepton Flavor Universality, which is discussed in section 3.4. This structure is simplified by diagonalizing the non-Hermitian Yukawa matrices and making changes of variables. It should be noted that in the up and down quark fields, the Yukawa couplings combine into a unitary matrix that cannot be canceled in the $W^\pm$ interaction terms. This unitary matrix is the V_{CKM} matrix, described in section 2.6.1.

2.8 The Standard Model

The Lagrangian[40] obeying the $SU(3)\times SU(2)\times U(1)$ Gauge Symmetry after Spontaneous Symmetry breaking is reduced to:

$$\mathcal{L} = -\frac{1}{4}\sum_\alpha F_{\mu\nu}^\alpha F^{\alpha\,\mu\nu} + m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu + \sum_f \psi_f(i\gamma^m u D_{f\mu} - m_f)\psi_f + \frac{1}{2}(\partial_\mu h)^2 - V(h) \quad (2.6)$$

[12] h.c. refers to the Hermitian Conjugate terms. An expression $A + B + \text{h.c.}$ is equivalent to $A + B + A^\dagger + B^\dagger$. Hermitian Conjugate terms are often not written explicitly to prevent visual clutter.

where the sum over α runs over the generators of $SU(3) \times SU(2) \times U(1)$ and the sum over f runs over quark and lepton flavors. The covariant derivatives $D_{f\mu}$:

$$D_{f\mu} = \partial_\mu - ieQ_f A_\mu - i\frac{g}{cos\theta_W} Q_{Zf} Z_\mu - ig_s A_\mu^\alpha t^\alpha$$

where Q_f is the charge of the fermion particle in units of e. The interactions of the Higgs boson field $h(x)$ are generated by replacing $v \rightarrow v+h(x)$ in the mass terms for the $W^\pm$, Z and fermions. This Lagrangian describes the Standard Model of Particle Physics. All terms except for the couplings of the $W^\pm$, Z conserve the Charge Conjugation $\hat{C}$, Parity $\hat{P}$ and Time Reversal $\hat{T}$ symmetries. Moreover, QED and QCD conserve all three individually.

3 Elements of b-Physics Theory

This section contains the necessary theoretical elements for the analysis of this thesis. The object of study is the K_S^0 strange meson, under the decay $b \to s\mu\mu$. Specifically, the decays:

$$B^0 \to \mu^+\mu^- K_S^0 \qquad \text{Rare Mode}$$
$$B^0 \to J/\psi(\mu\mu)K_S^0 \qquad \text{Resonant Mode}$$
$$B^0 \to \psi(2S)(\mu\mu)K_S^0 \qquad \text{Second Resonant Mode}$$

are of primary interest in the analysis. Next to the decays, their associations are showcased. Each of the aforementioned decays fall under the process $b \to sl^+l^-$, where $l = \mu$ only. The referenced *Rare Mode* is the decay under study. The former of the two will be used in order to validate results in the analysis.

Particle	B^0	K_S^0	J/ψ	$\psi(2S)$	μ
Mass (MeV)	5279.72	497.611	3096.900	3686.097	105.6583755
Uncertainty (MeV)	0.08	0.013	0.006	0.011	0.0000023

Table 3.1: Masses of important particles in the analysis[19]. Uncertainties are symmetrical.

The field of b-physics regards decays and searches for New Physics (NP) using the b quark. Searches include decays induced by QCD, QED and the Weak Interaction. Flavor changing processes refer to decay processes where a quark flavor is changed by emitting some additional particles, for example $b \to sX$. Flavor Changing Charged Current (FCCC) processes are permitted in first order (tree-level Feynman diagrams, figure 3.2), whereas Flavor Changing Neutral Current (FCNC) processes are forbidden in the Standard Model as there is no photon/Z^0 coupling between, for example, the b and the s quark. FCNC processes can be described by an approximate Effective Field Theory as point-like interactions by integrating out heavy particle degrees of freedom such as $t, W^\pm, Z$ in the one-order loop penguin Feynman diagrams (figure 3.1). This thesis considers FCNCs in the Weak Interaction and focus on this section on b-physics. The penguin-induced FCNC transition $b \to sX$ is among the most valuable probe of New Physics in flavor physics due to the presence of very massive particles ($t, W^\pm, Z$) within the loop, making these processes very sensitive to non-SM extensions. Therefore, a detailed theoretical description will be in section 3.2.

Electroweak Penguin Diagrams Penguin diagrams describe processes where a quark emits and re-absorbs a $W^\pm$ boson changing flavor two times, typically $b \to t \to q_2$, where $q_2 \in \{s, d\}$. $b \to sl^+l^-$ proceeds via three penguin diagrams: $b \to s\gamma^*(l^+l^-)$, $b \to sZ^*(l^+l^-)$, and the box diagram as shown in figure 3.1. The $W^{*\pm}, Z^*, \gamma^*$ are intermediate virtual states that cannot be measured by the experiment. In the penguin diagrams, they are denoted with dashed lines. This process is sensitive to the sign of the C_7, C_9 and C_{10} Wilson coefficients, which will be discussed later. Depending on how New Physics interacts with the Wilson coefficients, their magnitude and sign change, leading to the change of the rate of $b \to sl^+l^-$. The motivations of studying b-physics by using an Effective Field Theory will now be discussed.

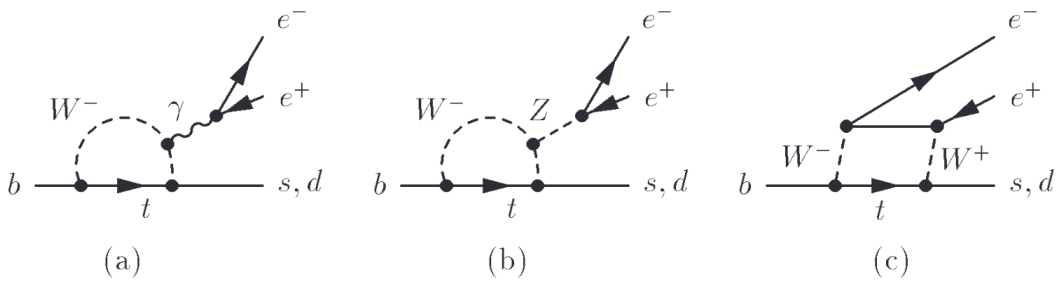


Figure 3.1: (a) photonic penguin (b) Z penguin and (c) box diagrams for the FCNC electroweak decay $b \to sl^+l^-$. The penguins are true for any l, not just e. If $l = \nu_l$, penguin diagram (a) does not contribute[42].

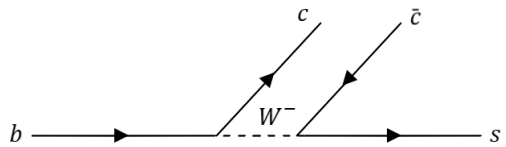


Figure 3.2: Tree level Feynman Diagram for the FCCC process $b \to sc\bar{c}$. The $c\bar{c}$ bounded meson system can most commonly be the particle J/ψ and $\psi(2S)$. This is a tree level diagram of $b \to sc\bar{c}$.

3.1 Motivation

As has been mentioned, penguin diagrams can give hints of New Physics at high energies since loops are very sensitive to heavy particles, such as $t, H, W^{\pm}, Z$ in the Standard Model. New contributions from new hypothetical particles from non-SM extensions can be present, such as charged Higgs, Supersymmetric (SUSY) particles or fourth generation leptons and quarks[42]. More specifically, effects from New Physics may come from:

- Models with Z^0 mediated FCNC can enhance various decay channels such as $B^0 \to l^+l^-$ by two orders of magnitude over the SM value and $B_s^0 \to l^+l^-$ by one order of magnitude. Furthermore, the electro-weak penguin dominated decays $B^+ \to \phi\pi^+$ and $B_s^0 \to \phi\pi^0$ are increased by two and one order of magnitude respectively without violating the current limits on $b \to sl^+l^-$[42]. In such models, one introduces an additional exotic vector singlet charge -1/3 quark and allows it to mix with the ordinary down-type quarks[43]. FCNCs involving the Z are induced due to the different weak isospin of the exotic quark. The CKM matrix ceases to become unitary, affecting $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ mixing, allowing the study of the mixing also explore New Physics.

- Multi-Higgs doublet models can enhance the rates of di-lepton processes such as $B^0 \to l^+l^-$ without considerably affecting $b \to sl^+l^-$ where in this case $l = \{e, \mu\}$. These models can be classified into two types[43]: models in which there are no FCNC and models in which FC interactions can be mediated by neutral scalars. In the former type of models, the phase of $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ mixing is unaffected, the CKM matrix continues to be unitary. In the latter type of models, $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ mixing is affected from tree-level neutral Higgs exchange amplitudes carrying new phases.

- In Supersymmetric models, one of which was studied in my Undergraduate thesis[44], contributions from charged Higgs sparticles can be cancelled by contributions from

charginos and gluinos, leaving the process $b \to s\gamma$ unaffected, but increasing the rate of $b \to sl^+l^-$. Supersymmetric Standard Models (SSMs) leave the gauge group unchanged but introduce and assign new particle partners to each existing Standard Model one. There are multiple frameworks of Supersymmetry (SUSY), and the simplest and most commonly known one is the Minimal Supersymmetric Standard Model (MSSM). Supersymmetry introduces many heavy particles in order to solve the Hierarchy problem, Grand Unification and dark matter. Such particles may appear and contribute in the penguin loops.

- Left-right symmetric models[43] typically assign a new gauge group to the SM, $SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times SU(2)_{\mathrm{R}} \times U(1)_{\bar{Y}}$ along with a discrete $L \leftrightarrow R$ symmetry. The consequence of this change in the group gauge is to introduce the right-handed neutrinos and a different definition of hypercharge. The new right-handed W_R boson participates in weak processes in the same way as the Standard Model $W^{\pm}$, can contribute to $B^0 - \bar{B}^0$ and $B_s^0 - \bar{B}_s^0$ mixing. Limits from the mass difference between K_L^0 and K_S^0 constrain the new hypothetical particle W_R to be heavier than 1.4TeV, potentially rendering its effects in the mixing negligible. The $L \leftrightarrow R$ symmetry relates a right-handed CKM matrix to a left-handed counterpart. Removing this symmetry can lead to the K_L^0 and K_S^0 mass difference constraints to vanish. Then, the effect of the B^0 and B_s^0 lifetime can be attributed to heaviness of the W_R and not through the CKM matrix.

Each theoretically proposed model affects $b \to X$ processes differently and measurements of multiple decay channels is an integral part for search of New Physics. What's more, apart from the search of New Physics, penguin diagrams can measure the parameters of the CKM matrix and are important in measuring the CP violation[45].

3.2 Penguin Diagram EFT

Flavor changing processes, such as decays of the form $b \to s\mu^+\mu^-$, occur at low energies, at scales $\mu \ll M_W$. Such decays provide an important test of possible new physics at the electroweak scale[46]. It is convenient to pass from the full theory of Electroweak Interactions to an Effective Field Theory (EFT) to perform theoretical calculations with ease, in order to compare with experimental data at those scales.

At the energy scales of $\mu \sim M_W \sim 80GeV$, $b \to s$ quark decays are governed by the aforementioned penguin Feynman Diagrams. To obtain an EFT at the scale $\mu \sim m_b \sim 5GeV$, the heavy particle degrees of freedom $(t, W^{\pm}, Z)$ must be integrated out to reduce the loops into point-like interactions. The EFT for this energy scale is a generalization of the Fermi EFT for beta decays[42]. After the integration, the particle degrees of freedom do not appear explicitly in the theory but are incorporated in the effective gauge coupling constants, running masses and the Wilson Coefficients C_i. The generalized EFT Lagrangian has the following form[46]:

$$\begin{aligned}\mathcal{L}_{\text{eff}} &= \mathcal{L}_{\text{QCD}\times\text{QED}}(u,d,s,c,b,e,\mu,\tau) \\ &+ \frac{4G_F}{\sqrt{2}}\left(V_{us}^* V_{ub}\left(C_1^c P_1^u + C_2^c P_2^u\right) + V_{cs}^* V_{cb}\left(C_1^c P_1^c + C_2^c P_2^c\right)\right) \\ &+ \frac{4G_F}{\sqrt{2}}\sum_{i=3}^{10}\left(\left(V_{us}^* V_{ub} + V_{cs}^* V_{cb}\right) C_i^c + V_{ts}^* V_{tb} C_i^t\right) P_i\end{aligned}$$

where V_{ij} are elements of the V_{CKM} matrix, $G_F = 1.166379 \cdot 10^{-5} GeV^{-2}$ is the Fermi constant and $\mathcal{L}_{\text{QCD}\times\text{QED}}$ consists of kinetic terms of light fermions plus QED and QCD interactions.

This Lagrangian density has been described in sections 2.4 and 2.5 respectively. The remaining terms represent $\Delta B = -\Delta S = 1$ transitions and consist of local operators of dimension ≤ 6. The form of the operators is:

$$P_1^u = (\bar{s}_L \gamma_\mu T^\alpha u_L)(\bar{u}_L \gamma^\mu T^\alpha b_L) \qquad P_2^u = (\bar{s}_L \gamma_\mu u_L)(\bar{u}_L \gamma^\mu b_L)$$
$$P_1^c = (\bar{s}_L \gamma_\mu T^\alpha c_L)(\bar{c}_L \gamma^\mu T^\alpha b_L) \qquad P_2^c = (\bar{s}_L \gamma_\mu c_L)(\bar{c}_L \gamma^\mu b_L)$$
$$P_3 = (\bar{s}_L \gamma_\mu c_L) \sum_q (\bar{q} \gamma^\mu q)$$
$$P_4 = \left(\bar{s}_L \gamma_\mu T^\alpha c_L\right) \sum_q (\bar{q} \gamma^\mu T^\alpha q)$$
$$P_5 = \left(\bar{s}_L \gamma_{\mu 1} \gamma_{\mu 2} \gamma_{\mu 3} b_L\right) \sum_q \left(\bar{q} \gamma^{\mu 1} \gamma^{\mu 2} \gamma^{\mu 3} q\right)$$
$$P_6 = \left(\bar{s}_L \gamma_{\mu 1} \gamma_{\mu 2} \gamma_{\mu 3} T^\alpha b_L\right) \sum_q \left(\bar{q} \gamma^{\mu 1} \gamma^{\mu 2} \gamma^{\mu 3} T^\alpha q\right)$$
$$P_7 = \frac{e}{g^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu} \qquad P_8 = \frac{1}{g} m_b (\bar{s}_L \sigma^{\mu\nu} T^\alpha b_R) G^\alpha_{\mu\nu}$$
$$P_9 = \frac{e^2}{g^2} \left(\bar{s}_L \gamma_\mu b_L\right) \sum_l (\bar{l} \gamma^\mu l) \qquad P_{10} = \frac{e^2}{g^2} \left(\bar{s}_L \gamma_\mu b_L\right) \sum_l (\bar{l} \gamma^\mu \gamma_5 l)$$

where the sums over q and l denote sums over all the light quarks (u, d, s, c, b) and leptons (e, μ, t) respectively. The opposite chirality $L \leftrightarrow R$ local light operators exist and are denoted in primed notation[13]. $F_{\mu\nu}$ and $G^\alpha_{\mu\nu}$ are the QED and QCD field strength tensors respectively and T^α are the eight generators of the symmetry group of QCD. m_b, e and $g = 4\pi a_s$ are the mass of the b quark, electron charge and QCD coupling strength respectively. The light (with respect to the b quark mass fermions and the W boson mass) operators is due to the fact that Effective Field Theories are first expressed in an energy scale energy renormalization limit μ, in this case, $\mu \sim m_W$, then evolved down to $\mu \sim m_b$ using Renormalization Theory techniques. The Wilson coefficients are essentially the coupling strengths of the local light field operators and encode information about New Physics (NP) as well as heavy particle degrees of freedom. They are expanded pertrubatively as[46][47]:

$$C_i^Q = C_i^{Q(0)} + \frac{a_s}{4\pi} C_i^{Q(1)} \left(\frac{a_s}{4\pi}\right)^2 C_i^{Q(2)} + O(a_s^3), \quad Q \in \{c, t\}$$

where $16\pi^2/g^2 = 4\pi/a_s(m_W)$. At tree level, all zero-th order Wilson Coefficients $C_i^{Q(0)}$ vanish, except for $C_2^{c(0)} = -1$. For more information about the procedure, as well the expression of higher-order terms, see[46]. An alternative formulation using Hamiltonians can be found in[47]. From the forms of the local light operators P_i, it can be seen due to the form of the operators P_9, P_{10} containing light lepton fields, that these two operators contribute in $b \to s l^+ l^-$ decays. P_9 is a vector current operator, whereas P_{10} is an axial vector current operator due to the presence of the γ_5 matrix[14]. Once the matching renormalization scale has been evolved from $\mu \sim m_W$ down to the effective region $\mu \sim m_b$, the Wilson coefficients mix with each other[47]. It is of importance to note the new effective form of C_9:

$$C_9^{\text{eff}} = \frac{4\pi}{a_s} C_9 + Y(q^2)$$

[13] See[47] for more details.

[14] Some authors follow the P_{9V} and P_{10A} notation to denote this. The field operators are also sometimes written as O_i in literature.

where $Y(q^2)$:

$$\begin{aligned} Y(q^2) = & \, h(q^2, m_c)\left(\frac{4}{3}C_1 + C_2 + 6C_3 + 60C_5\right) \\ & - \frac{1}{2}h(q^2, m_b)\left(7C_3 + \frac{4}{3}C_4 + 76C_5 + \frac{64}{3}C_6\right) \\ & - \frac{1}{2}h(q^2, 0)\left(C_3 + \frac{4}{3}C_4 + 16C_5 + \frac{64}{3}C_6\right) + \frac{4}{3}C_3 + \frac{64}{9}C_5 + \frac{64}{27}C_6 \end{aligned}$$

and $h(q^2, m_q)$:

$$h(q^2, m_q) = -\frac{4}{9}\left(\ln\frac{m_q^2}{\mu^2} - \frac{2}{3} - z\right) - \frac{4}{9}(2+z)\sqrt{|z-1|}\,\operatorname{atan}\frac{1}{\sqrt{z-1}}, \qquad z > 1$$

$$h(q^2, m_q) = -\frac{4}{9}\left(\ln\frac{m_q^2}{\mu^2} - \frac{2}{3} - z\right) - \frac{4}{9}(2+z)\sqrt{|z-1|}\,\ln\frac{1+\sqrt{1-z}}{\sqrt{z}} \qquad z \leq 1$$

where $z = 4m_q^2/q^2$, related to the basic fermion loop and μ is the matching scale of the interaction. The function h is discontinuous for $z = 1$ or equivalently $q^2 = 4m_q^2$. Such discontinuities can cause issues in the analysis[48] but corrections are not taken into account in Monte Carlo simulations and ad-hoc models are not included in this thesis. The operators P_i are used to extract the form factors, or matrix elements of a particular decay.

Effective Hamiltonian for $b \to sll$ The Effective Hamiltonian for $b \to s$ penguin decays has the following form:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}}V_{tb}V_{ts}^*\sum_{i=1}^{10}C_i(\mu)P_i(\mu)$$

where μ is the energy scale. Calculations are performed at a high energy scale $\mu \sim M_W$ and then evolved to a low energy scale $\mu \sim m_b$ using renormalization group equations. The evolution to the low energy scale mixes the operators and the dependence on the μ energy scale is canceled between the Wilson coefficients C_i and the light operators P_i, making any observable quantity independent of the normalization energy scale parameter μ[15]. Since the calculations are performed perturbatively, however, the theoretical predictions depend on the renormalization scale μ and can sometimes dominate over the theoretical uncertainty. To minimize this effect, calculations must be performed in higher orders. Wilson coefficients represent short distance (high-energy) Electroweak and Strong Interactions. The operator elements, on the other hand, are influenced by long distance (low energy) effects therefore they cannot be obtained pertrubatively due to the quark confinement phenomenon of QCD. The operator elements $\langle X|\, P_i \,|B\rangle$ where X denote the final state products (e.g. $\mu\mu K_S^0$) can be approximated to $\langle X|\, P_i \,|B\rangle \approx \langle s|\, P_i \,|b\rangle$ to leading order using expansions of powers of $1/m_b$ when X is an inclusive[16] final state[42]. When X is an exclusive final state, the calculation of the operator elements is difficult from first principles, however some methods can succeed depending on the presence of hadrons, leptons and the number of particles in the final state[42]. The simplest allowed final states are $B \to Kl^+l^-$ and $B \to K^*l^+l^-$, where B, K, K^* can be either charged or neutral. These final states constitute about $\sim 10\%$ of the total rate of $b \to sl^+l^-$.

[15] Any physical observable does not depend on the renormalization scale which is a number which allows theoretical simplifications in Renormalization Theory

[16] An exclusive/inclusive final state denotes that the momentum 4-vector of all/some of the decay products has been determined.

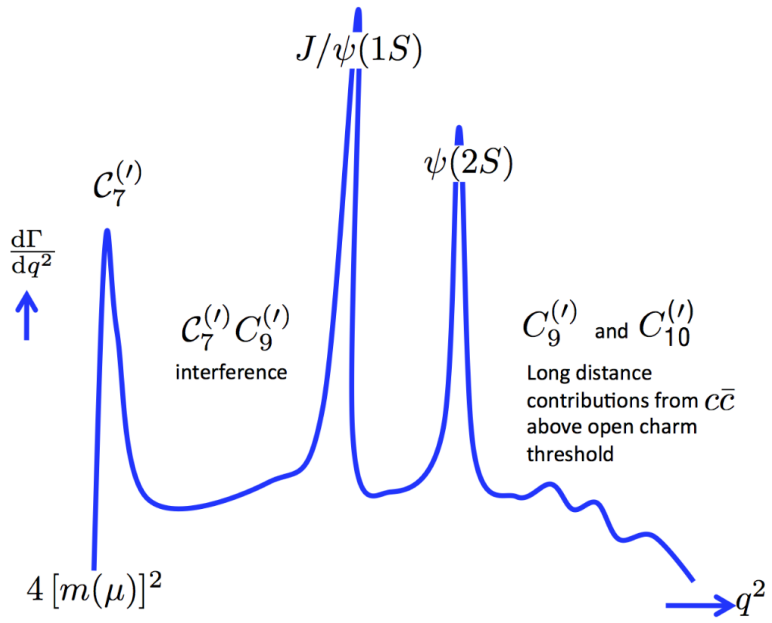


Figure 3.3: Differential decay rate in $q^2 = m(\mu\mu)$ for $b \to s\mu^+\mu^-$ decays. The figure qualitatively shows the two predominant resonances as well as the dominant contributions of the effective (primed) Wilson coefficients in different q^2 regions.

3.3 $K^0 - \bar{K}^0$ Mixing

Some mesons, like the $K^0, \bar{K}^0$ can convert into each other respectively [49] due to the lack of a conserved quantum number[17]. The $K^0, \bar{K}^0$ have definite quark content $d\bar{s}, s\bar{d}$ which have strangeness $S = +1, S = -1$ respectively, constituting them as definite particles participating in QCD interactions. Moreover, the pair has spin 0 and negative parity $P = -1$. Flavor Changing Weak Interactions do not conserve the Strangeness quantum number, allowing the particle pair to mix with itself. This phenomenon is called $K^0 - \bar{K}^0$ *mixing*. The charge conjugation operator $\hat{C}$ and the parity operator $\hat{P}$ can change the particles of the pair into each other, and the parity operator introduces an eigenvalue of -1 since the particles in the pair have negative parity.

$$\hat{C}\left|K^0\right\rangle = \left|\bar{K}^0\right\rangle, \quad \hat{C}\left|\bar{K}^0\right\rangle = \left|K^0\right\rangle$$
$$\hat{P}\left|K^0\right\rangle = -\left|K^0\right\rangle, \quad \hat{P}\left|\bar{K}^0\right\rangle = -\left|\bar{K}^0\right\rangle$$

Applying the two operators $\hat{C}$ and $\hat{P}$, one sees:

$$\hat{C}\hat{P}\left|K^0\right\rangle = -\left|\bar{K}^0\right\rangle, \quad \hat{C}\hat{P}\left|\bar{K}^0\right\rangle = -\left|K^0\right\rangle$$

demonstrating that the pair $K^0, \bar{K}^0$ is not a CP eigenstate pair. Moreover, two types of neutral kaons K_S and K_L[18], different from the aforementioned $K^0, \bar{K}^0$ are measured experimentally, with approximately the same mass of $498MeV$ but different decay modes and lifetimes of $\sim 9 \cdot 10^{-11}s$ and $\sim 5 \cdot 10^{-8}s$ respectively. The K_S decay modes are primarily:

$$K_S^0 \to \pi^0\pi^0\ (B = 31)\%, \qquad K_S^0 \to \pi^+\pi^-\ (B = 69)\%$$

containing two particles in the final state. For the K_L^0, three particles are identified:

$$K_L^0 \to \pi^0\pi^0\pi^0\ (B = 20)\%, \qquad K_L^0 \to \pi^0\pi^+\pi^-\ (B = 13)\% \qquad K_L^0 \to \pi^\pm l^\mp \nu_l(\bar{\nu}_l)\ (B = 67)\%$$

[17] For example, a π^+ cannot turn into a π^- due to the conservation of charge, a p^+ cannot turn into a n^0 due to the conservation of charge and baryon number and the π^0 is a Majorana particle.

[18] Pronounced K-short and K-Long respectively. This naming convention is due to the magnitude of their lifetimes, where the K_L's lifetime is larger.

However, it has been discovered[50] that the K_L^0 can decay into $\pi^+\pi^-$ with a $B \sim 0.1\%$, which is another experimental evidence of CP violation of the Weak Interaction. Therefore, the K_S^0 and K_L^0 particles are definitely not the definite QCD particles $K^0, \bar{K}^0$, or the $\hat{C}\hat{P}$ eigenstates K_1^0, K_2^0 but a specific mixture[49]. The mixture is defined with the help of the $\hat{C}\hat{P}$ eigenstates K_1^0, K_2^0 of the $K^0, \bar{K}^0$:

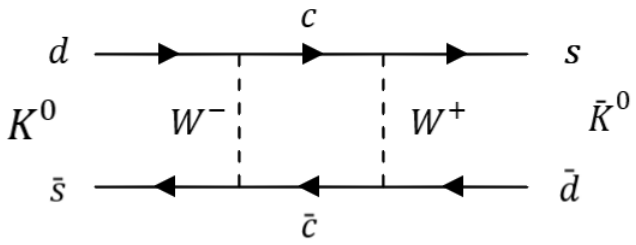


Figure 3.4: Feynman Diagram demonstrating $K^0 - \bar{K}^0$ mixing.

$$|K_1^0\rangle = \frac{1}{\sqrt{2}}\Big(|K^0\rangle - |\bar{K}^0\rangle\Big), \quad \hat{C}\hat{P}\,|K_1^0\rangle = |K_1^0\rangle$$
$$|K_2^0\rangle = \frac{1}{\sqrt{2}}\Big(|K^0\rangle + |\bar{K}^0\rangle\Big), \quad \hat{C}\hat{P}\,|K_2^0\rangle = -|K_2^0\rangle$$

where the K_S^0 and K_L^0 are expressed via a complex parameter ϵ, which can be determined experimentally $|\epsilon| \approx (2.232 \pm 0.007) \cdot 10^{-3}$[50]:

$$|K_S^0\rangle = \frac{1}{\sqrt{1+|\epsilon|^2}}\Big(|K_1^0\rangle - \epsilon\,|K_2^0\rangle\Big)$$

$$|K_L^0\rangle = \frac{1}{\sqrt{1+|\epsilon|^2}}\Big(\epsilon\,|K_1^0\rangle + |K_2^0\rangle\Big)$$

Other mesons mix the way the neutral kaons do, for example the $B^0, \bar{B}^0$ pair with quark content $d\bar{b}, b\bar{d}$, the $B_s^0, \bar{B}_s^0$ pair with quark content $s\bar{b}, s\bar{d}$, as well as the $D^0, \bar{D}^0$ pair.

3.4 Lepton Flavor Universality

Lepton Flavor Universality (LFU) refers to the assumption that all three lepton flavors have identical couplings to the three massive Electroweak bosons (Z^0, $W^\pm$)[51]. Testing the validity of the LFU is a direct test on the validity of the Standard model, as well as hint towards new NP scenarios as LFU can be violated in such scenarios, due to the higher mass of the third generation of leptons (τ, ν_τ) coupling to NP heavier particles much easier than lighter leptons. The LFU in the Standard Model comes from the absence of proportional matrices (such is the V_{CKM} for the quarks) in the Weak Interaction bosons between the fermions, see section 2.7.1.

3.4.1 The R_K Observable

In $b \to s l^+ l^-$ processes, an observable that may hint towards LFU violation is the R_K family of ratios:

$$R_K = \frac{\text{BR}(B \to \mu^+\mu^- K)}{\text{BR}(B \to e^+e^- K)}$$

where:

$$B \in \{B^{\pm}, B^0, B_S^0\}, \quad K \in \{K^{\pm}, K^{*\pm}, K_S^0, K_L^0, K^{*0}\}$$

In order to prevent experimental uncertainties stemming from the different nature of measuring μ and e within the CMS experiment, such as bremstrahlung radiation and trigger reconstruction efficiencies, the above measurable is modified in such a way that expresses a double ratio of the two leptonic modes by including a resonant channel for the leptons, in this case the J/ψ, so that $B \to J/\psi K$. The aforementioned uncertainties in the lepton pairs largely cancel and the measurement of the R_K ratio is not dominated by large uncertainties. The double ratio is expressed by:

$$R_K = \frac{\mathrm{BR}(B \to \mu^+\mu^- K)}{\mathrm{BR}(B \to J/\psi(\mu^+\mu^-)K)} \Big/ \frac{\mathrm{BR}(B \to e^+e^- K)}{\mathrm{BR}(B \to J/\psi(e^+e^-)K)}$$

It is convenient to write the above as:

$$R_K = R_K(\mu)/R_K(e)$$

The decay channel of $J/\psi \to l^+l^-$, where $l \in \{e, \mu\}$ is largely studied. In this thesis, the choice of the K particle is K_S^0. Furthermore, since the reconstruction efficiency of electrons is a rather complicated task, in this thesis only the ratio $R_K(\mu)$ is calculated. Therefore, the object of study within the scope of this thesis is explicitly:

$$R_{K_S^0}(\mu) = \frac{\mathrm{BR}(B^0 \to \mu^+\mu^- K_S^0)}{\mathrm{BR}(B^0 \to J/\psi(\mu^+\mu^-)K_S^0)}$$

To determine the $R_{K_S^0}(\mu)$ observable experimentally, one needs to account for detector geometry and particle reconstruction effects. The yield of the decay channel $B^0 \to X$ $N(B^0 \to X)_{\mathrm{exp}}$ is determined as:

$$N(B^0 \to X)_{\mathrm{exp}} = N(B^0) \cdot BR(B^0 \to X) \cdot \alpha(B^0 \to X) \cdot \epsilon(B^0 \to X) \tag{3.1}$$

where $N(B^0)$ is the total number of B^0 particles and $\alpha(B^0 \to X)$, $\epsilon(B^0 \to X)$ is the acceptance and efficiency factors, accounting for detector geometry and particle reconstruction, thoroughly explained in section 4.3.

$$N(B^0) \cdot \mathrm{BR}(B^0 \to X) = \frac{N(B^0 \to X)_{\mathrm{exp}}}{\alpha(B^0 \to X) \cdot \epsilon(B^0 \to X)}$$

Taking the expression above and then dividing it by a similar expression but with a decay channel with the same mother particle B^0 so that the total number of mother particles, or total yield $N(B^0)$ can cancel, one can arrive into the experimental determination of the $R_{K_S^0}(\mu)$ ratio. By taking equation 3.1 for $X = \mu^+\mu^- K_S^0$ and the resonant channel $X = J/\psi(\mu\mu)K_S^0$, where $J/\psi(\mu\mu)$ is an abbreviation of the decay channel $J/\psi \to \mu^+\mu^-$ and dividing by parts:

$$R_{K_S^0}(\mu) = \frac{N(B^0 \to \mu^+\mu^- K_S^0)_{\mathrm{exp}}}{\alpha(B^0 \to \mu^+\mu^- K_S^0) \cdot \epsilon(B^0 \to \mu^+\mu^- K_S^0)} \Big/ \frac{N(B^0 \to J/\psi(\mu\mu)K_S^0)_{\mathrm{exp}}}{\alpha(B^0 \to J/\psi(\mu\mu)K_S^0) \cdot \epsilon(B^0 \to J/\psi(\mu\mu)K_S^0)}$$

Or equivalently, by re-arranging:

$$R_{K_S^0}(\mu) = \frac{N(B^0 \to \mu^+\mu^- K_S^0)_{\mathrm{exp}} \cdot \alpha(B^0 \to J/\psi(\mu\mu)K_S^0) \cdot \epsilon(B^0 \to J/\psi(\mu\mu)K_S^0)}{N(B^0 \to J/\psi(\mu\mu)K_S^0)_{\mathrm{exp}} \cdot \alpha(B^0 \to \mu^+\mu^- K_S^0) \cdot \epsilon(B^0 \to \mu^+\mu^- K_S^0)} \tag{3.2}$$

Later on, as a cross-check the determination of the same observable for the Second Resonant mode $B^0 \to \psi(2S)K_S^0$ in place of the Rare Mode will also be calculated.:

$$R_{K_S^0}(\psi(2S)) = \frac{N(B^0 \to \psi(2S)(\mu\mu)K_S^0)_{\text{exp}} \cdot \alpha(B^0 \to J/\psi(\mu\mu)K_S^0) \cdot \epsilon(B^0 \to J/\psi(\mu\mu)K_S^0)}{N(B^0 \to J/\psi(\mu\mu)K_S^0)_{\text{exp}} \cdot \alpha(B^0 \to \psi(2S)(\mu\mu)K_S^0) \cdot \epsilon(B^0 \to \psi(2S)(\mu\mu)K_S^0)} \tag{3.3}$$

The quantities above can be determined experimentally.

3.4.2 The Differential Branching Function dB/dq^2

The differential Branching function for a specific decay process relative to another can also be determined experimentally by taking equation 3.1 for two modes and dividing them by parts as done in equations 3.2 and 3.3. However, in one of the decay channels, the Branching ratio can be expressed in terms of q^2 as:

$$\text{BR}(B^0 \to sX) = \frac{\Delta\text{BR}(B^0 \to sX)}{\Delta q^2}\Delta q^2$$

Instead of measuring the $R_{K_S^0}(\mu)$ observable, the differential branching function of the Rare Mode may be measured instead for various q^2 regions. The differential branching function is given by:

$$\frac{d\text{BR}(B^0 \to \mu^+\mu^- K_S^0)}{dq^2} = \frac{N(B^0 \to \mu^+\mu^- K_S^0)_{\text{exp}} \cdot \alpha(B^0 \to J/\psi(\mu\mu)K_S^0) \cdot \epsilon(B^0 \to J/\psi(\mu\mu)K_S^0)}{N(B^0 \to J/\psi(\mu\mu)K_S^0)_{\text{exp}} \cdot \alpha(B^0 \to \mu^+\mu^- K_S^0) \cdot \epsilon(B^0 \to \mu^+\mu^- K_S^0)} \cdot \frac{\text{BR}(B^0 \to J/\psi(\mu^+\mu^-)K_S^0)}{\Delta q^2}$$

or equivalently:

$$\frac{d\text{BR}(B^0 \to \mu^+\mu^- K_S^0)}{dq^2} = R_{K_S^0}(\mu) \cdot \frac{\text{BR}(B^0 \to J/\psi(\mu^+\mu^-)K_S^0)}{\Delta q^2} \tag{3.4}$$

The switch $dq^2 \to \Delta q^2$ affects the quality of this analysis. Ideally, the bin sizes Δq^2 used in the analysis should be as small as possible in order to differentiate effects such as different decays spanning across a q^2 bin. However, the problem lies in the low number of statistics available. This requires the use of broader q^2 bins.

