USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

Miguel
Fernández Gómez

Tese de doutoramento

The light meson guide to New Physics: Latest searches of rare decays using LHCb data

Santiago de Compostela, 2025

Programa de Doutoramento en Física Nuclear e de Partículas

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

TESIS DOCTORAL

THE LIGHT MESON GUIDE TO NEW PHYSICS: LATEST SEARCHES OF RARE DECAYS USING LHCb DATA

Autor

Miguel Fernández Gómez

Director: Xabier Cid Vidal

Diego Martínez Santos

Tutor/a: Xabier Cid Vidal

PROGRAMA DE DOCTORADO EN FÍSICA NUCLEAR Y DE PARTÍCULAS

SANTIAGO DE COMPOSTELA / LUGO

Here's to the fools who dream.

Abstract

This thesis presents the first-ever study of the decay of a neutral kaon into four muons, $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$. Using 5.1 fb^{-1} of data collected by the LHCb experiment between 2016 and 2018, no signal was found. Upper limits of $\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}$ and $\mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}$ were set. These results also represent, respectively, the first time a sensitivity of the order of 10^{-12} is reached by a Large Hadron Collider experiment, and the first result of a K^0_L mode at the LHC.

Additionally, the study of four decays of η and η' mesons are presented, using the same set of LHCb data. The thesis introduces the first-ever experimental study of $\eta' \to \mu^+\mu^-$, with an expected significance of 2.85σ as well as new studies of the branching fractions of $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \mu^+\mu^-$. Respectively, the new uncertainties on the branching fractions are expected to be 9.6% and 5.3%. The mode $\eta \to \pi^+\pi^-\mu^+\mu^-$ was also studied in this thesis for the first time at LHCb.

Limiar

Esta tese presenta o primeiro estudo da desintegración dun kaón neutro en catro muóns, $K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-$. Usando 5.1 fb^{-1} de datos recollidos polo experimento LHCb entre 2016 e 2018, non se atopou ningún sinal. No seu lugar, fixáronse cotas superiores de $\mathcal{B}(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}$ e $\mathcal{B}(K^0_{\mathrm{L}} \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}$. Estes resultados also representan, respectivamente, a primeira vez que se acada unha sensitividade da orde de 10^{-12} nalgún experimento do Gran Colisionador de Hadróns, así como o primeiro resultado da desintegración dun mesón K^0_{L} no mesmo colisionador.

Ademais, tamén se presenta o estudo de catro desintegracións de mesóns η e η' usando o mesmo conxunto de datos recollido no LHCb. Esta tese introduce o primeiro estudo experimental da desintegración $\eta' \to \mu^+\mu^-$, cunha significancia esperada de 2.85σ, así novos estudos das fraccións de desintegración dos modos $\eta' \to \pi^+\pi^-\mu^+\mu^-$ e $\eta \to \mu^+\mu^-$. Respectivamente, as incertezas medidas das fraccións de desintegración espéranse que sexan 9.6% e 5.3%. O modo $\eta \to \pi^+\pi^-\mu^+\mu^-$ tamén se estudou nesta tese con datos do LHCb por primeira vez.

Resumen

Esta tesis presenta el primer estudio de la desintegración de un kaón neutro en cuatro muones, $K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-$. Usando 5.1 fb^{-1} de datos recogidos por el experimento LHCb entre 2016 y 2018, no se encontró ninguna señal. En su lugar, se fijaron cotas superiores de $\mathcal{B}(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}$ y $\mathcal{B}(K^0_{\mathrm{L}} \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}$. Estos resultados también representan, respectivamente, la primera vez que se alcanza una sensitividad del orden de 10^{-12} en cualquiera de los experimentos del Gran Colisionador de Hadrones, así como el primer resultado de la desintegración de un mesón K^0_{L} en el mismo colisionador.

Además, también se presenta un estudio de cuatro desintegraciones de mesones η y η' usando el mismo conjunto de datos recogido por el LHCb. Esta tesis introduce el primer estudio experimental de la desintegración $\eta' \to \mu^+\mu^-$, con una significancia esperada de 2.85σ, así nuevos estudios de las fracciones de desintegración de los modos $\eta' \to \pi^+\pi^-\mu^+\mu^-$ y $\eta \to \mu^+\mu^-$ que anticipan una mejora de sus precisiones experimentales. Respectivamente, se espera que las incertidumbres finales de estas fracciones de desintegración sean 9.6% y 5.3%. El modo $\eta \to \pi^+\pi^-\mu^+\mu^-$ también se ha estudiado en esta tesis por primera vez en el LHCb.

Acknowledgements

La lista de personas a las que agradecer que estos últimos tres años y medio hayan sido tan remarcables como lo han sido, junto con lo que me gustaría decir sobre cada una, probablemente ocuparía tanto como toda la tesis que hay a continuación. El cómo ordenarlas es quizás tan complejo como el contenido científico de ésta, por lo que no debe tomarse de forma literal.

Aún así, no podría sino empezar por agradecer a Xabier y a Diego por la infinita paciencia y la dedicación que han puesto en cada paso de mi trayectoria y producción científica, parte de la cual quedará documentada a continuación. Gracias por haberme acogido y apoyado, y respondido a todas las preguntas estúpidas que uno podría haber hecho en casi cuatro años. Espero que el resultado haya cumplido mínimamente las expectativas.

El ambiente de trabajo que han creado estos años, junto con el que ya tenían Alexandre, Adrián y Saúl (quienes también fueron una importante guía en varios momentos) permitieron que esta tesis se desarrollase de forma tan fluida y con, relativamente, tan pocos tropiezos. Especial gracias a Emilio, por haber seguido cultivando ese ambiente conmigo y por todos los momentos que compartimos según nos aventurábamos en este mundo juntos. Y a todos los que llegaron posteriormente, especialmente a María, Pablo y Cristina, por el aire fresco que han traído.

En cuanto a apoyo, no ha habido ninguno más importante que el de mis padres y mi hermana, que han soportado con capa y espada los momentos más estresantes de estos últimos años, y han estado ahí para celebrar cada momento.

Cuando miro hacia atrás y pienso en este tiempo, no me cabe ninguna duda de que la persona que más me alegro de que haya entrado en mi vida es Verónica, que a diario durante más de dos años, y aún hoy en la distancia continúa a ser una fuente de inspiración y apreciación por cada momento. Gracias por todo lo que compartimos y por seguir siendo así.

Volveré a la gente que ha entrado recientemente, pero no podría sino mencionar a aquellos que ya llevan conmigo desde segundo de carrera. A Doval, quien posiblemente tenga el récord mundial de aguantar mis tonterías, mis obsesiones infundadas e inexplicables, y mi falta de conocimiento absoluto sobre algunas cosas por las que le resultó imposible no reírse; y aún así, seguir aquí conmigo. Igualmente, gracias a Martina por seguir a mi lado de forma tan fiel después de tantos años, y por hacer que cada día a su lado sea inolvidable de una forma distinta. A Lucía, por ser un ejemplo de persistencia y de buscarle el lado positivo a cada momento, con una sonrisa en la cara y preparada para afrontar otro día más. A Sara, cuyos logros en estos últimos años me dan más ganas de seguir superándome cada día. A Xabi (García), por hacer que, aunque pasen los años, es como si no hubiera pasado un día entre cada vez que nos vemos. Y a todos ellos en conjunto por seguir queriendo hacer clubitos conmigo. (Queda aquí constancia de que me deben un viaje a Bora Bora desde 2018.)

Gracias a Clara, por, ocho años después de nuestra primera conversación, seguir siendo tan importante en mi vida, y por la alegría que esto me produce. Gracias a Carlos, no sólo por ser una pozo infinito de sabiduría sino también por ser el amigo que todos deberíamos aspirar a ser. A Andrea (Iglesias), por su insistencia en continuar formando parte de mi vida y por ser una fuente de desconexión en cada uno de los pocos ratos en que nos vemos. Y no menos importante, a Dani, porque aún estando a miles de kilómetros, encuentra el momento para mantenerme en su día a día, demostrando que hay amistades que no se deben perder.

Prácticamente todo lo dicho anteriormente podría aplicarse a Blanca, cuya presencia en mi vida sigo considerando un regalo. Por todo lo que hemos compartido, por todo su apoyo y comprensión, y por ser una de esas personas con las que simplemente estoy cómodo siendo yo mismo, mil gracias.

Siguiendo con las personas que llegaron a mi vida estos últimos años, gracias a Arantxa por aún ser capaz de meterme alegría en el cuerpo cada vez que me acuerdo de ella. Gracias a Victoria, por lo imprescindible que se ha convertido para mí en este último año, y por la paciencia y comprensión que ha tenido en tantas ocasiones ya. A David y a Bea, por alegrarme los días de formas únicas pero igualmente importantes. A Ana, a Victor y a Georgina, por cada momento juntos, y en especial por todos los recuerdos compartidos en París y Alemania.

Me he dejado a Iván, por supuesto, porque él fue uno de los principales artífices detrás de los tres meses inolvidables que pasamos en CERN. En casi igual medida, al grupo de la UB, y en especial a Albert, por ser una de esas personas en las que sé que puedo contar siempre. A Núria, por todos los vasos de papel que mutiló mientras descansábamos el cerebro después de cada comida. Y a Sergio, no sólo por ser un buen perdedor en un concurso de comer brownie de chocolate, ni por inspirarme cada día a trabajar aún más duro, sino por todas las conversaciones tan personales que hemos compartido en este último año.

Uno de los momentos más especiales de estos tres años y medio sucedió en 2022 al recibir la felicitación de cumpleaños más inesperada. Gracias a Uxía por ello y por haber vuelto a mi vida. En esta línea, gracias a Jorge, por ser la persona de toda esta lista que lleva aguantándome más tiempo fuera de mi familia, y por todos los años que nos quedan. Y gracias a Laura, por creer siempre en mí.

Y hablando de aguantarme durante años, gracias a Andrea (García), que aun no se ha dado cuenta que es la única persona por la que lo dejaría todo si lo pidiera.

Switching now into English, I would like to start by saying that there is no other single person responsible for how good this thesis turned out than Titus (the less impressive parts are all my fault), possibly one of the most giving and remarkable people I have met (and also one of the most intrinsically funny). I thank him for every minute of his time he sacrificed to make my various research projects so much better they could have ever been, and making me a better physicist than I could have ever hoped. (To illustrate the point, he even came up with the title of the thesis.)

Speaking of Germans, I also thank Felicia for providing me with some of my favorite memories of this past year and for her phone calls becoming an indispensible part of my week. For her understanding and her trust, and for lighting my life with a ray of sunshine.

I thank every member of the Collaboration who took part in my research, whether it was actively partaking in it, consulting, or reviewing the work. Most notably, of course, the Kaon dream team at Lausanne, with Luismi and Rado, who have become a deep source of knowledge and learning, but also have made our meetings and the days working together so much easier more fun than I could have hoped for.

Y por último y más importante, a Raquel, porque combina todo lo anteriormente dicho y lo eleva para convertirse en la mejor compañera de vida. Ella ha estado presente desde el primer minuto hasta el último, apoyándome incondicionalmente y sacrificando casi tanto como yo para llegar aquí.

Gracias a Arnau, que había aceptado informalmente estar en mi tribunal de tesis más de un año antes de terminarla, y meses antes de dejarnos un hueco en el corazón.

Contents

1

An introduction to particle physics

It is said that the Scottish physicist William Thomson, Lord Kelvin, claimed in 1897 that “There is nothing new to be discovered in physics now. All that remains is more and more precise measurement.” [1] A few years later, Albert Einstein would turn everything upside down with the Theory of Special Relativity [2] (which had already been somewhat conceptualized by earlier works in the field of electromagnetism [3, 4]), and even more so, with the publication of the Theory of General Relativity in 1915 [5].

By then, it had also been proven that matter is made up of tiny elements known as atoms [6], which include a positively-charged nucleus and negatively-charged particles moving around it [7]. The physics at this scale, however, seemingly behaved differently than at human scales. The new paradigm would later be dubbed quantum mechanics [8].

Einstein was awarded the Nobel Prize in 1921 for his understanding of the photoelectric effect [9], historically one of the first times that a quantum mechanical effect was ever studied. The formulation of quantum mechanics into a mathematical theory would take no longer than two decades, and was the subject of a tight race among some of the top minds working in physics at the beginning of the 20th century [10, 11].

The nucleus contains both protons (positively charged) and neutrons (no electrical charge). The particles surrounding it are electrons (negative charge). In the early 1930s, Anderson identified a particle [12] that shared the exact same properties as the electron, but had an oppositely-signed electric charge. It was called the positron, and it represents the first time an anti-particle was observed.

The floodgates of research and discovery opened up following the end of World War II. The topic of understanding the subatomic world would fascinate physicists for decades to come, both at the theoretical and the experimental level. One of the key ingredients that make quantum mechanics so different from classical physics (an umbrella term used to refer to pre-20th century formulations) was the fact that two things can be true at the same time, as long as a measurement is not performed. This is called the superposition principle: a (quantum) cat with a 50% chance of having been poisoned is both alive and dead, as long as a medical exam is not performed.

This idea was one of the key ingredients to the early-days mathematical formulation of quantum mechanics, done by Schrödinger [10] and Heisenberg [11], among others [13–15]. It led to very precise measurements of subatomic quantities, like the energy spectrum of an electron orbiting a proton (the Hydrogen atom). Understanding the movements of particles at this level, though, would soon require an even larger framework, as they could easily acquire speeds close to that of light. This phenomenon is described at the macroscopic level by Einstein's Special Relativity.

Special Relativity and Quantum Mechanics were united to create Quantum Field Theories (QFTs) [16, 17], a set of mathematical formulations that have since led to the best agreements between theoretical predictions and experimental observations ever. They also describe the interactions between particles with a set of tools known as Feynman diagrams [17, 18]. Around the same time these were being developed, experimental particle physics was inside a period known as the "Particle Zoo" [19]. New particles were being discovered [20–23] at a higher rate than they were being understood as part of a mathematical model [24].

This theoretical framework, however, has been formulated since then. It was both informed by experimental observations [25, 26] and capable of making predictions of its own that were later proven right at physics detectors [27]. It is the theory known as the Standard Model (SM) [28].

1.1 The Standard Model of Particle Physics

The Standard Model of Particle Physics is a mathematical theory that studies the fundamental particles that make up matter and how they interact between them. (See Refs. [29–31] for extensive overviews.)

Inside the SM, particles are split into two groups. There are fermions (with half-integer spin) and bosons (integer spin), which mediate the interactions between fermions. Fermions themselves are also split into two subgroups: quarks and leptons. Both of these subgroups are organized in three families, or generations, each one heavier than the previous one.

Quarks cannot exist by themselves, they need to be paired up with an antiquark (making up a meson) or with two other quarks (a baryon). This results in the creation of a conserved quantity for particles known as color charge: by combining three colors, we achieve a color-neutral bound state, which we also reach by combining a color and an anti-color.

Particles made up of quarks are hadrons. Protons are baryons, for instance: they are composed of two up quarks, u, and one down quark, d. Quarks u and d make up the first generation. The first lepton generation includes the electron, e, and the electronic neutrino, ν_e a particle so light it hardly interacts with matter).

Subsequent generations of quarks include charm, c, and strange s quarks (second), and top, t, and bottom, b (third). The rest of the leptons are the muon, μ, and the muonic neutrino, ν_μ (second generation), and the tau, τ, with its corresponding neutrino ν_τ (third).

Therefore, there are six types, or *flavors*, of quarks, and six types (flavors) of leptons.

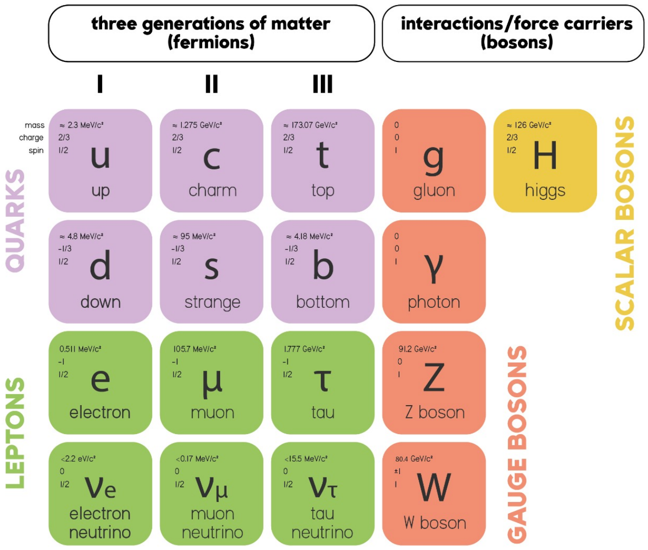


Figure 1.1: Visual representation of the elementary particles and the interacting bosons that conform the Standard Model.

The study of the properties of quarks and fermions is called *flavor physics*. A visual scheme of the Standard Model is presented in Fig. 1.1.

1.1.1 Quantum Electrodynamics

The interactions between leptons (e, μ, τ) are mediated by photons, and the QFT describing them is called Quantum Electrodynamics (QED). Fermions are subject to a local phase invariance, $\psi(x) \to e^{i\theta(x)}\psi(x)$, so by the grace of Noether's theorem [32], QED must be symmetric under these transformations. The symmetry group most appropriate seems to be $U(1) = \{e^{i\theta} | \theta \in \mathbb{R}\}$.

The strength of the electromagnetic interaction is given by the coupling constant of the theory, α, which is predicted by perturbation theory to be $\alpha = \frac{e^2}{4\pi\varepsilon_0\hbar c} \simeq \frac{1}{137}$.

The theory is characterized by the Lagrangian, a function that can yield the equations of motion of the theory and describe the interactions between the elements in it. Since the leptons are fermions, they are described by Dirac spinors and their equations of motion are given by the Dirac equation, $(i\gamma^\mu\partial_\mu - m)\psi = 0$. The Lagrangian term is easy to derive from there, $\mathcal{L}_f = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi$.

The photon is described by the gauge field A_μ, and the kinetic term associated to it is given by $\mathcal{L}_\gamma = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. However, to ensure the gauge invariance of the Lagrangian, and therefore, of the theory, we must add a term that also turns out to describe the interaction between photons and electrons, $\mathcal{L} = -e\overline{\psi}(\gamma^\mu A_\mu)\psi$. We can use this term to define a covariant derivative operator, $D_\mu = \partial_\mu + ieA_\mu$, such that

$$\mathcal{L}_{QED} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.1}$$

1.1.2 Quantum Chromodynamics

The interactions between quarks are mediated by gluons, and the QFT describing those interactions is called Quantum Chromodynamics (QCD). The name is given by the role that the color charge plays in it. To account for symmetries under local transformations, QCD must be a gauge theory and thus be described by a symmetry group. The easiest choice, to account for all three color charges[*], is $SU(3)$, which contains all unitary, 3×3 matrices with a determinant equal to 1. There are eight of them, known as the Gell-Mann matrices, which in turn yield the eight generators of the group, T_a.

In a gauge theory, the generators are associated to the mediating (or gauge) bosons. This means that in QCD there are eight possible gluons[†]. The color charge is known as a local symmetry, since it is applied to each point in spacetime independently of what happens elsewhere. The fact that there is a color-neutral state in point A does not mean there cannot be another one in point B.

The strength of interactions in a gauge theory is determined by the coupling constant, α_s. We can define the beta function, $\beta(\alpha_s)$, as the dependency of the coupling constant with the energy scale, μ:

$$\beta(\alpha_s) = \mu \frac{\partial \alpha_s}{\partial \mu}.$$

For QCD, one can compute, via one-loop Feynman diagrams [29],

$$\beta(\alpha_s) = -\left(11 - \frac{2}{3} n_f\right) \frac{\alpha_s^2}{2\pi}, \tag{1.2}$$

where n_f is the number of quark flavors at the energy scale μ. Since in QCD, $n_f \leq 16$, $\beta(\alpha_s) < 0$, which means that the coupling becomes really large at low energies, and therefore, there is no room for a perturbation theory to describe QCD. This also causes the confinement of quarks inside hadrons. As we increase the energy, or decrease the distance, however, the coupling becomes much smaller, a phenomenon known as asymptotic freedom. The boundary between both regimes is given by the fundamental scale of QCD, Λ_{QCD}, known only experimentally, $\Lambda_{QCD} \simeq 200 - 300\,\text{MeV}$ [33, 34]. (This is for $n_f = 5$, meaning up to the b mass scale.)

The missing piece in this breakdown of Quantum Chromodynamics is the introduction of the Lagrangian. Our theory constituents are the quarks, which are fermions described by a Dirac spinor, and gluons, which are bosons that can self-interact. The Lagrangian is constructed by adding together the kinetic terms of both types of particles and a third term describing how they interact.

We can also fold the interaction between quarks, described by Dirac spinors, ψ, and gluons, described by gauge G_μ^a fields, inside the quark kinetic term through the introduction

---

[*] It is important to note that the concept of color charge is an illusion, an artifice used by mathematics to describe the grouping of quarks in threes. Quarks of a specific flavor are not green or red based on that flavor; the color is given by the presence of other quarks. In reality, electric charge, which is much more ingrained in our culture, is no different. It is simply a sign that we give a particle based on how it reacts in the presence of an electric field.

[†] In general, the special unitary group $SU(n)$ has $n^2 - 1$ generators. It is easy to see why: an $n \times n$ unitary matrix has n^2 complex entries, meaning $2n^2$ real degrees of freedom. The unitary condition, $U^\dagger U = 1$, imposes n^2 constraints. On top of that, $det(U) = 1$ adds another constraint, so the number of degrees of freedom of $SU(n)$, its number of generators, is given by $2n^2 - n^2 - 1 = n^2 - 1$.

of a covariant derivative, D_μ:

$$\mathcal{L}_{quark} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi \quad , \qquad D_\mu = \partial_\mu - i g_s T^a G^a_\mu. \tag{1.3}$$

Here, $g_s = \sqrt{4\pi\alpha_s}$ is the strong coupling constant.

Similarly, the gluon term is given by

$$\mathcal{L}_{gluon} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad , \qquad F_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu. \tag{1.4}$$

Here, f^{abc} are the structure constants of $SU(3)$, the QCD analogous to the Levi-Civita symbol. With all of the above,

$$\mathcal{L}_{QCD} = \overline{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.5}$$

1.1.3 Weak interactions

All quarks and leptons are subject to the weak interaction, which rules transitions between them. In 1957, when studying the β-decay of ^{60}Co, C.S. Wu [35] found that in this decay, neither parity nor charge conjugation were conserved. Up until then, the theory governing β decays had been proposed by Fermi in 1933 [36], and it covered parity conservation.

A new theory was formulated later in 1958, known as the $V - A$ model [37,38], which also solved the other problem that Fermi's theory had: it was not renormalizable. Most of the knowledge we have of weak interactions and the theory that dictates them has been empirically accumulated.

The helicity is defined as the projection of the particle's spin onto its momentum,

$$h = \frac{\vec{S} \cdot \vec{p}}{|\vec{p}|}, \tag{1.6}$$

which means that massive particles can be boosted to invert its helicity, while massless particles must always have the same helicity state. Wu's experiment showed a significant excess of electrons with helicity -1 (left-handed) over electrons with helicity $+1$ (right-handed), indicating that the weak interaction couples to left-handed particles and right-handed anti-particles.

Following the argument above, neutrinos are always produced left-handed, while anti-neutrinos are right-handed. Consider the $\pi^- \to \ell^-\overline{\nu}$ decay. If neutrinos are always right handed, their spin is always aligned with their momentum. In the pion's center of mass, the neutrino and the lepton are produced back-to-back, so they fly away with opposite-signed momentum directions. By means of angular momentum conservation, the lepton's spin must *also* be aligned with its momentum, which means that it must also be right-handed.

Since the electron mass is much lower than the muon's, using the argument above, it will be much more complicated for them to invert their helicity state, which means that $\pi^- \to \mu^-\overline{\nu}_\mu$ is much more likely to happen than $\pi^- \to e^-\overline{\nu}_e$. The ratio of probabilities between both decays can be computed very precisely theoretically, and has been ratified experimentally.

Also experimentally we see that the fermions adopt a doublet structure. Leptons and their corresponding neutrinos are always seen together in weak decays, but processes involving an even number of leptons of different flavors have not been observed. Similarly, quark transitions also seem to favor a doublet structure:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \quad \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \tag{1.7}$$

All of the evidence above seems to point towards a $SU(2)_L$ symmetry group lying underneath the weak interactions. Furthermore, this would predict three spin–1 force mediators, which have been observed: $W^\pm$, governing charged decays [39, 40], and Z^0, mediating neutral transitions [41, 42]. The conserved quantity in weak interactions is known as isospin, T. Its projection, T_3, is $T_3 = \frac{1}{2}$ for particles in the upper side of the doublets, and $T_3 = -\frac{1}{2}$ in the lower side.

Based on the interacting bosons, weak currents can be either charged, when they are mediated by $W^\pm$, or neutral, when they are mediated by Z^0. Charged currents, in addition to being allowed to violate parity conservation, are also the only interactions permitted to change a particle's flavor.

Neutral currents, however, cannot. This means that processes like $q + \overline{q} \to \ell + \overline{\ell}$, where q is a quark and ℓ is a lepton, cannot happen at tree level inside the SM, since they would require the exchange of a spin–1 neutral boson. They are called Flavor Changing Neutral Currents (FCNCs) and they *are* allowed at loop level inside the theory, though they are highly suppressed. This is one of the main areas of research in experimental particle physics, since small deviations from the SM at such low scales would be triggered by physics beyond our current understanding. One of the decays studied in this thesis is an FCNC (see Sect. 5).

Mathematically, the generators of the theory are given by the Pauli matrices, $\vec{\tau} = (\sigma_1, \sigma_2, \sigma_3)$, while the fields of describing the gauge bosons are given by $\vec{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$, where W^+ is described by $W_\mu = \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu)$, W^- is described by $W^\dagger_\mu = \frac{1}{\sqrt{2}}(W^1_\mu - iW^2_\mu)$, and the Z boson is described by W^3_μ. The coupling constant is given by

$$\alpha_s = \frac{g_W^2}{4\pi} \simeq \frac{1}{30}, \tag{1.8}$$

where g_W can be linked to the Fermi constant that was the first weak coupling introduced,

$$\frac{G_F}{\sqrt{2}} = \frac{g_W^2}{8m_W^2},$$

and where m_W is the mass of the W boson.

The covariant derivative that enters the Lagrangian of this theory is given by

$$D_\mu = \partial_\mu - ig\vec{\tau} \cdot \vec{W}_\mu.$$

For a left-handed fermion doublet, χ_L, we have the following kinetic term:

$$\mathcal{L}_{kinetic} = \overline{\chi}_L i\gamma^\mu D_\mu \chi_L$$

With this in mind, and considering a left-handed fermion doublet given by χ_L, we can write

$$\mathcal{L}_{weak} = \overline{\chi}_L i\gamma^\mu D_\mu \chi_L - \frac{1}{4}\vec{F}_{\mu\nu}\vec{F}^{\mu\nu}, \tag{1.9}$$

where $\vec{F}_{\mu\nu} = \partial_\mu \vec{W}_\nu - \partial_\nu \vec{W}_\mu + g\vec{W}_\mu \times \vec{W}_\nu$.

It is important to note here how there is no mass term for either the fermions of the theory, i.e., a term of the form $m\overline{\chi}\chi$. The same happens for the gauge bosons. This would break gauge symmetry and yield an invalid Lagrangian. The paradox is solved once the Higgs mechanism is incorporated, which was achieved after the unification with quantum electrodynamics.

1.1.4 The electroweak unification

QED was successfully folded into the theory describing weak interactions, $V - A$, during the 1960s by Sheldon Glashow, Steven Weinberg, and Abdus Salam [43–45]. The result was a theory capable of describing both weak and electromagnetic interactions, which accepts all four gauge bosons, $W^\pm$, Z^0, and photons, as real particles that can be absorbed and emitted during transitions.

The missing piece is that the electric charge, Q, is not a fundamental generator of a unified $SU(2)_L \times U(1)$, since it does not preserve gauge symmetry. Instead, we define the weak hypercharge, $Y = 2(Q - T_3)$, where T_3 is the weak isospin, one of the generators of $SU(2)_L$, as the basis for $U(1)_Y$, which can now yield a symmetry group, $SU(2)_L \times U(1)_Y$, that is gauge invariant for all electroweak interactions. Both Y and T_3 are conserved quantities inside this group, which means that only terms with neutral Y and T_3 can enter the Lagrangian.

The new field representing the photon is B_μ, and the new coupling is g'. To construct the Lagrangian, we simply have the usual two terms. First, the gauge term, that is extended from QED and the weak interaction theories:

$$\mathcal{L}_{gauge} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.10}$$

Local gauge invariance is preserved in the fermionic kinetic term if we define $D_\mu = \partial_\mu - \frac{ig}{2}\vec{\tau}\vec{W}_\mu - ig'\frac{Y}{2}B_\mu$:

$$\mathcal{L}_{fermion} = \sum_{j=1}^{3} i\overline{\psi}_j(x)\gamma^\mu D_\mu \psi_j(x), \tag{1.11}$$

where, for the first quark family, for instance,

$$\psi_1 = \begin{pmatrix} u \\ d \end{pmatrix}_L \quad , \quad \psi_2 = u_R \quad , \quad \psi_3 = d_R.$$

The Lagrangian $\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion}$ is able to describe the kinematics of all fermions and gauge bosons in the theory, as well as the interactions between them. However, this expression does not incorporate mass terms for any of the particles involved, which would be of the form $m\overline{\psi}\psi$. We know empirically that the leptons are massive, for instance. So are the $W^\pm$ and Z^0 bosons.

Mass terms can be added if we include an additional, complex doublet field, which in turn hides four real, scalar fields:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^+ + i\phi_2^+ \\ \phi_3^0 + i\phi_4^0 \end{pmatrix}.$$

It is found that, considering a potential of the form $V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$, and $-\mu^2 < 0$, $\lambda > 0$, the Higgs field (ϕ_3^0), develops a vacuum expectation value:

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix} \quad , \quad v = \sqrt{-\frac{\mu^2}{\lambda}}.$$

This is known as spontaneous symmetry breaking (SSB), or the Higgs mechanism, in honor of Peter Higgs, one of the theoretical physicists who proposed this in 1964 [27], when applied to this particular topic.

Replacing $\phi = (0, v/\sqrt{2})$ into the Higgs Lagrangian term,

$$\mathcal{L}_{Higgs} = (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2,$$

we find mass terms for the W and Z bosons, but not the photon, consistent with everything that was known up to this point. The SM also yields precise predicted values for the masses of these bosons. Measuring differences with respect to these predictions could be an indicator of physics beyond our current understanding.

A similar Lagrangian can be formulated to give masses to fermions [44]:

$$\mathcal{L}_{\text{Yukawa}} = -y_u\bar{Q}_L\tilde{\phi}u_R - y_d\bar{Q}_L\phi d_R + \text{h.c.},$$

where $\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}$, $\tilde{\phi} = i\tau^2\phi^* = \begin{pmatrix}\phi^{0*}\\ -\phi^-\end{pmatrix}$, and y_f are the Yukawa couplings to each fermion[‡] and τ^2 is the second Pauli matrix.

After SSB, we find

$$M_f = \frac{y_f v}{\sqrt{2}}. \tag{1.12}$$

The presence of mixed-flavor terms in the Lagrangian is not forbidden by our theory, which means that M_f could adopt a 3×3 matrix form (there are three generations). And because these masses are measurable quantities, M_f must be diagonalizable, *i.e.*, there are unitary matrices $\mathcal{U}_{L,R}$ such that

$$(\mathcal{U}_L^U)^\dagger M_U \mathcal{U}_R^U = M_u = \text{diag}(m_u, m_c, m_t), \tag{1.13}$$

$$(\mathcal{U}_L^D)^\dagger M_D \mathcal{U}_R^D = M_u = \text{diag}(m_d, m_s, m_b). \tag{1.14}$$

We shall use capital letters to describe flavor (weak) eigenstates, U_L, D_L, while lowercase letters denote mass eigenstates, u_L, d_L:

$$U_{L,R} = \mathcal{U}_{L,R}^U u_{L,R} \quad , \quad D_{L,R} = \mathcal{U}_{L,R}^D d_{L,R}, \tag{1.15}$$

where

$$u_{L,R} = \begin{pmatrix}u\\ c\\ t\end{pmatrix}_{L,R} \quad , \quad d_{L,R} = \begin{pmatrix}d\\ s\\ b\end{pmatrix}_{L,R}.$$

[‡] The Yukawa couplings relate the fermions to the ϕ field, and are allowed (in fact, measured) to be different for each lepton. Leptons, however, share the same weak coupling constant in the electroweak model, g. This is known as Lepton Flavor Universality, and it is an essential part of the SM. Small deviations between lepton couplings could point to the need of a larger theory, so it is currently an active field of study by experimentalists.

Notice that this means that massive quarks are a mixture of different flavors eigenstates. We shall define the Cabibbo, Kobayashi, Maskawa (CKM) matrix [46],

$$V_{CKM} = (\mathcal{U}_L^U)^\dagger \mathcal{U}_L^D, \tag{1.16}$$

which is a fundamental piece of the SM for various reasons. First, it predicts cross-generation transitions in weak decays, which have been observed and well measured. It is also the gateway for CP violation into the SM.

This happens because we can always parametrize a n-dimensional unitary matrices with n^2 parameters, which are split between mixing angles and phases. The mixing angles are described inside the framework of the symmetry subgroup $SO(n)$, which says that a real rotation in n dimensions in parametrized by $n(n-1)/2$ angles. The rest of the parameters, $n(n+1)/2$, are phases.

For $n = 3$, we have 3 angles (θ_{12}, θ_{13}, θ_{23}) and 6 phases. They must correspond to the six quark flavors, and in general, can be reabsorbed into the quark fields. In fact, we can always do a global phase shift that affects all quarks in the theory and leaves it unchanged. This means that there are only five independent quark phases, which would be unmeasurable, as they vanish for $\overline{q}q$. It leaves us with only one phase, δ, which is CP-violating since it survives in $\overline{q}q$ (matter-antimatter) terms.

With this in mind, we can write the CKM matrix as follows:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix}, \tag{1.17}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$.

The CKM matrix enters the Lagrangian through the charged-currents part, contained inside the covariant derivative,

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}}\overline{u}_L\gamma^\mu V_{CKM} d_L W_\mu^+ + h.c. \tag{1.18}$$

So far, it is only known experimentally [47]:

$$\begin{aligned} V_{\text{CKM}} &= \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \\ &= \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix}. \end{aligned} \tag{1.19}$$

Transitions inside the same generation are very much favored, while transitions between generations are increasingly less likely as the shift between generations widens. We also have measurements for the three angles and the CP-violating phase:

$$\begin{aligned} \sin\theta_{12} = 0.22650 \pm 0.00048 \quad &, \quad \sin\theta_{13} = 0.00361^{+0.00011}_{-0.00009} \\ \sin\theta_{23} = 0.04053^{+0.00083}_{-0.00061} \quad &, \quad \delta = 1.196^{+0.045}_{-0.043}. \end{aligned} \tag{1.20}$$

The SM predicts that V_{CKM} is unitary, which represents an excellent experimental test for the theory.

The CKM matrix was first proposed in 1973, when the authors of Ref. [46] used it to reason that there must be a total of three fermion families, and that the quark model was not a quartet, as it had been found until then. For the CP-violating phase δ to appear in the SM Lagrangian, there needed to be two more, then-undiscovered quarks. Across the history of the SM, theory and experiment have been ahead of the other in alternating situations.

1.2 Beyond Our Standard Model

The successes of the SM as a fundamental theory that explains the subatomic world have been many over the years. Chief among them remains the discovery of the Higgs boson in 2012 [48, 49], but as indicated, the model has also accurately predicted the number of quarks and leptons involved in it [46], as well as the existence and the mass values of three of the four electroweak bosons (later corroborated by experiments).

However, there remain some inconsistencies between some of the theoretical predictions and experimental observations, as seen in Sect. 1.2.1. Additionally, there are other aspects of the theory that remain incomplete or unsatisfying to the community (see Sect. 1.2.2.

1.2.1 Experimental tensions

1.2.1.1 The anomalous magnetic moments of leptons

QED has also been able to predict the value of the anomalous magnetic moment of the electron, a measure of the contribution of quantum mechanical effects to the magnetic moment of the particle, with a precision of 12 significant figures [50],

$$\frac{g_e}{2} = 1.001\ 159\ 652\ 180\ 59(13). \tag{1.21}$$

Naturally, given the level of precision, this is also a good target for searches of tiny deviations from the SM.

The Dirac equation predicts an exact value of $g_e = 2$ [51], but precise calculations in Quantum Electrodynamics using Feynman diagrams with up to four loops have contributed slight corrections that are later corroborated in experiment (see references inside Ref. [50])

In a similar fashion, QED can also compute precise corrections to the anomalous magnetic moment of the muon [52], which is also predicted by Dirac to be $g_\mu = 2$, but which has been found both experimentally and theoretically to be larger than 2. However, the excellent agreement found between theory and experiment for g_e has not been replicated for g_μ. Results published by Fermilab in 2023 [53] concluded a world-average experimental value of

$$a_\mu^{\text{exp}} = \frac{g_\mu - 2}{2} = 116\ 592\ 059(22) \times 10^{-11}, \tag{1.22}$$

which represented a 4.2σ deviation from the most up-to-date SM prediction (see (8.2) from Ref. [52] and references therein).

This is known as the $(g-2)_\mu$ puzzle, and has historically been a source of disagreement between the SM and experimental observations (although it must be noted that most of the discrepancies come from theoretical and systematic uncertainties, rather than being of statistical nature). It has inspired the creation of its own collaboration, the Muon $g-2$

Collaboration at Fermilab, which is expected to publish an even more precise value of this quantity in 2025. Naturally, it is believed that new physics could contribute to these deviations.

The theoretical calculations needed for the computation of a_μ are rather complex, and involve a lot of uncertainties of different sources, including experimental results and also computational power. The dominant source stems from the hadronic vacuum polarisation (HVP) at leading order (LO) in the fine-structure constant, which modifies the propagation of virtual photons in the vacuum. Ref. [54], published in July 2024 and following new measurements from Ref. [55], updated the LO-HVP contribution to a_μ with unprecedented accuracy. When adding this new value to the rest of SM contributions, they found

$$a_\mu^{\text{SM}} = 116\ 592\ 019(38)\ \times 10^{-11}. \tag{1.23}$$

This updated value differs from observations in

$$\frac{a_\mu^{\text{exp}} - a_\mu^{\text{SM}}}{\sqrt{(\sigma(a_\mu^{\text{exp}}))^2 + (\sigma(a_\mu^{\text{SM}}))^2}} = 0.9\sigma.$$

Another particularly challenging contribution to the theoretical computations is the hadronic light-by-light (HLbL) scattering, which involves a process in which four photons interact through a hadronic intermediate state. Notice how this is theoretically more complex than HVP, as the contributions involve a four-point function rather than a two-point one, as is the case with HVP.

At low energies, the exchanged mesons can be light pseudoscalars, like the π^0, η, or η'. Ref. [56] published an updated computation of the term using lattice QCD,

$$a_\mu^{\text{Hlbl}} = 106.8(14.7) \times 10^{-11}. \tag{1.24}$$

The 14% uncertainty, which adds in quadrature all systematics carried through the computations, may be further constrained with additional and more precise experimental measurements.

The door is not shut yet, though. The Muon $g-2$ Collaboration, as mentioned, is expected to update their results in 2025 [57]. Results from this thesis regarding precise measurements of branching ratios of η and η' decays, could contribute to improving theoretical calculations of transition form factors that play a key role in the computation of the hadronic light-by-light scattering.

1.2.1.2 CP violation and the matter-antimatter asymmetry

Shortly after the Big Bang, the universe was a quark-gluon plasma (QGP), which eventually evolved into the galaxies, planets, and stars we see 13 billion years later. One would assume that from that state, an equal number of particles and antiparticles emerged, which would preserve charge and parity.

However, we are only able to see objects made of matter, not anti-mater, which means that *CP* was violated at some point. Our SM gateway into *CP* violation is given by the corresponding phase in the CKM matrix, δ, as well as the θ term from the strong *CP* problem (see Sect. 1.2.3.2).

Neither term is relevant enough to explain such a large discrepancy in the early Universe. The LHCb experiment was originally conceptualized to tackle this problem, provide a better understanding of how CP violation works inside the SM, and possibly discover some tensions that may yield new physics. The understanding of the early-universe QGP is also one of the LHCb's goals (see more in Sect. 3).

1.2.1.3 Additional question marks

- **$b \to s\ell^+\ell^-$ transitions**: The branching ratio measurements of $B^+ \to K^+\mu^+\mu^-$ [58] and $B_s^0 \to \phi\mu^+\mu^-$ [59] present, respectively, $4.2{-}3.6\sigma$ deviations from SM predictions in the $\mu^+\mu^-$ bins below the J/ψ mass. The angular observable P_5' has also introduced $2.5 - 2.9\sigma$ tensions with respect to the theory [60].

- **The X17 anomaly**: The measurement of the relative angle between the e^+e^- pair emitted by the excited state of ^{8}Be showed a peak at the invariant mass of 17 MeV/c^2 with a statistical significance of 6σ [61], which could correspond to a new boson and even be related to the $(g-2)_\mu$ puzzle [62].

- **The W boson mass**: While some LHC experiements, including ATLAS [63], CMS [64] and LHCb [65] have measured the value of the W boson mass to be consistent with the SM prediction [66], in 2022 a result by the CDF II Collaboration, based on legacy data collected by the Fermilab Tevatron collider showed a 7σ tension with respect to the SM [67]. This remains one of the main focus of the electroweak programs at the LHC experiments, including LHCb.

- **Neutrino masses**: The Neutrino Solar problem was solved in the 20th century with a simple yet still fully unexplained reason. The rate of neutrinos reaching the Earth was exactly one third of the ones that were supposed to be produced on the surface of the Sun. This led to the conclusion that there were three families of neutrinos, each related to one of the leptons. Furthermore, they were able to oscillate between flavors on their way to Earth, something that can only be explained by quantum mechanics if they are allowed to have mass. However, the SM Lagrangian does not include a mass term for the neutrinos. Several BSM theories have presented different explanations for this phenomenon, but none of them have been proven yet.

- **Dark energy**: We know that baryonic matter makes up roughly 5% of the known energy in the universe, while dark matter is responsible for 28%. However, 68% of the known energy content in the universe is a mysterious form of energy, dubbed dark energy, which is believed to be responsible for the expansion of the universe. It is not directly observable, and its true nature is one of physics' largest blind spots.

1.2.2 Theoretical problems

- **The inclusion of gravity**: The SM is able to reconcile under one Lagrangian three of the four fundamental forces, even if QCD has not been unified with the electroweak theory. However, gravity, with a much weaker coupling constant, remains isolated from the rest. The effects of quantum gravity would be felt at energy scales way above our current standards for high-energy physics and detectors, so we are still a long ways off from being able to understand Einstein's relativity as a quantum field theory.

- **The hierarchy problem**: Going even further, there is a large discrepancy in the SM between the energies governing the weak interaction and gravity. This hierarchy puzzle can be refocused by looking at the Higgs mass. The observed value of this parameter, which is not protected by any symmetry in the SM, corresponds to that computed through tree-level Feynman diagrams. Loop-level corrections, however, can become arbitrarily large, and theorists have not found a suitable explanation to cancel them so that the actual mass the measurement.

- **The gauge couplings in the SM**: The SM coupling constants tend to have a tendency towards unification at an energy scale of ~ 16 GeV, but they do not quite match. This is due to the arbitrary number of free parameters that conform to the SM, and is solved by some Beyond the Standard Model theories, like supersymmetry (see Sect. 1.2.3.3).

1.2.3 Some New Physics models

1.2.3.1 Dark sectors

After decades of hypothesis about the universe possibly being made of more matter than is actually visible to us, a plethora of astrophysical observations in the 1970s led to the conclusion that only a small fraction of the matter we are able to see is baryonic matter. (See Ref. [68] and references therein.) The nature of the rest, dubbed *dark matter*, remains a mystery. It is able to interact with gravity, but so far we haven't observed any interactions with the SM, which implies at least very weak couplings to it, if any.

There are many simultaneous, ongoing efforts in the particle physics community to both understand the fundamental elements making up dark matter and also reconcile this new *hidden sector* with our SM. If particles from this hidden sector are to interact with the SM, they must do so through a mediator that would act as a portal between both sectors. See Refs. [69–72] for a detailed overview.

Several possibilities arise. One of them would be a spin–1 particle similar to the SM's photon, a dark photon, that would act as the force carrier of the dark matter equivalent of electromagnetism. This can be added into the SM Lagrangian by extending its symmetry group with a $U(1)_D$ gauge symmetry [71, 73],

$$\mathcal{L} \supset -\frac{1}{4}F'_{\mu\nu}F'^{\mu\nu} - \frac{1}{2}m^2_{A'}A'_\mu A'^\mu + \varepsilon e A'_\mu J^{\mathrm{EM}}_\mu, \tag{1.25}$$

where $m_{A'}$ is the mass of the hypothetical dark photon, and ε represents the kinetic mixing strength. These two are the fundamental parameters of the theory; the last term represents the portal between the SM and DM.

