

Search for Rare Charm Decays into Two Muons

by

Julius Heitkoetter

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE IN PHYSICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2025

© 2025 Julius Heitkoetter. This work is licensed under a [CC BY-NC-ND 4.0](#) license.
The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free
license to exercise any and all rights under copyright, including to reproduce, preserve,
distribute and publicly display copies of the thesis, or release the thesis under an
open-access license.

Authored by: Julius Heitkoetter
Department of Physics
May 9, 2025

Accepted by: Lindley Winslow
Associate Professor of Physics
Associate Head, Department of Physics

Search for Rare Charm Decays into Two Muons

by

Julius Heitkoetter

Submitted to the Department of Physics

on May 9, 2025 in partial fulfillment of the requirements for the degree of

BACHELOR OF SCIENCE IN PHYSICS

ABSTRACT

A search for the rare charm decay $D^0 \rightarrow \mu^+ \mu^-$ is presented. This decay is an example of a flavor changing neutral current, which serves as an excellent probe into new Beyond the Standard Model physics due to its suppressed Standard Model nature. The analysis is performed using proton-proton collision data collected from the CMS detector at the Large Hadron Collider in 2022 and 2023, with two datasets at a center of mass energy of $\sqrt{s} = 13.6$ TeV and integrated luminosity of 51.415 nb^{-1} and 64.525 fb^{-1} . The analysis uses D^0 mesons originating from $D^{*\pm} \rightarrow D^0 \pi^\pm$ decays, which allow for better signal selection and less contribution from combinatorial backgrounds. It uses rates which are normalized to the common $D^0 \rightarrow \pi^+ \pi^-$ decay rate to calculate its results, which allows for a strategic reduction of systematic uncertainties and lack of need for a D^0 cross-section. Pions decaying in-flight to muons cause a significant difficulty for the analysis which is carefully studied and understood. No obvious excess is observed. Upper limits on the $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-)$ decay at a 90(95)% confidence level are found to be $2.1(2.4) \times 10^{-6}$, outperforming the current world-best limit by 35%.

Thesis supervisor: Christoph Paus

Title: Professor of Physics

Acknowledgments

To begin, I would like to acknowledge that the analysis described in this thesis was a large effort carried out by many others at CERN beyond just myself. My primary focus during the analysis was in developing the muon fake rate, helping build the multivariate analysis tools, and analyzing backgrounds, which wouldn't have culminated to much had it not been for the work of others. To that end, I'd like to specifically thank Dima Kovalskyi, Christoph Paus, Zhangqier Wang, Kai-Feng Chen, Wei-Hsiang Huang, and Priyanka Sadangi for their contributions to the rest of the analysis. I hope this thesis properly represents the results of all your hard work.

I would like to specifically thank Christoph for his amazing role as a mentor to me during my time here at MIT. He certainly has had an enormous impact in kindling my curiosity and fascination of physics.

I'd also like to specifically thank Zhangqier for the many, many hours he has invested in me and for all his patience as I face the struggles an undergraduate typically faces while trying to understand particle physics.

The friends I've made at the PPC group have made every day I spent in lab enjoyable and memorable. Thanks Luca, Pietro, Ludo, Marti, Kyungseop, and Chad for all the good times.

I thank all the other groups I've briefly been a part of, including ETP at KIT and Phillip Harris' group at MIT. I've profited significantly from the diversity of talented mentorship I've received over the years from all of you.

My profound appreciation goes to the many professors and TAs who spent countless hours trying to teach me physics. A special thank you to Daniel Harlow, Netta Engelhardt, and Sam Alipour-fard for their amazing commitment to teaching and unbelievable insights.

Of course, no acknowledgment would be complete without thanking my family Vivian, Constanze, and Markus for their endless love and support; I am nothing if not for y'all.

Lastly, my love for physics has come from a love of learning from and hanging out with the people I have met along the way. Though I cannot name everyone here, I am deeply grateful to all of you.

Contents

Title page	1
Abstract	2
Acknowledgments	3
List of Figures	6
List of Tables	8
1 Introduction	9
2 Theory and Motivation	12
2.1 The Theory of Our Universe	12
2.2 Quantum Field Theory	13
2.2.1 Correlation Functions	14
2.2.2 Scattering	14
2.3 The Standard Model	16
2.3.1 Quantum Chromodynamics	16
2.3.2 Electroweak Interactions	18
2.4 The $D^0 \rightarrow \mu\mu$ decay	21
2.4.1 Flavor Changing Neutral Currents	21
2.4.2 The $D^0 \rightarrow \mu\mu$ decay as a FCNC	23
3 The CMS Detector	25
3.1 The Large Hadron Collider	25
3.2 The CMS Detector	27
3.2.1 Solenoid Magnet	28
3.2.2 Silicon Tracker	29
3.2.3 Electromagnetic Calorimeter	29
3.2.4 Hadronic Calorimeter	30
3.2.5 Muon Chambers	31
3.3 Triggers	32

4	$D^0 \rightarrow \mu\mu$ Branching Fraction Measurement	33
4.1	Analysis overview	33
4.2	Triggers and datasets	34
4.2.1	Data samples	34
4.2.2	Backgrounds	35
4.2.3	Monte Carlo samples	36
4.3	Selections and efficiency	37
4.3.1	Preselection	39
4.3.2	Baseline selection	39
4.3.3	Multivariate Analysis	41
4.4	Unbinned Maximum Likelihood Fits	46
4.4.1	Normalization Channel Fit	47
4.4.2	Signal Channel Fit	51
4.5	Muon Fake Rates	54
4.5.1	Data Samples and Selections	55
4.5.2	Measurement Procedure	56
4.6	Efficiency Corrections	59
4.6.1	Trigger Efficiency Correction	60
4.7	Results	61
4.7.1	The CL_s Method	61
4.7.2	Sensitivity Scan	65
4.7.3	Branching Fraction Limit	66
5	Conclusion	68
References		70

List of Figures

2.1	An example feynman diagram describing the $\mu \rightarrow \nu_\mu \nu_e e$ decay.	15
2.2	The particles of the Standard Model	17
2.3	The Mexican hat potential, given by $f(r, \theta) = (r^2 - v^2)$, demonstrating ground state symmetry breaking in a symmetric potential	20
2.4	Two Feynman diagrams contributing to rare D^0 decays: (left) the leading SD contribution in the form of a penguin diagram and (right) the leading LD contribution, depicting the two-photon contribution. Note that the $D^0/\gamma/\gamma$ vertex hides additional complexity in the Feynman diagram, which is calculated by BGHP in the form of many SD and LD contributions combined and is abstracted away in the above diagrams.	24
3.1	The CERN accelerator complex [12]	26
3.2	A labeled cutaway diagram of the CMS detector [15]	28
3.3	The CERN accelerator complex	29
3.4	A schematic example of an electromagnetic shower caused by bremmstrahlung and pair production.	30
4.1	Comparison of normalized frequencies for various peaking background $D^0 \rightarrow \mu\mu$ decays against the signal decay. Right: Reconstructed Δm distribution. Left: Reconstructed dimuon mass distribution, where decay products are forced to take the muon mass using <code>TGenPhaseSpace</code> , even for non-dimuon final states.	37
4.2	An example of the $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow \mu\mu$ vertex	40
4.3	Correlation matrix over the MVA training variables for training background events generated from data (left) and training signal events generated from MC (right)	43
4.4	Correlation matrix comparing the MVA classification parameter to Δm and the dimuon mass. Shown left is the training background events generated from data and shown right is the training signal events generated from MC.	44
4.5	A graphical summary of MVA_D cut efficiency at various diverse cut values on various samples used throughout the analysis.	45
4.6	The signal model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.	47
4.7	The $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow K^\pm \pi^\mp$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.	49

4.8	The combinatorial model fit on data side-bands, with $m(D^0)$ displayed left and Δm displayed right.	50
4.9	The full $D^0 \rightarrow \pi^\pm \pi^\pm$ model fit on HLT_ZeroBias data, with $m(D^0)$ displayed left and Δm displayed right.	50
4.10	The full $D^0 \rightarrow \pi^\pm \pi^\pm$ model fit on HLT_DoubleMu4_3_LowMass data, with $m(D^0)$ displayed left and Δm displayed right. The green line follows the combinatorial model, the blue line follows the signal model, the red line follows the $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow K^\pm \pi^\mp$ background model, the red dashed line following the non- $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow K^\pm \pi^\mp$ background model, and the blue dashed line follows the non- $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow \pi^\pm \pi^\pm$ background model.	51
4.11	The signal model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.	52
4.12	The $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow p i^\pm \mu \nu_\mu$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.	53
4.13	The $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu \nu_\mu \mu \bar{\nu}_\mu$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.	54
4.14	The fit to partial unblinded data before any MVA_D cut, demonstrating the stability of the combinatorial shape which dominates the overall shape.	55
4.15	Mass projection from $K_S^0 \rightarrow \pi^\pm \pi^\pm$ decays in the μ p_T range 4-40.0 GeV using 2022 data before (left) and after (right) soft muon MVA ID cut. The green dashed line is for the signal distribution, the red dotted line is for combinatorial background, and the result of the fit projection is the solid pink line. The data are shown with solid black circles.	57
4.16	Pion muon fake rates for Run2022 (top) and Run2023 (bottom) as a function muon p_T	57
4.17	Fake rate dependence on flight length (lxy) measured in cm for data and MC	58
4.18	Difference in fakerate between Egamma and Parking Electron data (top) and W plus jets, Drell-Yan, and $t\bar{t}$ MC samples (bottom) binned by flight length in cm (left) and pT in Gev (right)	59
4.19	A visualization of the CL_s method, using idealized data and a given signal strength, s	64
4.20	The sensitivity scan against various MVA cuts first within a broad range of MVA cut values (left) and then with a more concentrated range of MVA cut values (right).	66
4.21	The unblind fit on observed data shown in $m(D^0)$ (left) and Δm (right). Note that the peaking decays are visually scaled by a factor of 20.	67

List of Tables

2.1	Electric charge and mass of the six quark flavours. Masses are quoted in the schemes recommended by the Particle Data Group [13].	18
2.2	Weak isospin (T_3), weak hypercharge (Y), and electric charge (Q) for the chiral fermions of one Standard Model generation [13].	22
4.1	Integrated luminosity of data collected with the primary triggers.	35
4.2	The 4 most relevant MC simulation samples and their defining attributes. . .	38
4.3	Summary of MVA_D cut efficiencies at various cut values on various samples used throughout the analysis.	44
4.4	Fitted yields and associated uncertainties for the normalization UML fit. . .	51
4.5	Muon fake rate ratios for Data over MC simulation for 2022 and 2023 in the bins of meson p_{T}	60
4.6	Trigger efficiency measurements in Data and MC for 2022 and 2023.	61
4.7	The final observed yield, based on data.	66

Chapter 1

Introduction

The Standard Model (SM) of particle physics is one of the most successful scientific theories ever developed by humans. It accounts for the electromagnetic, weak, and strong forces, intertwining them in a single mathematical framework. It predicts the behavior of fundamental physics with incredible precision, as demonstrated in numerous predictions such as the discovery of the Higgs boson and the measurement of the magnetic moment of the electron. Furthermore, these predictions hold over a large range of energies and length scales, earmarking the success of the theory.

However, we know that the SM cannot be the full story. For example, it contains no explanation for dark matter, which has been found to make up the majority of matter in the universe. It also does not explain the large matter-antimatter symmetry that allows for the stable matter we rely on to exist. Lastly, and perhaps most alarming, it is inconsistent when combined with gravity in a quantum gravity model.

Even within the scope of the SM, it is unclear why there are 3 generations of quarks, why the particles each have the masses that they do, or why only certain interactions are allowed. Many of these questions can be explained away by considering the parameters they require (such as the quark masses) as fundamental constants of the universe. However, many of these puzzles could be clues into a deeper understanding of the structure of the SM. In the modern understanding of physics, this *structure* refers not to the equations of motion, but rather to the exact geometrical symmetries of our universe. For example, the SM can be derived from a $SU(3) \times SU(2) \times U(1)$ symmetry structure, as outlined later in this thesis. Finding deeper structures is therefore often equivalent to probing for new symmetries, such as is done with a common extension of the SM known as the Super Symmetry (SUSY) model.

Finding deeper structure does not always begin with large signature discoveries. Often, it begins with considering edges of phase space where the SM heavily suppresses processes, leading to not only robustly small but also incredibly fragile theoretical expectations. Even extremely small signals measured in these places can lead to incredible insight into the structure of our universe.

Therefore, one promising approach to searching for new physics is the study of rare processes that are highly suppressed in the Standard Model (SM). Such processes offer a unique opportunity to identify small deviations from SM predictions that may indicate contributions from Beyond the Standard Model (BSM) physics. One such suppressed process is Flavor Changing Neutral Currents (FCNC), which are hadronic decay processes not

allowed at leading order in the SM as a consequence of none of the two neutrally charged electroweak bosons (Z and γ) being flavor changing. However, there are many Beyond the Standard Model (BSM) theories that allow for FCNCs at tree level, such as SUSY or Flavor Non-Universal Z' Bosons.

Of the six quarks in the SM (u, d, c, s, t, b), the up/down quarks cannot decay because they are the lightest generation quarks. The top quark, due to its large mass, is extremely short-lived and decays into a W boson and a bottom quark before it can form hadrons. Decays of hadrons composed of bottom and strange quarks have been extensively researched, while decays of charmed hadrons receive less attention due to a smaller branching fraction in the $c \rightarrow u$ decay relative to the $b \rightarrow s$ or $s \rightarrow d$ decays. This makes it particularly interesting to study FCNCs of charmed, or D , mesons.

One such accelerator experiment is the Compact Muon Solenoid (CMS), one of two large general-purpose detectors at the Large Hadron Collider (LHC). CMS specializes in muon detection, making a dimuon ($\mu^+ \mu^-$) final state a useful choice for a rare decay. Therefore, this thesis looks to search for the rare decay of a neutrally charged D meson into a dimuon final state, or more specifically set an upper limit on $\mathcal{B}(D^0 \rightarrow \mu\mu)$ [1].

Previous work studying $\mathcal{B}(D^0 \rightarrow \mu\mu)$ has been most successfully done by the Large Hadron Collider beauty (LHCb) experiment, achieving an upper limit of the branching fraction at 3.1×10^{-9} at a 90% confidence level [2]. Current leading theoretical work places the SM prediction of the branching fraction at 3×10^{-13} [3], leaving an unexplored region of 4 orders of magnitude to probe for new physics.

In this thesis, I run an analysis to measure the upper limit on $\mathcal{B}(D^0 \rightarrow \mu\mu)$. The main challenge of the analysis is the large amount of background events, making it difficult to isolate signal events. The main approach to reduce the background is to look at the cascade decays of $D^{*\pm} \rightarrow D^0 \pi^\pm \rightarrow (\mu\mu)\pi^\pm$, which produces a characteristic additional pion that helps to tag and reconstruct the event more cleanly.

This technique, while increasing the signal-to-background ratio, has two main obstacles. The first is the muon fake rate, or the decay of $\pi \rightarrow \mu\nu$, where ν has low enough energy that the muon and pion have essentially the same 4-momentum, causing the detector to falsely reconstruct a muon as the product of the decay, instead of recognizing a pion. This causes the more common $D^0 \rightarrow \pi^\pm \pi^\pm$ decay to be misreconstructed as a $D^0 \rightarrow \mu\mu$ event. The second obstacle is that canonically, to get $\mathcal{B}(D^0 \rightarrow \mu\mu)$, one would count the number of $D^0 \rightarrow \mu\mu$ events, $N_{D^0 \rightarrow \mu\mu}$, and divide by the number of D^0 mesons produced in the detector, N_{D^0} . However, N_{D^0} is not well known at CMS. Instead, $\mathcal{B}(D^0 \rightarrow \pi^\pm \pi^\pm)$ is very well known, meaning that if one measures $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$ in a normalization channel, then one can construct

$$\mathcal{B}(D^0 \rightarrow \mu\mu) = \mathcal{B}(D^0 \rightarrow \pi^\pm \pi^\pm) \frac{N_{D^0 \rightarrow \mu\mu}}{N_{D^0 \rightarrow \pi^\pm \pi^\pm}} \times \text{efficiency correction ratio}$$

where the dominating challenge lies in properly tracking the efficiency correction ratio.

This concludes an introduction into rare charm decays. The rest thesis is structured as follows:

Chapter 2 of this thesis lays the theoretical groundwork on which the remainder of the thesis is built and gives insight to why the rare decay $D^0 \rightarrow \mu\mu$ is a valuable probe for new physics. It begins with a short survey of Quantum Field Theory (QFT), outlining how the

adoption of special relativity into quantum mechanics gives rise to a QFT characterized by a Lagrangian and resulting correlation functions that can be calculated perturbatively, using tools such as path integrals, Feynman diagrams, and the Lehmann Symanzik Zimmermann (LSZ) reduction formulation to extract scattering amplitudes. The chapter then describes the SM, beginning with its local gauge symmetry structure and building the fundamental particles and interactions from the resulting Lagrangian, with special attention paid to the electroweak sector and the mechanism by which the Higgs field gives mass to the W^\pm and Z bosons. This chapter also applies the theoretical groundwork to motivate the $D^0 \rightarrow \mu\mu$ search by showing that FCNCs are forbidden at tree level in the SM, with loop contributions predicting a branching fraction of $\simeq 10^{-13}$. It uses this discussion to conclude in outlining how the $D^0 \rightarrow \mu\mu$ decay can be used as a powerful probe of new physics.

Chapter 3 of this thesis describes the experimental foundation used to measure the $D^0 \rightarrow \mu\mu$ branching fraction, beginning with a summary of the Large Hadron Collider at CERN. The chapter then describes the CMS detector, starting with an overview of its geometry and coordinate system before detailing each of its subsystems. For each of these subsystems, the chapter outlines how the engineering of the detectors impacts the physics read-out from them. The next section explains how the CMS experiment uses a trigger system to balance the large volume of data with the need to capture rare particle interactions. Finally, the chapter discusses how raw data is reconstructed into particle events and how simulation is used for analysis and to benchmark detector response and acceptance.

Chapter 4 of this thesis describes the experiment performed to measure the branching fraction of the $D^0 \rightarrow \mu\mu$ decay. It begins with an overview of the analysis strategy, before diving into the datasets and simulation samples. It then describes the reconstruction and event selection methods used to generate the signal and normalization datasets, as well as details the multivariate analysis boosted decision tree framework used to better identify signal events. Then, the chapter covers the calculation of $N_{D^0 \rightarrow \mu\mu}$ and $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$ events using an unbinned maximum likelihood fit. Then, this chapter covers the efficiency corrections derived from simulation samples, namely the trigger efficiency, the muon reconstruction efficiency, and the muon fake rate. Lastly, the chapter systematic uncertainties as well as the CL_s method used to calculate the final result, before presenting the final result.

Chapter 5 of this thesis gives a short conclusion on the findings of the analysis performed in this thesis.

This thesis ends with a conclusion of the experimental study, summarizing and giving commentary on its results.

Chapter 2

Theory and Motivation

This chapter lays the theoretical groundwork on which the remainder of the thesis is built and gives insight into why the rare decay $D^0 \rightarrow \mu\mu$ is a valuable probe for the potential new physics. The chapter begins with a short survey of Quantum Field Theory (QFT), outlining how the adoption of special relativity into quantum mechanics gives rise to a QFT that is characterized by a Lagrangian and a resulting correlation function that can be calculated perturbatively. This framework then describes the computational tools for calculating these correlation functions, such as path integrals, Feynman diagrams, and the LSZ formulation, which allows the calculation of scattering amplitudes from correlation functions.

Next, this chapter describes the Standard Model (SM), beginning with its local gauge symmetry structure and then building the fundamental particles and interactions from the resulting Lagrangian. Special attention is paid to the electroweak portion of the Lagrangian, giving an outline of the mechanism by which the Higgs field gives mass to the W^\pm and Z bosons, important mediators in the $D^0 \rightarrow \mu\mu$ decay.

Lastly, this chapter applies the theoretical groundwork established in its first two sections as a motivation for the $D^0 \rightarrow \mu\mu$ search. Namely, FCNCs are defined and calculated to not be allowed at tree level in the SM. Loop contributions to the $D^0 \rightarrow \mu\mu$ decay are shown to bring the branching fraction of the decay rate to $\simeq 10^{-13}$. The chapter ends by giving motivation as to why this makes the $D^0 \rightarrow \mu\mu$ decay a probe of new physics and provides a brief survey of the new physics that this probe could illuminate.

2.1 The Theory of Our Universe

Since its formulation in the early 1970's, the Standard Model of Particle Physics (SM) has become one of the most successful physics theories ever conceived. Not only does it describe 3 out of the 4 fundamental forces in our universe, but in the past 50 years it has explained virtually all small length scale experimental results and has made some of the most precise predictions in all of physics. For example, it predicted the discovery of the Higgs Boson that occurred in 2012 [7][5] and it predicted the anomalous magnetic dipole moment as $a = 0.00115965218059(13)$ [10], which results in a prediction of the fine structure constant that has a precision of better than one part in a billion.

However we know that the SM cannot be the complete theory of our universe due to a

few, major shortcomings. Perhaps the largest of these is the lack of gravity in its description of physics, meaning it cannot describe any long distance cosmological observations. Other problems also persist at small length scales, such as the observation of neutrino oscillations [9] which cannot occur using the massless neutrinos that the SM predicts.

Therefore, physics analyses are often in search of BSM physics at small length scales to build a theory of particle physics that could solve many of the problems with the SM. Importantly, the SM is a quantum field theory,

2.2 Quantum Field Theory

Heuristically speaking, Quantum Field Theory (QFT)[11] is a mathematical framework developed to unify the theories of classical special relativity and non-relativistic quantum mechanics. In the late 1920s, non-relativistic quantum mechanics had developed to a mature theory, modeling many of the phenomena that the physics of the 19th century simply couldn't explain. However, when put into context with Einstein's theory of special relativity, there were two central problems:

1. Velocities and momenta are strictly non-relativistic
2. Interactions are instantaneous.

Fixing these two issues forces the fundamental building block of any QFT to be fields, not particles, that obey causality. Specifically, physics in a QFT is modeled as interactions between dynamical fields over spacetime coordinates.

In order to force compliance with special relativity and classical mechanics, the fields in any QFT must have three global symmetries:

1. Translational symmetry along both spatial and temporal coordinates
2. Rotational symmetry along two spatial coordinates
3. Rotational symmetry along a spatial coordinate and the temporal coordinate

The last symmetry gives rise to Lorentz transformation, which describe how physics is invariant under transformations between inertial reference frames, one of the central properties of special relativity. These three global symmetries are often grouped together as global Poincaré symmetry, which lays the framework for virtually all of QFT.

To construct a QFT, the Lagrangian is first written down in the most general form possible that satisfies global Poincaré symmetry as well as any local symmetries that should be preserved. Importantly, the Lagrangian must be renormalizable, meaning that divergences can be absorbed into the parameters of your model to give real, physical interpretations of the results.

2.2.1 Correlation Functions

Unfortunately, these Lagrangians can rarely be solved directly under the principle of least action. Instead, physicists study vacuum expectation values of time-ordered products of field operators, known as correlation functions. Using a path integral approach, one can write down the time ordered n -point correlation function of ϕ as

$$\langle T\phi(x_1) \dots \phi(x_n) \rangle = \frac{\int \mathcal{D}\phi \phi(x_1) \dots \phi(x_n) e^{iS_\epsilon}}{\int \mathcal{D}\phi e^{iS_\epsilon}} \quad (2.1)$$

where S_ϵ is the action retrieved by integrating the Lagrangian over all of spacetime while analytically continuing time $t \rightarrow t - i\epsilon$ in order to force the exponential to converge.

$$S_\epsilon = \int_{\infty(1-i\epsilon)}^{\infty(1-i\epsilon)} dt \int d^{d-1}x \mathcal{L} \quad (2.2)$$

The path integral is evaluated by using perturbation theory to Taylor-expand the integral with respect to any coupling terms, resulting in integrals of the form

$$\frac{\int d^{d-1}x x_{i_1} \dots x_{i_n} e^{-\frac{1}{2}x^T A x}}{\int d^{d-1}x e^{-\frac{1}{2}x^T A x}} \quad (2.3)$$

where A is a symmetric matrix determined by the action S_ϵ . This can be solved using derivatives in the complex vector, B , or the moments of the following gaussian integral:

$$\int d^{d-1}x x e^{\frac{1}{2}x^T A x + B^T x} = \frac{1}{\sqrt{\det(\frac{A}{2\pi})}} e^{\frac{1}{2}B^T A^{-1} B} \quad (2.4)$$

The process of evaluating these integrals can be long and tedious. In the late 1940s, Richard Feynman introduced the Feynman Diagram as a graphical method for writing down these integrals along with Feynman Rules describing how to graphically evaluate correlation functions without needing to explicitly compute the path integral. While Feynman diagrams are often interpreted as collision diagrams, it is important to remember their one-to-one relationship with path integrals of correlation functions. Once the Feynman Rules have been derived from the path integral of a specific QFT, they can be applied to get correlation functions. An example of a Feynman diagram can be seen in Figure 2.1.