4 Analysis

This section of this thesis describes the Analysis procedure as a whole. Firstly, a description of the software and datasets are given in section 4.1 for the reader to familiarize with the framework used in the analysis. The analysis was mainly performed by using the ROOT framework, provided by CERN, to analyze both synthetic datasets from physics process generators as well as real data acquired from the CMS experiment. Following the aforementioned description, section 4.2.3 describes various physics objects that are present in the data samples which are essential for interpretation. Section 4.3 is the first essential step in determining the two observables of this analysis, which were described in section 3.4. Sections 4.4 and 4.5 give a thorough description of the BDT Machine Learning model and its training procedure for the analysis, which is also very essential in the analysis. Section 4.6 explains the procedure of the B^0 mass fits on the low-q^2 region after the BDT is trained in order to determine the value of the BDT cut which favors background rejection. Section 4.7 explains a technique which can extract variable distributions from a specific fit component of a fitted distribution, allowing the generated samples to be compared with the Signal search in the Data samples. Finally, section 4.8 describes the simultaneous fit on the B^0 mass across the Rare Mode bins in table 4.1, making use of increased statistics in order to determine the differential branching function observable. The results and summary of this analysis are in section 5. Figure 4.1 shows a qualitative graphic of the analysis strategy.

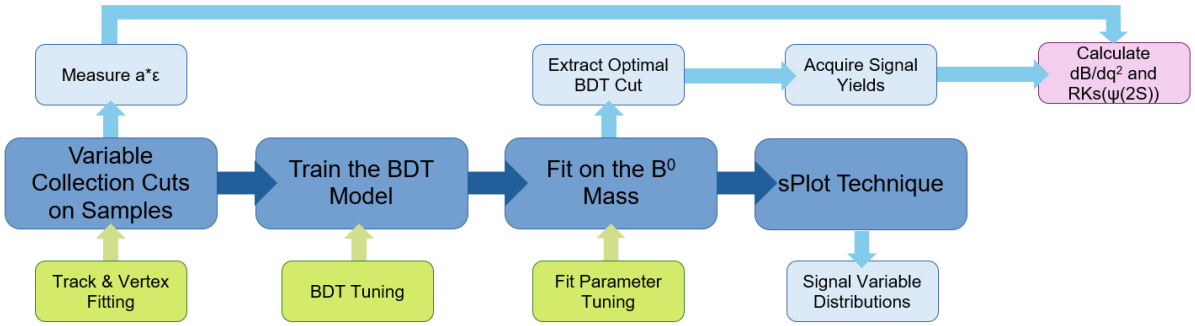


Figure 4.1: Qualitative analysis strategy graphic. The rectangular shapes in dark blue signify techniques and processes using various frameworks in the analysis. The green shapes signify refinement processes in order to improve the analysis. The light blue shapes signify results that are essential to the analysis. The final pink shape is the end result of this analysis.

4.1 Analysis Software

The analysis is done using software on an Linux CentOS able to run ROOT and various python packages. The aforementioned operating systems are hosted on the `lxplus7`, `lxplus8` and `lxplus9` nodes of the LXPLUS (Linux Public Login User Service) machine cluster. It is a collection of machines running Linux distributions where the analysis can be performed with considerable computing power while also providing a large amount of storage space. Each machine on the LXPLUS cluster runs on different versions of the operating system, with different versions of software. One limitation of the analysis was running the same software on different versions, therefore the software had to be adjusted

to prevent bugs and unpredictable behaviors. By the end of June 2024, `lxplus7` nodes were discontinued from operation, requiring further technical software maintenance so as to keep up with newer nodes.

4.1.1 Datasets

The available datasets[19] for analysis are separated into two distinct categories, *Monte Carlo samples* and *Data samples*. *Monte Carlo samples* are simulated samples that involve simulated physics processes, usually incorporating realistic detector simulation and measurement boundaries and created by generator software according to a set of Physics rules or theory (for example the Standard Model, or the MSSM). Within this thesis, if the Monte Carlo dataset incorporates detector simulation and/or includes a biased amount of a specific physics process or decay, it will be referred to as a *biased Monte Carlo sample*. These data samples must be requested and are not publicly available. The Monte Carlo samples for this thesis involve the following processes:

- Unbiased samples with $B^0 \to \mu\mu K^0_S$, $B^0 \to J/\psi(\mu\mu)K^0_S$ and $B^0 \to \psi(2S)(\mu\mu)K^0_S$ processes, as well as $B \to \mu\mu K^*(\pi K^0_S)$ background processes.
- Biased[20] samples with $B^0 \to \mu\mu K^0_S$, $B^0 \to J/\psi(\mu\mu)K^0_S$, $B^0 \to \psi(2S)(\mu\mu)K^0_S$, $B^0 \to \mu\mu K^*(K^0_S\pi)$ processes, as well as $B \to \mu\mu K^*(\pi K^0_S)$ background processes.

Data samples are experimental measurements acquired through the CMS detector, after some elementary preprocessing. This thesis regards and uses CMS Run 3 2022 Data samples. Specifically among the whole dataset, eras C, Dv1, Dv2, E, F, G and parts 0 to 7 were used. Each entry in the MC and Data samples corresponds to an event with a B^0 decay. These processed samples that are used in this analysis are also referred to as *tuples*.

4.1.2 PYTHIA

PYTHIA[52][53] has been used for the generation of events in high energy collisions. Version 8.2 of the program is written in C++ and covers a wide range of hard processes, models for parton showers and particle interactions and decays. It describes Physics that is derived from theory, based on phenomenological models, or options from yet to be confirmed models, like Supersymmetry.

The physics models of PYTHIA regard high-energy particle collisions with $\sqrt{s} > 10GeV$, corresponding to a proton-proton fixed target beam energy of $E \geq 50GeV$. This choice of energy cutoff is picked because PYTHIA incorporates the approximation of a continuum of final states to make cross section calculations easier. However, for $\sqrt{s} < 10GeV$, hadron-hadron cross section calculations become unreliable as the aforementioned approximation no longer holds. Furthermore, PYTHIA's models can be verified for $\sqrt{s} \sim 100TeV$, or otherwise $E < 10^{10}GeV$. Two limitations that are worth mentioning are the types of collisions between leptons and hadrons present in the program and the absence of detector material. PYTHIA works with hadron-hadron or lepton-lepton collisions at its newest version. There are no current plans to incorporate mixed hadron-lepton collisions, however, collisions including nuclei are not going to be programmed into PYTHIA in the

[19]The terms datasets and samples are used interchangeably. However, the term sample is more common within the Experimental High Energy Physics community.

[20]A biased MC sample is also referred to as a Filtered MC sample due to passing filters to the physics event generator that creates these samples.

foreseeable future. The second issue is the absence of detector material in the program's code. Luckily, an external interface can be created by a third party to simulate the desired detector material. PYTHIA event generation is also supported by other modules, such as MadGraph5. PYTHIA includes a large number of processes. The majority of processes included are reactions and scatterings of type $2 \to 1$, $2 \to 2$ and a few $2 \to 3$ and above. Processes that are widely known include:

- QCD soft and hard processes, for example jets.
- Electroweak processes, with photon production, production of the γ/Z and $W^{\pm}$ bosons. Inelastic scattering $\gamma\gamma \to f\bar{f}$ is also allowed.
- Top production in pairs or as a single final state.
- Various Higgs processes, including the SM production of the Higgs boson, as well as production of Higgs from Higgs-doublet models.
- SUSY processes, for example the pair production of SUSY particles.

The total proton-proton beam cross section is given by the Regge fitting to data as a function of the Mandelstam variable s in units of TeV^2:

$$\sigma_{\text{tot}}^{pp}(s) = (21.70s^{0.0808} + 56.08s^{-0.4525})mb$$

for $s = 13.5^2 TeV^2$, one gets the value for the proton-proton total cross section:

$$\sigma_{\text{tot}}^{pp}(13.5^2 TeV^2) = 100.917mb$$

which is consistent with the pp cross section of the LHC during Run 3 2022. Since PYTHIA runs in the C++ programming language, an object oriented programming language, it makes use of classes. Each event is generated and is described by the `Event` class, which is a wrapper for a vector of `Particles`, which is an array of all the particles of the event, where their properties can be accessed for analysis. For the `Particle` class, some of its key properties include:

- An assigned identification number (id), which is unique per particle type,
- its mother(s) particle(s),
- its color charge and electrical charge, Breit-Wigner width,
- its four-momentum, mass, spin, and lifetime on its CM frame,
- a four-vector representing the production vertex,
- whether the particle decay, or different types of statuses which describe the state of the particle.

Using this information, one can essentially perform analysis by hand-picking which particles are important, including their properties. This level of data access is referred to as the Generated (GEN) Level, where no attempts of a simulated physical detector to record and reconstruct particle tracks have been made. It is briefly reminded that this level does **not** represent the experimental reconstructed data that one gets from collision experiments such those that happen in the detectors of the LHC, as explained by the two dataset types in section 4.1.1. In short, the GEN level of the Monte Carlo sample is a dataset that corresponds to generated data by the PYTHIA software.

4.1.3 ROOT

ROOT[54] is an object oriented data analysis framework. It is used primarily for analyses in high-energy physics. The main components that will be discussed in this section include:

- a machine-independent, highly-compressed object-oriented database along with a C++ interpreter,
- advanced statistical analysis tools, namely fitting and minimization techniques (RooFit), multi-dimensional histogramming,
- visualization tools for 2D and 3D graphics,
- query mechanisms to select information in very large data sets, which are called ROOT Trees,
- GUI elements, in case the user wants to avoid cumbersome coding to extract results about their analysis (histogram styling, fitting, etc).

The ROOT framework has about 460 classes grouped by functionality into shared libraries. The libraries are designed in such a way that dependencies are minimized, so the user does not have to include a lot of unnecessary classes from an inefficient organized system of libraries, which would increase file size and computation time. The `core` and `CINT` libraries include the bare minimum necessities and the interpreting classes for the C++ interpreter respectively. There are also classes that communicate with PYTHIA via an interface for event generation. The C++ interpreter is used because of its efficiency over other interpreters like Perl and Python.

ROOT offers an easy overall handling of histograms. Apart from the standard manipulation of a histogram, which includes the bin size, minimum and maximum values of an axis and style, it provides operations between histograms: addition, division, normalization, integration, as well as providing various statistical quantities via the use of fitting and minimization functions. Since this is one of the main purposes of ROOT, the framework is simple, but can also become very extensive. There are various histogram classes for `1D`, `2D` and `3D` histograms whose classes depend on the byte size (precision and type) of the data. Visualization tools allow for interactively communicating with the histogram, changing parameters and finally saving the result in various file forms.

A ROOT file contains directories and objects in the form of a UNIX file directory. The ROOT file can also be accessed graphically for easier navigation. For example, histograms and various calculations can be saved within a root file, which is machine-independent, and be read on any machine. For large quantities of the same class objects, it is advised to use the `TTree` class, which is specifically designed for this purpose. It can handle all sorts of data in various structures, like objects and arrays while effectively reducing disk storage and increasing access speed. Standard analysis practice is to fill a `TTree` with events of interest and write the `TTree` to a file for further analysis A `TTree` is composed of branches, described by the class `TBranch`, which can be read independently from other branches. In practice, a `TBranch` contains information about a multitude of physics objects, for example the coordinates of a four-vector or the particle mass m, or even non-physics objects. Each variable is called a leaf, described by the class `TLeaf`. In analogy and context with other data science frameworks, such as `pandas` in the python programming language, a `TLeaf` is a column of a `TTree` dataframe. If desired, often due to speed processing reason, some variables can be nested together to be read from the computer memory for analysis.

This nesting can be done with the help of a `TBranch`. By nesting variables within a `TBranch` instead of keeping all variables unnested, the computation time is considerably affected.

4.1.4 Other Analysis Frameworks

The CMSSW (CMS Offline WorkBook) is an open-source collection of software that CMS uses to acquire, produce, process and analyze data on a specific architecture. It is written on the C++ programming language, but configuration is performed by the python programming language. The choice of architecture is primarily tied to the version of the operating system of the LXPLUS service. Further updates of the software on a specific architecture are also available. Connecting to an LXPLUS node with unsuitable architecture for a given project may cause bugs and unpredictable behavior, therefore good knowledge and features of the CMSSW version used in the project must be established.

The framework for analysis is hosted on private GitHub repositories. The analysis framework is a culmination of the work done by the previous scientists working on the analysis[55] and is not available publicly. Two private repositories exist: One repository responsible for processing raw CMS Run 3 2022 Data into the required format for analysis, a fork of CMG-Tools and the post-processing analysis framework itself. The former, written in C++, running on `CMSSW_12_4_8` (lxplus7-8 architecture) makes use of the CRAB (CMS Remote Analysis Builder) tool, a tool which enables the user to hide their job processes[21]. The latter is written in a mixture of C++ and python programming languages, running on `CMSSW_10_4_0` (lxplus7 architecture, later adapted to lxplus8) which both are high-level general purpose languages, and is the framework in which the analyst spends most time on. This framework largely uses the ROOT computational software to process events, perform analysis and extract results either visual or in the form of numerical data. What's more, the use of ROOT library RooFit, a statistical software that is primarily used for fitting functions to data is also used in the analysis. The portion of this framework that isn't related to ROOT is the use of Machine Learning algorithms and related functions, such as the Boosted Decision Tree, as ROOT does not include such functions as of yet.

4.2 Quantity Definitions

In this section, a brief familiarization with various quantities, either physical or technical used throughout the analysis is carried. A description of the coordinate basis is important in order to adapt to the conditions of the experiment. Furthermore, the notion of tracks, vertices and various track-related parameters can render some choices of preselection cuts in the analysis understandable. What's more, a clear reference of the present variables in analysis is crucial in understanding the implications of the conclusions.

4.2.1 Detector Coordinate System

The coordinate system that is primarily used in collider experiments such as CMS is described in figure Fig. 4.2.

[21] Often for private research reasons and security.

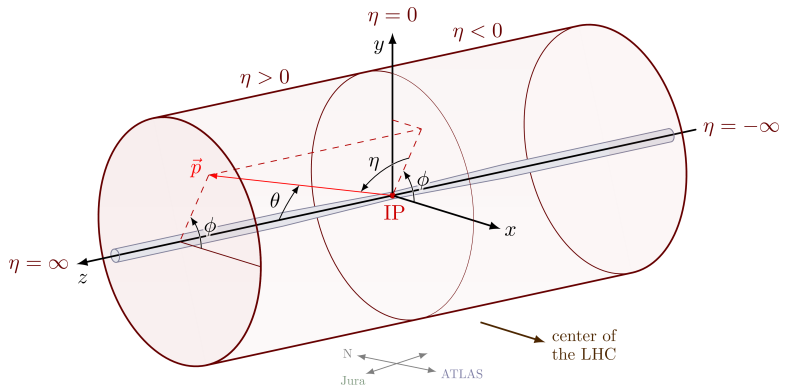


Figure 4.2: Definition of the coordinate system.

The plane normal to the beam axis is the transverse plane T. The azimuthial angle ϕ is defined from the positive x axis on the transverse plane T. The polar angle θ is defined with respect to the positive z axis. The components of the Lorentz 4-vector $p^\mu = (E, p_x, p_y, p_z)$ are difficult to measure. Therefore, a different description of the aforementioned Cartesian coordinates E and p_i $i \in \{j\}_1^3$ is attained depending on the required functionality:

$$p^\mu = (p_t, \rho, \theta, \phi)$$

where ρ is the distance in the transverse plane from the beam axis. It is important to understand an alternative and more experiment friendly description of the coordinate system. From the 4-product of the momentum of a particle $p^\mu p_\mu$ of mass m, expanded on the Cartesian coordinates in the Lab Frame, the transverse mass m_t can be defined:

$$E^2 = p_x^2 + p_y^2 + p_z^2 + m^2 \quad \Rightarrow \quad m_t^2 = p_x^2 + p_y^2 + m^2$$

or alternatively $m_t^2 = E^2 - p_z^2$. The Lab Frame boosts regarded in this analysis are on the z-axis only, therefore the transverse mass m_t is a Lorentz invariant. A useful quantity is the rapidity[56][57] y, defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{4.1}$$

An analogous definition of y is:

$$y = \tanh^{-1} \frac{p_z}{E}$$

With this expression, one can show that a Lorentz boost $\beta = u_z c$ parallel to the beam axis z, transforms the rapidity y into y' via the transformation $y' = y - \tanh^{-1} \beta$. The difference between two rapidities y_1, y_2 is a Lorentz invariant, $y_2' - y_1' = y_2 - y_1$. For a highly relativistic particle, the condition $E \approx p_z \gg m$ is approximately true, then from the definition of rapidity y (eq Eq. (4.1)), one yields an approximate form of y called the *pseudorapidity* η:

$$y \approx \eta = -\ln \tan \frac{\theta}{2}$$

where θ is defined via $\cos\theta = p_z/|\vec{p}|$, analogous to the definition of θ. Due to the challenging measurement of p_z, η replaces the measurement of y in highly energetic particles, for example in muons, as η is easier to measure due to the singular dependence on the polar angle θ. What's more, a new form of distance metric, the Angular Distance ΔR between two points (y_1, ϕ_1) and (y_2, ϕ_2) is defined:

$$\Delta R = \sqrt{(y_2 - y_1)^2 + (\phi_2 - \phi_1)^2} \tag{4.2}$$

The Angular Distance is a Lorentz Invariant only for boosts in the direction of the beam axis z, as the difference in rapidities is a Lorentz Invariant and ϕ lies in the transverse plane T. For highly energetic particles, $y \approx \eta$, therefore ΔR is approximately Lorentz invariant to a given precision of significant digits.

4.2.2 Tracks

In this thesis no focus was given in creating and fitting tracks to vertices, as the fitting software was already available. However, a brief explanation of some key concepts will be given. A vertex is a spatial component and is characterized by three coordinates x, y, z in the Lab Frame. Apart from z, these coordinates are of no much use to analysis, however they are important for performing kinematic fits to tracks and vertices. A track is a particle path with direction within the CMS geometry, characterized by a vertex at its beginning but no ending. When a kinematic fit is performed on a track, a quality value is assigned to it. Important vertices that are used in post-kinematic fit analysis are the Primary, Secondary and Tertiary Vertices. For this analysis, the Primary Vertex (PV) describes the genesis of a B^0 particle. The Secondary Vertex (SV) is the vertex where the B^0 decays and the genesis of the K_S^0 is extrapolated, (the K_S^0 is chargeless, and the kinematic fit on a chargeless track is difficult to accomplish without extrapolation methods) along with either the dimuon pair $\mu\mu$ or the intermediate states J/ψ and $\psi(2S)$. Sometimes, it is possible that the Dimuon $\mu\mu$ Vertex (DV) does not correspond to the Secondary Vertex. The Tertiary Vertex corresponds to the decay of the K_S^0 pseudotrack and the genesis of the $\pi\pi$ tracks.

It should be noted however, that in contrast to the strongly decaying intermediate states J/ψ and $\psi(2S)$, the K_S^0 decays weakly. Lifetimes mediated by the Weak Interaction are orders of magnitude greater than interactions mediated by the Strong and Electromagnetic Interaction. This causes the K_S^0 to be displaced, complicating the kinematic fit of the track as the vertex of the genesis of the decay products of the K_S^0, the dipion pair $\pi^+\pi^-$ are not in the same vertex as other B^0 decay products.

4.2.3 Sample Quantities

In this section, the commonplace variables within the available samples of any type are showcased. Understanding these variables is necessary for future reference and physical intuition as they come up in analysis, especially during the Machine Learning stage.

- p_t, η, ϕ, m, `charge` are the transverse momentum after refitting to the B^0 vertex, pseudorapidity, azimuthial angle, mass and charge of the corresponding particle respectively. For the two muons, μ_1 and μ_2, the former and latter are labeled as leading and subleading according to the higher and lower value of their p_T respectively.

- $m(\mu\mu)$ refers to the dimuon mass system. In the context of the analysis, it is defined as $q^2 = m(\mu\mu)$ due to the common usage of this quantity.

- $m_{\mathrm{MC}}(B)$ is the mass of the reconstructed B^0 with additional requirements that the mass resolution ($m(\mu\mu)$ is fixed around the mass of the J/ψ at 3.097GeV. The same applies for the $\psi(2S)$ particle at 3.686GeV). The label MC stands for Mass Constraint while referring to the mass quantity.

- $\cos_{2\mathrm{D}}\alpha$ is the cosine of the angle in the T plane between the line segment connecting the beamspot and the Primary Vertex of the B^0 and the momentum vector of the

B^0. Equivalently, it is the value of the cosine between the angle of the momentum vector of the B^0 candidate and its position vector.

- $p(B^0)$ or `Bprob` is a statistic of kinematic fit quality for the B^0 vertex and is derived from the normalized χ^2 of the vertex. Values of 0 denote poorest quality, whereas values of 1 denote best quality.
- L_{xy} is the decay length in the Lab Frame of the B^0 in the transverse plane T.
- τ_{xy} is the mean lifetime in the Lab Frame of the B^0 in the T plane, given by the relation: $\tau_{xy} = L_{xy} m(B)/p_t c$.
- $IP(K^0_S, 2D)$ and $IP(K^0_S, 3D)$ are the 2D and 3D impact parameters of the Secondary Vertex.
- sig(L_{xy}) is the decay length significance of the B^0 meson and is given by the ratio $L_{xy}/L_{xy\,\mathrm{unc}}$.
- sig(D_{xy}) is the absolute value of the 2D Impact Parameter of the dimuon Vertex and is given by the ratio $D_{xy}/D_{xy\,\mathrm{unc}}$.
- $\Delta R(K^0_S, \mu)$ refers to the minimum ΔR between the K^0_S particle and the leading or subleading μ.
- $\Delta\eta(\mu_1, \mu_2) = \eta_2 - \eta_1$ and $\Delta R(\mu_1, \mu_2)$ are the difference in pseudorapidity and angular distance between the leading and subleading muon as described in eq. Eq. (4.2).
- `MediumID` and `SoftID` are Particle Flow with additional track quality requirements and non-PF low-p_T muons respectively. `isTrg` refers to whether the muon has fired the `DoubleMu4_3_LowMass` trigger. All three variables are binary, with 0 corresponding to false and 1 to true.
- Iso(μ) refers to the PF relative isolation of the μ. The Particle Flow isolation is calculated as the sum of charged hadron, neutral components within $\Delta R < 0.4$ and it is the standard PF relative isolation.
- Δz is the absolute value of the difference of the z coordinate between the dimuon Vertex and the z coordinate of the distance of closest approach (DCA) of the K^0_S pseudotrack. The notion of pseudotrack is attributed to the neural K^0_S. Noncompactly: $\Delta z = \left| z\Big(\mathrm{vtx}(\mu\mu)\Big) - z\Big(\mathrm{dca}(K^0_S)\Big)\right|$.
- AP (Armenteros-Podolanski) q_T is the transverse momentum to the flight direction of the B^0[58].

4.3 Acceptance and Efficiency of the MC Samples

The analysis starts with the determination of the acceptance and efficiency for the Unbiased and Biased MC samples respectively, quantities that are essential in the measurement of the R_{K_S} observables. As has been explained, the MC samples include one process each, therefore each data sample can be studied extensively so that its behavior is well known. The aim of this study of examining the MC dataset well is to acquire a set of variable preselection cuts to be applied on the CMS Run 3 2022 Data samples because in these data samples the search for the $B^0 \to \mu^+\mu^- K^0_S$ signal and its

resonances will be performed. By cutting a lot of events unrelated to this signal search by predetermined selection of variable ranges, the analysis becomes more efficient in both terms of computing time and fitting techniques. Therefore, it is imperative to search for a set of cuts that mimic the MC signals as much as possible.

The MC samples contain GEN and RECO level quantities. GEN level quantities correspond to quantities of particles as have been generated by the generator. The RECO level quantities correspond to quantities of particles that have been reconstructed by the simulated detector. Naturally, the RECO level quantities generally differ from the GEN level quantities due to the limited detector resolution, efficiency and geometric acceptance. It is the RECO level quantities that are measured by the CMS detector in the Data samples, therefore the analysis cannot be done with GEN level quantities, but they are useful in determining the acceptance α.

First of all, the dimuon mass $m(\mu\mu)$ is binned into 9 bins, described in table 4.1. The bins are not all joined continuously and the reason for this is to set a clear distinction between the rare mode region and other resonances. The space between bins 1-2 is to avoid low-q^2 η and ϕ particle resonances. The space between bins 3-**4**-5 is to distinct the J/ψ resonance in bin **4**. The same applies to the spacing in bins 5-**6**-7 for the $\psi(2S)$ resonance, which is for the distinction of that particle in bin **6**. The two resonant decay channel data samples are used in their respective bins.

q^2-bins (GeV^2)			
1	[0, 0.98)	**6**	[12.60, 14.44)
2	[1.1,4.0)	7	[15.0,17.0)
3	[4.0,8.0)	8	[17.0,19.0)
4	[8.41,10.24)	9	[19.0, 23.0)
5	[11.0, 12.50)		

*Table 4.1: Bins in $m^2(\mu\mu) = q^2$. The bins in **bold text** correspond to the two resonant mode bins. The bins are left-inclusive and right-exclusive in q^2.*

The calculation of acceptance α in Unbiased MC samples and efficiency ϵ in Biased MC samples are performed after a collection of variable cuts is performed. Since there are different collections of variable cuts across different bins, the acceptances and efficiencies for a specific bin-dependent collection of variable cuts will be denoted as α_i^j and ϵ_i^j where i refers to the collection of variable cuts i and j refers to the bin in table 4.1 and should not be confused with an exponent. The calculation for these quantities is retrospective, defined as:

$$\alpha_i^j = \frac{M_i^j}{M_{i-1}^j}, i \in \{1,2\} \quad \epsilon_i^j = \frac{N_i^j}{N_{i-1}^j} \quad i \in \{3,4,5,6,7\} \tag{4.3}$$

where M_i^j, N_i^j is the number of events passing the Collection of variable cuts i in the jth bin in the Unbiased and Biased MC samples respectively, with M_0^j, N_0^j being the total events within their respective MC samples. In order to calculate these quantities, the collections of variable cuts must first be determined. They are shown in tables 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7. In these tables, μ_1, μ_2 correspond to the leading and subleading muon respectively, and μ corresponds to both muons. For simplicity, the acceptance or efficiency at a particular Collection of variable cuts i for all bins will also be denoted as α_i and ϵ_i respectively. The uncertainty of either acceptance $\sigma(\alpha_i^j)$ or efficiency $\sigma(\epsilon_i^j)$ on any

Collection i and bin j are calculated via the binomial uncertainty:

$$\sigma(\alpha_i^j) = \sqrt{\frac{\alpha_i^j\left(1-\alpha_i^j\right)}{M_{i-1}^j}} \quad \sigma(\epsilon_i^j) = \sqrt{\frac{\epsilon_i^j\left(1-\epsilon_i^j\right)}{N_{i-1}^j}} \tag{4.4}$$

The following tables show the description of Collection of variable cuts. There are six Collections in total across the two MC sample types.

Collection 1: GEN-μ Cuts
$p_T(\text{GEN } \mu_1) > 3.5$ GeV
$p_T(\text{GEN } \mu_2) > 2.5$ GeV
$\lvert\eta(\text{GEN } \mu)\rvert < 2.5$

Table 4.2: GEN level cuts for the two muons. These muons come directly from the B^0 decay, except in the resonant regions (bins 8 & 12), where they come from the two resonances $J/\psi \to \mu\mu$ and $\psi(2S) \to \mu\mu$.

Collection 2: GEN-π Cuts
$p_T(\text{GEN } \pi) > 0.7$ GeV
$\lvert\eta(\text{GEN } \pi)\rvert < 3.0$

Table 4.3: GEN level cuts for the two pions. These pions come from the decay of the $K_S^0 \to \pi\pi$.

The use of the Unbiased sample is purely to determine the acceptance after these two GEN level Collection of variable cuts. The reason for this is purely to measure the reduction of the Phase Space due to the geometry of the detector. As it can be seen from the plots in figure 4.4, the acceptance in each bin is less than 1% signifying that the Phase Space before reconstruction is on the order of magnitude of $< 1\%$. Therefore, if analysis were to continue with the Unbiased sample, it would require a tremendous amount of file size, complicating technical matters such as memory availability and sample production. Instead, the Biased MC sample contains an even larger amount of events while also effectively taking into account the Collections 1 & 2 when they have been artificially generated. This is the reason for the choice of $i \in \{1, 2\}$ in α_i and $i \geq 3$ in ϵ_i.

Collection 3: RECO-μ Cuts
$\Delta R(\text{GEN, RECO})_\mu < 0.03$
mediumId RECO (μ)
$p_T(\text{RECO } \mu_1) > 4$ GeV
$p_T(\text{RECO } \mu_2) > 3$ GeV
$\lvert\eta(\text{RECO } \mu)\rvert < 2.4$

Table 4.4: RECO level for the two muons. The ΔR variable matches the GEN level muon with the RECO level muon to a good degree in order to avoid bad reconstructions.

Collection 4: HLT Cuts
μ_1, μ_2 trigger HLT_DoubleMu4_3_LowMass

Table 4.5: The trigger used is a double muon trigger. The label 4_3 label corresponds to the minimum p_T of the two muons.

The choice of having a Collection of variable cuts representing the reconstruction of the muons and the Collection of a single cut with the Muon Trigger is to denote the proper approach for handling muons within the CMS experiment, as the muon trigger allows extremely good quality muons to be used in the analysis.

Collection 5: RECO-K_S^0 Cuts
$\Delta R(\text{GEN, RECO})_{K_S^0} < 0.1$
$p_T(\text{RECO } K_S^0) > 1$ GeV
$\lvert\eta(\text{RECO } K_S^0)\rvert < 2.4$

Table 4.6: RECO level for the K_S^0.

Collection 6: Preselection Cuts
$m(\text{RECO } B^0) \in (5, 5.6)$ GeV
$\cos_{\text{2D}}(\alpha) > 0.90$
$p(B^0) > 0.00001$
$\Delta z(\mu\mu, K_S^0) < 1cm$
IP SV3D(RECO K_S^0) $< 0.045cm$
AP $q_T > 0.105GeV$

Table 4.7: Preselection Cuts for the next stage of the analysis. Descriptions for these variables are found in section 4.2.3.

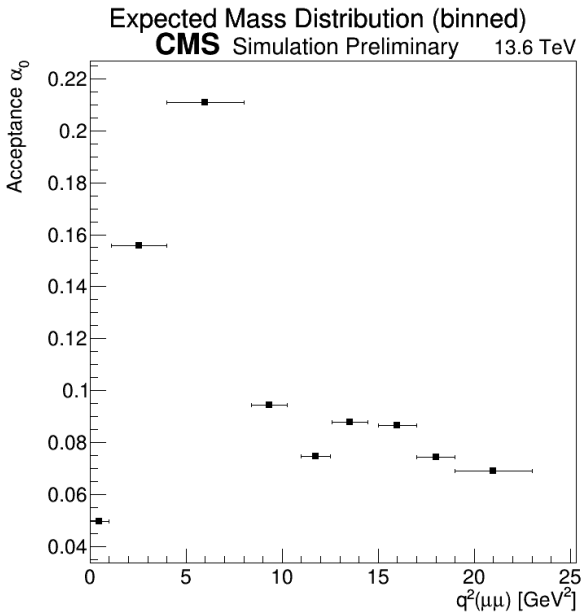
The reconstruction of the neutral K_S^0 pseudotrack is followed by further specific preselection cuts for the analysis, given that the muons have been properly reconstructed and identified by the trigger. The MC and Data samples following these cuts are used in the BDT training procedure, described in section 4.5. Collection 7 includes a variable cut on the BDT prediction score $x > c_b$ as described in section 4.4 which is necessary in order to differentiate unwanted processes from the signal search in the Data samples. This means that this Collection cannot be used in the BDT training procedure as it presupposes that it has been formulated. Furthermore, this collection also contains cuts that were introduced later in the analysis, but are incorporated in determining the $R_{K_S^0}$ observables.

Collection 7: BDT & Background Cuts
$x > c_b$, to be determined
$\lvert(m(B^0) - 5.27925GeV) - (m(\mu\mu) - 3.096900GeV)\rvert > 0.19GeV$, bin 4 only
$\lvert(m(B^0) - 5.27925GeV) - (m(\mu\mu) - 3.686097GeV)\rvert > 0.08GeV$, bin 6 only
$m(K_S^0\mu_2) > 2GeV$, bins 1-3 only
$\Delta R(\mu, \pi) > 0.03$, bins 1-3 only

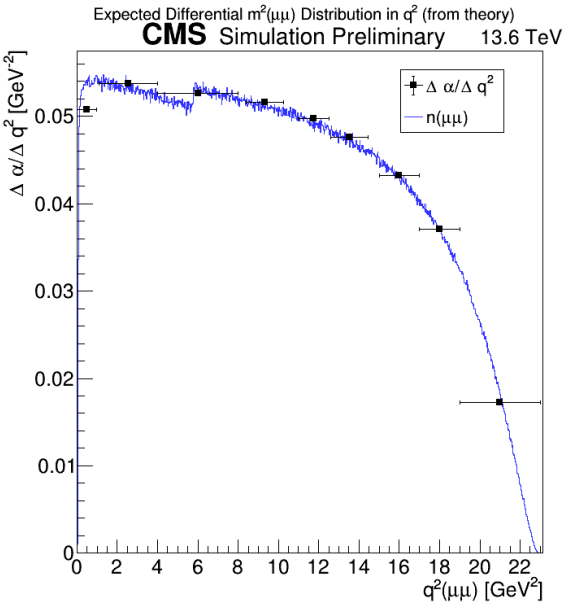
Table 4.8: BDT & Background Cuts for the next stage of the analysis. The variable cut $x > c_b$ corresponds to a cut on the BDT prediction value. The value of c_b is determined as explained in section 4.6.2.

The second and third cuts are referred to as bin leakages. Specifically, they regard the background leakage from the left tail of the two resonant mode B^0 mass peaks into the left $m(B^0)$ sideband, $m(B^0) \in (4.85, 5.05)GeV$. The fourth cut on the mass of the subleading μ and K_S^0 system is to reject more semileptonic decay background in the Data samples due to the left sideband. Lastly, the final cut exists to prevent events from using $\mu_1, \mu_2, \pi_1, \pi_2$

tracks twice in order to create the B^0 candidates. Now that the Collections have been fully determined, the values of α_i^j and ϵ_i^j are shown in figures 4.3, 4.4, 4.5 and 4.6.



(a) Mass Distribution of $m(\mu\mu)$, no collection of variable cuts applied.



(b) Normalized Mass Distribution of dimuon mass $n(\mu\mu)$, no collection of variable cuts applied.