If $m_{A'}$ is above the mass scale of the dark matter candidate, it will decay into it and will not be observable to us. However, if it is below, it will decay into SM particles and we might be able to observe it. Many searches across many different experiments have targeted A' decays, but could only get so far as placing stringent constraints on $m(A')$ and ε^2 (see Refs. [74–82]) The LHCb Collaboration conducted a wide-range search for $A' \to \mu^+\mu^-$ decays [83] with data collected between $2016 - 2018$, for both prompt-like and displaced di-muon resonances, but found no significant excess in $0.214 < m(A') < 70$ GeV/c^2.

One of the decays analyzed presented in this thesis, $K^0 \to \mu^+\mu^-\mu^+\mu^-$, was predicted by a dark matter model [70, 84, 85] to have its branching ratio enhanced by two orders of magnitude with respect to its SM value. Similar models could also solve some of the mysteries in the $(g-2)_\mu$ puzzle [85].

1.2.3.2 Axion-like particles

There is an infinite number of topological vacua in QCD, but the true, physical vacuum of the theory is a superposition of all of them [86, 87],

$$|\theta\rangle = \int e^{i\nu\theta}|\nu\rangle.$$

This term, however, translates into a component in the QCD Lagrangian,

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{QCD}} + \frac{\theta\alpha_s^2}{32\pi^2} F_{\mu\nu,a}\tilde{F}^{\mu\nu,a}, \tag{1.26}$$

where $\tilde{F}^{\mu\nu,a} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$. Experimentally, θ is found to be non-zero, but really small. The $\mathcal{L}_\theta$ term, however, is CP-violating. This is known as the strong CP problem.

Refs. [88, 89] introduced a possible solution, where they added a global symmetry $U(1)_{PQ}$ that, in turn, introduced an additional ϕ field. This new symmery forces θ to be a dynamic variable rather than a parameter once it is spontaneously broken. It now depends on the vacuum expectation value of the axion field, $a(x)$, which appears after SSB, and it can be reworked theoretically to be set to 0.

QCD axions are the Nambu-Goldstone bosons associated to this particular $U(1)_{PQ}$ SSB. Axion-like particles (ALPs) appear whenever a symmetry in a $U(1)_{PQ}$ theory is broken. They can acquire a mass, m_a, and can couple to SM particles. Being massive and feebly interacting with the SM, axions and axion-like particles are perfect candidates for dark matter.

The precision measurement of $\mathcal{B}(\eta' \to \pi^+\pi^-\mu^+\mu^-)$, presented in Sect. 6, will also include a search for potential bumps in the $\mu^+\mu^-$ invariant distribution that may be originated by $a \to \mu^+\mu^-$, where a is an ALP.

1.2.3.3 Supersymmetry

The supersymmetry (SUSY) model [90, 91] is based on the assumption that there exists a symmetry between the bosons and the fermions of the theory. As such, every particle would have a super-partner, and there would be a unified description independent of matter and interaction. The R-parity is defined as being $R = 1$ for SM particles and $R = -1$ for their supersymmetric partners.

There are many SUSY models, with the Minimal Supersymmetric Standard Model being the simplest example: a natural extension of the SM that includes supersymmetric partners to the known particles and without additional fine-tuning of the parameters.

Supersymmetry, if found, would be able to solve many of the SM problems listed above, including the hierarchy problem, dark matter, and the unification of the gauge constants. However, the faith in finding experimental evidence of supersymmetry with our current accelerators has faded over the years, since no supersymmetry particles have been observed. If SUSY is to solve the hierarchy problem, its particles should not have masses beyond the TeV scale.

2

Statistical tools in High Energy Physics

Experimental particle physics is a field built upon the interpretation of many measurements of processes that happen in nature. These measurements are generally independent of each other, and a proper framework needs to be put in place for correctly studying and understanding the collected data.

Before performing a measurement, we only know about the probabilities of each possible outcome. We are unable to predict with certainty what will happen. Kolmogorov defined the probability P as a real-valued function that satisfies a set of axioms inside a set S, called the sample space, and possible subsets A, B, ...

i) For every subset A in S, $P(A) \geq 0$.

ii) For disjoint subsets, $P(A \cup B) = P(A) + P(B)$.

iii) $P(S) = 1$.

Today, it is more common to speak about probabilities inside the frequentist and bayesian paradigms. For a frequentist, the probability is given by the number of times each outcome A, B,... of an experiment is yielded if it was performed an infinite number of times. For a bayesian, this is a more subjective quantity, as it would be the degree of belief that the hypothesis A is true.

2.1 Probability density functions

Let x be the possible outcome of a repeatable experiment. This is generally called a random variable by a frequentist. If x can take only one value inside a continuous range, the probability of measuring said random variable being inside $[x, x+dx]$, for an infinitely small dx, is given by $f(x;\theta)dx$. This function $f(x;\theta)$ is called the probability density function (*p.d.f.*), and it may additionally depend on a set of parameters θ.

If $x \in [a,b]$, the *p.d.f.* must be normalized inside that interval,

$$\int_a^b f(x;\theta)dx = 1. \tag{2.1}$$

Probability density functions will be used in this thesis to describe the behavior of the data that was obtained experimentally. Some examples include:

- **Binomial distribution**: A random process with only two possible outcomes that have fixed probabiilties (p for success and $1-p$ for failure) is called a Bernouilli process. The probability of obtaining r successes out of N independent trials is given by the binomial distribution, whose standard deviation is also indicated below:

  $$f(r;N,p) = \frac{N!}{r!(N-r)!}p^r(1-p)N-r \quad ; \quad \sigma = \sqrt{\frac{p(1-p)}{N}} \tag{2.2}$$

  This is often used to compute the uncertainty associated to the efficiency of a given variable selection.

- **Poisson distribution**: In the limiting case $p \to 0$, $N \to \infty$, $Np = \nu$, the binomial distribution becomes the Poisson distribution, which describes the probability of observing n independent events in a fixed interval x (space, time, etc.) at an average rate of ν:

  $$f(n;\nu) = \frac{\nu^n e^{-\nu}}{n!} \quad ; \quad \sigma = \nu. \tag{2.3}$$

  Measurements of the times a given process occurs (for instance, events inside a histogram) are modeled by a Poisson distribution.

- **Gaussian distribution**: When the average rate is large, the Poisson distribution approaches the Gaussian distribution,

  $$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{2.4}$$

  where σ is the associated standard deviation. A Gaussian *p.d.f.* may be used, as a first-order approximation, to model the invariant mass distribution of the decay products in a process where all the tracks are measurable.

- **Crystal Ball distribution**: The convolution of a Gaussian *p.d.f.* with an exponential that models one of the tails is known as a Crystal Ball. It is defined as [92]

  $$p(x;\mu,\sigma,\alpha,n) = \begin{cases} A \cdot (B - \frac{x-\mu}{\sigma})^{-n}, & \text{if } \frac{x-\mu}{\sigma} \leq -\alpha, \\ exp\big(-\frac{(x-\mu)^2}{2\sigma^2}\big), & \text{if } \frac{x-\mu}{\sigma} > -\alpha \end{cases} \tag{2.5}$$

  where

  $$A = \left(\frac{n}{|\alpha|}\right)^n exp\left(-\frac{|\alpha|^2}{2}\right) \quad ; \quad B = \frac{n}{|\alpha|} - |\alpha|.$$

This *p.d.f.* describes quite well processes that may feature energy losses due to bremsstrahlung radiation, for instance. The sign of the α parameter dictates whether the exponential tail is the left one (positive) or right one (negative).

An even more refined model would involve the superposition of two Crystal Balls, each with a different sign of α. This is known as a double-sided Crystal Ball and is often used in the thesis, especially in the analysis of $\eta^{(\prime)}$ decays. It can be useful in cases where the events have different uncertainties on the measured mass, which could distort the core of the Crystal Ball.

- **Hypatia distribution**: The Hypatia distribution [93] is a generalization of the Crystal Ball that brings the concept behind the superposition of various Crystal Ball distributions to the next level. Using the combination of various *p.d.f.s* is equivalent to assuming that the per-event uncertainty is a sum of as many delta functions. Per-event uncertainties are typically continuous functions very different from the sum of a limited number of deltas.

  The following *p.d.f.* was defined as a generalization of the Crystal Ball with the goal in mind of addressing these issues, and is used to model the K^0_{S} invariant mass distributions in Sect. 5:

$$\mathcal{I}(m,\mu,\sigma,\lambda,\zeta,\beta,a,n) = \\ \begin{cases} \left((m-\mu)^2 + A^2_\lambda(\zeta)\sigma^2\right)^{\frac{1}{2}\lambda-\frac{1}{4}} e^{\beta(m-\mu)} K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right) & , \text{ if } \frac{m-\mu}{\sigma} > -a \\ \frac{G(\mu-a\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{\left(1-m/\left(n\frac{G(\mu-a\sigma,\mu,\sigma,\lambda,\zeta,\beta)}{G'(\mu-a\sigma,\mu,\sigma,\lambda,\zeta,\beta)}-a\sigma\right)\right)^n} & , \text{ otherwise} \end{cases} \tag{2.6}$$

  where

$$G(m,\mu,\sigma,\lambda,\zeta,\beta) \propto \\ \left((m-\mu)^2 + A^2_\lambda(\zeta)\sigma^2\right)^{\frac{1}{2}\lambda-\frac{1}{4}} e^{\beta(m-\mu)} K_{\lambda-\frac{1}{2}}\left(\zeta\sqrt{1+\left(\frac{m-\mu}{A_\lambda(\zeta)\sigma}\right)^2}\right) \tag{2.7}$$

To understand the model that dictates how our measurable random variable x will behave we need to find the parameters θ that best describe our physics case. We can do this using the maximum likelihood (ML) method, which gives us the set θ for which $\mathcal{L}(\theta) \equiv P(x;\theta)$ is maximum. The problem is usually formulated, instead, as the minimization of $-\log\mathcal{L}(\theta)$, because computers have it easier to minimize than to maximize, and the properties of the logarithm lend themselves to simplifying complex calculations.

If x is a set of statistically independent quantities, $(x_1, x_2, ..., x_n)$, like measurements in an experiment would be, we can factorize

$$\mathcal{L}(\theta) = \prod_i f(x_i;\theta). \tag{2.8}$$

If $(x_1, ..., x_n)$ are independent Gaussian random variables,

$$z = \sum_{i=1}^{n} \frac{(x_i-\mu_i)^2}{\sigma_i^2} \tag{2.9}$$

follows the χ^2 distribution with n degrees of freedom:

$$f(z;n) = \frac{z^{n/2-1}e^{-z/2}}{2^{n/2}\Gamma(n/2)} \quad ; \quad \sigma = \frac{\sqrt{k}}{\lambda}. \tag{2.10}$$

This *p.d.f.* is often used to validate the how well a given hypothesis for a *p.d.f.* describes the observed data. This is an example of hypothesis testing.

2.2 Cumulative distribution functions

Every *p.d.f.*, P, has an associated cumulative distribution function (*c.d.f.*), F, defined as the probability that a random variable X takes a value less or equal to x,

$$F(x) = P(X \leq x). \tag{2.11}$$

The analysis described in Sect. 5 will use the concept behind cumulative distribution functions to make data-MC corrections.

The idea is to create a $1-1$ relation between a given variable, in both data and simulation, and its *c.d.f.*,

$$y_{\mathrm{MC,Data}} = \mathcal{P}_{\mathrm{MC,Data}}(x_{\mathrm{MC,Data}}), \tag{2.12}$$

where $x_{\mathrm{MC,Data}}$ is the variable in the corresponding dataset and $y_{\mathrm{MC,Data}}$ is the *c.d.f.* We then define the corrected variable, x_{corr}, as

$$x_{\mathrm{corr}} = \mathcal{P}^{-1}_{\mathrm{Data}}\left(\mathcal{P}_{\mathrm{MC}}(x_{\mathrm{MC}})\right). \tag{2.13}$$

Introduced in Sect. 3.8.1, the LHCb tool known as `PIDGen` also uses this technique to correct variables based on particle identification, though the procedure is a bit more complex since the correction is performed in a three-dimensional binning scheme.

2.3 Confidence intervals and hypothesis testing

In a frequentist approach, if we define a confidence interval $[a,b]$ as a boundary in which, if we repeated a certain experiment a large number of times, the parameter of interest θ would fall inside said interval a C.L. % of the times, where C.L. is defined as the Confidence Level. We can also write

$$\mathcal{L}(a < \theta < b) = C.L. \tag{2.14}$$

This can also be rephrased as finding θ inside $[a,b]$ with a probability of $1-\alpha$. Here, $[a,b]$ contains all values of θ that cannot be rejected at a significance level α. The hypothesis $H_0 : \theta = \theta_0$ is rejected at a significance level α if $\theta_0 \notin [a,b]$.

Note that θ can not be precisely known, but rather estimated. Let $\hat{\theta}$ be the parameter that maximizes the likelihood. Often times, we can use the profile likelihood ratio,

$$\lambda(\theta) = \frac{\mathcal{L}(\theta)}{\mathcal{L}(\hat{\theta})}, \tag{2.15}$$

to measure how likely a given θ is with respect to $\hat{\theta}$.

The null hypothesis, H_0, is a rule for which a set of data values x is rejected. The critical region w is that where there is no more than a given probability α, under H_0, of finding $x \in w$. If the data are observed in w, H_0 is rejected.

An immediate application of this reasoning, known as hypothesis testing, can be a new discovery. Let H_0 state that the observation of θ is compatible with the SM prediction. If the hypothesis can be rejected with a 99.73% C.L. , we would say there is evidence of new physics. This corresponds to θ being 3σ away from the SM expectation. A 5σ deviation would be categorized as an observation.

The alternative hypothesis, H_1, in this case would be a new physics case. Wilks' theorem [94] states that the test statistic

$$q_0 = -2 \log \lambda(x) = -2 \frac{\mathcal{L}(x|H_1)}{\mathcal{L}(x|H_0)} \tag{2.16}$$

follows a χ^2 distribution with one degree of freedom, which is equivalent to the square of a normal distribution, $\chi^2_1 = N(0,1)^2$. This means that the number of standard deviations, or significance of H_1, would be given by

$$S = \sqrt{-2\left(\mathcal{L}(x|H_1) - \mathcal{L}(x|H_0)\right)} \tag{2.17}$$

Wilks' theorem will be used in this thesis to estimate the significance of observing $\eta' \to \mu^+\mu^-$.

2.3.1 Setting upper limits on branching fractions

The definition of the confidence interval, as given above, can be rewritten as

$$C.L. = \int_a^b f(x|\theta)dx. \tag{2.18}$$

This thesis will attempt to measure the branching fractions of multiple decays. In the cases in which they are not observed, a confidence interval will be set for those decays. The lower bound will be 0, and the upper bound will be hereafter referred to as our upper limit for the given $\mathcal{B}$. Given this definition, we can probe possible values of b until we find that which provides the desired C.L. Here, $f(x|\theta) = f(\mathcal{B})$ is the likelihood of the branching fraction as provided by the fit.

2.4 An introduction to Machine Learning

In all of the analyses presented in this thesis, the dominant source of background is a combination of random tracks that have satisfied all prior fiducial selections. This is also the case across many studies in modern High Energy Physics, provided the large statistics. Computational tools developed for big data problems have therefore found their way into the treatment of the physics we measure. The predominant one is machine learning (ML).

In reality, machine learning has been present in experimental physics since its inception. Defining ML as studying the behavior of a dataset to later make educated predictions, one can differentiate between supervised ML, in which the model knows the correct solutions during its learning process, and unsupervised, where there are no labels or pre-defined answers.

Supervised models are the most common ones, and we have already discussed one of them. At its most basic level, a maximum-likelihood fit to a dataset simply attempts to find a set of parameters θ that will best describe the behavior of a dataset so that later we can make a prediction. Take a set of data points (x, y), which follow a linear model, $y = f(x; a, b) = a + bx$. The machine learning model attempts at finding a and b such that we can later provide an entirely new x and get the corresponding y. The optimal set of parameters is that which maximizes a likelihood function. Or, in other words, minimize a loss function.

The definition of a loss function is the key to introducing any new ML algorithm.

2.4.1 An aside on fits

As indicated above, the loss function used to fit a data set of n entries to a given *p.d.f.* $f(x;\theta)$ is given by the negative logarithm of the likelihood

$$\ell(\theta) = -\log(\mathcal{L}(\theta)) = -\sum_{i=1}^{n} \log(f(x_i;\theta)). \tag{2.19}$$

The process of fitting consists in finding the set of parameters θ that minimizes $\ell(\theta)$.

The minimization process can be done simultaneously for various sets of data, whether they share some parameters in θ or not. For data sets $\mathcal{D}_1$ and $\mathcal{D}_2$, with a set of parameters θ_1 and θ_2 and loss functions $\ell_1(\theta_1)$ and $\ell_2(\theta_2)$, a simultaneous fit would minimize

$$\ell(\theta_1, \theta_2) = \ell_1(\theta_1) + \ell_2(\theta_2). \tag{2.20}$$

In terms of likelihood, the total likelihood would be the product $\mathcal{L}_1(\theta_1) \times \mathcal{L}_2(\theta_2)$. All of the analyses in this thesis will feature a simultaneous fits to two categories.

In the case where multiple observables are targeted, for instance different mass variables, there can also be multi-dimensional fits. The *p.d.f.* would be the product of all individual *p.d.f.s* describing each of the observables, as long as these observables are not correlated. In the case of a bi-dimensional fit,

$$f(x, y|\theta_x, \theta_y) = f_1(x|\theta_x) \otimes f_2(y|\theta_y). \tag{2.21}$$

In this case,

$$\ell(\theta_x, \theta_y) = -\sum_{i=1}^{n} \log(f(x_i, y_i; \theta_x, \theta_y)). \tag{2.22}$$

Constraints can also be set on some of the parameters in θ, by adding an additional term to the loss function,

$$\ell_{\text{Total}}(\theta) = \ell_{\text{Data}}(\theta) + \ell_{\text{Constraint}}(\theta_j), \tag{2.23}$$

where θ_j is the constrained parameter. The only constraints applied on fits in this this thesis are Gaussian constraints, which originate from a Gaussian likelihood function.

Constraints themselves are independent measurements. In the case of Gaussian-constraining the parameter θ_j, the values estimated by this measurement are the mean of the *p.d.f.*, which is precisely θ_j. The variable that defines the Gaussian *p.d.f.* is given by the central value provided by previous knowledge used to constraint θ_j; the uncertainty is the width of the Gaussian *p.d.f.*

In this thesis, Gaussian constraints are used to incorporate information from fits to simulated into the fit to LHCb data. The central value given by simulation is taken as the x variable of the gaussian, while its uncertainty (sometimes multiplied by a scaling factor to ensure fit convergence) is used as the width of the Gaussian *p.d.f.* Systematic uncertainties are also introduced via Gaussian constraints on the final fits.

2.4.2 Decision trees

Supervised machine learning problems can be of either a classification or regression nature. Classification problems consist of labeling a given instance of our dataset as belonging to one of multiple pre-defined categories. Regression problems are, in essence, fits.

Decision trees are among the simplest instances of classification algorithms. They take in a set of input features, x, that should be descriptive of our dataset, and create a set of logical sequences of selections, $x_i > y$, on these.

The structure is similar to a flowchart, with a root node at the beginning on which the full dataset is asked a question about an input feature x_i. Events passing the query will go on to a decision node in a lower level, while events not passing it will go to a different node, on the same lower level. There are as many levels as the user wants, though adding many of them may result in training an algorithm that is specific to this data set, which can cause problems when generalizing.

Leaf nodes are at the end of the chain. They can be created because we have too many levels in our tree, or because the likelihood of instances from one of the categories reaching this node is overwhelming when compared to the rest.

The decision set up at each node is based on the minimization of the following loss function:

$$\mathcal{L}(x_i, y) = \frac{N(x_i < y)}{N_T} g(x_i < y) + \frac{N(x_i > y)}{N_T} g(x_i > y) \quad ; \qquad g(sel) = 1 - \sum_{j=1}^{k} \left(\frac{N_j(sel)}{N_T} \right)^2 , \tag{2.24}$$

where N_T is the total number of events that survived through the corresponding node, N_j is the number of events from the category j in that node, and k is the total number of categories.

Decision trees can later be combined to form more robust predictors. For instance, we can split our sample into sub-data sets and train different decision trees with each sub-sample. The final prediction would combine the predictions of each one. Boosted decision trees (BDTs) are an ensemble of several weak learners (algorithms with a slightly better performance than random guessing). The ensembling can be done through various techniques and weighted using a parameter known as the learning rate, which can accelerate or slow down the process.

BDTs are highly customizable thanks to the input hyperparameters, which address the levels of depth of each tree, the number of total decision trees, the learning rate, the minimum number of samples to define a leaf node, the maximum number of leaf nodes per tree, the size of the sub-samples used to train each decision tree in the ensembling, etc.

2.4.3 Testing the performance

In this thesis, BDTs will be used to classify between signal- and background-like events. Simulation will be used as a proxy for signal candidates, while combinatorial background will be used as the corresponding proxy. A large part of the data sets, in each case, will be used to train these models, *i.e.*, find the optimal decisions to be made at each node of each of the trees involved.

The performance of the models will have to be tested with a separate data set. This is to ensure that the model is not overtrained, that is, over-specific to the training sample. To ensure that the full data set is used in the training and performance testing, we use a method known as k-folding, or cross-validation. It divides the full sample into k subsamples and systematically rotates them for multiple uses. Besides training and testing, a third holdout sample is also created. The trained and tested classifying algorithm will be applied on it.

Several metrics can be used to evaluate the performance. The accuracy, for instance, is simply the fraction of correct predictions. The precision is the number of correct positive predictions over the number of positive predictions, while the recall is the number of positive predictions over the number of positive instances. The harmonic mean of the precision and recall is the F1 score. This score will be used in Sect. 6 to retrieve the optimal hyperparameters of the BDT.

The Receiver Operating Characteristic (ROC) curve is often seen as the most reliable performance indicator. It plots the true positive rate vs. the false positive rate. In physics terms, it would be the signal efficiency vs. the background efficiency. In this thesis, however, the horizontal axis will be inverted, and the ROC will represent the signal efficiency vs. background rejection. The associated metric is the area under the curve (AUC), which is supposed to be as close to 1 as possible.

3

Experimental conditions

3.1 The Large Hadron Collider

One of the key takeaways from Sect. 1 is the idea that, for fermions and gauge bosons to have mass in the SM, there must exist a field, ϕ_3, known as the Higgs field. This field would have an associated spin–0 particle, whose discovery would be the final piece of the puzzle that was once started when Thomson discovered the electron.

The search for the Higgs boson started shortly after its theoretical prediction, with the Large Electron-Positron Collider coming close to finding it at the end of the 20th century [95]. It was also the driving motivation behind the construction of the Large Hadron Collider (LHC) [96], a massive, 27 km-long circular accelerator built underneath the border between France and Switzerland, outside the city of Geneva. This is the largest physics laboratory in the world, and currently operates four experiments at the same time.

- **ATLAS** [97]: This general-purpose detector concentrates on studying high-p_{T} physics, as well as precision measurements on the higher-end mass spectrum of the SM, including the t quark and the Higgs boson. Along with CMS, it operates at a much higher luminosity that the other two LHC experiments, which allows it to access these regimes, at the cost of having worse particle identification capabilities. It reported the discovery of the Higgs boson in 2012 [48].
- **CMS** [98]: Its physics program has a high overlap with that of ATLAS, and often they serve as a way of corroborating the other's observations. This was the case when CMS also reported the discovery of the Higgs boson in 2012 [49].
- **ALICE** [99]: This experiment is oriented towards heavy-ion collisions and the studies of quark-gluon plasma to infer what happened in the early moments of the Universe. It operates at a much smaller instant luminosity than its LHC counterparts.

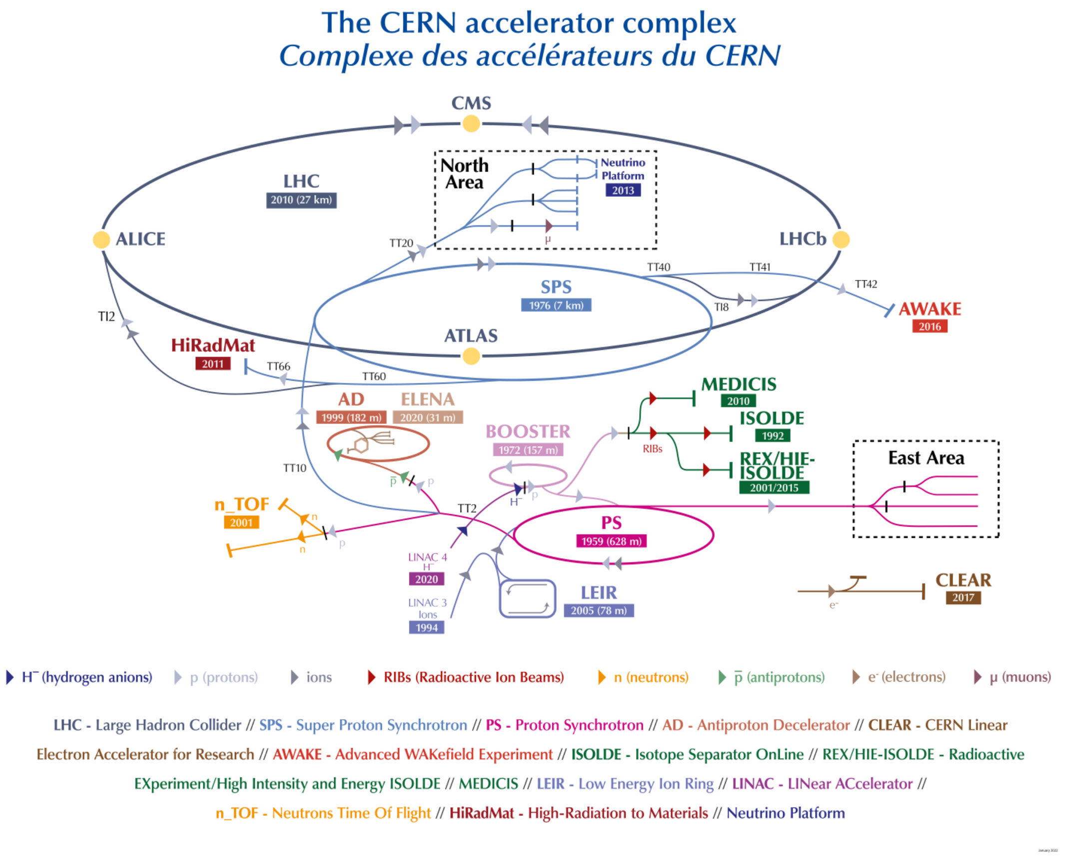


Figure 3.1: Visual representation of the chain of pre-accelerators that the proton beams go through before being injected into the LHC.

- **LHCb** [100]: The main goal of LHCb is the study of *CP* violation and precision measurements in flavor physics. This thesis will analyze data collected by this experiment between 2016 and 2018. More details are provided in this chapter.

The LHC collides proton bunches in each of these points of interaction. These protons are taken from hydrogen atoms and accelerated at the Linac2 [101] linear accelerator. They are able to reach an energy of 450 MeV, and are subsequently injected into a chain of circular accelerators that includes the Proton Synchroton Booster (PSB), the Proton Synchroton (PS), and the Super Proton Synchroton (SPS). The created proton bunches leave the SPS with an energy of 450 GeV, and enter directly the LHC. The chain of pre-accelerators is represented in Fig. 3.1.

Each collection of bunches make up a beam. Up to 2808 bunches can be injected into each beam. At the beginning of the LHC, the spacing between bunches was of 50 ns. Starting in 2015, the spacing was reduced to 25 ns, which allowed to reach higher luminosities.

Once they enter the LHC, magnetic fields of up to 8.3 T, created by 1232 superconducting dipole magnets cooled down to 1.9 K, are used to force the beams into their circular path. Beams are injected into the LHC through interaction points 2 and 8. Interaction point 4 includes eight superconducting radio-frequency cavities for each beam, which accelerate the protons from 450 GeV to their target center-of-mass energy: $7-8$ TeV in

2011 – 2012, and 13 TeV in 2015 – 2018. This was increased to 13.6 TeV in 2022.

At a rate of 40 MHz, the proton bunches ultimately collide at four interaction points inside particle physics detectors, associated to each of the experiments above. During specific periods of data-taking, one or both of the LHC beams can be filled with specific ion nuclei, as opposed to protons.

All of the LHC experiments have different goals, as explained earlier. The bottom line for most of them, however, remains the objective of making strenuous tests of our theory, the Standard Model, to try to poke holes in it that would act as experimental proof that there are still more boundaries to push in our understanding of the subatomic world. There are already some isolated pieces of evidence that the SM does not fully describe nature, though. In the following section we shall describe some, that will also serve as motivation for this thesis.

3.2 The LHCb detector

The LHCb detector [100, 102] was originally designed and optimized for the studies of particles containing b and c quarks. However, in its early days, it was found that it also acted as a factory of s quarks [103], prompting the Collaboration to focus on other types of physics as well. In particular, this thesis will introduce studies of decay channels that involve strange and charm mesons.

The original goal was to study *CP* violation in b and c decays, as well as rare decays involving b and c quarks. Over the years, studies involving rare strange decays, on top of semileptonic searches in b- and c-quark physics, heavy ion and fixed target collisions, electroweak processes, and exotica searches also turned out to be areas in which the LHCb detector, and by extension, the Collaboration, could contribute significantly to.

With time, it has evolved into a general physics detector [104, 105] with a special dedication to heavy flavor physics and expertise in the lower end of invariant-mass particles. While CMS [98] and ATLAS [97] can study the Higgs boson and processes happening in the > 20 GeV/c^2 invariant-mass region, the LHCb is especially interested in what happens below 10 GeV/c^2.

Its original goal of searching for *CP* violation in b-hadron decays is the main reason behind its geometry. Unlike its LHC counterparts (aside from Alice [99], which is solely focused on heavy ion physics), the LHCb is a forward spectrometer, with an angular coverage between 10 mrad and 300(250) mrad in the bending (non-bending) plane. In terms of pseudorapidity, a quantity defined as

$$\eta = -\log\tan\frac{\theta}{2} \tag{3.1}$$

that is often used to characterize particle physics detectors, the LHCb covers $2 < \eta < 5$.

This gives it a narrow window of common ground with other general-purpose detectors, but allows it to focus on a region in which the $b\bar{b}$ pairs are boosted, and thus the b-hadron decays are more displaced from the point in which they originated, allowing to be distinguished in an optimal way from other particles produced immediately after the collision (prompt).

In addition, the LHCb detector operates at a lower instantaneous luminosity than ATLAS and CMS, thanks to having a larger β^* (the amplitude modulation of the beam at the interaction point, a *de facto* measure of the beam size) than other LHC experiments. This yields less focused beams, but in turn allows the LHCb a much better reconstruction of the points in which the collisions took place, and subsequently, the points where short-lived particles produced promptly decay. These are referred to as primary and secondary vertices, respectively (PV and SV).

Over the years, the LHCb Collaboration has been able to observe the rare decay of the B_s^0 meson into a pair of oppositely-charged muons [106], has shown evidence of the pentaquark state [107, 108], observed decays of charmed baryons [109], and performed a wide range of studies of *CP* violation, including in decays from charm mesons [110] or the measurement of the *CP*-violating phase ϕ_s [111–114], among many other results. As discussed later in this thesis, it has also set world-best upper limits on very rare decays involving strange mesons [115, 116].

3.3 Data-taking

Data taking at the Large Hadron Collider is organized in periods of time, called Runs, in which the accelerator, as well as the individual detectors, do not suffer any major upgrades. Each Run lasts several years, with a months-long technical stop during the winter months. From year to year, minor updates to the electronics or the software may be implemented. The luminosity for each individual detector may also vary.

To date, two runs of data taking have been completed by the LHC. Run 1 span 2010-2012 [102], where LHCb collected 3 fb^{-1} of data at a center-of-mass energy of 7 and 8 TeV during 2011 and 2012, respectively. Run 2 started in 2015 and wrapped in 2018, tallying up 5.4 fb^{-1} [117] at $\sqrt{s} = 13$ TeV. In 2015, the integrated luminosity was 0.29 fb^{-1}, well below subsequent years. In this thesis, we shall only study data collected between 2016 and 2018. A sketch of the detector configuration during this data-taking period can be found in Fig. 3.2.

The LHCb Run 2 detector only records data such that at least one stable particle[*] produced in a given collision (event) satisfies at least one of multiple kinematic and multiplicity requirements. This is known as the hardware, or Level−0 (L0) trigger, and it is implemented at hardware level. Software-level selections are later introduced at the two High Level Trigger (HLT) stages. HLT1 includes a variety of lines (*i.e.*, a group of selections) that are of general purpose and that are implemented during data-taking, while HLT2 tends to get more decay-specific and is performed after the data has been collected[†].

Events for which at least one stable particle (or combination of) has successfully passed each of the three trigger levels is subsequently stored. There is an additional no-bias filter, where LHCb also stores one of $\sim 10^5$ events, with minimal trigger requirements implemented that simply ensure some multiplicity in the event. This is known as the Minimum Bias (MB) stream. All of this data is recorded on tape.

To access the data, further selections are implemented. A set of physics case-focused cuts is called *stripping*. A stripping line is able to isolate a specific decay from the rest of the data set, by combining the necessary particles and applying kinematic selections. A stripping line can also be inclusive of multiple decay channels, by simply imposing

---

[*] In LHCb, stable particles are muons, charged kaons, pions, electrons, photons, protons, and deuterons.

[†] In Run 3, the hardware trigger was removed and the detector is only collecting software-triggered collisions.

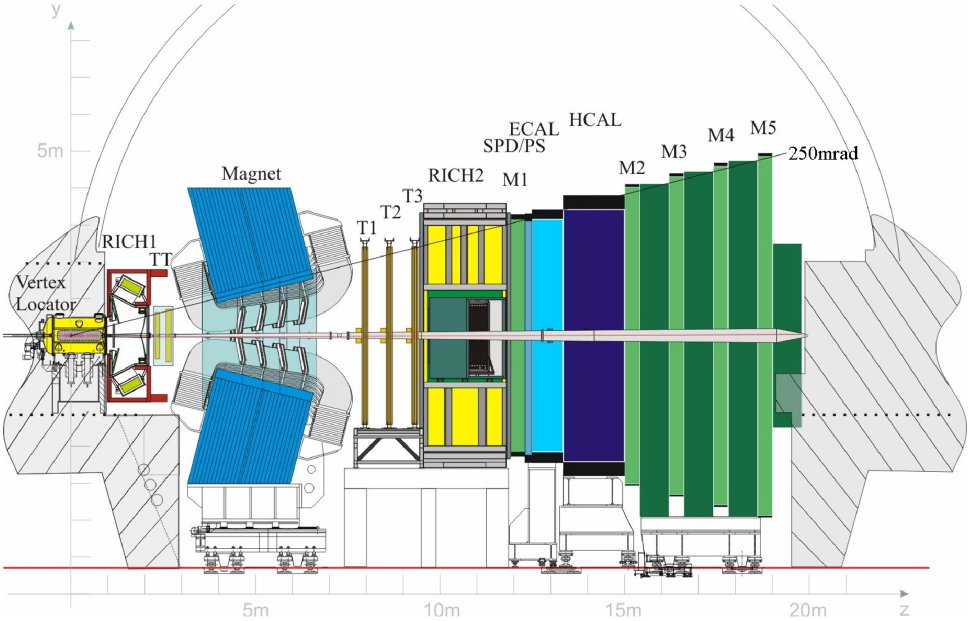


Figure 3.2: The LHCb detector used during Run 1 and 2 of data-taking, in its yz projection (non-bending plane). Figure taken from [118] under Creative Commons license.

requirements on combination of two or more particles that may later be grouped with further selections.

At the analysis stage, LHCb members have only access to data which has passed at least one stripping line. Updates on the stripping line set are performed periodically, leading to different versions of the stripping project for each year. In general, to study a decay at LHCb it must be available via one of these lines. Offline, further selections may be applied on the stripping-selected dataset.

3.4 The LHCb subdetectors

Specifics of the LHCb Run 1 and Run 2 detector are provided here, with a general overview of the changes that were implemented in Run 3, which started in 2022 and is planned to end mid-2026 [119]. The LHCb subdetectors include those dedicated to tracking reconstruction, to particle identification, as well as a dipole magnet.

3.4.1 The VErtex LOcator (VELO)

As it has already been mentioned, the LHCb experiment requires of a near-perfect reconstruction of the tracks produced immediately after the proton-proton collision, and this extends to the earliest moments after the bunch crossing. To achieve this, it needs an excellent tracking system surrounding the pp interaction point so that the primary vertices are known to analysts very precisely, as well as very good time resolution to account for CP-violating effects.

This is the VErtex LOcator (VELO) [120, 121], the first subdetector particles produced after the collision go through. In Run 1 and Run 2, it consisted of two halves with 21 stations each. Additionally, each of the stations had two silicon-strip half-circle modules,

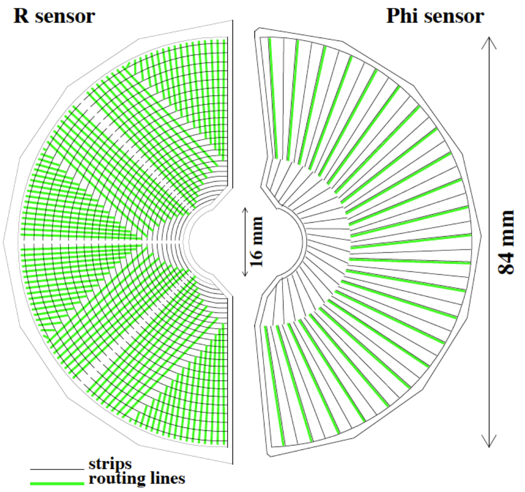


Figure 3.3: Close look at the $R - \Phi$ sensors. Taken from [121] under Creative Commons license.

surrounding the $\sigma = 5.3\,\mathrm{cm}$ interaction region. This is an aperture smaller than the one needed by the Large Hadron Collider when protons are being injected, and therefore the VELO must be retractable. The VELO must cover the angular acceptance of the detectors downstream of the magnet, too.

Each of the modules of the VELO contains an $R - \Phi$ sensor (see Fig. 3.3), which are then able to provide a (r, ϕ) set of coordinates for each particle leaving a *hit* as it flies through it. Simulations showed that a $r\phi$ geometry allows for a faster reconstruction in the LHCb software trigger with enough impact parameter resolution to efficiently select events with b-hadron [100].

The uncertainty on the primary vertex position is mainly driven by the number of tracks generated during a pp collision. On average, the resolution on the z-direction is $42\,\mu\mathrm{m}$ and $10\,\mu\mathrm{m}$ in the direction perpendicular to the beam [120]. Ignoring the influence of the PVs, the resolution of the impact parameter can reach $20\,\mu\mathrm{m}$ for large-momentum tracks.

The VELO sensors are separated from the LHC vacuum by a 0.5 mm-thin aluminium RF foil. This allows for shielding against RF pickup from the LHC beams, but it also introduces an additional layer of material that tracks must go through, which in turn degrades the resolution.

The VELO subdetector used during Runs 1 and 2 of data-taking was completely removed at the end of the second period and replaced with a brand-new piece of hardware. The new VELO, already in use for Run 3 of data taking, consists of 26 layers, each with two L-shaped modules [122].

3.4.2 Silicon Tracker

There are two subdetectors comprising the silicon tracking system at LHCb during Runs 1 and 2. Upstream of the magnet, and immediately after the VELO, the Tracker Turicensis (TT) [123] is composed of two rectangular modules with two layers each. On each of them, the silicon strips are arranged in the vertical direction, which enable a better momentum resolution in the bending plane. Each layer is reduced into different readout sectors, which allows for a reduction in the occupancy. The TT was replaced by a new subdetector called the Upstream Tracker (UT) at the start of Run 3 [124].

Downstream of the magnet there are three additional tracker stations with the purpose of measuring the position of the tracks after they are bent, T$1-3$. Based on occupancy, each of these stations is divided into two areas. Closer to the beam pipe, there is the Inner Tracker (IT) [125], while the Outer Tracker (OT) [126, 127] surrounds them. The T-stations were replaced in Run 3 by the Scintillating Fibre (SciFi) [124].

The process of track reconstruction, which uses information from the Silicon Tracker as well as the VELO, is explained in more detail in Sect. 3.7.

3.4.3 The dipole magnet

Measuring the momenta and the charge of the particles produced inside the LHCb is made possible thanks to the presence of a dipole magnet [128], with a bending power of $4\ \mathrm{T}\cdot\mathrm{m}$. The magnetic field is applied in the y direction, causing the particles' trajectories to be bent in the horizontal (x) plane.

This is also the reason why particle physics detectors are much better at reconstructing charged particles than they are neutral particles, like photons or neutrons. Neutral pions are often a challenge too. Analyses involving these particles can either attempt a full decay reconstruction or account for missing energy in the invariant mass spectra.

The dipole magnet also allows for the magnetic field to invert its direction, resulting in two different configurations, Up and Down, across data taking. This allows to minimize systematic biases associated that may be particularly important when studying CP violation.

3.4.4 The Ring Imaging Cherenkov detectors

The dipole magnet placed between the TT and the T$1-3$ stations bends the charged particles' trajectories, and allows for their reconstruction as tracks. At this point, however, the nature of each particle is not known to experimentalists. The particle identification (PID) is done via the subsystems placed after the tracking stations, with one exception.

The LHCb Run $1-2$ detector contains two Ring Imaging Čerenkov (RICH) detectors [129], one placed after the VELO (RICH1), and the other one after the T$1-3$ tracking stations (RICH2). See Fig. 3.4 for a scheme of both subdetectors. Čerenkov light is emitted when a particle travels through a medium with a velocity v that is higher than that of the light in that medium, c_n, where n is the refraction index. These emitted photons form a cone of angle θ given by

$$\cos\theta = \frac{c_n}{v\cdot n}. \tag{3.2}$$

RICH1 uses C_4F_{10} gas and covers the low-momentum range of $1-60$ GeV/c. RICH2, downstream of the magnet, uses CF_4 and covers the high-momentum range, $p \geq 15\,\mathrm{GeV}/c$. However, this second detector has an acceptance limited to the low-angle region, which means low-momentum particles will not go through. The radiation light is then redirected into photon detectors thanks to an array of mirrors. Using both detectors allows for PID determination in a wide momentum range, with the likelihood of correctly identifying charged hadrons being above 90% [130].

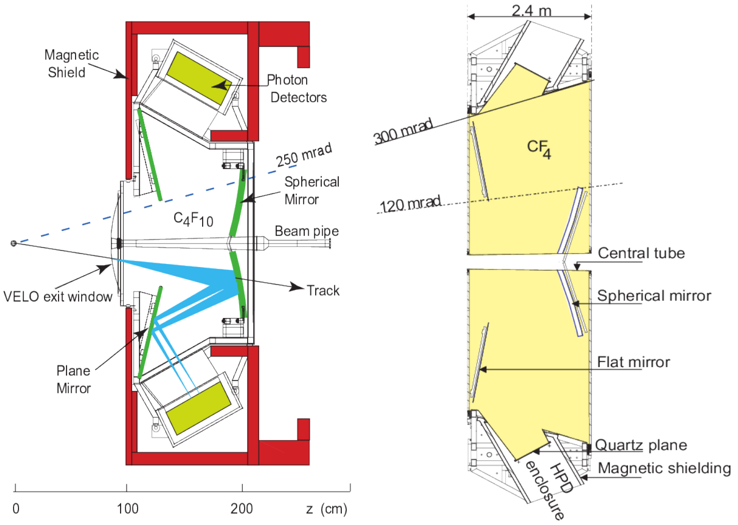


Figure 3.4: Schematic of the two RICH detectors responsible for particle identification at LHCb. RICH1 is placed downstream of the magnet, while RICH2 is placed upstream of the magnet. Figure taken from [131] under Creative Commons Attribution 4.0 license.

3.4.5 The calorimeter system

The calorimiter system of the LHCb [132, 133] is a set of subsystems in charge of retrieving the transverse energy of incoming hadrons, electrons, and photon candidates, on top of providing additional information for the identification of these particles, and measuring their energies and positions. This is necessary information by the detector's hardware trigger system (see Sect. 3.5) They also stop the incoming particles by placing the necessary layers of lead, forcing their electromagnetic showering in various stages.