2.2.2 Scattering

While correlation functions are the foundational building block of theoretical QFT calculations, it is not immediately obvious what to do with their result. Physicists think of particles instead of fields (such as in the SM) and are often interested in how they scatter off of one another. These scattering properties allow physicists to compute - Cross sections: roughly the rate at which a certain particle is produced. - Branching fractions: the probability of a specific decay occurring, like is the subject of this thesis. which are the two most frequently measured quantities in colliders. So far, correlation functions have only given vacuum expectation values of field operators.

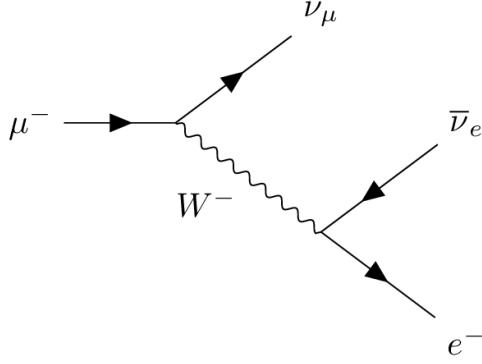


Figure 2.1: An example feynman diagram describing the $\mu \rightarrow \nu_\mu \nu_e e$ decay.

Formalizing particles in a theory built on fields results in considering particles as eigenstates of the Hamiltonian whose wave packets are non-interacting at either early or late times, called "in" and "out" states respectively. A QFT can thus be equipped with a scattering description only if the Hamiltonian has a complete set of "in" and "out" states. Luckily, the SM is one such QFT.

The scattering properties between these "in" and "out" states can be extracted from the S -matrix, or the amplitude to find the world in an "out" state β , given that it started in an "in" state α :

$$S_{\beta\alpha} = \langle \beta | \alpha \rangle \quad (2.5)$$

The method of constructing this S -matrix from correlation function is done using the LSZ Reduction Formula, which states that the existence of particles in a QFT leads to poles in the Fourier transform of its two-point functions (correlation functions between only two operators, also known as propagators). Therefore, computing the S -matrix amounts to computing correlation functions using Feynman Diagrams, taking the Fourier transform, taking all the external momentum to on-shell, and then computing the residue of the pole in the propagator.

This thesis concerns itself with the branching fraction of the $D^0 \rightarrow \mu\mu$ interaction. To compute the branching fraction from the scattering amplitude, it is first convenient to remove delta functions and constants always present in the scattering amplitude by defining the M -matrix as the matrix that satisfies

$$S_{\beta\alpha} = \delta(\beta - \alpha) + i \times (2\pi)^d \delta^d(p_\beta - p_\alpha) \mathcal{M}_{\beta\alpha} \quad (2.6)$$

where p_α, p_β are the 4-momentum of the states.

From this, one can calculate that the differential decay rate into a final state β is given by

$$d\Gamma(\alpha \rightarrow \beta) = (2\pi)^d \delta^d(p_\beta - p_\alpha) |\mathcal{M}_{\beta\alpha}|^2 d\beta \quad (2.7)$$

Specifically, to get a full decay width, one must integrate over all final states and their momentum that one cares about to get $\Gamma(\alpha \rightarrow \beta) = \int (2\pi)^d \delta^d(p_\beta - p_\alpha) |\mathcal{M}_{\beta\alpha}|^2 d\beta$.

This concludes the outline of how to arrive at a branching fraction for a general QFT. The next section covers the specifics of the Standard Model as a QFT and the last section outlines the computation of the $D^0 \rightarrow \mu\mu$ decay.

2.3 The Standard Model

Recall that a QFT is constructed by identifying a list of local symmetries and writing down the most general, renormalizable Lagrangian that satisfies the symmetries. The local symmetries of the SM are given by

$$SU(3) \times SU(2) \times U(1) \quad (2.8)$$

Like many QFTs, there are two classifications of fields in the SM: bosons and fermions. While bosons are symmetric under exchange, fermions are antisymmetric under exchange. Consequently, the spin-statistic theorem states that fermions must have half integer spin while bosons have integer spin. Additionally, the spin-statistics theorem dictates that no two fermions may occupy the same state, making it very natural to think of them as matter. The total spin of two interacting half integer spin fermions must be integer, making it natural to think of bosons as mediators of interactions between fermions.

The three local symmetries of the Standard Model are in fact gauge symmetries. They represent redundancy in the description of our model, similar to the classical electromagnetism gauge. Due to formalism beyond the scope of this thesis, the generators of these gauge groups induce gauge bosons. Specifically, the 8 generators of $SU(3)$ give rise to the eight gluons that describe Quantum Chromodynamics (QCD) and the other four generators of $SU(2) \times U(1)$ give rise to electroweak (EW) interactions. Due to this, it makes sense to break up the Lagrangian into QCD and EW terms, which we know can be written independently, due to the independence of their generators. Namely, we have that

$$\mathcal{L}_{SM} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} \quad (2.9)$$

In the following sections, we will analyze each of these Lagrangians independently, showing how they give rise to the various particles in the SM and their interactions. As a preview, the particles of the Standard Model and their quantum numbers are summarized in Figure 2.2.

2.3.1 Quantum Chromodynamics

The eight generators of $SU(3)$ are given by the Gell-Mann matrices and yield the 8 bosons of QCD, known as gluons. The three-dimensional basis of these matrices give rise to three colors, labeled r , g , and b , and three anti-colors, \bar{r} , \bar{g} , \bar{b} , which are the fundamental charges of QCD. The Gell-Mann matrices govern that each of the gluons can carry both a color and an anti-color charge. The corresponding fermions in this theory are named quarks and simply carry one color charge or one anti-color charge for anti-quarks. There are 6 different flavors of quarks, summarized in table 2.1.

By constructing a QFT that is invariant under $SU(3)$, one arrives at the Lagrangian

$$\mathcal{L}_{QCD} = \bar{q}^a \not{\partial}_{ab} q^b - m_{ab} \bar{q}^a q^b - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (2.10)$$

where q_i is the quark field, with i being the flavor index which is implicitly summed over all the quark flavors. $\not{\partial}$ is a contraction between the gauge covariant derivative and the gamma

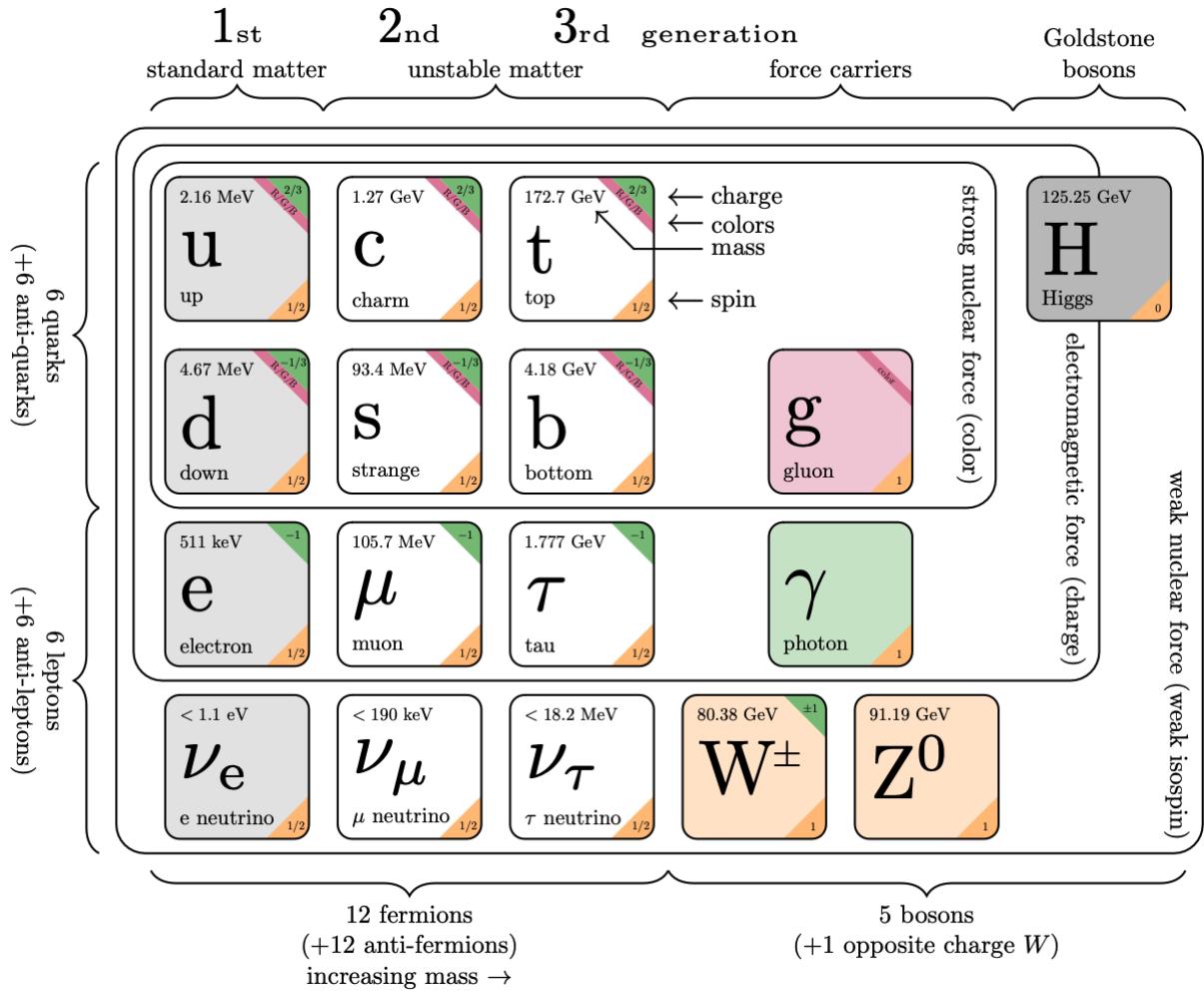


Figure 2.2: The particles of the Standard Model

matrices γ^μ which connect the spinor representation of the fields to the vector representation of the Lorentz group. Stated explicitly, we have

$$\not{D} = \gamma^\mu (D_\mu)_{ab} = \gamma^\mu (\partial_\mu \delta_{ab} - ig_s (G_\mu)_{ab}) \quad (2.11)$$

where G_μ is the gauge field and represented by the Glann-Mann matrices and g_s is the strong coupling constant found in renormalization. $G_{\mu\nu}$ is the associated field strength tensor, given by

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - ig_s [G_\mu, G_\nu] \quad (2.12)$$

Lastly, m_{ij} resembles the mass matrix of the various quark flavors and is only non-zero when $i = j$.

In general, perturbative calculations in QCD can be quite complex and difficult. Therefore, in practice, QCD is often place on a lattice, imposing a hard momentum cut-off and performing perturbations on the lattice.

Quark	Symbol	Charge Q/e	Mass (MeV/c^2)
Up	u	+2/3	$2.16^{+0.49}_{-0.26}$
Down	d	-1/3	$4.67^{+0.48}_{-0.17}$
Charm	c	+2/3	1270 ± 20
Strange	s	-1/3	93^{+11}_{-5}
Top	t	+2/3	$172\,760 \pm 300$
Bottom	b	-1/3	4180^{+30}_{-20}

Table 2.1: Electric charge and mass of the six quark flavours. Masses are quoted in the schemes recommended by the Particle Data Group [13].

2.3.2 Electroweak Interactions

The electromagnetic and weak nuclear forces are unified under EW interactions at energies above ≈ 250 GeV or temperatures above $\approx 10^{15}\text{°}K$, mediated by the $SU(2) \times U(1)$ symmetry. The generators of the $SU(2)$ group yield three gauge bosons of weak isospin, given by A_μ^i , and the generators of the $U(1)$ group yield three gauge bosons of weak hypercharge, given by B_μ . Importantly, $SU(2)$ is the chiral part of the electroweak symmetry, meaning it only affects left-handed fermions. Therefore, much of the discussion on electroweak interactions will differentiate between left-handed and right-handed particles.

Due to spontaneous symmetry breaking, the four vector bosons are mixed using the Weinberg angle, θ_W , to produce the 4 physical gauge bosons: γ , Z^0 , W^\pm . Note that the photon, γ is sometimes written as the electromagnetic field A_μ and is not to be confused the weak isospin gauge bosons, A_μ^i . Additionally, the Z_μ^0 is often just written as Z_μ , dropping the charge label. Specifically, the mixing under the Weinberg angle manifests as

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(A_\mu^1 \mp i A_\mu^2) \\ \begin{pmatrix} A_\mu \\ Z_\mu^0 \end{pmatrix} &= \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ A_\mu^3 \end{pmatrix} \end{aligned} \quad (2.13)$$

The left-handed fermions in this theory are given as doublets that transform under $SU(2) \times U(1)$ and can be categorized into leptons and quarks. Specifically, the doublet notation is used to manifestly ensure transformation under $SU(2)$. This is similar to a two-state quantum spin system with $|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ with spin operators that act as $SU(2)$ rotations on the spin vectors. The components of these doublets are the $U(1)$ abiding states, similar to those given by QED interactions.

Therefore, the left-handed leptons are given by

$$L_i = \begin{pmatrix} \nu_{iL} \\ l_{iL} \end{pmatrix} \quad (2.14)$$

where ν_{iL} are the three left-handed neutrinos and the l_{iL} particles are the three other left-handed leptons. Notationally, we use i to sum over all three flavors and L to denote the

left-handedness of the field. Left-handed quarks are given by

$$Q_i = \begin{pmatrix} u_{iL} \\ V_{ij} d_{jL} \end{pmatrix} \quad (2.15)$$

where u_{iL} are the three up quarks, d_{iL} are the three down quarks, and V_{ij} is the Cabibbo-Kobayashi-Maskawa (CKM) matrices that keep track of the Weinberg angle mixing coefficients.

The right-handed fermions in this theory are not affected by the $SU(2)$ symmetry; they only transform trivially due to the chirality assigned to the $SU(2)$ symmetry. Therefore, right-handed states are $SU(2)$ singlets that must only abide by $U(1)$ symmetry. We write these to be u_{iR} , d_{iR} , and e_{iR} . However, for notational ease, often the R is dropped since all the left fields are in doublets. Therefore, we write the quark fields as u_i , d_i , and e_i .

The Higgs field

Now, there is one more subtlety that must be addressed before introducing the Lagrangian for the electroweak interaction. Specifically, we know that in $d = 4$, renormalizable QFTs can only contain non-invariant dimension-four operators, meaning that it cannot allow a term that would give mass to the bosons, such as $m^2 B_\mu^a B^{a\mu}$ for the weak isospin gauge bosons. This is consistent with our formulation of QCD above due to the massless gluon, as well as consistent with our observation of the massless photon. However, the W^\pm and Z boson are observed to carry mass.

Therefore, we must introduce a scalar field that is allowed to break the $SU(2) \times U(1)$ symmetry and give rise to Z and W^\pm masses. This field is known as the Higgs field. While an in-depth discussion of the Higgs mechanism is out of the scope of this thesis, a short discussion should suffice to understand the EW Lagrangian.

In order to give mass to the W^\pm and Z bosons, the field we add must allow for the full Lagrangian to remain invariant while resulting in a ground state that is not. One good way to think of this is known as the Mexican hat potential, given by $f(r, \theta) = (r^2 - v^2)$ shown in Figure 2.3. This potential is completely rotationally symmetric, but the ground state lies away from the origin at $r = v$, meaning the ground state itself is not spherically symmetric. Additionally it satisfies all other properties of a valid potential by being continuous, well defined, and bounded from below. Therefore, we introduce a doublet under $SU(2)$, given by

$$h = \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} \quad (2.16)$$

and put it in the Mexican hat potential.

To see that indeed this gives the W and Z bosons mass, we write down the Lagrangian of this field in the presence of the W_μ^a and B_μ fields as

$$\mathcal{L}_h = (D_\mu h)(D^\mu h) - \lambda \left(|h|^2 - \frac{v}{2} \right)^2 \quad (2.17)$$

where $D_\mu = (\partial_\mu - \frac{ig}{2}\tau^a A_\mu^a - \frac{ig'Y}{2}B_\mu)$. Here τ^a (also known as the Pauli matrices) and YI are the generators of $SU(2)$ and $U(1)$. Their corresponding eigenvalues are labeled as T_a

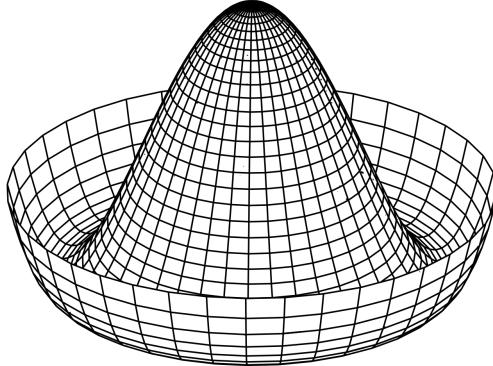


Figure 2.3: The Mexican hat potential, given by $f(r, \theta) = (r^2 - v^2)$, demonstrating ground state symmetry breaking in a symmetric potential

and Y ; they are known as the weak isospin and weak hypercharge quantum numbers and a convenient basis is often $T^2 = T_1^2 + T_2^2 + T_3^2$, T_3 , and Y . A small note that T_a is actually the eigenvalue of $\frac{\tau^a}{2}$, but can be discussed as an eigenvalue of τ^a . Separately, note that $\lambda \left(|h|^2 - \frac{v}{2} \right)^2$ is exactly a rescaled Mexican hat potential. Now, it is straightforward to see that this Lagrangian contains the terms

$$\frac{v^2}{8} [g^2 ((A_\mu^1)^2 + (A_\mu^2)^2) + (gA_\mu^3 - g'B_\mu)] \quad (2.18)$$

giving the intermediate weak isospin and weak hypercharge bosons mass. Using the definition of the Weinberg angle where $\tan(\theta_W) = g'/g$, the mass terms of the observed bosons become

$$M_{W^+} = M_{W^-} = \frac{gv}{2} \quad M_Z = \frac{gv}{2 \cos(\theta_W)} \quad M_\gamma = 0 \quad (2.19)$$

Full Electroweak Theory

Finally, by constructing a QFT involving these fields and one that is invariant under $SU(3)$, one arrives at the Lagrangian

$$\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_y \quad (2.20)$$

where \mathcal{L}_g is the gauge term, similar to the QCD gauge term, given by

$$\mathcal{L}_g = -\frac{1}{4} A_{\mu\nu}^i A^{i\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (2.21)$$

The electroweak field strength tensors $A_{\mu\nu}^i$ and $B_{\mu\nu}$ are defined similarly to how the gluon field tensor is defined for QCD.

\mathcal{L}_f is the fermion kinetic term, given by

$$\mathcal{L}_f = \bar{L}_i \not{D} L_i + \bar{e}_i \not{D} e_i + \bar{Q} \not{D} Q + \bar{u} \not{D} u + \bar{d} \not{D} d \quad (2.22)$$

where again \mathcal{D} is the covariant derivative in the above section, D_μ contracted with a gamma matrix, γ^μ and the fields are defined in the above section as left-handed doublets and right-handed singlets.

\mathcal{L}_h has already been defined in the previous section and \mathcal{L}_Y are the Yukawa coupling terms between the fermion fields and the Higgs scalar field. This is given by

$$\mathcal{L}_y = -Y_e \bar{L} h e + Y_u \bar{Q} h u + Y_d \bar{Q} h d + \text{h.c.} \quad (2.23)$$

Here, for notational convenience we have dropped the flavor indices. $Y_{e,u,d}$ are the three Yukawa coupling term matrices. Just as the Higgs field gives mass to the W^\pm and Z bosons, this mechanism provides mass to the fermions in electroweak theory.

2.4 The $D^0 \rightarrow \mu\mu$ decay

In principle, now that we have written down the SM Lagrangian, one could follow the standard QFT formalism described above to calculate perturbatively $\Gamma(D^0 \rightarrow \mu\mu)$. In practice, this is incredibly complex and involves decades of perturbative QFT formalisms. Therefore, in this section we will first categorize this decay as a FCNC and briefly summarize the heuristic structure of the calculation of $\mathcal{B}(D^0 \rightarrow \mu\mu)$ as performed by G. Burdman, E. Golowich, J. L. Hewett, and S. Pakvasa (BGHP) [3].

2.4.1 Flavor Changing Neutral Currents

Flavor changing neutral currents (FCNCs) are interactions that change the flavor of a fermion without altering its electric charge.

The SM forbids FCNCs explicitly at tree level. Processes that are flavor-changing with neutral currents must occur through loop processes that result in a neutral current, the details of which are discussed later. To illustrate this explicitly, we start with the gluon-fermion interaction terms of the Lagrangian which we derived earlier. Namely we have that the interaction terms can be pulled out of Equation 2.22 to give us

$$\mathcal{L}_{gf} = \sum_{f=\{Q,L,u,d,e\}} \bar{f}_i \gamma^\mu \left(\partial_\mu - \frac{ig}{2} \tau^a A_\mu^a - ig' \frac{Y}{2} B_\mu \right) f_i \quad (2.24)$$

We now rotate into the Weinberg basis, using $\tau^\pm = \frac{\tau^1 \pm i\tau^2}{\sqrt{2}}$ and W^\pm defined as in Equation 2.13. This gives us

$$\mathcal{L}_{gf} = \sum_{f=\{Q,L,u,d,e\}} \bar{f}_i \gamma^\mu \left(\partial_\mu - \frac{ig}{\sqrt{2}} (W_\mu^+ \tau^+ + W_\mu^- \tau^-) - \frac{g}{2} A_\mu^3 \tau^3 - \frac{g'}{2} B_\mu Y \right) f_i \quad (2.25)$$

From this, we can easily isolate \mathcal{L}_{gf} into neutral currents and charged currents, as τ^\pm is the charged operator. This fact is not explicitly proven here, but can be shown by writing out τ^\pm in the doublet basis and then noticing that τ^+ converts the lower component into the upper one while τ^- does the opposite. This causes the weak isospin eigenvalue to change

while keeping hypercharge constant, resulting a change in electric charge, $Q = I_3 + Y/2$. Therefore, we have that the neutral current interactions are given by

$$\mathcal{L}_{NC} = \sum_{f=\{Q,L,u,d,e\}} \bar{f}_i \gamma^\mu \left(-\frac{g}{2} A_\mu^3 \tau^3 - \frac{g'}{2} B_\mu Y \right) f_i \quad (2.26)$$

Now, we notice that the Lagrangian only has interactions between fermions with the same flavor index. Therefore, the only flavor mixing that can occur is between the doublets. Therefore, we can expand out to get the candidates for FCNC interactions as

$$\begin{aligned} \mathcal{L}_{FCNC \text{ Candidates}} = & -\frac{g}{2} (\bar{u}_{iL} \quad \bar{d}_{iL}) \gamma^\mu A_\mu^3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix} \\ & -\frac{g}{2} (\bar{\nu}_{iL} \quad \bar{l}_{iL}) \gamma^\mu A_\mu^3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \nu_{iL} \\ l_{iL} \end{pmatrix} \\ & -\frac{g'}{2} (\bar{u}_{iL} \quad \bar{d}_{iL}) \gamma^\mu B_\mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix} \\ & -\frac{g'}{2} (\bar{\nu}_{iL} \quad \bar{l}_{iL}) \gamma^\mu B_\mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_{iL} \\ l_{iL} \end{pmatrix} \end{aligned} \quad (2.27)$$

Now, we can consolidate this significantly by introducing $e = g \sin(\theta_W) = g' \cos(\theta_W)$ where the second equality can be verified by using $\tan(\theta_W) = \frac{g'}{g}$, $A_\mu^3 = \cos(\theta_W) Z_\mu + \sin(\theta_W) A_\mu$, and $B_\mu = -\sin(\theta_W) Z_\mu + \cos(\theta_W) A_\mu$. The second two equations can be derived from Equation 2.13. Additionally, we can use the weak isospin and weak hypercharge eigenvalues I_3 and Y , defined below Equation 2.17. It is also useful to define the electric charge eigenvalue, $Q = I_3 + Y/2$. Using this, we can greatly simplify Equation 2.27 to

$$\begin{aligned} \mathcal{L}_{FCNC \text{ Candidates}} = & - \sum_{f=u_L, d_L, l_L, \nu_L} \left[\frac{e}{\sin(\theta_W) \cos(\theta_W)} \bar{f}_i \gamma^\mu Z_\mu \left(T_3^{(f)} - Q^{(f)} \sin^2(\theta_W) \right) f_i \right] \\ & - \sum_{f=u_L, d_L, l_L, \nu_L} [e \bar{f}_i \gamma^\mu A_\mu Q^{(f)} f_i] \end{aligned} \quad (2.28)$$

where the values of $T_3^{(f)}$, $Y^{(f)}$ and $Q^{(f)}$ can be found in table 2.2.