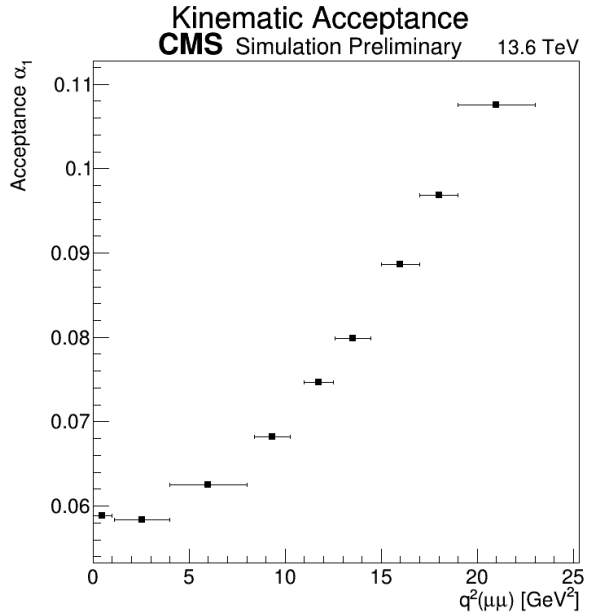
Figure 4.3: Plots of Mass and Normalized Mass Distributions of the Unbiased MC sample of $B^0 \to \mu^+\mu^- K_S^0$ across all 17 q^2 bins.

The Normalized Mass Distribution $n_k(\mu\mu)$ at bin k in figure 4.3b is calculated using the following expression:

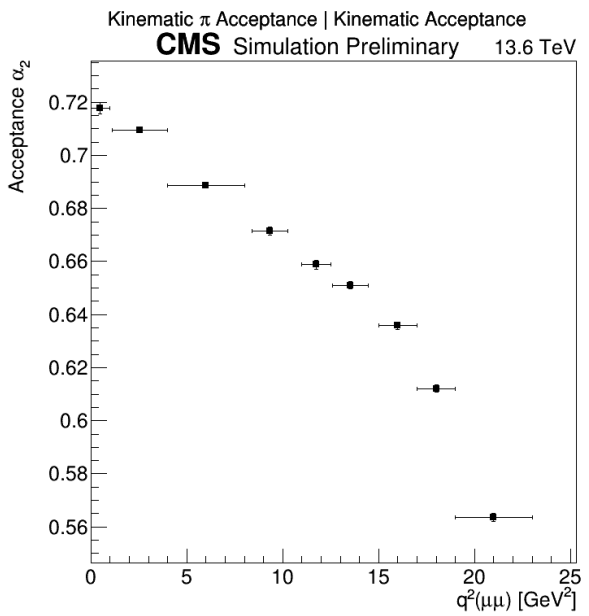
$$n_k(\mu\mu) = \frac{m_k(\mu\mu)}{\sum_{i=1}^{1000} m_i(\mu\mu)}$$

as it is binned in a histogram with 1000 bins. For the remainder of this thesis, thinly-binned smooth[22] distributions, with number of bins greater than 1000 will not be referred to by bin but as a continuous distribution. Therefore, the Normalized Mass Distribution will be referred to as $n(\mu\mu)$.

[22] with enough statistics to justify the large number of bins.

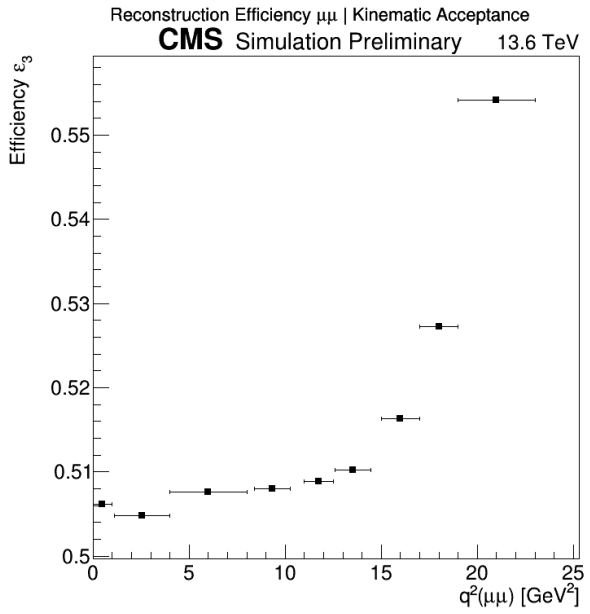


(a) Plot of acceptance α_1^j.

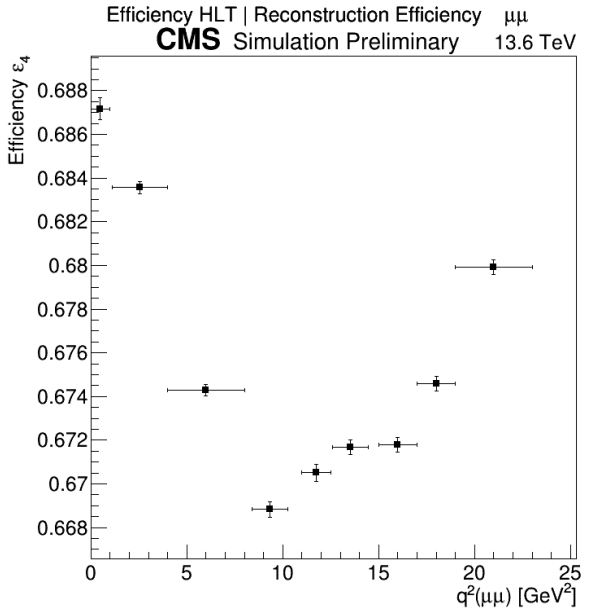


(b) Plot of acceptance α_2^j.

Figure 4.4: Plots of acceptances α_1^j and α_2^j across all 9 q^2 bins for the Unbiased MC sample of $B^0 \to \mu^+\mu^- K_S^0$.

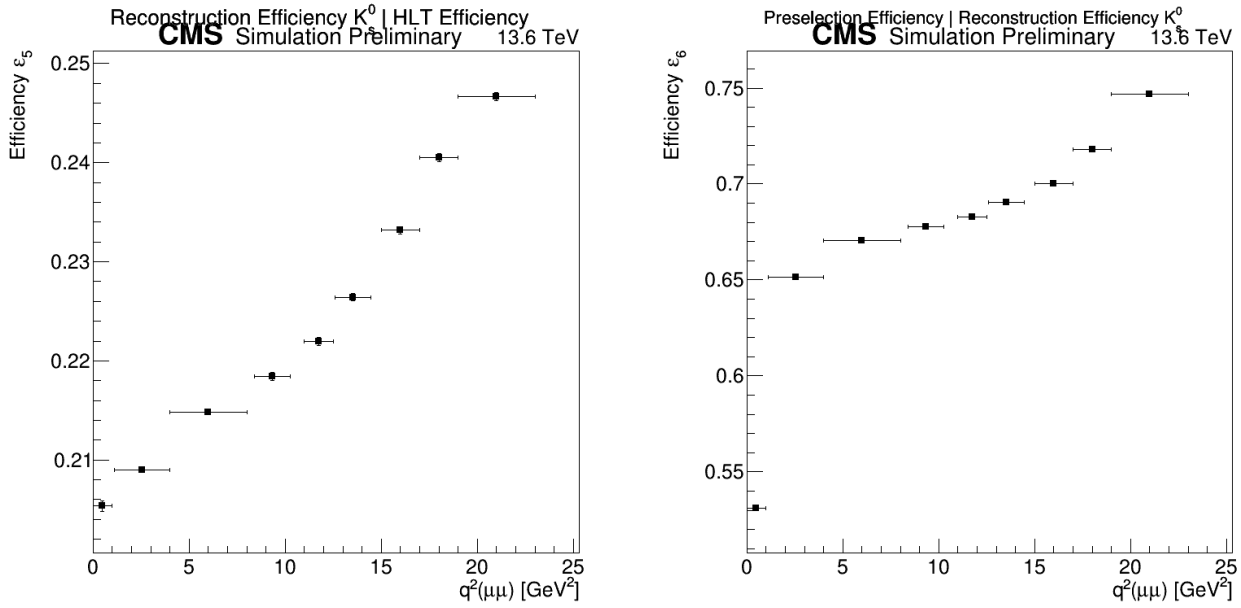


(a) Plot of efficiency e_3^j.



(b) Plot of efficiency e_4^j.

Figure 4.5: Plots of acceptances e_3^j and e_4^j across all 9 q^2 bins for the Biased MC sample of $B^0 \to \mu^+\mu^- K_S^0$.



(a) Plot of efficiency e_5^j. *(b) Plot of efficiency e_6^j.*

Figure 4.6: Plots of acceptances e_5^j and e_6^j across all 9 q^2 bins for the Biased MC sample of $B^0 \to \mu^+\mu^- K_S^0$.

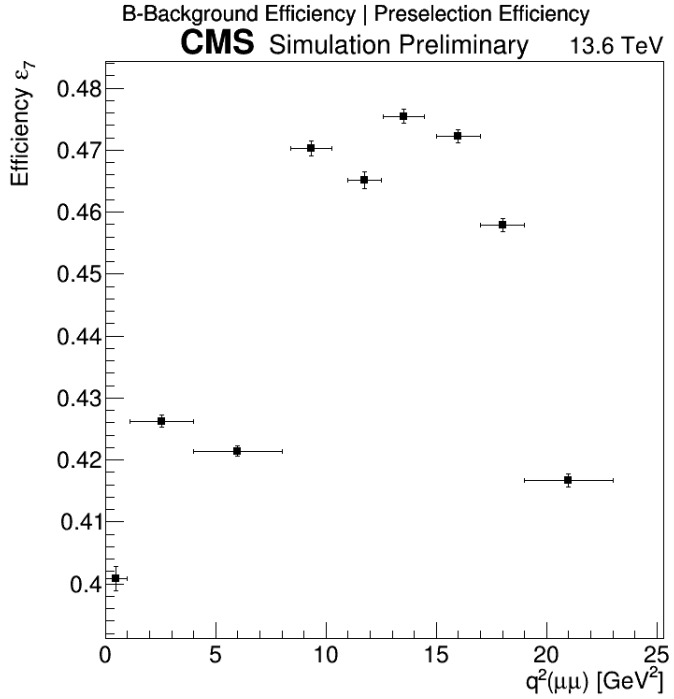


Figure 4.7: Plot of efficiency e_7^j across all 9 q^2 bins for the Biased MC sample of $B^0 \to \mu^+\mu^- K_S^0$, along with the optimal BDT value $x > 4.5$.

The large variations in figure 4.7 can be attributed to the BDT cut chosen for this analysis, according to section 4.6.2. The acceptance and efficiency values used in equations 3.3 and 3.4 are determined from the individual values of α_i^j and ϵ_i^j. The acceptance and efficiency used are equal to:

$$\alpha^j = \alpha_1^j \, \alpha_2^j, \quad \epsilon^j = \epsilon_3^j \, \epsilon_4^j \, \epsilon_5^j \, \epsilon_6^j \, \epsilon_7^j$$

Table 4.9 shows the values of α^j and ϵ^j. Table 4.10 shows the values of α^4 and ϵ^4 of the J/ψ resonance and α^6 and ϵ^6 of the $\psi(2S)$ second resonance respectively.

Quantity	Bin 1	Bin 2	Bin 3	Bin 5	Bin 7	Bin 8	Bin 9
Acceptance (10^{-4})	442 ± 2	414 ± 1	430 ± 1	492 ± 2	564 ± 2	593 ± 2	606 ± 2
Efficiency (10^{-5})	1520 ± 9	2077 ± 6	2077 ± 5	2406 ± 9	2676 ± 8	2814 ± 9	2892 ± 9

Table 4.9: α^j and ϵ^j values for 7 out of the 9 bins in table 4.1 of the Rare Mode. The resonant mode bins 4 & 6 are irrelevant in further analysis and are not given.

Quantity	Value
J/ψ Resonance	
Acceptance (10^{-5})	4648 ± 5
Efficiency (10^{-5})	2450 ± 5
$\psi(2S)$ Resonance	
Acceptance (10^{-5})	5231 ± 6
Efficiency (10^{-5})	2615 ± 6

Table 4.10: α^j and ϵ^j for bin 4 & 6 in table 4.1 of the J/ψ Resonant and $\psi(2S)$ Second Resonant mode MC samples. Each value corresponds to the respective bin of the resonant mode MC sample.

4.4 Decision Trees

Once Collection 6 has been applied to the Biased sample, the sample is then prepared for training a Boosted Decision Trees (BDTs) Machine Learning (ML) model, built on the principle of a simpler ML model, Decision Trees (DTs). The python package `xgboost` version 0.80 is used, along with the `scikit-learn` wrapper version 0.19.1. BDTs are used to provide a differentiation of the rare signal versus the background in the Data samples by assigning a score in each event in that Data samples. But first, a brief introduction to Decision Trees (DTs) and their correlation to BDTs will be given.

Decision Trees[59], first developed by Breiman et al[60] in the CART (Classification And Regression Trees) algorithm, are a ML technique first developed for non-HEP applications, such as data mining, pattern recognition, later adapted in medical diagnosis, insurance and recognition of handwritten text. In the context of this thesis, the trees are *binary*, meaning that two classes are considered, namely the signal, which is the Rare Mode, and background, which describes events failing the classification as signal. DTs classify events between these two classes in a node t using a selection criterion or *impurity measure* $i(t)$. This measure often must satisfy some properties in order for the DT algorithm to work:

- $i(t)$ should have a maximum value for an equal presence of the two classes (signal and background)
- $i(t)$ should have a minimum value for the absence of a class (presence of a single class)
- $i(t)$ should be symmetric in each of the two classes
- $i(t)$ should be strictly concave in order to avoid trapping the algorithm into states separating the two classes in local minima. This means that if $i(t)$ has local minima, the DT could make boolean statements starting from the root node leading into leaves with worse separation of the two classes.

The goal is to find the best boolean statement S among the set of different variables and different splitting values $\mathcal{S}$ that maximizes the decrease of impurity[23] $\Delta i(S,t)$ on the node t_j:

$$\Delta i(S,t_j) = i(t_j) - pi(t_{j,p}) - (1-p)i(t_{j,1-p})$$

where p is the fraction of events that pass the boolean statement and $1-p=q$ is the fraction of events that fail the boolean statement. It can be written:

$$\Delta i(S^*,t_j) = \max_{S\in\mathcal{S}} \Delta i(S,t_j)$$

Common impurity measures $i(t)$ that satisfy the criteria mentioned above are:

- The misclassification error $M(p) = 1 - \max(p, 1-p)$
- Entropy $S(p) = -p\log_2 p - (1-p)\log_2(1-p)$
- The Gini Index $G(p) = 1 - p^2 - (1-p)^2$

The description of a DT is rather simple and is showcased in figure 4.8.

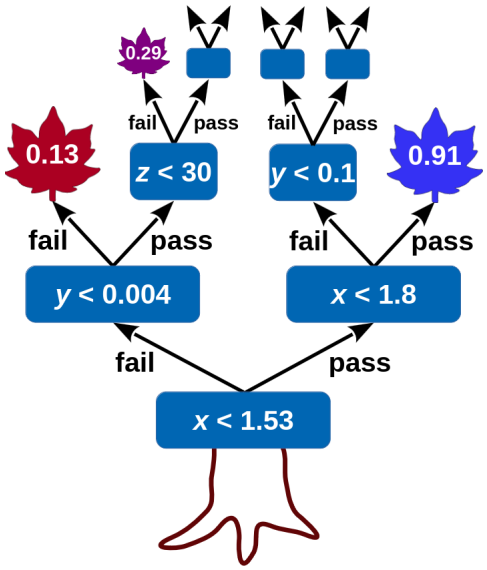


Figure 4.8: Graphical representation of a DT. The blue rectangles denote nodes, the directional black arrows denote branches, and the red-blue color-mapped leaves denote the terminal nodes along with their impurity measure. The root node is at the bottom of the figure, where a tree stump is drawn on the bottom side of the node. Each node assigns a boolean test on an arbitrary variable x, y, z[59].

The root node is the initial node. Each node is split up to two other branches, each leading to a new node until the splitting condition is not satisfied. The splitting condition can be partially controlled by the person or machine tuning the DT as well as by the algorithm itself. Nodes that do not lead to further branches are referred to as terminal nodes or leaves. The collection of nodes and branches constitutes the DT. The initialization of the algorithm happens at the root node. The algorithm can be summarized in the following steps.

1. If this node satisfies a stopping criterion, declare it as terminal and exit the algorithm

[23] Essentially the subtraction between the impurity measure of the node t_j and the child nodes.

2. For each variable (e.g. p_T, η of a particle) find the splitting value that separates events into the two classes (signal and background) the best. If the separation cannot be improved with respect to the previous node along the branch connecting to it, turn this node into a terminal node.

3. Select the variable and its splitting value leading to the best separation of the two classes. Turn this choice into a boolean statement and split the node into two new nodes, the nodes satisfying and failing this criterion.

4. Go back to step 1 and apply the steps for a new node (until all are turned into terminal nodes).

Once this processes has finished, the DT is said to be trained and ready for use. New samples of events, unknown to the tree can be now parsed into the DT, and the tree will now make a prediction based on its structure, where the event will be classified as one of the two classes by being tested on the boolean statements that constitute its structure. Lastly, it is imperative to note a few things regarding training DTs. First of all, CPU consumption scales as $nNlogN$ where n is the number of variables and N is the number of training events. Therefore, eliminating variables that would not contribute to the training of a BDT will make the training procedure faster. What's more, it is possible for a particular training variable to be particularly noisy. This means that it can inherently damage and complicate training and should be removed, however, this is not as common.

4.4.1 Boosted Decision Trees

Boosted Decision Trees are made by multiple trained individual DTs, grouped and combined in specific ways depending on the BDT algorithm chosen. BDTs come with a lot of benefits:

- they are trained fast and also predict very fast unlike other typical ML models in this area of study. In this thesis, all each BDT was trained with a maximum training time of about twenty five minutes.

- they are easy to tune in order to improve the quality of the training procedure and the prediction quality.

- they are not scale sensitive. The features that a BDT is trained on can be continuous data as well as categorical/non-number data encoded in the form of numbers.

- there is little need for data pre-processing. Variables do not need to be scaled (linearly or non-linearly) before a BDT is trained.

- they have good performance, meaning that training on the residuals gives very good accuracy in predicting.

- they are incorporated in numerous software across different programming languages (e.g. python, R) which have been widely used and tested in the past.

In this thesis, the BDT algorithm is `gbtree` which is a Gradient Boosting algorithm for training a BDT model. Using this algorithm, the BDT model F is expanded iteratively to cover for its imperfections. If at step k the model has the form F_k, a new component h_k is added to the model in such a way that aims to improve at the imperfections:

$$F_{k+1}(x) = F_k(x) + h_k(x)$$

where x denotes the events of the training sample. The guidance to improve the imperfections comes from the Loss function $L(x, y)$, y denoting the class label (signal or background), an element of ML models where it aims to minimize the error made in predictions. By minimizing the Loss function across all trees in the model:

$$\frac{\partial L(x, y)}{\partial F_k(x)}$$

one can express the new residuals to the model $h_k(x)$ as a function of the Loss function in the form of negative gradients of the Loss function:

$$h_k(x) = h_k\left(-\frac{\partial L(x)}{\partial F_k(x, y)}\right)$$

hence the name of the algorithm. One major issue with BDTs is that they are susceptible to a phenomenon called overfitting. An overfitted ML model such as the BDT is very often a bad predictor due to having learned the training dataset very well and expecting each testing or predicting dataset to be extremely similar to it in terms of behavior, yielding bad predictions. This is equivalent of the BDT model lacking any predicting power, relying solely on the behavior of the distributions supplied in the training dataset. Therefore, it is imperative that the BDT is properly tuned and measures to prevent overfitting are taken.

4.4.2 Complementary Methods for Training

BDT Training is a stochastic processes. The main element of randomness comes from the composition of the training sample.

Cross Validation Techniques Cross Validation (CV) is a technique that is a counter to overfitting. The training dataset $\mathcal{T}$ of the two classes is split into k folds of equal size and a BDT classifier T_i is trained on all folds except the ith fold, which is used for testing[59] during the training process in order to get a model exclusive error estimate:

$$\mathcal{T} = \bigcup_{i=1}^{k} \mathcal{T}_i$$

This procedure produces k different BDT classifiers. CV techniques give an estimate of the variability of the performance of the output. There are three CV techniques that have been used in this thesis for various processes and tasks. They are described shortly:

- simple k-fold Cross Validation (CV), as described above.
- n-Repeated k-fold Cross Validation (RCV), repeating the CV process n times, treating all k folds across all n repetitions equally to a total of $n \cdot k$ folds.
- Stratified n-Repeated k-fold Cross Validation (SRCV), performing RCV but taking a balanced ratio of the two classes across each fold and repetition instead of a completely random sample.

This method can reveal large or nonexistent variations in the output of the training procedure, serving as a guide to counter overfitting.

Feature Selection Feature Selection is a technique that determines and ranks the features (variables) in a dataset given a non-qualitative criterion as well as choosing a subset of the initial features that can replicate the results of a trained model with minimal information loss. Feature selection methods provide

- data interpretability,
- the removal of redundant, noisy and highly-correlated features

There are two main kinds of feature selection methods[61], filter and wrapper methods. Filter methods assign each selected feature an importance criterion based on its contribution to the class labels while not evaluating the performance metric of the classifier's model on the selected features, effectively allowing the distributions of the features in the dataset in a given model to be characterized by their individual connections to the class labels. Filter methods therefore explore the relevance between a feature and the class labels. This makes filter methods independent of the model used, reducing systematic errors and are very time effective due to the lack of the computational complexity supplied by evaluation of the performance metric of the underlying ML model. However, they are susceptible to the stopping criterion of the method itself. Wrapper methods employ a performance metric based on the model to evaluate a candidate feature subset. Based on the results of the evaluation metric, the optimal subset is chosen based on the results of the evaluation process. The evaluation is performed by using a distinct testing dataset in conjunction to the training dataset in which the performance metric is evaluated on. Wrapper methods can model the dependencies of feature distributions, be able to interact with the underlying model but consume a lot of computational resources and rely on the underlying models' training procedure, introducing a risk of overfitting.

4.5 Training the BDT

The Data samples are split so that a part of them may be used in training a BDT. Only $B^0 \rightarrow \mu^+\mu^- K^0_S$ MC Biased samples are used in the BDT training procedure in order to define the signal class, as no background is present in the MC samples, completely removing the risk of damaging the efficiency of the trained BDT. In order to provide as many background processes as possible, by eradicating the risk of various B^0 or other B meson decays present in the Run 3 2022 Data samples to be characterized as signal, a variable cut for the mass of the B^0 candidate is applied, namely: $m(B^0) \in (4.85, 5.05) \cup (5.48, 5.65) GeV$. As explained, the area between the aforementioned cut $(5.05, 5.48) GeV$ contains various B^0 decay processes which can be characterized as signal while searching in the Data samples. This region of the Data sample is called the Sideband region (figure 4.9) and is only used in the training of the BDT Model throughout this thesis. What's more, during BDT training an important question is also posed regarding the training q^2 of the BDTs. The 17 bins in table 4.1 are adjusted in the following four unique q^2 regions, plus an additional q^2 region made up from two unique ones:

Region Label	q^2 Range (GeV^2)
low-q^2	[1.10, 6.00]
J/ψ-q^2	[8.41, 10.24]
$\psi(2S)$-q^2	[12.60, 14.44]
high-q^2	[16.00, 23.04]
low+high-q^2	[1.10, 6.00]∪[16.00, 23.04]

Table 4.11: The five q^2 bins that are used for the BDT training and measurement analysis, as well as B^0 mass fitting (section 4.6). The low+high-q^2 bin is a combination of the low-q^2 and high-q^2 bins.

The reason for choosing these five regions is rather simple. The first and core reason is to test whether the BDT performs best in a particular q^2 region. To increase the statistics of the samples, some of the bins of table 4.1 were merged into 4 unique q^2 regions. The dimuon mass in the low-q^2 region does not come from unwanted background meson decays, except from η, ϕ meson decays at the lower q^2 at about 0-1GeV^2, so these values of q^2 have been omitted deliberately. In contrast to this, the high-q^2 region, there are more intermediate $c\bar{c}$ meson decays present which can contaminate the training of the BDT. The two resonant intermediate regions favor the identification of the two resonant states J/ψ and $\psi(2S)$. Lastly, the low+high-q^2 region is a combination of the two aforementioned regions, as an additional test for BDT training. The reason for including this q^2 region is to increase the statistics of the samples even more compared to just the low-q^2 and high-q^2 regions in case these regions alone could not train the BDT classifier to a satisfying degree.

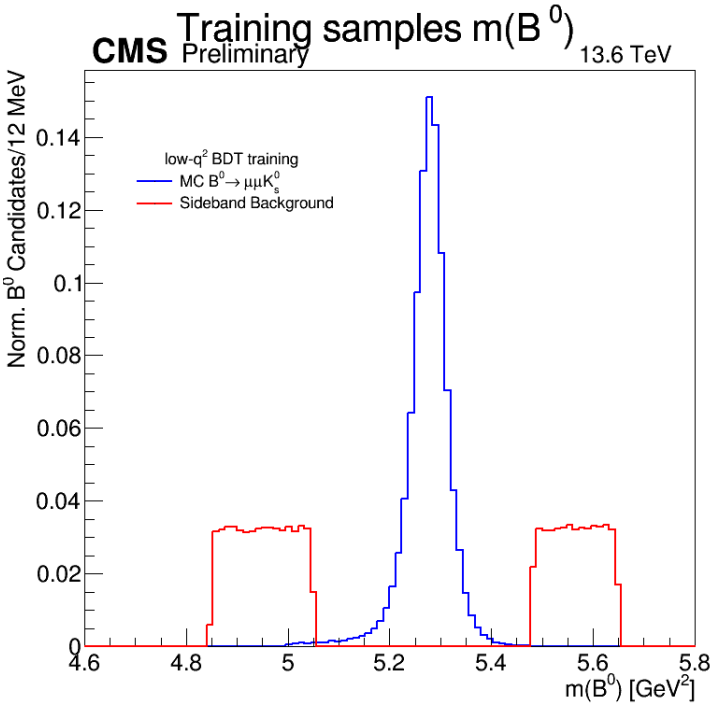


Figure 4.9: Training region for the BDT model. The Biased MC $B^0 \to \mu^+\mu^- K^0_S$ samples as well as the Data samples are trained in the low-q^2 region. The Data samples are subjected to the sidebands of the B^0 mass, $m(B^0) \in (4.85, 5.05) \cup (5.48, 5.65) GeV$. Both distributions are normalized to the sum of their bin content.

4.5.1 Choosing the q^2 Training Region

The plot of $m^2(\mu\mu) = q^2$ mass spectra for the Biased MC Signal and Sideband Background along with the four unique q^2 regions can be found in figure 4.10.

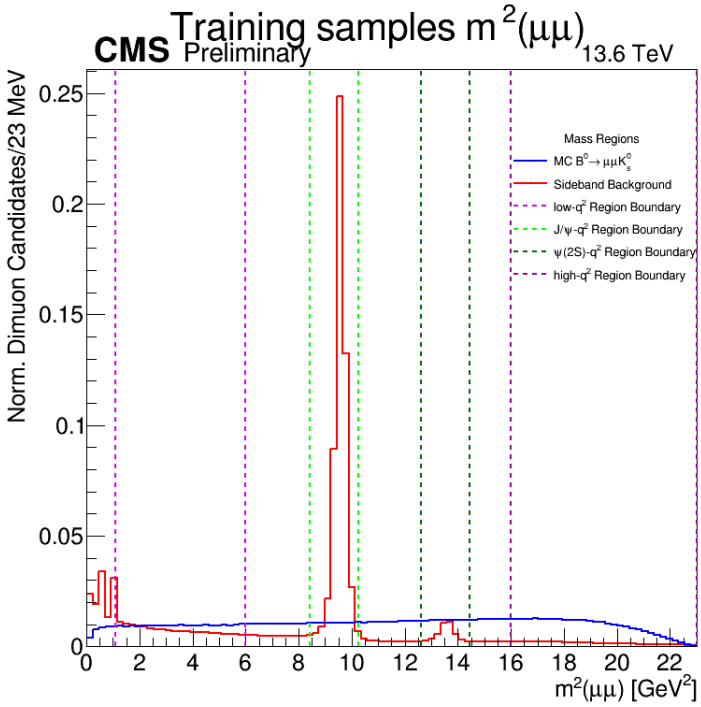


Figure 4.10: $m(\mu\mu)$ distribution spectra for the two samples, normalized to the sum of the bin contents. The plot also shows the regions of the four unique q^2 regions. Left to the boundary of the q^2 region, the η, ϕ meson resonances can be seen.

Since it is imperative to distinguish the Biased MC $B^0 \to \mu^+\mu^- K_S^0$ sample in real data acquired in the CMS experiment, the only possible training regions in q^2 for any BDT are the two unique low-q^2 and high-q^2 as well as the composite region low+high-q^2. Initially, three BDTs with the same hyperparameters (table 4.12) were trained on the same training samples and predictions were made on the same evaluation samples. Furthermore, the two class labels have had been weighted in order to improve the training procedure. Each of the two class labels has been assigned a weight w_i $i = \{1, 2\}$ according to the `compute_sample_weight()` function of `sklearn` wrapper:

$$w_i = \frac{Y_1 + Y_2}{2Y_i} \quad i = \{1, 2\}$$

where Y_i is the total number of events belonging to a class, $i = 1$ for the Biased MC $B^0 \to \mu^+\mu^- K_S^0$ signal and $i = 2$ for the Run 3 Data 2022 Sideband Background.

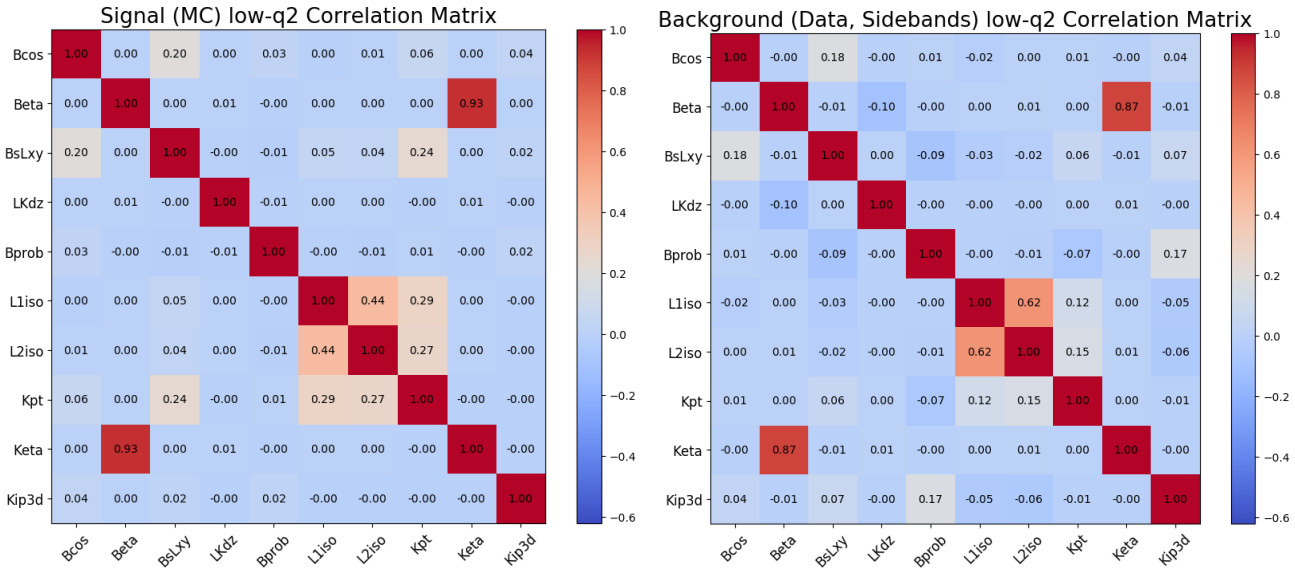
Hyperparameter	Value
gamma	3
learning_rate	0.1
max_depth	6
n_estimators	750
min_child_weight	1.0
scale_pos_weight	1.0
objective	binary:logitraw
subsample	1

Table 4.12: Initial parameters for BDT training on the low-q^2, high-q^2 and low+high-q^2 training regions.

The correlation between two variable distributions X and Y is performed using Pearson's coefficients:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma(X)\sigma(Y)} = \frac{E\big((X-\mu(X)(Y-\mu(Y)\big)}{\sigma(X)\sigma(Y)}$$

where $\text{cov}(X,Y)$ is the covariance between the variable distributions X and Y and $\sigma(X)$, $E(X)$ and $\mu(X)$ is the standard deviation and the mean value of the variable distribution X. The correlation matrix ρ plots for the BDT input variables are shown in figure 4.11.



(a) Correlation matrix for all the MC Signal BDT input variables.

(b) Correlation matrix for all the Data BDT input variables in the Sideband region.

Figure 4.11: Visual representation of the correlation matrices for the input variables for the two BDT classes.

The correlation matrices show the linear correlations between two variable distributions. In training a BDT, it is important to reduce correlations as much as possible, in order to prevent non-explainability in the nodes of the model. What's more, the importance between variable distributions increase when they are highly correlated or anti-correlated, complicating the training process as two correlated variables have higher weights than uncorrelated variables in the training process, weakening the bonds between other uncorrelated variables which leads to less explainability. The variables that were used in training the BDT model are shown in table 4.13.

Number	Variable	Number	Variable
1	$p(B^0)$	6	Iso(μ_1)
2	$cos_{2D}\alpha$	7	Iso(μ_2)
3	$\eta(B^0)$	8	$p_T(K^0_S)$
4	sig(L_{xy})	9	$\eta(K^0_S)$
5	Δz	10	IP(K^0_S, 3D)

Table 4.13: BDT training variables. Brief explanations of the variables are found in section 4.2.3

Tables 4.14 and 4.15 show information of sample sizes depending on the training q^2 region. The MC samples used from now on throughout this thesis are always Biased unless stated otherwise, therefore that label will be dropped. Furthermore, from the Run 3 Data 2022 samples only parts 0 and 1 were used across all eras, and for prediction parts 2 to 7 were used.

Type	Training	
Sample Label	MC $\mu\mu K^0_S$	Data 2022
pre-Collection 6	10,989K	6,344K
post-Collection 6	575K	1,216K
BDT Region	low-q^2	SB low-q^2
Entries	117K	75K
BDT Region	high-q^2	SB high-q^2
Entries	158K	22K
BDT Region	low+high-q^2	SB low+high-q^2
Entries	274K	98K

Table 4.14: Various sample information regarding the BDT training procedure. SB refers to the Sideband region.

Type	Prediction			
Sample Label	MC $\mu\mu K^0_S$	MC $J/\psi K^0_S$	MC $\psi(2S)K^0_S$	Data 2022
pre-Collection 6	38,479K	11,190K	6,786K	19,021K
post-Collection 6	2,027K	574K	372K	3,644K
BDT Region	low-q^2	J/ψ-q^2	$\psi(2S)$-q^2	SB low-q^2
Entries	411K	564K	359K	224K
BDT Region	high-q^2	J/ψ-q^2	$\psi(2S)$-q^2	SB high-q^2
Entries	554K	564K	359K	67K
BDT Region	low+high-q^2	J/ψ-q^2	$\psi(2S)$-q^2	SB low+high-q^2
Entries	965K	564K	359K	292K

Table 4.15: Various sample information regarding the BDT prediction procedure. SB refers to the Sideband region.