Four different subdetectors are included inside the calorimeter system. First, the scintillator pad detector (SPD) and the preshower detector (PS) [134], separated by a lead layer that is 2.5 radiation lengths thick (12 mm), but around 0.06 hadronic interaction lengths, which means photons and electrons will start showering in the PS, but not hadrons. Both subdetectors help the calorimeter with good background rejection while maintaining a reasonably high efficiency when detecting photons, either prompt or from $\pi^0 \to \gamma\gamma$, and electrons. The SPD allows to identify charged particles and isolate electrons from photons. The PS helps with identification of electromagnetic particles.

The electromagnetic calorimeter (ECAL) [135] is placed immediately after the PS, and alternates between scintillator and lead absorber plates, which help it contain the whole electromagnetic showers; this technology is known as *shashlik* [136, 137]. The ECAL is 25 radiation lengths thick, and is able to fully stop electrons and photons. It contains a $65 \times 65\,\text{cm}^2$ hole at the center due to the substantial radiation around the beam pipe. The dimensions of the ECAL match those of the tracking system, with $\theta_x < 300\,\text{mrad}$ and $\theta_y < 250\,\text{mrad}$.

The ECAL is subdivided into inner, middle, and outer sections, with 176, 448, and 2688 modules, and 9, 4, and 1 cells per module, respectively. The difference in granularity is to accommodate the varying multiplicity depending on the azimuthal angle, such that the gain in calorimeter cells is uniform with $E_T = E \sin\theta$. The nominal resolution, determined through electron reconstruction during the CERN test beam in 2007, was determined to be [133, 138]

$$\frac{\sigma_E}{E} = \frac{(9.0 \pm 0.5)\%}{\sqrt{E}} \oplus (0.8 \pm 0.2)\% \oplus \frac{0.003}{E_T},$$

where E is given in GeV. The first term is the stochastic measure, the second is the constant contribution due to mis-calibration, non-linearities, etc., and the third is simply noise of the electronics.

The last piece of the calorimeter system is the hadronic calorimeter (HCAL) [139], made up of iron and scintillating tiles that act as absorber and active material, respectively. In the lateral direction, tiles are placed between 1 cm iron spacers; in the longitudinal direction, both of their lengths correspond to the hadron interaction length in steel. The scintillation light is transmitted to photo multipliers (PMTs) through wavelength-shifting fibers.

The HCAL is positioned at 13.33 m from the interaction point, and is 8.4 m high, 6.8 m wide, and 1.65 m deep ($5.6\lambda_{int}$). It is segmented into two regions, called inner and outer sections. Similar to ECAL, their varying granularity fits the different multiplicity expected in both regions based on the proximity to the beam pipe.

The baseline resolution was measured to be [140]

$$\frac{\sigma(E)}{E} = \frac{(67 \pm 5)\%}{\sqrt{E}} \oplus (9 \pm 2)\%,$$

where the first term is the stochastic contribution, and the second, the constant one. This was measured during beam-test results for a prototype of the calorimeter in the year 2000. A thicker HCAL, with $L_0 = 7.3\lambda_{int}$, showed only a slight improvement in the resolution.

3.4.6 The muon stations

The muon system of the LHCb detector [141, 142] is comprised of five tracking stations, $\mathrm{M}1-5$, arranged along the beam axis. M1 is placed before the SPD and allows to improve the p_{T} measurements in the trigger, with the other four stations situated after the HCAL and separated between them by three 80 cm iron absorbers. M1 uses the Gas Electron Multiplier (GEM) technology, to account for higher expected multiplicities, while the last four stations use multi-wire proportional chambers (MWPC) to detect charged particles. In particular, those that are able to traverse more than one muon chamber are most likely to be muons, due to their penetration power and minimum ionizing properties.

The total absorber thickness, from M1 to M5 (including the calorimeter) is of around $20\lambda_{int}$. The detectors provide binary information to the trigger processor by partitioning the detector into rectangular logical pads. Stations $\mathrm{M}1-3$ have a high spatial resolution and play a crucial role in track reconstruction as well as thee muon p_{T} estimation for the hardware trigger.

Stations $\mathrm{M}4-5$ have a lower spatial resolution, as their main purpose is that of particle identification. Only muons with $p_{\mathrm{T}} > 6\,\mathrm{GeV}/c$ are able to reach them. Each muon station is divided into four regions, $\mathrm{R}1-4$, of different granularity that also depends on the distance to the beam pipe.

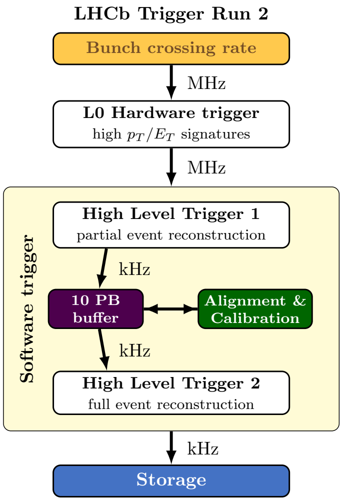


Figure 3.5: Overview of the LHCb trigger system in Run 2. Figure taken from Ref. [143] under CC-BY-4.0 license.

3.5 The LHCb Run 2 trigger system

During Run 2, the LHC ran at a center-of-mass energy of 13 TeV, which, for LHCb meant working at a frequency of 40 MHz. This means that every 25 ns the hardware trigger had to make a decision of whether to keep or discard an event (collision) based on limited information, since the detector cannot be read out at these high rates. The Level−0 trigger, L0, acts as a filter, and based on physical quantities like the transverse energy or momentum, along with the occupancy of the SPD chamber, it allows an event to pass through the High-Level Trigger (HLT) stage. Thanks to this filter, optimized to select typical signatures of b and c-hadron decays, the rate goes down to 1 MHz. Fig. 3.5 contains a visual scheme of how the Run 2 trigger system worked.

The L0 is divided into three independent triggers [143–145]. The muon trigger looks for straight-line tracks with a five-fold coincidence between all five stations within a time window smaller than 25 ns, to unanimously identify the bunch crossing.

The trigger decision can be based either the largest p_{T} value (as measured by their track direction) must be above the `L0Muon` threshold, or the product of the two largest p_{T} values must be above the `L0DiMuon` threshold. There are also targeted lines selecting events with low particle multiplicity, mainly with the goal of studying central exclusive production and inclusive jet trigger lines. Requirements are also established on the maximum number of SPD hits in most L0 trigger lines to reduce the complexity of events. Additionally, to minimize the chances of signals produced in the previous collision may last more than 25 ns, a requirement is set on the sum of the transverse energy of all of the clusters in the HCAL, `SumEtPrev` < 24 GeV [146], on all lines except for `L0DiMuon` during 2017 and 2018.

The hadron trigger, in the meantime, selects events containing particles with a high transverse energy deposit in the calorimeters and creates three types of candidates. The

`L0Hadron` candidate originates from a large E_{T}, 2×2-cell cluster in the HCAL, where

$$E_{\mathrm{T}} = \sum_{i=1}^{4} E_i \sin\theta_i, \tag{3.3}$$

such that E_i is the energy deposited in cell i and θ_i is the angle between the z-axis and a line from the cell center to the average pp interaction point. The `L0Photon` candidate is identified as the highest E_{T} cluster in the ECAL, while the `L0Electron` candidate requires, on top of that, at least one hit in an SPD cell.

Finally, the `L0PileUp` trigger is used for luminosity measurements.

Different trigger configurations are implemented throughout data-taking periods. They are labeled by Trigger Configuration Keys (TCKs), and they are useful offline to know under which trigger conditions the data was collected.

An event containing a track that has satisfied one of the L0 requirements is sent by the data acquisition network to a processor in the Event Filter Farm (EFF). There, the software-based High-Level Trigger (HLT) reduces the rate to a much more manageable 12.5 kHz.

This process is done in two steps. First, HLT1 performs a reconstruction on long tracks with a $p_{\mathrm{T}} > 500\,\mathrm{MeV}/c$. A precise reconstruction of the PV is also performed, and thanks to their clear signature, muons are also identified. Events passing at least one of the HLT1 algorithms are then stored in a buffering system. The HLT1 output rate is around 100 kHz.

The second step of the High-Level Trigger, HLT2, then uses the full detector information, including the RICH and the calorimeter systems, to perform a more complete track reconstruction. This was performed while the data-taking was on hold, meaning between LHC fills or during technical detector shutdowns [143]. This is also used to calibrate and align the detector for the next run.

HLT2 selections can either be inclusive, when they target events containing certain topologies, or exclusive, when they target specific final states. In the latter cases, only information from the signal particles is saved (as opposed to the full event), and it is fed directly into the *Turbo stream*. In the case of inclusive HLT2 lines, which all of the analyses in this thesis will use, the entire event is saved into the *Full stream* should it deliver a positive response on at least one of them.

To make the amount of data used by analysts much more manageable, an additional, fully offline selection step is introduced. This is called *stripping*, and it allows to filter through different physics cases by being optimized for each of them.

In Runs $1-2$, the majority of positive HLT2 lines were inclusive and most of the data was saved into the Full stream. This changed once Run 3 started, with approximately two thirds of the data being stored into the Turbo stream, following a collaboration-wide effort to vastly increase the number of HLT2 lines to target specific physics cases. A set of topological, inclusive lines was still kept to accommodate studies that were not considered before data-taking. Their information is then processed by a new version of the strripping pre-selection step, called *sprucing*. A visual scheme of the dataflow in Run 3 can be found in Fig. 3.6.

In Run 2, there were around 20 HLT1 lines, and 500 HLT2 lines [143]. Roughly 40% of the trigger output rate is dedicated to inclusive topological trigger lines, with another 40% dedicated to exclusive c-hadron trigger lines. The rest is divided among dimuon, electroweak physics, exotica, and exclusive trigger lines.

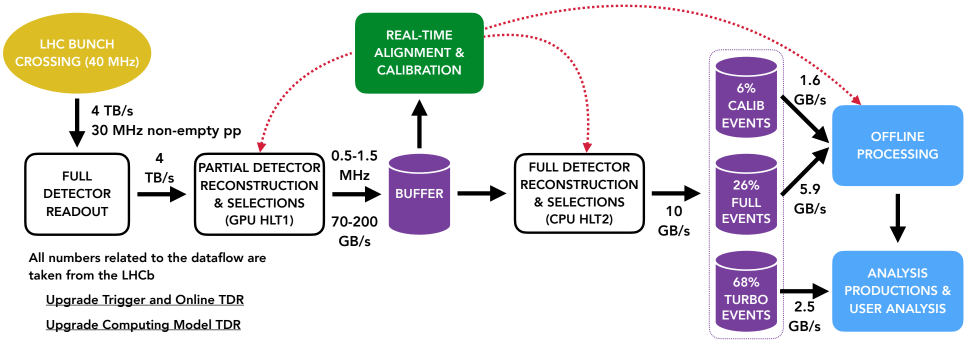


Figure 3.6: The LHCb upgrade dataflow, from bunch crossing to real-time analysis to offline analysis of the data. Figure taken from [147] under Creative Commons license.

The specific selections of the HLT lines used in the analysis presented in this thesis are included in the corresponding sections.

3.5.1 TIS and TOS

Information on which part of the detector fired a certain decision is kept. In LHCb jargon, we define triggered-on-signal (TOS) candidates when one of the signal particles (or group thereof, in specific cases) was responsible for yielding a positive trigger response. Whenever another particle outside our signal candidate did so, we speak of an event triggered independently of signal (TIS). There is a third option that can be most common in the case of multi-particle selections, where neither the signal candidate alone, nor the rest of the event by itself, are enough to justify the firing of the trigger. This is called Triggered-on-Both (TOB), but given its much lower resolution when compared to TIS and TOS, it is hardly ever used in analyses.

The TIS and TOS form a Venn diagram, such that there will be candidates that are exclusively triggered on signal, where only particles that fired the trigger are part of the signal candidate, candidates that are exclusively triggered independently of signal, where no particle in the signal candidate fired the trigger, and candidates where one ore more signal particles, on top of one or more non-signal particles, fired the trigger.

The efficiencies of the different trigger stages are computed through simulation. However, we know that we cannot perfectly model our detector's response, which often means that the simulation has to be validated and corrected for.

Trigger efficiencies often need to be ratified through real data. And though this process may not be obvious, Ref. [148] provided the steps to cross-check them. Take

$$\varepsilon_{Trig} \equiv \frac{N_{Trig|Sel}}{N_{Sel}} \tag{3.4}$$

where N_{Sel} is the number of events that have passed a given selection, and $N_{Trig|Sel}$ is the number of events that have passed a trigger decision after having passed the aforementioned selection. We can also define the following efficiencies:

$$\varepsilon_{TOS} \equiv \frac{N_{TOS|Sel}}{N_{Sel}} \quad , \quad \varepsilon_{TIS} \equiv \frac{N_{TIS|Sel}}{N_{Sel}} \quad , \quad \varepsilon_{TISTOS} \equiv \frac{N_{TISTOS|Sel}}{N_{Sel}}, \tag{3.5}$$

where $N_{TIS|Sel}$, $N_{TOS|Sel}$, $N_{TISTOS|Sel}$ are the number of events that have been flagged as TIS, TOS, or TIS *and* TOS, respectively.

In these definitions, N_{Sel} is not directly measurable on data, since only events that have passed the trigger are available, not the number that has been processed. However, one can play around with these definitions:

$$\varepsilon_{Trig} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \frac{N_{TIS|Sel}}{N_{Sel}} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \varepsilon_{TIS} \tag{3.6}$$

Assuming that the TIS efficiency of any subsample of the triggered events is the same as that of the whole subsample of selected events (in particular, the TOS subsample),

$$\frac{N_{TISTOS}}{N_{TOS}} \equiv \varepsilon_{TIS} \simeq \varepsilon_{TIS|TOS}, \tag{3.7}$$

and therefore,

$$\varepsilon_{Trig} = \frac{N_{Trig|Sel}}{N_{TIS|Sel}} \times \frac{N_{TISTOS}}{N_{TOS}} \tag{3.8}$$

All four of these quantities are directly measured on data. This method of validating the trigger efficiencies is called the TISTOS method and will be used in this thesis to estimate systematic uncertainties associated to trigger efficiencies.

3.6 LHCb simulation

The use of simulation is an essential part of any high-energy physics analysis. If we are looking for extremely rare signatures, we need to have an accurate expectation of what we are seeking, to perform a good reduction of the surrounding background contributions. We also need a good modelling of how our detector will respond to the passing of particles through it, as well as have a good sense of how the geometry will influence our efficiencies.

The process of simulating the signal channels in this thesis starts with the use of PYTHIA, which generates proton-proton collisions [149]. In particular, we use a specific LHCb configuration, the GAUSS framework [150], and the EVTGEN package [151] to reproduce the decays of hadronic particles, and force the decays of certain particles to a given final state. Final-state radiation is generated using PHOTOS [152].

EVTGEN may use several models to force the decays for each random generated event. The simplest option is the `PHSP` model, which generates instances such that the available phase space, as given by the mass of the mother particle, is uniformly populated. The angular distribution of the decay products is flat, and all spins of particles in the initial and final states are averaged.

The presence of vector mesons, such as the ϕ, introduces the need to use a model that takes into account the spin of the mother particle. Specifically, the $\phi \to \mu^+\mu^-$ has a special relevance to this thesis. The `VLL` model, which tackles decays of a vector meson to a pair of charged leptons, was used to properly model this decay. Custom models, covering special cases where the angular distributions may require of a specific theoretical description, can also be included. The $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ decays, studied in this thesis, needed special attention in this regard.

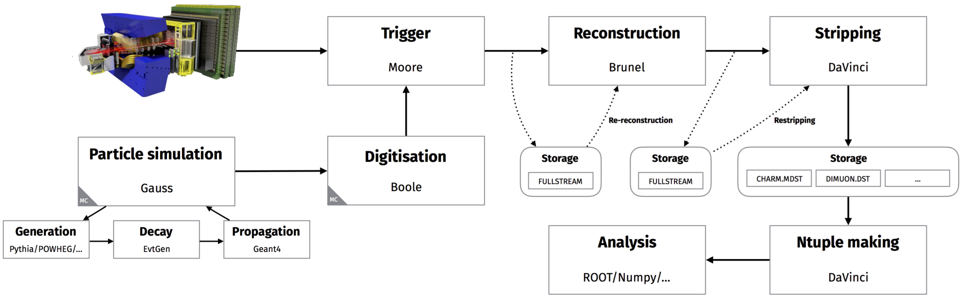


Figure 3.7: The full dataflow for LHCb during Run 2, combining the simulation process with data-taking. Figure taken from Ref. [158] under CC-BY-4.0 license.

This is called the generation process. To maximize the use of the resulting dataset, for instance to not include events which may fall outside of our acceptance, it is common to also include cuts at this level that will ignore instances that do not satisfy them.

The detector is then implemented through the GEANT4 package [153, 154], which includes a good modeling of our detector [155] under each of the conditions that were in place throughout the data-taking process, and is able to properly reproduce the interaction between particles and matter.

The simulation framework used in all of the analyses in this thesis is Sim09. This was replaced in 2022 by Sim10, with the main difference being the upgrade from GEANT4 v9 to GEANT4 v10. By the time the Collaboration transitioned into Sim10, one of the analysis had already been concluded, and the simulation for the other had already been requested.

The digitalization of our detector's response is performed by BOOLE [156, 157], a software that reads in the output of the detector simulation by GAUSS, adds hits from the Spillover events and LHC background, and applies the detector response. The output is digitized data that mimics the real data coming from the actual detector, and subsequent steps match the processing chain of actual collected data. The full dataflow, combining data-taking with simulation, that was implemented in Run 2, can be seen in Fig. 3.7.

In simulated data, we are able to access a lot more information than in data. For instance, we know exactly the *true* nature of our particles. But we also know that our reconstruction process is not perfect, and sometimes particles of a specific type are reconstructed as something else. This originates some background contributions in our MC, which ideally should be a clean signal.

The process of truth-matching consists in making sure that the particles that were reconstructed as a certain type are, in fact, of that nature. We do this by making sure their TRUEID label, assigned at the generation label as the identifier of the nature of the particle, matches that of the type that the track was reconstructed as. Information from the decay chain, including the identifiers of the mothers, may also be brought into the process.

3.7 Track reconstruction and vertexing at LHCb

As particles produced in the aftermath of pp collisions fly through the LHCb detector, they leave hits in the different tracking stations, which include the VELO, the T-stations, and the muon system. Track reconstruction is the process of combining those hits to make a picture of the trajectories that charged particles (tracks) followed after being produced.

In LHCb, a track combines a series of track states, which are defined by a state vector, $\vec{S} = (x, t_x, y, t_y, q/p)$, as well as a 5×5 covariant matrix. In the definition of $\vec{S}$, x and y are the coordinate positions of the track, while

$$t_x = \frac{\partial x}{\partial z} \quad , \qquad t_y = \frac{\partial y}{\partial z}$$

are the slopes of the track at a given z position; q is the charge, as given by the curvature of the track due to the magnet bending, and p is the momentum. The process to form this vector is as follows:

1. **Pattern recognition**: Hits likely to have come from the same traversing particle are grouped together. In the VELO, for instance, the iterative process begins by looking at pairs of hits on the VELO modules that are the furthest away from the interaction region. A straight-line extrapolation is then used to search for additional hits in earlier modules, and form what is known as a *track seed*. Track seeds with at least three hits are passed on to the next step.

2. **Track fitting**: VELO seeds are then associated to hits in the tracking stations to form tracks. Starting with a hit in one of the T stations, and neglecting the residual magnetic field, track candidates are confirmed by looking for narrow clusters with the same x position. (Low-momentum tracks, however, would produce much wider clusters, and the straight-line approximation does not hold as well.)

   The presence of the magnetic field and the bending of the tracks also allows for a determination of q and p. This is how long tracks are built (see below), and the associated hits are removed from the algorithm. The algorithm is called forward tracking. The alternate seeding algorithm starts from hits in the tracking stations and back-propagates to VELO seeds.

   Once the track is found, it is refitted using the Kalman filter (see Sect. 3.9 and Refs. [159, 160] for more details). Multiple algorithms might try to reconstruct the same physical track, which results in clone tracks.

3. **Cleaning up**: Poorly reconstructed tracks, as well as clone tracks are discarded. Fake tracks can also occur. Known as ghosts, they would be combinations of random, unrelated hits. They are also discarded here, but neural networks under the names of `TrackGhostProb` and `ProbNNghost` [161] can also take care of them during analysis.

The LHCb classifies tracks based on hits in the different tracking stations (see Fig. 3.8 for a visual explanation):

1. **VELO tracks**: They only leave hits inside the VELO, and fall out of LHCb acceptance immediately after.

2. **Upstream tracks**: They leave hits in the VELO and the TT, but do not leave hits beyond the magnet.

3. **Downstream tracks**: They have no hits in the VELO, only in the tracking stations. They can be reconstructed using the seeding algorithm.

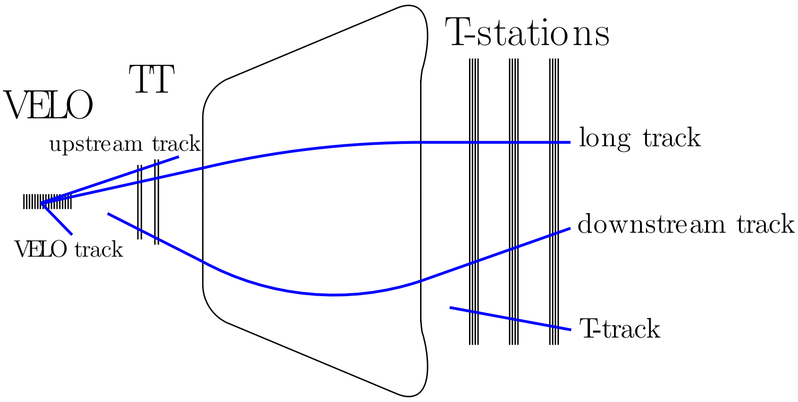


Figure 3.8: Different types of tracks defined for LHCb, based on the information collected by the different tracking subdetectors. Figure taken from Ref. [143] under CC-BY-4.0 license.

4. **T-tracks**: They only have hits in the T-stations stations. They can also be reconstructed via seeding.

5. **Long tracks**: They have hits on all tracking subdetectors.

All analyses present in this thesis use long tracks, as do the majority of studies at LHCb. This is due to a much higher momentum resolution, as well as their 95% reconstruction efficiency [158, 162]. This is evaluated using control channels such as $J/\psi \to \mu^+\mu^-$.

VELO tracks, however, are used to determine very precisely the location of the primary vertex. This is another important step of the reconstruction process at LHCb, and of what is known as vertexing, which consists on finding the 3D positions where the different particles present in the event were originated (vertices).

Secondary vertices (SVs) are the decay postions of unstable particles produced directly at the PV. They can be found by combining track candidates and fitting them to a common origin. The proper-time resolution achieved during Run 2 was of the order of 45 fs [143], which enabled the measurement of the B_s^0 mass oscillations (the lifetime of the B_s^0 is $\tau_{B_s^0} \sim 1500$ fs.)

3.8 Particle identification at LHCb

One of the great challenges of the LHCb experiment is correctly identifying the nature of the particles that are measured by the detector. Particle identification (PID) is crucial to properly distinguish between exclusive decay modes of an unstable particle. To perform this task, LHCb has developed over the years several techniques and algorithms based on the response of certain subdetectors.

The main challenge here is the high track multiplicity and background levels. For instance, this makes the identification of separate rings inside the RICH detectors rather complicated. The association between tracks and rings reconstructed inside the RICH system, combining information from other subdetectors, is done by a specialized algorithm that assigns a likelihood that each track is of a particular type.

Table 3.1: Muon stations required to trigger the `IsMuon` decision as a function of the momentum of the particles.

Momentum range	Muon stations
$3\ \text{GeV}/c < p < 6\ \text{GeV}/c$	M2+M3
$6\ \text{GeV}/c < p < 10\ \text{GeV}/c$	M2+M3+(M4.or.M5)
$p > 10\ \text{GeV}/c$	M2+M3+M4+M5

This likelihood is then compared to that of the pion, the most abundant particle produced at LHCb [102]. The logarithm of the ratio of the likelihoods is the basis for the `DLL` variables,

$$\texttt{DLLX} = \log\left[\frac{\mathcal{L}(X)}{\mathcal{L}(\pi)}\right]. \tag{3.9}$$

Neural networks (specifically, a multilayer perceptron with one hidden layer) were also trained to assign a $0-1$ probability that a track is of a given type, `ProbNNX` [102]. This is, in general, a more powerful discriminant than the difference in likelihoods.

A boolean variable, `IsMuon`, is also defined for muons with a momentum higher than $3\,\text{GeV}/c$ [163]. Based on their total momentum, the variable yields `True` if the track has hits on certain muon stations. See Table 3.1 for the varying thresholds needed to flag `IsMuon`. This is the most basic kind of muon identification.

3.8.1 PID Calibration

The response of the particle identification algorithms implemented at LHCb is not perfectly modeled in simulation. This leads to analysts having to implement offline corrections to the MC samples to make sure that the studies performed through them are as realistic as possible.

These corrections can take place in several ways. In cases where a simple, rectangular cut on a PID variable is implemented, correcting the efficiency of this selection should suffice. More complex cases, like those using particle identification in a multivariate classification, may require a full transformation of the corresponding variable.

For either procedure, data from control channels is selected without any PID requirements [164]. For muon identification, $B^+ \to J/\psi(\mu^+\mu^-)K^+$ candidates are used, while for pion identification, the primary source of calibration samples comes from $D^{*+} \to D^0(K^-\pi^+)\pi^+$. The residual background that may not be rejected through fiducial requirements is statistically subtracted through the *sPlot* technique [165].

`PIDCalib` [166, 167] uses the tag-and-probe method to compute the efficiencies. This involves tagging one of the tracks (one of the muons from the J/ψ or the soft pion from the D^{*+}) that is well identified, and using a list of probe tracks as our PID candidate, which has no particle identification selection applied. The efficiency of a given selection, for instance, `IsMuon`, will be given by the number of probe candidates that pass it over the number of total candidates. This efficiency is weighted by the `sWeight` variable provided by the *sPlot*.

`PIDGen` [167], instead, creates a new transformed variable by matching the (`sWeight`-ed) cumulative distribution functions of the PID variable to those in our own data sets.

Instead of performing these procedures on the full sample, both of these packages divide the calibration samples into kinematic bins, and perform their respective tasks inside each bin. The corrected efficiencies of our reference samples are then computed per bin, as are the transformed values of the PID variable.

These corrections will be applied on all of the analyses presented in this thesis.

3.9 Decay Tree Fitter

Let us consider decay chains that contain two generations, meaning a mother particle that decays into an intermediate, short-lived state, which then decays into some daughter particles. In high-energy physics detectors, the signature of this decay is reconstructed from the bottom up. We first focus on the final-state particles, and reconstruct the vertex where they originated. We then use the information from the intermediate state to reconstruct the upstream decay.

The parameters of the mother particle at each of these steps (*i.e.*, the resonance in the first case, and the mother at the top of the chain in the second case) are taken via a least-squares fit to the daughter particles, with the constraint that they originate from the same point. However, when doing the first fit, information from the upstream of a decay vertex is not implemented, despite the fact that it could be useful to improve the quality of the downstream fit.

Ref. [168] introduced an implementation of a least-squares fit that extracts the full extent of the parameters of a decay chain simultaneously, using a Kalman filter [159, 160], which they argued is a suitable technique to extract the necessary parameters of a decay chain (vertex positions, momenta, and decay times) and the corresponding covariance matrix from the external constraints. It was originally developed for the BaBar experiment, but has since been implemented by LHCb as well, under the name `DecayTreeFitter` [169].

The idea is to have a model whose parameters x are minimized via the function χ^2. This function contains the framework constraints, which can be external (like measurements of our data; they carry uncertainties) or internal (conservation of momentum at a decay vertex, for instance; these are exact). The Kalman filter helps with the minimization of the χ^2, which can get quite complicated.

For a measurement constraint k, we take an iterative process. We assume that the χ^2 with all the $k-1$ previous constraints has already been minimized, which leads to a prediction of x_{k-1} with covariance C_{k-1}. With this,

$$\chi^2_k = (x - x_{k-1})^T C^{-1}_{k-1}(x - x_{k-1}) + (h_k(x) - m_k)^T V_k^{-1}(h_k(x) - m_k),$$

where $h_k(x)$ is the measurement model, V_k is the covariance of the constraint k, and m_k is the measured quantity. We get the updated x_k by solving

$$\frac{\partial \chi^2}{\partial x} = 0 \Leftrightarrow C^{-1}_{k-1}(x - x_{k-1}) + H_k^T V_k^{-1}(h_k(x) - m_k) = 0,$$

where we have defined $H_k^T \equiv \frac{\partial h}{\partial x}\big|_{k-1}$.

We can solve the equation by assuming a linear model, $h_k(x) = h_k(x_{k-1}) + H_k(x - x_{k-1})$. We get:

$$x = x_{k-1} + K_k r_k^{k-1},$$

where

$$K_k = C_{k-1} H_k^T (R_k^{k-1})^{-1} \quad , \qquad r_k^{k-1} = m_k - h_k(x_{k-1}) \quad , \qquad R_k^{k-1} = V_k + H_k C_{k-1} H_k^T.$$

We can also define the covariant matrix as

$$C_k = C_{k-1} - K_k(2H_k C_{k-1} - R_k^{k-1} K_k^T),$$

and the χ^2 contribution is given by

$$\chi_k^2 = (r_k^{k-1})^T (R_k^{k-1})^{-1} r_k^{k-1}.$$

Perhaps the most interesting part of the `DecayTreeFitter` method is that, as indicated above, it allows us to exert certain constraints on our decay tree. These are the internal, or exact constraints, which we implement through Lagrange multipliers, $g(x)$. This can be momentum conservation at a decay vertex, a geometrical constraint, like matching the end position of a mother to the beginning of the daughter's reconstructed track, forcing the mother of the tree to come from the primary vertex, or even assign a mass hypothesis to it.

The process is similar to the measurement constraints, where we now have to assume a linear model on our Lagrange multipliers,

$$g_k(x) = g_k(x_{k-1}) + G_k(x - x_{k-1}) + ... \quad ; \qquad G_k \equiv \frac{\partial g}{\partial x}$$

The χ^2 contribution is given by

$$\chi_k^2 = (x - x_{k-1})^T C_{k-1}^{-1} (x_k - x_{k-1}) + 2\lambda_k^T g_k(x).$$

The minimization of this function yields

$$x_k = x_{k-1} - K_k g_k(x_{k-1}) \quad ; \text{where} \quad K_k = C_{k-1} G_k^T (G_k C_{k-1} G_k^{-1})^{-1}.$$

The negative sign in x_k is purely conventional. Finally, the covariant matrix, plus χ_k^2 are given by

$$C_k = (1 - K_k G_k) C_{k-1} (1 - K_k G_k)^T \quad ; \qquad \chi_k^2 = g_k(x_{k-1})^T (G_k C_{k-1} G_k^T)^{-1} g_k(x_{k-1}).$$

The full χ^2 function to be minimized during the fit is given by $\chi^2 = \sum_k \chi_k^2$.

We find that both types of constraints are equivalent under the transformations

$$h(x_{k-1}) - m_k \rightarrow g_k(x_{k-1}) \quad ; \qquad V_k + H_k C_{k-1} H_k^T \rightarrow G_k C_{k-1} G_k^T.$$

This method is implemented by LHCb as a way to improve the vertex fit in certain circumstances. It will be used in this thesis to improve the resolution of resonant mass peaks, which will help differentiate between two peaks close to each other.

4

Objectives and methodology

4.1 Objectives

This thesis is an extensive study of three of the lightest mesons that exist in the Standard Model (SM) of Particle Physics, and in particular, a search for some of their rarer decay modes into known particles.

In the case of Kaon studies, we attempt to measure for the first time the probability that a neutral, short-lived kaon, K^0_S , decays into four muons. This process is highly suppressed in the SM, with the chances of it happening estimated to be around 1 in 100 trillion. If the decay is not observed, an upper limit on its branching fraction will be set. An observation of the decay, and the determination that the probability of happening is higher than what the SM predicts would result in a direct experimental observation of Physics Beyond the Standard Model (BSM).

One of the longest-standing tensions between theory and experiment in particle physics is the disagreement in the value of the anomalous magnetic dipole moment of the muon, g_μ. Uncertainties in theoretical calculations involved in computing this quantity could be minimized thanks to a better understanding of the rare decays of η and η' decays.

The $\eta \to \mu^+\mu^-$ and $\eta' \to \pi^+\pi^-\mu^+\mu^-$ decays have both been observed experimentally, but the uncertainties in their branching ratio measurements are still rather large. This thesis aims to improve these studies. On the other hand, the decay of an η' meson into two muons has never been analyzed in particle physics experiments. Its chances of happening are estimated to be around one in 10 million. Its observation, which this thesis aims to do, would be considered a major feat in the study of these light mesons. Additionally, the even-less-likely process of an η meson decaying into two pions and two muons has also never been observed. It will also be under study in this thesis.

Finally, when studying the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ channel, searches for potential bumps in the $\mu^+\mu^-$ mass distribution that would correspond to a hypothetical scalar particle decaying into two muons, will be performed.

4.2 Methodology

To perform these measurements, the thesis will use data collected after proton-proton collisions inside the LHCb experiment of the Large Hadron Collider (LHC) between 2016 and 2018.

The analysis of $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ decays will be done with respect to $K^0_\mathrm{S} \to \pi^+\pi^-$ decays, one of the most common processes that happen in the detector. The analysis consists of implementing several selections on our variables, which can be either topological or multivariate, to reduce the combinatorial background contributions that may be hiding our signal. With a proper examination of the efficiencies of each of these selections, and after carrying out a detailed overview of the different systematic uncertainties that the methodology may carry, a global fit to the dataset, accounting for both a hypothetical signal peak and the background over which it may be standing, will provide us with the result of either a value of the $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ branching fraction or an upper limit for it.

In the case of the η and η' analyses, we will look for these mesons when they originate from a heavier charm meson that can be either a D^+_s or a D^+. The measurements will be done with respect to $D^+_{(s)} \to \phi(\mu^+\mu^-)\pi^+$ decays, which are also quite abundant at LHCb and well understood. The biggest challenge of these four analyses is isolating the $\eta' \to \mu^+\mu^-$ signal from the dominant $\phi \to \mu^+\mu^-$ peak in the $\mu^+\mu^-$ invariant mass distribution. In addition to constraining the combinatorial background leaks through different selection steps, similar to those used in the K^0_S analysis, detailed studies of the different systematic uncertainties that may arise will be carried out.

The procedure followed will be similar for the rest of the η and η' decays analyzed. The search for hypothetical, neutral particles decaying into a di-muon pair will also be performed in the fit of the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ channel.

5

Neutral kaon decays into four muons

5.1 An introduction to kaon mesons

Kaons [22] are the lightest possible mesons containing a s quark [170, 171]. The bound state $u\overline{s}$ ($s\overline{u}$) forms a K^+ (K^-), while $K^0 = [d\overline{s}]$, $\overline{K}^0 = [s\overline{d}]$. The weak interaction provides a mechanism through which neutral kaons can mix via $K^0 \leftrightarrow \overline{K}^0$ box diagrams (see Sect. 8.6.3 of Ref. [172] and Sect. 14.4 of Ref. [29], among others).

Because of this mixing process, a neutral kaon produced as a K^0 will also develop a $\overline{K}^0$ component, and the physical neutral kaon states are the stationary states given by the Hamiltonian of the whole $K^0 - \overline{K}^0$ system, which includes the weak interaction mixing Hamiltonian. Neutral kaons, thus, propagate as linear combinations. The physical states we see are known as the K-short (K^0_S) and the K-long (K^0_L). Their main differentiating factor is their lifetimes,

$$\tau(K^0_\mathrm{S}) = 0.89540(40) \times 10^{-10}\,\mathrm{s} \quad ; \quad \tau(K^0_\mathrm{L}) = 0.5116(21) \times 10^{-7}\,\mathrm{s}. \tag{5.1}$$

Kaon decays provided the first experimental evidence of CP violation [173], and remain an excellent probe in this field, since this is not an exact symmetry in the decays of these mesons. This can be used to demonstrate why the two most probable decays of the K^0_S are to $\pi^+\pi^-$ and $\pi^0\pi^0$ final states, while K^0_L decays mostly to $\pi^+\pi^-\pi^0$ and $\pi^0\pi^0\pi^0$ (see Sect. 14.4.1 from Ref. [29]).

Because of their low mass, $m(K^0) = 497.611(13)\,\mathrm{MeV}/c^2$, kaons can only decay to pions, muons, electrons, and neutrinos. Furthermore, kaon decays to a fully leptonic final state are forbidden by the SM at tree level, and are highly loop-suppressed, given they are flavor-changing neutral currents.

One of the most interesting and appealing aspects of kaon physics is that most of the decay modes have branching fractions of the order of 10^{-10} and below, according to the

Standard Model. This could make them more sensitive to BSM dynamics than other golden channels at LHCb, like $B_s^0 \to \mu^+\mu^-$ or even $B^0 \to \mu^+\mu^-$.

$K^0_\mathrm{S} \to \mu^+\mu^-$, for instance, is predicted to occur in the SM with a branching fraction of $\mathcal{O}(10^{-12})$, which makes it very sensitive to various BSM frameworks, including leptoquark models [174] and supersymmetry [175].

5.1.1 Very rare kaon decays

In general, kaon decays that involve flavor-changing neutral currents or loop diagrams tend to be a combination of short- and long-distance contributions (SD and LD, respectively) [176–178], depending on whether the intermediate virtual particle has a mass above or below the c quark, respectively. The SD component is governed by weak interactions and involves loop diagrams. The $V_{ts}^* V_{td}$ term in its effective Hamiltonian causes a strong suppression.

Meanwhile, the LD component arises from intermediary states, especially when they are mediated by photons. In the case of $K^0 \to \mu^+\mu^-$, for instance, we would have the kaon decaying into two virtual photons, which then decay into a dimuon pair. Since K^0_L is a *CP*-odd state, it has a larger coupling to two photons, which means the LD contribution is dominant in its decays. The LD term is highly negligible in K^0_S decays.

Ref. [179] used $K^0_\mathrm{L} \to \gamma\gamma$ to introduce a form factor that would help with the theoretical disentanglement of SD and LD,

$$\mathrm{F}\left(q_1^2, q_2^2\right) \equiv \mathrm{F}(0,0)\left[1+\alpha\left(\frac{q_1^2}{q_1^2-M_V^2}+\frac{q_2^2}{q_2^2-M_V^2}\right)+\beta\frac{q_1^2 q_2^2}{\left(q_1^2-M_V^2\right)\left(q_2^2-M_V^2\right)}\right]. \tag{5.2}$$

The parameters α and β can be retrieved experimentally via $K^0_\mathrm{L} \to \gamma\ell\ell$ and $K^0_\mathrm{L} \to e^+e^-\mu^+\mu^-$, and satisfy $1+2\alpha_L+\beta_L = 0.3$, in the case of K^0_L decays, and $1+2\alpha_S+\beta_S = 0$ in the case of K^0_S .

From a theoretical point of view, one of the main challenges in kaon physics at the moment is the determination of the sign of the amplitude $\mathcal{A}(K^0_\mathrm{L} \to \gamma\gamma)$ [180, 181], which is not clearly reachable by the theory due to the cancelation of the low-order terms. This affects, for instance, the computation of $\mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-)$ [179].

Experimentally, the sign is not an observable on its own, but rather relative to other interfering amplitudes. Branching fraction measurements are sensitive to the magnitude of the decay, which is proportional to the square of the amplitude. A precise determination of $K^0_\mathrm{L} \to \mu^+\mu^-$, however, could potentially add more clarity to it.

The sign represents a constructive or destructive interference between the short- and long-distance contributions in kaon decays, and could become relevant when studying *CP* violation experimentally [181, 182].

Taking the $K^0 \to \mu^+\mu^-$ picture a bit further, we can also study $K^0 \to \mu^+\mu^-\mu^+\mu^-$ decays, which are even more suppressed FCNCs. The Feynman diagrams describing the interaction are presented in Fig. 5.1. According to the SM, and depending on the values of α and β [183],

$$\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-) = (1-4)\times 10^{-14} \quad , \quad \mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-\mu^+\mu^-) = (5-9)\times 10^{-13}. \tag{5.3}$$

The current chapter will later present the first-ever experimental search for this decay, performed at LHCb and which appeared in arXiv in Dec. 2022 and which was published in Phys. Rev. D in August 2023 [116].

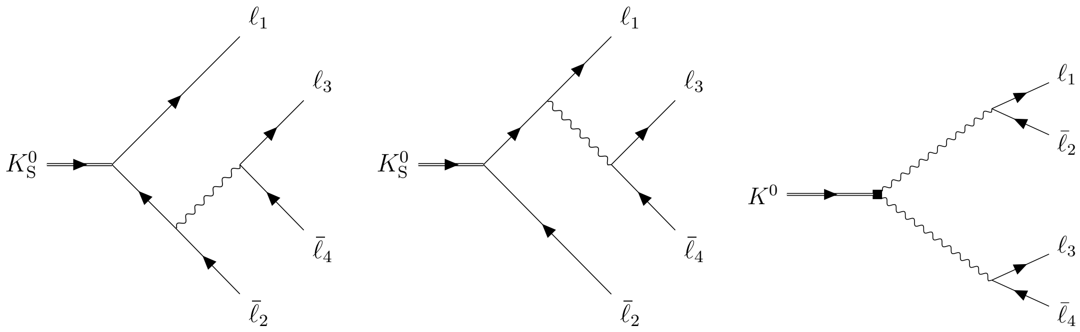


Figure 5.1: Feynman diagrams contributing to $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$.

5.1.2 $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ in BSM scenarios

An extension of the SM given by a dark matter (DM) model proposed in Ref. [84] could allow for an enhancement in the branching fraction of $K^0 \to \mu^+\mu^-\mu^+\mu^-$. The model adds a Higgs-like DM boson, Φ, to the SM Lagrangian, and also involves a dark charge-neutral scalar S:

$$\begin{aligned}\mathcal{L}_\mathrm{DS} = &\left|D^\mu\Phi\right|^2 - \frac{1}{4}F^{\mu\nu}_d F_{d\mu\nu} - \frac{\varepsilon}{2}F^{\mu\nu}_d F_{\mu\nu} - \mu^2\left(\Phi^\dagger\Phi\right) - \lambda\left(\Phi^\dagger\Phi\right)^2 - \lambda_d\left(\Phi^\dagger\Phi\right)\left(H^\dagger H\right)\\ &+ \partial_\mu S\partial^\mu S - m_S^2 S^2 - A_H H^\dagger H S - A_\phi \Phi^\dagger\Phi S.\end{aligned} \tag{5.4}$$

The inclusion of the S scalar is enforced by the presence of the quartic coupling term, $\lambda_d(\Phi^\dagger\Phi)(H^\dagger H)$, which can be shown to prevent any signatures in $K^0 \to A'A'$ in a purely 4-dimesional Higgs portal model. In the $\lambda_d \to 0$ limit, the only SM-DM portals are the kinetic mixing ε, which allows $A' \to \ell\ell$. Additionally, the couplings $A_{H,\phi}$ mediate the production on dark photons in kaon decays.

We can compute from the corresponding Feynman diagram the following expression,

$$\mathcal{B}\left(K^0_\mathrm{L} \to A'A'\right) = 5\times10^{-8} \times \left(\frac{\mathcal{B}\left(h \to A'A'\right)}{10\%}\right) \times \left(\frac{m_K^4}{\left(m_K^2 - m_S^2\right)^2 + m_S^2\Gamma_S^2}\right) \times f\left(\frac{m_{A'}}{m_K}\right). \tag{5.5}$$

Considering benchmark values of $m_{A'} = 230$ MeV/c^2 and $m_S = 800$ MeV/c^2, we get $\mathcal{B}\left(A' \to \mu\mu\right) = 0.22$, and

$$\mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-\mu^+\mu^-) = 2.5\times10^{-10}. \tag{5.6}$$

Since

$$\frac{\mathcal{B}\left(K^0_\mathrm{S} \to A'A'\right)}{\mathcal{B}\left(K^0_\mathrm{L} \to A'A'\right)} = \left(\frac{\mathcal{R}eV^*_{ts}V_{td}}{\mathcal{I}mV^*_{ts}V_{td}}\right)^2 \times \left(\frac{\tau_{K^0_\mathrm{S}}}{\tau_{K^0_\mathrm{L}}}\right) \simeq 9\times10^{-3}, \tag{5.7}$$

we can also reach

$$\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-) = 2.2\times10^{-12}. \tag{5.8}$$

5.2 Kaons at LHCb

As mentioned in Sect. 3, in addition to being a b factory, the LHCb detector has been found to be optimal for the study of s physics. In particular, it is estimated that on each collision at the VELO, 2 − 3 neutral kaons are produced [182]. They are the third most-common particle species produced directly from pp collisions, only surpassed by charged kaons and pions. The cross section times the multiplicity is approximately 0.6 barn (with K^0_S and K^0_L each taking 0.3 barn). The LHCb detector would be unable to detect most of K^0_L decays, since they fly through the detector leaving no trace. In turn, K^0_S mesons fly a few centimeters after being produced in the PV. 40% of them, according to simulation, decay inside the VELO.