Left-handed fermions				Right-handed fermions			
Field	T_3	Y	Q/e	Field	T_3	Y	Q/e
Up quark u_L	+1/2	+1/3	+2/3	Up quark u_R	0	+4/3	+2/3
Down quark d_L	-1/2	+1/3	-1/3	Down quark d_R	0	-2/3	-1/3
Neutrino ν_L	+1/2	-1	0	Charged lepton l_R	0	-2	-1
Charged lepton l_L	-1/2	-1	-1				

Table 2.2: Weak isospin (T_3), weak hypercharge (Y), and electric charge (Q) for the chiral fermions of one Standard Model generation [13].

More importantly, we confirm that no flavor-changing neutral current terms remain at tree level in the SM. More specifically, there are no terms that allow for interactions between fermions of two different flavors. Therefore, we have that $\mathcal{L}_{FCNC} = 0$ under the SM, preventing FCNC at tree level.

2.4.2 The $D^0 \rightarrow \mu\mu$ decay as a FCNC

Hadrons, such as the D^0 are collections of quarks in bound states. One defining property of QCD is that virtually all hadrons are in gluon-bound color-neutral states. These states therefore contain either a quark triplet (known as baryons) or a quark and an antiquark (known as mesons). D mesons are known as charm mesons and are any light meson that contains a charm quark. The D^0 is one such meson, being the only neutral D meson, composed of a charm and an anti-up quark. Note that there is also an anti- D^0 meson, being composed of an anti-charm and an up quark.

Now, since the D^0 meson is composed of a quark pair and the final state $\mu\mu$ contains no quarks, therefore implying a flavor change. However, importantly the D^0 is a neutral particle, as is the $\mu\mu$ final state. Therefore, for there to be a tree-level contribution, there must be a tree-level FCNC, which is not allowed by the Standard Model. Therefore, the $D^0 \rightarrow \mu\mu$ must proceed via higher-order loop processes, resulting in strong suppression in the SM.

G. Burdman, E. Golowich, J. L. Hewett, and S. Pakvasa (BGHP) have carried out the most complete SM analysis of the $D^0 \rightarrow \mu\mu$ decay that includes all short and long-distance effects.

To begin with the short distance (SD) effects, the quark level transition arises from electroweak penguin and box diagrams, an example of which is found in Feynman diagram in Figure 2.4. After integrating over all possible diagrams, BGHP find that

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{SD} \simeq 10^{-18} \quad (2.29)$$

Next, BGHP identify two distinct long distance (LD) mechanisms, single-particle unitary contributions and two-photon contributions. The single particle unitary contributions are dominated by weak-mixing with ground-state pseudoscalars, such as π^0, η, η' , followed by these pseudoscalars decaying into a dimuon final state, $P^0 \rightarrow \mu\mu$. The branching fraction of this calculation is given by

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{\pi^0, \eta, \eta'} \simeq 2.5 \times 10^{-18} \quad (2.30)$$

In principle, there could be mixing with heavier 0^\pm resonances such as $\pi(1800)$; however, the decays of $\pi(1800) \rightarrow \mu\mu$ occur with a branching fraction likely around 10^{-12} , leading to

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{\pi(1800)} \simeq 5.0 \times 10^{-17} \quad (2.31)$$

The two-photon contributions come from the dispersive two photon decay $D^0 \rightarrow \gamma\gamma \rightarrow \mu\mu$ shown in Figure 2.4. BGHP show that

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{\gamma\gamma} \simeq 2.7 \times 10^{-5} \mathcal{B}(D^0 \rightarrow \gamma\gamma) \quad (2.32)$$

where their own calculations show that, by considering short distance contributions, long distance vector meson dominant contributions, long distance single particle unitary contributions, and long distance two-particle unitary contributions, $\mathcal{B}(D^0 \rightarrow \gamma\gamma) \simeq 10^{-8}$. Therefore,

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{\gamma\gamma} \simeq 10^{-13} \quad (2.33)$$

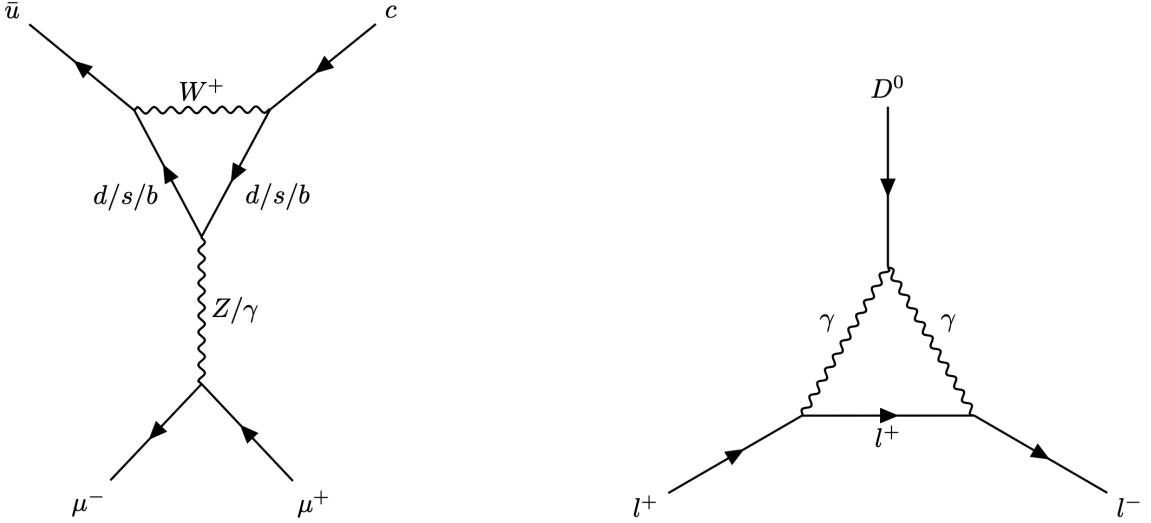


Figure 2.4: Two Feynman diagrams contributing to rare D^0 decays: (left) the leading SD contribution in the form of a penguin diagram and (right) the leading LD contribution, depicting the two-photon contribution. Note that the $D^0/\gamma/\gamma$ vertex hides additional complexity in the Feynman diagram, which is calculated by BGHP in the form of many SD and LD contributions combined and is abstracted away in the above diagrams.

When summing each of these contributions, one finds that $\mathcal{B}(D^0 \rightarrow \mu\mu)_{\gamma\gamma} \simeq 10^{-13}$ is by far the dominant contribution to the total branching fraction, therefore BGHP conclude that

$$\mathcal{B}(D^0 \rightarrow \mu\mu)_{\text{total}} \simeq 10^{-13} \quad (2.34)$$

It is also worth noting that BGHP show that various new physics models significantly increase this branching fraction. However, their work on this is far out of the scope of this thesis and outlines many different new physics models, including many commonly accepted candidates for Beyond the Standard Model (BSM) physics such as Flavor Non-Universal Z' Bosons and Super Symmetry (SUSY). One such model is Flavor Non-Universal Z' Bosons, which introduce neutral Z' bosons with flavor off diagonal couplings

Due to the complexity and number of these models, we simply refer to their work to give meaningful calculations in support of new physics should a branching fraction significantly larger than 10^{-13} be detected. It should be clear from these calculations that putting better limits on $\mathcal{B}(D^0 \rightarrow \mu\mu)$ acts as a meaningful probe into many new physics models, serving as the primary motivation of this work.

Chapter 3

The CMS Detector

This chapter describes the experimental framework underlying the measurement of the $D^0 \rightarrow \mu^+ \mu^-$ branching fraction. It begins with a summary of the Large Hadron Collider at CERN, covering the injector chain, magnet systems, running periods, and luminosity profile.

Then, this chapter describes the CMS detector beginning with a short overview of its geometry and coordinate system before diving into each of its subsystems, including the silicon tracker, electromagnetic calorimeter, hadronic calorimeter, superconducting solenoid, and muon chambers. For each of the subsystems, this chapter describes both the basic engineering needed for these detectors and the effect on the physics read-out from the detector.

Lastly, this chapter describes how the CMS experiment uses a trigger system to balance the large volume of data with the need to capture sufficient numbers of rare particle interactions.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) is one of the largest experiments in the world. It has a circumference of 26.7 km, spanning the border of Switzerland and France near Geneva, Switzerland. As the name might suggest, the LHC is a two-ring superconducting hadron accelerator focusing on proton-proton collisions with occasional heavy-ion collision runs. For this thesis, we will entirely focus on proton-proton collisions.

The protons originate from hydrogen atoms, from which electrons are stripped using a strong electric field to isolate the protons. From there, a series of accelerators called the CERN LHC injector chain accelerates the protons to 450 GeV before injecting them into the LHC in bunches. At any given time, there are 39 batches of 72 proton bunches in the LHC with a 25 ns bunch spacing to accommodate the LHC kicker's rise time.

Once injected, the proton bunches are accelerated to a center-of-mass energy of several TeV using two main types of magnets: (i) 1,232 dipole magnets with a field strength of 8.3 T, which bend and accelerate the protons along the circular beam pipes, and (ii) 492 superconducting quadrupole magnets, operated at 1.9 K, which focus the proton bunches to maintain beam stability.

Once the particles are accelerated to their target energy, they collide at four main inter-

action points within the LHC. These correspond to the four major LHC experiments: two general-purpose detectors capable of a broad range of physics analysis: ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid); one detector optimized for heavy-ion collisions: ALICE (A Large Ion Collider Experiment); and one detector focused on b -physics: LHCb. In addition, five smaller experiments are located at secondary sites in the LHC: two specialized in forward-scattered particles, one experiment near LHCb to look for magnetic monopoles, and two experiments near ATLAS specialized in light particles, such as neutrinos. The full CERN accelerator complex can be seen in Figure 3.1

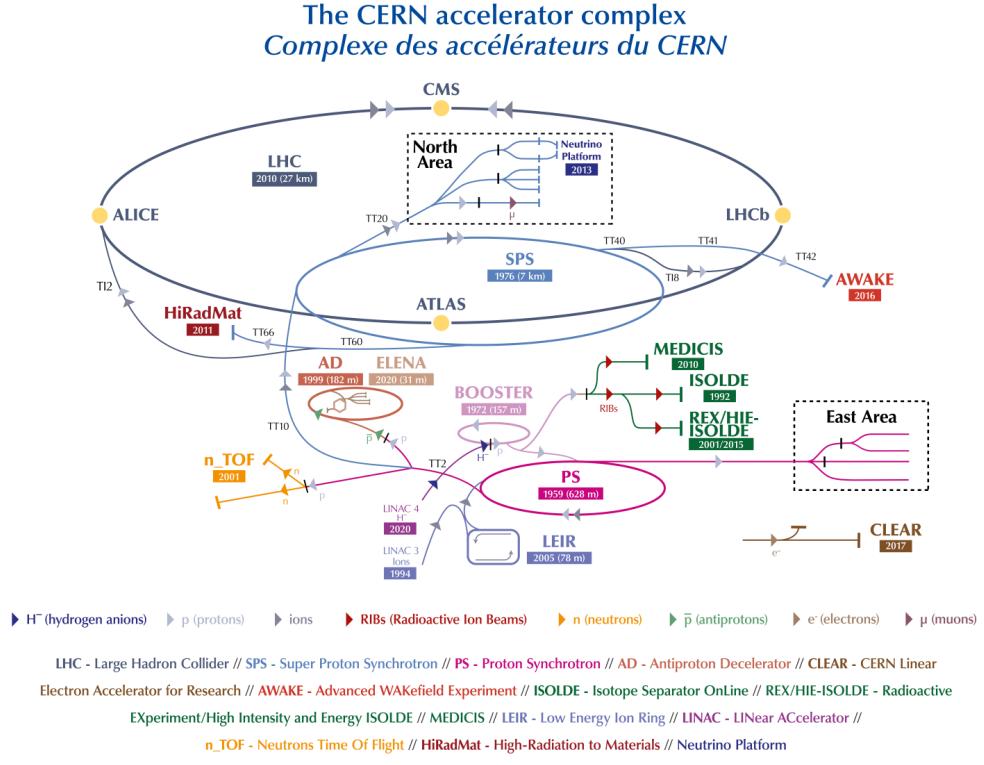


Figure 3.1: The CERN accelerator complex [12]

One of the main difficulties of circular colliders is the rapid energy loss of every through synchrotron radiation, the electromagnetic radiation emitted when relativistic charged particles move along a curved path. The total power radiated by synchrotron radiation in circular colliders is given by

$$P = \frac{q^2 p^4}{6\pi\epsilon_0 m^4 c^5 r^2} \quad (3.1)$$

where q is the charge of the particles, p is the momentum of the particle, m is the rest mass of the particle, and r is the radius of the collider. This provides the LHC with two key advantages in minimizing synchrotron radiation losses while achieving high beam energies: (i) compared to electron-positron colliders, such as the LHC's predecessor the Large Electron-Positron collider, the mass of protons is much larger and (ii) the radius of the LHC is larger than any other collider in the world.

Another difficulty of high energy colliders is the rarity of physically interesting collisions. Not only is it difficult to have two protons collide, but it is also rare to have a collision that will get included in an analysis. Due to these factors, the LHC operates at a very high beam intensity, measured by its luminosity. The luminosity of an accelerator is given by

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r F}{4\pi \epsilon_n \beta^*} \quad (3.2)$$

where N_b is the number of protons per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the Lorentz factor, F a geometric factor, and lastly ϵ_n and β^* determine the interaction rate at the interaction point. The high luminosity at the LHC is achieved largely with the $\approx 10^{11}$ protons per bunch. However, this high intensity introduces numerous challenges due to the high number of background collisions, which we will discuss further in Section 3.3.

Instead of running continuously, the LHC breaks up its data collecting into runs, each at a different center-of-mass energy and luminosity. The first run from 2010-2012 had a 8 TeV center-of-mass energy with a total luminosity of 29.45 fb^{-1} . Run 2 occurred from 2015-2018 and had a 13 TeV center-of-mass energy with a total luminosity of 163.6 fb^{-1} . Run 3 started in 2022 and is currently ongoing with a center-of-mass energy at 13.6 TeV and a targeted luminosity of 42 fb^{-1} .

3.2 The CMS Detector

The CMS detector is a large cylindrical detector that almost entirely wraps around the proton beam. It is 21 meters long and 15 meters in diameter, weighing 14,000 tonnes. It is composed of 5 main subsystems wrapped in layers around the beam pipe. These five layers are the silicon tracker, the electromagnetic calorimeter, the hadron calorimeter, and the superconducting solenoid, and the muon chambers. These layers are wrapped around the beam line through the center of the detector and can be seen in Figure 3.2. More discussion on the subsystems of CMS will proceed in following sections. However, one must first understand the coordinate system.

The coordinates used in CMS are ϕ , η , and r with the origin located at the interaction vertex at the center of the cylinder where r represents the distance from the point of interest to the origin. ϕ describes the azimuthal angle along the plane normal to the beam line and is often ignored due to the complete symmetry in ϕ . Lastly, η is known as the pseudorapidity, describing the angle of a particle relative to the beam line, θ , as

$$\eta \equiv -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (3.3)$$

For high energy particles, this is a good approximation of rapidity, given by

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (3.4)$$

where p_z is the momentum along the beam line. This is used instead of the conventional angle θ because rapidity is invariant under Lorentz boosts while the standard angle θ is not.

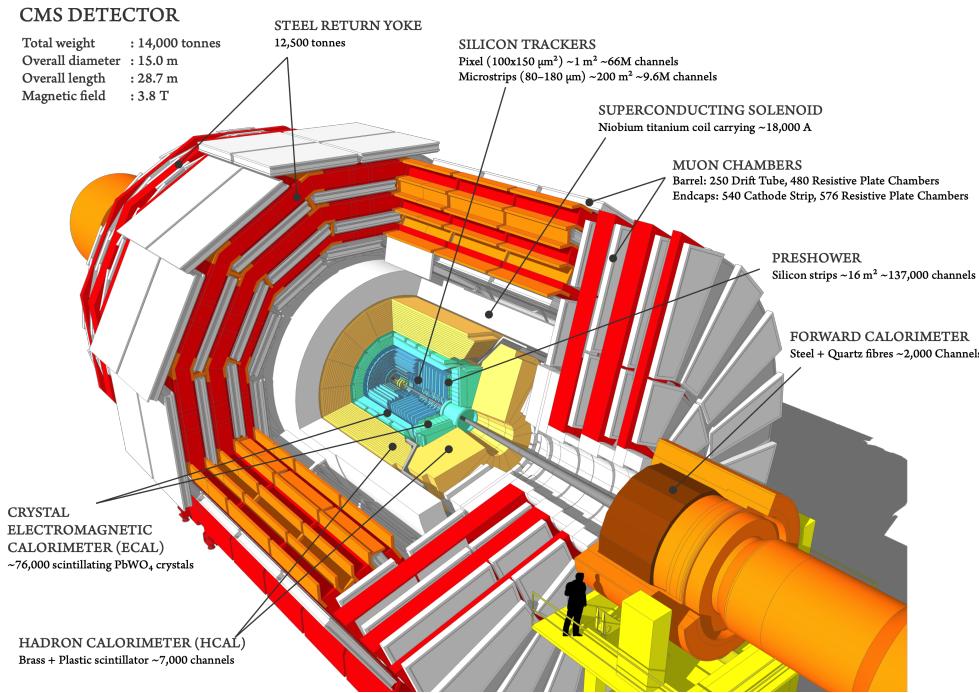


Figure 3.2: A labeled cutaway diagram of the CMS detector [15]

For example, if two particles have rapidities y_1 and y_2 and the difference $y_1 - y_2$ is invariant in any reference frame. The CMS detector has two main regions: the barrel which wraps around the beam pipe and the endcaps which cover the ends of the cylinder. The exact η value separating the barrel and endcap regions varies by subsystem at CMS, but needs to be carefully considered due to biases introduced because of detector geometry as you move from the barrel to the endcap.

One other quantity frequently used in describing particle kinematics is p_T , or the momentum of the particle transverse to the beam line given by

$$p_T = |\vec{p}| \operatorname{sech}(\eta) \quad (3.5)$$

The full set of coordinates and the CMS geometry can be seen in Figure 3.3.

3.2.1 Solenoid Magnet

One of the easiest ways to measure the electromagnetic charge of high energy particles is to introduce them into a strong magnetic field and measure the curvature of their tracks. Additionally, the curvature is also dependent on the momentum of the incoming particle, allowing the detector to reconstruct the momentum of charged particles from their tracks. Therefore, the presence of a strong magnetic field across all other subsystems is essential.

The CMS detector contains a 3.8 T solenoid magnet in the middle of the other subsystems. The solenoid is approximately 6 meters in diameter and 13 meters in length, making it one of the largest of its kind ever constructed. Its immense field strength enables CMS to accurately reconstruct the trajectories and momenta of particles produced in collisions, even at very high energies.

3.2.2 Silicon Tracker

Before the tracks of particles can be disturbed by interactions with other detector components, the silicon tracker reconstructs a precise track that records the particle's trajectory. This is most commonly used for locating their primary vertices, the exact location of their production, as well as assisting in kinematic reconstruction. In order to do this, silicon is configured in a reverse-biased p-n junction. As charged particle pass through the silicon, they ionize silicon atoms, creating electron-hole pairs. This electric field pushes electrons toward the n-type electrode and hole toward the p-type electrode, causing a current pulse. This current pulse is then read out by Application-Specific Integrated Circuits (ASICs). Because this detection relies on ionization, neutral particles do not produce signals in the tracker.

The detector is made of two layers of silicon detectors: small pixels located near the center of the tracker and larger strips located near the outside of the detector. The pixels have a size of $100 \times 150\mu m$, allowing for them to have an extremely fine resolution. Further out, the large strips arranged in a vertical/horizontal pattern reconstruct the track of charged particles by superimposing the vertical and horizontal signals. This allows for more precise modeling of the particle's position, but at the cost of resolution. Strips closer to the center of the detector are smaller while strips further outside are larger.

3.2.3 Electromagnetic Calorimeter

A calorimeter is any device that measures the heat exchanged between a process and its environment. In particle physics detectors, the term takes on a specialized meaning: it denotes a system that destructively measures a particle's energy by scattering it in dense material and recording the ensuing shower. Such devices therefore contain two essential components:

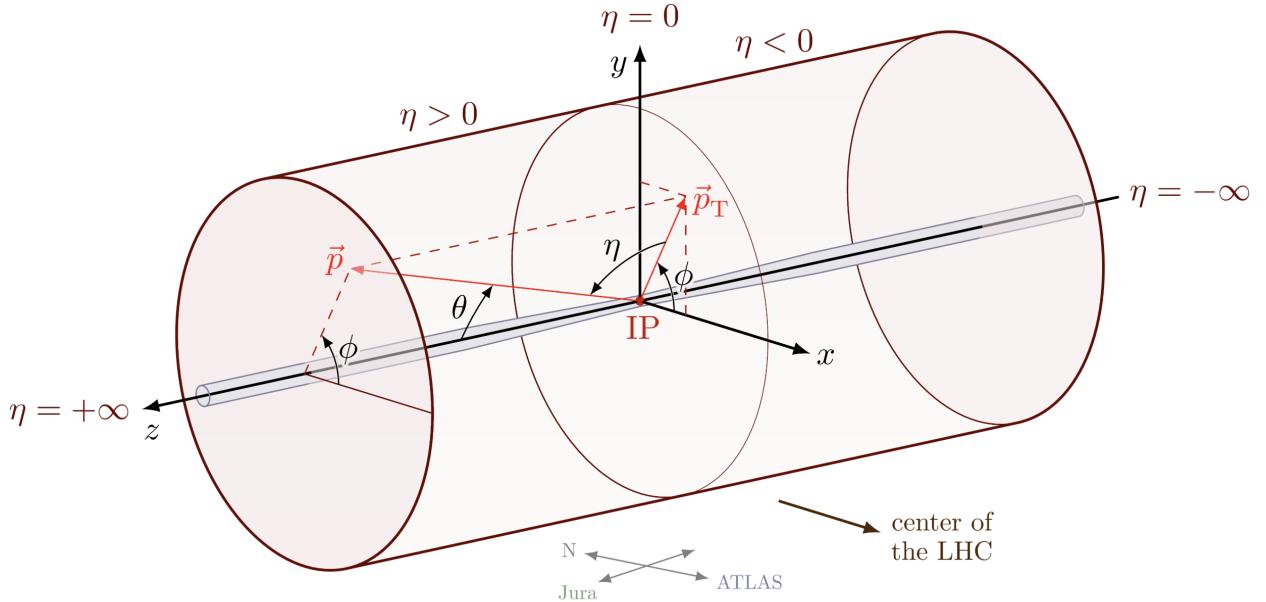


Figure 3.3: The CERN accelerator complex

(i) a passive absorber, which initiates the shower, and (ii) a scintillating medium, which converts the energy released by the shower into optical photons that can be read out.

Perhaps the most common example of a calorimeters is the electromagnetic calorimeter (ECAL), whose purpose is to stop and measure high-energy electrons and photons. Incoming γ/e^\pm particles lose energy almost exclusively through well-understood QED processes such as bremsstrahlung and pair production, causing a cascade to lower energy photons and electrons, causing an electromagnetic shower such as in Figure 3.4. Eventually the cascaded photons and electrons have a low enough energy that they are able to be easily absorbed and measured by the scintillator. The characteristic scale of this shower is the radiation length X_0 , the mean distance over which an electron's energy is reduced by a factor e^{-1} . In practice an ECAL is built to a depth of $\sim 20\text{-}30 X_0$ so that the entire shower is contained. The other important scale is the Molière radius, or the radius of the cylinder containing 90% of the shower's energy.

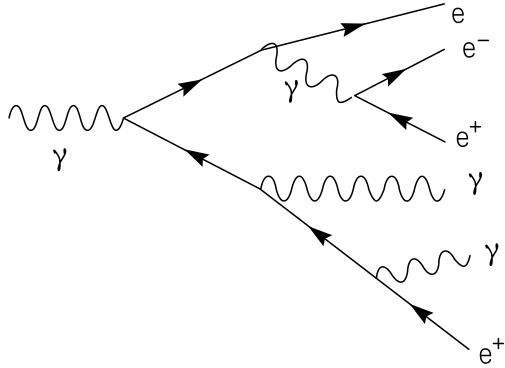


Figure 3.4: A schematic example of an electromagnetic shower caused by bremmstrahlung and pair production.

In CMS the ECAL is a homogeneous calorimeter, meaning the absorber and the scintillator are the same material (lead tungstate, PbWO_4) so that virtually no energy is lost to passive absorbers since they are the same as the scintillator. Additionally, homogeneous calorimeters allow for the same response from every angle due to their homogeneity. The barrel section ($|\eta| < 1.48$) and two end-caps ($1.48 < |\eta| < 3.0$) together comprise 75,848 crystals, each $22 \times 22 \text{ mm}^2$ at the front face and 230 mm long ($25.8 X_0$). The fast scintillation time of PbWO_4 ensures that $\sim 80\%$ of the light is collected within a single LHC bunch crossing, and the small Molière radius ($R_M \simeq 21 \text{ mm}$) localizes showers largely inside one crystal.

Radiation damage gradually darkens the crystals. Therefore, CMS injects laser light between fills to monitor transparency and applies time-dependent calibration constants. This careful calibration allows for precise reconstruction of an electromagnetic particle's energy.