The output of the prediction is the $o(x)$ BDT Output Distribution, where x is the value of the BDT score assigned to predicted values. Such distributions are shown in figures 4.12, 4.13a and 4.13b

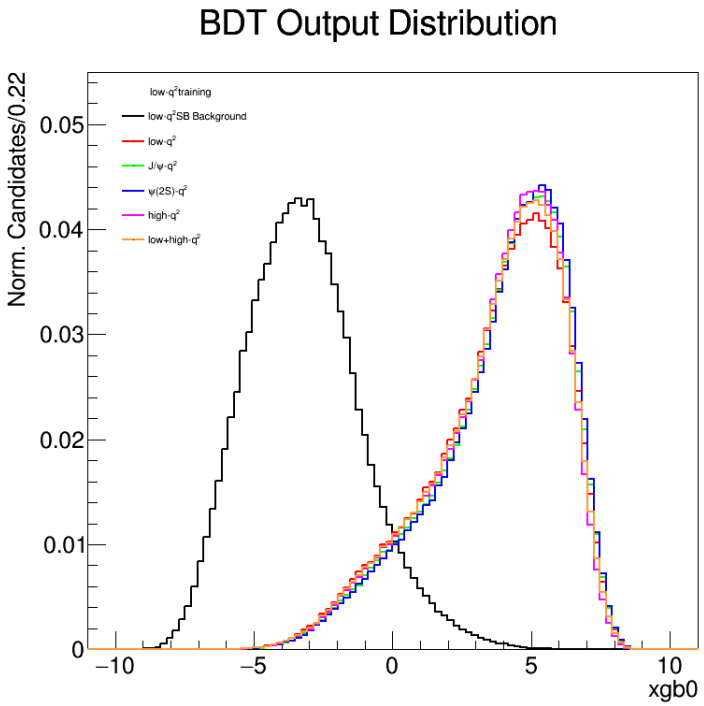
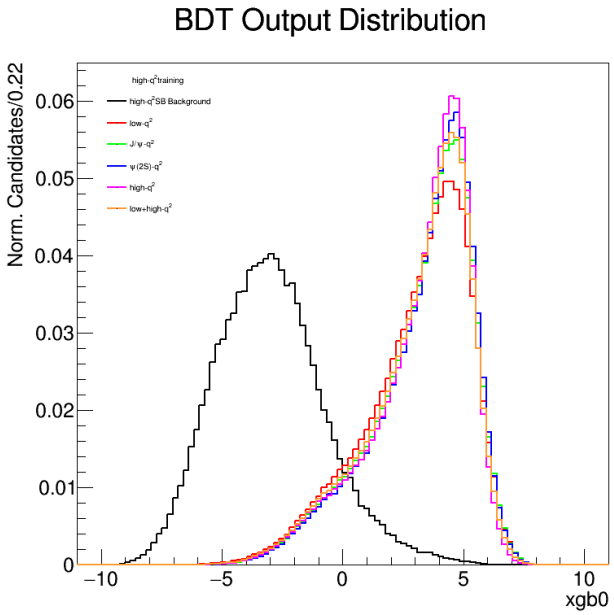
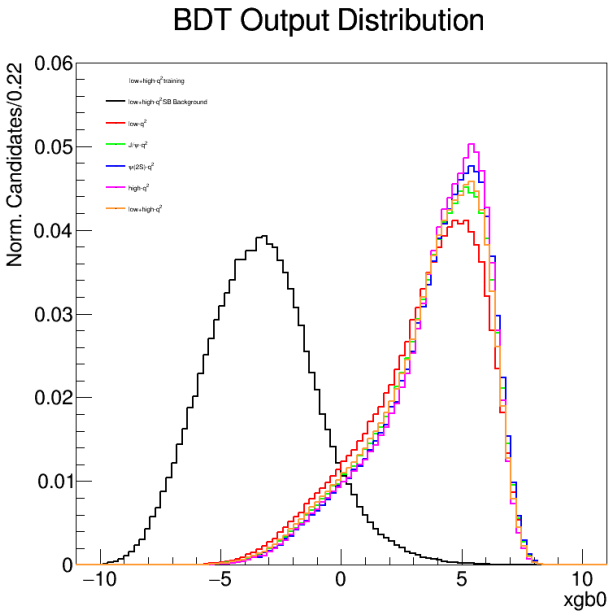


Figure 4.12: BDT Output plot for the low-q^2 trained BDT featuring the low-q^2 Sideband Background as well as all regions from table 4.11. This region was selected in the scope of this analysis.



(a) BDT Output Distributions for the high-q^2 trained BDT featuring the high-q^2 Sideband Background as well as all regions from table 4.11.



(b) BDT Output Distributions for the low+high-q^2 trained BDT featuring the low+high-q^2 Sideband Background as well as all regions from table 4.11.

Figure 4.13: BDT Output plots for the high-q^2 and the low+high-q^2 regions. These training regions were rejected from being used in this analysis.

The analysis to pick the proper q^2 region continues via the use of the BDT Efficiency plots in figures 4.14, 4.15a and 4.15b.The efficiency of any of the two class labels i in the q^2 region r given a BDT cut c is calculated as:

$$\epsilon_{i,r}(c) = \frac{\int_{-\infty}^{c} o_{i,r}(x)\,dx}{\int_{-\infty}^{\infty} o_{i,r}(x)\,dx} \quad i \in \{1,2\} \cong \{S,B\}$$

The BDT Efficiency plots show the relationship between $\epsilon_{S,r}(c)$ versus $1 - \epsilon_{B,r}(c)$, the former being Signal Efficiency and the latter being Background Rejection in the Sideband Region of the B^0 mass spectrum. They are crucial in determining how much Signal is preserved and how much Background is rejected given a BDT Output cut $x > c$. The ideal BDT would have a maximum of $1 - \epsilon_{B,r}$ especially for the non-resonant regions at values of $\epsilon_{S,r} \approx 1$. However, in practice, the curves describing the efficiency plots are strictly decreasing. Studying these curves along with the BDT Output distributions for all training regions is important to picking the optimal q^2 region.

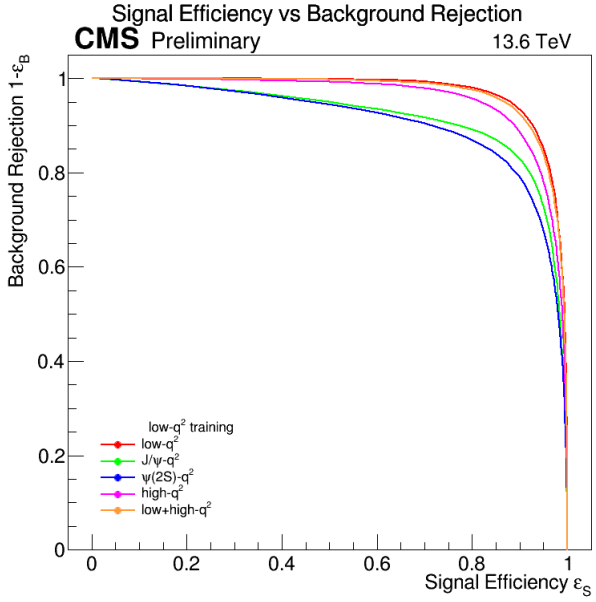
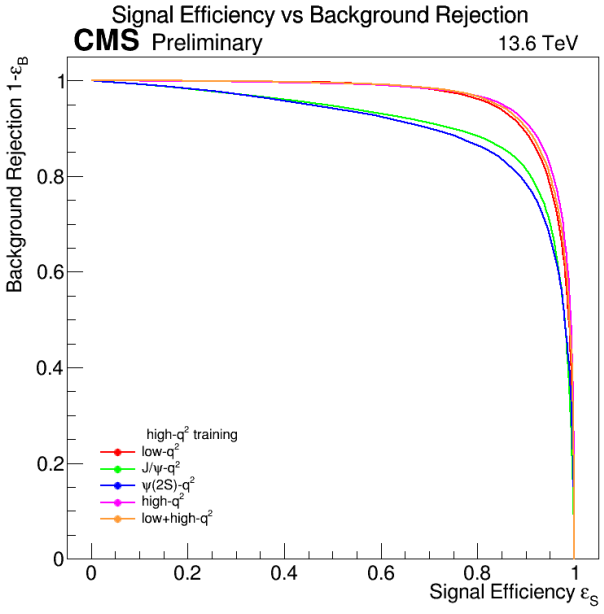
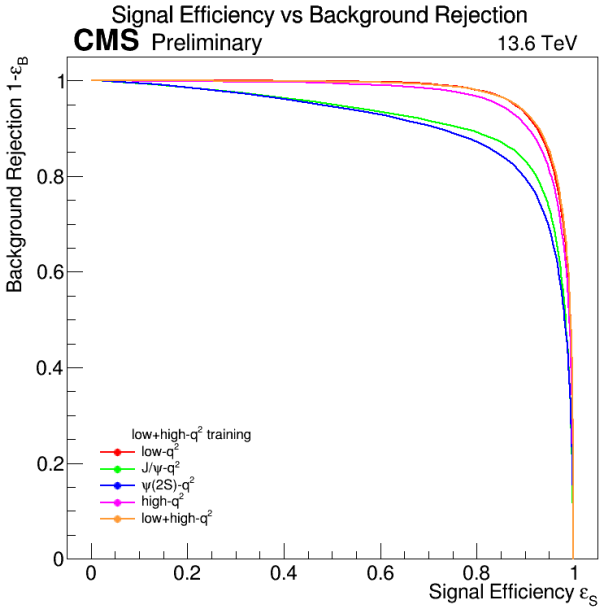


Figure 4.14: Efficiency plots for the low-q^2 region. This region was selected in the scope of this analysis.



(a) Efficiency plots for the high-q^2 region.



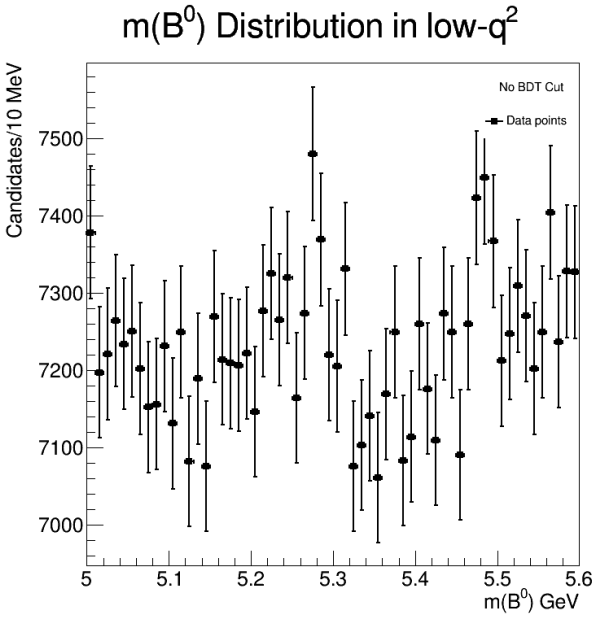
(b) Efficiency plots for the low+high-q^2 region.

Figure 4.15: Efficiency plots for the high-q^2 and the low+high-q^2 regions. These training regions were rejected from being used in this analysis.

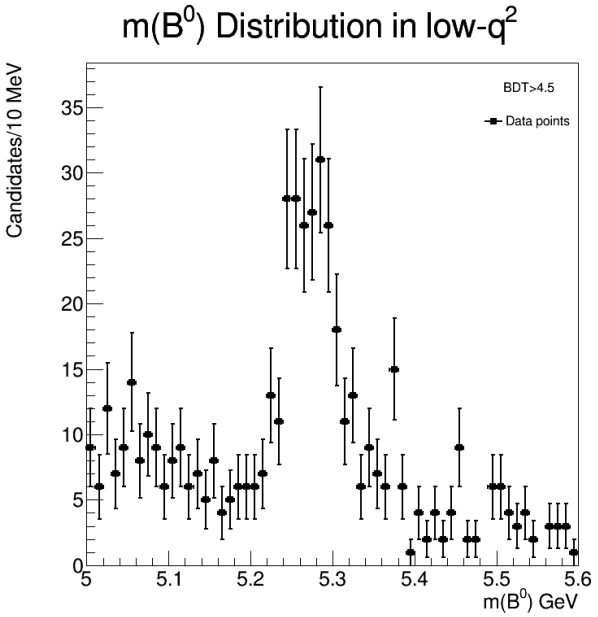
As it can be seen from the BDT Output Distributions in figures 4.12, 4.13a, 4.13b and the BDT Efficiency plots in figures 4.14, 4.15a and 4.15b, four aspects can be denoted:

- The region r the BDT is trained on also has the best predicting value as seen from the efficiency plots.
- The resonant regions underperform in background rejection. This is not an issue as no resonant decays were taken into account in the BDT training procedure, and the BDT predicts these resonant decays as background to the decay mode even in the Sideband region of the B^0 mass spectrum.
- The high-q^2 BDT Output Distribution has a different shape than the low-q^2 BDT Output Distribution when trained on the high-q^2 and low-q^2 regions respectively. While the MC samples do not include any high $c\bar{c}$ resonances with low amplitudes, the Sideband Background region does which could affect the prediction of the MC Samples of the Rare Mode and the two resonant modes.
- The low+high-q^2 region is predictive between the low-q^2 and high-q^2 region when the BDT is not trained on the low+high-q^2 region. This means that this region can be treated like an average - low-q^2 trained BDTs make the high-q^2 region underperform and vice versa. The combination region low+high-q^2 includes information about the high $c\bar{c}$ resonances and its increased statistics do not make a better prediction across all three different training regions for the BDT.

Now that some analysis was done in order to pick a proper q^2 region, the best choice would be the low-q^2 region due to the lack of low amplitude resonances in the dimuon q^2 mass spectrum. Furthermore, preliminary analysis in B^0 mass fits show that the 224K Data sample entries (table 4.15) of the low-q^2 BDT suffice. The 67K entries of the high-q^2 BDT do not increase the statistics of the Data samples by a large margin and induce further complications into the training procedure, prediction and the mass fit of the B^0 meson. Therefore, the high-q^2 and low+high-q^2 BDTs are dropped from this analysis going forward. The effect of applying a BDT cut to the Data samples is demonstrated in figures 4.16, 4.17 and 4.18.

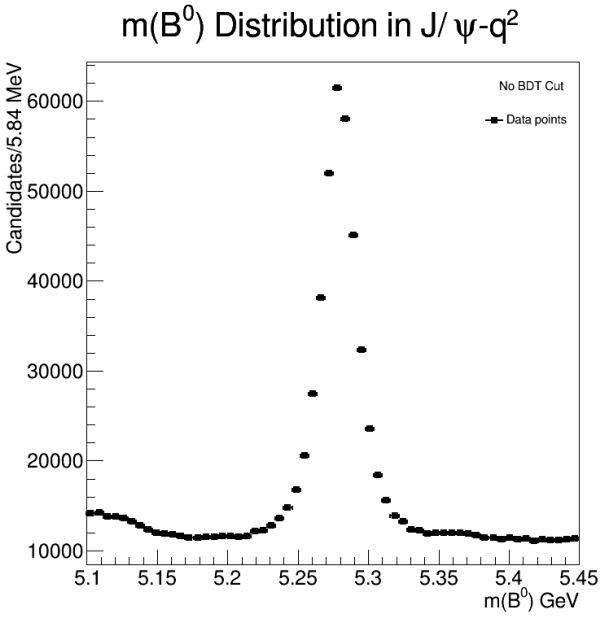


(a) $m(B^0)$ Distribution using $x > -\infty$.

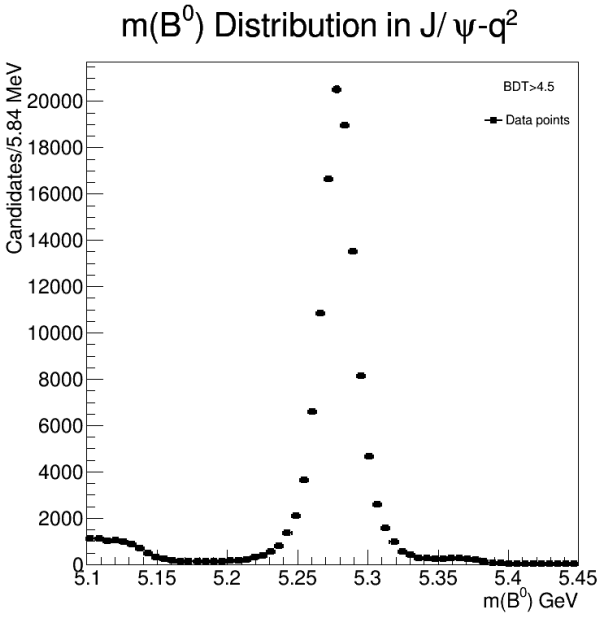


(b) $m(B^0)$ Distribution using $x > 4.5$.

Figure 4.16: $m(B^0)$ Mass Distribution in the Data Samples for $x > -\infty$ and $x > -4.5$ in the low-q^2 region.

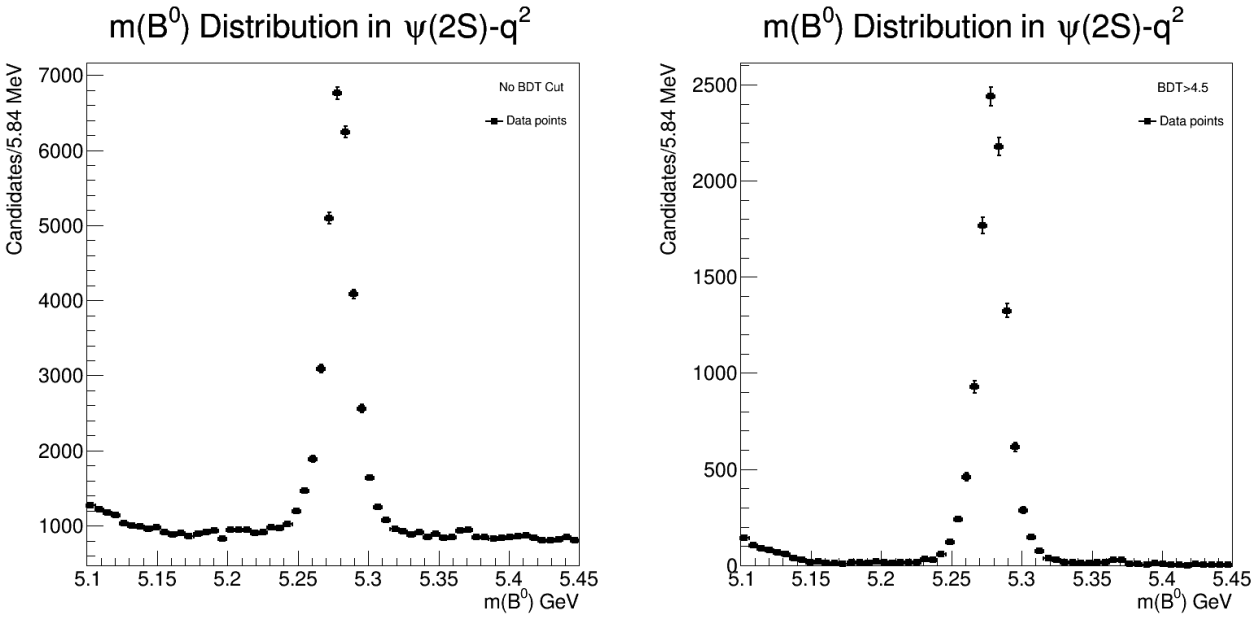


(a) $m(B^0)$ Distribution using $x > -\infty$.



(b) $m(B^0)$ Distribution using $x > 4.5$.

Figure 4.17: $m(B^0)$ Mass Distribution in the Data Samples for $x > -\infty$ and $x > -4.5$ in the J/ψ-q^2 region.



(a) $m(B^0)$ Distribution using $x > -\infty$. *(b) $m(B^0)$ Distribution using $x > 4.5$.*

Figure 4.18: $m(B^0)$ Mass Distribution in the Data Samples for $x > -\infty$ and $x > -4.5$ in the $\psi(2S)$-q^2 region.

4.6 B^0 Mass Fits

The Data samples described in table 4.15 have had a prediction score given a BDT describing their class. These Data samples were used to fit on the B^0 mass in various q^2 regions: low-q^2, J/ψ-q^2 and $\psi(2S)$-q^2. The former of the three regions is to search for the Rare Mode decay $B^0 \to \mu^+\mu^- K^0_S$ and the other two regions are fits for the two resonant regions. The yields that will be acquired from the fits N_{exp} are required in the determination of the differential branching function dB/dq^2 of the Rare Mode (eq. 3.2) as well as the $R_{K^0_S}(\psi(2S))$ ratio, (eq. 3.3). The J/ψ resonance has a large number of statistics in its q^2 region, as shown in figure 4.10. The $\psi(2S)$ resonance includes a lesser number of statistics relative to the First Resonant mode but more compared to the Rare Mode. The fits on the B^0 mass windows of the two resonant regions mass are performed as a sanity check to check the validity of the results, given that the decays of these two intermediate decays are well known. The Fitting procedure is described by the following steps:

1. Fit the B^0 Mass Distribution in the MC samples of the Rare Mode and the two resonant regions. Save the distributions for the next step.

2. Slightly adjust the fitting range of the fitting parameters of B^0 each mass distribution as determined in step 1 and refit on the Data samples along with additional distributions for other background processes.

3. Perform the sPlot technique (section 4.7) to extract the yields of the decays.

Before showcasing the values of the parameters of the distributions used in the fits, it is important to denote different background types:

- *Partially Reconstructed Background* are background processes whose decay products are captured by the analysis. This analysis requests three different tracks in the final state, being the two μ^+, μ^- and a neutral pseudotrack K^0_S under the Collection 6 variable cuts. Some processes, like the $B^0 \to K^*(K^0_S\pi)\mu\mu$ decay produce four or more

tracks. Not reconstructing all of the decay products but three has a consequence on the distribution of the B^0 mass and particularly its mean for that decay mode; it is shifted to the left of the expected B^0 mass due to the missing particles. A notable example in this analysis is the K^* decay mode of the B^0.

- *Combinatorial Background* is a type of background that regards misclassification of tracks. The CMS experiment does not measure particles but tracks. If the track fit is appropriate with some mass and momentum, a particle may be associated with it. However, tracks of bad quality allow for a larger range of values for both mass and momentum even after passing the corresponding trigger. For example, a charged pion could be classified as a muon and therefore become evident in the Data samples for the analysis.

4.6.1 B^0 Mass Distributions

The elementary distributions and combinations thereof used in the B^0 mass fits using the RooFit library of ROOT are shortly explained in this section. The elementary distributions are the *Exponential Decay*, $E(x;\tau)$ *Gaussian*, $G(x;\mu,\sigma)$ *Crystal Ball (CB)* and *Double Sided Crystal Ball (DSCB)*. The last two distributions share the non-fittable parameters which are used at the tails of the distributions, A_i and B_i:

$$A_i = \left(\frac{n_i}{|\alpha_i|}\right)^{n_i} \exp\left(-\frac{\alpha_i^2}{2}\right), \quad B_i = \frac{n_i}{|\alpha_i|} - |\alpha_i| \quad i \in \{1,2\}$$

The Crystal Ball function $C(x;\alpha_1,n_1,\mu,\sigma)$ is defined as:

$$\begin{aligned} C(x;\alpha_1,n_1,\mu,\sigma) &= A_1\left(B_1 - \frac{x-\mu}{\sigma}\right)^{-n_1} && -\infty < \frac{x-\mu}{\sigma} \leq -\alpha_1 \\ C(x;\mu,\sigma) &= \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) && -\alpha_1 < \frac{x-\mu}{\sigma} < +\infty \end{aligned}$$

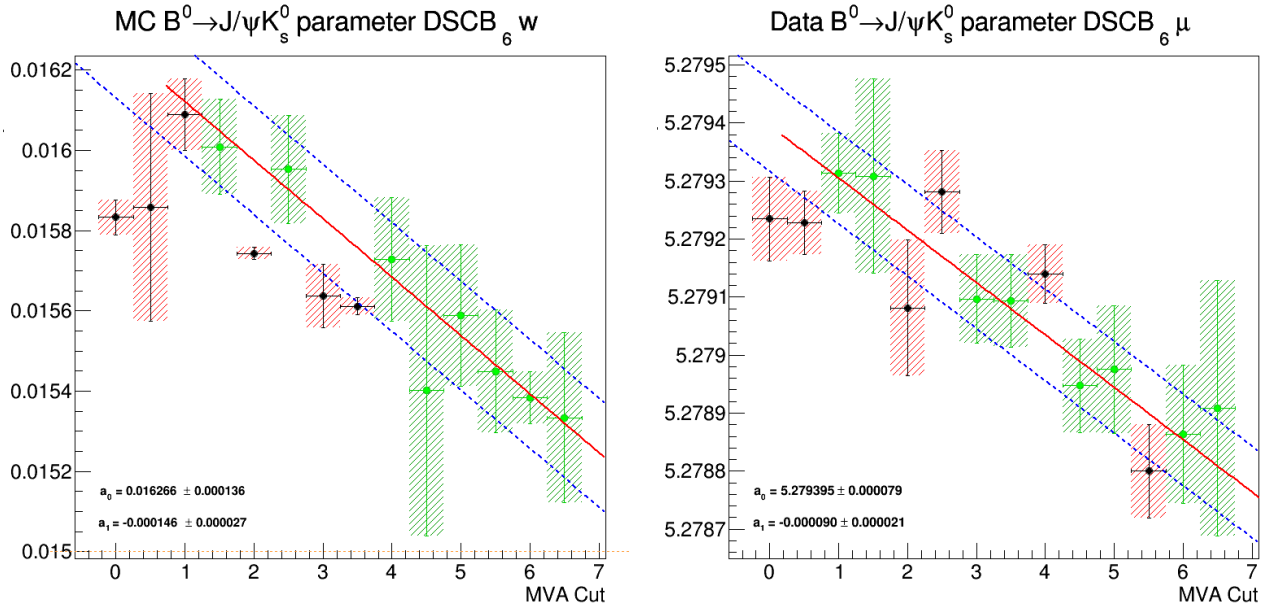
And the Double Sided Crystal Ball function $D(x;\alpha_1,n_1,\alpha_2,n_2,\mu,\sigma)$ is defined as:

$$\begin{aligned} D(x;\alpha_1,n_1,\mu,\sigma) &= A_1\left(B_1 - \frac{x-\mu}{\sigma}\right)^{-n_1} && -\infty < \frac{x-\mu}{\sigma} < -\alpha_1 \\ D(x;\mu,\sigma) &= \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) && -\alpha_1 < \frac{x-\mu}{\sigma} < a_2 \\ D(x;\alpha_2,n_2,\mu,\sigma) &= A_2\left(B_2 + \frac{x-\mu}{\sigma}\right)^{-n_2} && a_2 < \frac{x-\mu}{\sigma} < +\infty \end{aligned}$$

The CB and DSCB functions are continuous at the piece-wise points $x = \{-\alpha_1, \alpha_2\}$ as well differentiable, making them able to be used as a fitting function. The use case of these functions is to give more freedom to the fitting region of the tails of "Gaussian-like" distributions, taking into effect various detector effects. The CB function has a Gaussian core and right tail for $\frac{x-\mu}{\sigma} > -\alpha_1$ and the DSCB function has a Gaussian core between the bounds $-a_1$ and a_2. The CB and DSCB functions have the additional advantage of being very customizable. This is attributed to the form of their corresponding expressions at the tail regions due to the parameters n_i and α_i. Each distribution, including the Gaussian and Exponential Decay is normalized after it is fitted to range of the B^0 mass window. In general, the mass fits are performed by adding distributions together using the total fit function $T(x;\vec{\theta})$ using any of the elementary fit functions $F_i(x;\vec{\phi}_i)$ or combinations thereof described previously:

$$T(x;\vec{\theta}) = \sum_i^N r_i F_i(x;\vec{\phi}_i) \tag{4.5}$$

where r_i is the relative yield of the fit component distribution $F_i(x;\vec{\phi}_i)$, which limits the yield number of a distribution relative to another. The choice of the relative yield value leads to fit convergence as long as the range of the fitting parameters ϕ_i of each elementary distribution allow the parameters to converge to a stable value. Therefore, the relative yield values are initially picked and then the fitting ranges for the fitting parameters are optimized. These relative yields are converted into total yields for a fit component distribution once the total fit distribution $T(x;\vec{\theta})$ is normalized after the fit. In order to find optimal parameter ranges for various BDT output cuts, an extensive analysis was performed. By taking the values of the parameters of a B^0 mass fit of a particular distribution across different values of a BDT output cut $x > c$, a linear fit[24] $a_1x + a_0$ was performed in order to determine their ranges as $[a_1c + a_0 - \sigma(a_0), a_1c + a_0 - \sigma(a_0)]$ at the BDT Output cut $x = c$. Then, the fitting process on the B^0 mass was performed again and the parameter range was re-optimized until the B^0 mass fits converged to a satisfying degree, with $\chi^2/N_{\mathrm{DOF}} \sim 1$.



(a) Example of parameter range optimization.

(b) Example of parameter range optimization.

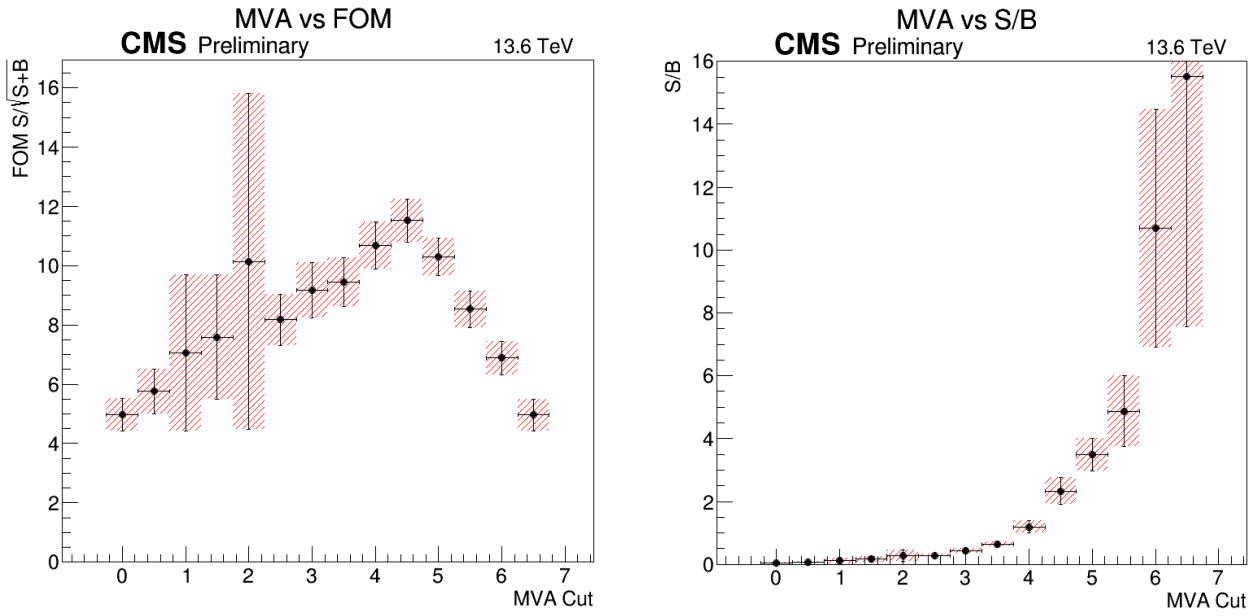
Figure 4.19: Examples of parameter range optimization for two different parameters. The green points were only used in determining the new fitting parameter ranges. The choice for whether a point should be used or rejected was due to large deviations or extremely small errors in the fitting parameter, often a sign of a bad fit. The legend in the bottom left corner shows the values of the linear fit $a_0 + a_1x$.

4.6.2 Figures of Merit

The choice of the appropriate set of mass fits given a cut $x > c$ on the BDT Output value is given by a *figure of merit*. The two figures of merit used in this analysis are the *Significance* $S/\sqrt{S+B}$ and S/B where S, B are the yields corresponding to two σ away from the μ of the DSCB fit function. The maximum value of the Significance $S/\sqrt{S+B}$ determines the appropriate set of mass fits to be used in extracting the yields. If no maximum value of the Significance exists, the choice is then determined using the maximum value of S/B. This Figure of Merit (FoM) is only applied to the Rare Mode mass fits. The plots

[24]This fit was not performed by RooFit but using the core built-in ROOT library Minuit.

showcasing the two FoMs are shown in figure 4.20.



(a) Significance $S/\sqrt{S+B}$ FoM.

(b) S/B FoM.

Figure 4.20: The Significance (left) and S/B (right) FoMs for the Rare Mode Data sample. The S/B FoM has no maximum. MVA is the BDT cut.

In figure 4.20a, the Significance figure of merit has large errors for $1 \leq x \leq 2$. This is due to the existence of large background and little signal presence. Thankfully, that area of BDT cuts contains a lot of background and fit parameter optimization in these regions was not given as much of a weight as higher BDT cuts. In figure 4.20b, The error is increasing for $x \geq 4$, and this is attributed to the low number of statistics present in the Data samples, as these BDT cuts are very tight in rejecting events. This leads to lower signal and background yields with large errors provided by the fit which carry over in the calculation of S/B.

4.6.3 Strategy and Results

The strategy of the fitting procedure in order to determine the best cut for the BDT prediction score $x > c_b$ will now be briefly described:

1. Pick a BDT Output cut $x > c$ and use the MC samples for the Rare Mode and the two Resonances, along with an additional MC Sample of $B^0 \to K^*(K^0_S\pi)\mu\mu$ decays.

2. Fit on the B^0 mass window. This window is different in each sample (Rare Mode and the two resonant q^2 regions). Each fit uses two distributions: a DSCB plus a Gaussian, except in the case of the K^* MC sample, where a DSCB plus a CB is used. The role of this fit is to acquire approximate values of the fitting parameters in order to be used in the Data sample fitting. The Rare Mode q^2 region is the low-q^2, and the two resonances J/ψ and $\psi(2S)$ correspond to their labeled q^2 regions.

3. Take the width parameter values as determined in step 1 and allow deviations up to 5% of their initial value (the means of the distributions have a deviation limit up to 1%) for the two resonant modes and the K^* MC sample and 30% for the widths of the two distributions in the Rare Mode.

4. Fit the B^0 mass window in the Rare Mode Data samples.

5. Use the parameters of the peak in the fitted B^0 mass window in the Rare Mode Data samples to fit a $B_S^0 \to \mu\mu K_S^0$ background process as well as the rest of the mass fit in the resonant modes, which is composed of their corresponding MC sample plus the K^* MC sample.

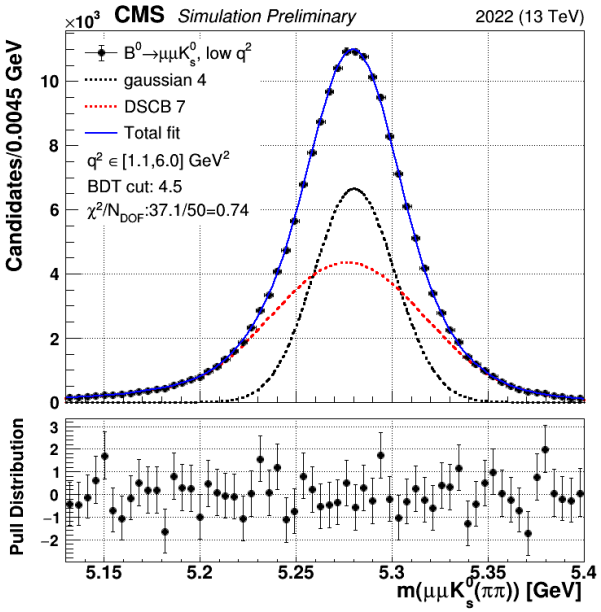
6. Perform the sPlot technique (section 4.7).

7. Save the yields and values of parameters in each step. Identify bad fits ($\chi^2/N_{\text{DOF}} \gg 1$) and tweak the fit parameters as explained (figure 4.19) and repeat the process to a satisfactory degree.