As of the submission of this thesis, the LHCb Collaboration has studied three very rare kaon modes. Searches for $K^0_\mathrm{S} \to \mu^+\mu^-$ have been performed with every new period of data-taking, with no signal found as of yet. The latest one, Ref. [115], pushed the expected limit

$$\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-) < 2.1 \times 10^{-10} \ (90\% \ C.L.\ ) \tag{5.9}$$

rather close to the SM prediction [181],

$$\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-)_{SM} = (5.18 \pm 1.50_{\mathrm{LD}} \pm 0.02_{\mathrm{SD}}) \times 10^{-12}. \tag{5.10}$$

Run 2 data taken during 2016 − 2018 was later used to search for $K^0 \to \mu^+\mu^-\mu^+\mu^-$ decays [116], and $K^0 \to \pi^+\pi^-\mu^+\mu^-$. Details of the study of $K^0 \to \mu^+\mu^-\mu^+\mu^-$ will be presented in this chapter. Neither the K^0_S nor the K^0_L decays have been seen before, with upper limits established for the first time for both decays. Though the study was optimized for the K^0_S mode, knowing the suppression factor of K^0_L at LHCb, the upper limit on the branching fraction of the K^0_L mode comes in almost for free.

5.3 Signal selections

The search for $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ decays (signal) at LHCb was performed relative to $K^0_\mathrm{S} \to \pi^+\pi^-$ processes. The latter will be hereafter referred to as our normalization mode, and has a measured branching fraction of [47]

$$\mathcal{B}(K^0_\mathrm{S} \to \pi^+\pi^-) = (69.20 \pm 0.05)\%. \tag{5.11}$$

$K^0_\mathrm{S} \to \pi^+\pi^-$ candidates are reconstructed using trigger-unbiased data collected during the same period as the triggered signal candidates.

Dedicated stripping lines were written for both decays, attempting to isolate the candidates from random combinations of tracks which may contaminate our signal. In the case of our signal, we are looking for four tracks identified as muons (they are required to satisfy `IsMuon`, a requirement that asks for hits on a subset of the LHCb muon stations, depending on the total momentum carried by the track) forming a vertex with an invariant mass lower than 600 MeV/c^2. They must be well reconstructed, with its χ^2 per degree of freedom below 3.

The particle from which they originate shall be called a K^0_S , and it is required to have a lifetime greater than ∼ 4.6 ps (which is the case for ∼ 90% of the K^0_S that are produced and which can, in turn, eliminate a lot of background contributions), as well as an impact parameter with respect to the PV to be below 1 mm and an IP_{χ^2} below 100, to maximize the chances of it coming from the PV. The flight direction of the K^0_S and the sum of the 3-momenta of the daughters must be aligned, which we ask for by requesting the cosine of the angle between the two vectors, called the direction angle (DIRA), to be greater than 0.9999.

The point where the K^0_S decays is called the secondary vertex, SV, and it is required to happen inside the VELO. The quality of the SV is ensured by a loose cut on its χ^2. The daughter tracks must also have a separation of, at most, 2 mm at the point they are closest to each other (which would be near the reconstructed SV). This is called the distance of closest approach (DOCA).

Meanwhile, candidates on the normalization channel consist of two tracks reconstructed as pions whose invariant mass falls inside the $(400-600)$ MeV/c^2 range, an χ^2_{IP} greater than 100, and a distance of closest approach between them below 0.1 mm.

They form a secondary vertex, a K^0_S , with a positive value on the cosine of its direction angle and an impact parameter below 0.4 mm. The decay time must be $0.06 \times \tau_{K^0_S}$, where $\tau_{K^0_S}$ is the average lifetime of the K^0_S .

There is a third channel used in the analysis as a control mode for trigger studies, $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$. This decay is also reconstructed from trigger-unbiased data, where tracks reconstructed as a pion and a muon are required to form a good secondary vertex, inside the VELO and displaced at least 4 mm from the beam pipe in the transversal plane. The tracks also have strict cuts on its ghost probability, a neural network-based quantity that estimates the chances that the reconstructed tracks are not real tracks, but rather random combinations of hits (ghosts).

Table 5.1 contains a full breakdown of the selections used in all three stripping lines. The least efficient selection on $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ is asking the muon candidates to satisfy `IsMuon`. This takes away roughly 50% of the signal, mostly due to the requirements on the total momentum of the tracks. In the case of $K^0_S \to \pi^+\pi^-$ candidates, the requirement on the DOCA is the least efficient one, at roughly 65%.

Offline selections are implemented on both channels to further suppress combinatorial and peaking backgrounds. In the case of the signal channel, cuts on track quality variables are implemented to reduce the number of ghosts, which tend to have larger impact parameters than real tracks.

As such, we require the χ^2 per degree of freedom of each track to be below 2.5. The ghost probability of all tracks must be below 0.3, and the `ProbNNghost` variable must be under 0.7. The particle identification variable referred to as `PIDmu` (see Sect. 3.8), is required to be positive.

The first three selections are aligned on both the signal and normalization channels. In the latter, two more cuts are implemented to improve the similarities. The transverse flight distance of the K^0_S (defined as the distance that the K^0_S flies away from the beam pipe in the transversal plane) is required to be greater than 5 mm. In addition, we minimize the inclusion of tracks that may originate from material interactions by creating and cutting on a variable that maps these. See more in Sect. 5.6.

Finally, one more selection is required in the normalization channel to remove the only peaking background present in this analysis, $\Lambda \to p\pi^-$. See more in Sect. 5.5.

In the case of $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$, additional selections were also applied to isolate the decay from misidentified $K^0_S \to \pi^+\pi^-$ candidates. The use of this channel, as well as its various selection steps, were optimized in the $K^0_S \to \mu^+\mu^-$ analysis from Ref. [115]. We take $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ candidates where the di-pion mass, $m_{\pi^+\pi^-}$, *i.e.*, the invariant mass of both tracks having a pion mass hypothesis, satisfies $m_{\pi^+\pi^-} < 400$ MeV/c^2 and $m_{\pi^+\pi^-} < (300 + (200 - E_{\text{T}}) \times 5/3)$ MeV/c^2. Here, the E_{T} is computed by assuming the

Table 5.1: Selections used at stripping level for each of the channels used in the analysis. Here, Ghost Prob. stands for the probability that the input tracks are ghosts. nDoF referes to the number of degrees of freedom. θ is the direction angle of the mother, computed as the angle between the estimated flight direction of the mother and the sum of the four-momentum of the tracks. DOCA stands for the distance of closest approach between the daughter tracks. $\mathrm{PV_Z}$ ($\mathrm{SV_Z}$) stands for the z coordinate of the primary (secondary) vertex, while SV_ρ stands for its radial position, $\sqrt{\mathrm{SV_X}^2 + \mathrm{SV_Y}^2}$.

Channel	$K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$	$K^0_\mathrm{S} \to \pi^+\pi^-$	$K^0 \to \pi^+\mu^-\bar{\nu}_\mu$
Input particles	Muons	Pions	Pions and Muons
Daughters IP	—	—	> 0.4 mm
Daughters IP_{χ^2}	—	> 100	—
Daughters Ghost prob.	< 0.4	—	< 0.1
Daughters Track χ^2/nDoF	<3	—	—
Daughters `IsMuon`	True	—	True (only for μ)
Mother mass	< 600 MeV/c^2	$\in [400, 600]$ MeV/c^2	< 800 MeV/c^2
Mother decay time	$> 4.59\,\mathrm{ps}$	$> 0.06 \times \tau_{K^0_\mathrm{S}}$	—
Mother IP	< 1 mm	< 0.4 mm	—
Mother $\cos\theta$	> 0.9999	> 0	> 0
Mother DOCA	< 2 mm	< 0.1 mm	< 0.1 mm
Mother $\mathrm{SV_Z}$	< 650 mm	< 650 mm	< 650 mm
Mother SV_ρ	—	—	> 4 mm
Mother SV_{χ^2}	< 50	—	< 9
Mother IP/($\mathrm{SV_Z}$- $\mathrm{PV_Z}$)	—	—	$< 1/60$
Mother IP_{χ^2}	< 100	—	—

K^0_S is produced at its closest PV and the flight direction is given by the reconstructed SV. Cuts on the VELO material veto, to minimize candidates coming from material interactions (see Sect. 5.6), as well as the $\mathrm{IP}_{\chi^2}(K^0_\mathrm{S}) > 256$, to ensure that the two-track combination alone is not pointing back to the PV, are also applied.

To avoid experimenter's bias when analyzing the collected data, the signal region was blinded during the main part of the analysis, and only restored once everything else was satisfactory. To this end, candidates in the interval $490 - 510$ MeV/c^2 were made unavailable during the bulk of these studies, including all steps of background removal. The remaining events make up the so-called sidebands, with candidates with $m(K^0_\mathrm{S}) < 490$ MeV/c^2 are known as the left sideband, and those with $m(K^0_\mathrm{S}) > 510$ MeV/c^2 are the right sideband.

5.4 Triggering $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ candidates

In the case of strange decays at LHCb during Run 2, it is observed in simulation that roughly 50% of the signal candidates picked up by the hardware trigger step, L0, do not contain tracks responsible for firing the trigger. This motivates the decision to split our dataset into two trigger categories that will be hereafter studied simultaneously.

On the one hand, the TIS sample contains candidates where other tracks in the event, other than the four muons selected by our stripping line, are responsible for firing the trigger. The TIS subsample is therefore made up of $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ candidates where L0 lines targeting muons, di-muon pairs, electrons, photons, and hadrons have been fired elsewhere in the event.

On the other hand, we have an exclusively triggered-on-signal dataset (xTOS), which contains candidates where at least one of our four stripping-selected muon tracks fired the trigger, and no other track in the event did. The vast majority of the candidates in the xTOS category come from muons firing the `L0Muon` line, with a small fraction of di-muon pairs triggering `L0DiMuon`, and a negligible number passing the low-multiplicity counterpart.

This split at the L0 stage, and kept throughout the entire analysis, until the final fit. Both datasets are then required to pass either one or two lines at HLT1. As indicated in Table 5.2 the difference between them lies in a higher total and transverse momentum threshold in the case of `DiMuonLowMass`, and a cut on the IP to ensure a detachment from the PV in the case of `DiMuonNoL0`. Most of the candidates pass `DiMuonNoL0`.

Finally, at HLT2 level, they are required to satisfy a specific line targeting di-muons coming from strange mesons, `Hlt2DiMuonSoft`. Their invariant mass must be below 1 GeV/c^2, they must form a good SV that is sufficiently detached from the beam pipe. Cuts on the probability and PID are also set in place, as is a requirement on the angle between the two tracks to remove clones (groups of hits reconstructed as two tracks but that in fact come from the same traversing particle).

At HLT2 level, there is also a requirement for the K^0_S vertex to be inside the VELO, which forces us to work exclusively with long tracks. This is beneficial to the analysis, since we need very good track reconstruction, but it also means we are losing a large portion of our signal, since more than half of the K^0_S produced at LHCb decay outside of the VELO.

All of the HLT selections are required to be fired by at least one of the di-muon pairs in our signal. The efficiencies of each trigger stage can be found in Table 5.3.

5.5 The Armenteros-Podolanski plane

In the 1950s, it was found that two-particle decays of neutral kaons and Lambdas may be subject to experimental mix-ups. The mis-identification of a proton and a pion may cause the mis-reconstruction of $K^0_\mathrm{S} \to \pi^+\pi^-$ as $\Lambda \to p\pi^-$, and vice versa. After understanding that the energy released in the decay, the Q-value, was not enough to separate the decays, Ref. [184] defined a new set of transformed kinematic variables to isolate the different decay schemes.

In the early days of Particle Physics, in cosmic ray experiments [22], particle decays were studied by looking into the signatures left by the daughter tracks in cloud chambers. Characteristic, V-shaped marks left behind by pairs of tracks led to the naming of V particles as their mother. They have since been classified as kaons, Lambda and Sigma baryons.

Let us consider a two-body decay $V^0 \to h_1^+ h_2^-$. In the center of mass of the V^0,

$$P^\mu_{V^0} = (M, 0) \quad , \quad P^\mu_1 = (E_1, \vec{p}^*) \quad , \quad P^\mu_2 = (E_2, -\vec{p}^*). \tag{5.12}$$

Operating, and using conservation of energy,

$$M = E_1 + E_2 = \sqrt{m_1^2 + \vec{p}^{*,2}} + \sqrt{m_2^2 + \vec{p}^{*,2}}, \tag{5.13}$$

Table 5.2: Cuts applied by the HLT1 and HLT2 lines used in this analysis. GEC means Global Event Cuts, a requirement on some of the parameters of the event, like the number of VELO hits. The Tight state allows up to 3000 VELO hits, while Loose allows up to 6000. θ is the angle between tracks.

	`Hlt1DiMuonNoL0`	`Hlt1DiMuonLowMass`	`Hlt2DiMuonSoft`
Daughters IP	> 0.4 mm	—	> 0.3 mm
Daughters p	> 3 GeV/c	> 6 GeV/c	—
Daughters p_{T}	> 80 MeV/c	> 200 MeV/c	—
Daughters IP_{χ^2}	> 9	> 9	> 9
Daughters Ghost Prob.	—	—	< 0.4
Daughters `ProbNNmu`	—	—	> 0.05
Daughters track χ^2/ nDoF	< 4	< 4	—
$m_{\mu^+\mu^-}$	> 220 MeV/c^2	$[220, 1000]$ MeV/c^2	< 1 GeV/c^2
Mother SV_{χ^2}	< 25	< 25	< 25
Mother SV_{ρ}	—	—	> 3 mm
Mother SV_{Z}	—	—	< 650 mm
DOCA	< 0.2 mm	< 0.2 mm	< 0.3 mm
$\mathrm{SV_Z} - \mathrm{PV_Z}$	—	—	> 0 mm
$\mathrm{IP}/(\mathrm{SV_Z} - \mathrm{PV_Z})$	—	—	$< 1./60$
$\cos\theta$	—	—	< 0.999998
GEC	Tight	Loose	—

Table 5.3: Efficiencies, in %, for each of the trigger stages on both the TIS and xTOS trigger categories.

Trigger Category	**L0**	**HLT1**	**HLT2**	**Total**
TIS	13.00	75.62	92.54	9.10
xTOS	12.93	80.71	93.10	9.72

which leads to

$$p^{*,2} = \frac{M^4 - 2M^2\left(m_1^2 + m_2^2\right) + \left(m_1^2 - m_2^2\right)^2}{4M^2}. \tag{5.14}$$

In the lab frame, due to the conservation of momentum, the p_{T} of h_1 and h_2 is identical, but with opposite signs, $p_{\mathrm{T}}(h_1) = -p_{\mathrm{T}}(h_2)$. This quantity is also invariant under a Lorentz boost to the V^0 center-of-mass frame, $p_{\mathrm{T}} = p_{\mathrm{T}}^*$, since the boost is performed along the V^0 flight direction, the longitudinal axis.

This is not the case, however, for the longitudinal components, which, under a Lorentz boost of strength β from the center of mass to the lab frame, transform as

$$p_{L,i} = \gamma(p_{L,i}^* + \beta E^*). \tag{5.15}$$

Using

$$p_L^* = \pm p^* \cos\theta \quad , \quad p_{\mathrm{T}}^* = p^* \sin\theta, \tag{5.16}$$

Table 5.4: Values that define the ellipses in the α vs. p_{T} plane, also known as the Armenteros-Podolanski plot. These ellipses provide a kinematic way of isolating $K^0_{\mathrm{S}} \to \pi^+\pi^-$ candidates from $\Lambda \to p\pi^-$ or its CP counterpart.

Decay	p^* [GeV/c]	α_0	r_α
$K^0_{\mathrm{S}} \to \pi^+\pi^-$	0.206	0	0.827
$\Lambda \to p\pi^- / \overline{\mathit{\Lambda}} \to \overline{p}\pi^+$	0.101	± 0.691	0.181

we can compute the asymmetry in the longitudinal components as

$$\alpha = \frac{p_{L,1} - p_{L,2}}{p_{L,1} + p_{L,2}} = \frac{2p}{\beta M} + \frac{m_1^2 - m_2^2}{M^2}, \tag{5.17}$$

using that $E_1^* + E_2^* = M$, and $\frac{E_1^* - E_2^*}{M} = \frac{m_1^2 - m_2^2}{M^2}$. In the ultra-relativistic case, $\beta \to 1$, notice how

$$\cos\theta = \frac{\alpha - \alpha_0}{r_\alpha} \quad ; \quad \alpha_0 = \frac{m_1^2 - m_2^2}{M^2} \quad , \quad r_\alpha = \frac{2p^*}{M}. \tag{5.18}$$

Recalling Eq. (5.16), we find that the identity $\cos^2\theta + \sin^2\theta = 1$ allows us to define an ellipsis that is unique to each V^0 decay:

$$\left(\frac{\alpha - \alpha_0}{r_\alpha}\right)^2 + \left(\frac{p_{\mathrm{T}}}{p^*}\right)^2 = 1. \tag{5.19}$$

In the specific case of this analysis, this provides a clear-cut way of isolating $K^0_{\mathrm{S}} \to \pi^+\pi^-$ candidates from $\Lambda \to p\pi^-$ misreconstructed decays. Table 5.4 provides values for each of the parameters on the decays of interest. As seen in Fig. 5.2, the following selection, fine-tuned via a $S/\sqrt{S+B}$ figure of merit in Ref. [115], removes the $\Lambda \to p\pi^-$ and $\overline{\mathit{\Lambda}} \to \overline{p}\pi^+$ contributions:

$$\left|\left[\left((\alpha \pm \alpha_0)\frac{M p_{K^0_{\mathrm{S}}}}{2p^*\sqrt{p^2_{K^0_{\mathrm{S}}} + M^2}}\right) + \frac{p_{\mathrm{T}}^2}{(p^*)^2}\right] - 1\right| > 0.3, \tag{5.20}$$

where now α_0 is given by the value from Table 5.4 for $\Lambda \to p\pi^-$ (the ± sign already covers both the $\mathit{\Lambda}$ or $\overline{\mathit{\Lambda}}$ cases), and M stands for the $\mathit{\Lambda}^0_b$ mass.

5.6 Material interactions

Due to the average lifetime of the K^0_{S} , which allows it to fly several centimeters away from the PV, and often not strictly along the z-axis, effects due to inellastic collisions with the VELO material must be considered. In the case of $K^0_{\mathrm{S}} \to \mu^+\mu^-$ this proved to be a highly discriminatory variable against background contributions [115], as these undesired interactions may result in two tracks that may be disguised as muons but which are not, in fact, coming from an actual K^0_{S} decay vertex.

In the case of $K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-$, where four tracks must form a good enough SV, it is also an important feature to study. To do so, we require an excellent parametrization of

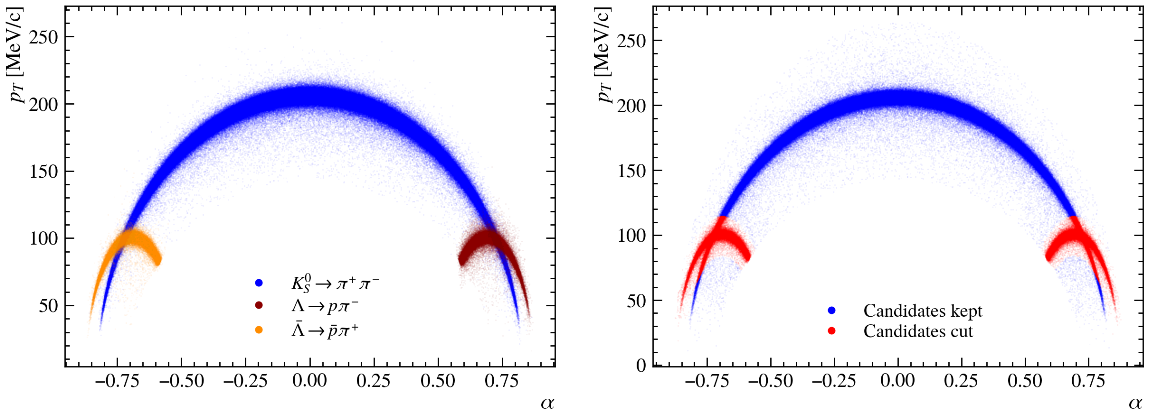


Figure 5.2: Left: Armenteros-Podolanski plot on truth-matched $K^0_S \to \pi^+\pi^-$ (blue), $\Lambda \to p\pi^-$ (green), and $\bar{\mathit{\Lambda}} \to \overline{p}\pi^+$ (orange) candidates. Right: Armenteros-Podolanski plot on minimum bias $K^0_S \to \pi^+\pi^-$ candidates. The cut to remove the $\mathit{\Lambda}$ ($\bar{\mathit{\Lambda}}$) decays is illustrated through the red marks. It has been optimized using a figure of merit on simulation.

the subdetector, which may allow us, combined with the decay vertex's precise position, to calculate the distance from the SV to the material. The parametrization of the VELO is documented in Ref. [185], and was performed using proton-gas events collected by LHCb. As seen in Fig. 5.3, the positions of the SVs perfectly reproduce the shape of the VELO material. They found that the best metric to test the hypothesis that the true SV location is compatible with a point in space occupied by the material was given by the harmonic mean of the uncertainty-weighted material distances,

$$d = 6\left(\sum_{i=1}^{6} d_i^{-1}\right),$$

where

$$d_i = min\left[\sqrt{\frac{(SV_x - x)^2}{\sigma_x^2} + \frac{(SV_y - y)^2}{\sigma_y^2} + \frac{(SV_z - z)^2}{\sigma_z^2}}\right].$$

The variable that results from performing these computations on our data samples can be seen in Fig. 5.4.

Two options arise here. On the one hand, we may select candidates with $d > d_0$, where d_0 would need to be optimized, to veto tracks originated close enough to the subdetector material. The better option, however, turns out to be including this variable as an input feature of a multivariate classification algorithm that will also reduce most of the combinatorial background (see Sect. 5.8).

5.7 MC corrections

5.7.1 Kaon physics corrections to simulation

Kaon physics are not accurately described by the MC tools used in LHCb. In particular, the momentum spectrum requires correcting, which may trickle down to an underestimation

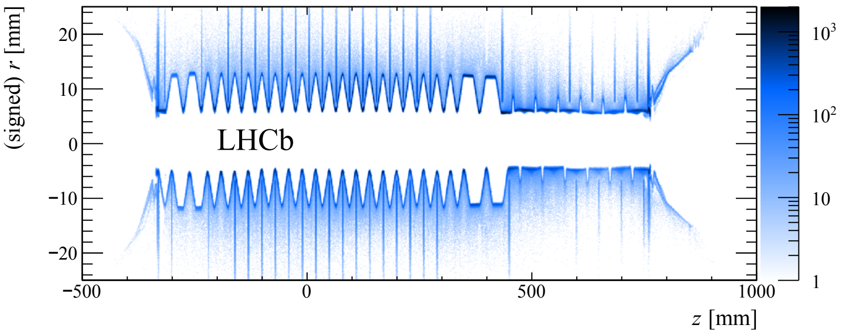


Figure 5.3: Parametrization of the VELO and RF-foil using secondary vertices from Run 1 data. This is then used to calculate the distance to the material from displaced decays, like those of K^0_S . Figure from Ref. [185] under Creative Commons Attribution 3.0 license.

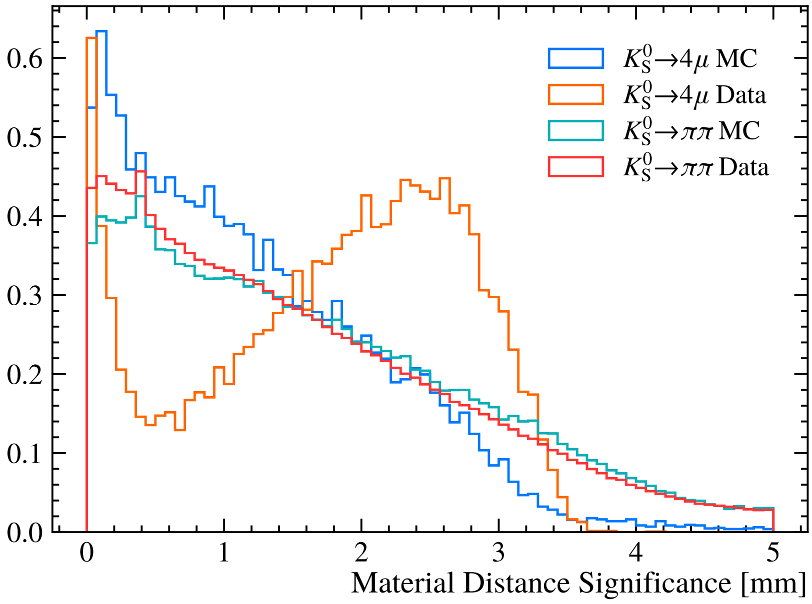


Figure 5.4: Distribution of the minimum uncertainty-weighted distance obtained with the VeloMaterial tool for the different data samples used in the analysis. Given the lower opening angle, the vertex resolution of $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ decays is worse than in $K^0_\mathrm{S} \to \pi^+\pi^-$, which causes the $K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-$ signal to drop much faster. The $K^0_\mathrm{S} \to \pi^+\pi^-$ candidates also have a cut on the distance of closest approach.

of the efficiencies when using the numbers from plain simulation. It is a problem born at the generation level, that is then carried over to the stripped datasets. Two methods are used in this analysis to account for this.

First, we use the accept-reject algorithm, first introduced by John von Neumann in a 1951 lecture [186, 187], to filter out simulated events for which the transverse momentum of the K^0_S is not well reproduced.

In this method, we have a target probability distribution, $f(x)$, and a proposal distribution, $g(x)$, such that $f(x) \leq Mg(x)$ for some constant M and for all x. We now

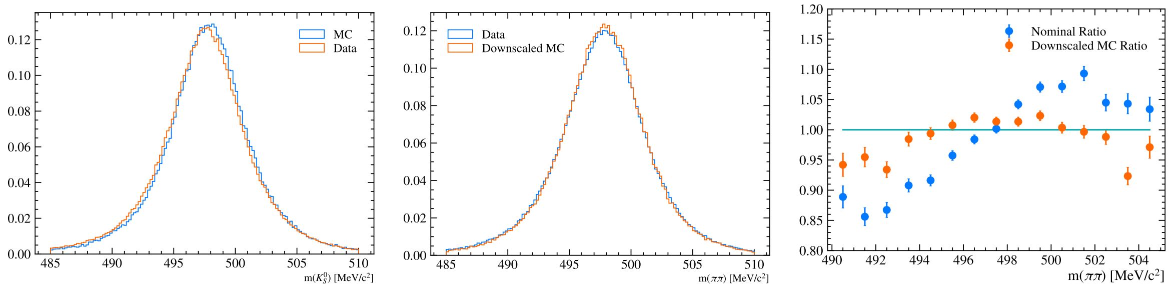


Figure 5.5: Comparison of the $K^0_S \to \pi^+\pi^-$ peaks as seen in MC and data, before (left) and after (center) rescaling the daughters' momenta. The per-bin ratio between both histograms can be seen on the right, where we notice that the agreement is better after the correction.

draw a candidate sample x_0, compute the acceptance probability,

$$r = \frac{f(x_0)}{Mg(x_0)},$$

and flip a coin with success probability r. If head turns up, we accept x_0.

In our case, our distributions are the K^0_S p_T from data (target) and simulation (proposal) from the $K^0_S \to \pi^+\pi^-$ control mode. We correct our simulation on each channel by drawing candidate samples from both $K^0_S \to \pi^+\pi^-$ and $K^0_S \to \mu^+\mu^-\mu^+\mu^-$. See the corrections on the p_T of the K^0_S in Fig. 5.6, top left.

Second, since the invariant mass of the K^0_S is also not well defined in simulation, we find, looking at $K^0_S \to \pi^+\pi^-$ data-MC comparisons, that the peak is slightly displaced towards higher masses. This can be corrected by recomputing the invariant mass after applying a small downscale in the daughters' momenta in simulation. In particular, we find that only a factor of $s \simeq 0.99942^{+0.00006}_{-0.00004}$ need be applied. See Fig. 5.5.

5.7.2 Tracking corrections

On top of all the above, some of the tracking variables in simulation tend not to be well defined. To correct this, we use $K^0_S \to \pi^+\pi^-$ once again as our control channel, and match the cumulative distribution functions of the desired variables between data and MC. These variables are the track χ^2 per degree of freedom, the track's ghost probability, as well as its `ProbNNghost`. The functions derived from matching the cumulative distribution functions in the control mode are then used to correct the variables on both the signal and normalization channels.

Tracking efficiencies are also not accurately described in MC. `TrackCalib2` is a tool developed by LHCb to help with this [188]. We consider, independently, all three years and both channels. One single correction factor is considered for TIS and xTOS. 20% of muon tracks, and 51% of $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ candidates fall outside the calibration spectrum (their momentum is below 5 GeV/c). To these, a correction factor of 1 ± 0.05 is applied. On top of that, each track is assigned a 0.8% systematic uncertainty. Since muons were used to make the calibration efficiency tables, pion tracks are assigned a 2% systematic.

The final correction factor to the ratio of efficiencies is

$$R = 1.002 \pm 0.012 \tag{5.21}$$

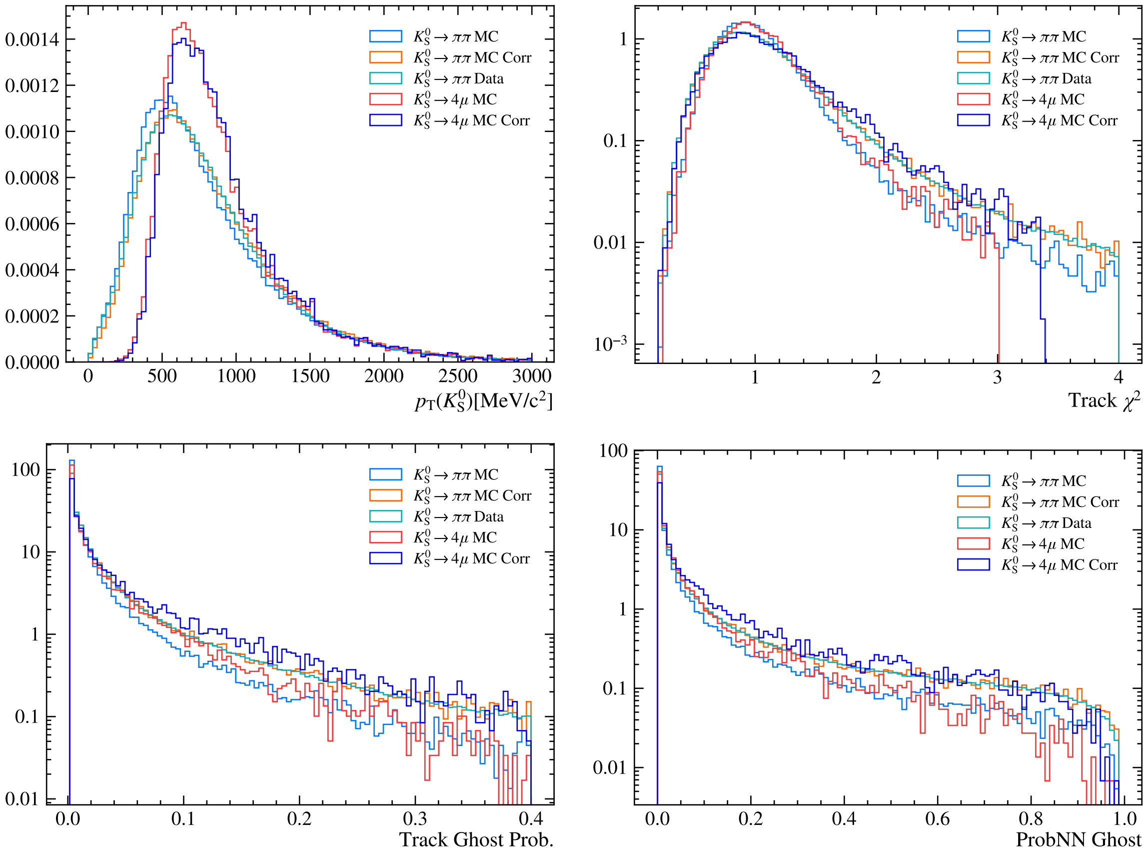


Figure 5.6: Corrections to the simulation using $K_S^0 \to \pi^+\pi^-$ as a control sample. Top left: p_T of the K_S^0 , which is corrected through the accept-reject method. Top right: Track χ^2 of the daughters, corrected by matching the CDFs of the simulation control sample to that of the data. Bottom: Ghost probability of the track (left) and `ProbNNghost` of the track (right), corrected the same way as the track χ^2.

To account for hypothetical acceptance drops in low opening angles that may appear in data, but not necessarily in MC, we compare the opening angle of the di-Kaon pair from $B_s^0 \to J/\psi(\mu^+\mu^-)\phi(KK)$ between simulation and sWeighted data, which is of the same order of magnitude as the $K_S^0 \to \mu^+\mu^-\mu^+\mu^-$ opening angle (see Fig. 5.7). When comparing these distributions, only tiny differences are observed, which may be attributed to lack of $f^0 \to KK$ decays in simulation, sWeights factorization, or vertex resolution.

5.7.3 PID corrections

The efficiencies computed in simulation corresponding to the different selections on particle identification need to be corrected, following the arguments presented in Sect. 3.8.1. The calibration was done using low-p_T muons from $B^+ \to J/\psi(\mu^+\mu^-)K^+$ data (see Fig. 5.8 for a comparison between the momentum distribution of the signal and control samples.)

Tracking and kinematic selections are aligned between the calibration samples and our nominal MC, to ensure that they look as closely as possible. To reinforce this, we also require our calibration muons to have a $p_T < 1\,\mathrm{GeV}/c$, which describes most of our signal and excludes higher-momentum tracks that may not apply to our kinematic spectrum.

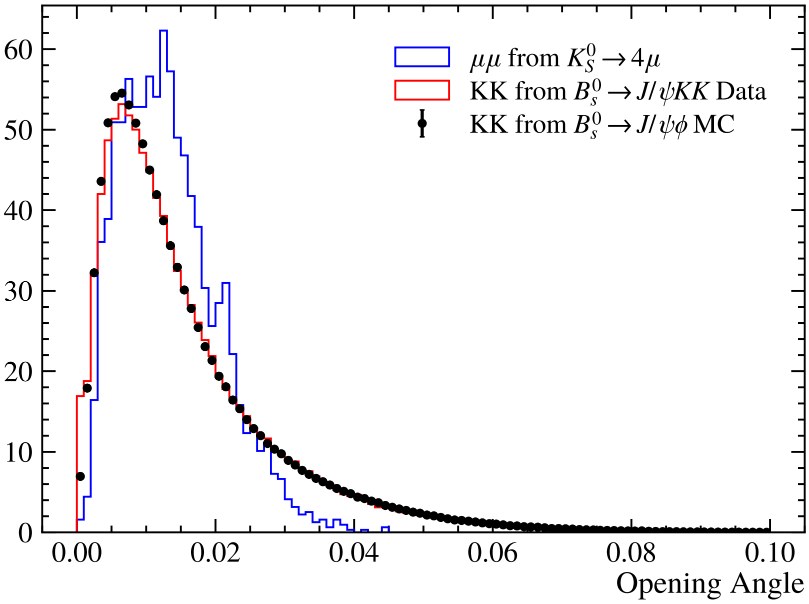


Figure 5.7: Opening angle distribution from two-track vertices. We compare the KK opening angle between simulation and sWeighted $B^0_s \to J/\psi(\mu^+\mu^-)\phi(KK)$ data, and only observe tiny differences that may be attributed for some unaccounted effects in simulation, like the lack of $f^0 \to KK$ decays, or vertex resolution. The di-muon opening angle from $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ is also plotted, to illustrate that it is in the same order of magnitude

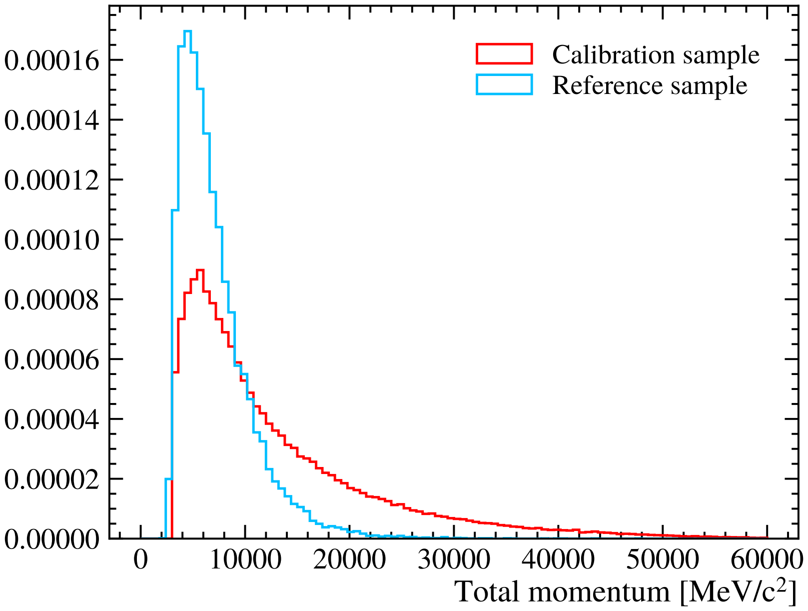


Figure 5.8: Comparison between the momentum of the final-state muons in the reference ($K^0_S \to \mu^+\mu^-\mu^+\mu^-$) and calibration ($J/\psi \to \mu^+\mu^-$) samples.

The calibration is done in bins of total momentum, the pseudorapidity, η, and the number of hits in the SPD subdetector. The sWeighted efficiency of the corresponding cuts (in our case, `IsMuon & PIDmu` > 0), is computed for the calibration data in each bin. Then, this per-bin efficiency is assigned to each event in the MC. The final, corrected efficiency, is the average of all the corrected efficiencies of the events.

Also to be considered is the fact that the calibration is done with a two-track channel, while our signal channel contains four tracks. Efficiencies were computed factorizing two muons, then three, and finally considering all four, and found no significant differences.

The final correction factor to the nominal MC efficiencies is found to be

$$R_{PID} = 1.23 \pm 0.04. \tag{5.22}$$

5.8 Multivariate Classifier

A Boosted Decision Tree (BDT) was trained in this analysis to filter out combinatorial background contributions. $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ simulated events were taken as a signal proxy, while data from the sidebands was taken as a background proxy. The following, discriminatory variables were taken as the input features on the algorithm,

- The distance of closest approach between all four tracks,
- The impact parameter of the K^0_S ,
- The minimum impact parameter of the four muons with respect to the PV,
- The transverse flight distance of the K^0_S ,
- The distance to the VELO material,
- The minimum angle between two muons.

Fig. 5.9 and Fig. 5.10 include the comparison between the signal and proxy backgrounds of all six variables that enter the BDT, from the TIS and xTOS samples, respectively.

The dataset is divided in three subsamples. One is used to train the model (training set). Another is used to examine the properties of the train model (test dataset). A final subsample is held out of the process until the algorithm is fully trained and characterized. At that point, it is applied on this third subsample, such that it is free of biases that may have arisen before.

Half the dataset is used for training and testing, and half is held aside. The split is done according to the parity of the `eventNumber`. The first half is, once again, split in two according to this variable. A `PIDmu`> 0 requirement is set on the signal proxy, while the training sample of the background proxy is asked to satisfy `PIDmu`< 0. The testing subsample is not required to satisfy this.

To avoid overtraining concerns, and to maximize the use of our full dataset, we do the training twice (two folds). The first time, we use the candidates with an even `eventNumber` for the holdout sample, and the second time, we use the odd subsample.

Various MVA methods have been considered:

- Boosted Decision Trees, with four different settings from TMVA [189]

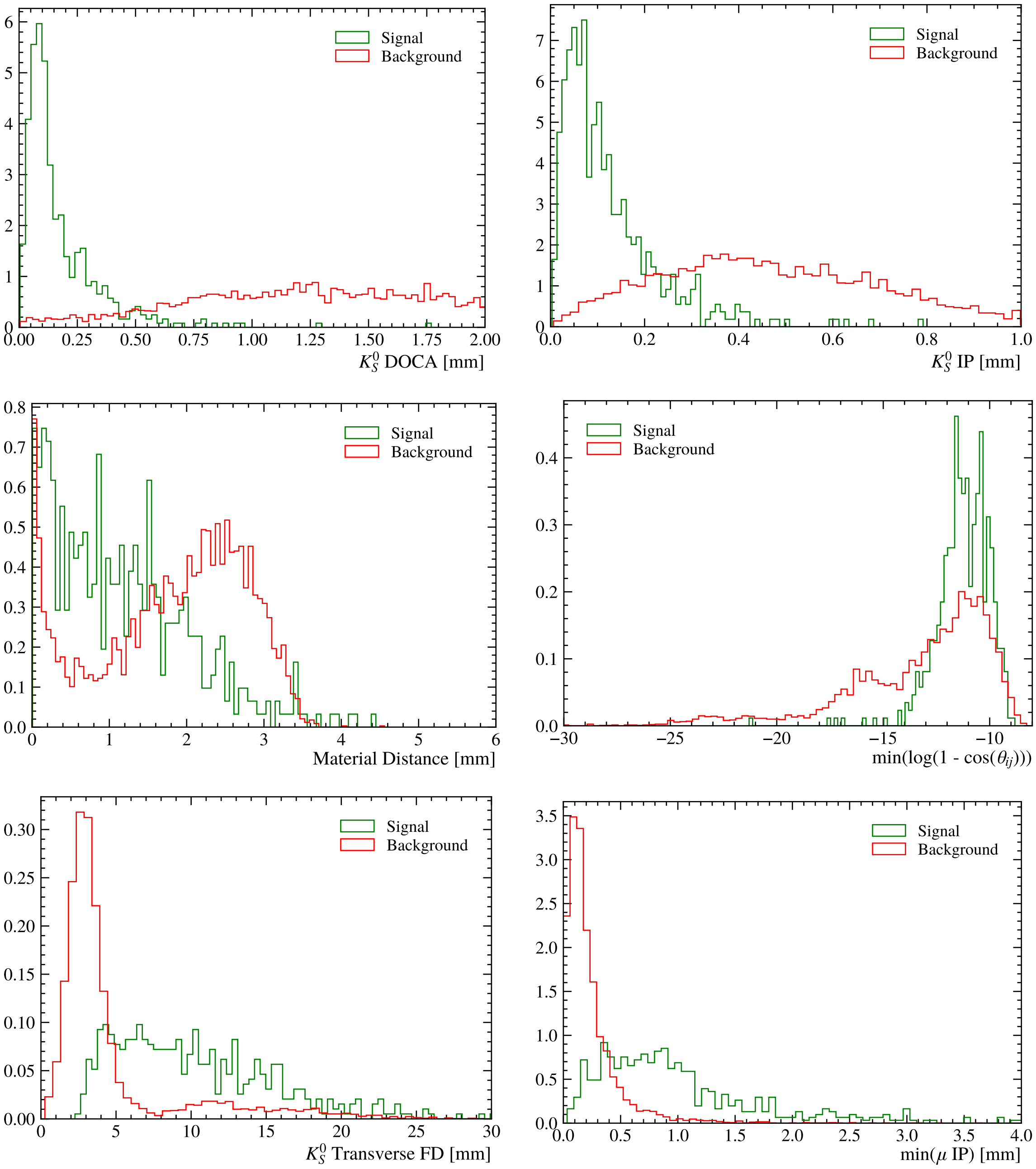


Figure 5.9: Representation of the variables in the TIS sample used as an input of the BDT. Top: Distance of closest approach between the four tracks (left) and impact parameter of the K_S^0 (right). Middle: Distance to the material, as provided by the tool described in Sect. 5.6 (left) and the minimum angle between pairs of muons, transformed as $\log(1 - \cos\theta_{ij})$ (right). Bottom: Transversal flight distance of the K_S^0 (left) and minimum impact parameter of the four muons with respect to the PV (right).