3.2.4 Hadronic Calorimeter

While the ECAL measures purely electromagnetic showers, the hadronic calorimeter (HCAL) captures and quantifies the energy of neutral and charged hadrons. The underlying physics is

much more complex. When an incoming hadron interacts strongly with the passive absorber of the HCAL, it produces a hadron shower. The hadrons in the shower then produce electrons and photons in three main ways: (i) through neutral pions which decay to charged kaons or protons, which interact to produce two photons through a three vertex Feynman diagram, (ii) through charged hadrons ionizing the passive absorber, and (iii) through evaporation and spallation of the nuclei of the passive absorber. Then, these photons and electrons produce showers which are measured, similar to the ECAL. While neutral pion decay accounts for most of the energy converted to electromagnetic showers, the specific fraction of hadron energy that is converted to photon/electron energy is highly dependent on the incoming hadron's momentum and energy. This leads to a lower energy resolution compared to ECALs since it is difficult to estimate how much energy is not converted to photons and electrons. Additionally, because of the addition of hadron showers in HCALs, the radiation length of the HCAL is longer, leading to a larger detector compared to ECALs.

CMS uses a sampling HCAL. Alternating layers of passive brass absorber and plastic scintillator tiles form a barrel ($|\eta| < 1.3$), two end-caps ($1.3 < |\eta| < 3.0$), and a forward calorimeter extending to $|\eta| < 5.2$. Radiation and aging also affect HCAL response, similar to ECALs. CMS calibrates the system continuously using embedded radioactive sources, dedicated laser/LED pulses, and ϕ -symmetry methods.

3.2.5 Muon Chambers

Lepton detection is critical to many analyses performed at CMS, including this one, which focuses on muon and electron final states. Neutrinos are too light to be detected by anything in the detector and τ leptons are so unstable they always decay into other products which we measure. Electrons are well measured in the ECAL, but muons are heavy enough that they almost always don't get absorbed by the ECAL and therefore are not detected. Therefore, there is a need to construct large muon chambers, specifically built to detect muons.

The muon chambers work by measuring the muon's track as it leaves the detector. The magnetic field is still strong enough that the momentum and sign of the charge of the muon can directly be reconstructed from the curvature of the track. In this way, the muon chambers can be thought of as very large trackers.

There are three different types of gaseous detectors used in the muon system: drift tubes, cathode strip chambers, and resistive plate chambers. The drift tubes are 4cm wide tubes containing a positively charged stretched wire and a gas volume. When muons interact with the gas, they release electrons which drift toward the positively charged wire, causing a measurable pulse on the wire. Due to the length of the wires, drift tubes work very well in environments with uniform magnetic field and a low muon rate and are therefore found at the barrel. Cathode strip chambers contain an array of positively charged wires arranged in a grid. They work similarly to drift tubes, but the array allows for better measurements in the high muon rate and magnetic field variation environment of the endcap. Lastly, resistive plate chambers are gaseous parallel-plate detector consisting of one positively charged and one negatively charged plate. While they operate similarly to drift tubes and cathode strip chambers, they have detecting strips instead of wires to measure the electrical signal of released electrons. This produces a much more precise time resolution but a fairly poor spatial resolution. Therefore, these detectors are present throughout the barrel and the endcap,

complementing the drift tube and cathode strip detectors to improve the timing resolution of the muon chambers.

3.3 Triggers

As mentioned earlier, high energy colliders are faced with the problem that collisions that are interesting for physics analysis are rare. This results in a need for a very high rate of data production and a large amount of background. At the LHC, bunch crossing occur every 25 ns and result in an average of 20 collisions. Processing and storing the data from all these collisions is not technologically possible. Therefore, a trigger system is built to "trigger" on interesting physics events while ignoring all others. This trigger reduces the event rate from 800 MHz to roughly a few kHz. The quality of data output requires a computing farm running advanced algorithms, known as the High Level Trigger (HLT). However, the raw data rate is too large to give to any computer and must therefore be filtered by hardware, forming the Level-1 (L1) trigger. The L1 trigger reduces the event rate to 100kHz, where it can be processed by the HLT.

The L1 trigger is entirely built on Field Programmable Gate Arrays (FPGAs), custom integrated circuits that can be configured even after manufacturing, allowing for CMS to tune its hardware triggers without replacing hardware. These L1 triggers are segregated in terms of detector sub-systems, meaning that they only trigger once an interesting ECAL energy deposit has been found, an interesting muon detection, etc. but not on the event as a whole.

In contrast to the L1 trigger, the HLT is a large computing center with over 30,000 cores. One of the most important features of the HLT are the large arrays of buffers that can temporarily hold information during or prior to HLT processing, ensuring that the HLT has enough time to consider the event as a whole. The HLT system can be abstracted as a series of paths that each specialize in a specific type of event. For example, one might be interested in dimuon events with high muon energy. Carefully selecting the trigger paths is one of the most important aspects of any analysis.

One of the largest difficulties that come with using triggers is avoiding the bias that comes with picking a specific HLT trigger path. Numerous methods are used to prevent large trigger bias in analysis, including normalization channels and measuring trigger efficiency across whatever variable you are interested in. To allow these calculations to occur, a MinBias dataset is created, which is a random small sample of the raw data read out. A discussion on the triggers used for this analysis and their bias can be found in Section [4.2.1](#).

Chapter 4

$D^0 \rightarrow \mu\mu$ Branching Fraction Measurement

4.1 Analysis overview

The goal of this analysis is to measure the branching fraction of the $D^0 \rightarrow \mu\mu$ decay, written as $\mathcal{B}(D^0 \rightarrow \mu\mu)$. In performing this calculation, there are three major challenges that this analysis is faced with.

The first, and most fundamental, of these challenges is a small signal presence compared to a large combinatorial background, due to rarity of the $D^0 \rightarrow \mu\mu$ decay. In order to address this, we use D^0 mesons that originated from $D^{*\pm} \rightarrow D^0\pi^\pm$ decays, allowing us to use the additional soft pion to better reconstruct and select for signal events. By using $D^{*\pm}$ decay, we decrease the number of available $D^0 \rightarrow \mu\mu$ signal events, but also reduce the combinatorial background by a much larger factor, improving the overall quality of the analysis.

The second of these challenges is that the $D^{*\pm}$ cross-section is not well known at CMS, making it difficult to compute the number of D^0 mesons in our dataset directly. Therefore, instead we use a normalization channel approach, where we count the number of events in a normalization channel with a well known branching fraction, which we pick to be $D^0 \rightarrow \pi^\pm\pi^\pm$ due to the kinematic similarity to the $D^0 \rightarrow \mu\mu$ decay. Therefore, the result for this analysis is obtained by extracting out $\mathcal{B}(D^0 \rightarrow \mu\mu)$ from the following formula.

$$\frac{N_{D^0 \rightarrow \mu\mu}}{N_{D^0 \rightarrow \pi^\pm\pi^\pm}} = \frac{\mathcal{B}(D^0 \rightarrow \mu\mu)}{\mathcal{B}(D^0 \rightarrow \pi^\pm\pi^\pm)} \times \frac{\epsilon_{D^{*\pm}, D^0 \rightarrow \mu\mu}}{\epsilon_{D^{*\pm}, D^0 \rightarrow \pi\pi}} \times S_{ZB} \times \text{MVA}_D \times T_{\text{corr}} \quad (4.1)$$

where $\mathcal{B}(D^0 \rightarrow \pi^\pm\pi^\pm)$ is well known [13], $N_{D^0 \rightarrow \mu\mu}$ and $N_{D^0 \rightarrow \pi^\pm\pi^\pm}$ are calculated using fits to data as described in Section 4.4, ϵ_X is the efficiency of our selection calculated from Monte Carlo described in Section 4.3, S_{ZB} is a scale factor of the data discussed in Section 4.2, and $\text{MVA}_D, T_{\text{corr}}$ are corrections described in Section 4.6. This solution also has a nice property of reducing many of the errors that exist in both channels, reducing overall systematic error and providing a simpler analysis strategy.

The third of these challenges is caused by in-flight $\pi \rightarrow \mu\nu$ decays, which cause pions to be misreconstructed or "faked" as muons. Due to the nature of this experiment in measuring pion and muon final states, the effect of the fake rate impacts elements of the entire analysis.

To address the fake muons, this analysis performs an extensive analysis of the muon fake rate in Section 4.5.

Lastly, to prevent analysis design from biasing the result through effects such as overfitting to statistical fluctuations or selection bias, the entirety of the analysis is designed while being blind to the signal in data. Due to the small expected signal yield, the analysis is able to consider an eighth of the full dataset during its design while still being considered blind. Only once the entire analysis has been laid out, the full dataset is considered, which is discussed in the final Section of this analysis, Section 4.7.3.

4.2 Triggers and datasets

4.2.1 Data samples

The events used for this analysis are collected from proton-proton collisions at the LHC at a center of mass of 13.6 TeV. Specifically, we use data from the CMS detector during the years 2022 and 2023. The CMS collaboration marks specific run ranges as good runs and groups them into datasets marked with letters. The data we use is denoted as 2022C, 2022D, 2022E, 2022F, 2022G, 2023C, and 2023D, with letters used to mark specific eras, or ranges of runs where the detector conditions were relatively unchanged.

There are two triggers used for the analysis:

1. The `HLT_DoubleMu4_3_LowMass` trigger is used to collect $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \mu\mu$ signal events by triggering on the dimuon final product. During the collection of the data samples of the analysis, the trigger was virtually unchanged and unprescaled, making it convenient to use for signal collection.
2. The `HLT_ZeroBias` trigger was used to collect $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm$ normalization events. Unlike the signal trigger, this trigger does not filter for a specific signature. However, the branching fraction of $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm$ is large enough to generate a sufficient number of desired events. During the collection of the data samples of the analysis, the trigger was virtually unchanged.

Due to the fact that the `HLT_ZeroBias` trigger does not filter for a specific signature, it is exposed to much more data than any other dataset. The trigger selects data on every bunch crossing that occurs at CMS. As discussed in Section 3.3, this is far too much data to be processed by any modern day computer. Therefore, a *prescaling* is used to keep only a fraction of the events. This prescaling factor is incredibly important; a prescale factor of 100 means that there exists 100 times more events that satisfy the `HLT_ZeroBias` requirements than reported in the dataset. The prescaling factor is set to be roughly 10^6 and then measured experimentally by comparing the integrated luminosity of the data.

Specifically, for this experiment we calculate the ZeroBias scale factor, S_{ZB} as

$$S_{ZB} = \frac{\text{total luminosity}}{\text{ZeroBias luminosity}}. \quad (4.2)$$

Table 4.1 shows the measured integrated luminosity for each era. Using this, we can calculate that the total scale factor is $S_{ZB} = 1.255 \times 10^6$.

	Integrated luminosity	
	HLT_DoubleMu4_3_LowMass, fb^{-1}	HLT_ZeroBias, nb^{-1}
Run2022C	6.155	6.476
Run2022D	3.283	1.395
Run2022E	5.939	4.756
Run2022F	18.124	13.482
Run2022G	3.084	2.256
Run2022 Total	36.585	28.365
Run2023C	18.300	15.08
Run2023D	9.640	7.970
Run2023 Total	27.940	23.050
Run2022 + Run2023 Total	64.525	51.415

Table 4.1: Integrated luminosity of data collected with the primary triggers.

The events collected by these two triggers and grouped into two datasets, ZeroBias and Parking DoubleMuonLowMass, respectively. We use the NanoAOD file format, designed by CMS to contain ntuples of per-event information and used for most analysis at CMS. Specifically, we use the NanoAODv12 recipe which is processed from MiniAOD. In this processing, we apply the muon data certification to ensure high quality muon objects.

4.2.2 Backgrounds

A key component of the analysis is the accurate modeling of background, specifically because of how small the anticipated limit of the signal strength is. There are two main types of backgrounds to consider:

1. Combinatorial backgrounds: backgrounds which are characterized by noise from random combinations of particles produced in secondary particle collisions.
2. Peaking backgrounds: backgrounds which are characterized by real particle decays that are either fully or partially reconstructed.

Combinatorial Backgrounds

Combinatorial backgrounds are signals that originate from random combinations of particles that do not come from a common physical process. The existence of combinatorial backgrounds is present in virtually every experiment performed at the LHC due to the large amounts of events generated in each proton-proton bunch collision. Combinatorial backgrounds are often characterized by smooth, non-peaking shapes due to the lack of a unifying physical process to generate a peak. There are many shapes that are used to fit to combinatorial backgrounds, with the most common being exponential functions or polynomial functions. The functions allow for the needed degrees of freedom and overall shape to converge on the smooth combinatorial shape. In other applications, more specific combinatorial functions are derived to properly model the background.

Peaking Backgrounds

Unlike combinatorial backgrounds, peaking backgrounds must be carefully studied to ensure that the background peak in the signal region is not modeled as signal. These peaking backgrounds occur most commonly from detector misreconstruction. For example, pions can be misreconstructed as muons, resulting in some $D^0 \rightarrow \pi^\pm \pi^\pm$ events being reconstructed as $D^0 \rightarrow \mu\mu$ events, causing a peaking background near the signal peak. Specifically, we know that pions and kaons are the only 2 particles which are (1) common decay products of the D^0 and (2) are relatively frequently misreconstructed as muons. See Section 4.5 on Muon Fake Rates. Therefore, the candidates for peaking background are two body hadronic decays of the D^0 meson and non- $D^{*\pm} \rightarrow D^0 \pi^\pm$ decays.

For each of the peaking backgrounds, we check their Δm and dimuon mass against the Δm and dimuon mass of the signal. While the non- $D^{*\pm} \rightarrow D^0 \pi^\pm$ decays exhibit a dimuon mass distribution which is virtually identical to the signal, their Δm contribution is expected to be combinatorial so they do not overlap with the signal's signature. Therefore, the dominate two body hadronic decays of the D^0 meson are the only contributions to peaking background and are:

1. $D^0 \rightarrow \pi^\pm \pi^\pm$,
2. $D^0 \rightarrow K^\pm \pi^\mp$,
3. $D^0 \rightarrow K^\pm K^\pm$,
4. $D^0 \rightarrow \pi^\pm \mu\nu_\mu$ ¹.

For each of these, we use TGenPhaseSpace to simulate the effects of the misreconstructed decays on the peaking background distributions in both Δm and the dimuon mass and compare them to the signal distribution. As is drawn in Figure 4.1, every decay except $D^0 \rightarrow \pi^\pm \pi^\pm$ gives a distribution which is distinct from the signal distribution in either Δm or the dimuon mass. Therefore, the only peaking distribution which we must model and account for is $D^0 \rightarrow \pi^\pm \pi^\pm$.

4.2.3 Monte Carlo samples

The main component of constructing the Monte Carlo (MC) simulation samples is properly modeling the $D^{*\pm}$ meson by replicating their production in the CMS detector. The $D^{*\pm}$ has two primary production mechanisms at the LHC:

1. production through hadronization of charm quark directly from the proton-proton collision, creating prompt $D^{*\pm}$ mesons close to the primary vertex;
2. production through decay of other B hadrons, creating displaced $D^{*\pm}$ mesons from a secondary vertex.

¹Due to the semileptonic nature of this decay, it only has a peaking structure in $m(D^0)$, only in Δm

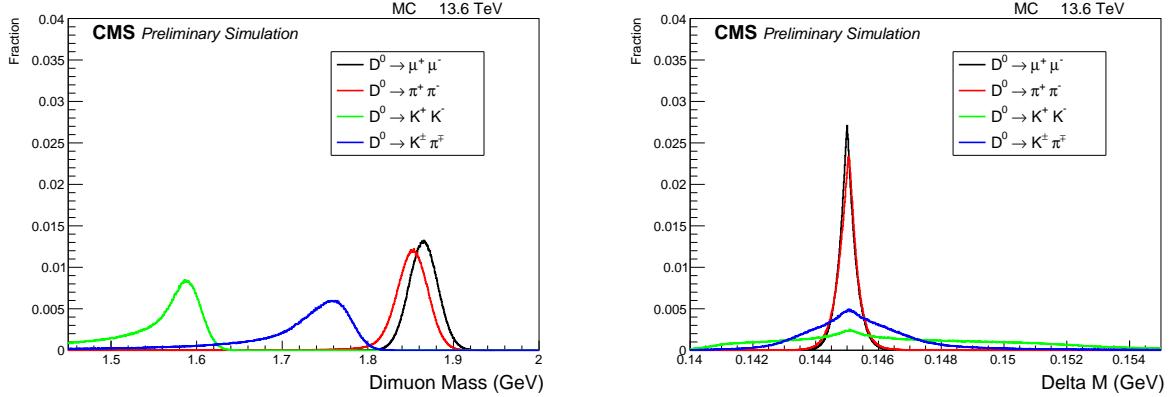


Figure 4.1: Comparison of normalized frequencies for various peaking background $D^0 \rightarrow \mu\mu$ decays against the signal decay. Right: Reconstructed Δm distribution. Left: Reconstructed dimuon mass distribution, where decay products are forced to take the muon mass using TGenPhaseSpace, even for non-dimuon final states.

It is important to model both of these production mechanisms of the $D^{*\pm}$ decay to generate a reliable MC simulation sample, therefore both are included in the MC sample. Specifically, the first method produces a prompt soft pion in the $D^{*\pm} \rightarrow D^0 \pi^\pm$ process while the second method produces a displaced soft pion, resulting in significant differences in vertex parameters used in reconstruction and classification. The full complexity of the MC samples extends beyond the scope of this thesis, but Table 4.2 displays some of the most used samples.

4.3 Selections and efficiency

Once we have attained the data and MC samples needed for this analysis, we filter through them using event selections, carefully tracking the *acceptance* and *efficiency* for each selection. We define *acceptance* as the purely kinematic fraction of signal events that fall inside the geometric phase space of the analysis and *efficiency* as the analysis' performance in identifying signal events that are inside the geometric region. The selection process is broken down into three main stages:

1. The *preselection* stage provides the selection criteria which are tied to trigger requirements, reconstruction requirements, and dataset size limitations due to prescaling.
2. The *baseline selection* stage is used primarily to reject background events while keeping the efficiency high, the sidebands of the signal region large, and the signal shape unperturbed.
3. The *multivariate analysis (MVA) selection* stage uses machine learning methods to optimize the background rejection.

Sample Name	Attributes
DstarToD0Pi_ D0To2Mu_ MuFilter	<ul style="list-style-type: none"> – SoftQCD:nonDiffractive (MinBias) – force $D^{*\pm} \rightarrow D^0\pi$, $D^0 \rightarrow \mu\mu$ decay – $\eta(D^{*\pm}) < 3.0$ – $\eta(D^0) < 3.0$ for D^0 from $D^{*\pm}$ – $\eta < 2.6$, $p_T > 3.0$ for muons from D^0
DstarToD0Pi_ D0To2Pi_ PiFilter	<ul style="list-style-type: none"> – SoftQCD:nonDiffractive (MinBias) – force $D^{*\pm} \rightarrow D^0\pi$, $D^0 \rightarrow \pi^\pm\pi^\pm$ decay – $\eta(D^{*\pm}) < 3.0$ – $\eta(D^0) < 3.0$ for D^0 from $D^{*\pm}$ – $\eta < 2.6$, $p_T > 3.0$ for pions from D^0
DstarToD0Pi_ D0ToKPi_ KPiFilter	<ul style="list-style-type: none"> – SoftQCD:nonDiffractive (MinBias) – force $D^{*\pm} \rightarrow D^0\pi$, $D^0 \rightarrow K^-\pi^+$ decay – $\eta(D^{*\pm}) < 3.0$ – $\eta(D^0) < 3.0$ for D^0 from $D^{*\pm}$ – $\eta < 2.6$, $p_T > 3.0$ for pions from D^0
DstarToD0Pi_ D0To2Pi_ PiToMuPiFilter_ PiLifetime0p02	<ul style="list-style-type: none"> – SoftQCD:nonDiffractive (MinBias) – force $D^{*\pm} \rightarrow D^0\pi$, $D^0 \rightarrow \pi^\pm\pi^\pm \rightarrow \mu^+\nu_\mu\mu\nu_\mu$ – reduce pion $c\tau$ from 7.8 m to 15.6 cm – pion decay limits: $R < 2$ m, $z < 4$ m – $\eta(D^{*\pm}) < 3.0$ – $\eta(D^0) < 3.0$ for D^0 from $D^{*\pm}$ – $\eta < 2.6$, $p_T > 3.0$ for pions from D^0 – $\eta < 2.6$, $p_T(\mu) > 3.0$ for muons from pions

Table 4.2: The 4 most relevant MC simulation samples and their defining attributes.

4.3.1 Preselection

The preselection is used to create reconstructed event candidates which pass the triggers discussed in Section 4.2.1. To stay consistent between signal and normalization events, we keep a similar preselection process for both $D^{*\pm} \rightarrow D^0\pi^\pm$, $D^0 \rightarrow \mu\mu$ and $D^{*\pm} \rightarrow D^0\pi^\pm$, $D^0 \rightarrow \pi^\pm\pi^\pm$ events.

In order to properly reconstruct the $D^{*\pm} \rightarrow D^0\pi^\pm$, $D^0 \rightarrow \mu\mu$ and $D^{*\pm} \rightarrow D^0\pi^\pm$, $D^0 \rightarrow \pi^\pm\pi^\pm$ events, we first must reconstruct their decay products: the muons and pion. Pions are reconstructed from charged tracks found in the tracker. These tracks are reconstructed using particle flow (PF) algorithms. The primary goal of these algorithms is to reconstruct individual particles using the data read out from the detector itself. This is made especially difficult because of the high luminosities of Run3 resulting in a large amount of *pileup*, a phenomenon that occurs due to multiple proton-proton collisions happening within a very short time frame resulting in multiple collisions per event. Due to the fact that the pions used in the analysis come from the PF algorithm, they are labeled as PF candidates. Muons are reconstructed primarily using detector read out from the muon chambers. We use the well established CMS reconstruction algorithms `TrackerMuon` and `GlobalMuon` as well as use the collaboration’s `LooseMuonID` for muon identification [16]. To reduce background noise and increase detector resolution, we also cut at $p_T > 4$ GeV and require a `highPurity` inner track in the tracker for both pions and muons.

Once we have the muon and/or pion candidates for the D^0 decay, we use vertex reconstruction to reconstruct the full decay candidate. A *vertex* is the location in 3D space where a process occurred and the *primary vertex* is the location of the interaction of the two colliding protons, which in our case produce the $D^{*\pm}$. Due to the short mean lifetime of the $D^{*\pm}$ ($6.9 \pm 1.9 \times 10^{-21}$ s) [13], we can label the $D^{*\pm} \rightarrow D^0\pi^\pm$ vertex using the primary vertex. The kinematic vertex reconstruction begins by identifying the dimuon or dipion decay candidates fit to a common vertex (i.e two pions or two muons which came from the same decay). The two 4-momentum vectors of these two candidates are added together to get a dimuon or dipion 4-momentum vector. We call this dimuon/dipion system a D^0 candidate. Using the D^0 candidate 4-momentum vector, we calculate the transverse momentum of D^0 candidate and extrapolate it to its intersection with the beamline. Then, in order to determine which primary vertex the $D^{*\pm} \rightarrow D^0\pi^\pm$ decay came from, we calculate the 3D distance between each primary vertex candidate and the extrapolated intersection point, known as the *3D-impact parameter*, and find the primary vertex which minimizes this parameter. This is difficult because there are often many primary vertices in one event due to pileup. Lastly, since the decay at the primary vertex is $D^{*\pm} \rightarrow D^0\pi^\pm$, we check if there exists a soft pion which came from the selected primary vertex [14]. This kinematic vertex reconstruction is then used to gather reconstructed signal and normalization events as well as refit the D^0 candidates to a common vertex using a kinematic vertex fitting tool, generating *refitted* candidates. The events which pass this reconstruction are thus the events that pass the preselection.