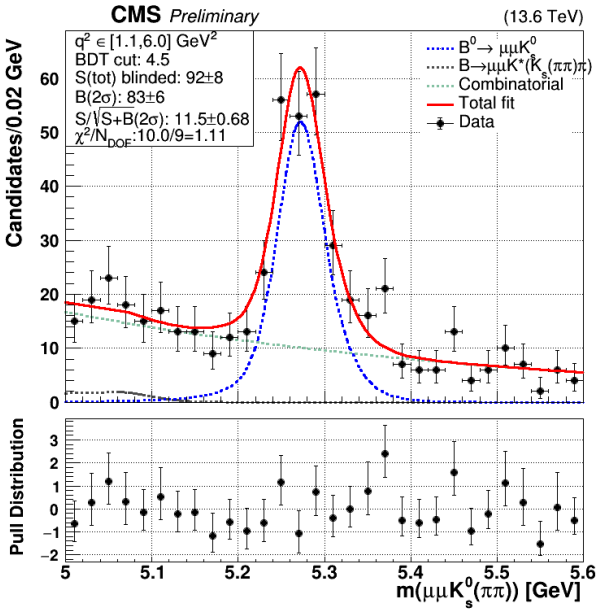
8. Repeat this process for a range of BDT Output cuts $x > c$, close to the MC Signal Distribution (minimum cut is at $x > 0$, right distribution in figure 4.12).

9. Pick the best value of the BDT Output cut c_b using the Figures of Merit, described in section 4.6.2 according to the Rare Mode B^0 mass fit on the Data samples.

According to the values of the FoMs in figure 4.20, there is a maximum c_b BDT cut determined in the last step in the enumeration above, which is equal to $c_b = 4.5$. This value is picked to determine the appropriate choice of the B^0 mass fits across all three modes, as well as determining the efficiency ϵ_7 of the Collection 7 of variable cuts, which is used in determining the differential branching function, as well as the $R_{K_S^0}(\psi(2S))$ observables.

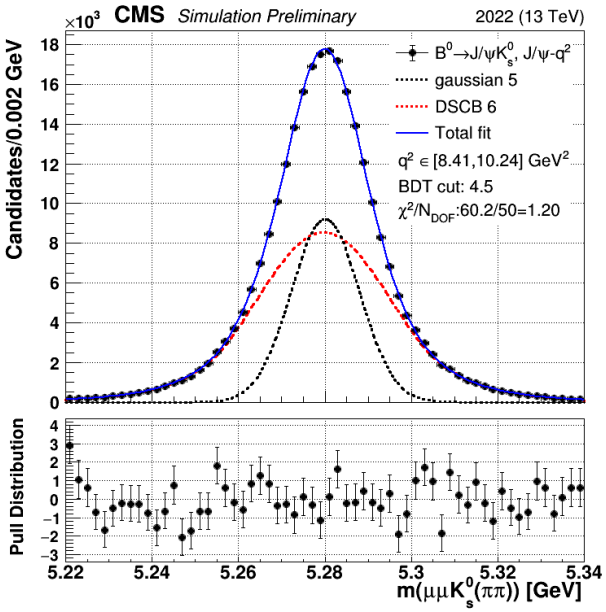


(a) MC sample B^0 mass fit in the low-q^2 region of the Rare Mode $B^0 \to K_S^0\mu\mu$.

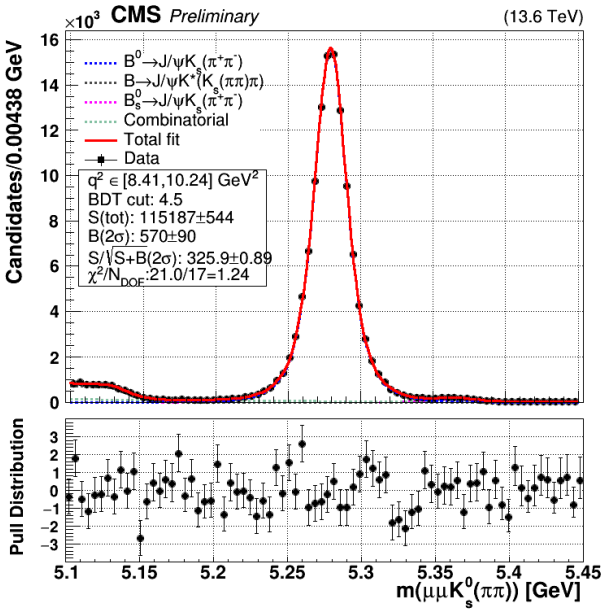


(b) Data sample B^0 mass fit in the low-q^2 region of the $B^0 \to K_S^0\mu\mu$.

Figure 4.21: MC and Data sample B^0 mass fits for the Rare Mode in the low-q^2 region.

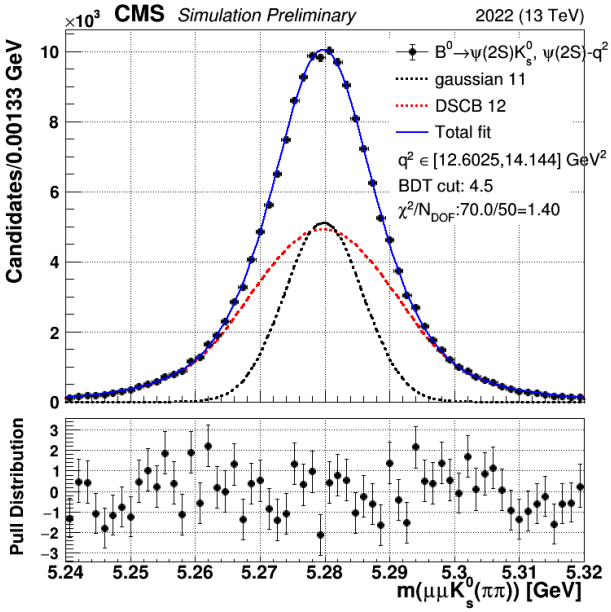


(a) MC sample B^0 mass fit in the $J\psi$-q^2 region of the Resonant Mode $B^0 \to J/\psi K^0_S$.

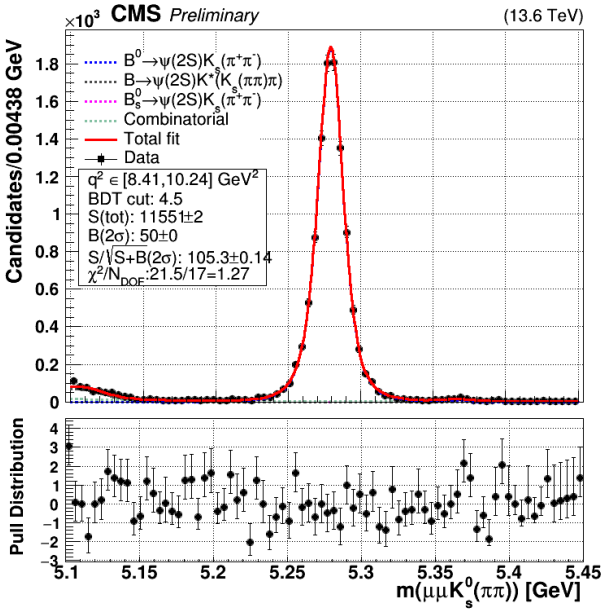


(b) Data sample B^0 mass fit in the $J\psi$-q^2 region of the Resonant Mode $B^0 \to J/\psi K^0_S$.

Figure 4.22: MC and Data sample B^0 mass fits for the Resonant Mode in the $J\psi$-q^2 region

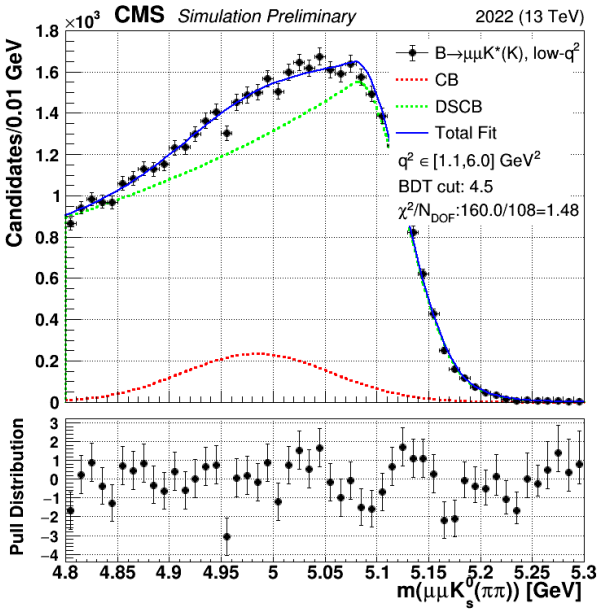


(a) MC sample B^0 mass fit in the $\psi(2S)$-q^2 region of the Second Resonant Mode $B^0 \to \psi(2S)K^0_S$.

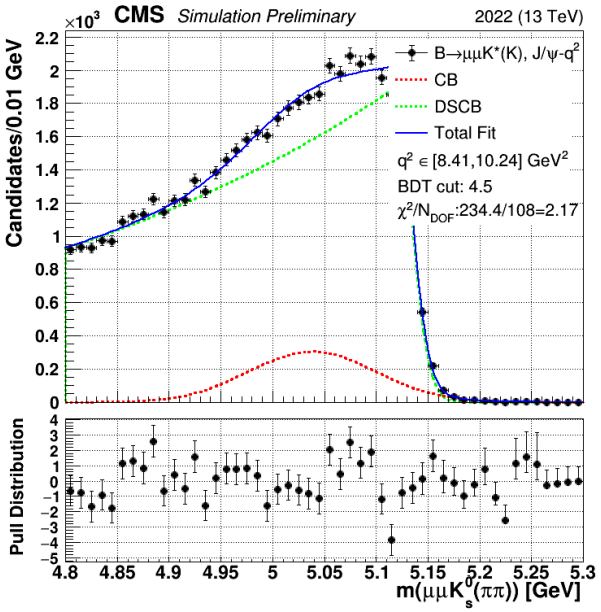


(b) Data sample B^0 mass fit in the $\psi(2S)$-q^2 region of the Second Resonant Mode $B^0 \to \psi(2S)K^0_S$.

Figure 4.23: Data sample B^0 mass fits for the two resonant modes in their corresponding q^2 regions.



(a) MC sample B^0 mass fit in the low-q^2 region of the $B^0 \to K^*(K^0_S\pi)\mu\mu$ partially reconstructed background process.



(b) MC sample B^0 mass fit in the J/ψ-q^2 region of the $B^0 \to K^*(K^0_S\pi)J/\psi$ partially reconstructed background process.

Figure 4.24: Data sample B^0 mass fits for the two resonant modes in their corresponding q^2 regions.

Therefore, according to equation 4.5, the MC sample B^0 mass fits are all fitted on the composite fit function P, which is composed by a Gaussian and a Double Sided Crystal Ball function:

$$P(x;\vec{\phi}_m) = G(x;\mu_{Gm},\sigma_{Gm}) + r_i D(x;\alpha_{1m},\alpha_{2m},n_{1m},n_{2m},\mu_{Dm},\sigma_{Dm})$$

where $m \in \{1,2,3\}$ corresponds to the fit function of the Rare Mode, Resonance and Second Resonance respectively. In the case of the MC samples of the partially reconstructed K^* background, the fit function is composed by a Crystal Ball and Double Sided Crystal Ball:

$$L(x;\vec{\phi}_m) = C(x;\alpha_m,n_i,\mu_{Cm},\sigma_{Cm}) + r_i D(x;\alpha_{1m},\alpha_{2m},n_{1m},n_{2m},\mu_{Dm},\sigma_{Dm})$$

where $m = \{4,5,6\}$ regards the corresponding K^* background to the Rare Mode, Resonant and Second Resonant final states. Finally, for the Data fits, the total fit functions are composed of an exponential decay function E for the combinatorial background, the L composite function of the partially K^* background, and two terms of the composite function P, one for the signal and one for an additional B^0_S background in the Resonant Regions only:

$$T(x;\vec{\theta}_m) = N_{Pm}P(x;\vec{\phi}_m) + N_{L\,m+3}L(x,\vec{\phi}_{m+3}) + N_{Em}E(x;\tau_m) + (1-\delta_{1m})N_{Pm}P(x;\vec{\phi}'_m) \qquad (4.6)$$

where δ_{ij} is the Kronecker delta and the last term in the total fit function, $P(x;\vec{\phi}'_m)$ corresponds to distinct background process not seen in the Rare Mode, the B^0_S background process. The fit function of this process is identical to the MC of the corresponding mode, with two changes. The mean of the Gaussian is set at the right of the signal peak, and the width of the Gaussian may vary between 50% and 130% of the fitted value of the MC sample fit function $P(x;\vec{\phi}_m)$. The N_P, N_L, N_E, N_B factors are the yields of each fit component. The reason for the absence of the B^0_S background process in the Rare Mode B^0 mass fit on the Data samples is Physics related, promoted by the small number of statistics; its yield must be constrained to 1% of the value of the yield of the Rare Mode. This is due to quark

fragmentation when hadronizing the B^0 and B_S^0 mesons. This constraint is given by the ratio of fragmentation factors times the ratio of the branching ratios[25] of the two decay modes. The constraint λ is equal to:

$$\lambda = \frac{f_s}{f_d} \cdot \frac{\mathrm{BR}(B_S^0 \to \mu\mu K_S^0)}{\mathrm{BR}(B^0 \to \mu\mu K_S^0)} = 0.009 \approx 1\%$$

where $f_s = 0.100 \pm 0.008$, $f_d = 0.408 \pm 0.007$ are the quark fragmentation factors for strange type B mesons: it is a measure of the probability of collecting a strange or down quark given a b quark is available to form a B_S^0 or B^0 respectively. The ratio of branching ratios computes to a value of 0.04. This constraint on the yield of the B_S^0 effectively hides it into the combinatorial background. To avoid fit convergence issues and errors in the sPlot technique due to this small value of the yield, it is not included in the Rare Mode mass fit. In the case of the two resonant mode B^0 mass fits, the yield of the B_S^0 is *floating*; it is not constrained to a value.
Lastly, the pull distribution $p(m)$ at bin k at the bottom of the mass fits is calculated as:

$$p_k(m) = \frac{N_k - T(m;\vec{\theta})}{\sigma(N_k)}$$

where N_k is the number of entries in the $m(B^0)$ distribution at bin k, $\sigma(N_k)$ is the statistical uncertainty at bin k and $T(m;\vec{\theta})$ is the value of the total fitted mass distribution. The ideal distribution would be a uniform distribution across the B^0 fitted mass range with a value of 0 and as small statistical uncertainties as possible. In the legend of any fitted mass of a Data sample, the Significance FoM is calculated through the measured yields. In the Rare Mode mass fits legends, the signal yield `S(tot)` is written as **blinded**, which means that the total signal yield is multiplied by a random number. The reason for doing this is to prevent known published results (such as known variable cuts) from other analyses determining the result of this analysis.

4.7 The sPlot Technique

sPlot[62] is a statistical tool dedicated to the exploration of datasets. This technique assigns a weight to each event in the Data samples according to the fit components $F_i(x;\vec{\phi}_i)$ of the total fit function $T(x;\vec{\theta})$ used on the B^0 mass fit, instead of labelling each event as being signal-like or background-like. While the labeling would be effective for events falling many means way from the distribution of the signal region, events that fall a few means away instead would be incredibly ambiguous to be labeled. This means that the whole Data sample can be used in plotting various physics quantity distributions that describe the signal and each entry will be weighted accordingly. sPlot also introduces the notion of *discriminating* and *control* variables.

The set of discriminating variables y is assumed to have their physics quantity distributions (such as η, p_T) known, whereas the set of control variables z does not, or at least is assumed not to be. It is convenient to write that the total variable set x (which is described in section 4.2.3) is equal to $y \cup z$. The foundation of the sPlot technique is attributed to the Maximum Likelihood method which extracts parameters from a data sample, along with the knowledge of the behavior of the discriminating variable set y only, which is uncorrelated to the control variable set z. The log-Likelihood is expressed

[25] which is also referred to as Cabbibo suppression.

as:

$$\mathcal{L} = \sum_{e=1}^{N} \ln\Big(\sum_{i=1}^{N_s} N_i F_i(y_e)\Big) - \sum_{i=1}^{N_s} N_i$$

where

- N is the total number of events in the Data samples.
- N_s is the number of fit components.
- N_i is the yield of the ith component. Proper normalization of the fit components F_i in the fitting state (section 4.6.1) converts the relative yield r_i into the yield of the ith component.
- y is the set of discriminating variables. Not to be confused with the total variable set x and control variable set z which are absent explicitly.
- f_i is the Probability Density Function (PDF) or fit component of the discriminating variables for the ith species. The value $F_i(y_e)$ is the value of the PDF F_i for the set of discriminating variables y at event e.

The weights $w_i(y_e)$ of the event e for the fit component i in this case are calculated by[62]:

$$w_i(y_e) = \frac{\sum_{j=1}^{N_s} V_{ij} F_j(y_e)}{\sum_{j=1}^{N_s} N_j F_j(y_e)}$$

where V_{ij} is the covariance matrix. The elements of the inverse covariance matrix V_{ij}^{-1} are given by:

$$V_{ij}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_i \partial N_j} = \sum_{e=1}^{N} \frac{F_i(y_e) F_j(y_e)}{\left(\sum_{k=1}^{N_s} N_k F_k(y_e)\right)^2}$$

Using these weights, one can fill a weighted histogram of a control variable. The sPlot technique also assigns uncertainties to binned histograms. The uncertainty σ_i^l of the set of control variable distributions z on the bin l is given by:

$$\sigma_i^l = \sqrt{\sum_{e \text{ in bin } l} \Big(w_i(y_e)\Big)^2} \tag{4.7}$$

In this analysis, the set of discriminating variables y includes the B^0 mass $m(B^0)$ only, making the the control variable set z equal to $x/\{m(B^0)\}$. sPlot will be an effective technique only if $m(B^0)$ is uncorrelated with the rest of the control variables z. Figure 4.25 shows correlation of several variables in the Data samples that are used in the B^0 mass fits. It can be seen that all variables but one are uncorrelated. Therefore, the sPlot technique will be performed for these variables using the signal fit component function. The distributions for some of these variables can be seen in Appendix A. From this point on, if a reference to the Data samples using sPlot has to be made, those samples will be referred to as Weighted Data samples.

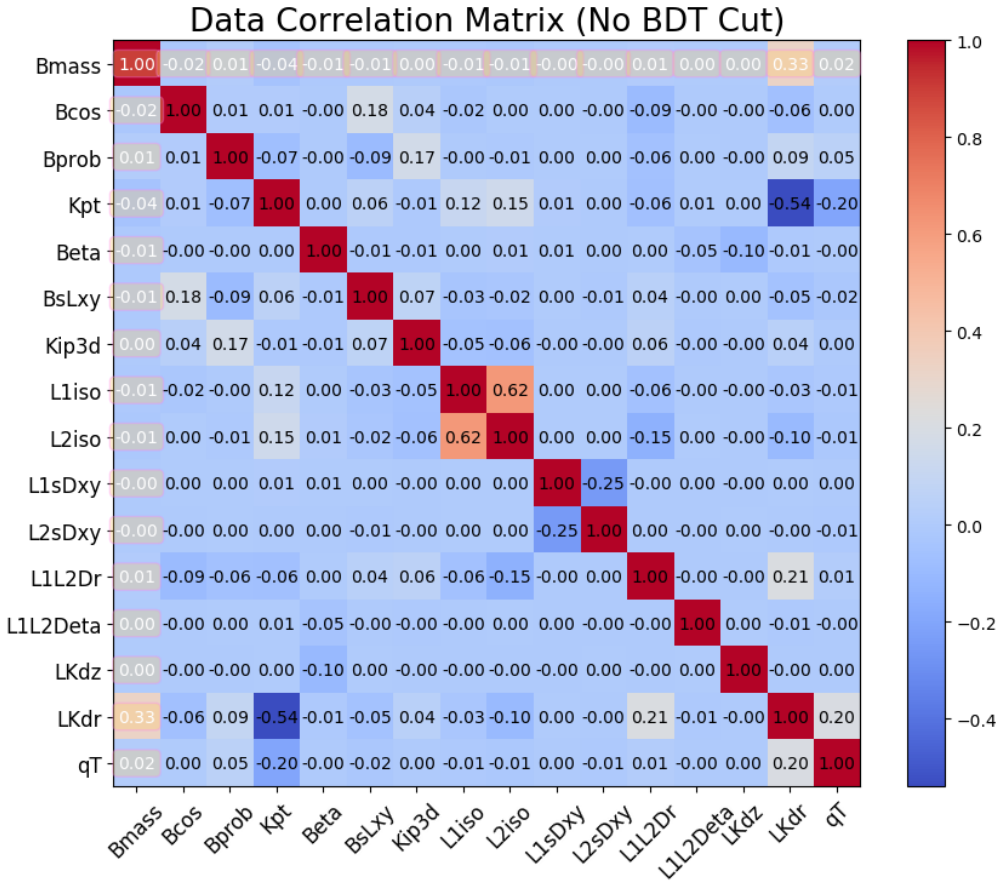
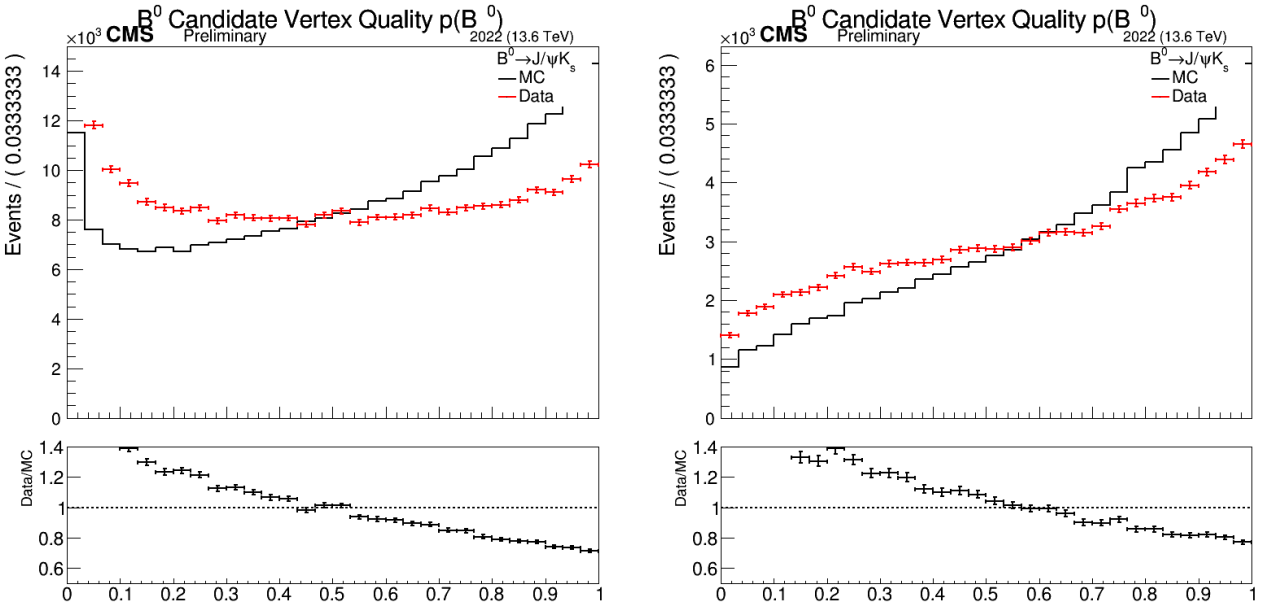


Figure 4.25: Pearson correlation coefficients between two variable distributions for the Data samples used in the B^0 mass fits. The $m(B)$ correlation coefficients are highlighted. No BDT cuts have been applied.

It is important to denote some differences in some variable distributions that become evident due to the sPlot technique; namely the $p(B^0)$ distribution, which is a measure of candidate quality of a reconstructed B^0 meson. Figure 4.26 showcases the difference for two different arbitrary BDT value cuts, $x > -5$ and $x > 5$ in the J/ψ-q^2 Resonant Mode $B^0 \rightarrow J/\psi K^0_S$ region. The differences in the distributions from the Weighted Data samples are easier to distinguish in the Resonant region due to the increased statistics.

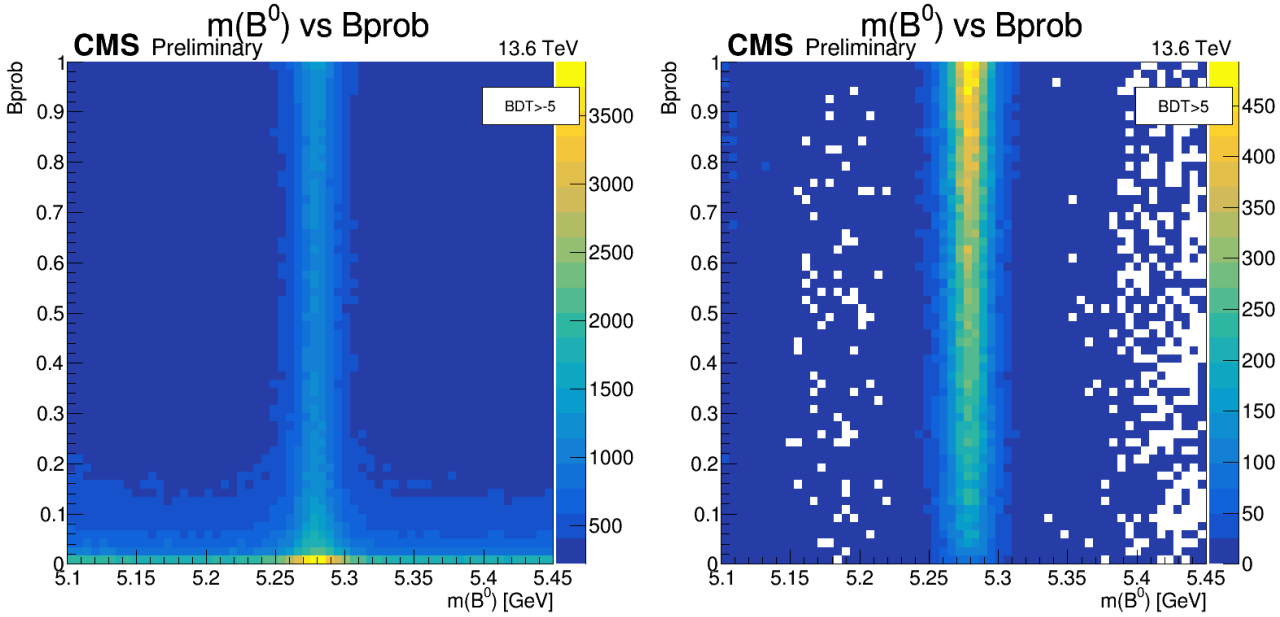


(a) $p(B^0)$ *sPlot of the Resonant Mode* $B^0 \rightarrow J/\psi K^0_S$ *for* $BDT > -5$.

(b) $p(B^0)$ *sPlot of the Resonant Mode* $B^0 \rightarrow J/\psi K^0_S$ *for* $BDT > 5$.

Figure 4.26: Weighted Data sample $p(B^0)$ *sPlot in the* J/ψ-q^2 *region of the Resonant Mode* $B^0 \rightarrow J/\psi K^0_S$. *The ratio plot in the bottom is the division between the weighted Data sample and corresponding MC sample bin contents.*

An extensive analysis has been made to see the divergence of the signal search in the Weighted Data samples acquired from the CMS experiment in contrast to the synthetically generated MC samples. Analysis shows that the distribution of $p(B^0)$ as well as all other variables in the Data samples under the $x > -5$ cut and under no imposition of such cut ($x > -\infty$) is very similar. Furthermore, the $x > -5$ cut marks the minimum value on the BDT value of the MC samples, therefore including lower values than -5 for the BDT cut would also theoretically introduce certain unwanted background processes, contaminating the results. It can be inferred from the sPlots in 4.26 and 4.27 that bad quality B^0 candidates are removed after the BDT is applied. The $p(B^0)$ distribution has shown to be the most different, showcasing that the MC data samples are different to the Data samples, imposing difficulties in the analysis, such as systematic errors which are hard to measure. It is to be reminded that the $p(B^0)$ distribution is one of the BDT training variables (table 4.13), which could also impose further systematics in the BDT scoring metric.



(a) 2D plot of $m(B^0)$ vs $p(B^0)$ of the Resonant Mode $B^0 \to J/\psi K^0_S$ for BDT> -5.

(b) 2D plot of $m(B^0)$ vs $p(B^0)$ of the Resonant Mode $B^0 \to J/\psi K^0_S$ for BDT> 5.

Figure 4.27: Data sample 2D plot of $m(B^0)$ vs $p(B^0)$ in the J/ψ-q^2 region of the Resonant Mode $B^0 \to J/\psi K^0_S$. The application of the BDT model clearly shows the rejection of zero bad quality combinatorial background B^0 candidates.

Half of the Weighted Data sample distributions on the J/ψ Resonant Region show complete divergence from the expected unitary ratio, such as the aforementioned $p(B^0)$, BDT Output $o(x)$, IP(K^0_S, 2D), Iso(μ_1), Iso(μ_2), ΔR and the muon sig(D_{xy}) distributions while other distributions such as sig(L_{xy}) and $p_T(K^0_S)$ show divergences from the unitary ratio distribution in specific regions (see Appendix A.2). The same trend also applies to the $\psi(2S)$ Second Resonant Region (see Appendix A.3). The sPlots of the Rare mode (Appendix A.1) suffer from large uncertainties due to the low number of statistics, making the comparison between the Weighted Data samples and MC samples difficult. Improvements of this sector of the analysis are discussed in section 5.3.

4.8 Simultaneous B^0 Mass Fits

The B^0 mass fits explained in section 4.6 are also performed in all q^2 bins (table 4.1) to increase the Data sample statistics in order to determine the differential branching function dB/dq^2. The result of this analysis is blinded, meaning that a random number γ multiplies the signal yields in each bin. This is to hide the true result and prevent the analysis from being affected by the measured values, as this thesis is also part of an ongoing project. Once the analysis is complete, the analysis unblinds, and the random number factor is removed. The simultaneous mass fit plots are showcased in Appendix B. The term Simultaneous refers to the existence of parameters that are shared by multiple fitting functions when the fitting procedure is performed. Specifically, there are 7 different fit functions for 7 distinct B^0 mass fits with three common shared parameters. There are three differences introduced in the Simultaneous B^0 mass fits in contrast to the B^0 mass fit in the low-q^2 region as performed previously in section 4.6.1:

1. The MC & Data Samples are subjected in the Collection 7 of variable cuts (table 4.8) instead of Collection 6 (table 4.7). This is primarily due to these cuts introduced later in the analysis that do not considerably affect the B^0 mass fits on the low-q^2

region that can reject a lot of background processes across the 7 Rare Mode bins in the Data samples (table 4.1).

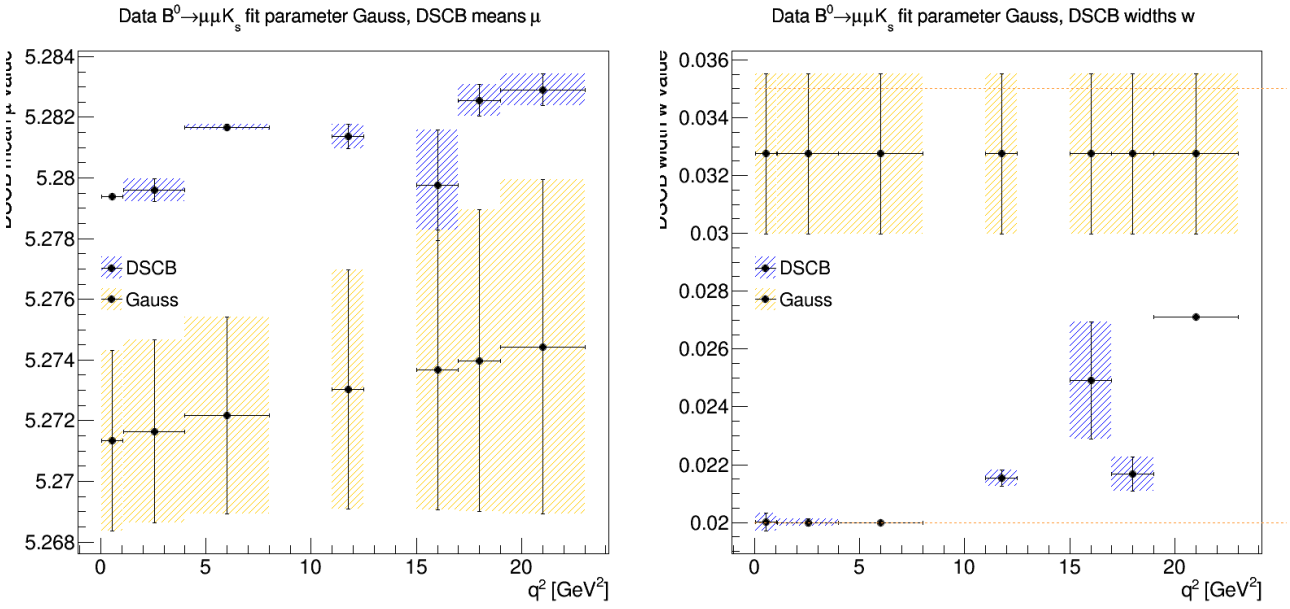
2. The Rare Mode signal search in the Data Samples has a constraint on the $B^0 \to K^*(K^0_S\pi)\mu\mu$ partially reconstructed Background to control its yield based on the PDG values. The yield factor takes into account the relative branching ratio of the Rare Mode over the branching ratio of the aforementioned partially reconstructed background, as well as their $\alpha \cdot \epsilon$ ratios up to Collection 7. This is a better practice than leaving the K^* yield relative to the Rare Mode and the two Resonances floating.

3. The three common shared parameters is a single Gaussian width σ_{G1} and a linear dependence of the Gaussian mean μ_{G1} with respect to the mean value of the q^2 region in the Rare Mode Data B^0 mass fit. The fit function is the same as in equation 4.6, including a constraint factor multiplying the relative yield as explained in the previous point. The fit parameter vector $\vec{\phi}_1$ across all q^2 regions is changed so that the σ_{G1} parameter to be commonly shared across all 7 fits. Furthermore the μ_{G1} parameter is modified as:

$$\mu_{G1} = a\frac{q^2_{\text{high}} + q^2_{\text{low}}}{2} + b$$

where a, b are commonly shared across all 7 fits establishing a linear dependence between $m(B^0)$ and $m(\mu\mu) = q^2$. The reason for doing this is to further impose constraints of the fitting procedure, in order to eliminate some independence across the q^2 regions. The errors have been calculated using the propagation formula (eq. 5.1).

Figure 4.28 shows the values of the DSCB and Gauss means as well as the DSCB widths and Gauss width in the Simultaneous Mass Fit. Tables 4.16 and 4.17 show the measured yields of the modes in the Data samples.



(a) DSCB and Gauss values of the means μ_{D1}, μ_{G1} for BDT> 4.5.

(b) DSCB and Gauss values of the widths w_{D1}, w_{G1} for BDT> 4.5.

Figure 4.28: DSCB and Gauss value mean and width fit parameters along with their errors as computed by the fitting procedure for BDT> 4.5. The orange dotted lines show the limits of the DSCB fit parameter ranges.

Quantity	Bin 1	Bin 2	Bin 3	Bin 5	Bin 7	Bin 8	Bin 9
Yields	16 ± 5	81 ± 9	88 ± 12	37 ± 5	34 ± 6	40 ± 6	9 ± 3

Table 4.16: N_j Yield values for 7 out of the 9 bins in table 4.1 of the Rare Mode in the Data samples.