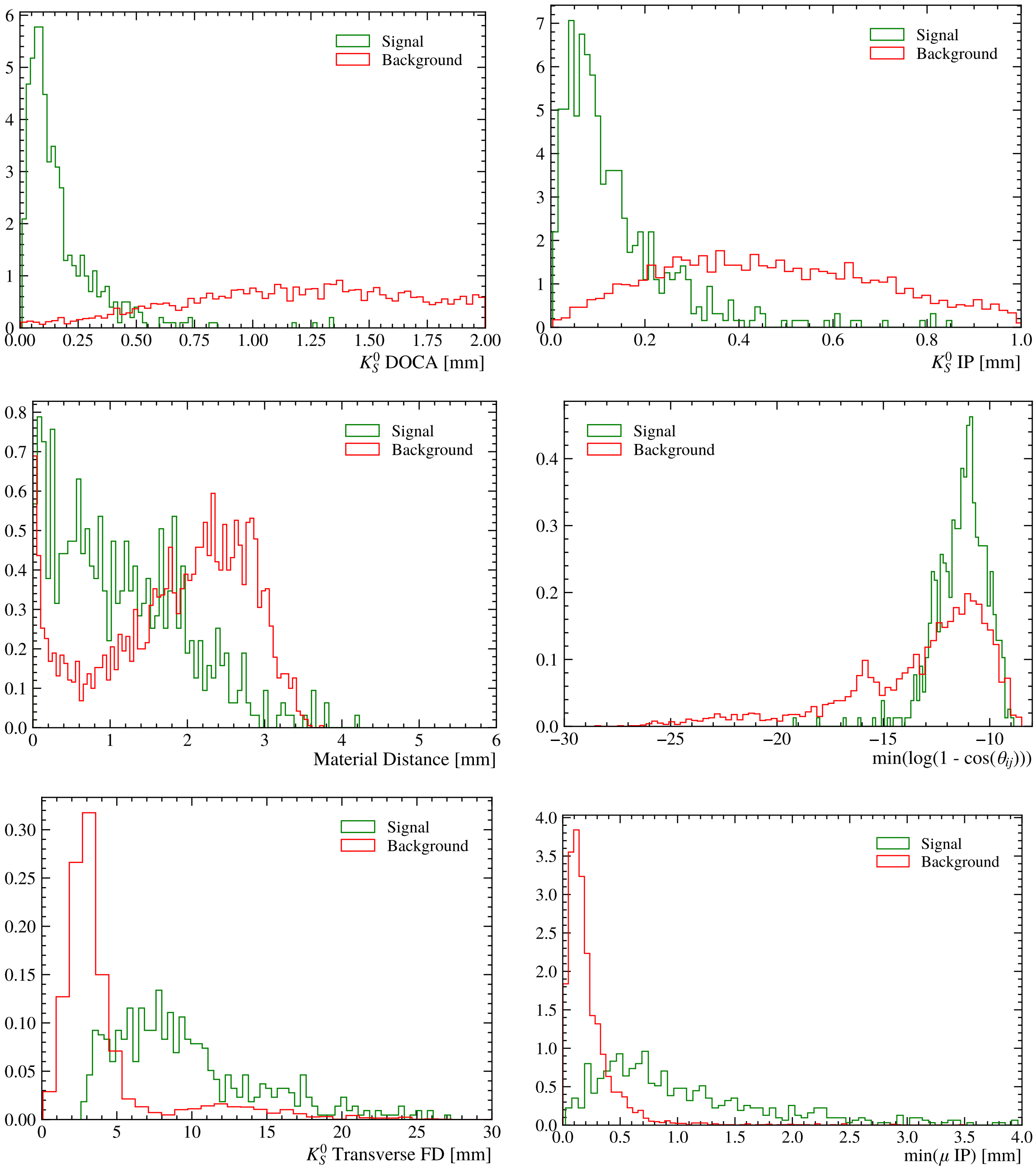


Figure 5.10: Representation of the variables in the xTOS sample used as an input of the BDT. The ordering is the same as that in Fig. 5.9.

- Fisher discriminant
- k-Nearest Neighbor
- TMVA Likelihood
- Geometrical Likelihood from the Urania package [190]
- hMatrix

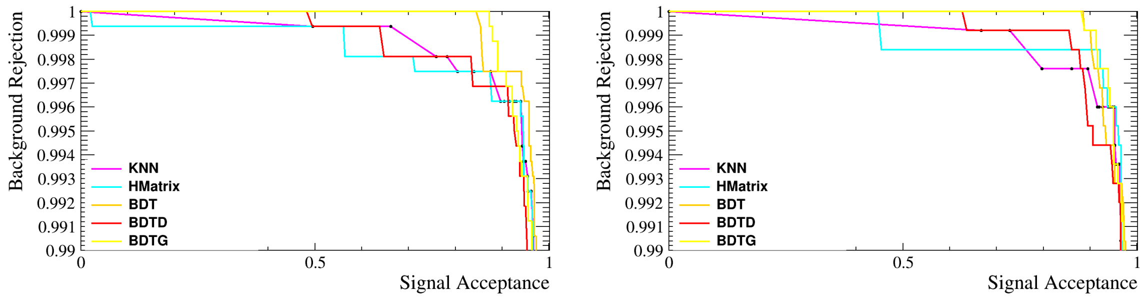
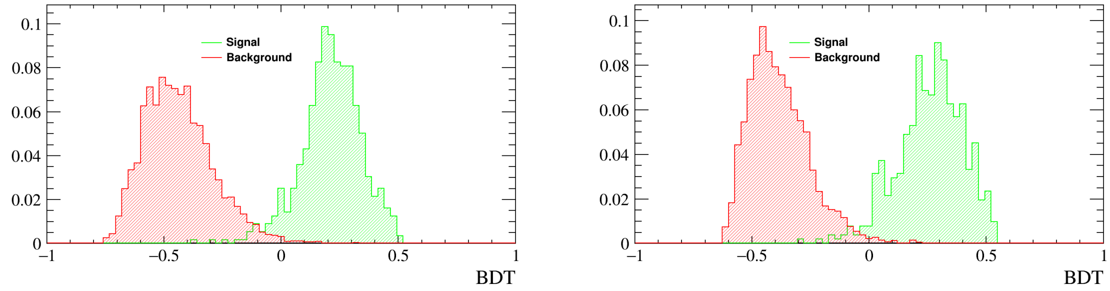
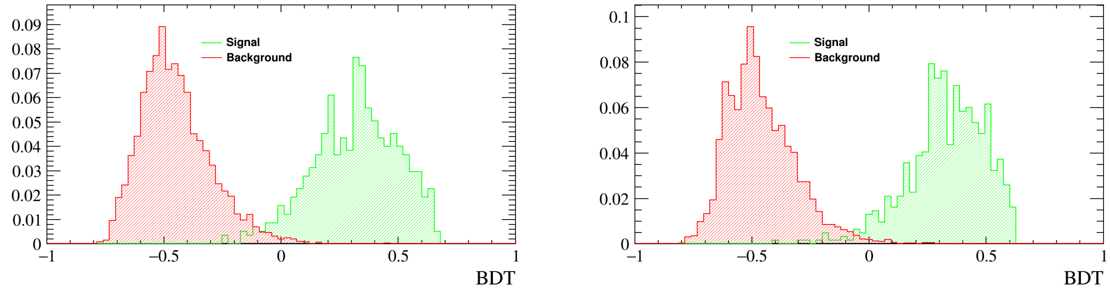


Figure 5.11: ROC curves for the two trigger categories, comparing the response from the different classifying algorithms tested.



(a) TIS, `eventNumber` odd (b) TIS, `eventNumber` even



(c) xTOS, `eventNumber` odd (d) xTOS, `eventNumber` even

Figure 5.12: Distributions of the output BDT variable for each of the two folds and trigger categories.

The most discriminating variables are found to be the DOCA, the K_S^0 impact parameter, and the minimum IP of the tracks. The response of the different classifiers tested can be seen in the ROC curves, in Fig. 5.11. The distributions of the output BDT variables for each of the folds are included in Fig. 5.12.

A cut on the output BDT variable is chosen using a figure of merit such that 80% of the signal survives. With it, only three background candidates on the TIS subsample remain. All background candidates on the xTOS subsample are removed.

5.9 Background studies

All background contributions in this analysis are caused by random combination of tracks around the K^0_S mass window. Very few kaon decays could actually threaten this statement, and no other SM process would. The η, the only known particle in the mass window studied in the analysis, does not satisfy the lifetime requirements if taken as a prompt decay, and does not satisfy the impact parameter requirements if coming from a heavier meson.

There are only a handful of possibilities for a K^0 to decay into four particles. The most likely one is $K^0_S \to \pi^+\pi^-e^+e^-$, which has a measured branching fraction in the order of 10^{-5}. The difficulty of reconstructing the soft di-electron pair, combined with a quadruple misidentification, which also involves a displacement of the mass peak several sigmas away from our signal region, makes it a negligible contribution.

A similar explanation can be used to rule out $K^0 \to \mu^+\mu^-e^+e^-$. The very rare decay $K^0_S \to \pi^+\pi^-\mu^+\mu^-$ is even less likely to occur, according to SM predictions, than $K^0_S \to \mu^+\mu^-\mu^+\mu^-$.

5.9.1 The case of $K^0_L \to \mu^+\mu^-\mu^+\mu^-$

All of the above processes, due to double misIDs, would fall inside the sideband regions of the analysis. Theoretically, $K^0_L \to \mu^+\mu^-\mu^+\mu^-$ is more likely to occur than $K^0_S \to \mu^+\mu^-\mu^+\mu^-$, and it would be confused with signal. Should a signal contribution had been found, it would have been catalogued as $K^0 \to \mu^+\mu^-\mu^+\mu^-$, as we would not have been able to differentiate the specific nature of the mother. However, when studying $K^0_S \to \mu^+\mu^-\mu^+\mu^-$, the K^0_L decay contributions must be accounted for.

According to the current LHCb acceptance, though, K^0_L mesons are suppressed by a factor $\sim 10^{-3}$ with respect to K^0_S, so the effective branching fraction of $K^0_L \to \mu^+\mu^-\mu^+\mu^-$ as seen by the detector would be

$$\begin{aligned}\mathcal{B}(K^0_L \to \mu^+\mu^-\mu^+\mu^-)_{eff} &= \mathcal{B}(K^0_S \to \mu^+\mu^-\mu^+\mu^-) \times \frac{\varepsilon_L}{\varepsilon_S} \\ &\simeq (8.78 \times 10^{-13}) \times (2.3 \times 10^{-3}) \simeq 2 \times 10^{-15},\end{aligned} \tag{5.23}$$

where $\varepsilon_{L(S)}$ is the efficiency of the K^0_L (K^0_S) decay, respectively. With this in mind, we will interpret our final limit in terms of both $\mathcal{B}(K^0_S \to \mu^+\mu^-\mu^+\mu^-)$ and $\mathcal{B}(K^0_L \to \mu^+\mu^-\mu^+\mu^-)$.

Finally, there is the interference between both modes to account for. Theoretically, the amplitude of the neutral-kaon decay is given by

$$\mathcal{A}(K^0 \to \mu^+\mu^-\mu^+\mu^-) = \mathcal{A}(K^0_S \to \mu^+\mu^-\mu^+\mu^-) + \mathcal{A}(K^0_L \to \mu^+\mu^-\mu^+\mu^-), \tag{5.24}$$

so the branching fraction, taken as $\mathcal{B}(K^0_S \to \mu^+\mu^-\mu^+\mu^-) = |\mathcal{A}(K^0_S \to \mu^+\mu^-\mu^+\mu^-)|^2$, includes a term known as interference, the product of both amplitudes. This term has a lifetime much shorter than the K^0_L, being only twice that of the K^0_S [191], and its contribution is only nonzero whenever there is a large $K^0 - \overline{K}^0$ production asymmetry, which was not the case for the luminosity collected during Run 2 [115, 181]. For this reason, it is neglected.

Table 5.5: Resulting values for the minimum bias scale factor for 2016, 2017 and 2018.

	2016	**2017**	**2018**
$\mathbf{s}_{\mathrm{MB}}$	5.66509×10^{-7}	5.30689×10^{-7}	4.43192×10^{-7}

5.10 Normalization

The number of signal decays is given by

$$N(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-) = \mathcal{L} \times \sigma_{K^0_{\mathrm{S}}} \times \mathcal{B}(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-)\varepsilon_{sig}. \tag{5.25}$$

Since we do not know very precisely the production cross section of the K^0_{S} , we are better off comparing this number to the number of decays produced via a different channel, which we are calling the normalization mode. $K^0_{\mathrm{S}} \to \pi^+\pi^-$ is very well suited for this, as its abundance at LHCb allows us to probe very low sensitivities. The difference in the number of tracks between both channels is accounted for by the use of `TrackCalib`.

The $K^0_{\mathrm{S}} \to \pi^+\pi^-$ candidates are taken from Minimum Bias data, a different stream of data than our signal which is heavily prescaled during data-taking. This prescaling factor, s_{MB}, must be factored into the equation, to get

$$\mathcal{B}(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-) = \alpha N(K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-), \tag{5.26}$$

where α, defined as the single event sensitivity (SES), or normalization factor, is given by

$$\alpha = \mathcal{B}(K^0_{\mathrm{S}} \to \pi^+\pi^-) \times \frac{s_{\mathrm{MB}}}{N^{\mathrm{MB}}_{K^0_{\mathrm{S}} \to \pi^+\pi^-}} \times \frac{\varepsilon_{K^0_{\mathrm{S}} \to \pi^+\pi^-}}{\varepsilon_{K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-}}, \tag{5.27}$$

where $N^{\mathrm{MB}}_{K^0_{\mathrm{S}} \to \pi^+\pi^-}$ represents the number of observed $K^0_{\mathrm{S}} \to \pi^+\pi^-$ decays in MB data.

The prescaling factor of the Minimum Bias stream s_{MB} varies from year to year, as seen in Table 5.5. This is defined as the ratio of the integrated luminosity observed by the minimum bias stream over the total luminosity,

$$s_{\mathrm{MB}} = \frac{\mathcal{L}^{\mathrm{int}}_{\mathrm{MB}}}{\mathcal{L}^{\mathrm{int}}}.$$

A fit to the $\pi^+\pi^-$ invariant mass using a Hypatia *p.d.f.* shows that the background contributions (accounted for via an exponential) are below 1%. Therefore, we can simply count the total number of entries in the dataset and add a 1% systematic to the final number. The efficiencies are split into subsequent steps, following the guidelines indicated above:

$$\varepsilon_{K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-} = \varepsilon_{gen} \times \varepsilon_{reco} \times \varepsilon_{strip} \times \varepsilon_{trig} \times \varepsilon_{sel} \times \varepsilon_{BDT} \times R_{PID}, \tag{5.28}$$

where the correction factor to the PID efficiencies, R_{PID}, as given by Eq. (5.22), has been included. Similarly,

$$\varepsilon_{K^0_{\mathrm{S}} \to \pi^+\pi^-} = \varepsilon_{reco} \times \varepsilon_{strip} \times \varepsilon_{sel}. \tag{5.29}$$

One normalization factor α is computed per trigger category. The breakdown of the efficiencies can be found in Table 5.6.

Table 5.6: Efficiencies for $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ and $K^0_S \to \pi^+\pi^-$ decays from corrected MC. Minimum bias MC is used for $K^0_S \to \pi^+\pi^-$. The L0 TIS efficiency is corrected according to $K^0_S \to \pi^+\pi^-$ data. The efficiencies correspond to a weighted average between 2016, 2017 and 2018 MC, taking into account the luminosity of each year. The uncertainties are statistical only.

	$K^0_S \to \mu^+\mu^-\mu^+\mu^-$	$K^0_S \to \pi^+\pi^-$
Gen. + reco. + sel.	$2.5138(23) \times 10^{-3}$	$4.585(53) \times 10^{-3}$
TIS| SEL	0.0746(26)	—
xTOS| SEL	0.0965(27)	—
MVA	0.8	–

Table 5.7: L0 TIS efficiencies computed via $K^0_S \to \pi^+\pi^-$ MB data. They are in good agreement with the same efficiencies computed for the signal channel. A correction factor, accounting for the discrepancies in the presacaling factor of the MB stream, is also included. Since the L0 thresholds did not change throughout 2018, no correction is needed.

	2016	2017	2018
$\varepsilon^{Data}_{L0TIS}(\%)$	11.57 ± 0.04	11.36 ± 0.04	9.29 ± 0.04
r	0.9924 ± 0.0038	0.9819 ± 0.0040	1

5.11 Systematic uncertainties

The largest source from systematic uncertainties comes from the computation of the trigger efficiencies. The L0 TIS efficiency depends, for the most part, on the underlying event, and is not accurately described in simulation. Assuming that the L0 TIS efficiency is the same for signal than for $K^0_S \to \pi^+\pi^-$, we can compute directly on data using the MB stream, as

$$\varepsilon_{L0TIS} = \frac{N^{K^0_S \to \pi^+\pi^-}_{L0TIS}}{N^{K^0_S \to \pi^+\pi^-}}. \tag{5.30}$$

A correction factor, r, is also applied to account for the fact that the minimum bias scale factor is not consistent throughout the same year:

$$r = \frac{\varepsilon^{L0TIS}_{w}}{\varepsilon^{L0TIS}} \quad , \quad \varepsilon^{L0TIS}_{w} = \sum_{TCK} \frac{d\mathcal{L}_{TCK}}{\mathcal{L}} \frac{N^{L0TIS}_{TCK}}{N_{TCK}} \quad , \quad \varepsilon^{L0TIS} = \frac{N^{L0TIS}}{N}. \tag{5.31}$$

Here, $d\mathcal{L}_{TCK}$ represents the luminosity of each trigger configuration key (TCK) and $\mathcal{L}$, the total luminosity. Since the L0 thresholds did not change throughout 2018, no correction is needed. The results are shows in Table 5.7

The aforementioned assumption of L0 TIS efficiencies being similar for the signal and control channels, however, does not come freely. Forced decays in MC have different multiplicities than if generated via MB*, which affects the computation of the efficiency. L0 TIS efficiencies computed via MB agree well between $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ and $K^0_S \to \pi^+\pi^-$. However, the forced-decay efficiency of $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ disagrees with the MB value. Since $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ is always forced, we take a 10% systematic to account for these differences.

Meanwhile, the L0 xTOS efficiency is validated using the TISTOS method (see Sect. 3.5.1), with $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ as a control mode, where the pion has been misidentified. We take

$$\varepsilon^{xTOS} = \varepsilon^{TOS} \times \frac{N^{xTOS}}{N^{TOS}} \simeq \frac{N^{TIS\&TOS}}{N^{TIS}} \times \frac{N^{xTOS}}{N^{TOS}}. \tag{5.32}$$

The systematic is computed by comparing this quantity in both data (asking for both tracks to be good muons) and $K^0_\mathrm{S} \to \mu^+\mu^-$ MC. An 11% difference appears. However, this is a two-track to two-track comparison, and our signal mode has four muon tracks. At leading order, since most of the events fire `L0Muon`,

$$\varepsilon_{4\mu} = \varepsilon_{2\mu} + (1 - \varepsilon_{2\mu}) \times \varepsilon_{2\mu}, \tag{5.33}$$

and therefore,

$$s(\varepsilon_{4\mu}) = (2 - 2\varepsilon_{2\mu}) \times s(\varepsilon_{2\mu}) = 21\% \tag{5.34}$$

is our L0 xTOS systematic uncertainty.

To validate the HLT efficiencies, we also use the TISTOS method with $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ as our control mode. Though we see a good agreement between data and MC, we take the 11% statistical uncertainty from our MC sample as our systematic. To extend this into the four-track case, we need to take into account that there are multiple dimuon combinations that can trigger the HLT dimuon lines, so in this case,

$$\varepsilon_{4\mu} = \varepsilon_{2\mu} + (1 - \varepsilon_{2\mu}) \times \varepsilon_{2\mu} + (1 - \varepsilon_{2\mu})^2 \times \varepsilon_{2\mu} + (1 - \varepsilon_{2\mu})^3 \times \varepsilon_{2\mu} = 3\%. \tag{5.35}$$

We conservatively keep the 11%.

The difference between the corrected and nominal MC efficiencies is also taken into account for the systematics. There is an 8.3% variation before the BDT, and an additional 3% at the MVA level. We take half the correction size as an uncertainty, 4.4%.

Finally, possible K^0_L contributions are also considered. The efficiency ratio between K^0_L and K^0_S is computed by selecting simulated events and weighting them to have the K^0_L lifetime instead. The ratios are shown in Table 5.8. A systematic covering the full range is assigned, finding

$$\frac{\varepsilon_{K^0_\mathrm{L}}}{\varepsilon_{K^0_\mathrm{S}}} = (2.14 \pm 0.03) \times 10^{-3} \tag{5.36}$$

Table 5.9 contains a breakdown of all systematics going into the computation of the single event sensitivity.

5.12 Results

With all of the above, we can compute

$$\alpha_{TIS} = (4.72 \pm 0.75) \times 10^{-12} \quad , \qquad \alpha_{xTOS} = (3.77 \pm 0.92) \times 10^{-12}. \tag{5.37}$$

---

[*] This is because Kaon decays in Pythia are also generated via minimum bias simulation. The difference is that those simulations we refer to as forced decays come from samples where Pythia forces the decay of one of the K^0_S produced into the requested final state. However, to reach $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$ candidates from straight MB simulation, a large number of K^0_S is required, given its $\mathcal{O}(10^{-5})$ branching fraction. This leads to much higher multiplicities.

Table 5.8: Efficiency ratios between $K^0_{\rm L} \to \mu^+\mu^-\mu^+\mu^-$ and $K^0_{\rm S} \to \mu^+\mu^-\mu^+\mu^-$ according to lifetime-weighted simulation.

Sample	2016	2017	2018
TIS	2.15	2.14	2.12
xTOS	2.15	2.14	2.13

Table 5.9: Full breakdown of the systematic uncertainties going into the computation of the single event sensitivity, α. We are largely dominated by the L0 trigger.

Source	Relative effect (%)
$\mathcal{B}(K^0_{\rm S} \to \pi^+\pi^-)$	0.07
MC Eff. Corrections	4.4
L0 Trigger Eff.	10 (TIS) , 21(xTOS)
HLT Trigger Eff.	11
Muon ID	3.3
Tracking	1.2
$K^0_{\rm S} \to \pi^+\pi^-$ Signal yield	1

We fit the 4μ invariant mass of both datasets simultaneously to a probability density function that includes a signal and a background component. They are weighted by respective yield parameters, $n_{sig,bkg}$, such that $n_{sig} + n_{bkg} = n_T$, where n_T is the total number of entries in the dataset. In addition, the signal yield is defined as $n_{sig} = \mathcal{B}/\alpha$, where $\mathcal{B}$ is a shared parameter across both fit categories, but α is the value indicated above for each one. The systematics are incorporated into the fit as a Gaussian constraint on the α parameters. The fit can be found in Fig. 5.13.

No signal is observed. We can set an upper limit on our branching fraction by profiling the likelihood of the $\mathcal{B}$ parameter, and integrating up to 90% of the area. As a result, we obtain

$$\mathcal{B}(K^0_{\rm S} \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}. \tag{5.38}$$

Using Eq. (5.36), we can turn our $K^0_{\rm S}$ normalization factors into those corresponding to $K^0_{\rm L} \to \mu^+\mu^-\mu^+\mu^-$. We can use this to also compute the expected limit of this decay, obtaining

$$\mathcal{B}(K^0_{\rm L} \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}. \tag{5.39}$$

The $K^0_{\rm S} \to \mu^+\mu^-\mu^+\mu^-$ result is the most stringent upper limit set by the LHC to date. If anything, it represents the reaching power of the LHCb detector by the end of Run 2, and is an encouraging sign for very rare decays searches moving forward.

Prospects plots for future runs of the LHC are provided in Fig. 5.14. By the end of Run 4 of data-taking, the LHCb experiment is expected to have collected 50 fb^{-1} [192], a dataset for which the dark-sector model prediction will be in reach. By the end of Run 6, the LHCb detector is planned to have reached 300 fb^{-1} of collected data, which will be enough to test the SM prediction.

On the other hand, it is not expected that the increase in luminosity will improve the chances of observing $K^0_{\rm L} \to \mu^+\mu^-\mu^+\mu^-$, as predicted by the Standard Model, even at the

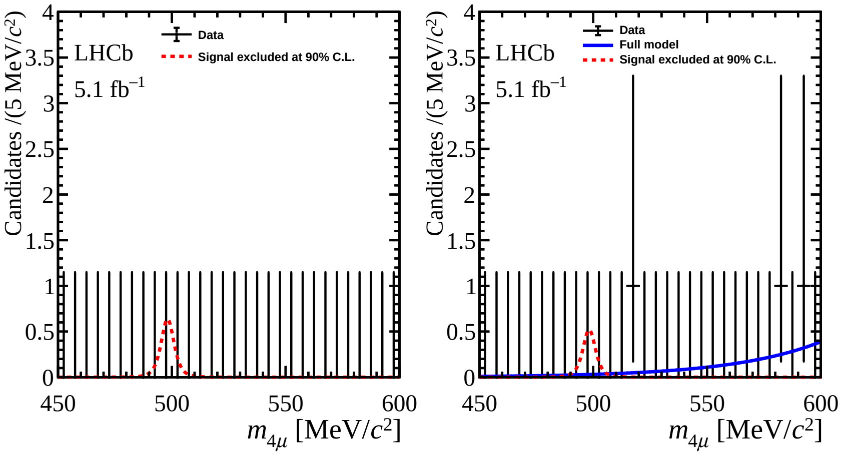


Figure 5.13: Mass distribution of the unblind data. Mass distribution of the observed candidates in data. Left (Right): $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ candidates in the xTOS (TIS) trigger category. The blue line is a simultaneous exponential fit to both categories, while the red line (null) accounts for the signal events. The plots above include the selected candidate events for each category, while the plots below are the corresponding pull histograms. Figure taken from [116] under Creative Commons license.

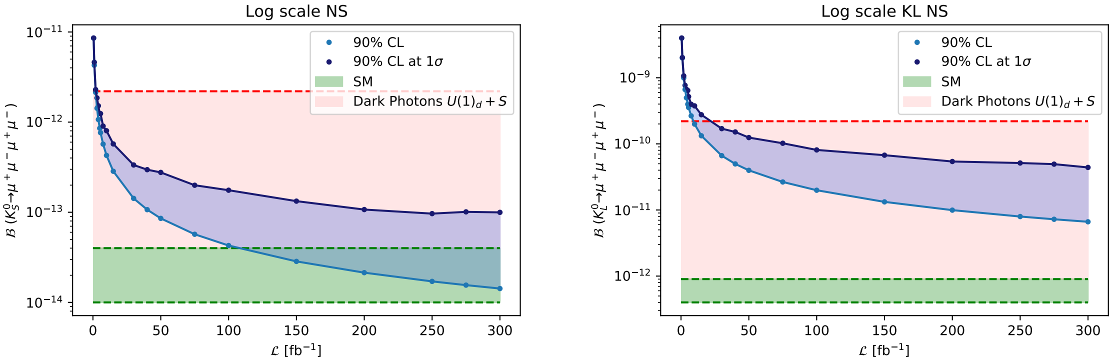


Figure 5.14: Extrapolation of the expected limit for the LHCb upgrades, taking into account that the trigger efficiency is expected to be very close to 100%. The red band marks approximately the region where Dark Photon contributions are large, while the green region shows the SM prediction. In this extrapolation, systematics are assumed to be reduced to a negligible level. Left: $K^0_S \to \mu^+\mu^-\mu^+\mu^-$, right: $K^0_L \to \mu^+\mu^-\mu^+\mu^-$.

end of Run 6, due to the lifetime requirements of the K^0_L .

6

Studies of anomalous η and η' decays at LHCb

6.1 Introduction to light-meson decays

6.1.1 Theoretical context

The η and η' mesons are an admixture of different possible combinations of quarks. In particular, $\eta^{(\prime)} = c_1(u\overline{u} + d\overline{d}) + c_2(s\overline{s})$ [47], where $c_1(\eta) \simeq \frac{1}{\sqrt{6}}$, $c_2(\eta) \simeq -\frac{2}{\sqrt{6}}$, and $c_1(\eta') \simeq c_2(\eta') \simeq \frac{1}{\sqrt{3}}$ [193].

Chiral perturbation theory, χPT, has been proven to be an adequate perturbative effective theory to describe interactions at low energies, especially those involving light quarks. The model is built upon neglecting the masses of the lighter quarks, u, d, and s. While this can be rather accurate for particles containing just u and d quarks, interactions with particles containing an s, like the η or η' mesons, can be harder to describe theoretically. This makes the study of decays involving these light mesons an excellent probe for the strong interaction at low energies.

Neglecting the masses of the light quarks means that the left-handed and right-handed quark fields are decoupled from each other in the QCD Lagrangian, which in classical field theory means that the axial current $J^\mu_5 = \overline{\psi}\gamma^\mu\gamma^5\psi$ is conserved, $\partial_\mu J^\mu_5 = 0$. By Noether's theorem, this yields the axial symmetry. However, at the quantum level, and involving loop calculations, the axial current J^μ_5 is not conserved due to interactions with the gauge fields $F_{\mu\nu}$,

$$\partial_\mu J^\mu_5 = \frac{e^2}{16\pi^2}\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}.$$

This is known as the chiral anomaly.

Two types of Feynman diagrams contribute to this anomaly: three-point (triangle) or

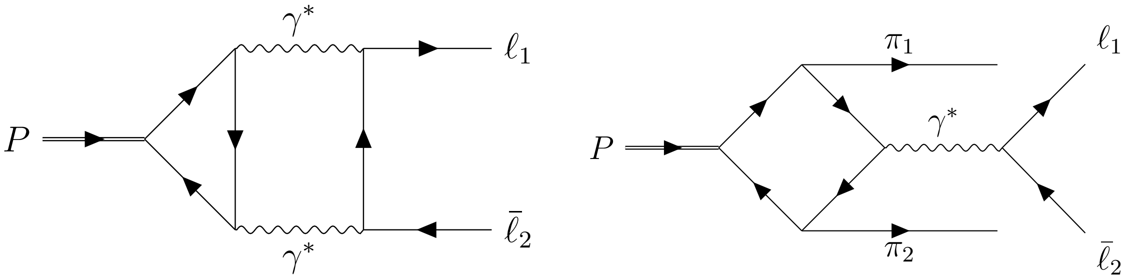


Figure 6.1: Some of the Feynman diagrams concerning the two- and four-body decays of pseudoscalar mesons η and η', generically represented by P.

four-point (box) diagrams (see Fig. 6.1). The triangle anomaly, for instance, explains the $\pi^0 \to \gamma\gamma$ decay, which would be forbidden if the current was conserved.

The $\eta \to \mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$ decays are also induced by the triangle anomaly, as seen on the left diagram of Fig. 6.1. They have very small predicted branching fractions in the SM, which makes both of them sensitive to Beyond the SM (BSM) physics effects. There are theoretical ambitions to provide a single explanation to the observed branching fractions and spectra of different η and η' radiative decays within the framework of χPT [194]. BSM explanations involving dark photons could help solve this puzzle [195].

The theoretical prediction of the $\eta \to \mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$ branching ratios is challenging, involving complex soft-QCD calculations, and has been subject to an intense study in the literature [196–209]. Respectively, they range in $(4-5)\times 10^{-6}$ and $(1-2)\times 10^{-7}$. We will focus here in the most accurate prediction to date, given in Ref. [209], which uses the machinery of Canterbury approximants, providing a systematic data-driven description of the underlying hadronic physics and partly challenging the conclusions reached by a more simple use of χPT. The predictions from Ref. [209] can be found in Tab. 6.1.

Concerning BSM physics-related effects, these decays are sensitive to both new axial and pseudoscalar contributions [209]. Interestingly here, these are more likely to appear in $\mu^+\mu^-$ final states compared to e^+e^-. Furthermore, the axial contribution is expected to contribute equally for the η and η' decays, while the pseudoscalar one will depend on the relative mass of the η' or η compared to the mass of the new particle mediating the decay. Examples of models which could potentially modify the branching fraction of the $\eta \to \mu^+\mu^-$ or $\eta' \to \mu^+\mu^-$ decays are provided in Refs. [210–214].

The $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ decays are induced by the box anomaly (see right diagram on Fig. 6.1) and can be described through the hidden gauge model [215], the chiral unitary approach, which extends the use of χPT [216] or, more simply, by using the vector–meson dominance model [217]. The prediction for the $\mathcal{B}$ of these decays obtained by Ref. [216] can be found in Tab. 6.1. The expectations by all three references for $\mathcal{B}(\eta' \to \pi^+\pi^-\mu^+\mu^-)$ are in the $(1.5-2.5)\times 10^{-5}$ range, while $\mathcal{B}(\eta \to \pi^+\pi^-\mu^+\mu^-)$ predictions range between $(7.5-12)\times 10^{-9}$.

As mentioned above, both of these decays exhibit a contribution from the box anomaly, as well as a dominant $\rho \to \pi^+\pi^-$ contribution. This has implications for the angular distribution of the final-state particles, and will require the use of a specific generator model [218], implemented in EvtGen under the name `ETA_LLPIPI` [219].

But perhaps the most interesting application of the studies of η and η' decays concerns their contributions to solving the $(g-2)_\mu$ puzzle, as it was alluded in Sect. 1.2.1.1. The pseudoscalar transition form factor (TFF) that describes the coupling of $\eta^{(\prime)}$ to two virtual photons are very prominent in SM quantum corrections to $(g-2)_\mu$ [52, 220, 221]. These mesons, along with the π^0, are the lightest intermediate resonances contributing to the hadronic light-by-light scattering in $(g-2)_\mu$. The precision on TFFs can be significantly improved by measuring branching fractions of η and η' anomalous decays [206, 221–223].

The dominant contribution to the $\eta \to \mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$ decays arises from an intermediate two-virtual photon state. This connects them directly with the TFFs, as described above, therefore directly making the measurements their $\mathcal{B}$ useful to measure those and provide important info on the theoretical determination of $(g-2)_\mu$. Both $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ also provide useful information for the theoretical calculation of $(g-2)_\mu$, given this decay is related to the TFFs in the context of dispersion theory [221].

The current chapter of the thesis will present searches for four decays: $\eta \to \mu^+\mu^-$, $\eta' \to \mu^+\mu^-$, $\eta \to \pi^+\pi^-\mu^+\mu^-$, and $\eta' \to \pi^+\pi^-\mu^+\mu^-$.

6.1.2 Experimental status

Experimentally, the first observation of the $\eta \to \mu^+\mu^-$ decay dates back to 1969 [224]. Its branching fraction has been measured with increased accuracy since then [225, 226], with the current average being fully dominated by the most recent results of SATURNE II [227], which measured $\mathcal{B}(\eta \to \mu^+\mu^-) = (5.7 \pm 0.7 \pm 0.5) \times 10^{-6}$, where the first uncertainty is statistical and the second systematic. This result is in statistical agreement with the theoretical prediction mentioned earlier, with the central value around one sigma above the prediction. Note the theoretical uncertainty is smaller than the experimental one. As for $\eta' \to \mu^+\mu^-$, no experimental search, has been performed to date.

Both of these four-body decays have been investigated by the BESIII Collaboration in recent years. In 2021, they measured $\mathcal{B}(\eta' \to \pi^+\pi^-\mu^+\mu^-) = (1.97 \pm 0.33 \pm 0.18) \times 10^{-5}$ [228], where the first uncertainty is statistical and the second systematic. This measurement improved previous searches performed by BESIII [229] and CLEO [230], in the same paper where the $\eta' \to \pi^+\pi^-e^+e^-$ decay was measured for the first time.

In January 2025, BESIII also published a new search for $\eta \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-e^+e^-$ [231]. They set an upper limit for the for $\eta \to \pi^+\pi^-\mu^+\mu^-$, $\mathcal{B}(\eta \to \pi^+\pi^-\mu^+\mu^-) < 4.0 \times 10^{-7}$, and measured $\mathcal{B}(\eta \to \pi^+\pi^-e^+e^-) = (3.07 \pm 0.12 \pm 0.19) \times 10^{-4}$.

Apart from this, in their results, BESIII also measured the *CP*-odd $\pi^+\pi^-e^+e^-$ decay planes angular asymmetry for both the η and η' decays. They found that in both cases it was consistent with 0. Ref. [221] suggests these measurements could be extended to the muonic η' sector, as a way to test new unconventional *CP*-odd mechanisms, beyond the Kobayashi-Maskawa paradigm.

Finally, these η and η' four-body decays could also be mediated by BSM physics through decays such as $\eta' \to \pi^+\pi^-A'$, where A' would be a dark photon [232] or an axion-like particle [233, 234], decaying to a pair of muons ($A' \to \mu^+\mu^-$). In their recent result of $\mathcal{B}(\eta \to \pi^+\pi^-e^+e^-)$ [231], BESIII performed a search for these resonances in the di-electron spectrum, setting upper limits on the branching fraction of $\eta \to \pi^+\pi^-a$, $a \to e^+e^-$, with a being an axion-like particle in the mass range of $5 - 200$ MeV/c^2.

The measurement of $\eta' \to \pi^+\pi^-\mu^+\mu^-$ performed in this analysis will also include a search for potential BSM resonances in the $\mu^+\mu^-$ spectrum. Note the analysis of the $\mu^+\mu^-$ spectra will be also useful for the determination of the TFFs described above [216, 221].

Table 6.1: Theoretical predictions and best measurements of the branching fractions of the decays of interest in this analysis. The corresponding references, explained in the text, appear next to each result.

Channel	**Theoretical prediction**	**Best measurement of $\mathcal{B}$**
$\eta \to \mu^+\mu^-$	$(4.72^{+0.05}_{-0.21}) \times 10^{-6}$ [209]	$(5.7 \pm 0.9) \times 10^{-6}$ [227]
$\eta' \to \mu^+\mu^-$	$(1.36^{+0.29}_{-0.26}) \times 10^{-7}$ [209]	-
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	$(1.57^{+0.40}_{-0.47}) \times 10^{-5}$ [216]	$(1.97 \pm 0.38) \times 10^{-5}$ [228]
$\eta \to \pi^+\pi^-\mu^+\mu^-$	$(7.5^{+1.8}_{-0.7}) \times 10^{-9}$ [216]	$< 4.0 \times 10^{-7}$ [231]

6.2 η and η' decays at LHCb

Ref. [222] defines an optimal strategy to measure the $\eta \to \mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$ decays at LHCb. This consists in searching for these channels through charm meson decays, which are also available as an excellent source of normalization channels.

Indeed, the high branching fractions of $D_s^+ \to \pi^+\eta$, $D_s^+ \to \pi^+\eta'$ as well as, to a lesser extent, $D^+ \to \pi^+\eta$ and $D^+ \to \pi^+\eta'$, provide a high-statistics, low-background source of η and η' mesons. In the same regard, the $D_s^+ \to \pi^+\phi(\mu^+\mu^-)$ and $D^+ \to \pi^+\phi(\mu^+\mu^-)$ decays are excellent for normalization, since their $\mathcal{B}$ is known with good precision and they include the same final-state particles as the signals, which helps cancel most of the systematic uncertainties.

A similar strategy can be also used to search for $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$, although the signal channel, $D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$ includes two additional particles compared to the normalization, $D_s^+ \to \pi^+\phi(\mu^+\mu^-)$. $D_s^+ \to \pi^+\pi^-\pi^+\phi$ was also studied as a normalization option for the measurement of the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ decays, since both signal and normalization channels contain the same final-state particles and its $\mathcal{B}$ is known with reasonable precision. However, the angular distribution of $D_s^+ \to \phi(\mu^+\mu^-)\pi^+\pi^-\pi^+$ is hard to parameterize, since the decay contains a significant $D_s^+ \to \rho(\pi^+\pi^-)\phi(\mu^+\mu^-)\pi^+$ contribution (among others) with no proper theoretical description to our knowledge, which makes the systematic uncertainty on the reconstruction and trigger efficiencies too large to cope with.

The auxiliary branching fractions needed for the determination of the signal ones are displayed in Tab. 6.2.

Table 6.2: List of $\mathcal{B}$ of the decays relevant for the normalization of the signal channels

Decay	$\mathcal{B}$
$D_s^+ \to \pi^+\eta$	$(1.69 \pm 0.03)\%$ [235]
$D_s^+ \to \pi^+\eta'$	$(3.95 \pm 0.08)\%$ [235]
$D_s^+ \to \pi^+\phi$	$(4.5 \pm 0.4)\%$ [47]
$D^+ \to \pi^+\eta$	$(3.77 \pm 0.09) \times 10^{-3}$ [47]
$D^+ \to \pi^+\eta'$	$(4.97 \pm 0.19) \times 10^{-3}$ [47]
$D^+ \to \pi^+\phi$	$(5.70 \pm 0.14) \times 10^{-3}$ [47]
$\phi \to \mu^+\mu^-$	$(3.045 \pm 0.156) \times 10^{-4}$ [236]

This analysis will use data collected by the LHCb experiment between the years 2016 and 2018, adding up to $5.1\,\mathrm{fb}^{-1}$. Given its small luminosity and high hardware trigger requirements for low-p_{T} particles, 2015 was not considered.

The signal regions will be removed from the main part of the study to avoid the experimenter's bias. Candidates with an invariant η mass in the $(535, 565)$ MeV/c^2 interval, or an invariant η' mass in the $(940, 970)$ MeV/c^2 range, were not considered until the green light was given by an internal reviewing committee (we say we blinded the signal regions). Before this happened, tests to ensure the robustness of the final fit were performed using toy experiments dictated by SM predictions of the $\mathcal{B}$.

In the case of the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ decay, the di-muon invariant mass was also studied to perform a search for potential bumps in the $\mu^+\mu^-$ distribution that could indicate new physics contributions.

The goal of the analysis is to improve the precision of $\mathcal{B}(\eta \to \mu^+\mu^-)$ and $\mathcal{B}(\eta' \to \pi^+\pi^-\mu^+\mu^-)$, as well as measuring, for the first time, $\mathcal{B}(\eta' \to \mu^+\mu^-)$. If there is not enough sensitivity for this, an upper limit will be provided. The same applies to $\mathcal{B}(\eta \to \pi^+\pi^-\mu^+\mu^-)$.

6.3 Selecting signal candidates

Signal candidates in the $\mu^+\mu^-$ channels are built through a stripping line targeting $D^+_{(s)} \to \pi^+\mu^+\mu^-$ signatures, and isolating cases where the $\mu^+\mu^-$ invariant mass is in the vicinity of either the η or the η'. The same is true for candidates from the normalization channel, which are selected by isolating the $\mu^+\mu^-$ invariant mass region around $m(\phi) \simeq 1020$ MeV/c^2.

Dedicated stripping lines were written for the $\pi^+\pi^-\mu^+\mu^-$ decays.

In terms of trigger selections, only candidates containing either a muon or a pion that have fired the L0 trigger are considered. Specifically, to align our selections between signal and control samples, only the pion coming from the $D^+_{(s)}$ meson is required to have passed the trigger in the case of $\pi^+\pi^-\mu^+\mu^-$ channels.

The most efficient HLT1 strategy is by requiring a positive, on-signal decision on an inclusive line targeting two-track signatures. A dedicated HLT2 line targeting $D^+_{(s)} \to \pi^+\mu^+\mu^-$ decays was written before data-taking. It was used in the case of the $\mu^+\mu^-$ channels as the only HLT2 requirement.

In the case of the $\pi^+\pi^-\mu^+\mu^-$ channels, a combination of two topological lines targeting decays containing decays containing two muons or three-body final states containing muons, along with a line targeting displaced dimuon signatures, were used (see Table 6.3). Once again, to align selections between signal and control samples, $D^+_{(s)} \to (\phi \to \mu^+\mu^-)\pi^+$ datasets acting as normalization to $\pi^+\pi^-\mu^+\mu^-$ signal channels were required to pass the same HLT2 selections as their $\eta^{(\prime)}$ counterparts.

Since there are, as mentioned, two stripping lines, each targeting a different final-state set of particles, all of the selections across channels are further aligned offline by applying the non-corresponding stripping cuts to each sample. This is found to be an efficient strategy to fight combinatorial background.

Additional cuts, including one on the absolute value of the helicity angle of the $\mu^+\mu^-$ structure to be below 0.97, and a requirement for the pion coming from the $D^+_{(s)}$ not not pass `IsMuon`, are also applied.

Table 6.3: Cuts applied by the HLT lines used in this analysis. DLS means Decay Length Significance. GEC means Global Event Cuts, a requirement on some of the parameters of the event, like the number of VELO hits. The Tight state allows up to 3000 VELO hits, while Loose allows up to 6000. The topological lines are best explained in Ref. [237].
The HLT1 line selects pairs of tracks that do not come from the PV, as given by the IP_{χ^2} requirement, but that do come from a common secondary vertex (though they do not have to be the only tracks coming from that SV). There is also a selection on a classifying variable to differentiate them from combinatorial background.
The `Hlt2RareCharmD2PiMuMuOS` line selects $\mu^+\mu^-\pi^\pm$ coming from a D meson, as indicated by the invariant-mass requirements, while `Hlt2DiMuonDetached` selects $\mu^+\mu^-$ candidates reconstructed as good tracks coming from a common parent.