4.3.2 Baseline selection

Using the reconstruction described in Section 4.3.1, we are able to extract several kinematic variables from $D^{*\pm} \rightarrow D^0\pi^\pm$ candidates outputted by the preselection. These variables are

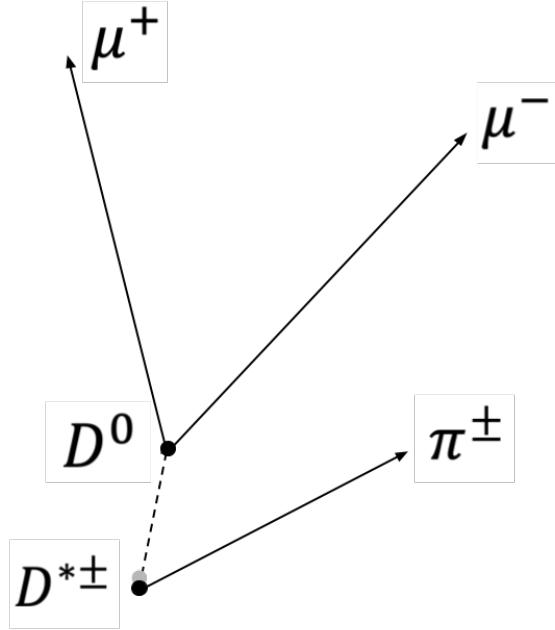


Figure 4.2: An example of the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \mu\mu$ vertex

used in the baseline selection as well as in other parts of the analysis. They are defined here as follows:

1. Reconstructed D^0 mass: the mass calculated from the addition of the two 4-momentum vectors of the D^0 's product candidates (either dimuon or dipion).
2. Refitted D^0 mass: the mass calculated from the addition of the two 4-momentum vectors of the D^0 's product candidates (either dimuon or dipion) once they have been refitted using the kinematic vertex fitting.
3. Reconstructed Δm : the mass difference between the reconstructed $D^{*\pm}$ candidates and the reconstructed D^0 candidates.
4. Refitted Δm : the mass difference between the $D^{*\pm}$ and the D^0 candidates once they have been refitted using the kinematic vertex fitting.
5. δ_{3D} : the 3D impact parameter as defined in Section 4.3.1.
6. $\delta_{3D}/\sigma(\delta_{3D})$: the significance of the 3D impact parameter. This is calculated by taking the value of the 3D impact parameter and dividing it by the square root of the expected variance of the parameter.
7. l_{3D} : the 3D distance between the vertex of the $D^{*\pm} \rightarrow D^0\pi^\pm$ decay (primary vertex) and the vertex of the $D^0 \rightarrow ll$ decay. This can also be called the *flight length* of the D^0 meson.

8. $l_{3D}/\sigma(l_{3D})$: the significance of the 3D distance between the vertex of the $D^{*\pm} \rightarrow D^0\pi^\pm$ decay (primary vertex) and the vertex of the $D^0 \rightarrow ll$ decay. Calculated by taking the value of the distance and dividing it by the square root of the expected variance of the distance.
9. l_{xy} : the distance in the xy plane (perpendicular to the beam line) between the vertex of the $D^{*\pm} \rightarrow D^0\pi^\pm$ decay (primary vertex) and the vertex of the $D^0 \rightarrow ll$ decay. This can also be called the *transverse flight length* of the D^0 meson.
10. $l_{xy}/\sigma(l_{xy})$: the significance of the distance in the xy plane between the vertex of the $D^{*\pm} \rightarrow D^0\pi^\pm$ decay (primary vertex) and the vertex of the $D^0 \rightarrow ll$ decay. Calculated by taking the value of the distance and dividing it by the square root of the expected variance of the distance.
11. α_{3D} : the angle between the D^0 momentum and flight direction.
12. D^0 vertex probability: the probability given by the χ^2 fit which reconstructs the D^0 vertex.
13. $D^{*\pm}$ vertex probability: the probability given by the χ^2 fit which reconstructs the $D^{*\pm}$ vertex.

Note that, unless otherwise stated, the refitted variables are used over the reconstructed ones.

Using these variables, we impart a baseline selection on the preselected events. The goal of the baseline selection is to reject much of the background while keeping the signal efficiency high and not perturbing the signal shape. To achieve this, we select a D^0 reconstructed mass in the range of [1.75, 1.95], D^0 refitted mass in the range of [1.81, 1.94], reconstructed Δm in the range of [0.135, 0.160], and refitted Δm in the range [0.140, 0.150]. These ranges are picked such that there are large sidebands on the signal, keeping signal efficiency high.

The other set of baseline selections are on the vertices themselves. We require the $D^{*\pm}$ vertex probability to be greater than 0.1 and the D^0 vertex probability to be greater than 0.01. This is done such that there is some confidence in the vertex reconstruction and such that we can match the double muon trigger requirement of 0.005. To gain further confidence in the vertex reconstruction, we limit $\alpha_{3D} < 0.1$ radians and the flight length significance to be greater than 3. Lastly, to keep the normalization channel (which is gathered from a ZeroBias trigger) under the same selection as the signal channel (which is gathered from a HLT_DoubleMuon trigger), we require the event in the normalization channel to have fired the HLT_DoubleMu4_3_LowMass trigger. Note, this does not mean that the specific decay we reconstruct fired the trigger, in fact usually some other particle has fired the trigger.

4.3.3 Multivariate Analysis

The preselection and baseline selections optimize signal efficiency, but not overall analysis sensitivity. In order to optimize the analysis sensitivity, we train a classifier using a decision tree model driven multivariate analysis (MVA). This single classifying parameter is used as a selection parameter named MVA_D and the cut is tuned to optimize analysis sensitivity.

The decision tree model used is based on the XGBoost (Extreme Gradient Boosting) library [4], which builds a forest of regression trees trained sequentially to minimize a regularized objective function. Each tree in the sequence focuses on correcting the errors of the previous tree and the scores from the trees are combined to get a final prediction on the classification of the event from the forest. Each tree is constructed using a greedy algorithm that selects splits based on gain, with regularization terms penalizing model complexity to prevent overfitting. The loss function is binary logistic loss, which is optimized using boosting under second-order gradient information with an evaluation metric based on the area under the receiver-operating characteristic (ROC) curve, known as AUC in the literature. Each tree has a maximum depth of three and is trained using a learning rate of $\eta = 0.1$. An additional regularization is applied with an L1 trigger penalty. A minimum loss reduction threshold for tree splits, which only allows additional trees to form if their contribution causes the loss to decrease by some amount (set in this analysis to be 2.0), and a sub sample ratio, which is the fraction of training data that each tree sees (kept at 60 percent in this analysis, as is standard), are employed to additionally prevent overfitting. The model is trained for over 4000 epochs in each training round.

The signal events used for training are taken from simulated $D^0 \rightarrow \mu\mu$ MC samples. The background events are taken from the data sidebands using the Δm parameter with a distance of over 5σ right of the expected value ($\Delta m \in [0.150, 0.155]$ GeV) and dimuon mass in the signal range of $[1.81, 2.45]$ GeV. The dimuon mass is kept in the signal range to align the training data with not obviously rejected events. Should the background training data be selected outside the signal dimuon mass window, the classification of signal and background events would become trivial, leading to a less effective model.

The variables used in the training are

1. the p_T of both, muons and the soft pion;
2. the D^0 vertex parameters, including point angle, flight length significance, vertex probability, 3D impact parameter, and significance of the 3D impact parameter;
3. the $D^{*\pm}$ vertex probability;
4. the D^0 mass resolution over the D^0 mass;

The correlation matrix for both the signal and background are drawn in Figure 4.3. Importantly, there are some positively correlated features, and no two features are strongly negatively correlated, meaning we expect stability in learning due to feature redundancy and simpler interactions, while still maintaining good model generalization.

Due to the relatively small amount of data available, the data is split into five groups and five separate models are training, each using a different data group as testing data sets and the remaining four as training datasets. This ensures the models have been exposed on the entire dataset while not overfitting on any particular events. Once the models have been trained, the event number, which is independent of the contents of the event, is used to decide which model to use for classification of that event.

It is important that the classification parameter is not correlated with any variables used later in the analysis for fitting (namely Δm and the dimuon mass) so as to not skew the

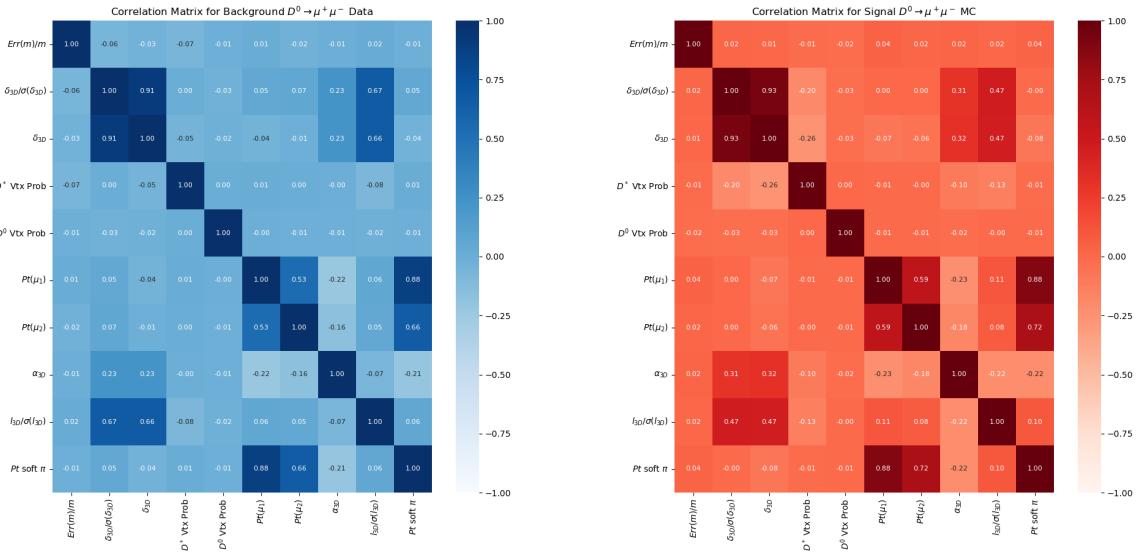


Figure 4.3: Correlation matrix over the MVA training variables for training background events generated from data (left) and training signal events generated from MC (right)

fit. Due to this, the kinematic variable given to the model are restricted to p_T and vertex parameters, so that it is not possible to reconstruct masses. To check that the correlations do not exist, a correlation matrix between the classification parameter, Δm , and the dimuon mass is created and shown in Figure 4.4. As expected, there is no correlation.

It is also important to note that the classifier is being trained and tested on simulated signal samples, yet for the analysis the efficiency of the classifier on real data is needed. Due to mismodeling effect in the simulation, the simulation efficiency is not necessarily the same as the data efficiency. Getting the efficiency is difficult to measure directly because the expected branching fraction of $D^0 \rightarrow \mu\mu$ is so low that we do not have a reliable data sample for the signal. Instead, the efficiency is derived from the $D^0 \rightarrow \pi^\pm \pi^\pm$ decay using the zero bias dataset and the $D^0 \rightarrow \pi^\pm \pi^\pm$ MC simulation samples. The $D^0 \rightarrow \pi^\pm \pi^\pm$ decay behaves similarly to the $D^0 \rightarrow \mu\mu$ decay in terms of reconstruction/vertex parameters and the MVA classifier is independent of Δm and dimuon mass in both cases, therefore the efficiency derived from $D^0 \rightarrow \pi^\pm \pi^\pm$ is similar to the efficiency of $D^0 \rightarrow \mu\mu$. The same classifier with the same cut is used on both the normalization channel and the signal channel, causing the MVA_D efficiencies to mostly cancel in the branching fraction equation (Equation 4.1). We calculate the efficiency in data on the $D^0 \rightarrow \pi^\pm \pi^\pm$ decay by performing the UML fit outlined in Section 4.4.1 on different MVA_D cut values and comparing to the `ZeroBias` dataset. In simulated data, we are able to directly tag the signal events that were created from the simulation, so calculating the efficiency is trivial. A list of efficiency values at specific MVA_D cuts are found in Table 4.3 and a graphical representation over a more diverse MVA_D cut range is found in 4.5.

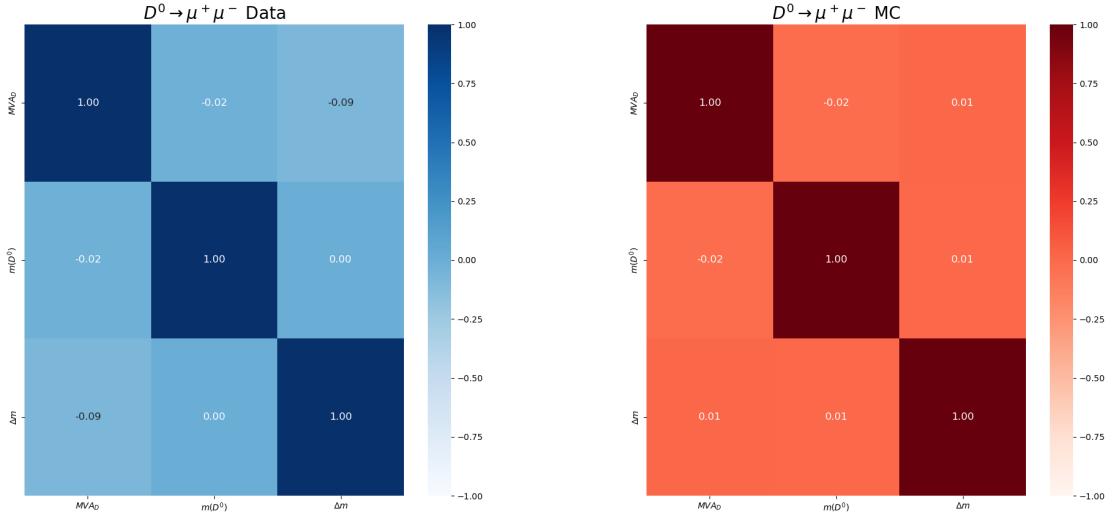


Figure 4.4: Correlation matrix comparing the MVA classification parameter to Δm and the dimuon mass. Shown left is the training background events generated from data and shown right is the training signal events generated from MC.

MVA _D cut value	0.74	0.76	0.78
$D^0 \rightarrow \pi^\pm \pi^\pm$ Data Efficiency (from Zero Bias)	0.722 ± 0.088	0.707 ± 0.087	0.681 ± 0.086
$D^0 \rightarrow \pi^\pm \pi^\pm$ MC Efficiency	0.817	0.801	0.783
$D^0 \rightarrow \mu\mu$ MC Efficiency	0.824	0.809	0.792
$D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\bar{\nu}_\mu$ MC Efficiency	0.806	-	-

Table 4.3: Summary of MVA_D cut efficiencies at various cut values on various samples used throughout the analysis.

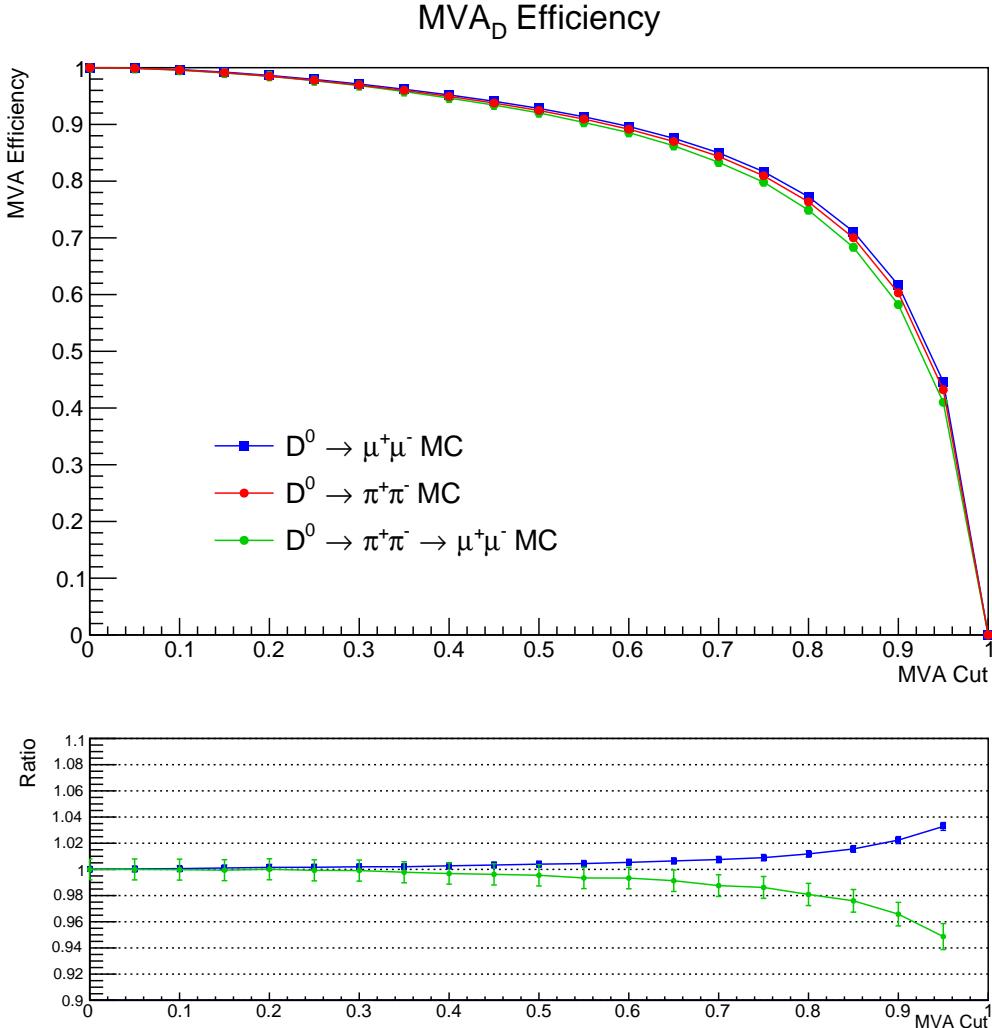


Figure 4.5: A graphical summary of MVA_D cut efficiency at various diverse cut values on various samples used throughout the analysis.

As is drawn in Table 4.3 as well as Figure 4.5, while the efficiency values are, as expected, extremely close in the MC samples, there is still a small difference between the three decay channels. To account for this small difference, we calculate a corrective efficiency factor, derived from MC. We name this corrective factor $MVA_{D,cor}$ and define it as

$$\begin{aligned} MVA_{D,cor}(D^0 \rightarrow \mu\mu) &= \frac{\epsilon_{D^0 \rightarrow \mu\mu} \text{ (simulation)}}{\epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm} \text{ (simulation)}} \\ MVA_{D,cor}(D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\nu_\mu) &= \frac{\epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\nu_\mu} \text{ (simulation)}}{\epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm} \text{ (simulation)}} \end{aligned} \quad (4.3)$$

From these definitions, one extrapolates to data due to the proximity of the efficiency values,

meaning

$$\begin{aligned}\epsilon_{D^0 \rightarrow \mu\mu}(\text{data}) &= \epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm}(\text{data}) \times \text{MVA}_{D,\text{cor}}(D^0 \rightarrow \mu\mu) \\ \epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\nu_\mu}(\text{data}) &= \epsilon_{D^0 \rightarrow \pi^\pm \pi^\pm}(\text{data}) \times \text{MVA}_{D,\text{cor}}(D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\nu_\mu)\end{aligned}\quad (4.4)$$

Lastly, a very conservative systematic uncertainty is assigned to the $\text{MVA}_{D,\text{cor}}$ factors of $|1 - \text{MVA}_{D,\text{cor}}|$. Therefore, we get

$$\begin{aligned}\text{MVA}_{D,\text{cor}}(D^0 \rightarrow \mu\mu) &= 1.009 \pm 0.012 \text{ (stat)} \pm 0.009 \text{ (sys)} \\ \text{MVA}_{D,\text{cor}}(D^0 \rightarrow \pi^\pm \pi^\pm \rightarrow \mu\nu_\mu \mu\nu_\mu) &= 0.987 \pm 0.020 \text{ (stat)} \pm 0.013 \text{ (sys)}\end{aligned}\quad (4.5)$$

4.4 Unbinned Maximum Likelihood Fits

Given the accurate descriptions of the datasets and selection processes used in this analysis, we are now ready to extract the two main values needed for this analysis, the number of $D^0 \rightarrow \pi^\pm \pi^\pm$ events, denoted $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$, the number of $D^0 \rightarrow \mu\mu$ events, denoted $N_{D^0 \rightarrow \mu\mu}$.

In order to do this, we perform two unbinned maximum likelihood (UML) fits on the events in our data that passes the preselection, baseline selection, and MVA selection processes. Maximum likelihood (ML) fits are statistical methods used to estimate parameters of a model given observed data. A ML fit is a UML fit when the data is not binned into a histogram, but rather the specific data values for each event separately are used. UML fits define a model $f(\vec{x}; \vec{\theta}, \vec{\theta}_N)$, which is a probability density function (PDF) over some set of observed variables, \vec{x} and some model parameters $\vec{\theta}$ and $\vec{\theta}_N$. The goal of a UML fit is to extract model parameters $\vec{\theta}$ given some observed data $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$. $\vec{\theta}_N$ are considered nuisance parameters and are important to defining the model, but are not parameters of interest to the analysis. Once a model has been defined, a likelihood is defined as

$$\mathcal{L}(\vec{\theta}, \vec{\theta}_N) = \prod_{i=1}^N f(\vec{x}_i; \vec{\theta}, \vec{\theta}_N) \quad (4.6)$$

In this formulation, the UML fit is represented as finding $\hat{\vec{\theta}} = \text{argmax}_{\vec{\theta}, \vec{\theta}_N} \mathcal{L}(\vec{\theta}, \vec{\theta}_N)$. In practice, this is not quite true. Often in applications with lots of nuisance parameters, UML fits operate using profile likelihoods. However, the point of this discussion is merely an overview of UML and the complexities will be ignored until Section 4.7.1.

In our analysis, $\vec{\theta}$ is the true number of signal events (in the normalization channel this is $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$ and in our signal channel this is $N_{D^0 \rightarrow \mu\mu}$), \vec{x} is the Δm and $m(D^0)$ of our data, and $\vec{\theta}_N$ are the variance parameters needed to construct the signal and background PDFs. As is drawn in Figure 4.4, there exists practically no correlation between Δm and $m(D^0)$. This tells us that the 2D model is a product of 1D models. More specifically, we have that

$$f(\vec{x}; \vec{\theta}, \vec{\theta}_N) = f_{m(D^0)}(m(D^0); \vec{\theta}, \vec{\theta}_N) \times f_{\Delta m}(\Delta m; \vec{\theta}, \vec{\theta}_N) \quad (4.7)$$

A more precise discussion of this construction follows in the sections below. For each of the two channels we begin by developing the models for both the signal and the background. Then, we outline the specifics of the fit and any corrections applied to it. Lastly, we outline the results of the fit.

4.4.1 Normalization Channel Fit

The goal of the normalization channel is to extract $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$. This is done using the `HLT_ZeroBias` trigger dataset, as is described in Section 4.2.1. Perhaps the largest difficulty of the UML fit on the normalization channel is that the `HLT_ZeroBias` trigger dataset does not contain many of the $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow \pi^\pm \pi^\pm$ signal events due the large prescaling factor of the trigger, discussed in Section 4.2.1. The below Section, in part, describes in detail how this difficulty is overcome by using MC samples to inform the fit and correcting for MC mismodeling effects when needed in the final fit.

Signal Model

The signal model for the normalization channel describes the Δm and $m(D^0)$ of the $D^{*\pm} \rightarrow D^0 \pi^\pm, D^0 \rightarrow \pi^\pm \pi^\pm$ decay. As is common to do for masses in signal distributions, the model for $m(D^0)$ is a sum of two Gaussian distributions forced to share a common mean. Similarly, the model for Δm is a sum of three Gaussian distributions forced to share a common mean.

Due to the small number of signal events in data, the shape of this model is determined by fits to simulated MC samples. This is because the shape will be much more stable in MC compared to data, simply just because there are more events in MC. Once the shape is found using MC, it is frozen and only the number of events is allowed to float when the fit is applied to data. The signal model fit to MC is displayed in Figure 4.6.

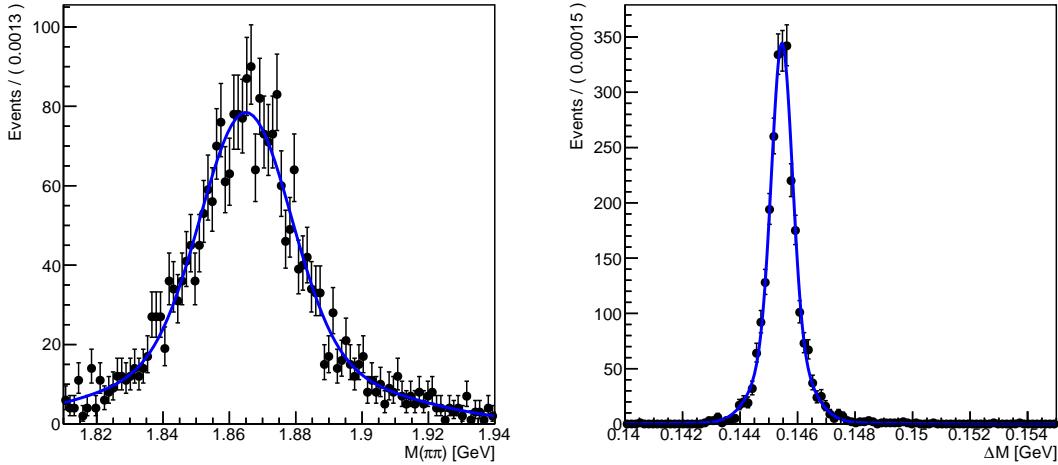


Figure 4.6: The signal model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.

It is possible, however, that the shape varies slightly between the MC samples and data, due to mismodeling. To account for this slight mismodeling affect, we derive correction factors μ_{corr} and σ_{corr} such that

$$\begin{aligned}\mu_{\text{data}} &= \mu_{\text{MC}} \times (1 + \mu_{\text{corr}}) \\ \sigma_{\text{data}}^i &= \sigma_{\text{MC}} \times (1 + \sigma_{\text{corr}})\end{aligned}\tag{4.8}$$

where i is used to denote that each of the models has multiple width parameters (one for each Gaussian distribution that is summed) but only one mean parameter.