Quantity	Value
J/ψ Resonance	
Yield	115277 ± 357
$\psi(2S)$ Resonance	
Yield	11565 ± 110

Table 4.17: N_j Yield values for bin 4 & 6 in table 4.1 of the J/ψ Resonant and $\psi(2S)$ Second Resonant q^2 regions in the Data samples.

5 Summary & Conclusion

In this section the results of this analysis are shown and discussed. Throughout this analysis, no systematic errors are calculated, as the work required for their estimation is time consuming and is out of the scope of this thesis. The statistical uncertainties σ_q of a quantity $q(\vec{\phi})$ depending on the parameters, measurements and other quantities $\vec{\phi} = (\phi_1, \phi_2, ..., \phi_n)^T$ with each ϕ_i having an error σ_{ϕ_i} are always calculated via the error propagation formula:

$$\sigma_q = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial q}{\partial \phi_i} \sigma_{\phi_i} \right)^2} \tag{5.1}$$

except in the case of the α, ϵ factors (binomial errors, eq. 4.4), sPlot (eq. 4.7) and the bin contents of quantity distributions, where the error provided by the ROOT framework plotter is equal to $\sigma_j = \sqrt{N_j}$ where N_j are the entries in the jth bin.

5.1 Results

The culmination of this analysis and its end result is the determination of the two observables, the differential branching function for the Rare Mode (section 3.4.2, which expresses the $R_{K_S^0}(\mu)$ observable (section 3.4.1) in the 7 Rare Mode binds of table 4.1. The computation of $R_{K_S^0}(\psi(2S))$ is also performed as a sanity check the validity of the analysis strategy (fig. 4.1). The values in tables 4.9, 4.10, 4.16 and 4.17 are used in determining the measurements of the observables. The list of statistical errors that have been propagated to these results is shown as follows:

- Yield errors provided by the B^0 mass fits for all three modes (Rare, Resonant and Second Resonant). These errors are given by the fitting procedure.
- The binomial errors of the α, ϵ factors.

Systematic errors that are present in these results but are not estimated come from the following effects:

- BDT training related systematics, such as randomness from the shuffling of the training data and randomness from the procedure (section 4.5). Re-training a BDT with no particular seed for a random number generator (RNG) will affect the training procedure slightly, which in turn changes the prediction score.
- The fitting procedure of the B^0 mass fits, such as B^0 mass window calculations (integrals, normalizations, range), the algorithm (Log Likelihood minimizer) and sensitivity of using binned (MC) versus unbinned (Data) versions of the samples. This effect is prominent in the trails of the MC sample distributions. The separation is performed in order for sPlot to be optimized.
- Differences between the MC and Data samples, as shown in figures 4.26 and 4.27. These differences are beyond the control of this analysis.
- Physics objects values and characteristics from the software that creates these values from raw data from the CMS experiment (trigger and selections). This affects the quality of the tracks which are then processed and attributed to a particle candidate. Muons are exceptionally immune to misvaluement in contrast to all other candidate tracks (B^0, K_S^0, etc) due to the nature of the experiment. These effects are also beyond the control of this analysis.

It should be noted however, due to the experimental determination of the form of the observables (eqs. 3.3, 3.4), systematic errors due physics background processes and trigger reconstruction efficiencies cancel out for two different decay modes as previously explained in section 3.4.1.

5.1.1 Measurement of the $B^0 \to K^0_S \mu^+ \mu^-$ Differential Branching Function

In order to provide a comparison measure for the differential branching function, it needs to be compared to another quantity. The `flavio`[63] python package was used, which provides Standard Model theoretical calculations for various flavor physics and other tests. `flavio` can provide a calculation of the differential branching function as the theoretical uncertainty to this calculation. Figure 5.1 shows the dB/dq^2 quantity. The blinding nature of the result prohibits the complete comparison of the measurement and its deviation of the theoretical calculation.

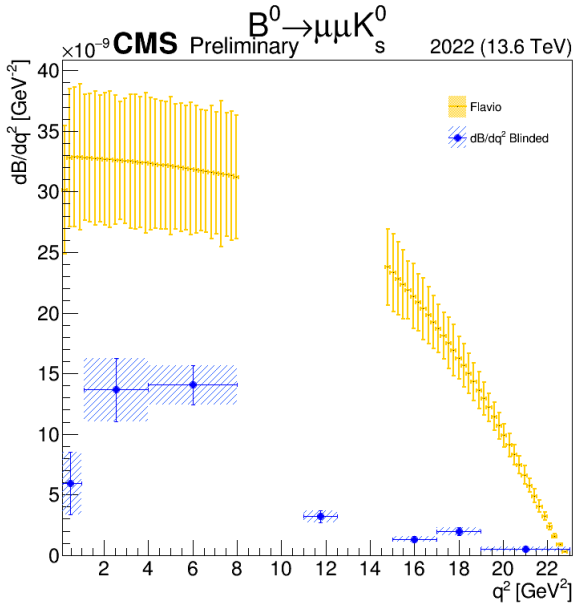


Figure 5.1: Measurement of the Differential Branching Function (blue) versus the theoretical calculation (yellow). The measurement is performed on the 7 Rare Mode bins of table 4.1.

5.1.2 Measurement of the $R_{K^0_S}(\psi(2S))$ Ratio

The experimental measurement of the $R_{K^0_S}(\psi(2S))$ ratio (eq. 3.3 is performed as a sanity check of the analysis. As per the differential branching function, the $R_{K^0_S}(\psi(2S))$ ratio is compared to the value provided by the PDG[19]. The values acquired from that resource are the branching ratios for $B^0 \to J/\psi K^0_S$ and $B^0 \to \psi(2S)K^0_S$, as well as the muonic decay channels of the resonant particles $J/\psi \to \mu\mu$ and $\psi(2S) \to \mu\mu$, multiplied to get the branching ratios in equation 3.3 respectively. PDG also provides experimental errors which include systematic errors from their respective analyses. Table 5.1 shows the values for this ratio.

Type	$R_{K^0_S}(\psi(2S))$ Value
Measurement	$(8.35 \pm 0.09) \cdot 10^{-2}$
PDG	$(8.74 \pm 1.02) \cdot 10^{-2}$

Table 5.1: Measurement of the $R_{K^0_S}(\psi(2S))$ ratio and the value provided by PDG.

5.2 Discussion

The results provided in sections 5.1.1 and 5.1.2 spark a lot of points for discussion, as well as possible refinement techniques to improve the nature of the result. Both of these results will be discussed.

Differential Branching Function By looking at figure 5.1, discussion of the q^2 bins 1, 4 and 6 (table 4.1) will be made, as these bins show divergence from the trend provided by the theoretical calculation.

- bin 1 $[0.1, 0.98]GeV^2$ is shown to diverge from the trend provided by the theoretical calculation by `flavio`, and this can be attributed to the low available statistics in the simultaneous B^0 mass fit bins (Appendix B). What's more, the Data samples are contaminated with low $c\bar{c}$ resonances in that bin (figure 4.10).
- There is no bin 4 theoretical calculation given. However, it should be noted that this bin is between the two resonant modes (figure 4.10) where event leakages from the tails of the two distributions can still have an undetermined effect.
- According to the trend provided by the theoretical calculation, the dB/dq^2 given in bin 6 should be between its two adjacent bins. However, it should be noted that this is the high-q^2 region as described and studied with the BDT performance in section 4.5, where the training region considerably affects the shape of the $o(x)$ distributions (figures 4.12, 4.13a). What's more, there is a small difference between the shape of the low-q^2 and high-q^2 $o(x)$ distributions on the low-q^2 trained BDT. In the case of the simultaneous B^0 mass fits, the high-q^2 range is broken down to three bins, each containing less statistics than the previous one, as the q^2 distribution is monotonically decreasing in the high-q^2 region (figure 4.10).

It is evident by these observations on the differential branching function that two attempts should be made to improve on these shortcomings. Firstly, the statistics of the Data Sample should be increased in order to reduce statistical errors and divergences. An attempt to increase the available statistics using the CMS Run 3 2023 Data in conjunction with the CMS Run 3 2022 Data is explained in section 5.3.1. What's more, the differences between the MC and Data Samples that became evident in section 4.7 thanks to the sPlot technique should be addressed by the groups providing these samples, as well as possible errors in the software. Further analysis of the Collection cuts could also be proven fruitful.
Secondly, an extensive sensitivity analysis of the BDT training procedure as well as the training q^2 regions should be performed to see the effect of improving the ML aspect of this analysis. The BDT is not only responsible for eliminating unwanted background processes, but also is crucial in determining the optimal BDT cut $x > c_b$ via a FoM metric. Different FoMs as well as lesser step sizes in the BDT cut than 0.5 would lead to a sensitivity analysis for the BDT cut value. It should also be noted that the BDT has been trained in the Collection 6 of variable cuts and not Collection 7, which has little but measurable effect on the low-q^2 training region. Discussion about the BDT tuning process, is furthered in section 5.3.2.

$R_{K^0_S}(\psi(2S))$ **Sanity Check** Table 5.1 provides the measurement of this thesis and compares it to the value of the PDG. Comparing the experimental measurement of this analysis to the PDG value, the experimental result lies within 1σ of the PDG value,

where σ is the uncertainty provided by the PDG. This analysis does not aim to find deviations from the SM by looking at the two resonant modes but uses this measurement in order to facilitate a sanity check which can validate the procedure of the analysis as described in figure 4.1 at the beginning of section 4. The equation:

$$R_{K_S^0}(\psi(2S))_{\text{exp}} = R_{K_S^0}(\psi(2S))_{\text{PDG}} + n\sigma$$

is satisfied for $n = -0.38$. A possible reason for the value of the measurement being lower than that provided from PDG can be attributed to the calculation of the yields from the B^0 mass fits, as the yields are calculated within $\pm 2\sigma_D$ of the DSCB distribution fit parameter and not from the whole integration range of the B^0 mass.

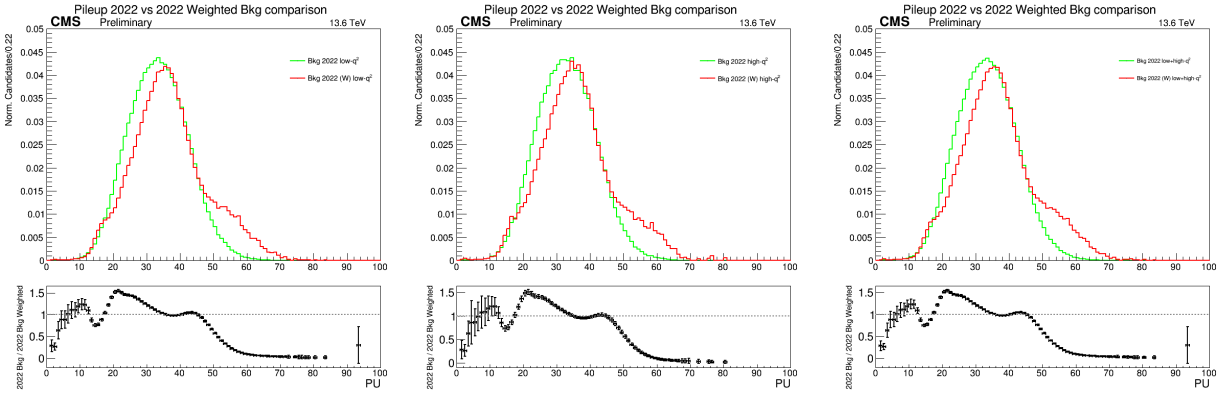
5.3 Analysis Refinement

In this section, discussion on what the analysis can be improved upon is contained. Firstly, two features are discussed, an attempt to increase the sample sizes by incorporating the CMS Run 3 2023 Data samples into the analysis in section 5.3.1 and the BDT Tuning process in section 5.3.2. Ideas for possible further refinement after the submission of this thesis can be found in section 5.4.

5.3.1 Increasing the Sample Sizes

As mentioned in section 5.2, the increase of the Sample Sizes of both MC and Data samples from year 2022 has been attempted. In order to incorporate this size increase using the Run 3 2023 samples in conjunction with the Run 3 2022 samples, both samples across the two years must have identical distributions, with the only uncertainty allowed to show the deviation being the statistical uncertainty of the bin contents. In order to check if this process is possible, one needs to check the Pileup distribution in the Data Samples of the two years. It is known that the 2023 year's Pileup distribution has a higher mean and median, which allows more processes to happen per beam collision. During the duration of a Run in a year in the LHC, some of the beam parameters vary, yielding different measurements as the machine ramps up and down of operation. This inherently requires different trigger strategies, which in turn affects the processed samples that are used in the analysis.

Unfortunately, the Pileup Distribution of the Data 2023 samples was unavailable for comparison. A direct comparison between the Data 2022 and Data 2023 samples requires these two distributions. The next step in determining whether these samples can be merged was to look at the BDT Output distribution $o(x)$ of the two year Data samples. Fortunately, the scale factors u_1, u_2 were available in order to reweight the 2022 Data into the 2023 Data, using the Pileup distributions as shown in figure 5.2. From these two scale factors, u_1 corresponds to the scale of MC 2022 to Data 2022 samples based on the total number of Primary Vertices in each collision and u_2 corresponds to the scale of Data 2022 to Data 2023 samples based on the total number of Primary Vertices in each collision. From this point on, if a sample is rewighted based on this method, it will be abbreviated as (W), for example the Weighted Data 2022 will be referenced as Data (W) 2022. A second unfortunate problem was the complete absence of the MC 2023 samples. If no issues between the 2022 (W) and 2023 Data samples arose, the events in the MC 2023 samples were to be weighted via the scale factor $u_1 \cdot u_2$, and an extensive analysis in training a hybrid BDT using weighted entries had to be performed.



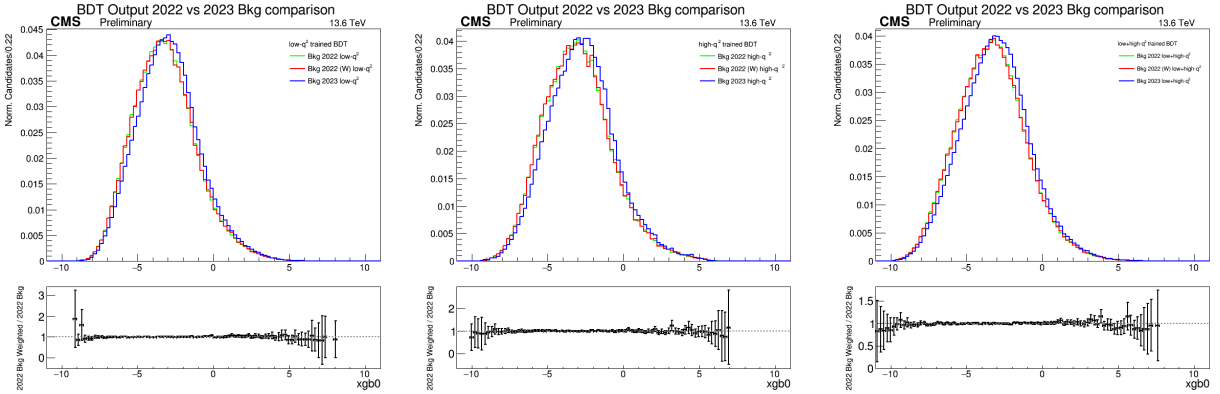
(a) Pileup Distributions in the low-q^2 Sideband Region.

(b) Pileup Distributions in the high-q^2 Sideband Region.

(c) Pileup Distributions in the low+high-q^2 Sideband Region.

Figure 5.2: Pileup Distributions in the low-q^2, high-q^2 and low+high-q^2 Sideband Region $m(B^0) \in (4.85, 5.05) \cup (5.48, 5.65)$GeV of the Data samples 2022 and (W) 2022.

In the process of checking if the BDT Output $o(x)$ distributions between Data (W) 2022 and Data 2023 were different, the training of the three initial BDTs (low-q^2, high-q^2 and low+high-q^2) was not changed to incorporate any of the 2023 year samples. The result is shown in figure 5.3. However, the deviation between the Data (W) 2022 and Data 2023 samples $o(x)$ distributions show a distinction that merging these samples together is not possible due to the difference. This also implies that the MC 2022 samples should not be reweighted according to the $u_1 \cdot u_2$ scale factors. The large distinction between the Data 2022 and MC 2022 also became evident later in the analysis due to the $p(B^0)$ Distribution thanks to the sPlot technique (figures 4.26, 4.27). Therefore the sample size expansion was abandoned.



(a) $o(x)$ Distributions in the low-q^2 of the low-q^2 trained BDT.

(b) $o(x)$ Distributions in the high-q^2 of the high-q^2 trained BDT.

(c) $o(x)$ Distributions in the low+high-q^2 of the low+high-q^2 trained BDT.

Figure 5.3: BDT Output $o(x)$ Distributions of low-q^2, high-q^2 and low+high-q^2 trained BDTs in their corresponding high-performing q^2 regions of the Sideband Region $m(B^0) \in (4.85, 5.05) \cup (5.48, 5.65)$GeV of the Data samples 2022, (W) 2022 and 2023.

5.3.2 Tuning the BDT

In section 5.2, a discussion arose regarding the tuning of the BDT. The BDTs used throughout this thesis have already been given as working optimally and no further exploration in a feasibility study regarding the optimization was made in the initial stage of the training procedure. During the revision process of this thesis, a small feasibility

study regarding the behavior of the training procedure was made and is presented in this section.
It is to be noted that this feasibility study is a very preliminary analysis and no focus was given throughout this thesis and its purpose is to set the stone for a possible future project.

Parameter Tuning The training procedure was augmented with testing additional parameters to those of table 4.12, as well as an early stopping method for adding additional estimators (hyperparameter `n_estimators`) to the training model. The early stopping method prevents the model from overfitting to the training data and requires an evaluation metric and sample to evaluate a possible saturation in the predictive power of the model. The evaluation metric was chosen to be AUC (Area Under the Curve) between the true positive rate and false positive rate. If the training score does not improve in over 35 estimators, the training procedure stops prematurely. Furthermore, the training procedure was subjected to the Stratified n-Repeated k-fold Cross Validation (SRCV) technique in order to repeat the training procedure for a given set of hyperparameters in order to reduce statistical fluctuations across the evaluation metric results. Each iteration of the tuning process was supplied with a hyperparameter grid which lists all possible combinations of values for the BDT hyperparameters as well as a list of training features.

A measure was needed in order to compare BDTs and their efficiency in evaluating the models' predictive power. Two scatter plots were essential in determining good sets of hyperparameters. The first plot is a set of points of the mean training score m_T versus the mean testing (evaluation) m_E score of the trained BDTs. These scores come from averaging the corresponding scores across folds and repetitions in the SRCV method. The latter plot is a variation of the former, showing the model ranking across all models versus the mean testing score m_E. The BDT ranking is defined as a decreasing integer scale of mean testing scores, where rank 1 refers to the model with the highest testing score, constituting this plots as a decreasing function of m_E versus the ranking. This combination of plots are essential in recognizing patterns in hyperparameter behavior and providing a sense of fluctuation in the distribution of the mean testing score.

Figures 5.4 and 5.5 show examples of this method using the an extensive list of training variables $\cos_{2D}\alpha$, $\text{sig}(L_{xy})$, $p(B^0), \text{Iso}(\mu_1), \text{Iso}(\mu_2)$, $p_T(K_S^0), \eta(\mu_1), \eta(\mu_2)$, $\eta(K_S^0), \Delta\eta(\mu_1,\mu_2)$, $\text{sig}(D_{xy},\mu_1)$, $\text{sig}(D_{xy},\mu_2)$. The BDT version in these figures refers to an internal code assigned to each tuning procedure which trains the BDT models with the same hyperparameter grid and features.

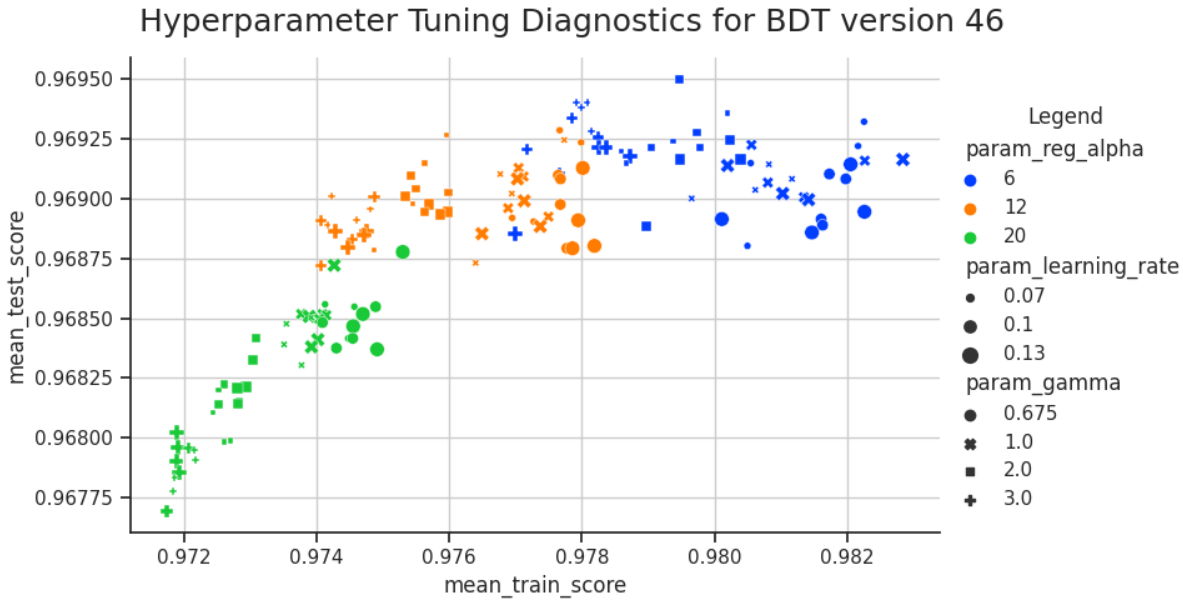


Figure 5.4: Mean training score versus mean testing score. Three of the values of the hyperparameters are shown in the figure. Each point corresponds to a distinct model.

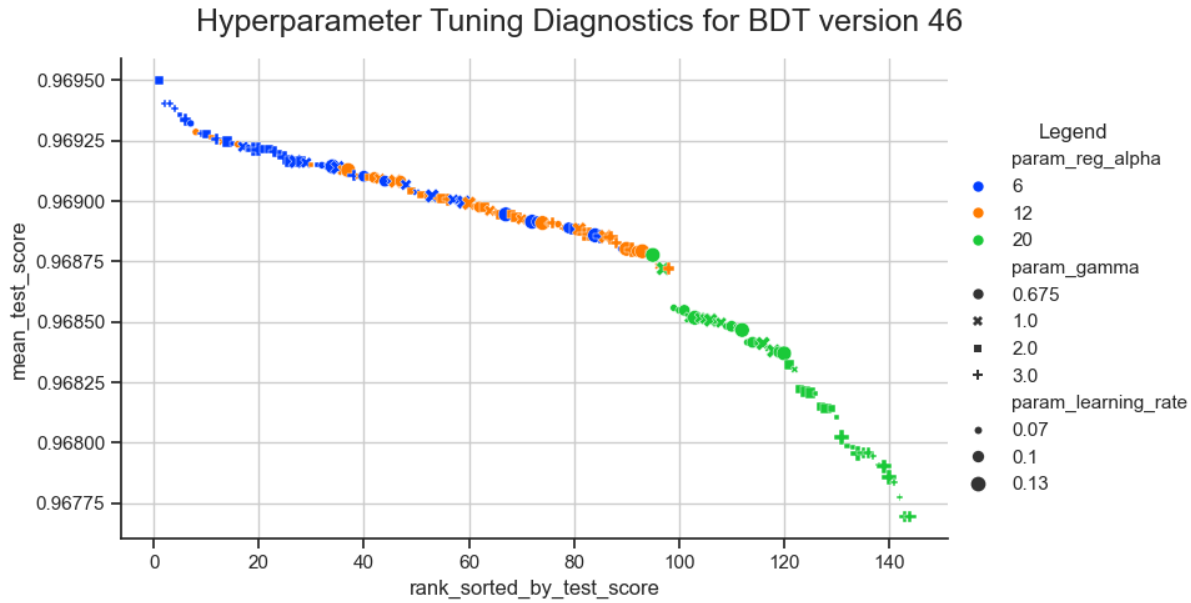
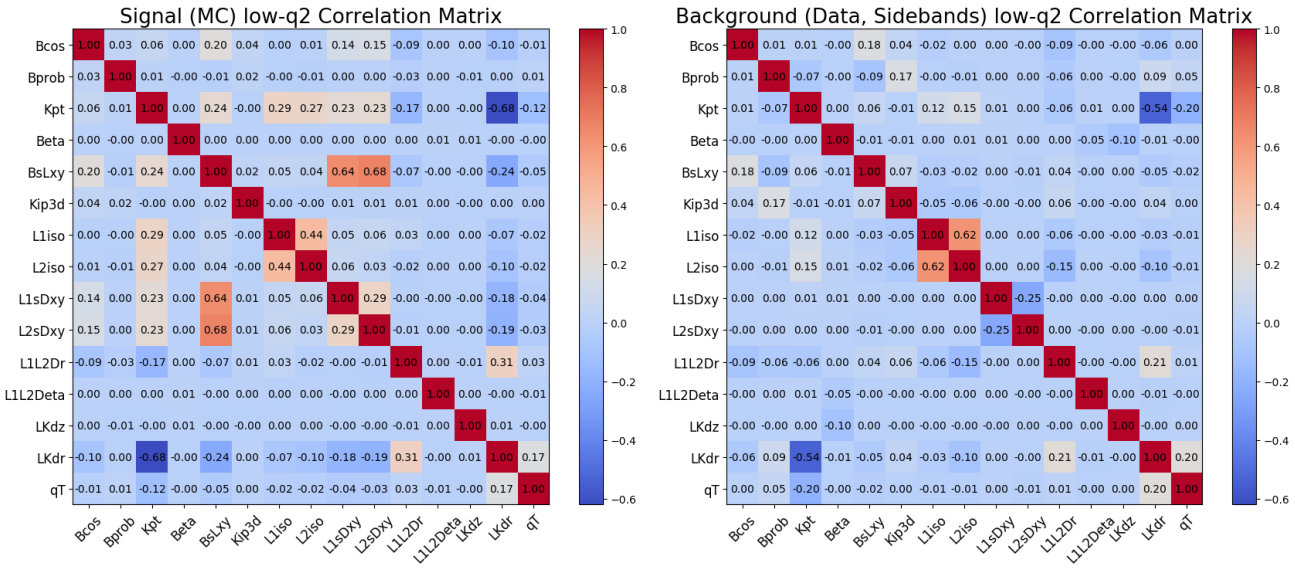


Figure 5.5: Ranking score versus mean testing score. Three of the values of the hyperparameters are shown in the figure. Each point corresponds to a distinct model.

From figures 5.4 and 5.5 it can be seen how a set of hyperparameters displays patterns in these plots and how these patterns provide insight in repeating the tuning process. If two mean training scores m_{T1} and m_{T2} satisfy $m_{T1} < m_{T2}$ and are equal to a mean testing score m_E, then the models with mean training score m_{T2} are rejected as candidates of overfitting. Ideally, areas containing points with high m_T and low m_E are very bad BDT candidates to be used in the analysis as they are subject to overfitting. These candidates typically fall at high ranks and their combinations of hyperparameters are rejected from further iterations of hyperparameter grid searches across the BDT training feature list. It should also be noted, in order to reduce correlations, the correlation matrices (figure 5.6b) of the training variables in the BDT training samples is a crucial metric in order to minimize correlations in the training procedure:



(a) Correlation matrix for a set of non-correlated the MC Signal samples variables.

(b) Correlation matrix for a set of non-correlated the Sideband Region in the Data samples variables.

Figure 5.6: Visual representation of the correlation matrices for the a set of uncorrelated variables for the two BDT classes in the training samples.

It is to be reminded that the difference in the values of the correlations between figures 5.6a and 5.6b are not a sign of data incompatibility across the samples as the training datasets are not subjected to background removal methods. This feasibility study has shown some basic behaviors of the hyperparameters in the BDT training procedure.

- The mean test score m_E improves when the `subsample` hyperparameter is set between the values of 0.85-0.95 relative to the default value of 1. This value corresponds to the percentage of the training dataset that is used in the training procedure for each new estimator added to the model. Rejecting a small portion is a counter to overfitting in the tuning procedure, allowing for high values of the evaluation metric.

- The hyperparameters `reg_alpha` and `reg_lambda` are also a counter to overfitting by applying a penalty to the loss function. Specifically, they regards L1 and L2 regularization respectively. The former has an optimal region between 0-10, with 0 denoting its default value and best performing value being at around 4.5 with no definite deviation and the latter provides best results if left to its default value of 1.

- The number of estimators `n_estimators` increases the computation time a lot and the early stopping technique yields satisfactory results in decreasing the computation time for this feasibility study. However, it has little effect outside the default optimal range of 700-1000.

- The hyperparameter `learning_rate` offers best results in the range of values 0.06-0.1. Slower values have higher computation times and yield slightly better results.

Due to limited resources while working on this project, this feasibility study was not expanded upon in order to measure any of the observables, as more effort had to be put upon in order to sanity check these results further. Due to the time constraints of this thesis, this was deemed unnecessary.

5.4 Future Developments & Closing Remarks

In this section, a brief discussion on future developments on this analysis will be given by the perspective of the whole procedure, separated into core analysis improvements and sensitivity analyses. The former category of improvements would change the results by a considerable amount, whereas the latter may or may not contribute to the analysis procedure and should be investigated after core analysis improvements have been made. Such improvements are:

1. The calculation of systematic errors found in section 5.1.
2. The introduction of more Data and MC samples into the analysis in order to increase the available statistics. In order to do this, the Pileup Distribution in the 2023 Data samples is required, as explained in 5.3.1.
3. The correction of data incompatibilities between the Data and MC samples, as demonstrated by figure 4.26.
4. A further analysis on the Collection of variable cuts, by looking to reject even more background processes and to increase the signal yields before applying a BDT cut $x > c$. Strong cuts will change the fitting parameters of the B^0 mass distributions.
5. A re-iteration of the q^2 bins in table 4.1 and the five q^2 regions in table 4.11. A further understanding of the limits in these bins and regions can lead into rejecting or including more statistics that are appropriate for analysis. The spaces between Rare Mode bins and the two Resonant regions can be adjusted and their effects and background leakages measured. The three high-q^2 bins in table 4.1 can be combined into one or two bins.
6. The low-q^2 training region of the BDT can be adjusted to include all cuts in Collection 7 (table 4.8, sans the BDT cut, in order to measure a possible improvement of the BDT efficiency and shape of the $o(x)$ Distribution. Furthermore, the upper boundary of the BDT training region can also be increased from 6.0GeV^2 to a value conservative relative to the lower boundary of the J/ψ-q^2 region to prevent accidental leakage and increase the B^0 mass fit statistics of the already established samples.
7. A re-evaluation of the calculation of the fitting yields as mentioned in section 4.6.2 and discussed in 5.2, to see whether the change of the calculation of measured yields of the modes affects the values of the observables.

Examples of sensitivity analyses for this analysis include:

1. An extensive analysis of the BDT training procedure by including complete hyper-parameter tuning, feature selection methods, dataset and class label weighting and their effects on the measurement of the differential branching function as well as the $R_{K^0_S}(\psi(2S))$ observable. Furthermore, other modern data science methods can be used in order to explore the data sets, such as Dimensionality Reduction techniques to observe data patterns which could help in identifying more cuts for the Collections of variable cuts.
2. The application of different ML models or Artificial Intelligence models will provide a layer of robustness into the analysis. Using additional ML models in the analysis in order to determine the two observable will hint at possible shortcomings and systematic errors of the BDT procedure that are currently unknown. Furthermore, a different approach may be more suitable for the samples used in this thesis.

3. An analysis on the parameter grid of the B^0 mass fits, as shown in section 4.6.1 and 4.8. The aim of such a procedure is to reduce the yield errors, as well as model the behavior of the $m(B^0)$ distribution given a q^2 selection criterion. One way to do this is to use a ML model or even a Deep Neural Network (DNN). The procedure would start by initializing the fit function $T(x;\vec{\theta})$ (eq. 4.5) along with an initial set of parameters $\vec{\theta}_i$, for $i \in \{k\}_1^N$ generating N different fit functions. By grouping these initial functions into a training dataset which would include the parameters $\vec{\theta}_i$ as well as a goodness of fit measure like χ^2/N_{DOF}, the performance metric can be chosen accordingly in order to minimize the goodness of fit measure as close to 1 as possible. A new testing dataset can be made using slight variations of the training dataset, which will allow an easy and simple analysis of the parameter grid of the fitting function $T(x;\vec{\theta})$. Furthermore, testing the boundaries of the extrapolation predictive power of this model would assist in identifying distributions of fitting parameters given the goodness of fit measure or another parameter subgrid but is an individual analysis of its own.

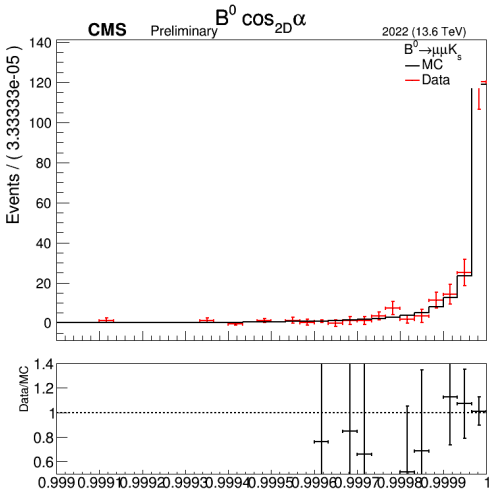
4. The application of different fit functions applied in the B^0 mass fits would simplify the interpretability of the $m(B^0)$ distribution. Different pairs of the elementary functions given in section 4.6.1 as well as new fit functions entirely.

The analysis shown in this thesis can also be done for other $b \to sll$ processes with different mediators. It is imperative to note that the $R_{K_S^0}(\mu)$ observable has been determined within the procedures of the scope in this analysis and the next step is to determine $R_{K_S^0}(e)$ in order to present the $R_{K_S^0}$ ratio as described in section 3.4.1 and compare with the Standard Model prediction. However, determining $R_{K_S^0}(e)$ is not as simple as presented in this thesis, due to the electrons being harder to measure in the CMS trigger. This section marks the end of this project. A new project aiming to improve the aforementioned inconsistencies and rough areas is of extreme interest, particularly the BDT tuning for the training procedure and the optimization of the B^0 mass fits.

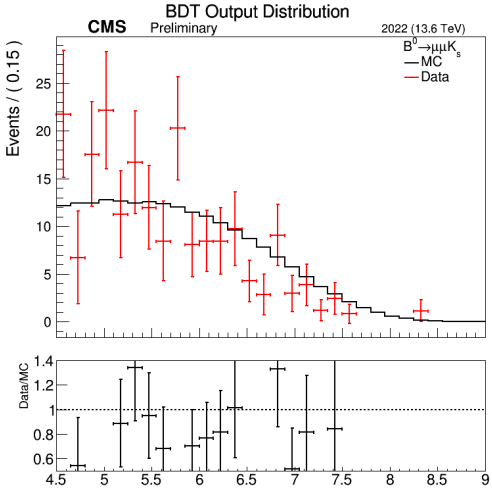
A sPlot Variable Distributions for the Signal Search

This appendix includes the sPlots in the Weighted Data samples for $x > 4.5$ as determined by the analysis across all three modes of interest. Each sPlot corresponds to a distribution of the control variable from the Weighted Data samples as well as its MC sample distribution. A brief explanation of the variable quantities can be found in section 4.2.3.