HLT cut	`Hlt1TwoTrackMVA`	`Hlt2RareCharmD2PiMuMuOS`	`Hlt2DiMuonDetached`
Inv. mass daughters	—	> 1763 MeV/c^2	—
Inv. mass $\mu^+\mu^-$	—	> 250 MeV/c^2	—
Daughters p	> 5000 MeV/c	> 2000 MeV/c	—
Daughters p_T	> 500 MeV/c	> 300 MeV/c	—
Daughters IP_{χ^2}	> 4	> 5	> 9
Daughters Ghost Prob.	< 999	—	—
Daughters track χ^2	< 2.5	< 3	—
Daughters Track Types	—	Long	—
Daughters DLS	—	—	> 7
Muon `IsMuon`	—	True	True
Muon p_T	—	—	> 300 MeV/c
$\mu^+\mu^-$ p_T	—	—	> 600 MeV/c
$\mu^+\mu^-$ SV_{χ^2}/ndof	—	—	< 9
Muon Track χ^2	—	—	< 5
Mother `ADAMASS`	—	< 200 MeV/c^2	—
Mother `DOCA`	—	< 0.15 mm	—
Mother SV_{χ^2}/ndof	< 10	< 5	—
Mother `FD` χ^2	—	> 20	—
Mother p_T	> 2000 MeV/c	—	—
Mother IP_{χ^2}	—	< 25	—
Mother DIRA	—	> 0.9999	—
ETA	> 2, < 5	—	—
DIRA	> 0	—	—
GEC	Loose	—	—
BDT Cut	> 0.95	—	—

The combinatorial background is further removed by using a multivariate decision algorithm, which will be explained in Sect. 6.4.4. This is only needed in the case of three of the four signal channels. Prior selections on the $\eta \to \pi^+\pi^-\mu^+\mu^-$ decay already remove most of the background contributions to the sidebands, and an optimized selection on the PID of the muon tracks suffices to reduce it even further.

The efficiencies of all of these selections were computed using dedicated simulated samples for each of the channel, and can be found in Table 6.4 and Table 6.5.

Table 6.4: Breakdown of the efficiencies for the $\mu^+\mu^-$ channels. The efficiency of each step is computed on top of that of the previous row. The last two rows are split into two because there are two different sets of normalization samples for each D meson. That indicated with (η) has the classifier from the corresponding η applied, while the (η') row has the classifier from that channel.

Step	$D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$	$D^+ \to \pi^+\eta'(\mu^+\mu^-)$	$D_s^+ \to \pi^+\eta(\mu^+\mu^-)$	$D^+ \to \pi^+\eta(\mu^+\mu^-)$	$D_s^+ \to \pi^+\phi(\mu^+\mu^-)$	$D^+ \to \pi^+\phi(\mu^+\mu^-)$
Generation	0.19489(10)	0.19560(10)	0.19794(10)	0.19918(12)	0.11749(18)	0.19363(27)
Reconstruction	0.39119(13)	0.38495(13)	0.36670(33)	0.35871(38)	0.46709	0.39458(12)
Stripping	0.11406(14)	0.19041(18)	0.08806(33)	0.14316(47)	0.15388(21)	0.19989(16)
L0	0.52112(68)	0.47208(53)	0.5266(19)	0.4878(18)	0.53930(74)	0.47742(45)
HLT1	0.81363(73)	0.81443(59)	0.8270(20)	0.8344(19)	0.53930(74)	0.81652(50)
HLT2	0.80258(58)	0.80132(48)	0.8027(18)	0.8019(15)	0.80319(65)	0.80222(41)
Offline Selection	0.7123(11)	0.80634(83)	0.7343(32)	0.8260(25)	0.7470(11)	0.81912(68)
Classifier (η)	0.7951(11)	0.6726(11)	0.9514(18)	0.9110(21)	0.8226(11)	0.67787(91)
Classifier (η')					0.96278(59)	0.91200(55)
Total Efficiency (η)	$167.93(41) \times 10^{-5}$	$240.99(64) \times 10^{-5}$	$156.3(1.2) \times 10^{-5}$	$252.7(1.8) \times 10^{-5}$	$152.26(52) \times 10^{-5}$	$266.71(7.1) \times 10^{-5}$
Total Efficiency (η')					$215.84(71) \times 10^{-5}$	$359.33(87) \times 10^{-5}$

Table 6.5: Breakdown of the efficiencies for the $\pi^+\pi^-\mu^+\mu^-$ channels. The efficiency of each step is computed on top of that of the previous row. The $\phi \to \mu^+\mu^-$ modes only differ from Table 6.5 at the HLT2 level. The subsequent offline selection is the same as in the other Table, but the efficiency of the cuts over HLT2 differs.

Step	$D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	$D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	$D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	$D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	$D_s^+ \to \pi^+\phi(\mu^+\mu^-)$	$D^+ \to \pi^+\phi(\mu^+\mu^-)$
Generation	0.18956(10)	0.19020(17)	0.20936(19)	0.21008(19)	0.11749(18)	0.19363(28)
Reconstruction	0.121083(87)	0.117383(87)	0.131696(98)	0.126980(84)	0.46709	0.39457(12)
Stripping	0.05092(17)	0.12160(25)	0.06434(20)	0.14192(25)	0.15388(21)	0.19988(16)
L0	0.3215(16)	0.26561(99)	0.2381(13)	0.19702(75)	0.53930(74)	0.47742(45)
HLT1	0.8422(22)	0.7877(18)	0.6738(31)	0.5819(21)	0.8381(75)	0.85303(46)
HLT2	0.2034(26)	0.3093(23)	0.1058(23)	0.1771(20)	0.27743(99)	0.37340(68)
Offline Selection	0.6938(73)	0.6888(47)	0.554(13)	0.5727(77)	0.97024(72)	0.95352(49)
Classifier	0.9173(51)	0.9204(32)	0.9245(21)	0.9580(72)	—	—
Total Efficiency	$4.111(74) \times 10^{-5}$	$11.29(12) \times 10^{-5}$	$1.531(50) \times 10^{-5}$	$4.111(74) \times 10^{-5}$	$102.51(46) \times 10^{-5}$	$222.87(61) \times 10^{-5}$

6.4 Background contributions

6.4.1 Background sources of $\eta' \to \mu^+\mu^-$

The $\phi \to \mu^+\mu^-$ decay is 2000 times more likely to occur than $\eta' \to \mu^+\mu^-$, according to the SM. Both channels are kinematically identical, which makes them even harder to separate. Selections on the helicity angle of the di-muon system were not found to be effective, despite the ϕ being a vector meson and the η' a pseudoscalar, which yields different angular structures in their daughters' momentum distributions.

Instead, the main effort to fight this background involves the kinematic refitting of the resonant invariant mass peak known as Decay Tree Fitter (see Sect. 3.9). As seen in Fig. 6.2, this provides a good separation of the $\phi \to \mu^+\mu^-$ peak from the blind region of the $\eta' \to \mu^+\mu^-$ decay. The contribution, which corresponds to the left tail of the peak, is later modeled through an exponential *p.d.f.* (added on top of another exponential accounting for the combinatorial background).

The radiative decay $\eta' \to \mu^+\mu^-\gamma$ is also several orders of magnitude more likely to occur than $\eta' \to \mu^+\mu^-$, according to the SM [47, 209]. However, since only the muons are reconstructed, the parent D meson would not point back to the PV, and thus would have a larger impact parameter. The aforementioned selections on this variable are expected to minimize the leakage into the left sideband of the distribution, as is the use of the Decay

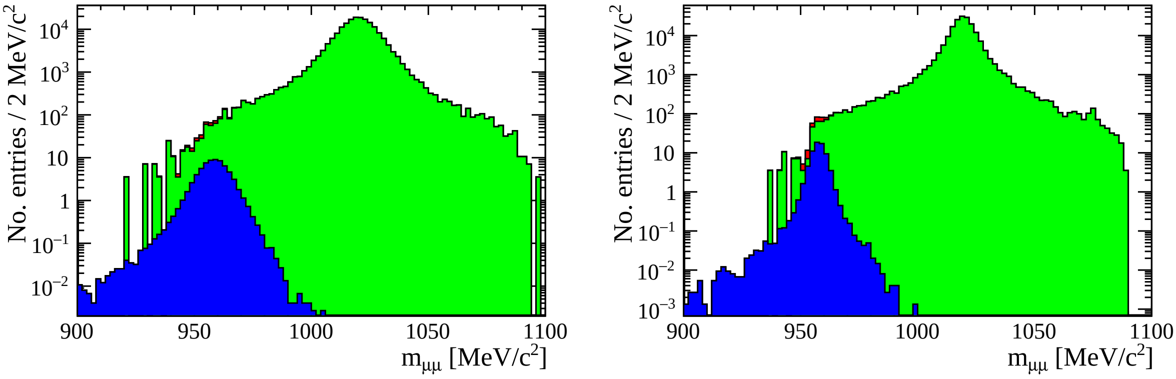


Figure 6.2: Invariant mass distribution of $\eta' \to \mu^+\mu^-$ (blue) and $\phi \to \mu^+\mu^-$ (green) when using the standard invariant mass (left) and that provided by the DecayTreeFitter (right). The red curve represents the sum of both components. The relative normalization of the yields is approximately the one expected if the SM prediction for $\eta' \to \mu^+\mu^-$ holds. The figure was computed using 2016 simulated events.

Tree Fitter variable.

6.4.2 Background sources of $\eta' \to \pi^+\pi^-\mu^+\mu^-$

The decay $\eta' \to \pi^+\pi^-(\eta \to \mu^+\mu^-)$ directly impacts our signal region. Provided $\mathcal{B}(\eta' \to \pi^+\pi^-\eta) = (42.05 \pm 0.5)\%$ [47], we find that

$$\mathcal{B}(\eta' \to \pi^+\pi^-(\eta \to \mu^+\mu^-)) = (2.4 \pm 0.38) \times 10^{-6}. \tag{6.1}$$

However, a cut on $m(\mu^+\mu^-) < 520$ MeV/c^2 is found to be 97% efficient and completely removes the background contribution.

6.4.3 Background sources of $\eta \to \mu^+\mu^-$

Similarly to what happened in $\eta' \to \mu^+\mu^-$, the partially reconstructed channel $\eta \to \mu^+\mu^-\gamma$ would populate the left sideband of the $\eta \to \mu^+\mu^-$ invariant mass distribution, given its branching ratio that is roughly 50 times larger than $\mathcal{B}(\eta \to \mu^+\mu^-)$. Though its contributions are expected to be negligible for the same reasons as in $\eta' \to \mu^+\mu^-$, out of an abundance of caution, the left sideband is taken out of the BDT training.

In turn, this was not done on the η' case, which also has a radiative decay potentially populating its left sideband. This is because otherwise we would not have a combinatorial background proxy to train the BDT.

6.4.4 Combinatorial background contributions

Aside from physical backgrounds, the main challenge of the analysis is to reduce random combinations of tracks that populate our invariant mass distributions. Dedicated multivariate classifiers were used to address this problem.

An XGBoost algorithm from the Python library of the same name [238] was trained on each of the six data samples to enhance the separation between signal and combinatorial background components: $\eta' \to \mu^+\mu^-$, $\eta \to \mu^+\mu^-$, $\eta' \to \pi^+\pi^-\mu^+\mu^-$, coming from either a D_s^+ or a D^+.

Simulated samples of the corresponding decay were used as a proxy for signal, while data from the mass sidebands of each decay was used as a proxy for background. In the case of $\eta' \to \mu^+\mu^-$ channels, only the left sideband was used, due to leaks from $\phi \to \mu^+\mu^-$ on the upper mass region, while only the right sideband on $\eta \to \mu^+\mu^-$ was used out of caution due to potential leaks from $\eta \to \mu^+\mu^-\gamma$.

The simulation samples were previously reweighted using the corresponding $\phi \to \mu^+\mu^-$ channels as control samples, to ensure that the simulation properly reproduces the behavior observed in data.

A six-fold cross-validation scheme was implemented, where, in each rotation, two thirds of the samples were used in the training, and the other third was split into a testing sample and a holdout set on which the algorithm is applied after being tested.

From a technical point of view, one of the main accomplishments of this analysis and this thesis was the development and use of novel techniques to optimize the hyperparameters of the classifying algorithm in such a way that would maximize the performance and minimize overtraining.

The optimization package OPTUNA [239] was used for the first time by an LHCb analysis for this purpose. The procedure consists in finding two flagship identifiers of separation power between the two samples, as well as the overtraining, measured by the level of agreement between the training and test samples. These parameters are the F1 score of the classifier and the p-value extracted from comparing the training and test data sets.

The measure function used by OPTUNA in the optimization procedure is the harmonic mean of these two parameters. This penalizes hyperparameter configurations where one of them is really good and the other one is not. The samples used by OPTUNA for this method are the training data sets. Half the sample is used to train different hyperparameter configurations, while the other half is used to validate it.

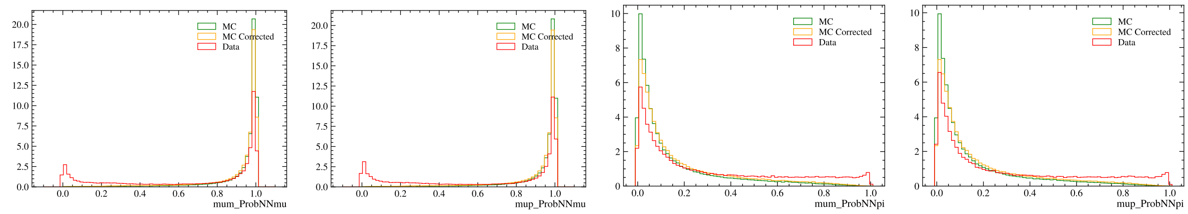
Once the set of hyperparameters is chosen, we train the algorithm with the full training sample. The set of variables used in the training mixes up kinematic, fit quality, particle identification, and isolation variables, and can be found in Table 6.6. They are aligned across all six channels, with the exception of two variables that are only used in the training of the $\mu^+\mu^-$ channels.

The PID variables, instead of being reweighted, are corrected using `PIDGen` (see Sect. 3.8.1). The isolation variables are computed by taking one (or two) additional tracks in the event and creating a new vertex with it. The difference between the χ^2 value of both the old vertex and the new vertex, using the extra track for which this difference is the smallest, yields the One (Two) Track χ^2 isolation variables, while the invariant mass of all of the tracks builds the Mass One (Two) Track χ^2 isolation variables.

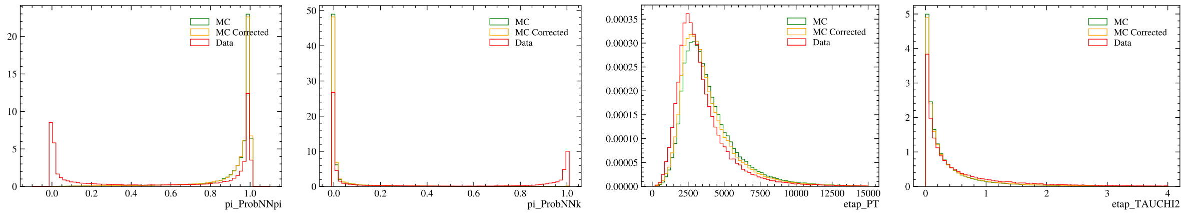
Fig. 6.3 shows the distributions of the signal and background proxies (as well as the comparison with the reweighted signal distributions) for one exemplary channel, $D^+ \to \pi^+\eta'(\mu^+\mu^-)$.

The ROC curves, as well as the distributions of the output variables, are computed using the test samples on each of the folds. The output variable is then used to implement a selection that would remove the highest amount of background while keeping as much signal as possible. See Fig. 6.4 and Fig. 6.5.

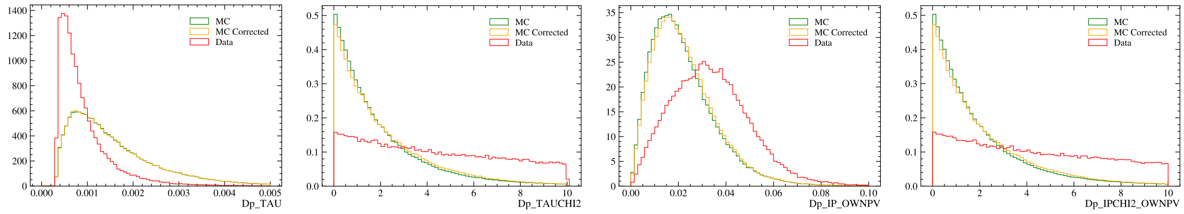
Instead of optimizing this selection, the subsequent part of the analysis was performed



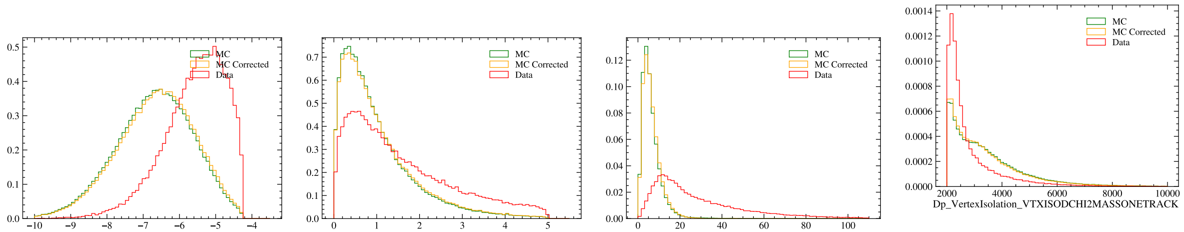
(a) μ^- `ProbNNmu` (b) μ^+ `ProbNNmu` (c) μ^- `ProbNNmu` (d) μ^+ `ProbNNmu`



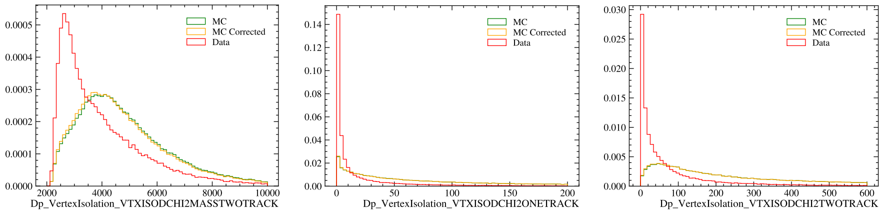
(e) π^+ `ProbNNpi` (f) π^+ `ProbNNk` (g) η PT (h) η lifetime χ^2



(i) D^+ lifetime (j) D^+ lifetime χ^2 (k) D^+ IP (l) D^+ IP$_{\chi^2}$

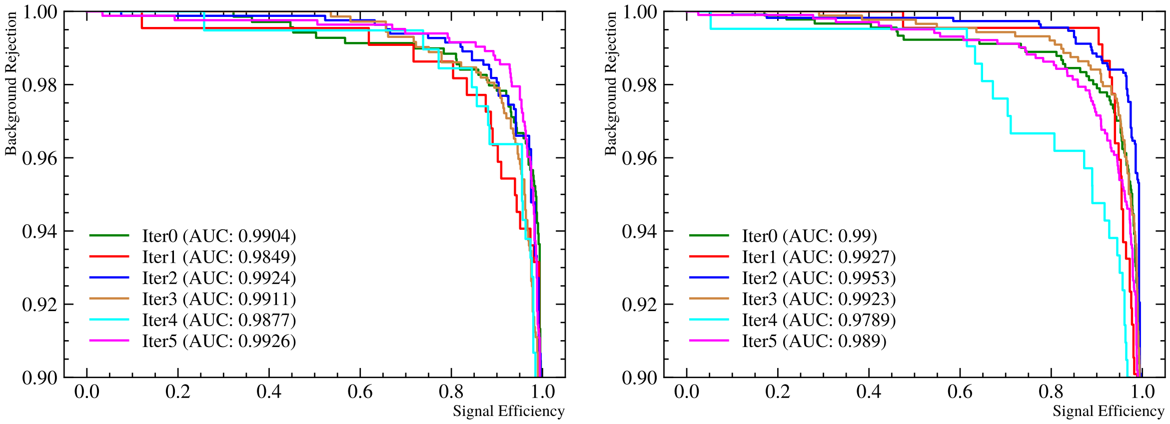


(m) D^+ DIRA (n) D^+ Vertex χ^2 (o) D^+ DTF χ^2 (p) D^+ Vertex Isolation χ^2 Mass One Track

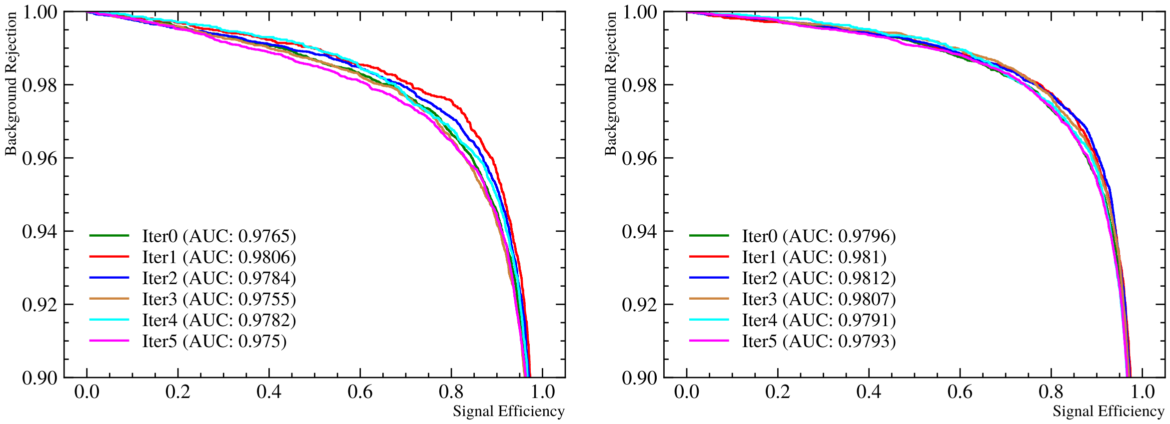


(q) D^+ Vertex Isolation χ^2 Mass Two Track (r) D^+ Vertex Isolation χ^2 One Track (s) D^+ Vertex Isolation χ^2 Two Track

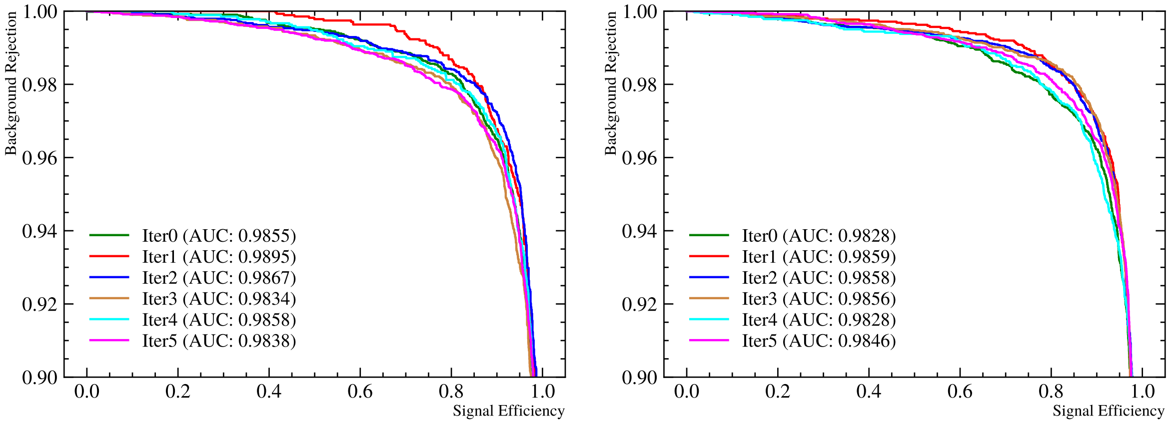
Figure 6.3: Distributions of all variables used to train $D^+ \to \pi^+\eta'(\mu^+\mu^-)$



(a) $D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$ (b) $D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$



(c) $D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$ (d) $D^+ \to \pi^+\eta'(\mu^+\mu^-)$



(e) $D_s^+ \to \pi^+\eta(\mu^+\mu^-)$ (f) $D^+ \to \pi^+\eta(\mu^+\mu^-)$

Figure 6.4: ROC curves for each of the channels

Table 6.6: Variables used in the BDT training. The variables in the first block are used separately for $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\mu^+\mu^-$ channels, whereas the rest of the variables are shared across all channels.

$\eta' \to \pi^+\pi^-\mu^+\mu^-$	$\mu\mu$ channels
	$D^+_{(s)}$ and $\eta^{(\prime)}$ lifetime χ^2
	Both of the muons' `ProbNNk`
Both of the muons' `ProbNNmu`	
Minimum π `ProbNNpi`	
π from $D^+_{(s)}$'s `ProbNNk`	
$D^+_{(s)}$ lifetime	
$D^+_{(s)}$ direction angle	
$D^+_{(s)}$ Vertex χ^2	
$D^+_{(s)}$ IP and IP χ^2	
$\eta^{(\prime)}$ PT	
$D^+_{(s)}$ Vertex Isolation χ^2 Mass One Track	
$D^+_{(s)}$ Vertex Isolation χ^2 Mass Two Track	
$D^+_{(s)}$ Vertex Isolation χ^2 One Track	
$D^+_{(s)}$ Vertex Isolation χ^2 Two Track	

using different possibilities for the cut on this variable. We choose the optimal value for the selection on $\eta' \to \mu^+\mu^-$ as that which gives the highest expected signal sensitivity. The optimal selection for both $\eta \to \mu^+\mu^-$ and $\eta' \to \pi^+\pi^-\mu^+\mu^-$ is that for which the total expected uncertainty is minimal.

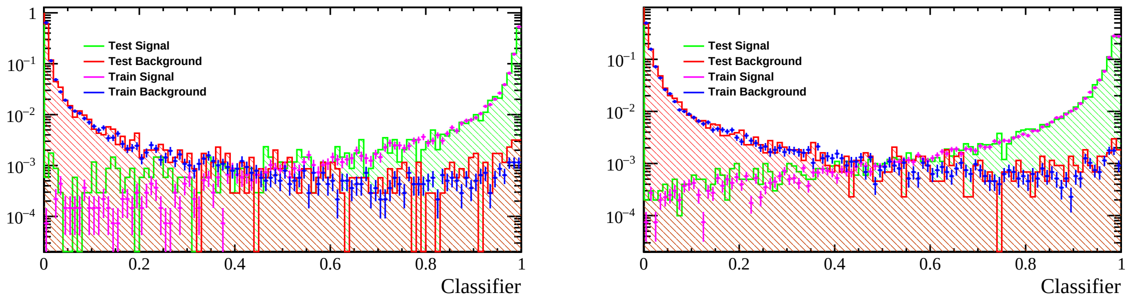
Finally, as mentioned earlier, instead of training a Boosted Decision Tree to remove the combinatorial background from the $\eta \to \pi^+\pi^-\mu^+\mu^-$ channels, it was seen that a cut on the `ProbNNmu` of the muons was good enough to perform the signal-background separation. A similar procedure for optimizing this cut was followed, wherein different variations were tried, and the optimal cut was given by that which provides the lowest value of the expected upper limit.

The figures of merit used to optimize these selections are two-dimensional, as there are selections on both the D_s^+ and D^+ to be taken into account simultaneously. The procedures to compute the expected values of the signal sensitivity, the total uncertainty, and the upper limit are provided in Sect. 6.7. The plots can be found in Fig. 6.6.

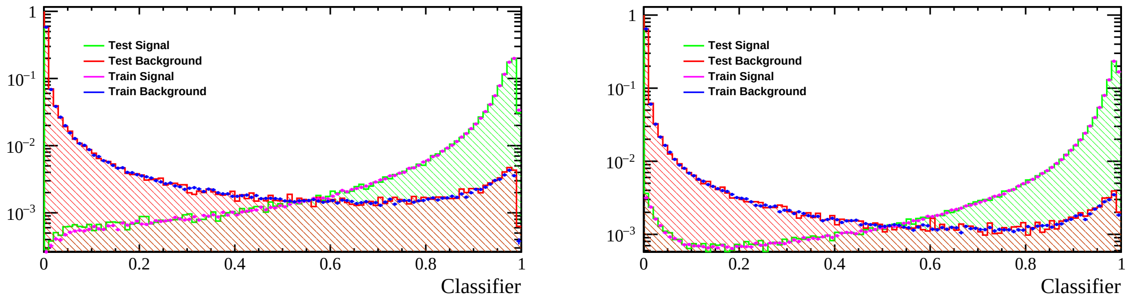
6.5 Normalization

The single event sensitivity, as also seen in Eq. (5.27), incorporates the number of candidates from the normalization channel, the ratio of efficiencies between normalization and signal, and the known branching fractions involved in the decay,

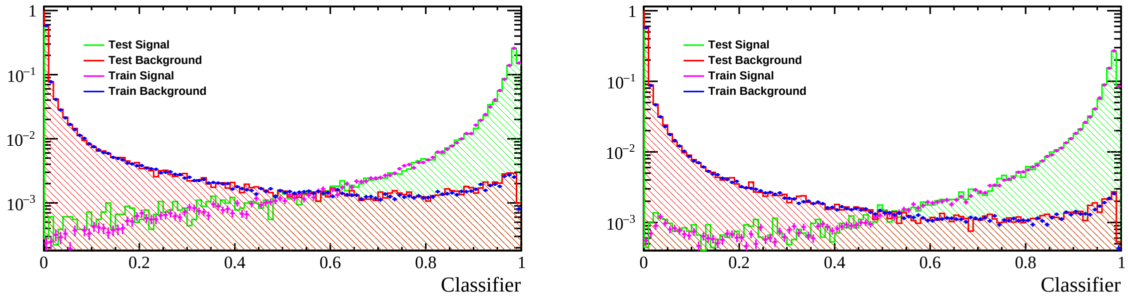
$$\alpha = \frac{1}{N_{\text{norm}}} \times \frac{\varepsilon_{\text{norm}}}{\varepsilon_{\text{sig}}} \times \frac{\mathcal{B}(D^+_{(s)} \to \phi\pi^+) \times \mathcal{B}(\phi \to \mu^+\mu^-)}{\mathcal{B}(D^+_{(s)} \to \eta^{(\prime)}\pi^+)}, \tag{6.2}$$



(a) $D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$ (b) $D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$



(c) $D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$ (d) $D^+ \to \pi^+\eta'(\mu^+\mu^-)$



(e) $D_s^+ \to \pi^+\eta(\mu^+\mu^-)$ (f) $D^+ \to \pi^+\eta(\mu^+\mu^-)$

Figure 6.5: Distribution of the output BDT variable for each of the channels

such that

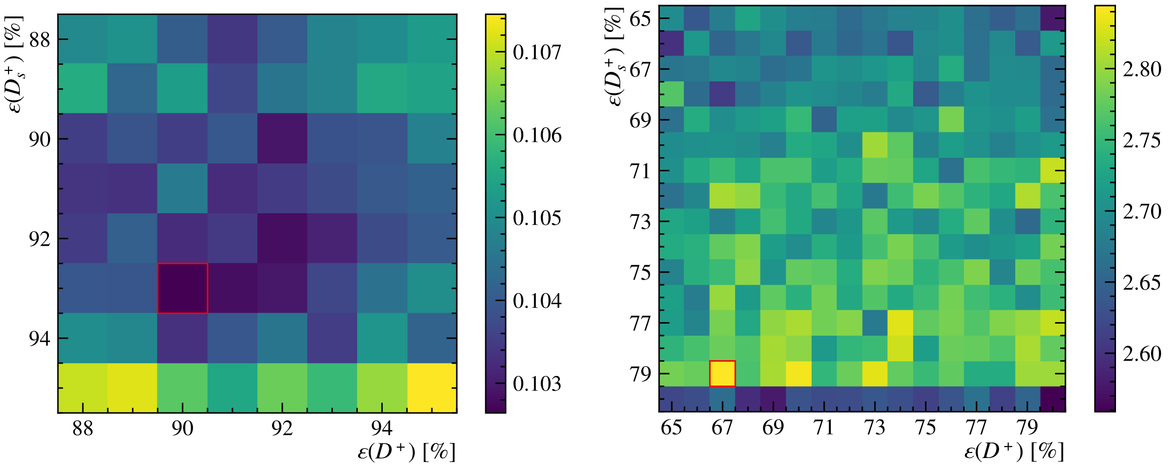
$$\mathcal{B}(\eta^{(\prime)} \to \mu^+\mu^-(\pi^+\pi^-)) = \alpha \times N_{\eta^{(\prime)} \to \mu^+\mu^-(\pi^+\pi^-)}. \tag{6.3}$$

The ratio of efficiencies can be computed from Table 6.4 and Table 6.5 for each channel.

The number of candidates in the normalization channel, N_{norm}, is computed independently for each topology.

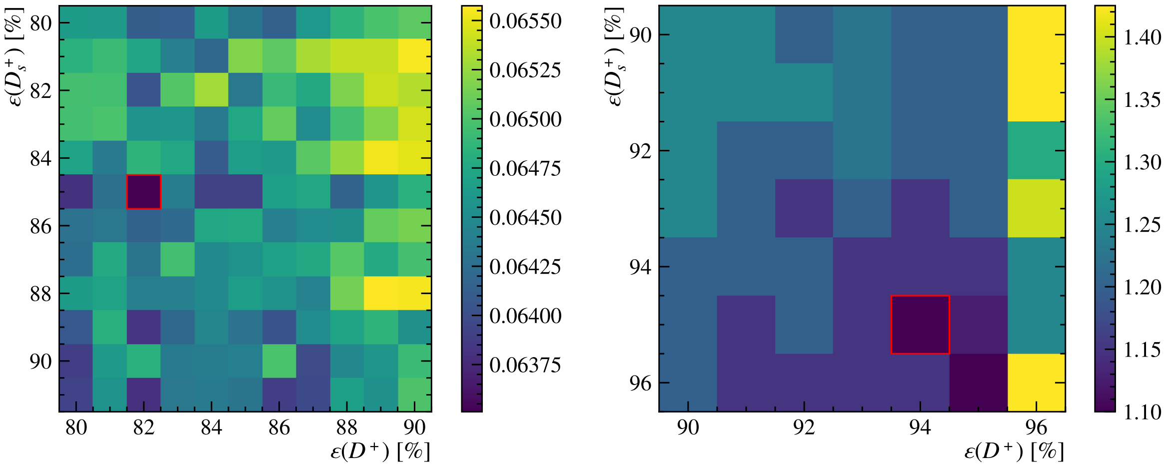
In the case of the $\mu^+\mu^-$ channels, since we apply the BDT from the corresponding signal channels (we now have four normalization samples, since we are splitting the $D_{(s)}^+$ modes into two, where each has the corresponding $\eta^{(\prime)}$ BDT applied), the background contributions are practically removed by the classifier. In the case of $\eta' \to \mu^+\mu^-$, the remaining background is completely negligible, while in the case of $\eta \to \mu^+\mu^-$, where there are some kinematic differences between signal and normalization channels (which leaves some combinatorial background behind on the latter, picked up by the fit to the ϕ mass), it is not.

In the case of the $\pi^+\pi^-\mu^+\mu^-$ channels, where no BDT is applied on the normalization



(a) $\eta' \to \pi^+\pi^-\mu^+\mu^-$

(b) $\eta' \to \mu^+\mu^-$



(c) $\eta \to \mu^+\mu^-$

(d) $\eta \to \pi^+\pi^-\mu^+\mu^-$

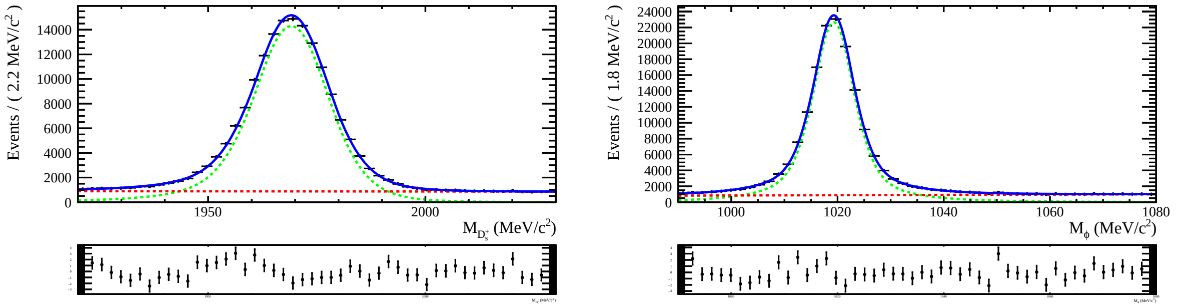
Figure 6.6: Figures of merit for each of the channels. In the case of $\eta \to \mu^+\mu^-$ and $\eta' \to \pi^+\pi^-\mu^+\mu^-$, our goal is to minimize the total uncertainty (in the z axis, in $0-1$ scale), so we base our FoM on this quantity and choose the cut that yields the minimum value. In the case of $\eta' \to \mu^+\mu^-$, we want to maximize the expected significance (in the z axis). In the case of $\eta \to \pi^+\pi^-\mu^+\mu^-$, we want to achieve the lower possible value for the upper limit on the $\mathcal{B}$ (in the z axis, with a 10^{-6} scaling factor omitted).

channel, the procedure is the same but the background contributions are larger on both fits. We take N_{norm} as the signal yield of the ϕ fit, and we assign to it a systematic that accounts for the difference with the $D^+_{(s)}$ mass fit.

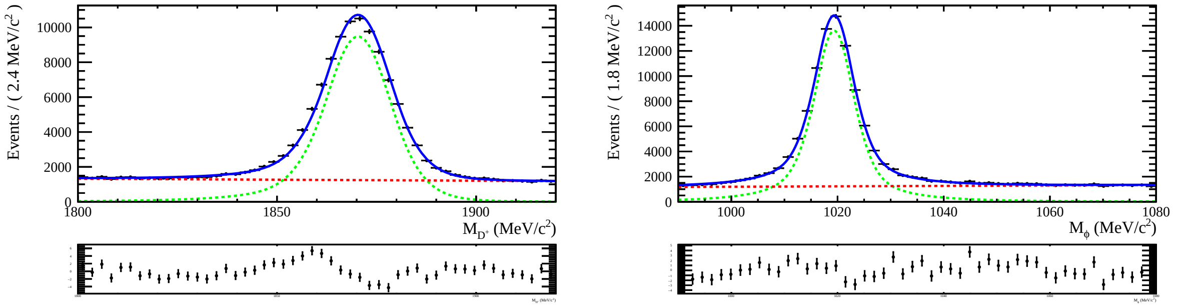
Fig. 6.7 contains the fits to the normalization channel data. See Table 6.7 for a breakdown of the fit yields for both $D^+_{(s)}$ and ϕ.

Table 6.7: Signal yields for the normalization channels, when the fit is done on the D meson mass, or the ϕ mass. The nominal value taken will be from the ϕ fit and the difference is taken as a systematic. Numbers on the systematic uncertainty of the ϕ yield are also provided in the last column.

Channel	$D^+_{(s)}$	$D^+_{(s)}$ **Norm yield**	ϕ **Norm yield**	**Syst.**	**Stat. Unc.**
$\eta' \to \mu^+\mu^-$	D^+_s	$184.4630(13) \times 10^3$	$183.98(44) \times 10^3$	0.26%	0.33%
	D^+	$74.8351(15) \times 10^3$	$74.73(28) \times 10^3$	0.14%	0.37%
$\eta \to \mu^+\mu^-$	D^+_s	$229.3380(28) \times 10^6$	$224.96(56) \times 10^3$	1.9%	0.25%
	D^+	$107.6350(28) \times 10^3$	$104.94(28) \times 10^3$	2.56%	0.36%
$\pi^+\pi^-\mu^+\mu^-$	D^+_s	$147.31(48) \times 10^3$	$146.14(48) \times 10^3$	0.86%	0.33%
	D^+	$84.93(37) \times 10^3$	$84.94(32) \times 10^3$	0.45%	0.38%



(a) Fit to D_s^+ mass for $D_s^+ \to \pi^+\phi(\mu^+\mu^-)$ (b) Fit to ϕ mass for $D_s^+ \to \pi^+\phi(\mu^+\mu^-)$



(c) Fit to D^+ mass for $D^+ \to \pi^+\phi(\mu^+\mu^-)$ (d) Fit to ϕ mass for $D^+ \to \pi^+\phi(\mu^+\mu^-)$

Figure 6.7: Fits to the D meson and ϕ meson invariant masses, for the samples used as normalization for $\pi^+\pi^-\mu^+\mu^-$ channels. These samples have the HLT2 selections aligned with those of the signal modes (see Sect. 6.3).

Table 6.8: Values and systematic uncertainty associated with each of the branching fractions used in Eq. (6.2). They are equally applied to three- and five-daughter final-state channels.

Channel	Ratio of BF	Uncertainty
$D_s^+ \to \eta\pi$	$(79.1 \pm 2.8) \times 10^{-5}$	6.15%
$D_s^+ \to \eta'\pi$	$(45.0 \pm 1.6) \times 10^{-5}$	6.23%
$D^+ \to \eta\pi$	$(33.9 \pm 1.3) \times 10^{-5}$	6.16%
$D^+ \to \eta'\pi$	$(34.2 \pm 1.6) \times 10^{-5}$	6.85%

6.6 Systematic uncertainties

The dominant source of systematic uncertainties in this analysis comes from the branching fractions of the decays that have already been measured. Most relevantly, Ref. [236] recently estimated

$$\mathcal{B}(\phi \to \mu^+\mu^-) = (3.045 \pm 0.156) \times 10^{-4}, \tag{6.4}$$

which carries an uncertainty of 5.1%. Additionally [47],

$$\mathcal{B}(D_s^+ \to \pi^+\phi) = \frac{\mathcal{B}(D_s^+ \to \phi(KK)\pi^+)}{\mathcal{B}(\phi \to KK)} = 4.50(13)\% \tag{6.5}$$

has an associated 2.9% uncertainty. The full breakdown of the final ratios of branching fractions for each of the channels and their corresponding systematic uncertainties is given in Table 6.8.

Another source of systematic uncertainty comes from the computation of the ratio of efficiencies between normalization and signal channels. These efficiencies are computed through simulated samples, and the associated binomial uncertainty is dependent on the size of the sample. This uncertainty is negligible for the $\mu^+\mu^-$ channels, but becomes larger for $\pi^+\pi^-\mu^+\mu^-$ modes given the lower efficiency of the trigger selections on these channels (see Table 6.5).

The value of N_{norm} also has an associated systematic, as provided in Table 6.7.

6.6.1 Systematics associated to data-MC corrections

Three different sources of potential differences between simulation and data were studied, with corresponding systematics associated to each of them.

6.6.1.1 Trigger efficiencies

To account for possible discrepancies between the ratio of trigger efficiencies between the signal and normalization channels as computed in MC with respect to data, we compute a corrected trigger efficiency using the TISTOS method (see Sect. 3.5.1).

We use datasets where all selection requirements are applied except for the trigger selection. We will split the procedures into trigger levels, and neglect HLT2 corrections, since selections at this level are performed offline. As such, correction factors for both L0 and HLT1 will be computed. To do so, we compute the following efficiency on our control channel data:

$$\frac{n_{TIS\&TOS}}{n_{TIS}} \tag{6.6}$$

In the case of the L0 correction factor, the TIS selection asks for any firing of the trigger at any level by at least one particle in the event outside of the signal. In the case of the HLT1 correction factor, we first ask for L0 TOS. On top of it, we require that at least one particle in the event outside of the signal has fired the software trigger (global TIS selection).

The TISTOS method requires the p_{T} spectrum to be binned, and the ratio Eq. (6.6) to be computed on data for each bin. Since our L0 selection mixes up `L0Hadron` and `L0Muon`, the choice for the binning is not immediately clear. Binning on the maximum p_{T} of these particles may be misleading, since a hadron with a higher p_{T} than the muons may not be responsible for firing the trigger, given the different thresholds on the respective lines.

Because of that, we split our L0 efficiency into two terms, accounting for the efficiency of each particle species. Using the formula of the probability of the union,

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = p(A) + p(B) - p(A) \cdot p(B), \tag{6.7}$$

where the last equality holds if there is no correlation, we can say:

$$\varepsilon(\texttt{L0Muon} \cup \texttt{L0Hadron}) = \varepsilon(\texttt{L0Muon}) + \varepsilon(\texttt{L0Hadron}) - \varepsilon(\texttt{L0Muon})\varepsilon(\texttt{L0Hadron}). \tag{6.8}$$

This way, we can bin on the muons and pions' p_{T} separately, and freely.

At the HLT1 level there is only one line, which is targeting two-track signatures, so we compute the combined p_{T} of the different pairs of particles and bin on the maximum of the three.