We use the `HLT_DoubleMu4_3_LowMass` trigger to construct a dataset with enough signal samples to construct a reliable data-driven fit shape. We then run the full normalization channel UML fit (including the background models) on this data. Comparing that to the shape obtained from MC, we can easily derive the μ_{corr} from Equation 4.8.

Initially, it might seem that one could apply the same procedure to find σ_{corr} . However, this dimuon trigger enhances the amount of $b\bar{b}$ production in the sample, which biases the width of the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm$ decay. Therefore, one must first properly understand that bias before applying the correction in the same way was done for the mass. To understand the bias effect between the Dimuon and Zerobias triggers, we compare $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$ fits from the Dimuon and Zerobias dataset. This is possible because the Kaon decay mode is much more frequent than the pion decay mode. The fitting process is the same as before and the width correction is obtained as

$$1 + \sigma_{\text{corr}} = \left(1 + \sigma_{\text{corr}, \text{DoubleMuon}}^{D^0 \rightarrow \pi^\pm\pi^\pm} \right) \times \frac{1 + \sigma_{\text{corr}, \text{ZeroBias}}^{D^0 \rightarrow \pi^\pm\pi^\mp}}{1 + \sigma_{\text{corr}, \text{DoubleMuon}}^{D^0 \rightarrow K^\pm\pi^\mp}} \quad (4.9)$$

Background Model

As described in Section 4.2.2, there are three peaking backgrounds and a combinatorial background to consider in the normalization channel.

To begin, we describe the three peaking backgrounds. The first, as most prominent, of these three backgrounds is the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$. As shown in Figure 4.1, the $m(D^0)$ peak is shifted significantly to the left due to incorrect mass of the pion imposed on the kaon during reconstruction. In fact, the mean is shifted so far that only the tail of the Gaussian distribution persists in the mass region of our fit. Since the tail of a Gaussian distribution is approximately an exponential function, we use an exponential function to fit the $m(D^0)$ parameter. This fit proves to be significantly more stable as the exponential function only has one shape parameter that requires fitting while the Gaussian distribution has two. In contrast, the Δm distribution of the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$ decay closely follows the shape of the signal. Therefore, we similarly use a sum of three Gaussian distributions with a common mean to model the Δm distribution. The convergence of the fit is verified using MC samples and is displayed in Figure 4.7. The data fit remains stable even when these shape parameters are allowed to vary. Therefore, we allow the shape of this peaking background to vary when fit to data.

The other two peaking backgrounds correspond to $D^0 \rightarrow \mu\mu$ and $D^0 \rightarrow K^\pm\pi^\mp$ decays where the D^0 meson does not originate from a $D^{*\pm}$ meson. Importantly, these have the same $m(D^0)$ shape as their $D^{*\pm} \rightarrow D^0\pi^\pm$ counterparts, so the shape derived from their counterparts is copied exactly into their models. The Δm of these decays, however, does not represent a peak but rather a combinatorial signature. This is due to the combinatorial nature of the production mechanisms of the D^0 meson that pass the selection criterion. Therefore, we fit a combinatorial background model for the Δm values of both non- $D^{*\pm} \rightarrow$

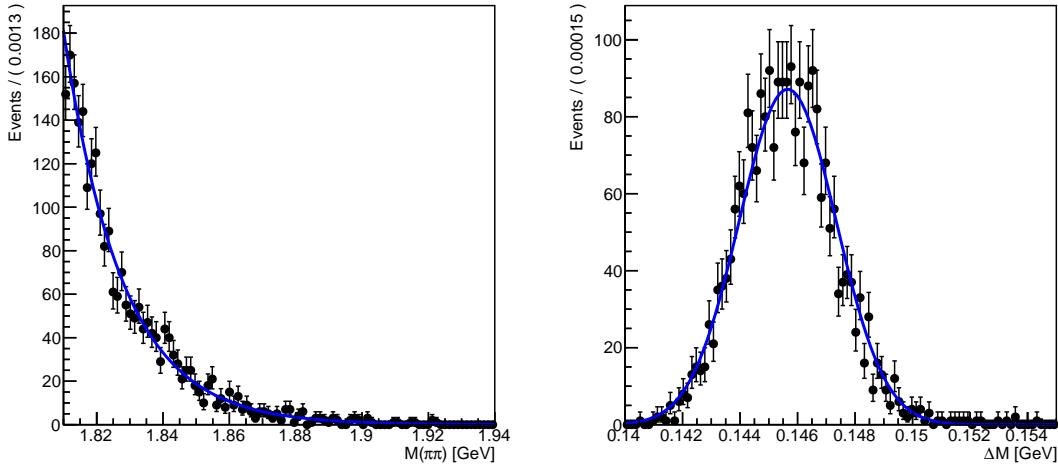


Figure 4.7: The $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.

$D^0\pi^\pm$ decays.

The combinatorial background for $m(D^0)$ is an exponential function, as is common for modeling combinatorial backgrounds of reconstructed rest masses. The combinatorial background for Δm , however, has a more complex shape than just an exponential. In this work, we use the `RooDstDOBg` function [17] given by

$$P(m|m_0, A, B, C) = \left(1 - \exp\left(-\frac{m - m_0}{C}\right)\right) \left(\frac{m}{m_0}\right)^A \quad (4.10)$$

The shape of this function is derived from fits to the data side-band regions where the distance from the signal resonance guarantees virtually no peaking background contribution. This shape is frozen to allow for fit stability when the model is used in the final data fit. Figure 4.8 shows the convergence of the shape on the data side-bands.

Fit Results

The final model used for the fit is a sum of signal model and all the background models. We check the convergence of this model on the `HLT_DoubleMu4_3_LowMass` data, which contains significantly more signal events than the `HLT_ZeroBias` trigger samples, allowing us to visually inspect the stability of the fit, as is seen in Figure 4.10 and Table 4.4. Importantly, the $D^0 \rightarrow K^\pm\pi^\mp$ decays where the D^0 meson does not originate from a $D^{*\pm}$ meson contributes virtually no events to the fit. Therefore, to increase the stability of the fit, we force the integral under this non- $D^{*\pm}$, $D^0 \rightarrow K^\pm\pi^\mp$ model to 0.

Finally, the fit is applied to the `HLT_ZeroBias` trigger samples. A graphical representation of the fit is drawn in Figure 4.9 and the results are summarized in Table 4.4.

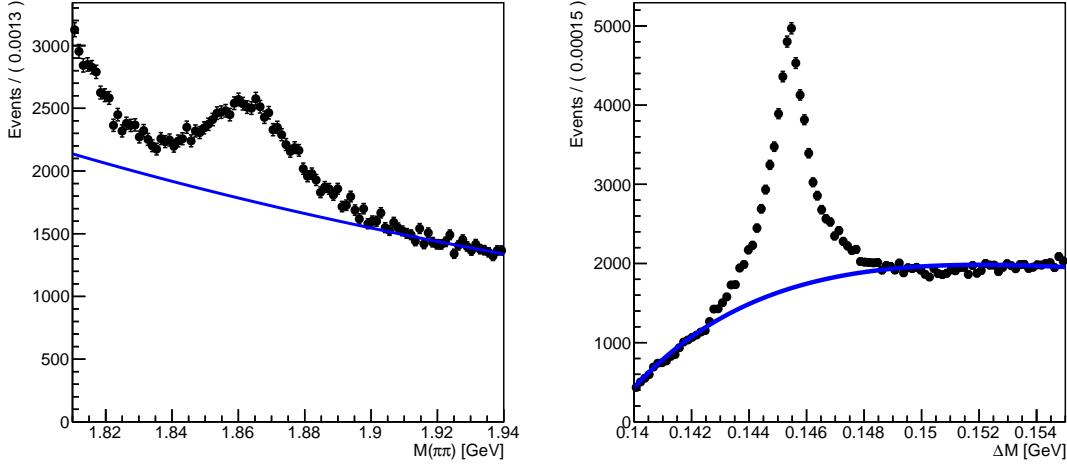


Figure 4.8: The combinatorial model fit on data side-bands, with $m(D^0)$ displayed left and Δm displayed right.

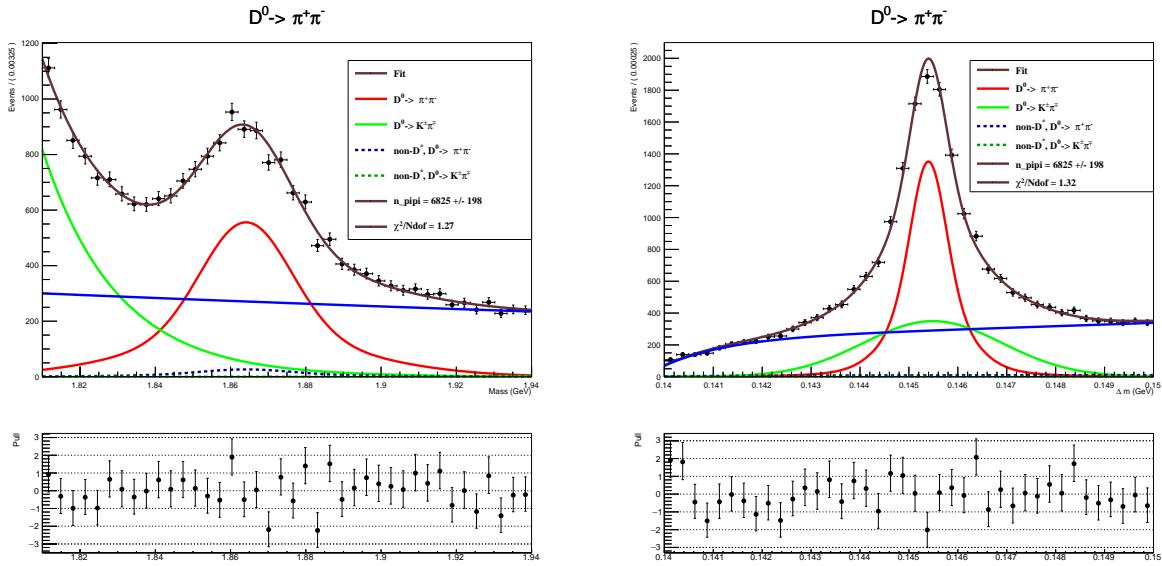


Figure 4.9: The full $D^0 \rightarrow \pi^\pm\pi^\pm$ model fit on HLT_ZeroBias data, with $m(D^0)$ displayed left and Δm displayed right.

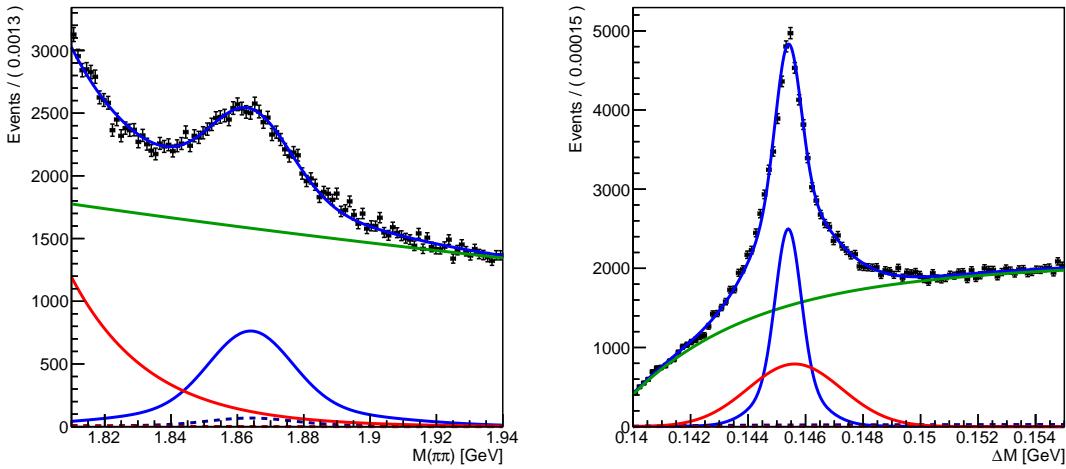


Figure 4.10: The full $D^0 \rightarrow \pi^\pm\pi^\pm$ model fit on HLT_DoubleMu4_3_LowMass data, with $m(D^0)$ displayed left and Δm displayed right. The green line follows the combinatorial model, the blue line follows the signal model, the red line follows the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$ background model, the red dashed line following the non- $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow K^\pm\pi^\mp$ background model, and the blue dashed line follows the non- $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm$ background model.

HLT_DoubleMu4_3_LowMass data	Fitted count	Fitted error
$D^{*\pm}, D^0 \rightarrow \pi\pi$ Yield	23787	± 509
$D^{*\pm}, D^0 \rightarrow K\pi$ Yield	21558	± 749
Combinatorial Yield	155088	± 1722
non- $D^{*\pm}, D^0 \rightarrow \pi\pi$ Yield	2140	± 685
non- $D^{*\pm}, D^0 \rightarrow K\pi$ Yield	181	± 1410

HLT_ZeroBias data	Fitted count	Fitted error
$D^{*\pm}, D^0 \rightarrow \pi\pi$ Yield	195	± 17
$D^{*\pm}, D^0 \rightarrow K\pi$ Yield	74	± 21
Combinatorial Yield	140	± 20
non- $D^{*\pm}, D^0 \rightarrow \pi\pi$ Yield	0	± 11
non- $D^{*\pm}, D^0 \rightarrow K\pi$ Yield	0	± 0

Table 4.4: Fitted yields and associated uncertainties for the normalization UML fit.

4.4.2 Signal Channel Fit

The goal of the signal channel is to extract $N_{D^0 \rightarrow \mu\mu}$. This is done using the HLT_DoubleMu4_3_LowMass trigger dataset, as is described in Section 4.2.2. As with the normalization channel, the largest difficult of this UML fit is the lack of signal events due to

the small branching fraction of the D^0 meson. Similarly to the normalization channel, the signal channel uses MC samples to inform the shapes for data, applying corrections when needed. Additionally, the results of the normalization channel are used to fix the background yields. These techniques are described in the remainder of the section.

It is worth mentioning that much of the machinery needed for this fit has been developed in Section 4.4.1 and is therefore not repeated here. Namely, this section focuses on the differences in the signal channel that require special attention.

Signal Model

The signal model for the normalization channel describes the Δm and $m(D^0)$ of the $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm$ decay. The same procedure for determining the normalization channel signal model is used in determining this signal model, with the mean and width corrections derived from the normalization channel.

The signal model fit to MC is drawn in Figure 4.11.

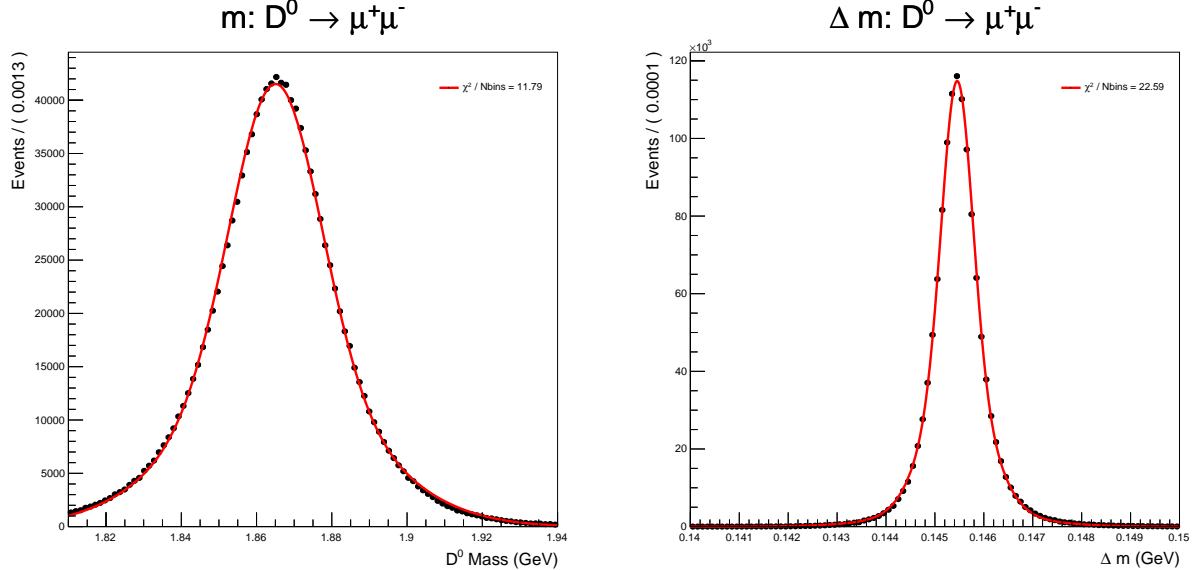


Figure 4.11: The signal model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.

Background Model

As is evident from Section 4.2.2, there are 2 $D^{*\pm} \rightarrow D^0\pi^\pm$ peaking backgrounds and a combinatorial background to consider in the normalization channel. Note that the non- $D^{*\pm} \rightarrow D^0\pi^\pm$ backgrounds which were discussed extensively in the normalization channel are not relevant here due to the large uncertainty of the combinatorial background compared to the expected yield from these backgrounds. Therefore, any non- $D^{*\pm} \rightarrow D^0\pi^\pm$ backgrounds are considered combinatorial and considered in the combinatorial model.

The two peaking backgrounds are $D^{*\pm} \rightarrow D^0\pi^\pm$ processes where the D^0 decays via either $D^0 \rightarrow \pi^\pm\pi^\pm \rightarrow \mu\nu_\mu\mu\bar{\nu}_\mu$ (hadronic mode) or via $D^0 \rightarrow pi^\pm\mu\nu_\mu$ (semileptonic mode). Both of these models are developed using the same process as the signal mode. Namely, the models are built out of sums of Gaussian distributions just like in the signal channel. Their shape is determined by fitting to MC samples, and their means and widths are corrected for using the correction derived in the normalization channel fit.

In the normalization channel, the yields of these models were left floating. In this channel, there are so few events that allowing for this would destabilize the fit. However, we have already calculated $N_{D^0 \rightarrow \pi^\pm\pi^\pm} = 195 \pm 17$ in Section 4.4.1 and the branching fractions of $\mathcal{B}(D^0 \rightarrow \pi^\pm\pi^\pm) = (1.454 \pm 0.024) \times 10^{-3}$ and $\mathcal{B}(D^0 \rightarrow pi^\pm\mu\nu_\mu) = (2.67 \pm 0.12) \times 10^{-3}$ are well known [13]. Lastly, in Section 4.5, we calculate the fake rate, $f_{\pi \rightarrow \mu}$, or the rate at which pions decay to muons in the detector, which occurs once in the semi-leptonic mode and twice in the hadronic mode.

Using these qualities, as well as efficiency and MC corrections, we can write

$$\begin{aligned} \frac{N_{D^0 \rightarrow \pi^+\pi^-\mu^+\nu_\mu\mu\bar{\nu}_\mu}}{N_{D^0 \rightarrow \pi^+\pi^\pm}} &= (f_{\pi \rightarrow \mu})^2 \times \frac{\epsilon_{D^*,D^0 \rightarrow \pi\pi \rightarrow \mu\mu}}{\epsilon_{D^*,D^0 \rightarrow \pi\pi}} \times S_{ZB} \times \text{MVA}_D \times T_{\text{corr}} \\ \frac{N_{D^0 \rightarrow \pi^\pm\mu^+\nu_\mu}}{N_{D^0 \rightarrow \pi^+\pi^\pm}} &= \frac{\mathcal{B}_{D^0 \rightarrow \pi^\pm\mu_\mu^\nu}}{\mathcal{B}_{D^0 \rightarrow \pi^+\pi^\pm}} \times (f_{\pi \rightarrow \mu}) \times \frac{\epsilon_{D^*,D^0 \rightarrow \pi\mu\nu}}{\epsilon_{D^*,D^0 \rightarrow \pi\pi}} \times S_{ZB} \times \text{MVA}_D \times T_{\text{corr}} \end{aligned} \quad (4.11)$$

where S_{ZB} is the trigger prescaling factor discussed in Section 4.2.1, ϵ is the efficiency discussed in Section 4.3.2, MVA_D is the MVA correction discussed in Section 4.3.3, and T_{corr} are the trigger corrections discussed in Section 4.6.1. The convergence of the two fits in MC is shown in Figure 4.12 and Figure 4.13

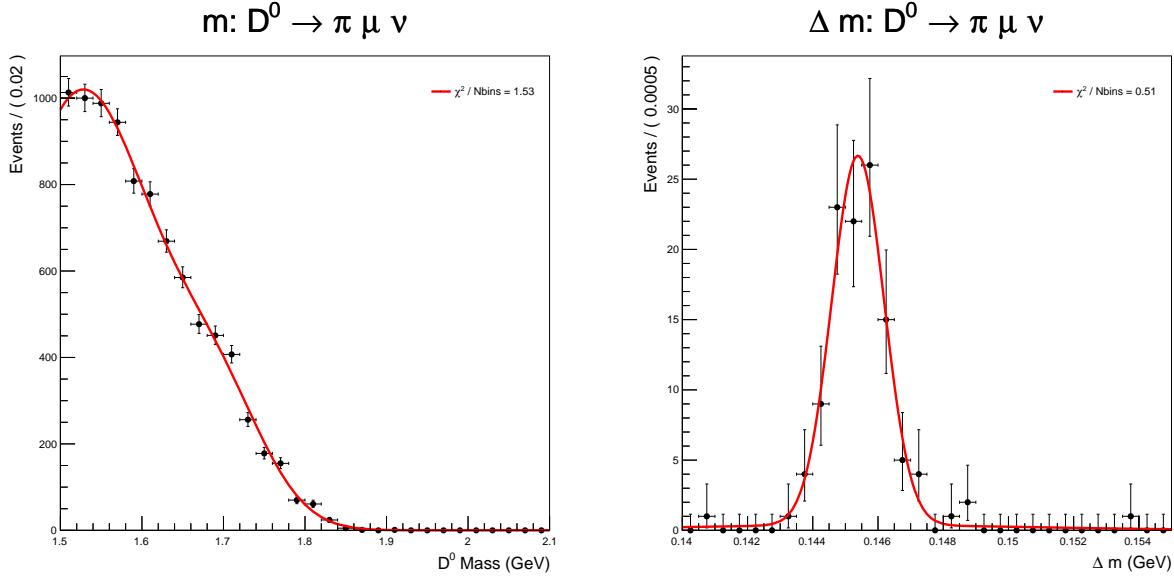


Figure 4.12: The $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow pi^\pm\mu\nu_\mu$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.

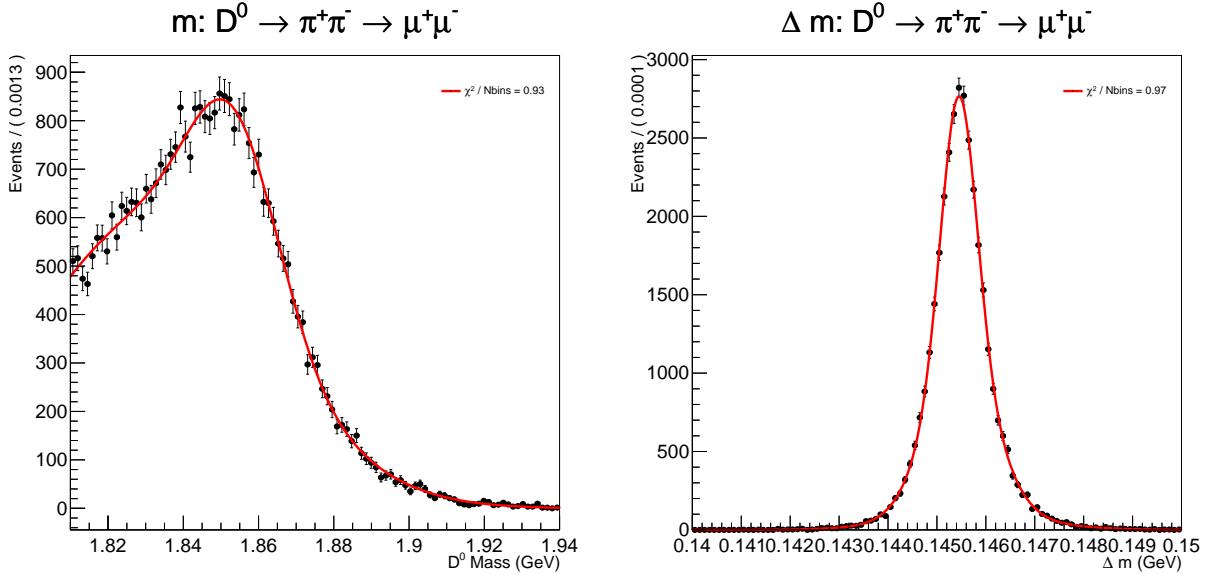


Figure 4.13: The $D^{*\pm} \rightarrow D^0\pi^\pm, D^0 \rightarrow \pi^\pm\pi^\pm \rightarrow \mu\nu_\mu\mu\bar{\nu}_\mu$ model fit on MC samples, with $m(D^0)$ displayed left and Δm displayed right.