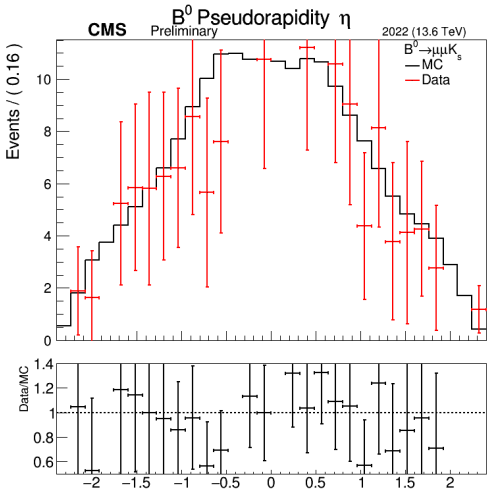
A.1 Rare Mode $B^0 \to \mu^+\mu^- K^0_S$



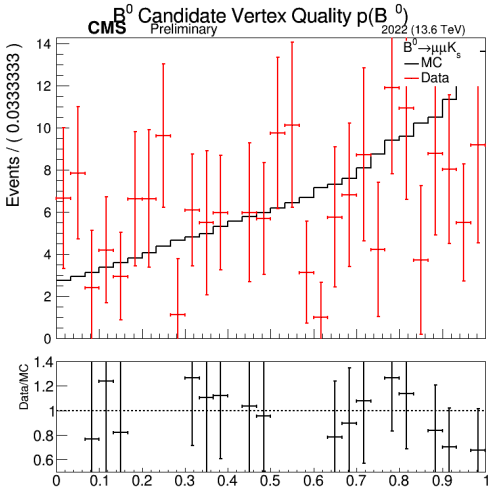
(a) sPlot of the $\cos_{2D}\,\alpha$ variable distribution.



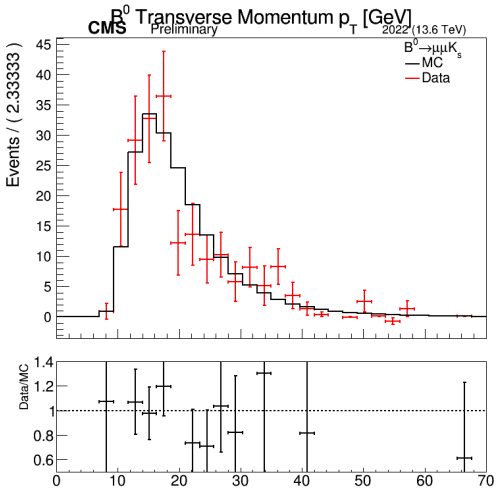
(b) sPlot of the BDT Output score $o(x)$ variable distribution.



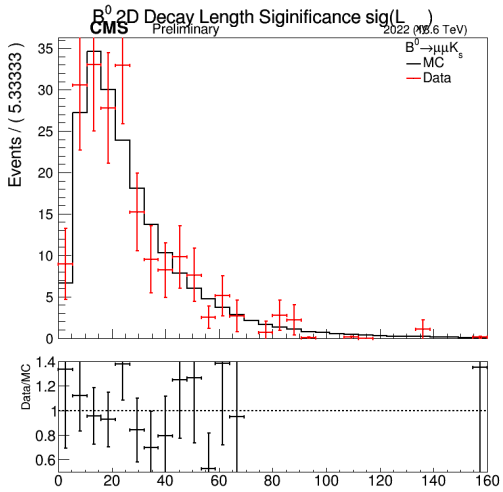
(a) sPlot of the $\eta(B^0)$ variable distribution.



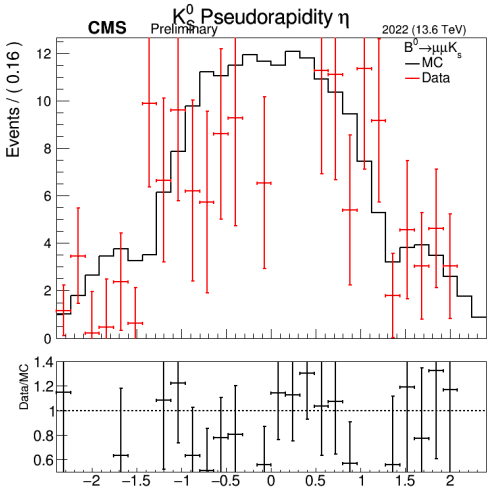
(b) sPlot of the $p(B^0)$ variable distribution.



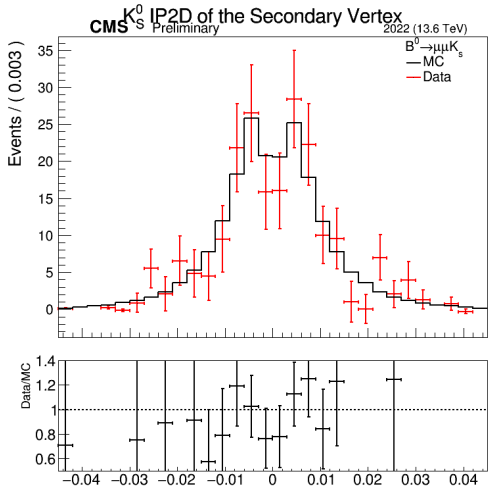
(a) sPlot of the $p_T(B^0)$ variable distribution.



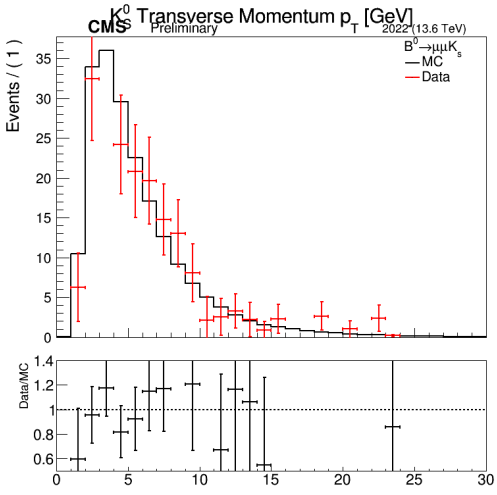
(b) sPlot of the $sig(L_{xy})$ variable distribution.



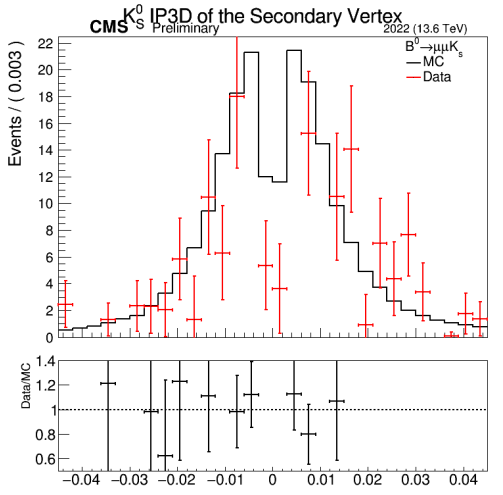
(a) sPlot of the $\eta(K^0_S)$ variable distribution.



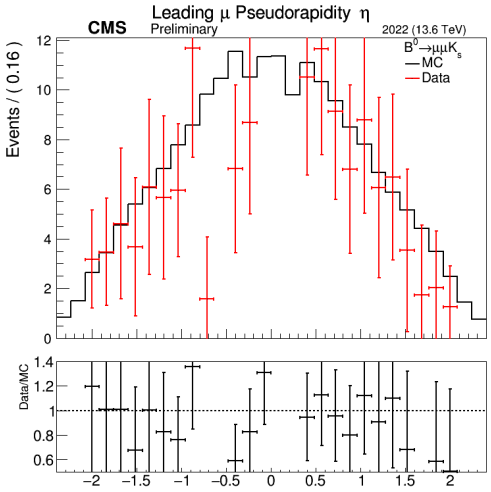
(b) sPlot of the $IP(K^0_S, 2D)$ variable distribution.



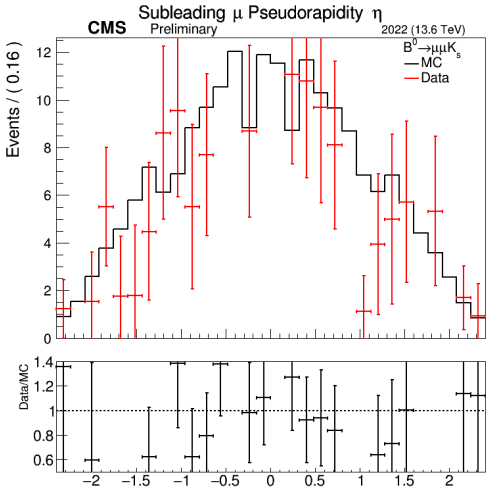
(a) sPlot of the $p_T(K^0_S)$ variable distribution.



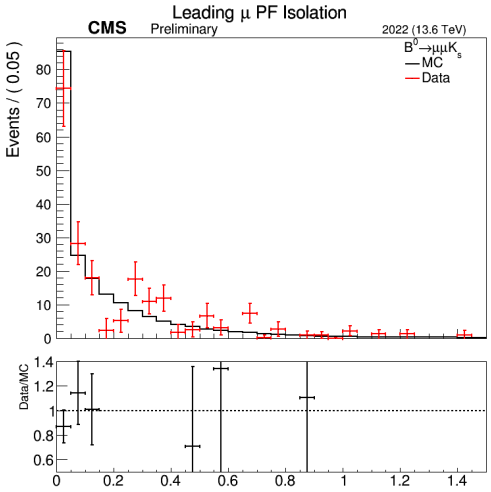
(b) sPlot of the $IP(K^0_S, 3D)$ variable distribution.



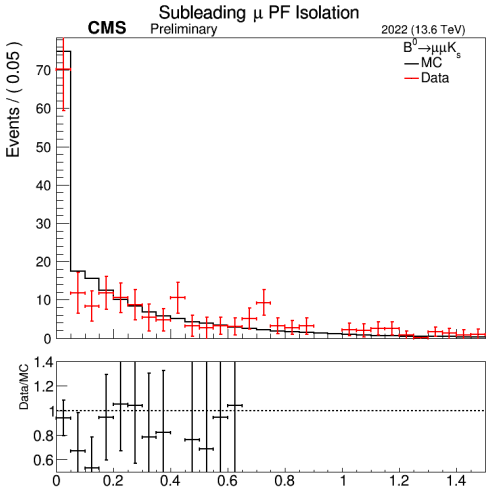
(a) sPlot of the $\eta(\mu_1)$ variable distribution.



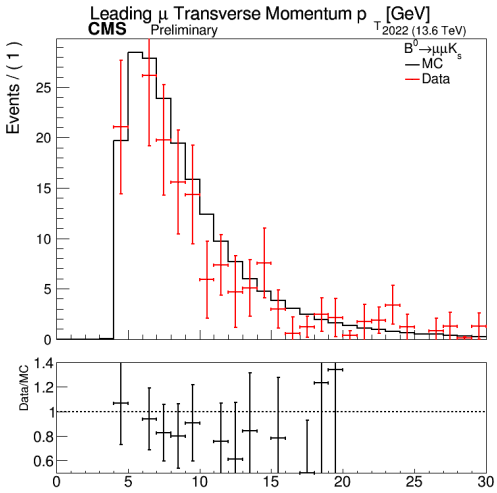
(b) sPlot of the $\eta(\mu_2)$ variable distribution.



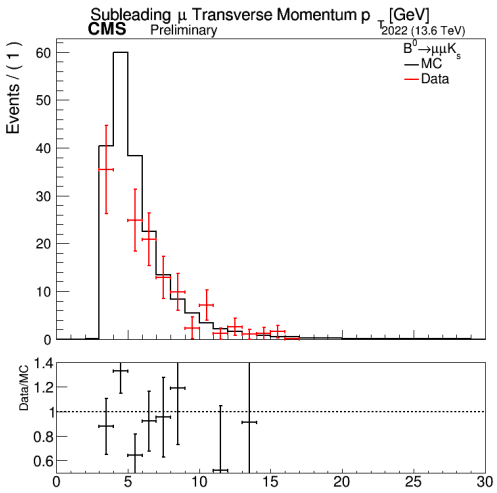
(a) sPlot of the $Iso(\mu_1)$ variable distribution.



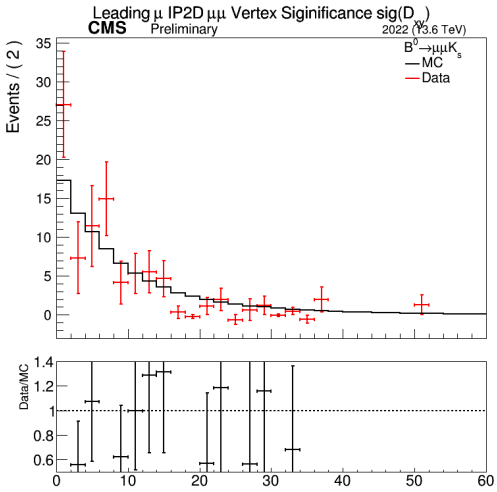
(b) sPlot of the $Iso(\mu_2)$ variable distribution.



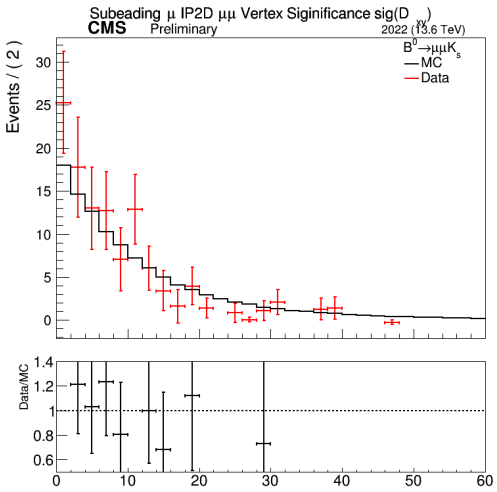
(a) sPlot of the $p_T(\mu_1)$ variable distribution.



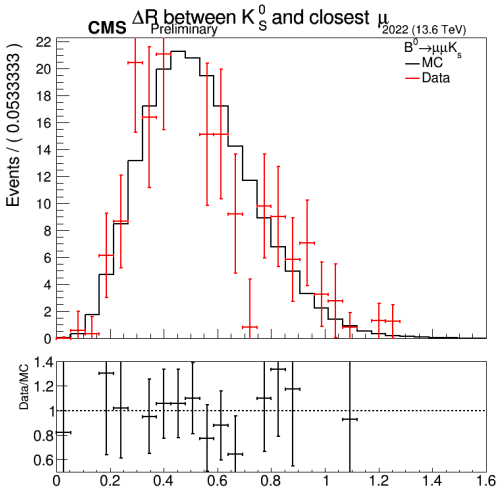
(b) sPlot of the $p_T(\mu_2)$ variable distribution.



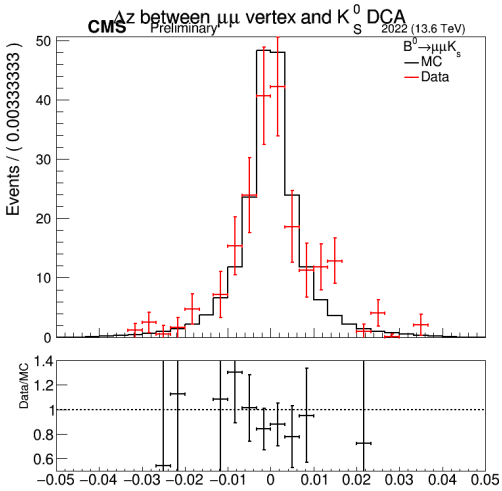
(a) sPlot of the $sig(D_{xy})$ leading muon variable distribution.



(b) sPlot of the $sig(D_{xy})$ subleading muon variable distribution.

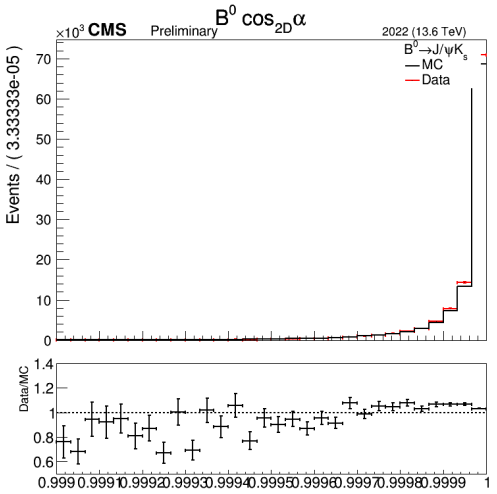


(a) sPlot of the $\Delta R(\mu, K^0_S)$ variable distribution.

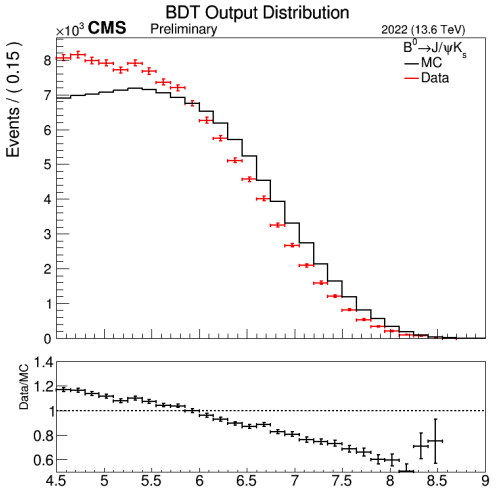


(b) sPlot of the Δz variable distribution.

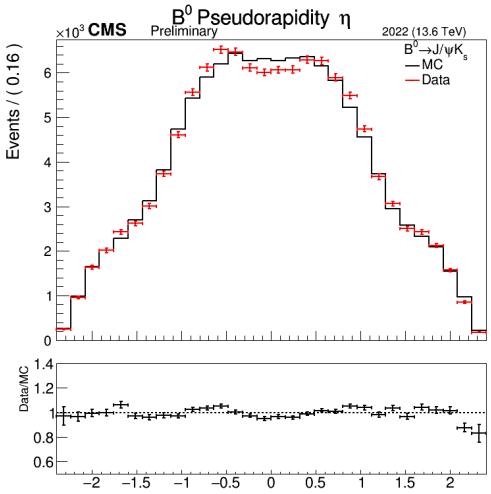
A.2 Resonant Mode $B^0 \to J/\psi K^0_S$



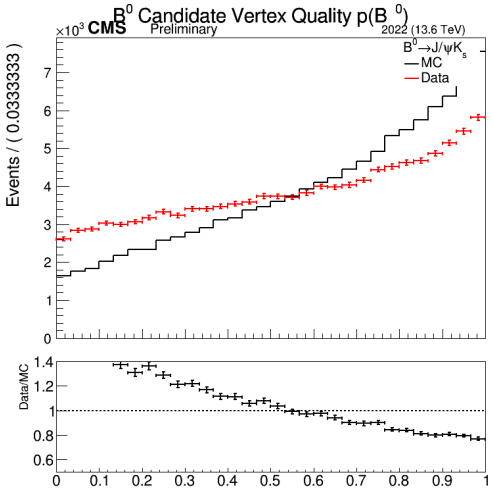
(a) sPlot of the $\cos_{2D}\alpha$ variable distribution.



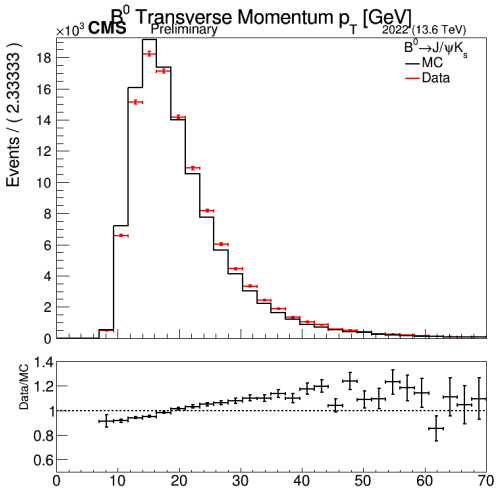
(b) sPlot of the BDT Output score $o(x)$ variable distribution.



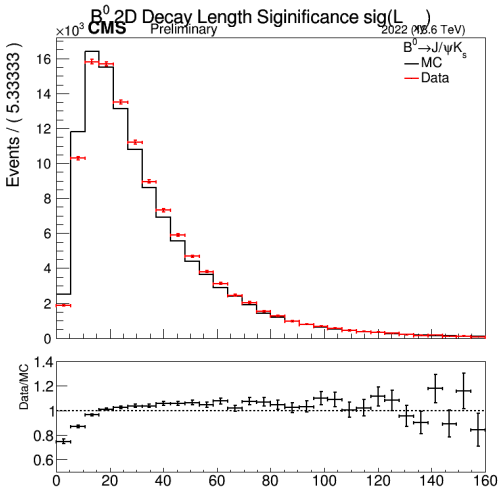
(a) sPlot of the $\eta(B^0)$ variable distribution.



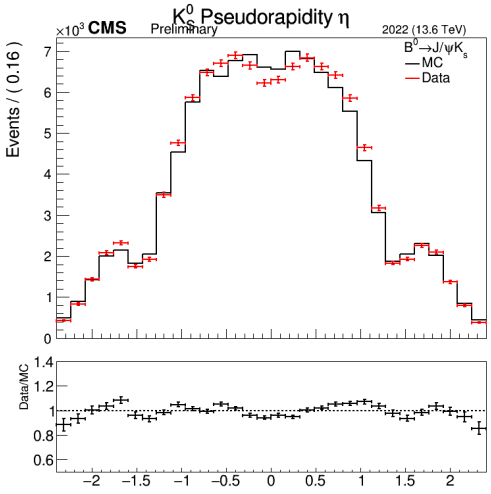
(b) sPlot of the $p(B^0)$ variable distribution.



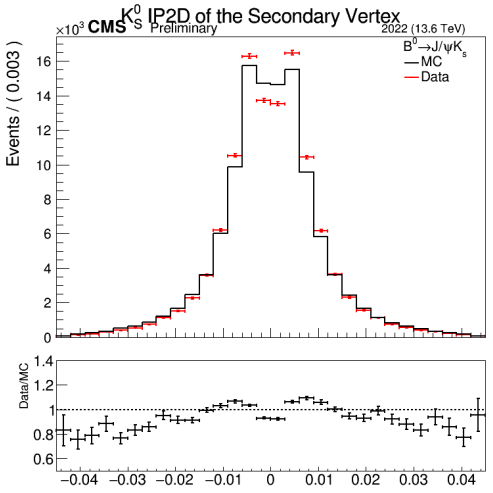
(a) sPlot of the $p_T(B^0)$ variable distribution.



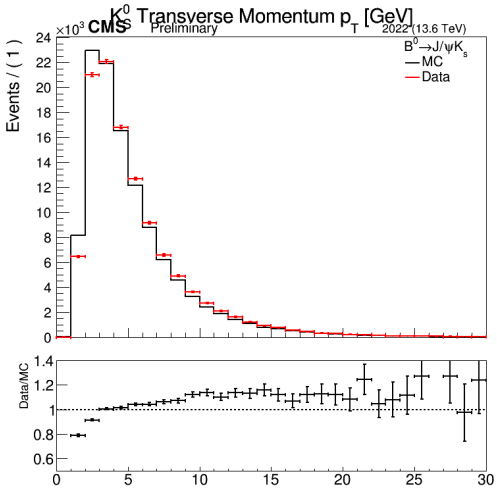
(b) sPlot of the $sig(L_{xy})$ variable distribution.



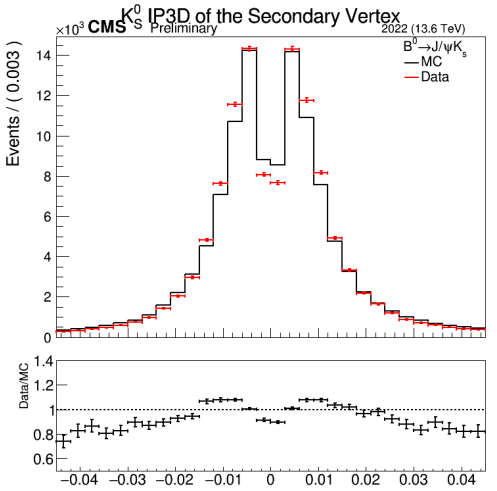
(a) sPlot of the $\eta(K^0_S)$ variable distribution.



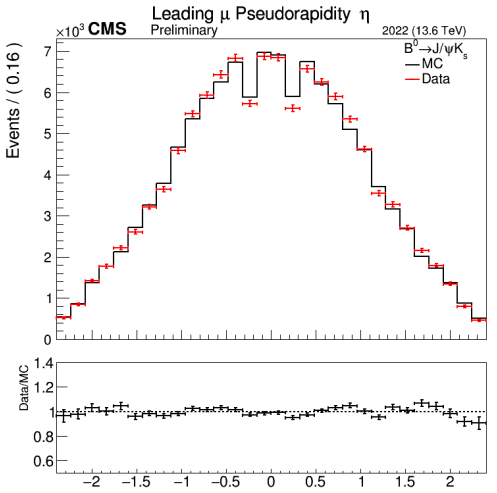
(b) sPlot of the $IP(K^0_S, 2D)$ variable distribution.



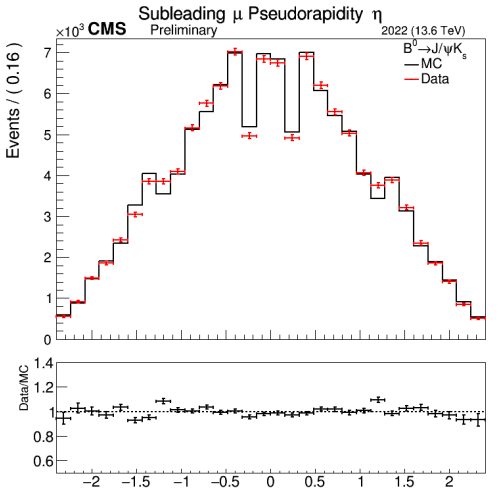
(a) sPlot of the $p_T(K^0_S)$ variable distribution.



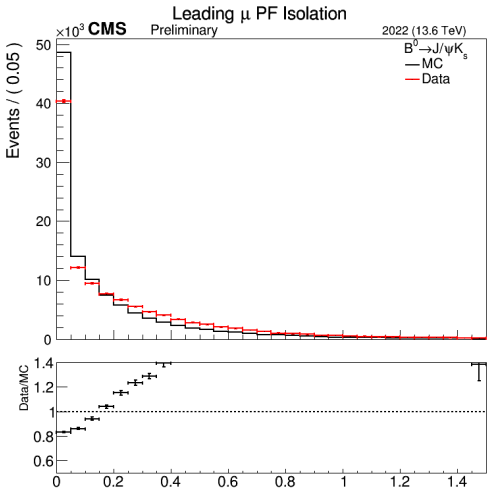
(b) sPlot of the $IP(K^0_S, 3D)$ variable distribution.



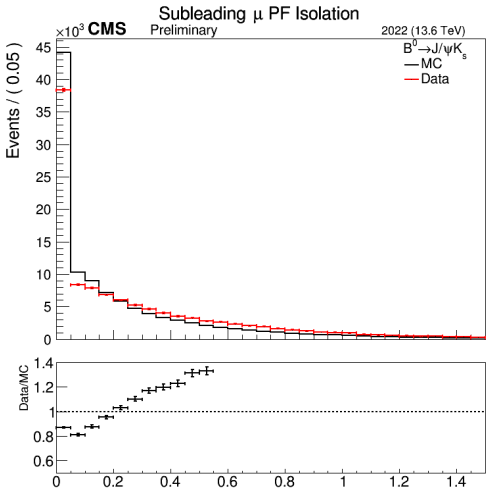
(a) sPlot of the $\eta(\mu_1)$ variable distribution.



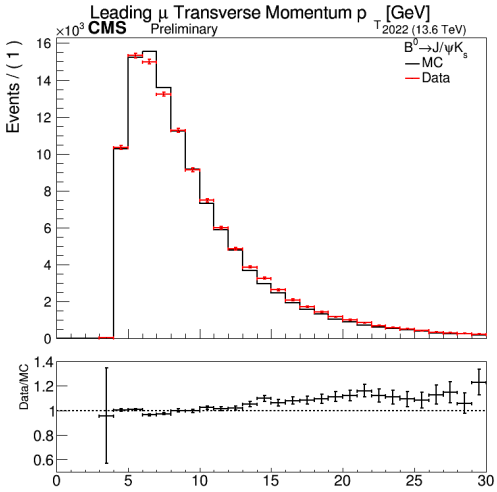
(b) sPlot of the $\eta(\mu_2)$ variable distribution.



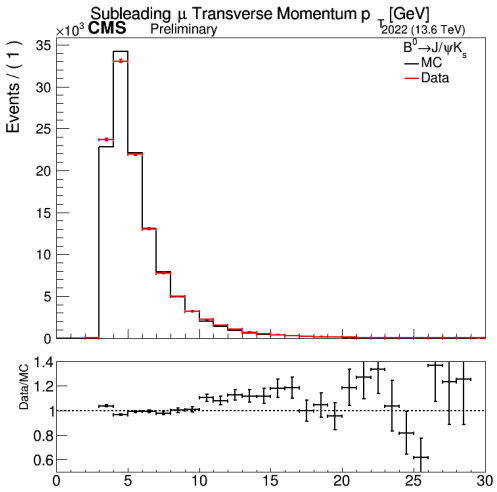
(a) sPlot of the $Iso(\mu_1)$ variable distribution.



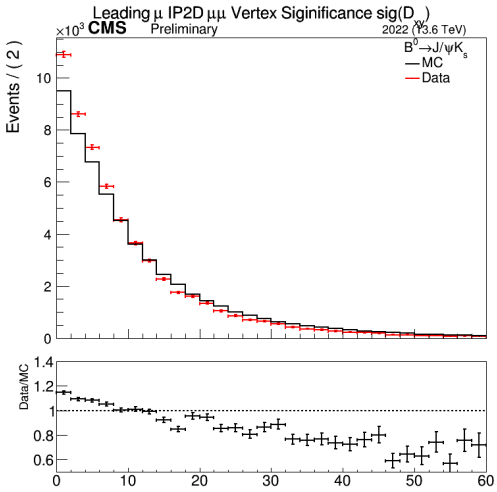
(b) sPlot of the $Iso(\mu_2)$ variable distribution.



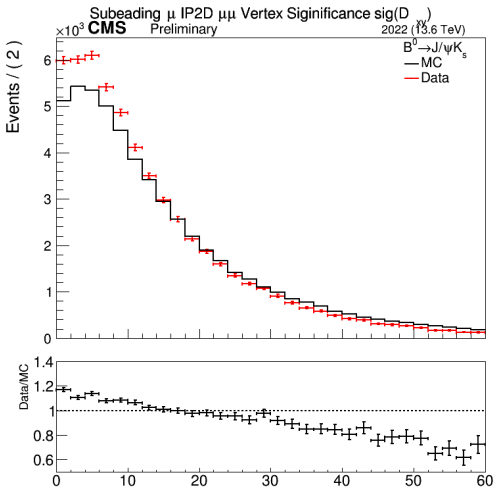
(a) sPlot of the $p_T(\mu_1)$ variable distribution.



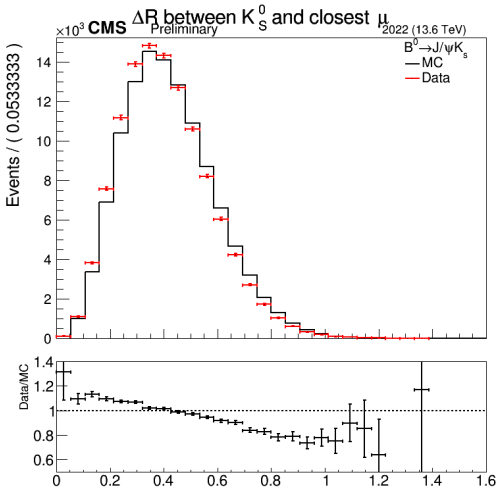
(b) sPlot of the $p_T(\mu_2)$ variable distribution.



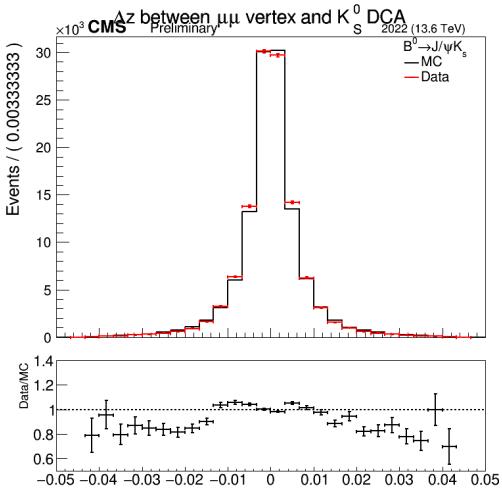
(a) sPlot of the $sig(D_{xy})$ leading muon variable distribution.



(b) sPlot of the $sig(D_{xy})$ subleading muon variable distribution.

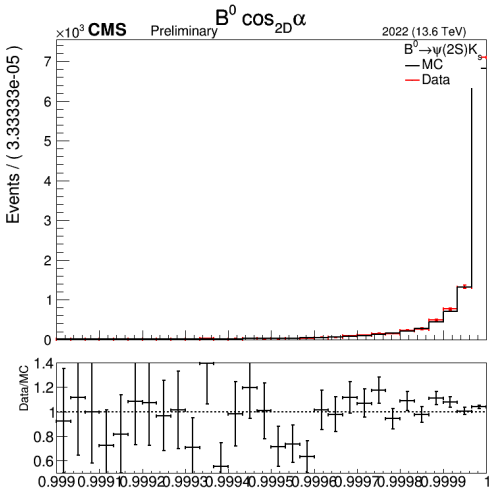


(a) sPlot of the $\Delta R(\mu, K^0_S)$ variable distribution.

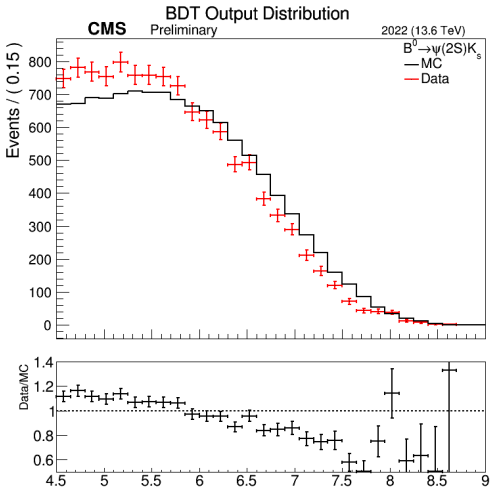


(b) sPlot of the Δz variable distribution.

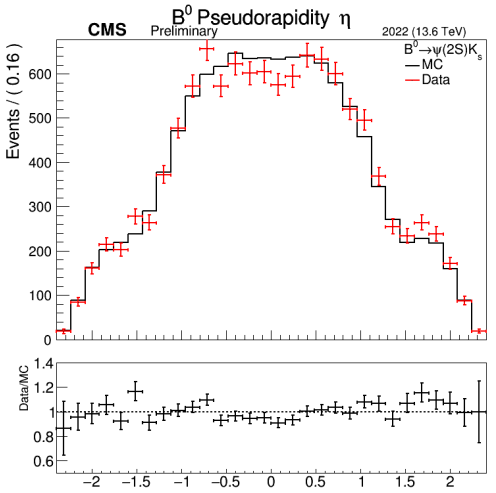
A.3 Resonant Mode $B^0 \to \psi(2S)K^0_S$



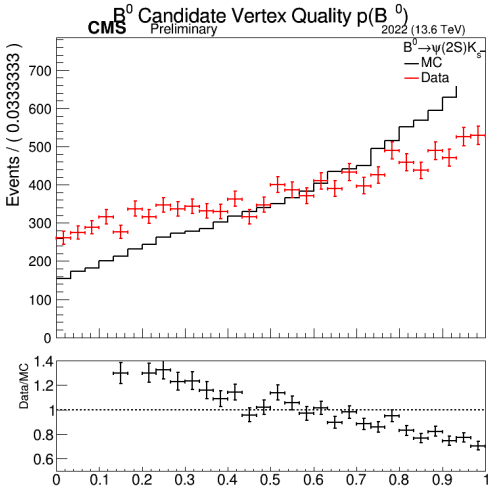
(a) sPlot of the $\cos_{2D}\,\alpha$ variable distribution.