For each p_{T} bin, we compute the TISTOS efficiency on data. We assign to each MC event the efficiency corresponding to the p_{T} bin it belongs to, and average all efficiencies. As such, we can write the following expression for the corrected trigger efficiency of the normalization channel:

$$\varepsilon_{trig,corr} = \frac{\sum_{i=1}^{N} \varepsilon_i(p_{\mathrm{T}}^{bin})}{N} = \sum_{i=1}^{N} \frac{\left[\frac{\varepsilon^{Data}}{\varepsilon_i^{MC}}(p_{\mathrm{T}}^{bin})\right] \times \left[\varepsilon_i^{MC}(p_{\mathrm{T}}^{bin})\right]}{N} = \sum_{i=1}^{N} \frac{\left[w_i(p_{\mathrm{T}}^{bin})\right] \times \left[\varepsilon_i^{MC}(p_{\mathrm{T}}^{bin})\right]}{N} \tag{6.9}$$

where N is the total number of MC events, and $\varepsilon(p_{\mathrm{T}}^{bin})$ is the TISTOS efficiency computed (either on data or MC) inside each p_{T} bin. This method relies on the assumption that $w_i(p_{\mathrm{T}}^{bin})$ will be consistent between the signal and control channels, so we can use the last equality to also compute the corrected trigger efficiency on the signal channel, where we do not have access to the data.

With both corrected efficiencies, we compute the ratio of corrected efficiencies per year. The correction factor we then introduce is defined as $R = \frac{r_{corr}}{r_{nom}}$, where r_{corr} is the aforementioned corrected ratio, and r_{nom} is the nominal ratio computed in MC.

A systematic is computed by varying the number of PT bins, and assigned as the difference between the maximum and minimum values of R for configurations with $5 - 10$ PT bins.

Results are presented in Table 6.9.

Table 6.9: Full results of the trigger correction factor, R. The error in $\mu^+\mu^-$ channels is computed as the difference between the maximum value of R and the minimum for different binning schemes.

Channel	D_s^+			D^+		
	L0	HLT1	Total	L0	HLT1	Total
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	1.3579(80)	0.996(14)	1.352(20)	1.604(15)	1.081(28)	1.734(48)
$\eta \to \pi^+\pi^-\mu^+\mu^-$	1.6200(92)	1.2825(91)	2.078(19)	1.985(37)	1.544(44)	3.070(10)
$\eta' \to \mu^+\mu^-$	0.9599(16)	1.0036(19)	0.9633(25)	1.01155(47)	1.0040(17)	1.0156(18)
$\eta \to \mu^+\mu^-$	0.9647(13)	0.9804(33)	0.9457(35)	0.9920(13)	0.9578(90)	0.9501(90)

6.6.1.2 BDT efficiencies

To compute the systematic uncertainties of the BDT optimal selection, we shall use a control channel whose MC we will reweight to look as close as possible to the data, and compute the weighted efficiency of the BDT variable for said cut. In the case of $\mu^+\mu^-$ channels, the control channel is the same as the normalization channel we use in the analysis, $D^+_{(s)} \to (\phi \to \mu^+\mu^-)\pi^+$.

The input MC datasets of the $\mu^+\mu^-$ channels have already been reweighted to account for possible MC-data differences. The BDT systematic is computed as the difference between computing the ratio of normalization over signal channel nominal efficiencies (as given directly by the MC) and the ratio of the same GBweighted efficiencies. The results are presented in Table 6.10.

Because of the differences between this normalization channel and $D^+_{(s)} \to (\eta' \to \pi^+\pi^-\mu^+\mu^-)\pi^+$, it is unusable as a control mode for this step. In turn, we will be using to $D_s^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$, which we discarded as a normalization possibility because of the complicated angular distribution as mentioned in Sect. 6.2. Moreover, the phase space-generated MC does not present a very accurate description of the decay, but we will assume it is good enough for this purpose.

In this case, the classifier from the signal channel is applied on both the MC and the data samples from $D_s^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$, and the systematic is taken as the difference between the efficiencies on data and simulation. This is because the simulation on the control channel does not properly reproduce the angular distributions of the data, and therefore, using it to reweight our signal MC would not be representative of a good correction. We only have a stripping line to select $D_s^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$ candidates, but not $D^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$. For that reason, we will calculate the systematic on the D_s^+ mode and use it on the D^+ mode. The results are also presented in Table 6.10.

6.6.1.3 PID efficiencies

We have three PID selections whose efficiencies we are to correct: `IsMuon & PIDmu > -5`, on both muons, as well as `!IsMuon` on the pion coming from the $D^+_{(s)}$. In the case of $\eta \to \pi^+\pi^-\mu^+\mu^-$, the muons also have a selection on `ProbNNmu>0.11`. To do so, we will compute a corrected efficiency for these selections using `PIDCalib` samples: $J/\psi \to \mu\mu$ with no cut on the muons' PT in the case of the muon selections and $D^0 \to K^-\pi^+$ for the pion.

Table 6.10: Full results of the BDT systematic computations. The systematic on the $\mu^+\mu^-$ channels is computed by comparing the ratio of efficiencies (control over signal channels) when the samples are reweighted and when they are not. The systematic on the $\pi^+\pi^-\mu^+\mu^-$ channels is computed by using $D_s^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$ as a control mode, and comparing the effect of the optimal cut from the signal channel on the control mode's data and simulation. No stripping line selects $D^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$, so we assume that $D_s^+ \to \pi^+\pi^-\pi^+\phi(\mu^+\mu^-)$ is valid for both. The numbers between the D_s^+ and D^+ modes, in this case, are different because the algorithms applied are different, as are the cutting points. No BDT was trained on $\eta \to \pi^+\pi^-\mu^+\mu^-$ modes.

Channel	D_s^+	D^+
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	1.00%	0.36%
$\eta' \to \mu^+\mu^-$	1.29%	0.023%
$\eta \to \mu^+\mu^-$	0.37%	1.02%

We use two variables for our binning, the track's total momentum, `P`, and its pseudorapidity, `ETA`. To compute the `P` bins, we divide the spectrum of the variable into a given number of bins, such that each has a similar sum of `sWeights`. Subsequently, the variable `ETA` is divided into another given number of bins, following the same strategy for each individual `P` bin, i.e., for candidates inside the first `P` bin, we once again divide the spectrum of the variable into a given number of bins, such that each has a similar sum of `sWeights`. This is to account for correlations between both variables, which leave empty bins in the higher `P` and lower `ETA` regions, causing problems should a rectangular binning be applied.

We assign a weight to each bin equal to the efficiency of the PID cuts computed on the control data sample.

In our MC, we have reconstruction-only datasets, meaning without any stripping cuts applied. We apply our stripping selections, except for the PID requirements. Then, we assign to each event the corresponding efficiency from the calibration according to the 2D bin it belongs to. Finally, we compute the corrected efficiency by averaging these weights.

Doing this on our signal and normalization channels, we compute a corrected ratio of the corresponding PID efficiencies for each of the track type (muon or pion). The global corrected ratio is the product of the corrected ratios of the muons and the pion.

Multiple binning schemes were tested for both variables, with the number of `P` bins ranging from 3 to 9, and the number of `ETA` bins ranging from 3 to 7. Half the difference between the maximum and the minimum corrections is taken as the systematic. The central value is assigned as the correction factor computed for 6 `P` bins and 5 `ETA` bins, since it is somewhere in the middle of ranges. The final results are given in Table 6.11.

6.6.2 Mass model systematics

To evaluate a systematic uncertainty associated with the choice of a double-sided Crystal Ball to model the distribution of the data set, we generated 1500 toys using an alternate model, a sum of two Gaussian *p.d.f.s*, and fit them to the same double-sided Crystal Ball *p.d.f.*. This is meant to describe a scenario in which the data follows a different distribution, and to evaluate the error that using our nominal *p.d.f.* would have.

The systematic is taken by looking at the residuals of the toys, divided by the SM $\mathcal{B}$

Table 6.11: Full results of the PID systematic computations.

Channel	μ **Corrected ratio**	π **Corrected ratio**	**PID Corrected ratio**	**PID systematic**
$D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	0.975(11)	1.00036(59)	0.975(11)	1.09%
$D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	1.003(10)	1.00037(76)	1.003(10)	1.02%
$D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	0.955(16)	1.00248(91)	0.958(17)	1.73%
$D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	0.982(16)	1.00299(10)	0.985(17)	1.68%
$D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$	0.9699(21)	0.99895(50)	0.9688(22)	0.22%
$D^+ \to \pi^+\eta'(\mu^+\mu^-)$	0.9951(11)	0.99999(46)	0.9951(11)	0.13%
$D_s^+ \to \pi^+\eta(\mu^+\mu^-)$	0.9718(23)	0.9991(11)	0.9709(26)	0.27%
$D^+ \to \pi^+\eta(\mu^+\mu^-)$	0.9983(14)	1.0002(13)	0.9985(19)	0.19%

Table 6.12: Systematics associated to the wrong estimation of the model describing the signal data set. They are computed by generating 1500 toys with two different *p.d.f.s*, a double Gaussian and a double-sided Crystal Ball, and fitting both options to the latter model. The deviation of the residuals from 0 is taken as a systematic.

Channel	**Systematic**
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	0.64%
$\eta' \to \mu^+\mu^-$	2.6%
$\eta \to \mu^+\mu^-$	0.063%

prediction. The systematic is the deviation of the mean of the pulls with respect to 0.

This is done for the three channels where some signal candidates are expected in the blind region. As seen in Table 6.12, the contribution is negligible. For $\eta \to \pi^+\pi^-\mu^+\mu^-$, this method is more complicated to implement. Since no signal candidates are expected in the blind region, we instead ran the expected limit with either option modeling the signal component. The result was the same, so we will neglect the systematic. Fig. 6.8 contains the residual plots for each of the channels.

As far as the normalization channels are concerned, the nominal values of N_{norm} are taken from a fit to the ϕ mass to a double-sided Crystal Ball. Since the ϕ has a natural width, a good alternative is typically a Voigtian *p.d.f.*, which is the convolution of a Breit-Wigner distribution, which takes into account said natural width, and a Gaussian function, which models the resolution.

The values of N_{norm} were recomputed for the $\pi^+\pi^-\mu^+\mu^-$ normalization channels, since, as seen in Sect. 6.5, the fits to the ϕ mass in the $\mu^+\mu^-$ channels have a background component compatible with 0, and therefore, a fit for this mass variable is not strictly needed.

Table 6.13 contains a comparison of N_{norm} using the different *p.d.f.s*, and includes the systematic associated to it. The systematic is added quadratically to that computed in Sect. 6.5.

6.6.3 Breakdown of all systematics

The main error contribution to the computation of the single event sensitivity comes from the systematic uncertainties. Only the yields from the fits to the normalization channels add a statistical uncertainty to α, and since it is highly negligible when compared to the

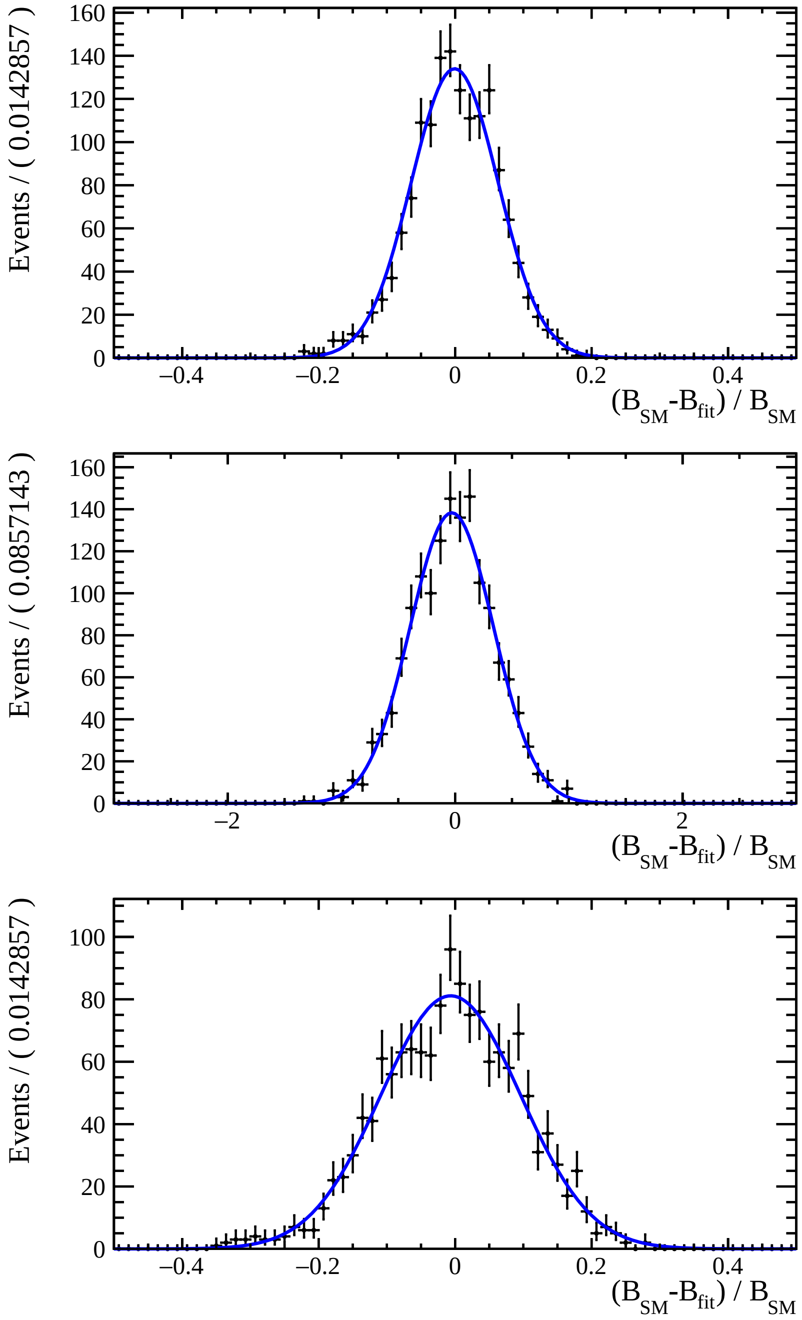


Figure 6.8: Residual plots comparing the expected SM value of the $\mathcal{B}$ for each channel, with the one given by the fit, for 1500 toys generated using a sum of two Gaussians. We divide by the SM prediction to estimate the associated systematic.

Table 6.13: Signal yields on the normalization channels of the $\pi^+\pi^-\mu^+\mu^-$ samples (*i.e.*, $D^+_{(s)} \to (\phi \to \mu^+\mu^-)\pi^+$ with the corresponding HLT2 selection) for the nominal *p.d.f.*, a double-sided Crystal Ball (DSCB for short in the table), and a Voigtian. The relative difference is taken as a systematic.

Channel	N_{norm} **from DSCB**	N_{norm} **from Voigtian**	**Mass model systematic**	**Total** N_{norm} **Systematic**
$D^+_s \to \pi^+\phi(\mu^+\mu^-)$	125.57×10^3	122.96×10^3	2.1%	2.29%
$D^+ \to \pi^+\phi(\mu^+\mu^-)$	77.11×10^3	74.77×10^3	3.0%	3.03%

Table 6.14: Breakdown of the systematic uncertainties contributing to the computation of the single event sensitivity. The main source of systematic comes from the branching fraction ratio. The Norm fits column refers to the uncertainty assigned to the difference between normalization fits to D_s^+ and ϕ invariant masses. In the case of $\mu^+\mu^-$ channels, we are applying the BDT from the corresponding signal channel, so the background is removed and we simply count the number of surviving candidates. We take a conservative 1% systematic.

Channel	**BR Ratio**	**Trigger**	**BDT**	**PID**	**Norm fits**	**Eff. Ratio**	**Total**
$D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	6.23%	1.5%	1.0%	1.09%	0.86%	1.84%	6.88%
$D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	6.85%	2.77%	0.36%	1.02%	0.45%	1.12%	7.56%
$D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	6.15%	0.91%	—	1.73%	0.86%	3.26%	7.28%
$D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	6.16%	3.37%	—	1.68%	0.45%	1.72%	7.44%
$D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$	6.23%	0.49%	1.29%	0.22%	1.0%	0.47%	6.48%
$D^+ \to \pi^+\eta'(\mu^+\mu^-)$	6.85%	0.36%	0.02%	0.13%	1.0%	0.37%	6.94%
$D_s^+ \to \pi^+\eta(\mu^+\mu^-)$	6.15%	0.81%	0.38%	0.27%	1.0%	0.84%	6.35%
$D^+ \to \pi^+\eta(\mu^+\mu^-)$	6.16%	1.09%	1.02%	0.19%	1.0%	0.75%	6.47%

systematics, we will not take into account. Table 6.14 contains a breakdown of all major systematic uncertainties contributing to the SES computations.

6.7 The final fit

Each of the four signal channels we are studying the intermediate resonance coming from both a D_s^+ and a D^+ meson. The enhancement in statistics this provides us is reflected by performing a simultaneous fit on both datasets.

In this section, we will explain the procedure for the final, post-unblinding invariant mass fit of the resonance.

The full $p.d.f.$ is a double-sided Crystal Ball, which accounts for the signal contribution), and a corresponding background $p.d.f.$: a single exponential, except for $\eta' \to \mu^+\mu^-$, which has two exponentials, one of them modeling the $\phi \to \mu^+\mu^-$ contributions and the other the combinatorial background.

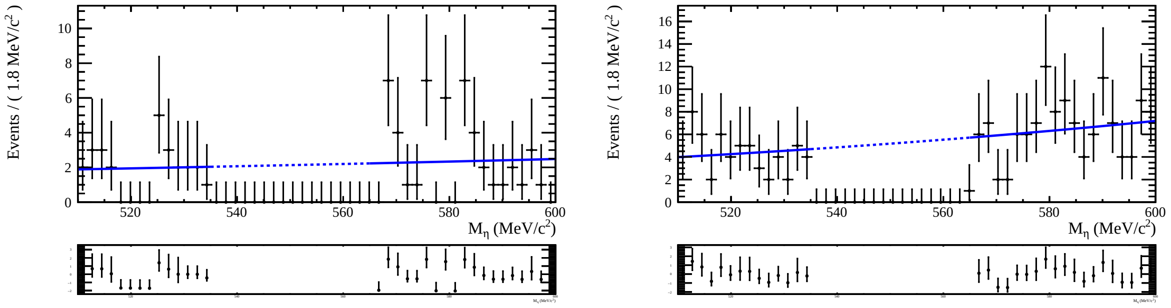
The tails from the signal $p.d.f.$ are fixed from simulation, and both the width and the mean are Gaussian-constrained to float around their values from simulation.

The width of the signal $p.d.f.$, however, is not always well modeled in simulation. To extract a more accurate number, we once again take the ϕ control sample and fit it to a Voigtian $p.d.f.$ instead. This is because the ϕ meson has a natural width, which is accounted for by the Breit-Wigner model that is then convoluted with a Gaussian $p.d.f.$.

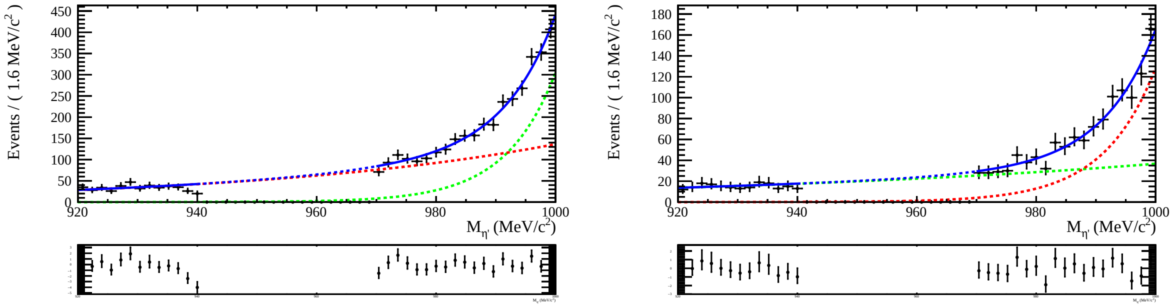
This allows us to extract the σ associated to the ϕ in data and simulation. By then getting the same parameter from the associated signal channel Monte Carlo, we can infer the expected value of the width of the signal $p.d.f.$ in data.

Finally, instead of introducing a signal yield for each of the two simultaneous fits, we use $\frac{\mathcal{B}}{\alpha_{D^+_{(s)}}}$, where $\mathcal{B}$ is a common parameter to both fits, and $\alpha_{D^+_{(s)}}$ is the single event sensitivity of the corresponding decay, as given by Table 6.19 in Sect. 6.9.1. Both variables are allowed to float, but the latter is constrained to a Gaussian $p.d.f.$ whose width equals the uncertainty of α (again, as given by Table 6.19).

As mentioned earlier in the note, in the case of the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ final fit, each of



(a) $D_s^+ \to \pi^+\eta(\mu^+\mu^-)$ (b) $D^+ \to \pi^+\eta(\mu^+\mu^-)$



(c) $D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$ (d) $D^+ \to \pi^+\eta'(\mu^+\mu^-)$

Figure 6.9: Sidebands fits to $\mu^+\mu^-$ channels.

the D_s^+ and D^+ samples will also be subject to a two-dimensional fit, to the $\mu^+\mu^-$ and $\pi^+\pi^-\mu^+\mu^-$ invariant masses, to account for possible resonances in the di-muon spectrum.

The fits to the blinded samples, containing only the modeling of the background components, can be found in Fig. 6.9 and Fig. 6.10.

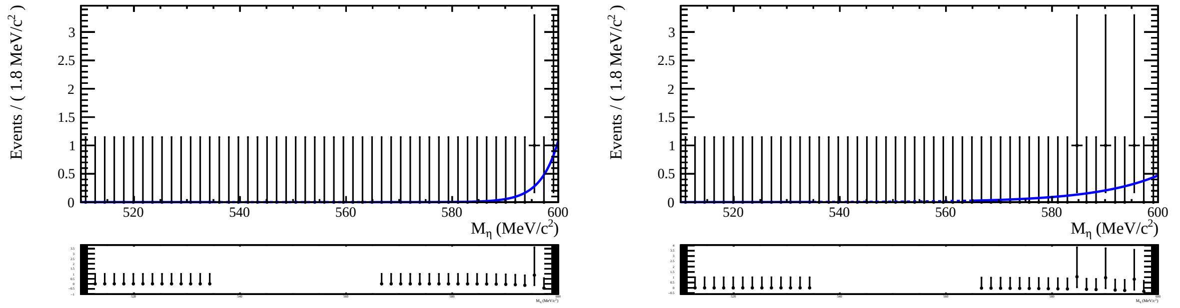
6.8 Search for BSM $\mu^+\mu^-$ resonances in $\eta' \to \pi^+\pi^-\mu^+\mu^-$

In addition to a direct measurement of $\mathcal{B}(\eta' \to \pi^+\pi^-\mu^+\mu^-)$, we will perform a search for possible bumps in the $\mu^+\mu^-$ distribution caused by an axion-like particle (ALP), $\eta' \to \pi^+\pi^-(a \to \mu^+\mu^-)$. Studies were performed for $m(a) \in [250, 300, 350, 400, 450, 500]$ MeV/c^2.

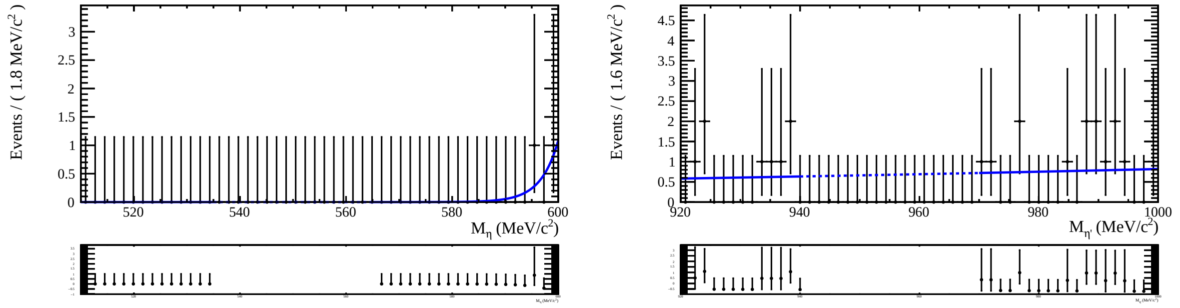
The selections applied on these simulated samples are the same as those applied to $\eta' \to \pi^+\pi^-\mu^+\mu^-$, including the BDT. The normalization samples will be the same as those used in the $\mathcal{B}$ measurement. The parameter of interest in the search is $\mathcal{B}(\eta' \to \pi^+\pi^- a) \times \mathcal{B}(a \to \mu^+\mu^-) = \alpha N(\eta' \to \pi^+\pi^-(a \to \mu^+\mu^-))$, where α is given by Eq. (6.2).

The only difference with respect to the single event sensitivity of $\eta' \to \pi^+\pi^-\mu^+\mu^-$ will be in the signal efficiency. The values for each possible mass value are provided in Table 6.15.

The search will be done through a two-dimensional fit to both the $\mu^+\mu^-$ and $\pi^+\pi^-\mu^+\mu^-$ masses (the latter is the DTF η' mass). In each mass case, the window of the $\mu^+\mu^-$ dimension will be three times that of the width of the MC fit. Since no signal is expected, all parameters of the signal $p.d.f.$ will be fixed from simulation. The components of the



(a) $D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$ (b) $D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$



(c) $D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$ (d) $D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$

Figure 6.10: Sidebands fits to $\pi^+\pi^-\mu^+\mu^-$ channels.

Table 6.15: Signal efficiencies and single event sensitivity values for each of the considered mass values of the scalar particle in the search for $\mu^+\mu^-$ BSM bumps.

Mass	ε_{sig} (D_s^+)	α (D_s^+)	ε_{sig} (D^+)	α (D^+)
250	3.987×10^{-5}	$29.80(11) \times 10^{-5}$	10.466×10^{-5}	$41.90(19) \times 10^{-5}$
300	5.358×10^{-5}	$22.14(83) \times 10^{-5}$	14.225×10^{-5}	$30.80(14) \times 10^{-5}$
350	6.394×10^{-5}	$18.56(69) \times 10^{-5}$	17.056×10^{-5}	$25.70(12) \times 10^{-5}$
400	7.680×10^{-5}	$15.45(58) \times 10^{-5}$	19.826×10^{-5}	$22.10(10) \times 10^{-5}$
450	9.042×10^{-5}	$13.12(49) \times 10^{-5}$	22.741×10^{-5}	$19.28(88) \times 10^{-5}$
500	10.116×10^{-5}	$11.73(44) \times 10^{-5}$	25.114×10^{-5}	$17.45(80) \times 10^{-5}$

two-dimensional $p.d.f.$ are given in Table 6.16.

After the final fit is performed, an upper limit will be computed for each a mass value explored, and an interpolation will be made for values in between. The procedure

Table 6.16: Different $p.d.f.s$ used in the two-dimensional fit to both the $\pi^+\pi^-\mu^+\mu^-$ and $\mu^+\mu^-$ masses, in the search for $a \to \mu^+\mu^-$ bumps in the $\mu^+\mu^-$ spectrum of $\eta' \to \pi^+\pi^-\mu^+\mu^-$.

Physical component	$\pi^+\pi^-\mu^+\mu^-$ **mass**	$\mu^+\mu^-$ **mass**
$\eta' \to \pi^+\pi^-(a \to \mu^+\mu^-)$	Double Crystal Ball	Double Crystal Ball
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	Double Crystal Ball	Exponential
$\pi^+\pi^-\mu^+\mu^-$ Combinatorial	Exponential	Exponential

to compute the limit is the same as the one used for $\eta' \to \mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ (see Sect. 6.9.5).

As mentioned, there are three components in each of the categories of the fit:

1. **Purely combinatorial background**: This is modeled with exponential $p.d.f.s$ on both the $\mu^+\mu^-$ and $\pi^+\pi^-\mu^+\mu^-$ mass dimensions. We extrapolate the expected number of background candidates from the blind fit to the $\pi^+\pi^-\mu^+\mu^-$ mass, and generate as many randomly distributed according to the $p.d.f.$ from the blind fit, on each dimension. (In reality, we generate a random number of candidates given by a Poisson distribution centered on the expected number of candidates.) The expected number of background candidates for each of the a mass values and D mesons are presented in Table 6.17.

2. **Background from** $\eta' \to \pi^+\pi^-\mu^+\mu^-$: While this was a signal channel for us in the bulk of the analysis, the decay $\eta' \to \pi^+\pi^-\mu^+\mu^-$ is an irreducible background in this search. It is modeled using the signal $p.d.f.$ from the nominal search, in the $\pi^+\pi^-\mu^+\mu^-$ mass, and an exponential in the $\mu^+\mu^-$ mass, as given by simulation. The parameters are allowed to float within a small range of their values from simulation.

   Before unblinding, there are no candidates from this background in our sample. Table 6.17 presents an estimation of how many are expect for each of the fit categories and a mass values, by taking into account the efficiency of each mass range (for $m(a) = 250$ MeV/c^2, for instance, we estimate the efficiency of the $225 < m(\mu^+\mu^-) < 275$ MeV/c^2 cut). Just like in the previous case, the numbers from the table are used as the central value of a random Poisson variable, that is then used as the number of expected $\eta' \to \pi^+\pi^-\mu^+\mu^-$ candidates in each case.

3. **Signal** $\eta' \to \pi^+\pi^- a(\to \mu^+\mu^-)$ **candidates**: Using simulation samples for each of the mass values and D mesons, we extract the signal shapes for both mass dimensions. We model the peaks as double-sided Crystal Balls. No signal candidates are generated when computing the expected limit. The parameters of the $p.d.f.s$ are fixed to the values from simulation, and only the signal yield, disguised as $\mathcal{B}/\alpha$, is allowed to float.

Using this procedure, we compute the expected limits as provided in Table 6.18.

6.9 Results

After reducing background contributions through all the aforementioned steps and computing the systematic uncertainties of our measurements, we are ready to proceed with the results of the analysis. Sect. 6.9.1 compiles all of the computations performed so far, and prints out the expected signal (according to SM predictions) and background events in the blind region of each decay.

In the case of the $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \mu^+\mu^-$ channels, which have been observed in other experiments as indicated in Sect. 6.1.2, our goal is to improve upon existing measurements. Since we expect to see a clear signal peak after unblinding, our main concern at this point is being able to properly count how many events in the blind region are signal-like and how many are background-like. See Sect. 6.9.3 for more details on that.

Table 6.17: Expected number of purely combinatorial background candidates (second and third columns) and $\eta' \to \pi^+\pi^-\mu^+\mu^-$ candidates (last two columns) for each of the studied mass values for a hypothetical scalar particle a decaying into two muons.

Scalar mass value $[\,\mathbf{MeV}/c^2\,]$	**Expected** N_{bkg}		**Expected** $N_{\eta' \to \pi^+\pi^-\mu^+\mu^-}$	
	D_s^+	D^+	D_s^+	D^+
250	2.46	6.52	43.3	26.4
300	7.46	3.69	37.0	20.6
350	4.11	7.55	22.6	13.1
400	6.16	8.67	13.9	7.1
450	5.52	5.17	6.7	3.4
500	6.27	2.50	2.1	1.5

Table 6.18: Expected limits for $\mathcal{B}(\eta' \to \pi^+\pi^- a(\to \mu^+\mu^-))$, for various possible values of the mass of the scalar particle.

Scalar mass value $[\,\mathbf{MeV}/c^2\,]$	250	300	350	400	450	500
$\mathcal{B}(\eta' \to \pi^+\pi^- a(\to \mu^+\mu^-))$ $[\times 10^{-5}]$	< 126.4	< 106.0	< 77.5	< 61.6	< 39.0	< 28.0

In the case of $\eta' \to \mu^+\mu^-$, based on how many events we are expecting in the signal region, we want to study whether or not we are sensitive to observing this decay for the first time. See Sect. 6.9.4 for more. In addition, we also want to check the quality of the fit to convince ourselves that, should we be sensitive, we would be able to properly characterize it with the models we are using.

Finally, we do not expect to be sensitive to observing $\eta \to \pi^+\pi^-\mu^+\mu^-$. However, we can introduce an upper threshold for the branching fraction of this decay, as detailed in Sect. 6.9.5.

6.9.1 Single Event Sensitivity results

The final number for the single event sensitivity for each of the channels is presented in Table 6.19. With this in mind, we can also compute the expected number of signal events for each channel according to the SM predictions. Moreover, by integrating the area under an exponential fit to the sidebands of each channel[*], we can also predict how many background events we expect to see after unblinding. Results of the expected signal and background events in the blind region are shown in Table 6.20.

6.9.2 Running toy experiments

Pre-unblinding checks involve a series of steps to ensure that our existing machinery is able to properly characterize the results we expect. In all four channels, we expect a background component, which is dominant in the cases of $\eta' \to \mu^+\mu^-$ and $\eta \to \pi^+\pi^-\mu^+\mu^-$ and which

---

[*] The sidebands are modeled through a single exponential in all channels except for $\eta' \to \mu^+\mu^-$, which includes a second exponential to account for $\phi \to \mu^+\mu^-$ contributions to the right sideband.

Table 6.19: Breakdown of the SES computations for all channels. N_{norm} is computed as explained in Sect. 6.5, using a fit for the $\pi^+\pi^-\mu^+\mu^-$ channels, and simply counting candidates in the $\mu^+\mu^-$ channels.

Channel	**Eff. Ratio**	**BR Ratio**	N_{norm}	α
$D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	36.21(66)	$34.7(22) \times 10^{-5}$	$146.14(13) \times 10^3$	$1.134(78) \times 10^{-7}$
$D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	28.20(32)	$34.9(24) \times 10^{-5}$	$84.97(38) \times 10^3$	$2.020(14) \times 10^{-7}$
$D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	97.2(32)	$81.1(50) \times 10^{-5}$	$146.14(13) \times 10^3$	$1.074(24) \times 10^{-6}$
$D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	80.8(14)	$46.0(28) \times 10^{-5}$	$84.97(38) \times 10^3$	$1.320(22) \times 10^{-6}$
$D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$	1.0940(52)	$34.7(22) \times 10^{-5}$	$188.96(19) \times 10^3$	$1.96(12) \times 10^{-9}$
$D^+ \to \pi^+\eta'(\mu^+\mu^-)$	1.1067(41)	$34.9(24) \times 10^{-5}$	$76.82(77) \times 10^3$	$5.01(35) \times 10^{-9}$
$D_s^+ \to \pi^+\eta(\mu^+\mu^-)$	1.381(12)	$81.1(50) \times 10^{-5}$	$239.10(24) \times 10^3$	$4.22(27) \times 10^{-9}$
$D^+ \to \pi^+\eta(\mu^+\mu^-)$	1.422(11)	$46.0(28) \times 10^{-5}$	$113.37(11) \times 10^3$	$5.49(36) \times 10^{-9}$

Table 6.20: Breakdown of the number of signal and background expected events for each of the channels in the blinded region.

Channel	N_{sig} **Expected**	N_{bkg} **Expected**	N_{sig} / N_{bkg}
$D_s^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	138.5(95)	37.5(49)	3.70(54)
$D^+ \to \pi^+\eta'(\pi^+\pi^-\mu^+\mu^-)$	77.9(59)	37.7(51)	2.07(32)
$D_s^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	0.00699(51)	0.000001(12)	2432(11)
$D^+ \to \pi^+\eta(\pi^+\pi^-\mu^+\mu^-)$	0.00568(42)	0.78(83)	0.0073(78)
$D_s^+ \to \pi^+\eta'(\mu^+\mu^-)$	69.4(45)	1791(66)	0.0388(28)
$D^+ \to \pi^+\eta'(\mu^+\mu^-)$	27.1(19)	803(55)	0.0338(33)
$D_s^+ \to \pi^+\eta(\mu^+\mu^-)$	1118(71)	788(21)	1.418(97)
$D^+ \to \pi^+\eta(\mu^+\mu^-)$	860(56)	559(17)	1.54(11)

we model through an exponential (or a double exponential, in the case of $\eta' \to \mu^+\mu^-$), and a signal component, modeled by a double-sided Crystal Ball.

These models have been set by fitting either the Monte Carlo distributions, in the case of the signal shapes, or the sidebands, in the case of the background PDFs. In the latter case, we are able to exclude the blind region.

With this in mind, RooFit allows us to use these two models to generate however many events we wish in the given intervals, which means we are able to simulate what the unblinding may look like, should our expected number of signal- and background-like events be accurate.

We use this technique to produce a large number of toy experiments, wherein we generate N_{sig} signal events according to the MC-fixed double-sided Crystal Ball, and N_{bkg} events according to the sidebands-fixed exponential model. Here, both N_{sig} and N_{bkg} are random numbers that follow a Poisson distribution where the mean is given by the numbers in Table 6.20. This production of toy experiments is done for both D_s^+ and D^+ channels. The simultaneous fit, as explained in Sect. 6.7, is applied to the toy samples, which would include a toy version of the expected signal shape.

A total of 1500 toy experiments were performed for each of the channels.

6.9.3 Signal fit quality

To check the quality of the final fit and assess the need for extra corrections on the final value of the branching fraction as computed by the simultaneous fit to D_s^+ and D^+, we compute the difference between the SM prediction of the $\mathcal{B}$ and the $\mathcal{B}$ as measured by the simultaneous fit, $\mathcal{B}_{fit}$.

With this, we can either compute the residuals plot, which evaluates this difference plainly as $\mathcal{B}_{SM} - \mathcal{B}_{fit}$, or the pulls plot, which divides the difference by the error of $\mathcal{B}_{fit}$, $(\mathcal{B}_{SM} - \mathcal{B}_{fit})/\sigma(\mathcal{B}_{fit})$

In an ideal case, both distributions would follow a Gaussian distribution with mean $\mu = 0$. Additionally, the width of the pulls plot should be close to $\sigma = 1$.

The residuals plots, and thus the results from their fit to a Gaussian (see Table 6.21), can help us introduce a correction to the $\mathcal{B}$ eventually given by the unblinded fit. Table 6.22 shows the results from the pulls plots. Fig. 6.11 contain these plots for all channels in which this test is performed.

Table 6.21: Parameters of the Gaussian fit to the residuals plot, i.e., the difference $\delta = \mathcal{B}_{SM} - \mathcal{B}_{fit}$ for each channel. The scale is removed, so that for $\eta' \to \mu^+\mu^-$, the numbers should be multiplied by 10^{-7}, and similarly for other decays.

Channel	Mean	Width
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	$-9.7(39) \times 10^{-3}$	0.1505(28)
$\eta' \to \mu^+\mu^-$	$-0.119(13)$	0.5094(93)
$\eta \to \mu^+\mu^-$	$-2.3(6.8) \times 10^{-4}$	0.2594(52)

Table 6.22: Parameters of the Gaussian fit to the pulls plot, i.e., $p = (\mathcal{B}_{SM} - \mathcal{B}_{fit})/s(\mathcal{B}_{fit})$ for each channel.

Channel	Mean	Width
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	0.021(26)	0.989(19)
$\eta' \to \mu^+\mu^-$	$-0.191(25)$	0.955(19)
$\eta \to \mu^+\mu^-$	0.012(27)	1.015(20)

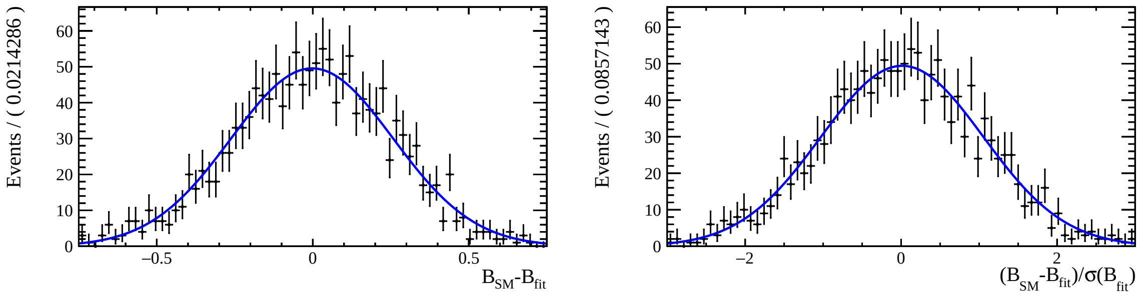
6.9.4 Sensitivity to $\eta' \to \mu^+\mu^-$

To evaluate whether or not we are sensitive to observing $\eta' \to \mu^+\mu^-$ after unblinding, we compute the likelihood profile of the $\mathcal{B}$ variable in the fit.

On each of 1000 pseudoexperiments, we input the number of expected signal events according to the SM prediction for the $\mathcal{B}$, and later evaluate the significance of the signal peak using Wilks' theorem. The median of the distribution is at 2.85σ.

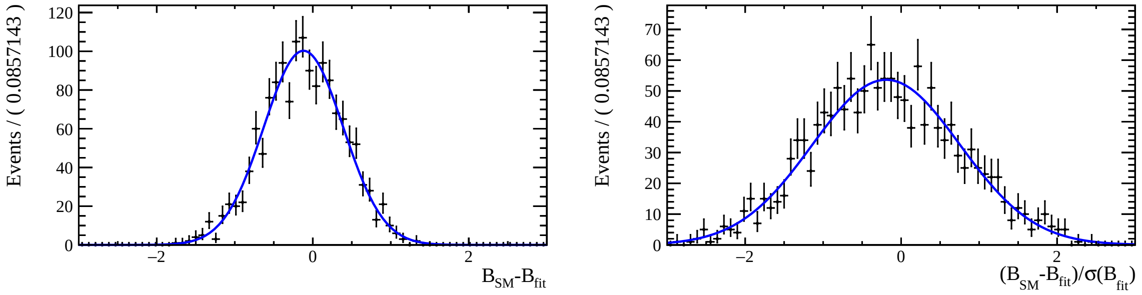
6.9.5 Expected limits

In the case of $\eta \to \pi^+\pi^-\mu^+\mu^-$, where we do not expect to be sensitive to finding signal, and in $\eta' \to \mu^+\mu^-$, where we are not certain yet, we integrate 90% of the positive side of



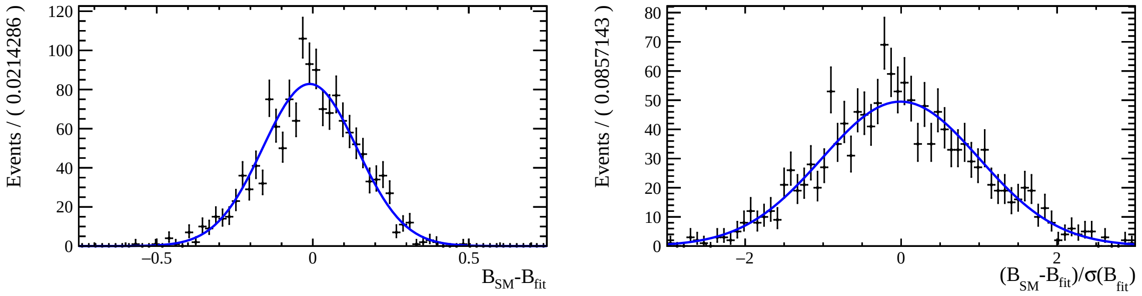
(a) $\eta \to \mu^+\mu^-$ Residuals

(b) $\eta \to \mu^+\mu^-$ Pulls



(c) $\eta' \to \mu^+\mu^-$ Residuals

(d) $\eta' \to \mu^+\mu^-$ Pulls



(e) $\eta' \to \pi^+\pi^-\mu^+\mu^-$ Residuals

(f) $\eta' \to \pi^+\pi^-\mu^+\mu^-$ Pulls

Figure 6.11: Results of fit-quality tests for all three channels in which we expect to be signal-sensitive. Left plots are residuals ($\mathcal{B}_{SM} - \mathcal{B}_{fit}$), where the scale of the $\mathcal{B}$ for each decay is removed (i.e., $\eta' \to \mu^+\mu^-$ is at 10^{-7}, $\eta' \to \pi^+\pi^-\mu^+\mu^-$ is at 10^{-5}, and $\eta \to \mu^+\mu^-$ is at 10^{-6}), while right plots are pulls, $(\mathcal{B}_{SM} - \mathcal{B}_{fit})/\sigma(\mathcal{B}_{fit})$

the likelihood profile of the $\mathcal{B}$ variable, which gives us the expected branching fraction for each of the pseudoexperiments. Computing the median of all pseudoexperiments, we establish an upper limit for both of their respective branching fractions:

$$\mathcal{B}(\eta \to \pi^+\pi^-\mu^+\mu^-) < 1.4 \times 10^{-6} \tag{6.10}$$

$$\mathcal{B}(\eta' \to \mu^+\mu^-) < 2.25 \times 10^{-7} \tag{6.11}$$

Fig. 6.13 shows the distribution of the $\mathcal{B}$ results for all toys.