The process for determining the combinatorial model is kept the same as for the normalization channel. The only difference is that due to the added complexity of the backgrounds, the $m(D^0)$ pdf function is a combination of a Bernstein polynomial, power law, and exponential function, instead of just an exponential function as was used in the normalization channel [8]. Similar to the normalization channel, the shape and yield is determined using data side-bands, leaving no combinatorial variables left to fit in the main data region. The convergence of the combinatorial fit using partially unblinded data sidebands is drawn in Figure 4.14.

Fit Results

The final model used for the fit is a sum of signal model and all the background models. One important note is to emphasize that the signal shape and peaking background shape are entirely deriving from MC, the peaking background yield is determined from Equation 4.11, and combinatorial background shape and yield are determined by the data side-bands. Therefore, the only parameter of this fit is the actual parameter of interest, namely the $N_{D^0 \rightarrow \pi^\pm\pi^\pm}$ events.

The results of this fit are only seen after unblinding the data in as discussed in Section 4.1. The results are presented in Section 4.7.3.

4.5 Muon Fake Rates

A muon fake occurs in this analysis when a pion decay product is measured as muon. This most commonly occurs due to in-flight pion decays, $\pi \rightarrow \mu\nu$, but can also occur due to

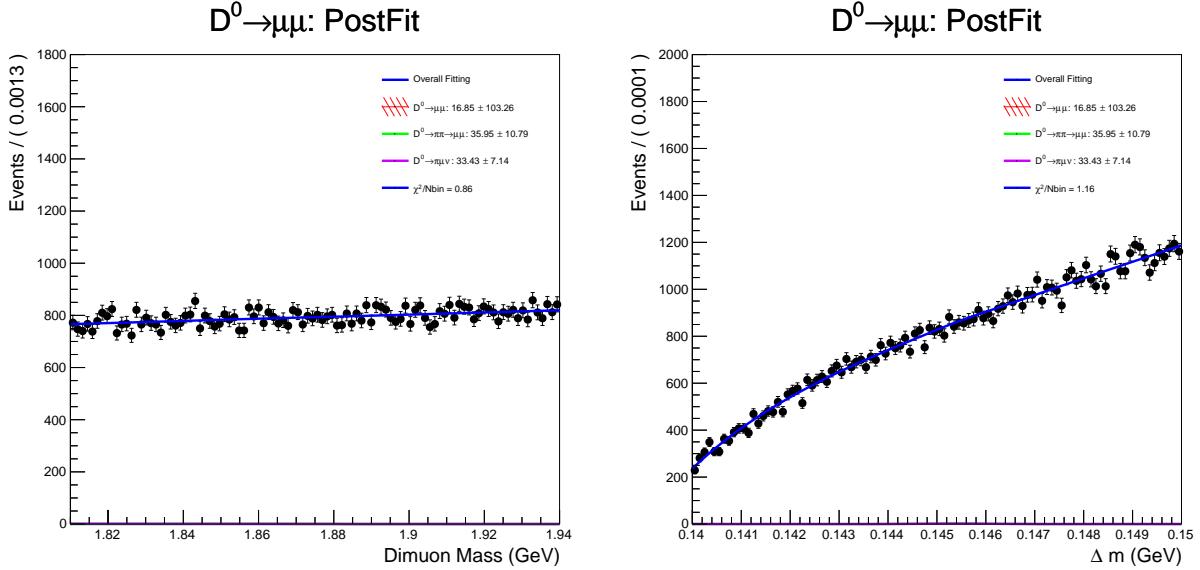


Figure 4.14: The fit to partial unblinded data before any MVA_D cut, demonstrating the stability of the combinatorial shape which dominates the overall shape.

detector misreconstruction. As a result, the more common $D^0 \rightarrow \pi^\pm \pi^\pm$ process can mimic genuine $D^0 \rightarrow \mu\mu$ events when the pions decay in flight, producing muons that are misidentified as signal and causing a significant challenge to this analysis. Therefore, a clear and precise understanding of the muon fake rate must be achieved to be used throughout the rest of the analysis. The below method was first developed to search for the $B^0 \rightarrow \mu\mu$ decay [6]. Below we summarize the most relevant aspects of the procedure and results for the Loose muon identification system, the `highPurity` inner track quality requirements, tracker muon, and global muon requirements. These requirements reflect the selection criterion for the rest of the analysis described in Section 4.3.1 and Section 4.3.2.

4.5.1 Data Samples and Selections

To get a reliable prediction for muon fake rates we need to be able to simulate pion decays in flight. We also needed to be able to simulate reconstructing the pion as a muon candidate through a muon identification scheme. While pion decays into muons are simulated with adequate precision in MC, the reconstruction and muon identification parts are harder to simulate correctly. Therefore, we need to be able to measure the fake rates in data, then use those results to correct MC predictions. To increase the precision of the fake rate measurement, we use control samples that clearly identify hadrons in data and MC. Additionally, to avoid a dependence on the trigger, we pick samples with triggers that only fire on electrons. Using this, we remove trigger dependence on muons in the sample. Thus, the data samples used in this analysis are `EGamma` (2022), `EGamma0` (2023), `Egamma1` (2023), and `ParkingDoubleElectronLowMass`. The `EGamma` datasets require events to pass single-electron triggers, while `ParkingDoubleElectronLowMass` relies on low-mass dielec-

tron triggers with reduced prescales, enabling efficient selection of low-momentum hadronic decays.

For MC studies of muon fake rates, several background samples are used to provide a realistic modeling of processes that can produce hadrons misidentified as muons. Drell-Yan samples with dileptons and two associated jets are included, both in the low-mass (10-50 GeV) and high-mass (above 50 GeV) regimes, generated using the `amcatnloFXFX` and `madgraphMLM` frameworks interfaced with `pythia8`. A dedicated sample of Drell-Yan to dielectrons is also included to help disentangle muon-specific effects. Additionally, $W \rightarrow l\nu$ events with jets and $t\bar{t}$ events in the lepton+jets final state are included to capture fake muons arising in high-activity environments with real leptons and multiple jets. All samples are produced at $\sqrt{s} = 13.6$ TeV and use the `TuneCP5` parton shower configuration.

To study muon fakes originating from pions we use $K_S^0 \rightarrow \pi^\pm \pi^\pm$ control sample. The selection requirements are:

- $p_T^{\pi_1} > 1$ GeV,
- $p_T^{\pi_2} > 4$ GeV,
- pion track is highPurity,
- $m_{K_S^0} \in [0.45, 0.55]$,
- $\frac{L_{xy}}{\sigma_L} > 3$,
- vertex probability greater than 0.001,
- vertex displacement in XY plane wrt Beam Spot less than 8,
- cosine of pointing angle in XY wrt BS greater than 0.999,
- impact parameter significance of the candidate trajectory in 3D wrt PV less than 3,
- 2D impact parameter significance for Track 1 and 2 wrt Beam Spot greater than 5,
- kinematic vertex fit χ^2/dof of the two track less than 3, and
- fire the HLT Electron30 trigger.

4.5.2 Measurement Procedure

To measure the fake rate we perform an extended binned maximum likelihood fit to $\pi^+ \pi^-$ and $\mu^\pm \pi^\pm$ mass distribution to extract the event yield of $K_S^0 \rightarrow \pi^\pm \pi^\pm$ and $K_S^0 \rightarrow \mu^\pm \pi^\mp$ events. The model used to describe the signal and combinatorial background component are a Crystall Ball added to a Gaussian centered at the same mean and a 2nd order Bernstein polynomial function, respectively. The signal shape in the $K_S^0 \rightarrow \mu^\pm \pi^\mp$ distribution is fixed from the $K_S^0 \rightarrow \pi^\pm \pi^\pm$ distribution to avoid the different signal shape due to the low statistics. One of the fit projection is shown in the Figure 4.15

The fake rate is estimated in bins of muon p_T . For each $K_S^0 \rightarrow \pi^\pm \pi^\pm$ candidate we check p_T and acceptance for both pions and add the K_S^0 candidate to the appropriate p_T bin (4.16).

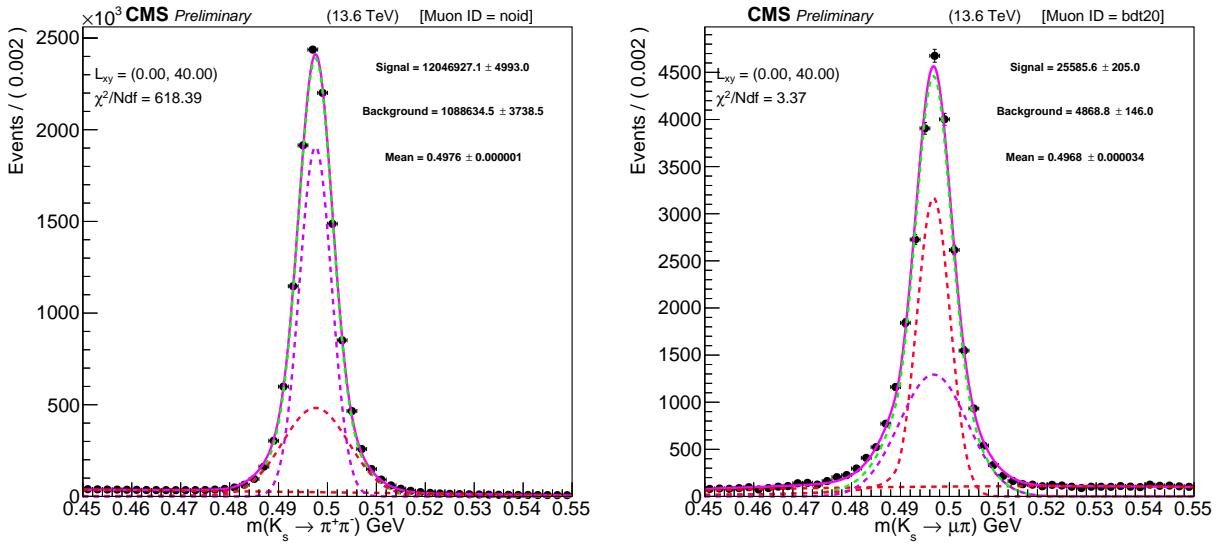


Figure 4.15: Mass projection from $K_S^0 \rightarrow \pi^\pm \pi^\pm$ decays in the μp_T range 4-40.0 GeV using 2022 data before (left) and after (right) soft muon MVA ID cut. The green dashed line is for the signal distribution, the red dotted line is for combinatorial background, and the result of the fit projection is the solid pink line. The data are shown with solid black circles.

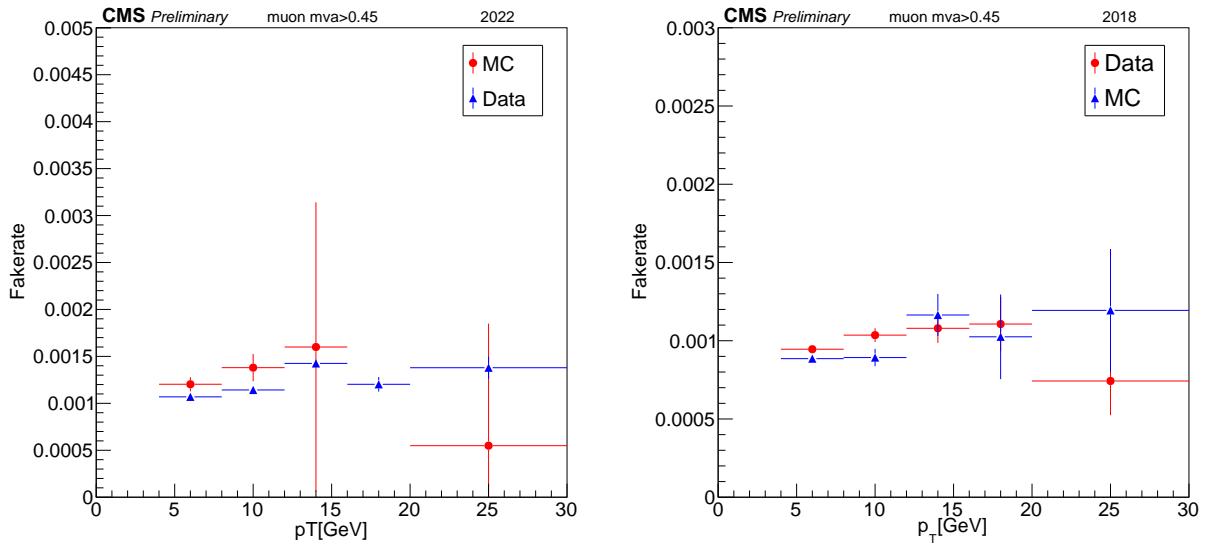


Figure 4.16: Pion muon fake rates for Run2022 (top) and Run2023 (bottom) as a function muon p_T .

The fake rate estimated from K_S^0 decays also has a dependence on the K_S^0 flight length (Figure 4.17). There are two different trends. The first is the fake rate increase with the flight length, which is most pronounced for the loose muon selection. The second one shows a

dramatic decrease in the fake rate for muon candidates passing the Soft Muon MVA selection when they originate outside of the pixel detector (flight length greater than 8 cm).

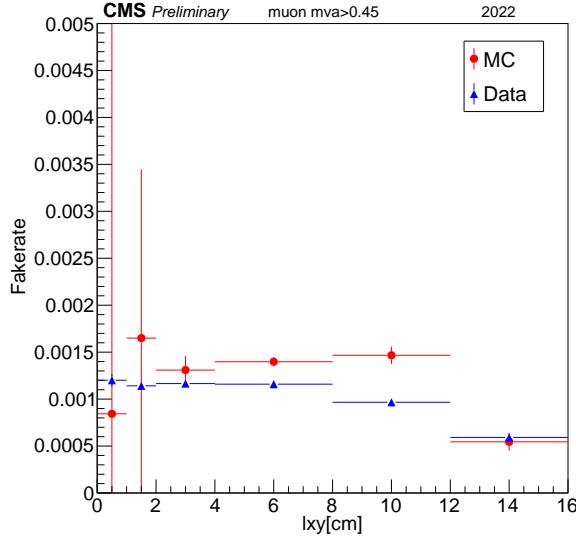


Figure 4.17: Fake rate dependence on flight length (l_{xy}) measured in cm for data and MC

Both effects are related to the number of pixel and silicon strip hits for the inner track, which is known to be smaller for fake muons. In the case of the Loose selection this information is not used directly for the muon identification, while for the Soft Muon MVA it is one of the most powerful discriminators. The fact that the fake rate has a non-trivial dependence on the flight length means that we have to match the flight length of the fakes used in the physics analysis to the one available in the control sample. Otherwise, this dependence can lead to a biased measurement of the fake rate since most K_S^0 have much larger flight length than the D mesons used in this analysis. Therefore, we restrict the transverse decay length to 8 cm to match the D mesons studied in the analysis. In this region, the fakerate seems to relatively flat and there is enough data to make precise estimates of the fake rate in this region.

Systematic error for the fake rates measurements in MC samples are derived from taking the weighted differences between the various MC samples used, namely W+Jets, Drell-Yan, and TT bar. Furthermore, systematic error for the fake rate measurements in data is mainly derived from the differences between EGamma and Parking Electron datasets. Agreement between these datasets is drawn in Figure 4.18. To account for the variation in terms of the l_{xy} value, an additional 2% of systematic uncertainty is assigned, motivated by the scale factor variation in terms of p_T and l_{xy} between data collected in 2022 and data collected in 2023.

The final fake rate results are summarized in Table 4.5.

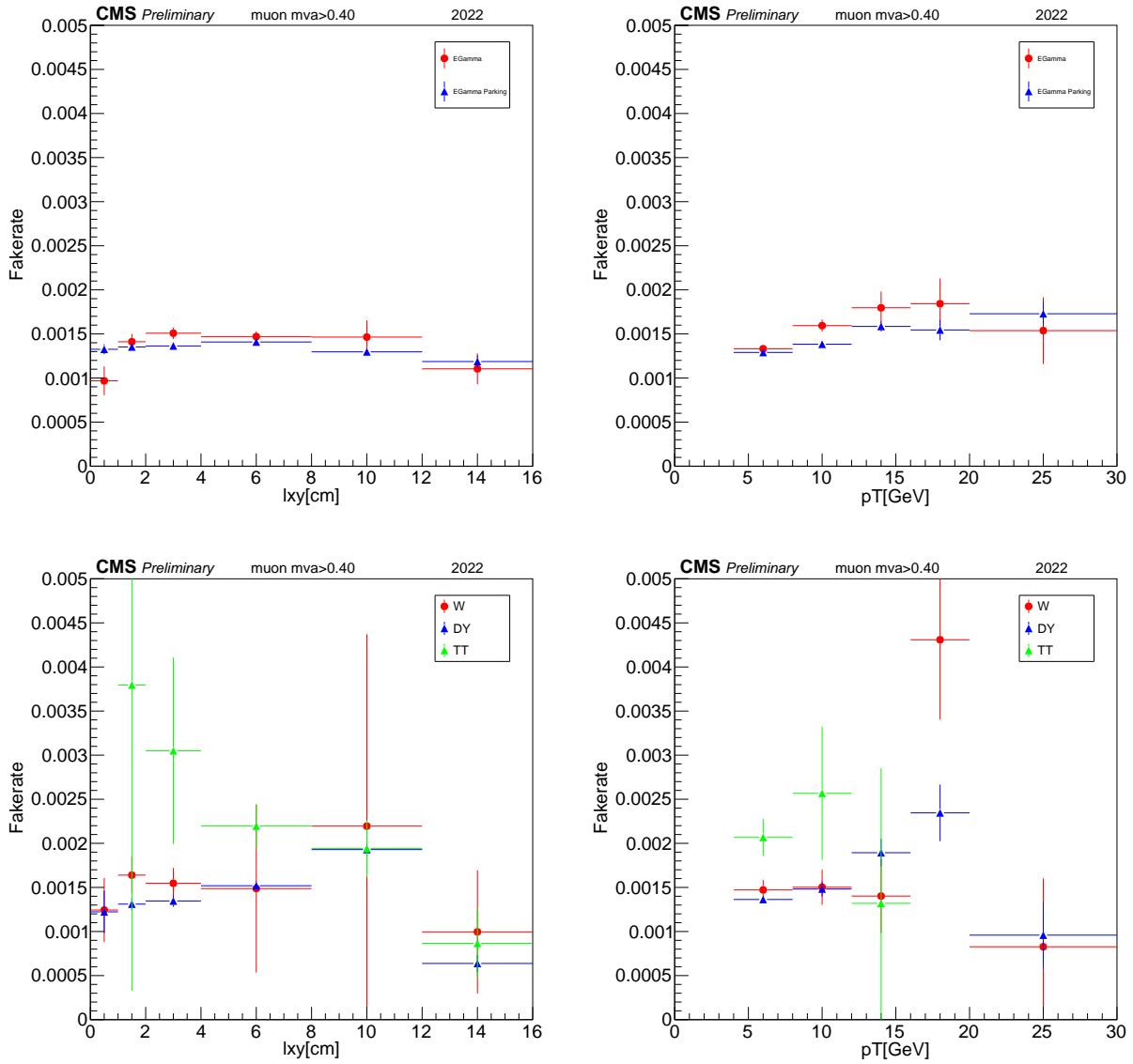


Figure 4.18: Difference in fakerate between Egamma and Parking Electron data (top) and W plus jets, Drell-Yan, and $t\bar{t}$ MC samples (bottom) binned by flight length in cm (left) and pT in GeV (right)

4.6 Efficiency Corrections

Virtually all selection efficiencies used for this analysis are measured in MC. Normally, this would be troublesome for the results of the analysis due to MC mismodeling effects that cause the efficiency of MC to be different from the efficiency of the data. In terms of a formula, we get that

$$\epsilon_{\text{data}} = \epsilon_{\text{MC}} \times \epsilon_{\text{corr}} \quad (4.12)$$

with mismodeling occurring when $\epsilon_{\text{corr}} \neq 1$.

	4–8 GeV	8–12 GeV	12–16 GeV	16–20 GeV	20–30 GeV	Total
Run2022	1.05 ± 0.03	0.98 ± 0.04	1.11 ± 0.09	1.00 ± 0.15	1.38 ± 0.46	1.06 ± 0.07
Run2023	1.09 ± 0.04	0.95 ± 0.04	1.04 ± 0.13	1.08 ± 0.17	1.27 ± 0.34	1.03 ± 0.07

Table 4.5: Muon fake rate ratios for Data over MC simulation for 2022 and 2023 in the bins of meson p_T .

However, the normalization channel approach used in this analysis means that in all calculations needed for the final result, we only care about a *ratio* of efficiencies. One notices that if the ϵ_{corr} is the same for both the normalization and signal channels, it cancels when considering a ratio between normalization and signal channel efficiencies. In fact, we notice that except for trigger selection and MVA_D selection, the selections are kept exactly the same between signal and normalization channels. Since the MVA_D and trigger selections are independent of the other selections and independent of each other, we can let the other selection efficiency corrections cancel and only consider corrections to MVA_D selection and trigger selection independently. The calculation of the MVA_D correction, denoted $\text{MVA}_{D,\text{corr}}$ has already been discussed extensively in Section 4.3.3. Therefore, all that is left to consider is the trigger efficiency correction for both the `HLT_ZeroBias` trigger and the `HLT_DoubleMu4_3_LowMass` trigger.

4.6.1 Trigger Efficiency Correction

Trigger efficiency quantifies how often the trigger fires on events that the trigger is built to capture. For example, the efficiency of the `HLT_DoubleMu4_3_LowMass` trigger can be calculated as the percentage of low-mass dimuon events that are successfully captured by the trigger. As discussed in Section 4.6, the trigger efficiency correction is the ratio of the trigger efficiency in data over the trigger efficiency in MC, used to account for MC mismodeling effects. For our analysis, the trigger efficiency correction of both the `HLT_ZeroBias` trigger and the `HLT_DoubleMu4_3_LowMass` trigger must be found. However, by definition, the `HLT_ZeroBias` trigger selects events with no bias, therefore it has 100% efficiency in both data and MC, causing its trigger efficiency correction to be 1. Therefore, in this analysis we only calculate the trigger efficiency correction of the `HLT_DoubleMu4_3_LowMass` trigger and label it T_{cor} . As is covered in Section 3.3, the triggers are split up into HLT and L1 triggering systems. Therefore, it is convenient for us to calculate their efficiency corrections separately and then combine them to get $T_{\text{cor}} = T_{\text{cor}, \text{L1}} \times T_{\text{cor}, \text{HLT}}$.

Calculating trigger efficiency in MC samples is done by simply calculating how many of the generated events that fit the trigger criterion actually fired the trigger. In data, we do not know how many events there actually were that fit a certain criterion, making this straight-forward calculation impossible. Instead, in data we calculate the trigger efficiency of one trigger with respect to another which we know has virtually 100% efficiency. Specifically, we calculate what percentage of the time the trigger of interest fired for all events where the reference trigger has also fired. For consistency, then, the MC trigger efficiency is calculated using this same method, instead of the straight forward method initially described. Note that when calculating trigger efficiencies, one must carefully take into account trigger

prescales, making sure to equate the scaling of both the trigger of interest and reference trigger. Additionally, since the triggers were changed slightly between 2022 and 2023, we calculate the trigger efficiency corrections in each year separately and then combine to get a total efficiency correction.

We begin with calculating the L1 trigger efficiency correction. The `HLT_DoubleMu4_3_LowMass` trigger has two L1 triggers it uses for selecting muons, named `HLT_Mu4_L1DoubleMu` and `HLT_Mu0_L1DoubleMu`. For the L1 trigger efficiency correction calculation, we pick using `HLT_Mu0_L1DoubleMu` and use the other trigger to find the HLT trigger efficiency correction. For the reference trigger, we use `HLT_Mu3_PFJet40` and `HLT_Mu8`, both intended for L1 trigger efficiency studies. These triggers both have a single muon efficiency of 0.8714, meaning that the dimuon efficiency is calculated to be $1 - (1 - \sqrt{0.8714})^2 = 0.996 \approx 1$.

Next, we calculate the HLT trigger efficiency correction, or the fraction of events that pass `HLT_DoubleMu4_3_LowMass` given that they've passed one of the L1 triggers, `HLT_Mu4_L1DoubleMu`. By definition, this means that we can use the `HLT_DoubleMu4_3_LowMass` trigger with respect to the `HLT_Mu4_L1DoubleMu` trigger to calculate the HLT trigger efficiency.

The results of these calculations are summarized in Table 4.6. Putting these results together, the correction derived for the L1 trigger is 0.990 ± 0.005 and the correction derived for the HLT trigger is 0.953 ± 0.005 . Put together, this results in a total corrective factor of $T_{\text{cor}} = 0.943 \pm 0.007$.