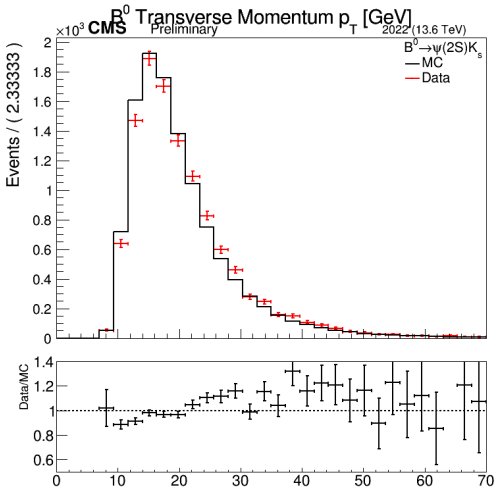
(b) sPlot of the BDT Output score $o(x)$ variable distribution.



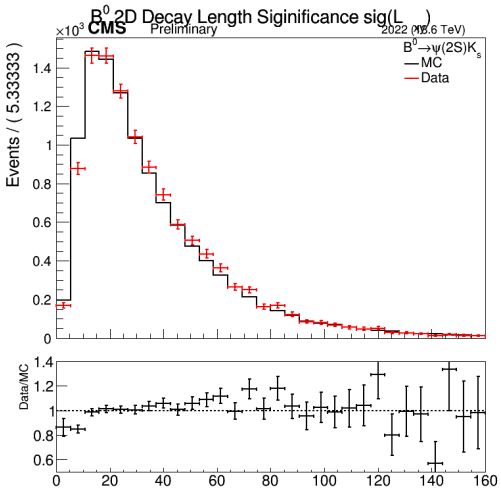
(a) sPlot of the $\eta(B^0)$ variable distribution.



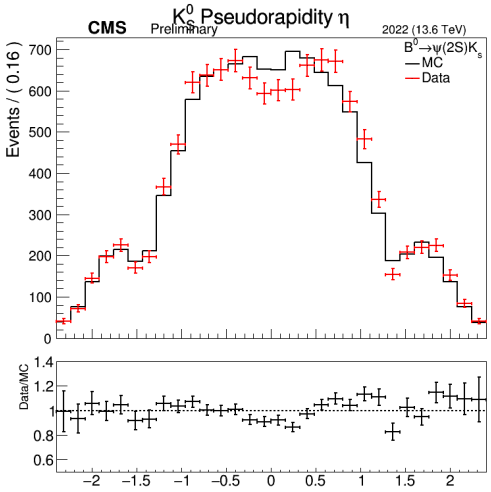
(b) sPlot of the $p(B^0)$ variable distribution.



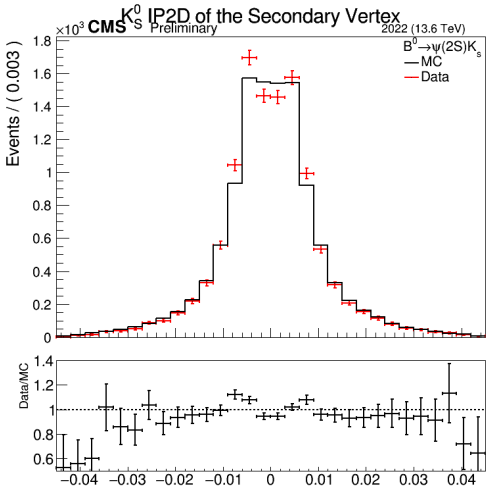
(a) sPlot of the $p_T(B^0)$ variable distribution.



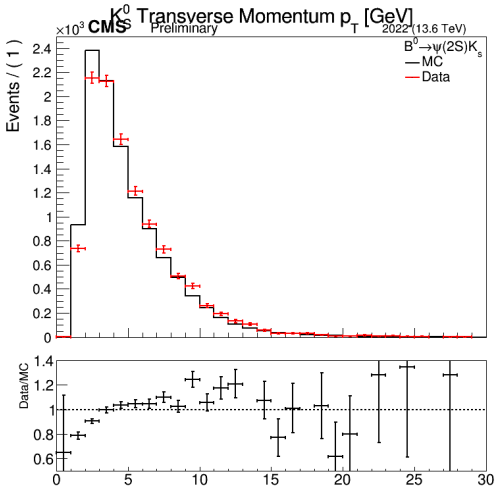
(b) sPlot of the $sig(L_{xy})$ variable distribution.



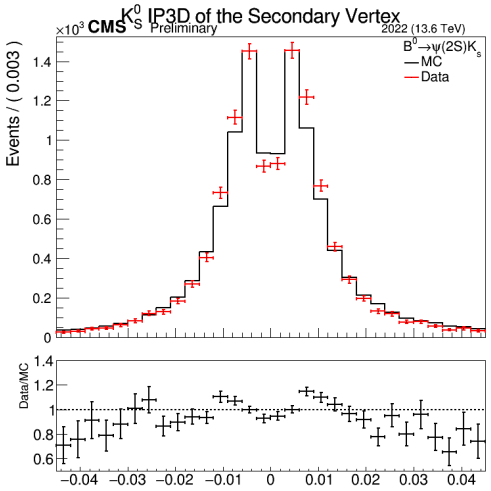
(a) sPlot of the $\eta(K^0_S)$ variable distribution.



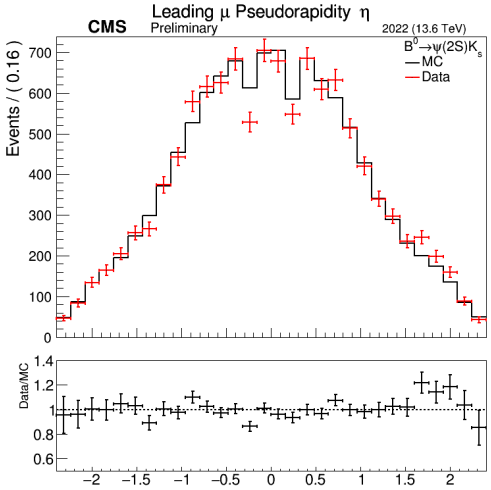
(b) sPlot of the $IP(K^0_S, 2D)$ variable distribution.



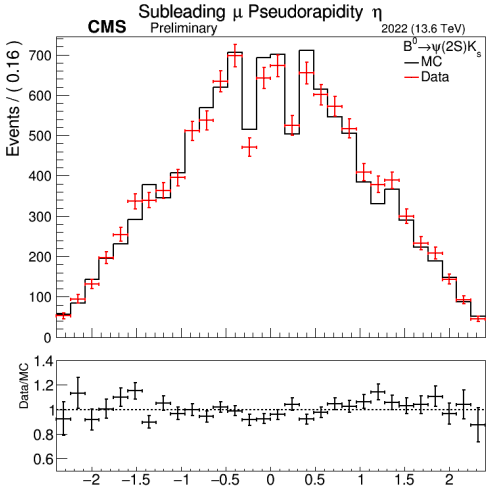
(a) sPlot of the $p_T(K^0_S)$ variable distribution.



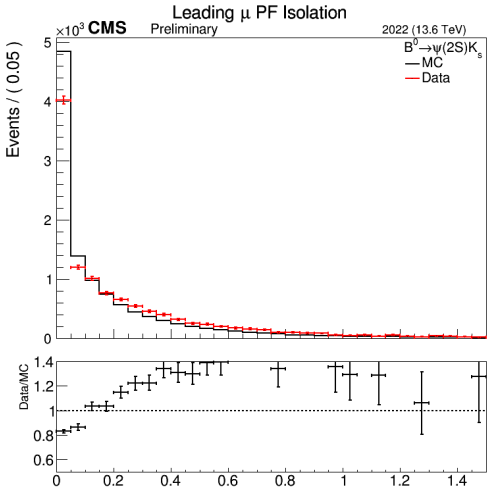
(b) sPlot of the $IP(K^0_S, 3D)$ variable distribution.



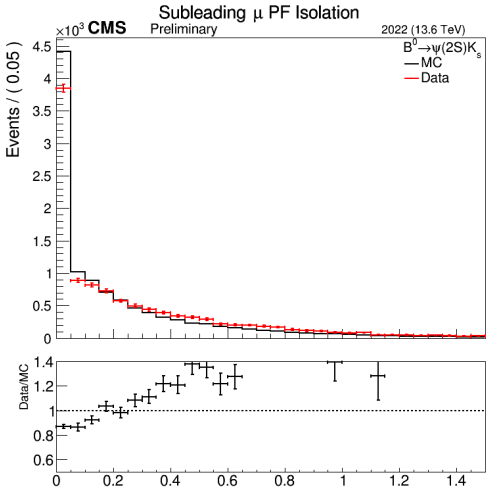
(a) sPlot of the $\eta(\mu_1)$ variable distribution.



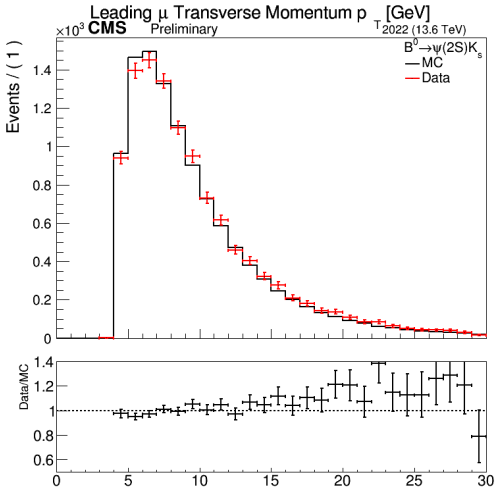
(b) sPlot of the $\eta(\mu_2)$ variable distribution.



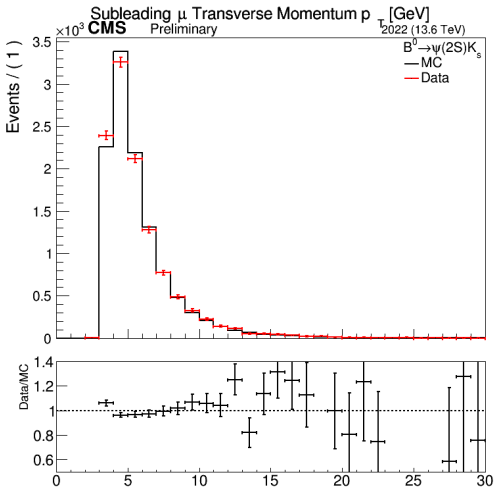
(a) sPlot of the Iso(μ_1) variable distribution.



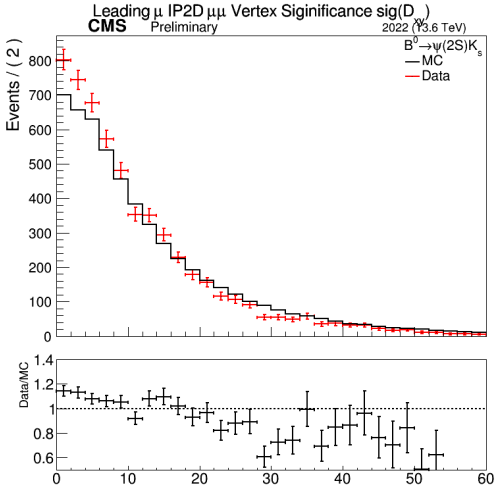
(b) sPlot of the Iso(μ_2) variable distribution.



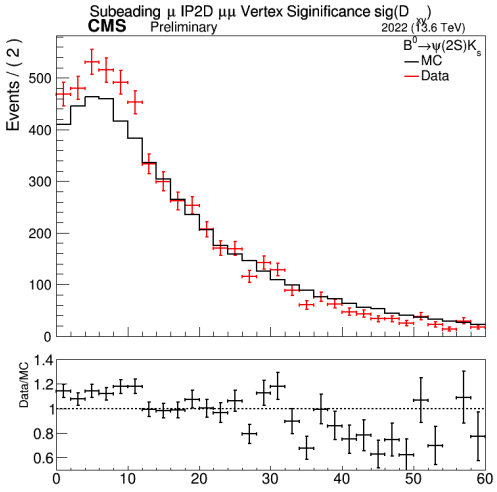
(a) sPlot of the $p_T(\mu_1)$ variable distribution.



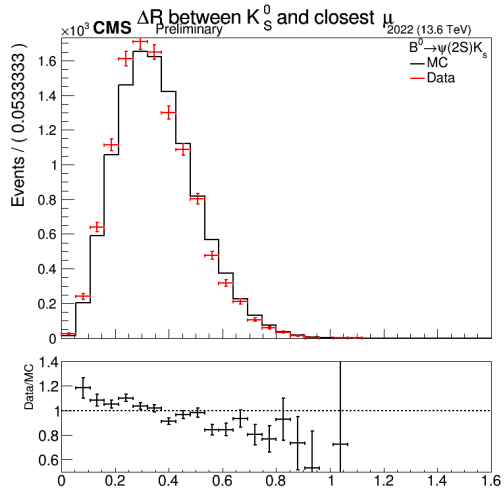
(b) sPlot of the $p_T(\mu_2)$ variable distribution.



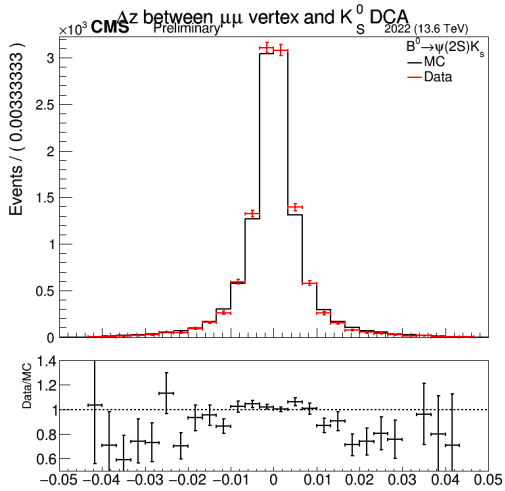
(a) sPlot of the sig(D_{xy}) leading muon variable distribution.



(b) sPlot of the sig(D_{xy}) subleading muon variable distribution.



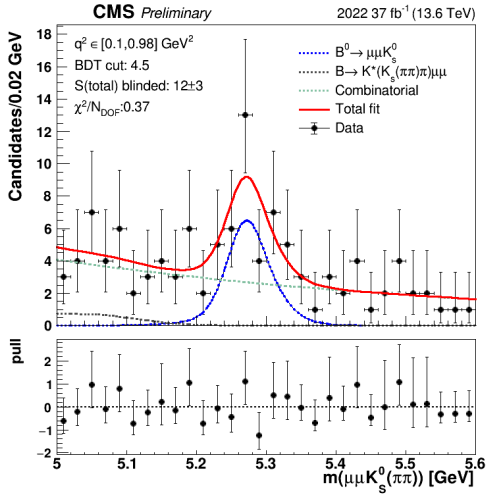
(a) sPlot of the $\Delta R(\mu, K^0_S)$ variable distribution.



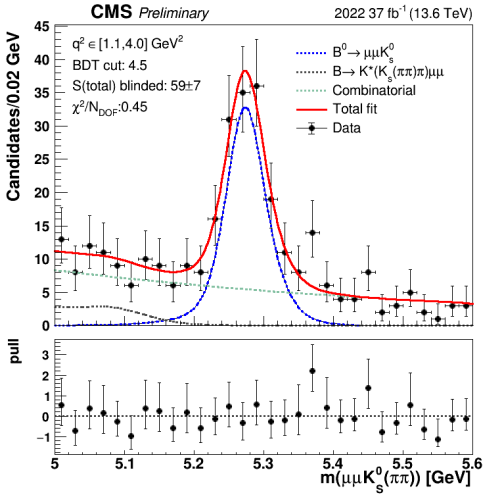
(b) sPlot of the Δz variable distribution.

B Simultaneous B^0 Mass Fits for the Rare Mode Signal Search

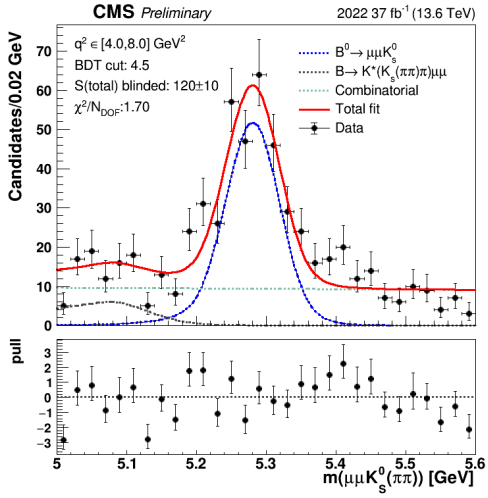
This section includes the Simultaneous B^0 mass fits for the Rare Mode in the Data Samples. The 7 bins of the search of the Rare Mode are found in table 4.1, where the bins in **bold** are excluded due to the dominant presence of the Resonant modes.



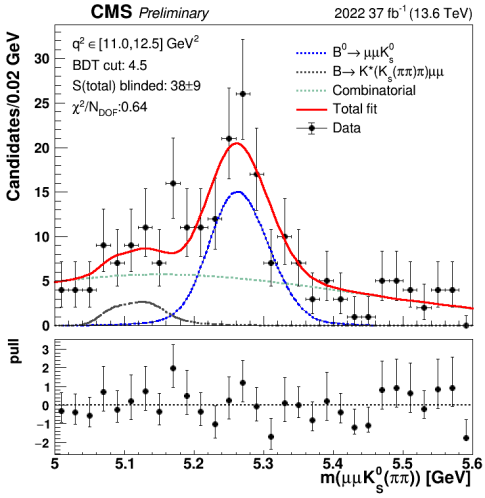
(a) Rare Mode B^0 mass fit in the [0.1, 0.98] GeV^2 q^2 bin.



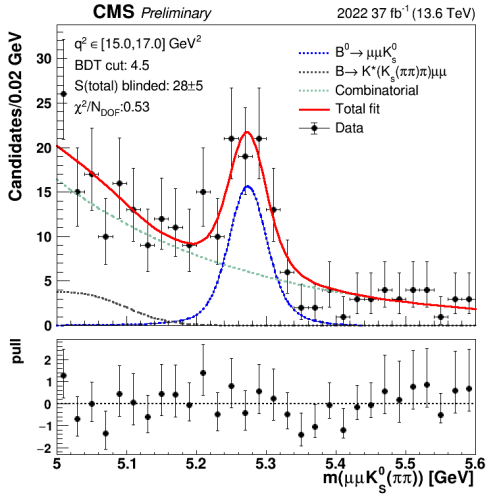
(b) Rare Mode B^0 mass fit in the [1.0, 4.0] GeV^2 q^2 bin.



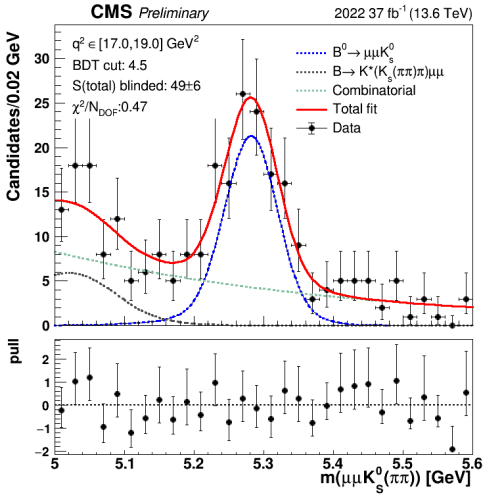
(a) Rare Mode B^0 mass fit in the [4.0, 8.0] GeV^2 q^2 bin.



(b) Rare Mode B^0 mass fit in the [11.0, 12.5] GeV^2 q^2 bin.



(a) Rare Mode B^0 mass fit in the [15.0, 17.0] GeV^2 q^2 bin.



(b) Rare Mode B^0 mass fit in the [17.0, 19.0] GeV^2 q^2 bin.

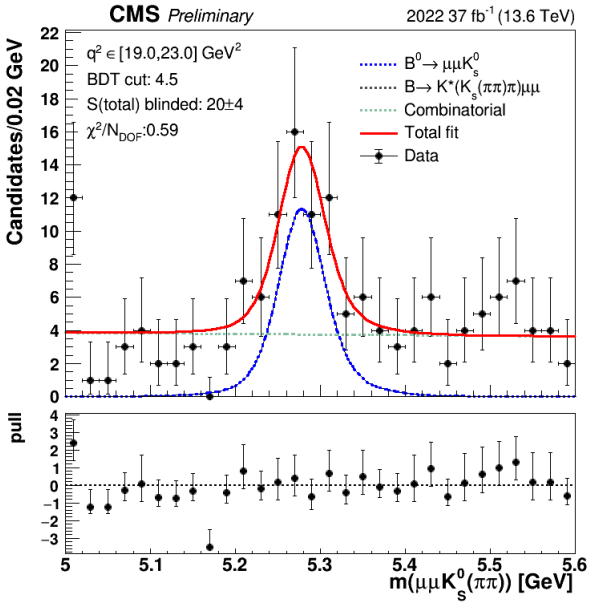


Figure B.4: Rare Mode B^0 mass fit in the [19.0, 23.0] GeV^2 q^2 bin.

C Acknowledgments

This thesis wouldn't have been possible without the help of my supervisor, prof. Costas Fountas for granting me the opportunity to work on such an analysis. He is the reason that I had a chance to work on such a project among my first two initial choices, the other being continuing work on experimental searches of Supersymmetry. I would also like to thank prof. Paris Sphicas for his extraordinary advice and guidance throughout this journey, which has helped me understand numerous aspects of Data Analysis in physics. I'd also like to give big thanks to Georgios Melachroinos for his technical and physics-related advice as well as paving the road for the analysis, step by step. Being part of a group with existing prior knowledge of an ongoing analysis has allowed multiple opportunities to flourish. Without the help of these people I wouldn't have made a debut in CMS Data Analysis as well as acquiring more knowledge regarding the Standard Model and future expansions, to complement my previous undergraduate thesis on Supersymmetry and searches in the CMS experiment.

I'd also like to express my gratitude to my friends and family for being extremely patient while I was devoting my time working on this thesis. I cannot fathom to imagine the sheer amount of patience they must've had while I kept cancelling trips and excursions, plans and get-togethers. My two most sincerest acts of gratitude go to Alex and Vasilis. Alex, you have been a fantastic cure of my tiredness and stress throughout the timeframe of working on this project, often providing a pillar of motivation that I can lean on. Vasilis, I cannot ever thank you enough for your unique sense of hangouts that took place between working on this project, providing an excellence source of relaxation. I would really love to keep listing everyone's contributions, starting from Christos, Manolis, Thanasis, Achilleas, the list goes on and on, each person offering something unique to improve on myself, from Machine Learning & A.I. advice and discussions, to how to deal with a big project such as this one. I'd also like to thank my close family, my brother and parents for being extremely patient while our interactions have become less frequent over time while working on this thesis. To all those people, once the thesis is submitted, I will get back to you three-fold at least.

Lastly, I'd like to say that working and writing this thesis in the midst of an illness and helping in managing a new Board Game Club at the University of Ioannina has been a remarkable achievement of its own. I have learned so many things and built connections to so many unknowns that beg for me to explore them, all while I have been learning where my limits are and how to push them, as well as my capabilities. I fell in love with the nature of Machine Learning models and Data Science even more that now I have a clearer path as to what I want to follow.

References

[1] Lyndon Evans. “The Large Hadron Collider”. In: *SLAC Summer Institute of Particle Physics (SSI)* (2006).

[2] CERN Education Communications & Outreach group. *CERN Annual Report 2020*. Tech. rep. CERN, 2020.

[3] Oliver Bruning et al. *LHC Design Report*. 2004.

[4] CERN. *The Large Hadron Collider*. https://home.cern/science/accelerators/large-hadron-collider. Accessed: 2023-12-25. 2023.

[5] CERN. *Linear Accelerator 4*. https://home.cern/science/accelerators/linear-accelerator-4. Accessed: 2023-12-25. 2023.

[6] CERN. *The Proton Synchrotron Booster*. https://home.cern/science/accelerators/proton-synchrotron-booster. Accessed: 2023-12-25. 2023.

[7] CERN. *The Proton Synchrotron*. https://home.cern/science/accelerators/proton-synchrotron. Accessed: 2023-12-25. 2023.

[8] CERN. *The Super Proton Synchrotron*. https://home.cern/science/accelerators/super-proton-synchrotron. Accessed: 2023-12-25. 2023.

[9] CERN. *Accelerating: Radiofrequency cavities*. https://home.cern/science/engineering/accelerating-radiofrequency-cavities. Accessed: 2023-12-25. 2023.

[10] David J. Griffiths. *Introduction to Classical Electrodynamics Fourth Edition*. Pearson Education, 2013.

[11] G. L. Bayatian et al. “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. In: (2006).

[12] Oliver S. Bruning. “LHC luminosity and energy upgrade: A feasibility study”. In: (Dec. 2002). Ed. by R. Cappi et al.

[13] Dave Barney. “An overview of the CMS experiment for CERN guides”. In: *CMS Document Server* (2003).

[14] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Institute of Physics Publishing and SISSA* 3.S08004 (2008).

[15] I .L. Horvath et al. “The CMS Conductor”. In: *IEEE Transactions On Applied Superconductivity* 10.1 (Mar. 2000).

[16] The CMS Collaboration. “The CMS Trigger System”. In: *Journal of Instrumentation* (2017). arXiv: 1609.02366v2 [physics.ins-det].

[17] Cristina Biino. “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections”. In: *Journal of Physics* 587.012001 (2015).

[18] Petar Adzic et al. *Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter*. Tech. rep. Geneva: CERN, 2007. DOI: 10.1088/1748-0221/2/04/P04004. URL: https://cds.cern.ch/record/1009081.

[19] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[20] P Krivkova and R Leitner. *Measurement of the interaction length of pions and protons in the TileCal Module 0*. Tech. rep. Geneva: CERN, 1999. URL: http://cds.cern.ch/record/683812.

[21] G Baiatian et al. *Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges*. Tech. rep. 1. Geneva: CERN, 2008. DOI: 10.1140/epjc/s10052-008-0573-y. URL: https://cds.cern.ch/record/1049915.

[22] S Chatrchyan et al. "The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment". In: *JINST* 3 (2008). Also published by CERN Geneva in 2010, S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://cds.cern.ch/record/1129810.

[23] Francesca Cavallari. "Performance of calorimeters at the LHC". In: *Journal of Physics: Conference Series* 293.1 (Apr. 2011), p. 012001. DOI: 10.1088/1742-6596/293/1/012001. URL: https://dx.doi.org/10.1088/1742-6596/293/1/012001.

[24] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[25] CERN. *Drift Tubes*. https://cms.cern/detector/detecting-muons/muon-drift-tubes. Accessed: 2024-3-15. 2023.

[26] CERN. *Cathode Strip Chambers*. https://cms.cern/detector/detecting-muons/cathode-strip-chambers. Accessed: 2024-3-15. 2023.

[27] CERN. *Resistive Plate Chambers*. https://cmsexperiment.web.cern.ch/index.php/detector/detecting-muons/resistive-plate-chambers. Accessed: 2024-3-15. 2023.

[28] CERN. *Gas Electron Multiplier*. https://cms.cern/detector/detecting-muons/gas-electron-multiplier. Accessed: 2024-3-16. 2023.

[29] A Colaleo et al. *CMS Technical Design Report for the Muon Endcap GEM Upgrade*. Tech. rep. 2015. URL: https://cds.cern.ch/record/2021453.

[30] *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. This is the final version, approved by the LHCC. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2283189.

[31] A.M. Sirunyan, A. Tumasyan and W. Adam. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$TeV". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. URL: https://dx.doi.org/10.1088/1748-0221/15/10/P10017.

[32] Daniele Trocino. "The CMS High Level Trigger". In: *Journal of Physics: Conference Series* 513.1 (June 2014), p. 012036. DOI: 10.1088/1742-6596/513/1/012036. URL: https://dx.doi.org/10.1088/1742-6596/513/1/012036.

[33] *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*. Tech. rep. 2010. URL: http://cds.cern.ch/record/1247373.

[34] Milos Dordevic. "The CMS High Level Trigger System". In: *PoS* BPU11 (2023), p. 088. DOI: 10.22323/1.427.0088.

[35] Mark Thomson. *Modern Particle Physics*. John Wiley and Sons, 2013.

[36] Timo Weigand. *Quantum Field Theort I + II*. http://www.physics.umd.edu/grt/taj/624b/WeigandQFT.pdf.

[37] UAI Collaboration. "Experimental Observation of isolated large transverse energy electrons with associated missing energy at sqrt(s) = 540 GeV". In: *Z. Phys. C - Particles and Fields* 44.15-61 (1989).

[38] UA Collaboration. “Studies of intermediate vector boson production and decayin UAI at the CERN proton-antiproton collider”. In: *Physics Letters* 122B.1 (1983).

[39] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Physics Review* 105.1413 (1957).

[40] M.E. Peskin. *Concepts of Elementary Particle Physics*. Clarendon Press, 2019.

[41] Tom Lancaster and Stephen J. Blundell. *Quantum Field Theory for the Gifted Amateur*. Oxford University Press, 2014.

[42] Karen Lingel, Tomasz Skwarnicki and James G. Smith. “PENGUIN DECAYS OF B MESONS”. In: *Annual Review of Nuclear and Particle Science* 48.1 (Dec. 1998), pp. 253–306. ISSN: 1545-4134. DOI: 10.1146/annurev.nucl.48.1.253. URL: http://dx.doi.org/10.1146/annurev.nucl.48.1.253.

[43] Michael Gronau and David London. “New physics in CP asymmetries and rare B decays”. In: *Physical Review D* 55.5 (Mar. 1997), pp. 2845–2861. ISSN: 1089-4918. DOI: 10.1103/physrevd.55.2845. URL: http://dx.doi.org/10.1103/PhysRevD.55.2845.

[44] Spyros Nessis. *Feasibility study of displaced $\mu^{\pm}$ final states from compressed electroweak SUSY $\tilde{\chi}_1^{\pm}$ decays in Monte Carlo generated pp beam collisions*. No public link available.

[45] A. Ali. “Flavour changing neutral current processes in B decays”. In: *Nuclear Physics B - Proceedings Supplements* 59.1–3 (Nov. 1997), pp. 86–100. ISSN: 0920-5632. DOI: 10.1016/s0920-5632(97)00431-3. URL: http://dx.doi.org/10.1016/S0920-5632(97)00431-3.

[46] Christoph Bobeth, Mikolaj Misiak and Joerg Urban. “Photonic penguins at two loops and m-dependence of BR(B to Xl+l-)”. In: *Nuclear Physics B* 574.1–2 (May 2000), pp. 291–330. ISSN: 0550-3213. DOI: 10.1016/s0550-3213(00)00007-9. URL: http://dx.doi.org/10.1016/S0550-3213(00)00007-9.

[47] Wolfgang Altmannshofer et al. “Symmetries and asymmetries of B to K*$\mu^+\mu^-$ decays in the Standard Model and beyond”. In: *Journal of High Energy Physics* 2009.01 (Jan. 2009), pp. 019–019. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2009/01/019. URL: http://dx.doi.org/10.1088/1126-6708/2009/01/019.

[48] Georgios Melachroinos. *Questions regarding dBR/dq^2 distribution for $B+ \to \mu\mu K^+$ decay*. https://github.com/flav-io/flavio/issues/198f.

[49] B.R. Martin and G. Shaw. *Particle Physics Third Edition*. John Wiley and Sons, 2008.

[50] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[51] Alex Seuthe. *Lepton flavour universality tests at LHCb*. 2023. arXiv: 2305.08216 [hep-ex]. URL: https://arxiv.org/abs/2305.08216.

[52] Torbjorn Sjostrand, Stephen Mrenna and Peter Skands. “PYTHIA 6.4 Physics and Manual”. In: *Journal of High Energy Physics* 2006.05 (2006).

[53] Torbjorn Sjostrand et al. “An Introduction to PYTHIA 8.2”. In: *Computer Physics Communications* (2015). arXiv: 1410.3012.

[54] Rene Brun, Fons Rademarkers and Suzanne Panacek. “ROOT, An Objected Oriented Data Analysis Framework”. In: *inspireHEP* (2000).

[55] Robert Bainbridge and Oliver Buchmueller and Vincenzo Cacchio and Francesca Cavallari and Sara Fiorendi and Georgios Karathanasis and Greg Landsberg and Ka Tung Lau and Riccardo Manzoni and Nancy Marinelli and Arabella Martelli and Georgios Melachroinos and Chiara Rovelli and Paris Sphicas. *Measurement of the R(K) ratio with the 2018 CMS Parked Data*. Not available for public viewing. Request access my sending a message. 2021.

[56] E. Daw. *Lecture 7 - Rapidity and Pseudorapidity*. https://www.hep.shef.ac.uk/edaw/PHY206/Site/2012_course_files/phy206rlec7.pdf. 2012.

[57] Raghunath Sahoo. "Relativistic Kinematics". In: *Indian Institute of Technology Indore* (2016). arXiv: 1604.02651v1 [nucl-ex].

[58] D Martinez Santos. "Study of the very rare decay $B_s \to \mu^+\mu^-$ in LHCb". Presented on 05 May 2010. Santiago de Compostela, Universidade de Santiago de Compostela, 2010. URL: https://cds.cern.ch/record/1264603.

[59] Yann Coadou. "Boosted Decision Trees". In: *Artificial Intelligence for High Energy Physics*. WORLD SCIENTIFIC, Feb. 2022, pp. 9–58. ISBN: 9789811234033. DOI: 10.1142/9789811234033_0002. URL: http://dx.doi.org/10.1142/9789811234033_0002.

[60] L. Breiman et al. "Classification and Regression Trees". In: *Biometrics* 40 (1984), p. 874. URL: https://api.semanticscholar.org/CorpusID:29458883.

[61] Miao Rong, Dunwei Gong and Xiaozhi Gao. "Feature Selection and Its Use in Big Data: Challenges, Methods, and Trends". In: *IEEE Access* 7 (2019), pp. 19709–19725. DOI: 10.1109/ACCESS.2019.2894366.

[62] M. Pivk and F.R. Le Diberder. "$_s$Plot: A statistical tool to unfold data distributions". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1–2 (Dec. 2005), pp. 356–369. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.08.106. URL: http://dx.doi.org/10.1016/j.nima.2005.08.106.

[63] flavio team. *flavio*. https://flav-io.github.io/. Accessed: 2024-8-13. 2024.

[64] Robert Bainbridge et al. *Measurement of the R(K) ratio with the 2018 CMS Parked Data, version 13*. No public link available. 2021.

[65] The CMS Collaboration. "Missing transverse energy performance of the CMS detector". In: *Journal of Instrumentation* 6.09 (2011).

[66] M.E. Peskin and D.V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 2018.

[67] Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Cambridge University Press, 1984.

[68] Jacob Scwichtenberg. *Physics from Symmetry*. 2nd ed. Springer, 2018.