Measurement	Efficiency (%)		Ratio
	Data	MC	
2022			
<code>HLT_DoubleMu4_3_LowMass wrt HLT_Mu4_L1DoubleMu</code>	90.08 ± 0.15	94.91 ± 0.09	0.949 ± 0.002
<code>HLT_Mu0_L1DoubleMu wrt HLT_Mu3_PFJet40</code>	88.61 ± 0.60	88.72 ± 0.43	0.999 ± 0.008
<code>HLT_Mu0_L1DoubleMu wrt HLT_Mu8</code>	90.51 ± 0.25	91.39 ± 0.18	0.990 ± 0.003
2023			
<code>HLT_DoubleMu4_3_LowMass wrt HLT_Mu4_L1DoubleMu</code>	90.76 ± 0.15	94.91 ± 0.09	0.956 ± 0.002
<code>HLT_Mu0_L1DoubleMu wrt HLT_Mu3_PFJet40</code>	87.88 ± 0.67	88.72 ± 0.43	0.990 ± 0.009
<code>HLT_Mu0_L1DoubleMu wrt HLT_Mu8</code>	90.10 ± 0.20	91.39 ± 0.18	0.986 ± 0.003

Table 4.6: Trigger efficiency measurements in Data and MC for 2022 and 2023.

4.7 Results

4.7.1 The CL_s Method

The final result of this thesis is presented as a CL_s limit. Therefore, before discussing the results of the thesis I will describe briefly what the CL_s limit is, why it is used, and how to calculate it.

Modern statistics utilizes two main frameworks for interpreting search results. The first of these frameworks is a Bayesian framework, developed by calculating posteriors using likelihoods generated from data and inferred priors. However, defining priors over complex or untested hypothesis proves difficult and becomes very subjective to the method in which researchers assign priors. Therefore, for many applications, such as this one, the Bayesian approach is incomplete. The second of these frameworks is the frequentist approach, which draws conclusions on the compatibility of data with a given theory. However, in nearly all physical applications, researchers aim to generate conclusions on theory given data. Due to the convergence of Bayesian and frequentist conclusions in experiments with large signals and high backgrounds, researchers often allow themselves to draw conclusions on theory with a frequentist approach. However, this method will not hold here due to the incredibly small signal we are searching for.

To figure out what technology to use instead of these methods, we analyze carefully what goes wrong with a standard confidence interval test. The fundamental problem of these tests is that without proper prior weighting of hypothesis, one rejects hypothesis which the experiment had no sensitivity to. More specifically, confidence intervals can tell you about the validity of a hypothesis without considering how likely it is that the hypothesis can be proven wrong. This is what the CL_s method fixes; it takes the p-value of the signal+background model used in confidence intervals and divides it by the one minus p-value of the background. Therefore, if the model for the signal+background hypothesis is close to the model for the background-only hypothesis, the CL_s method imposes a penalty due to the lack of sensitivity in the experiment.

The fundamental building block for the CL_s method is the likelihood ratio test statistic,

$$Q = -2 \ln \left(\frac{\mathcal{L}(s+b)}{\mathcal{L}(b)} \right) \quad (4.13)$$

where $\mathcal{L}(s+b)$ is the likelihood calculated by the UML fit described in Section 4.4 and $\mathcal{L}(b)$ is the likelihood calculated using the same fit but excluding the signal model. This means that small Q values correspond to fits that indicate a stronger signal+background fit than background-only fit.

Now, recall from Section 4.4 that the likelihood is dependent on the nuisance parameters as well as the parameter of interest. Therefore, we adjust our notation briefly to reflect this, writing $\mathcal{L}_{(b)}(\vec{\theta}, \vec{\theta}_N)$ and $\mathcal{L}_{(s+b)}(\vec{\theta}, \vec{\theta}_N)$ to represent the likelihoods. In this experiment, like many others, there are many nuisance parameters that complicate the shape of the likelihood, increase computational complexity, and cause us to track an unmanageable amount of uncertainties. To avoid these issues, we introduce the profile likelihood, $\mathcal{L}_p(\vec{\theta}) = \max_{\vec{\theta}_N} \mathcal{L}(\vec{\theta}, \vec{\theta}_N)$. Note that this is not a real likelihood due to the fact that it does not reflect uncertainties in nuisance parameters. However, the test statistic Q_p obtained from the profile likelihood ratio is equivalent to the full likelihood ratio test statistic. For this reason, in virtually every analysis (including this one) the profile likelihood is used. Additionally, due to the lack of impact on the validity of the test statistic, we drop that \mathcal{L}_p and Q_p notation for the remainder of the discussions in this thesis. However, it is important to remember that we are always using profile likelihoods instead of true likelihoods in this analysis.

Now that we have generated the basic formalism needed to discuss CL_s methods, there are two main applications of CL_s that are used in every experiment. These applications are

(i) scanning analysis variables for the best expected limit and (ii) finding the bound of the parameter of interest based on unblinded data. Both these applications are discussed below and later applied to the analysis in Section 4.7.2 and Section 4.7.3.

Limit Calculation Using Real Data

In many CL_s experiments, the goal is to put a limit on a parameter of a physical theory using observed data. This is typically done once the entire analysis has been frozen and only the unblinding is left. Before starting the CL_s method, one picks the confidence level that you would like to have, denoted CL . Common CL values are 90% and 95%.

To begin, n toy experiments are generated under a background-only hypothesis and for each of these toys, a Q value is generated, giving a distribution of n , Q_b values. In the case of this analysis, this is straight forward since our background model is in the form of a PDF as described in Section 4.4. Therefore, one simply samples from the background PDF to generate background-only toy data.

Once the background-only test statistic distribution is complete, one generates a singular Q_{obs} value based on the actual, observed data. The area under the curve of the b distribution to the left of Q_{obs} is then calculated and is the p_b value.

Next, n more toy experiments are generated under a signal+background hypothesis, where the signal strength, s , is left floating. These toy experiments are constructed using a similar method to the backgrounds, just with the signal+background model for some signal strength, s . We denote this distribution of n points $Q_{s+b}(s)$, highlighting its dependence on s , the signal strength. From this, the area under the curve of $s + b$ distribution to the right of Q_{obs} is the p-value (or $p_{s+b}(s)$ value, again denoting the dependence on s).

Then, the $CL_s(s)$ value is determined by calculating

$$CL_s(s) = \frac{p_{s+b}(s)}{1 - p_b} \quad (4.14)$$

Lastly, the value of s is found such that $CL_s(s) = 1 - CL$. It may seem computationally difficult to do this naively by scanning over s values and generating n toys for each value of s . While optimizations exists in certain cases, employing CL_s methods is in practice computationally challenging. This value of s is then reported as the upper limit. An example of this is drawn in Figure 4.19.

Sensitivity Scanning Using Expected Limits

In many CL_s experiments (including this one), one is interested in tuning a parameter, α , in the analysis to get the best experimental outcome. However, this cannot be done directly once the data is unblinded, since this would introduce bias by allowing the observed data to influence the optimization, leading to overfitting. Instead, sensitivity scans are performed using expected limits. To begin, one picks several discrete values of parameter α to test, labeled α_i . As before, one additionally picks desired confidence level.

For simplicity, let's assume, for now, that we are not interested in keeping our data blinded during the sensitivity scan and directly have access to the data and therefore Q_{obs} . In this case, the sensitivity scan becomes a straight forward extension of the previous section.

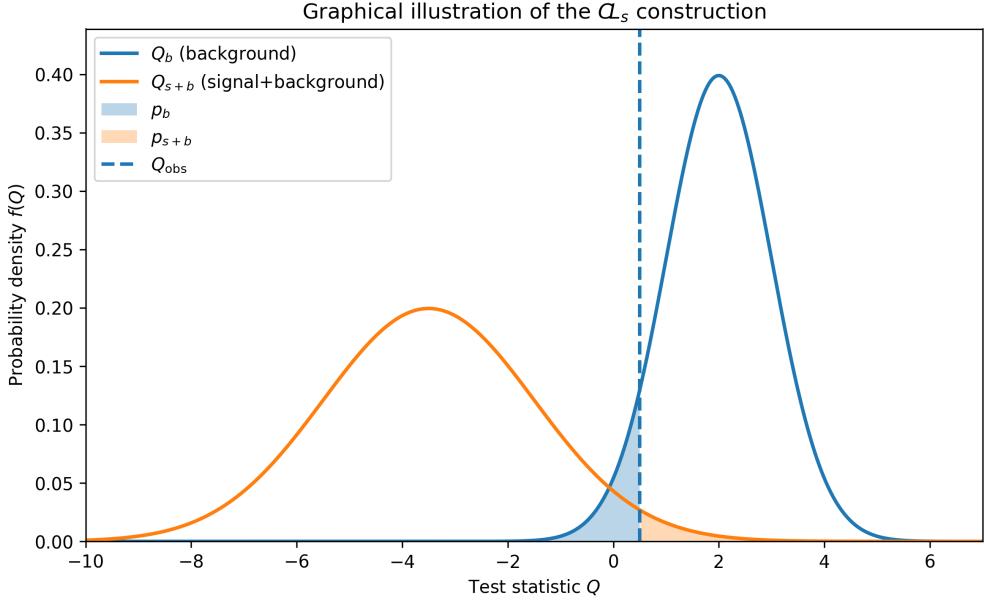


Figure 4.19: A visualization of the CL_s method, using idealized data and a given signal strength, s .

Namely, for each α_i , distributions of Q_b and $Q_{s+b}(s)$ are generated, the value of Q_{obs} is obtained given α_i , the $CL_s(s)$ value is calculated, and the value of s is found such that $CL_s(s) = 1 - CL$. Then, the specific α_i is picked that generates the most sensitive s .

In the case where we do not have access to the real data, we must compute an expected result given simulated data under a given hypothesis. It is common for searches to use background-only hypothesis for this, given that there is typically much evidence for the background-only hypothesis, though more complex choices are possible. One may be concerned about the freedom of choice for a reference hypothesis. However, recall that this doesn't affect the final limit calculation, just the choice of α that generates the best model sensitivity. If one picks the "wrong" reference hypothesis, one simply finds a worse choice of α . In practice, however, most reasonable choices work here to find a good value of α . There are then two options for computing $Q_{\text{pseudo-observed}}$ to be used instead of Q_{obs} . The first is to again generate toys under the hypothesis of choice to get a distribution of Q values, then take the median of that distribution to get $Q_{\text{pseudo-observed}}$. The second is to generate an Asimov dataset, by simply generating a "perfect" dataset with no statistical fluctuations on the background only model and computing the Q value of this "perfect" dataset. Importantly, this is much cheaper computationally, but requires the conclusion that the median of the toy results is equivalent to the toy result of the median, which requires a Gaussian approximation. In practice, this approximation is typically valid (as is the case in this analysis), so the Asimov dataset approach is used.

There is one more subtlety that is worth addressing to complete the CL_s picture. Wilks' theorem states that the profile likelihood ratio, Q , distributes asymptotically as a χ^2 distribution with degrees of freedom equal to the difference in the number of floating variables in

the signal+background model compared to the background-only model. This implies that, under the asymptotic and null assumption, one doesn't need to generate distributions of Q_b using toys. Instead, one can always use Wilks' theorem to derive the shape of the Q_b distribution from a χ^2 distribution. In practice, this approximation is also well-founded and used in most sensitivity scans, dramatically reducing the number of toys needed.

4.7.2 Sensitivity Scan

Now that the statistical formulation for the final result is complete, we turn back to the calculation of the final result.

As mentioned in Section 4.3.3, one of the major benefits of using an MVA_D cut parameter is that we can tune it to best improve analysis sensitivity. This is done using the technology developed in Section 4.7.1 with two important distinctions to highlight.

Firstly, one might assume that the parameter of interest is the output of the signal channel UML fit, $N_{D^0 \rightarrow \mu\mu}$. However, this is not quite true as the parameter that we wish to put a limit on is the actual branching fraction $\mathcal{B}(D^0 \rightarrow \mu\mu)$. The discussion in Section 4.4.2 is extended to the calculation of $\mathcal{B}(D^0 \rightarrow \mu\mu)$, resulting in

$$\frac{N_{D^0 \rightarrow \mu\mu}}{N_{D^0 \rightarrow \pi^\pm \pi^\pm}} = \frac{\mathcal{B}(D^0 \rightarrow \mu\mu)}{\mathcal{B}(D^0 \rightarrow \pi^\pm \pi^\pm)} \times \frac{\epsilon_{D^{*\pm}, D^0 \rightarrow \mu\mu}}{\epsilon_{D^{*\pm}, D^0 \rightarrow \pi\pi}} \times S_{ZB} \times MVA_D \times T_{corr} \quad (4.15)$$

where the variables are the same as defined in Equation 4.11. Therefore, to perform the sensitivity scan we do not just calculate $N_{D^0 \rightarrow \mu\mu}$ under different toy experiments, but rather perform the entire calculation required to derive $\mathcal{B}(D^0 \rightarrow \mu\mu)$, including the calculation of ϵ_X and $N_{D^0 \rightarrow \pi^\pm \pi^\pm}$ at various MVA_D working points. Luckily, there are some parameters that do not change, such as T_{corr} or S_{ZB} , so some computation is saved.

The other important distinction is that under the `HLT_ZeroBias` trigger, there are simply too few data points to produce an effective sensitivity scan result. Therefore, to avoid being biased by data fluctuations that prevail when signal events are rare, we include five additional zero bias data. The five triggers corresponding to the datasets are:

1. `HLT_ZeroBias_FirstBXAfterTrain`,
2. `HLT_ZeroBias_FirstCollisionAfterAbortGap`,
3. `HLT_ZeroBias_FirstCollisionInTrain`,
4. `HLT_ZeroBias_IsolatedBunches`, and
5. `HLT_ZeroBias_LastCollisionInTrain`.

An important note is that these additional datasets are not used in the calculation of the final limit, due to an incomplete understanding of how their additional requirements could affect the analysis. However, the effects are assumed to be small enough that they still provide valuable insight to determine what MVA_D working point to use.

The results of this scan are shown in Figure 4.20. The best MVA_D cut values are in the range 0.74 to 0.84. In order to include as many signal events as possible, we choose the loosest MVA_D cut in this range at 0.74 for the final value to be used to maximize analysis sensitivity.

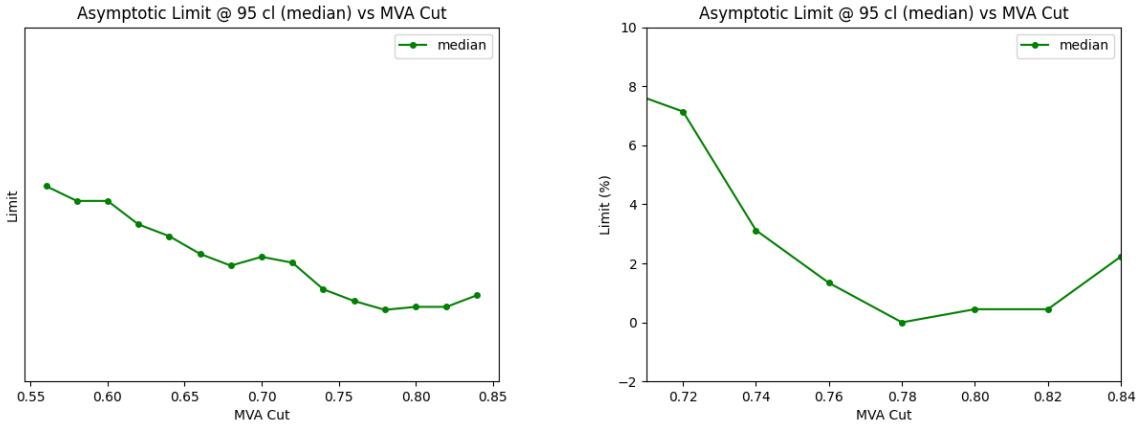


Figure 4.20: The sensitivity scan against various MVA cuts first within a broad range of MVA cut values (left) and then with a more concentrated range of MVA cut values (right).

4.7.3 Branching Fraction Limit

Finally, we present the final result. After applying the formalism developed in Section 4.7.1 and unblinding the data, we find that the upper limit on the branching fraction at a 90(95)% confidence level is

$$\mathcal{B}(D^0 \rightarrow \mu\mu) < 2.1(2.4) \times 10^{-9} \quad (4.16)$$

Table 4.7 shows the central values of the observed yields for each decay channel studied in this analysis and Figure 4.21 displays the final fit on the observed data.

Observed Yields	Values
$N_{D^0 \rightarrow \mu\mu}$	139 ± 123
$N_{D^0 \rightarrow \pi^\pm \pi^\pm}$	220 ± 58
$N_{D^0 \rightarrow \pi^\pm \mu\nu_\mu}$	207 ± 40
N_{comb}	126185 ± 366
N_{total}	126752

Table 4.7: The final observed yield, based on data.

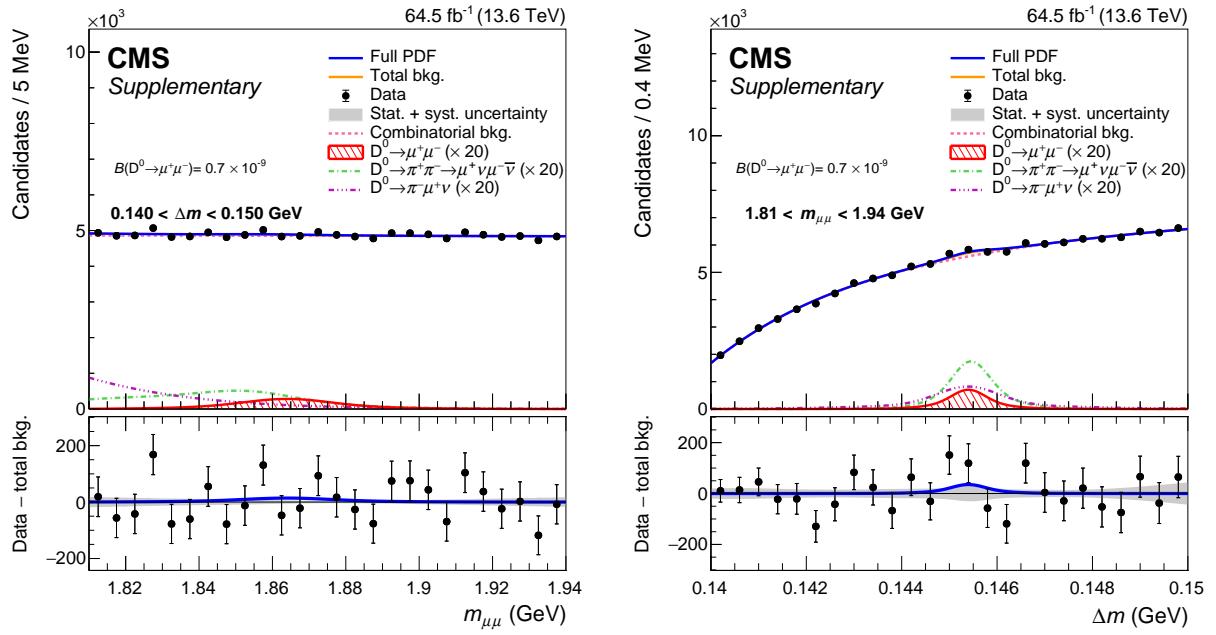


Figure 4.21: The unblind fit on observed data shown in $m(D^0)$ (left) and Δm (right). Note that the peaking decays are visually scaled by a factor of 20.

Chapter 5

Conclusion

A search for rare charm decays is presented in this thesis. These rare charm decays are examples of FCNCs, which serve as excellent probes into BSM physics models due to their suppressed SM nature. Previous FCNC studies often omit charm decays and focus on strange or bottom decays. Specifically, this thesis calculates a limit on the branching fraction of the $D^0 \rightarrow \mu^+ \mu^-$ decay, denoted $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-)$.

The analysis in this thesis has two major challenges. The first is the large amount of combinatorial background compared to the small amount of signal. The second is the presence of muon fakes, causing pions to be misreconstructed as muons and faking $D^0 \rightarrow \mu^+ \mu^-$ events. To address these obstacles, this thesis focuses on D^0 mesons that decay from D^* mesons. This allows for a more robust reconstruction of the signal events, which keeps combinatorial backgrounds low. Additionally, this thesis leverages the well-known quantity $\mathcal{B}(D^0 \rightarrow \pi^+ \pi^-)$ to use the $D^0 \rightarrow \pi^+ \pi^-$ decay as a normalization channel. This allows for the ability to calculate $\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-)$ without actually knowing how many D^0 mesons are produced. Furthermore, this normalization channel causes a reduction in many systematic uncertainties, allowing for a more precise measurement. The primary data used in this analysis consist of proton-proton collision data samples taken during 2022 and 2023 using the CMS detector at the LHC. One sample is constructed using a prescaled zero-bias trigger while the other is collected using a dimuon trigger with a center-of-mass energy at $\sqrt{s} = 13.6$ TeV and corresponding to an integrated luminosity of 51.4 nb^{-1} and 64.5 fb^{-1} , respectively.

No obvious excess is observed. The final upper limit on the branching fraction at a 90(95)% confidence level is found to be

$$\mathcal{B}(D^0 \rightarrow \mu^+ \mu^-) < 2.1(2.4) \times 10^{-9} \quad (5.1)$$

This upper limit outperforms the previous world-best limit calculated by LHCb at a 90% confidence level of 3.1×10^{-9} [2], placing new constraints on BSM physics.

In addition, this analysis makes other key contributions in the process of deriving its final result. Namely, this analysis performs one of the most exhaustive fake-rate studies to date, presenting a result and continuing development on a method which will be able to be used by many other analyses that involve muon fake rates. This analysis is also the first to use the new inclusive dimuon trigger developed at CMS. It demonstrates the benefits of the new trigger, illuminating the opportunities for its use in low-mass dimuon data.

References

- [1] For the duration of this text, charge conjugates are always implied unless otherwise stated.
- [2] R. Aaij and A. S. W. and Abdelmotteleb. Search for rare decays of d0 mesons into two muons. *Physical Review Letters*, 131(4), July 2023. ISSN 1079-7114. doi:[10.1103/physrevlett.131.041804](https://doi.org/10.1103/physrevlett.131.041804). URL <http://dx.doi.org/10.1103/PhysRevLett.131.041804>.
- [3] Gustavo Burdman, Eugene Golowich, JoAnne Hewett, and Sandip Pakvasa. Rare charm decays in the standard model and beyond. *Physical Review D*, 66(1), July 2002. ISSN 1089-4918. doi:[10.1103/physrevd.66.014009](https://doi.org/10.1103/physrevd.66.014009). URL <http://dx.doi.org/10.1103/PhysRevD.66.014009>.
- [4] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pages 785–794, New York, NY, USA, 2016. ACM. doi:[10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785). URL <https://doi.org/10.1145/2939672.2939785>.
- [5] Atlas Collaboration et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *arXiv preprint arXiv:1207.7214*, 2012.
- [6] CMS Collaboration. Measurement of the $b^0 \rightarrow \mu^+ \mu^-$ decay properties and search for the $b_s^0 \rightarrow \mu^+ \mu^-$ decay in proton-proton collisions at $\sqrt{s} = 13$ tev. *Phys. Lett. B*, 842: 137955, 2023. doi:[10.1016/j.physletb.2023.137955](https://doi.org/10.1016/j.physletb.2023.137955).
- [7] Cms Collaboration et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *arXiv preprint arXiv:1207.7235*, 2012.
- [8] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes: the discrete profiling method. *Journal of Instrumentation*, 10: P04015, 2015. doi:[10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015). URL <https://doi.org/10.1088/1748-0221/10/04/P04015>.
- [9] Huaiyu Duan, George M Fuller, and Yong-Zhong Qian. Collective neutrino oscillations. *Annual Review of Nuclear and Particle Science*, 60(1):569–594, 2010.
- [10] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse. Measurement of the electron magnetic moment. *Physical Review Letters*, 130(7), February 2023. ISSN 1079-7114. doi:[10.1103/physrevlett.130.071801](https://doi.org/10.1103/physrevlett.130.071801). URL <http://dx.doi.org/10.1103/PhysRevLett.130.071801>.
- [11] Daniel Harlow. Quantum field theory i and ii. MIT Lecture Notes, 2024. Spring 2024.

- [12] Ewa Lopienska. The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022. 2022. URL <https://cds.cern.ch/record/2800984>. General Photo.
- [13] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024. doi:[10.1103/PhysRevD.110.030001](https://doi.org/10.1103/PhysRevD.110.030001). URL <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>.
- [14] K. Prokofiev and T. Speer. A kinematic and a decay chain reconstruction library. In *Proceedings of the International Conference on Computing in High Energy Physics (CHEP '05)*, page 411. CERN, 2005. doi:[10.5170/CERN-2005-002.411](https://doi.org/10.5170/CERN-2005-002.411). URL <https://cds.cern.ch/record/865614>.
- [15] Tai Sakuma. Cutaway diagrams of CMS detector. 2019. URL <https://cds.cern.ch/record/2665537>.
- [16] etc. Sirunyan, A.M. Performance of the cms muon detector and muon reconstruction with proton-proton collisions at 13 tev. *Journal of Instrumentation*, 13(06):P06015–P06015, June 2018. ISSN 1748-0221. doi:[10.1088/1748-0221/13/06/p06015](https://doi.org/10.1088/1748-0221/13/06/p06015). URL <http://dx.doi.org/10.1088/1748-0221/13/06/P06015>.
- [17] W. Verkerke and D. P. Kirkby. The roofit toolkit for data modeling. In *eConf C0303241*, page MOLT007, 2003.