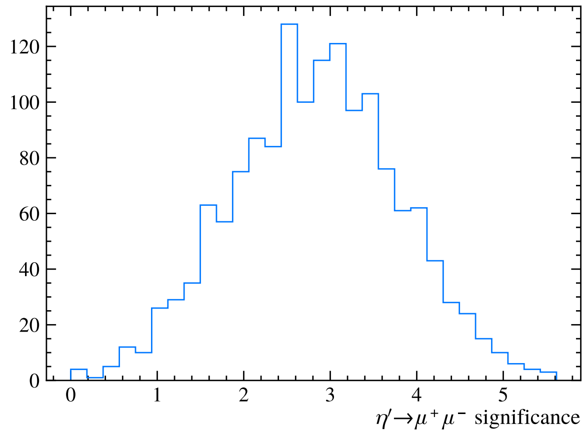
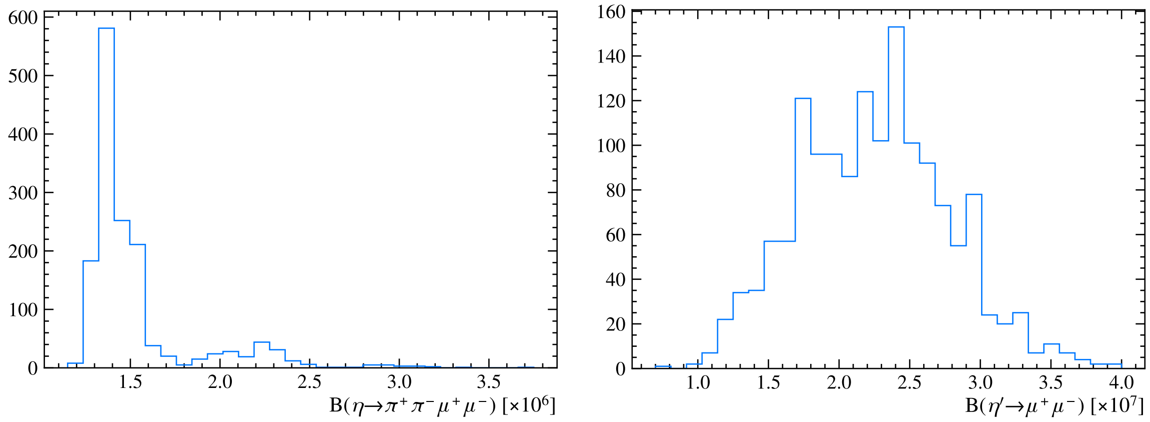


Figure 6.12: Results significance of $\eta' \to \mu^+\mu^-$ in 1500 pseudoexperiments.



(a) $\eta \to \pi^+\pi^-\mu^+\mu^-$

(b) $\eta' \to \mu^+\mu^-$

Figure 6.13: Results of the $\mathcal{B}$ expected upper limit from 1500 toys for both $\eta \to \pi^+\pi^-\mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$.

6.9.6 Final results

The final results of the analysis, pre-unblinding and including all systematic uncertainties computed so far, are included in Table 6.23. As seen, we expect to improve upon $\eta' \to \pi^+\pi^-\mu^+\mu^-$'s 19% uncertainty (now computed to be around 7.7%) and $\eta \to \mu^+\mu^-$'s 15% (5.5%). We will also establish the first expected limit for $\eta' \to \mu^+\mu^-$.The limit obtained for $\eta \to \pi^+\pi^-\mu^+\mu^-$ does not, however, represent an improvement upon the BESIII result [231].

Table 6.23: Expected results for the measured $\mathcal{B}$ of all four decays. The results remain unblinded, so all that can be provided are expectations according to toy studies. In the cases of $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \mu^+\mu^-$, xx is a placeholder for the final value of the $\mathcal{B}$.

Channel	$\mathcal{B}$	Total unc.	PDG results	PDG unc.	Theory
$\eta \to \pi^+\pi^-\mu^+\mu^-$	$< 1.4 \times 10^{-6}$	—	$< 4.0 \times 10^{-7}$	—	$(7.5^{+1.8}_{-0.7}) \times 10^{-9}$
$\eta' \to \pi^+\pi^-\mu^+\mu^-$	$(xx \pm 8.0\% \pm 5.2\%) \times 10^{-5}$	9.6%	$(1.97 \pm 0.33 \pm 0.18) \times 10^{-5}$	19.08%	$(1.57^{+0.40}_{-0.47}) \times 10^{-5}$
$\eta' \to \mu^+\mu^-$	$< 2.25 \times 10^{-7}$	—	—	—	$(1.36^{+0.29}_{-0.26}) \times 10^{-7}$
$\eta \to \mu^+\mu^-$	$(xx \pm 2.7\% \pm 4.5\%) \times 10^{-6}$	5.3%	$(5.7 \pm 0.7 \pm 0.5) \times 10^{-6}$	15.09%	$(4.72^{+0.05}_{-0.21}) \times 10^{-6}$

7

Results and conclusions

In this thesis, the first-ever experimental result of the decay of a neutral kaon into four muons was presented. Though no signal was observed, the analysis was able to set a stringent upper limit of $\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}$, the first time this scale has been reached by an experiment at the Large Hadron Collider. The complementary upper limit of $\mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}$ also represents the first K^0_L result at the LHC.

Furthermore, the thesis also presented the first set of measurements of η and η' decays by LHCb. Taking the resonances as coming from charm mesons, the first-ever experimental study of $\eta' \to \mu^+\mu^-$ was provided, as well as updated studies on the branching fractions of both $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \mu^+\mu^-$.

As of the submission of this thesis, the final results are yet to be unblinded. The expectation is to measure $\mathcal{B}(\eta' \to \mu^+\mu^-)$ with a significance of 2.8σ, and to improve the experimental precision of the branching ratios of $\eta' \to \pi^+\pi^-\mu^+\mu^-$ and $\eta \to \mu^+\mu^-$. Respectively, the total uncertainties are expected to be, after unblinding, of 9.6% and 5.3%. Additionally, a search of $\eta' \to \pi^+\pi^-(a \to \mu^+\mu^-)$, with a representing an axion-like particle, is introduced, and limits on the branching fraction of the decay for different mass values of a are presented.

The decay mode $\eta \to \pi^+\pi^-\mu^+\mu^-$ was also studied, with the final result expected to be a validation of previous studies by BESIII.

Lord Kelvin might have never actually stated that the only way forward for physics would be through more precise measurements, but the sentiment that everything that was to be found, already has been, has prevailed through the scientific community over a century later.

The underlying message of this thesis is that there is no shortage of areas to be explored in high energy physics in the early 21st century. Whether it is to provide more precise measurements that could shed light on some unresolved puzzles, or continue to search for new processes that have not been observed yet, the global picture of the Standard Model of Particle Physics has not been completed yet.

The Large Hadron Collider provides a unique way for physicists to take on this challenge, and thanks to the $K_S^0 \to \mu^+\mu^-\mu^+\mu^-$ study presented here, we know the current ceiling in scale that the detectors were operated under until 2018. With new data being collected and the detectors continue to get better, this ceiling will move even further up, possibly increasing the chances of seeing processes that cannot be explained under the current theory.

8

Resumo en galego

8.1 O Modelo Estándar da Física de Partículas

Dende a súa introdución a comezos do século XX, a Física Cuántica acadou os resultados máis precisos na historia da ciencia, sendo capaz de predicir correctamente resultados experimentais con ata doce cifras significativas [50]. As Teorías Cuánticas de Campos (QFTs) que se derivaron a partir de modelos matemáticos presentados por Dirac [13, 14] e Schrödinger [10] converxeron arredor dos anos 60 e 70 na formalización do Modelo Estándar da Física de Partículas (SM) [28], unha teoría capaz de estudar con excelente precisión as partículas subatómicas e as súas interaccións.

Pese aos seus éxitos, os cales inclúen, especialmente, o descubrimento do bosón de Higgs en 2012 [48, 49], considerado a peza que faltaba do puzle do SM, aínda quedan algunhas frontes abertas neste campo. Os neutrinos, por exemplo, non deberían ter masa, segundo a teoría, o cal choca con resultados experimentais na materia. Observacións astrofísicas tamén concluíron a finais do século pasado que a meirande parte da enerxía e materia presente no noso universo é de natureza actualmente descoñecida (materia escura) [68, 69]. O momento xiromagnético do muón, ademais, arrastra un desacordo histórico entre as predicións teóricas feitas con QFTs e os resultados experimentais [52, 53].

Estas ideas, xunto con outras pequenas incoherencias atopadas experimentalmente nos últimos anos, motivan a busca de procesos cuánticos que puideran rexirse por novos modelos físicos máis alá do coñecemento actual. Os denominados fotóns escuros, equivalentes aos mediadores da forza electromagnética en modelos teóricos de materia escura, poderían actuar como intermediarios en procesos moi raros no SM [69]. Outras tensións teóricas na teoría que rixe a física de quarks e gluóns, QCD, poderían esconder a existencia de novas partículas chamadas axións [88, 89], sobre as cales aínda non hai evidencias experimentais.

8.2 Condicións experimentais

A busca do bosón de Higgs motivou a finais do século XX a construción do Gran Colisionador de Hadrones (LHC) [96], un acelerador toroidal de 27 km que fai colisionar feixes de protóns en catro puntos do seu perímetro, correspondentes con catro grandes experimentos deseñados para explorar distintas áreas de coñecemento na física de partículas. ATLAS [97] e CMS [98], por exemplo, descubriron o bosón de Higgs en 2012 de xeito independente. ALICE [99] estuda colisións de protóns contra núcleos para tentar reproducir as condicións do plasma de quarks e gluóns que existiu nos primeiros instantes do universo.

LHCb [100], o experimento cuxos datos analizáronse nesta tese, construíuse coa intención de estudar a violación CP e desintegracións raras de mesóns que conteñen quarks b e c. Dende entón, o seu programa de física estendeuse moito máis aló ata considerarse agora un detector (e experimento) de carácter xeral, e cubre tamén desintegracións de mesóns con quarks s, colisións con ións pesados, e a busca de procesos exóticos, como poden ser novos estados cuánticos coma o tetraquark.

O detector LHCb está composto por distintos subcompoñentes encargados de realizar distintas tarefas relacionadas co sistema de rastreo das partículas que o atravesan, e a identificación das mesmas. O VErtex LOcator (VELO) [120] é o aparello que rodea o punto de colisión dos feixes de protóns (vértice primario, PV), e realiza unha primeira labor de seguimento das traxectorias das partículas aí producidas.

Estas partículas logo atravesan a estación de rastreo coñecida como Tracking Turicensis (TT) [123], que garda máis impactos das partículas no seu percorrido. A continuación, as traxectorias das partículas cargadas son curvadas grazas á presencia dun imán [128]; isto tamén axuda ao cálculo do momento e mellora a calidade da reconstrucción das traxectorias, analizando subseguintes impactos nas estacións T e nos calorímetros [135, 139], onde hadróns, electróns, e fotóns depositan a súa enerxía e se frean.

Os calorímetros, xunto cos subdetectores RICH [129], baseados na radiación Čerenkov, realizan tarefas de identificación das partículas. As estacións de muóns [141], situadas ao final do detector, só recollen as partículas máis penetrantes; en xeral, os muóns. Estes subdetectores actúan como identificadores e tamén sistemas de rastreo.

A identificación de partículas no LHCb faise de distintos xeitos. Pódense calcular variables baseadas en comparar a probabilidade de que unha partícula sexa de especie X coa probabilidade de que sexa un pión (as partículas máis abundantes no detector). Son as chamadas `DLL` variables, e calcúlanse coas diferenzas dos logaritmos destas dúas probabilidades, que son devoltas polo RICH e os calorímetros.

Os muóns tamén teñen a súa variable booleana propia `IsMuon` [163], que devolve unha resposta positiva en función dos impactos que teñan nas distintas estacións os candidatos a muóns con distintos rangos de momento.

O LHCb tamén adestra algoritmos de aprendizaxe automática, como redes neuronais, usando información dos detectores para calcular unha resposta máis complexa. Son as variables `ProbNN` [102]. A resposta dos distintos algoritmos de identificación de partículas non sempre está ben representada nos nosos eventos simulados, polo que debe calibrarse usando modos específicos de calibración que deberían devolver unha resposta parecida ao noso sinal. No caso dos muóns, por exemplo, úsase $J/\psi \to \mu^+\mu^-$.

Durante os períodos de toma de datos, o LHCb non pode almacenar toda a información producida nas colisións protón-protón. Para iso, desenvolveu un sistema de disparo que ten lugar en varias etapas, que toma decisións sobre que produtos das colisións gardar e cales non [144]. O primeiro paso ten lugar ao máis baixo nivel (Level–0, L0), e tómase segundo os datos se producen. En función de se en cada colisión hai algún produto que supere determinados umbrais de momento ou enerxía transversa, o evento entero gardarase ou non.

Esta resposta é moi rápida; ten lugar cada 25 ns, e non pode utilizar ningún tipo de reconstrucción. Os eventos que pasan esta primeira etapa son logo analizados computacionalmente. Nese momento, reconstruccións parciais (primeiro) e totais (segundo) teñen lugar, permitindo ao sistema de disparo de alto nivel (HLT) tomar decisións máis centradas na física das desintegracións que puideran ter lugar. Respectivamente, son os niveis HLT1 e HLT2.

Dado que as eficiencias asociadas ás seleccións de disparo non están sempre ben reproducidas na simulación, o LHCb desenvolveu un método, coñecido como TISTOS [148], que permite a súa evaluación usando datos reais dunha canle de control, que debería ser topoloxicamente similar ao noso sinal.

8.3 Métodos estatísticos na física de partículas

Nesta tese úsanse diversas ferramentas estatísiticas, incluíndo funcións de distribución de probabilidade, así como intervalos e niveis de confianza. Usaranse tamén diversos modelos de aprendizaxe automático, dende axustes de distribucións de datos a distintas funcións, ata algoritmos de clasificación baseados en árbores de decisión para distinguir entre candidatos de tipo sinal e de tipo fondo.

8.4 Desintegracións de kaóns neutros no LHCb

Os kaóns neutros poden ter dous autoestados físicos, segundo a vida media: K^0_S (kaón curto) e K^0_L (kaón longo). O LHCb produce unha gran cantidade de K^0_S , o que proporciona ao experimento cunhas capacidades únicas para medir os modos de desintegracións máis raros posibles segundo o Modelo Estándar. Entre eles, está a desintegración do kaón neutro en catro muóns, $K^0 \to \mu^+\mu^-\mu^+\mu^-$, que ten unha probabilidade de acontecer, segundo este modelo teórico, de [183]

$$\mathcal{B}(K^0_\mathrm{S} \to \mu^+\mu^-\mu^+\mu^-) = (1-4)\times 10^{-14} \quad , \quad \mathcal{B}(K^0_\mathrm{L} \to \mu^+\mu^-\mu^+\mu^-) = (5-9)\times 10^{-13}. \tag{8.1}$$

Esta tese contén o primeiro estudo experimental deste modo de desintegración, usando datos recollidos polo LHCb entre 2016 e 2018, que representan arredor de 5.1 fb^{-1}. Teorías alén do Modelo Estándar, coma un modelo de materia escura ao que se lle engade o equivalente ao bosón de Higgs (véxase Ref. [84]) predin unha probabilidade de desintegración dúas ordes de magnitude máis alta, o cal representaría unha evidencia clara de nova física.

A selección dos candidatos a sinal consiste en seleccionar catro trazas ben reconstruídas como muóns que formen un bo vértice secundario, dentro do VELO, e cunha masa invariante arredor de $m(K^0_\mathrm{S}) = 497.6\ \mathrm{MeV}/c^2$. A partícula nai, asociada a este vértice, terá unha vida media superior a ~ 4.6 ps, un desprazamento no plano perpendicular ao eixo z (definido como aquel polo que circulan os protóns que logo colisionan), e un parámetro de impacto con respecto ao PV inferior a 1 mm, o que maximiza a probabilidade de que o K^0_{S} se orixinase na colisión protón-protón.

A medida de $K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-$ realízase con respecto ao número de $K^0_{\mathrm{S}} \to \pi^+\pi^-$ producidos no LHCb durante o mesmo tempo de toma de datos. Os pións desta canle deben ter un IP_{χ^2} moi alto, de xeito que saibamos que non se produciron na colisión, e unha masa invariante $400 < m(K^0_{\mathrm{S}}) < 600$ MeV/c^2. A partícula da que veñen os pións tamén terá unha vida media cun valor preto da medida para o K^0_{S}, e a distancia máis pequena entre as dúas trazas de pións debe ser inferior a 0.1 mm.

Durante o tratamento dos datos tamén se aplican cortes na calidade das trazas para minimizar o caso no que sexan pantasmas (conxuntos de impactos que non se corresponden cunha única traza) ou partículas mal reconstruídas. Por último, durante a meirande parte da análise, os candidatos cunha masa invariante na rexión de sinal, $m(K^0_{\mathrm{S}}) \in [490, 510]$ MeV/c^2, non se consideraron para minimizar posibles sesgos por parte do analista.

As seleccións de disparo L0 no canal de sinal divídense en dúas categorías. Nunha delas, referida como TIS, polo menos unha partícula fóra do noso candidato de sinal debe estar rexistrada como que lanzou un sinal de disparo. Na outra, referida como TOS exclusivo (xTOS), unicamente trazas no noso sinal serán responsables do disparo L0. A nivel HLT1, pídese pasar unha de dúas seleccións que teñen como obxectivo capturar pares de muóns cunha masa invariante ou momento transverso baixo. Por último, en HLT2, deberán pasar unha liña deseñada especificamente para pares de muóns que veñan de kaóns.

Cada certo número de colisións protón-protón no LHCb, gárdanse eventos aos que non se lles require pasar ningunha selección de disparo. Coñécese como datos con mínimo sesgo (en inglés, *minimum bias*) faise para poder ter comparativas entre datos que pasan e non pasan o sistema de disparo, facer medicións sobre a luminosidade, ou comprobacións sobre o funcionamento do detector. Tamén é unha gran fonte de datos en canais de desintegración moi comúns, como $K^0_{\mathrm{S}} \to \pi^+\pi^-$ nesta análise.

O único fondo de orixe física relevante nesta análise é $\Lambda \to p\pi^-$, que contamina a construción de candidatos $K^0_{\mathrm{S}} \to \pi^+\pi^-$. Para minimizalo, Armenteros e Podolanski desenvolveron [184] un razoamento matemático que permite illar os dous modos de desintegración cunha sinxela selección sobre os momentos das partículas fillas e o momento transverso da nai.

Debuxando a asimetría nos momentos lineais das fillas fronte ao momento transverso da partícula nai, ambas desintegracións (así como a conxugada *CP* $\overline{\mathit{\Lambda}} \to \overline{p}\pi^+$) forman semi-elipses (véxase a Figura 5.2), de xeito que a selección Eq. (5.20) é eficiente ao 99% en $K^0_{\mathrm{S}} \to \pi^+\pi^-$ e elimina todo rastro de desintegracións de λ e $\overline{\mathit{\Lambda}}$.

Ao mesmo tempo, o único fondo relevante na desintegración $K^0_{\mathrm{S}} \to \mu^+\mu^-\mu^+\mu^-$ é de tipo combinatorio, é dicir, trazas de muóns que pasaron as seleccións anteriores pero que non necesariamente veñen dun K^0_{S}. Para minimizar estas contribucións, adestráronse varios algoritmos de clasificación que usaron a información do parámetro do impacto do K^0_{S}, a mínima distancia entre as catro trazas, o parámetro de impacto dos muóns con respecto ao vértice primario, a distancia de voo do K^0_{S} no plano transverso ao eixo z, o mínimo ángulo entre os distintos pares de muóns, e a distancia do punto de desintegración do K^0_{S} ao material do VELO, para distinguir entre candidatos de fondo e candidatos de sinal.

Cunha eficiencia no sinal do 80% conseguiuse eliminar todo o fondo combinatorio da rexión de sinal.

As mostras de simulación, tanto de $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ como de $K^0_S \to \pi^+\pi^-$, corrixíronse usando a información obtida nos datos de $K^0_S \to \pi^+\pi^-$. En particular, o momento transverso do K^0_S non está ben descrito na simulación, e as variables relativas á calidade das trazas tamén requiren un tratamento individualizado que consiste en usar as funcións de distribución cumulativas en datos das correspondentes variables. O $p_T(K^0_S)$ corríxese usando o método Aceptar-Rexeitar desenvolvido por John von Neumann [186, 187].

Usáronse tamén as ferramentas de LHCb `TrackCalib` e `PIDCalib` para corrixir as eficiencias das seleccións sobre as trazas, así como o feito de normalizar un modo de catro trazas usando outro de dúas, e sobre as variables de identificación de partículas, respectivamente. Respectivamente, estes dous métodos arrastran uns erros sistemáticos do 1% e 3.3%, respectivamente.

A maior fonte de erros sistemáticos orixínase nas seleccións de disparo, especialmente no L0. Estas seleccións ratifícanse usando como canles de control $K^0_S \to \pi^+\pi^-$ e $K^0 \to \pi^+\mu^-\bar{\nu}_\mu$, onde ao pión se lle piden seleccións de identificación de muóns de xeito que, a nivel de disparo, é unha canle con dous muóns. A validación das eficiencias deste xeito logo terá que ser extrapolada ao caso no que haxa catro muóns no estado final, o que introduce erros sistemáticos do 11% (para TIS) e 21% (para xTOS).

O axuste final aos datos é de tipo simultáneo ás dúas categorías TIS e xTOS, de xeito que se minimizan ao mesmo tempo as funcións de perdas de cada conxunto de datos. O fondo combinatorio (unicamente tres candidatos sobreviven na categoría TIS, e ningún na xTOS) está modelado por unha función de distribución exponencial, mentres que o posible sinal ven descrito por unha función Hypatia cuxos parámetros quedan fixos na simulación.

Constantes proporcionais relativas ao número de candidatos de sinal e de fondo moderan a combinación das dúas compoñentes. En particular, o número de candidatos de sinal pódese reescribir coma $n_{\rm sig} = \mathcal{B}/\alpha$, onde α é a chamada sensibilidade dun único evento, calculada segundo Eq. (5.27) para cada categoría. Os erros sistemáticos introdúcense como ligaduras de tipo Gaussiano sobre α. O parámetro $\mathcal{B}$ é común ás dúas categorías, e o seu resultado tras a minimización será a nosa medida.

O límite superior sobre $\mathcal{B}(K^0_S \to \mu^+\mu^-\mu^+\mu^-)$ mídese integrando o 90% da área baixo a curva que deixa o perfil da probabilidade do parámetro $\mathcal{B}$ tras a minimización. Obtívose, cun 90% $C.L.$,

$$\mathcal{B}(K^0_S \to \mu^+\mu^-\mu^+\mu^-) < 5.1 \times 10^{-12}. \tag{8.2}$$

A maiores, aplicando un factor de conversión de $\simeq 2 \times 10^{-3}$ sobre as α, relativo á aceptancia do K^0_L fronte ao K^0_S no LHCb, pódese calcular o límite

$$\mathcal{B}(K^0_L \to \mu^+\mu^-\mu^+\mu^-) < 2.3 \times 10^{-9}. \tag{8.3}$$

Ambos resultados son as primeiras cotas superiores sobre os modos de desintegración correspondentes. A maiores, o límite sobre $K^0_S \to \mu^+\mu^-\mu^+\mu^-$ resulta ser a primeira vez que se acada unha sensibilidade de 10^{-12} no LHC, mentres que $\mathcal{B}(K^0_L \to \mu^+\mu^-\mu^+\mu^-)$ é o primeiro resultado dunha desintegración con K^0_L no LHCb. Ningún dos dous resultados permite rexeitar as predicciónss teóricas nin do Modelo Estándar nin da teoría de materia escura mencionada.

A Figura 5.14 presenta estimacións sobre futuros estudos no LHCb de ambos modos. Calcúlase que cara ao final da súa vida, ao acadar luminosidades de $150-300$ fb^{-1}, poderase medir con precisión o Modelo Estándar.

8.5 Desintegracións anómalas de mesóns η e η'

A outra análise presentada nesta tese inclúe catro modos de desintegración dos mesóns η e η', sendo os estados finais $\mu^+\mu^-$ e $\pi^+\pi^-\mu^+\mu^-$. Usáronse novamente datos do LHCb correspondentes a 5.1 fb^{-1} tomados entre 2016 e 2018. A particularidade adicional é que estes mesóns, de moi curta vida media, procederán de mesóns D^+ ou D_s^+. Isto é para minimizar os posibles fondos de carácter físico que aparecen ao tomar candidatos que veñan directamente do vértice primario, ou de mesóns máis pesados.

Dos catro modos, $\eta' \to \mu^+\mu^-$ non ten rexistrados resultados experimentais no momento da escritura desta tese de doutoramento. Os modos $\eta' \to \pi^+\pi^-\mu^+\mu^-$ [228] e $\eta \to \mu^+\mu^-$ [227] están medidos con incertezas do 19% e 15% respectivamente. Por último, $\eta \to \pi^+\pi^-\mu^+\mu^-$ estudouse recentemente en Ref. [231].A Táboa 6.1 presenta unha comparativa das predicións teóricas e experimentais de cada un destes modos.

Ademais de ser modos de desintegración raros, cunhas fraccións de desintegración entre $10^{-5} - 10^{-7}$, o que significa que poderían estar facilmente suxeitos a desviacións con respecto ao Modelo Estándar, as medidas de precisión destes modos pode aportar información ao puzle do $(g-2)_\mu$, debido a que os seus mecanismos de desintegración están intrinsicamente conectados aos factores de forma de transición que gobernan as correccións cuánticas do Modelo Estándar nos cálculos do momento xiromagnético do muón.

Os candidatos de cada canle de sinal constrúense combinando trazas de muóns e pións na rexión cinemática do mesón $\eta^{(\prime)}$ correspondente. Este mesón, xunto con outro pión, veñen á súa vez dun mesón D^+ ou D_s^+. As canles de normalización en todos os casos son $D_{(s)}^+ \to \pi^+\phi(\mu^+\mu^-)$.

Requírese que polo menos un dos muóns pasase unha selección do disparo, ou ben o pión que vén do mesón D a nivel L0. Unha única selección é requerida a nivel HLT1, que busca combinacións de dúas trazas ben reconstruídas e que veñan do mesmo punto. A selección HLT2 é máis complexa. No caso dos modos $\mu^+\mu^-$, existe unha liña que selecciona exclusivamente desintegracións do tipo $D \to \pi^+\mu^+\mu^-$, e que é case 100% eficiente nestes canais. Unha combinación de varias liñas topolóxicas, e unha liña que busca pares de muóns con baixo p_T, forman a selección HLT2 nos modos $\pi^+\pi^-\mu^+\mu^-$.

A análise novamente faise “cega”, de xeito que se eliminan os candidatos con masa invariante $535 < m(\eta) < 565$ MeV/c^2 e $940 < m(\eta') < 970$ MeV/c^2. Estes intervalos son as rexións de sinal, ou rexións cegas.

O fondo máis importante nesta análise é $\phi \to \mu^+\mu^-$, que é varias ordes de magnitude máis probable que $\eta' \to \mu^+\mu^-$ e que contamina a rexión superior da masa invariante di-muónica deste modo. A maiores, candidatos de $\eta \to \mu^+\mu^-\gamma$ poderían popular a banda esquerda da masa invariante de $\eta \to \mu^+\mu^-$, pero os cortes introducidos para seleccionar $\eta \to \mu^+\mu^-$ deberían eliminar case todo rastro deste modo. Adicionalmente, un corte $m(\mu^+\mu^-) < 520$ MeV/c^2 introdúcese na canle $\eta' \to \pi^+\pi^-\mu^+\mu^-$ para eliminar contribucións de $\eta' \to \pi^+\pi^-(\eta \to \mu^+\mu^-)$.

Finalmente, o fondo combinatorio é relevante en todos os modos, aínda que con menor peso no $\eta \to \pi^+\pi^-\mu^+\mu^-$. Neste caso, unha selección sobre a identificación dos muóns permite eliminar case todas as contribucións non desexadas. No resto dos casos, adestráronse algoritmos de clasificación para filtrar estas contaminacións.

Estas árbores de decisión usaron un conxunto de variables, dispoñible na Táboa 6.6, que combina cinemática, identificación de partículas, e o illamento das trazas do sinal con respecto ao resto do evento. As variables en simulación corrixíronse a base de comparar o seu comportamento entre datos e simulación no modo de control $D^+_{(s)} \to \pi^+\phi(\mu^+\mu^-)$.

O corte óptimo na figura de mérito obtense estimando o parámetro obxectivo en cada unha das búsquedas. No caso de $\eta' \to \mu^+\mu^-$, téntase optimizar a sensitividade de observar a desintegración. No caso de $\eta' \to \pi^+\pi^-\mu^+\mu^-$ e $\eta \to \mu^+\mu^-$, búscase minimizar os erros totais da medida do $\mathcal{B}$, mentres que para $\eta \to \pi^+\pi^-\mu^+\mu^-$ o obxectivo é minimizar o límite superior do $\mathcal{B}$. Para isto, escanéase unha serie de posibles cortes na variable clasificadora, tanto no modo do D^+ como no do D^+_s. A selección óptima na variable será aquela para o que o parámetro obxectivo sexa máximo ou mínimo, segundo o que requira a búsqueda.

Os erros sistemáticos das medidas inclúen erros no cálculo binomial das eficiencias, así coma os debidos polas diferencias entre datos e MC no disparo, a identificación de partículas, e o algoritmo clasificador. Estas tres valídanse comparando datos e simulación nos correspondentes canais de control. A fonte principal de erro sistemático, non obstante, atópase en medidas previas dos distintos coeficientes de ramificación relevantes na análise. En concreto, $\mathcal{B}(\phi \to \mu^+\mu^-)$ introduce un erro do 5% que é dominante en todos os canais.

O total dos sistemáticos está detallado na Táboa 6.14. Asimesmo, na Táboa 6.20 preséntanse os resultados esperados para cada modo de desintegración.

O axuste final aos datos é novamente de tipo simultáneo, neste caso aos modos D^+ e D^+_s, combinando unha función exponencial que modela o fondo combinatorio (no caso de $\eta' \to \mu^+\mu^-$ engádese unha segunda exponencial que describe a cola esquerda de $\phi \to \mu^+\mu^-$) cunha función Crystal Ball dobre modelando o sinal. O parámetro que pesa a compoñente de sinal, n_{sig}, rescríbese novamente como $n_{\text{sig}} = \mathcal{B}/\alpha$, onde $\mathcal{B}$ é o parámetro de interese en cada un dos canais.

A calidade do axuste valídase co uso de 1500 pseudo-experimentos, en cada un dos cales estímase, a partires de extrapolacións da simulación e o comportamento dos datos arredor da rexión de sinal, o comportamento esperado para os datos na rexión cega.

Os resultados esperados preséntanse na Táboa 6.23.

Bibliography

[1] J. Scharr, *The Drama of Quantum Physics Takes to the Stage*, 2013. `https://www.nbcnews.com/id/wbna52111904`.

[2] A. Einstein, *On the electrodynamics of moving bodies*, Annalen Phys. **17** (1905) 891.

[3] J. C. Maxwell, *A dynamical theory of the electromagnetic field*, Phil. Trans. Roy. Soc. Lond. **155** (1865) 459.

[4] H. A. Lorentz, *La théorie électromagnétique de Maxwell et son application aux corps mouvants*, E.J. Brill, Leide, 1892.

[5] A. Einstein, *The Field Equations of Gravitation*, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1915** (1915) 844.

[6] E. Rutherford, *The scattering of alpha and beta particles by matter and the structure of the atom*, Phil. Mag. Ser. 6 **21** (1911) 669.

[7] J. J. Thomson, *Cathode rays*, Phil. Mag. Ser. 5 **44** (1897) 293.

[8] M. Born and P. Jordan, *Zur Quantenmechanik*, Z. Phys. **34** (1925) 858.

[9] A. Einstein, *Concerning an heuristic point of view toward the emission and transformation of light*, Annalen Phys. **17** (1905) 132.

[10] E. Schrödinger, *An Undulatory Theory of the Mechanics of Atoms and Molecules*, Phys. Rev. **28** (1926) 1049.

[11] W. a Heisenberg, *Uber den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*, Z. Phys. **43** (1927) 172.

[12] C. D. Anderson, *The Positive Electron*, Phys. Rev. **43** (1933) 491.

[13] P. A. M. Dirac, *The quantum theory of the electron*, Proc. Roy. Soc. Lond. A **117** (1928) 610.

[14] P. A. M. Dirac, *A new notation for quantum mechanics*, Mathematical Proceedings of the Cambridge Philosophical Society **35** (1939) 416–418.

[15] E. Fermi, *Quantum Theory of Radiation*, Rev. Mod. Phys. **4** (1932) 87.

[16] P. A. M. Dirac, *Quantum theory of emission and absorption of radiation*, Proc. Roy. Soc. Lond. A **114** (1927) 243.

[17] A. Zee, *Quantum field theory in a nutshell*, 2003.

[18] R. P. Feynman, *Space - time approach to quantum electrodynamics*, Phys. Rev. **76** (1949) 769.

[19] J. R. Oppenheimer *et al.*, *Theoretical interpretation of new particles*, in *6th Annual Rochester Conference on High energy nuclear physics*, VIII.1–36, 1956.

[20] J. C. Street and E. C. Stevenson, *New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron*, Phys. Rev. **52** (1937) 1003.

[21] C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *PROCESSES INVOLVING CHARGED MESONS*, Nature **159** (1947) 694.

[22] G. D. Rochester and C. C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*, Nature **160** (1947) 855.

[23] G. D. Rochester and C. C. Butler, *Evidence for the Existence of New Unstable Elementary Particles*, Nature **160** (1947) 855.

[24] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, doi: 10.2172/4008239.

[25] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214.

[26] G. Zweig, in *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, D. B. Lichtenberg and S. P. Rosen, eds., pp. 22–101, 1964.

[27] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[28] A. Pais and S. B. Treiman, *How Many Charm Quantum Numbers Are There?*, Phys. Rev. Lett. **35** (1975) 1556.

[29] M. Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.

[30] D. Goldberg, *The Standard Model in a Nutshell*, Princeton University Press, 2017.

[31] C. P. Burgess and G. D. Moore, *The standard model: A primer*, Cambridge University Press, 2006.

[32] E. Noether, *Invariante variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) 235.

[33] G. M. Prosperi, M. Raciti, and C. Simolo, *On the running coupling constant in QCD*, Prog. Part. Nucl. Phys. **58** (2007) 387, `arXiv:hep-ph/0607209`.

[34] A. I. Alekseev, *Strong coupling constant to four loops in the analytic approach to QCD*, Few Body Syst. **32** (2003) 193, `arXiv:hep-ph/0211339`.

[35] C. S. Wu *et al.*, *Experimental Test of Parity Conservation in β Decay*, Phys. Rev. **105** (1957) 1413.

[36] E. Fermi, *Tentativo di una teoria dell'emissione dei raggi beta*, Ric. Sci. **4** (1933) 491.

[37] R. P. Feynman and M. Gell-Mann, *Theory of Fermi interaction*, Phys. Rev. **109** (1958) 193.

[38] E. C. G. Sudarshan and R. e. Marshak, *Chirality invariance and the universal Fermi interaction*, Phys. Rev. **109** (1958) 1860.

[39] UA1 collaboration, G. Arnison *et al.*, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV*, Phys. Lett. B **122** (1983) 103.

[40] UA2 collaboration, M. Banner *et al.*, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. B **122** (1983) 476.

[41] UA1 collaboration, G. Arnison *et al.*, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. B **126** (1983) 398.

[42] L3 collaboration, B. Adeva *et al.*, *A Determination of the Properties of the Neutral Intermediate Vector Boson Z^0*, Phys. Lett. B **231** (1989) 509.

[43] S. L. Glashow, *The renormalizability of vector meson interactions*, Nucl. Phys. **10** (1959) 107.

[44] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[45] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, Nuovo Cim. **11** (1959) 568.

[46] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652.

[47] Particle Data Group, N. S. *et al.*, *Review of particle physics*, to be published in Phys. Rev **D110** (2024) 030001.

[48] ATLAS collaboration, G. Aad *et al.*, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, `arXiv:1207.7214`.

[49] CMS collaboration, S. Chatrchyan *et al.*, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30, `arXiv:1207.7235`.

[50] X. Fan, T. G. Myers, B. A. D. Sukra, and G. Gabrielse, *Measurement of the electron magnetic moment*, Phys. Rev. Lett. **130** (2023) 071801.

[51] P. A. M. Dirac, *The quantum theory of the electron. part ii*, Proc. R. Soc. Lond. A **118** (1928) 351.

[52] T. Aoyama *et al.*, *The anomalous magnetic moment of the muon in the Standard Model*, Phys. Rept. **887** (2020) 1, `arXiv:2006.04822`.

[53] Muon g-2 collaboration, D. P. Aguillard *et al.*, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm*, Phys. Rev. Lett. **131** (2023) 161802, `arXiv:2308.06230`.

[54] A. Boccaletti *et al.*, *High precision calculation of the hadronic vacuum polarisation contribution to the muon anomaly*, `arXiv:2407.10913`.

[55] CMD-3 collaboration, F. V. Ignatov *et al.*, *Measurement of the Pion Form Factor with CMD-3 Detector and its Implication to the Hadronic Contribution to Muon (g-2)*, Phys. Rev. Lett. **132** (2024) 231903, `arXiv:2309.12910`.

[56] E.-H. Chao *et al.*, *Hadronic light-by-light contribution to* $(g-2)_\mu$ *from lattice QCD: a complete calculation*, Eur. Phys. J. C **81** (2021) 651, `arXiv:2104.02632`.

[57] Muon g-2 collaboration, T. Marc, *Muon g-2 doubles down with latest measurement, explores uncharted territory in search of new physics*, 2023. `https://news.fnal.gov/2023/08/muon-g-2-doubles-down-with-latest-measurement/`.

[58] HPQCD collaboration, W. G. Parrott, C. Bouchard, and C. T. H. Davies, *Standard Model predictions for* $B \to K\ell^+\ell^-$, $B \to K\ell_1^+\ell_2^-$ *and* $B \to K\nu\bar{\nu}$ *using form factors from* $N_f = 2+1+1$ *lattice QCD*, Phys. Rev. D **107** (2023) 014511, `arXiv:2207.13371`, [Erratum: Phys.Rev.D 107, 119903 (2023)].

[59] LHCb collaboration, R. Aaij *et al.*, *Branching fraction measurements of the rare* $B_s^0 \to \phi\mu^+\mu^-$ *and* $B_s^0 \to f_2'(1525)\mu^+\mu^-$ *decays*, Phys. Rev. Lett. **127** (2021) 151801, `arXiv:2105.14007`.

[60] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP-averaged observables in the* $B^0 \to K^{*0}\mu^+\mu^-$ *decay*, Phys. Rev. Lett. **125** (2020) 011802, `arXiv:2003.04831`.

[61] A. J. Krasznahorkay *et al.*, *Observation of Anomalous Internal Pair Creation in Be8 : A Possible Indication of a Light, Neutral Boson*, Phys. Rev. Lett. **116** (2016) 042501, `arXiv:1504.01527`.

[62] T. Nomura and P. Sanyal, *Explaining Atomki anomaly and muon* $g-2$ *in* $U(1)_X$ *extended flavour violating two Higgs doublet model*, JHEP **05** (2021) 232, `arXiv:2010.04266`.

[63] ATLAS collaboration, M. Aaboud *et al.*, *Measurement of the W-boson mass in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Eur. Phys. J. C **78** (2018) 110, `arXiv:1701.07240`, [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[64] CMS collaboration, V. Chekhovsky *et al.*, *High-precision measurement of the W boson mass with the CMS experiment at the LHC*, `arXiv:2412.13872`.

[65] LHCb collaboration, R. Aaij *et al.*, *Measurement of the W boson mass*, JHEP **01** (2022) 036, `arXiv:2109.01113`.

[66] M. Awramik, M. Czakon, A. Freitas, and G. Weiglein, *Precise Prediction for the W Boson Mass in the Standard Model*, Phys. Rev. D **69** (2004) 053006, `arXiv:hep-ph/0311148`.

[67] CDF collaboration, T. Aaltonen *et al.*, *High-precision measurement of the W boson mass with the CDF II detector*, Science **376** (2022) 170.

[68] G. Bertone and D. Hooper, *History of dark matter*, Rev. Mod. Phys. **90** (2018) 045002, `arXiv:1605.04909`.

[69] J. Alexander *et al.*, *Dark Sectors 2016 Workshop: Community Report*, `arXiv:1608.08632`.

[70] M. Ahlers, J. Jaeckel, J. Redondo, and A. Ringwald, *Probing Hidden Sector Photons through the Higgs Window*, Phys. Rev. D **78** (2008) 075005, `arXiv:0807.4143`.

[71] M. Fabbrichesi, E. Gabrielli, and G. Lanfranchi, *The Dark Photon*, doi: 10.1007/978-3-030-62519-1 `arXiv:2005.01515`.

[72] S. Knapen, T. Lin, and K. M. Zurek, *Light Dark Matter: Models and Constraints*, Phys. Rev. D **96** (2017) 115021, `arXiv:1709.07882`.

[73] H. Ruegg and M. Ruiz-Altaba, *The Stueckelberg field*, Int. J. Mod. Phys. A **19** (2004) 3265, `arXiv:hep-th/0304245`.

[74] J. D. Bjorken, R. Essig, P. Schuster, and N. Toro, *New Fixed-Target Experiments to Search for Dark Gauge Forces*, Phys. Rev. D **80** (2009) 075018, `arXiv:0906.0580`.

[75] CHARM collaboration, F. Bergsma *et al.*, *A Search for Decays of Heavy Neutrinos in the Mass Range 0.5-GeV to 2.8-GeV*, Phys. Lett. B **166** (1986) 473.

[76] E787 collaboration, S. Adler *et al.*, *Further search for the decay K+ —> pi+ nu anti-nu in the momentum region $P <$ 195-MeV/c*, Phys. Rev. D **70** (2004) 037102, `arXiv:hep-ex/0403034`.

[77] APEX collaboration, S. Abrahamyan *et al.*, *Search for a New Gauge Boson in Electron-Nucleus Fixed-Target Scattering by the APEX Experiment*, Phys. Rev. Lett. **107** (2011) 191804, `arXiv:1108.2750`.

[78] H. Merkel *et al.*, *Search at the Mainz Microtron for Light Massive Gauge Bosons Relevant for the Muon g-2 Anomaly*, Phys. Rev. Lett. **112** (2014) 221802, `arXiv:1404.5502`.

[79] BaBar collaboration, B. Aubert *et al.*, *Search for Dimuon Decays of a Light Scalar Boson in Radiative Transitions Upsilon —> gamma A0*, Phys. Rev. Lett. **103** (2009) 081803, `arXiv:0905.4539`.

[80] D. Curtin *et al.*, *Exotic decays of the 125 GeV Higgs boson*, Phys. Rev. D **90** (2014) 075004, `arXiv:1312.4992`.

[81] KLOE-2 collaboration, F. Archilli *et al.*, *Search for a vector gauge boson in ϕ meson decays with the KLOE detector*, Phys. Lett. B **706** (2012) 251, `arXiv:1110.0411`.

[82] HADES collaboration, G. Agakishiev *et al.*, *Searching a Dark Photon with HADES*, Phys. Lett. B **731** (2014) 265, `arXiv:1311.0216`.

[83] LHCb collaboration, R. Aaij *et al.*, *Search for $A' \to \mu^+\mu^-$ decays*, Phys. Rev. Lett. **124** (2020) 041801, `arXiv:1910.06926`.

[84] E. Goudzovski *et al.*, *New physics searches at kaon and hyperon factories*, Rept. Prog. Phys. **86** (2023) 016201, `arXiv:2201.07805`.

[85] M. Pospelov, *Secluded U(1) below the weak scale*, Phys. Rev. D **80** (2009) 095002, `arXiv:0811.1030`.

[86] S. Pokorski, *GAUGE FIELD THEORIES*, Cambridge University Press, 2005.

[87] D.-d. Wu, *A Brief Introduction to the Strong CP Problem*, Z. Naturforsch. A **52** (1997) 179.

[88] R. D. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, Phys. Rev. Lett. **38** (1977) 1440.

[89] R. D. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, Phys. Rev. D **16** (1977) 1791.

[90] M. F. Sohnius, *Introducing Supersymmetry*, Phys. Rept. **128** (1985) 39.

[91] J. Wess and J. Bagger, *Supersymmetry and supergravity*, Princeton University Press, Princeton, NJ, USA, 1992.

[92] T. Skwarnicki, *A study of the radiative CASCADE transitions between the Upsilon-Prime and Upsilon resonances*, PhD thesis, Cracow, INP, 1986.

[93] D. Martínez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. A **764** (2014) 150, `arXiv:1312.5000`.

[94] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60.

[95] ALEPH, DELPHI, L3, OPAL collaboration, P. Bock *et al.*, *Searches for Higgs bosons: Preliminary combined results using LEP data collected at energies up to 202-GeV*, .

[96] *LHC Machine*, JINST **3** (2008) S08001.

[97] ATLAS collaboration, G. Aad *et al.*, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[98] CMS collaboration, S. Chatrchyan *et al.*, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[99] ALICE collaboration, K. Aamodt *et al.*, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[100] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[101] E. Boltezar *et al.*, *THE NEW CERN 50-MEV LINAC.*, in *10th International Linear Accelerator Conference*, S2–2, 1980.

[102] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, `arXiv:1412.6352`.

[103] LHCb collaboration, R. Aaij *et al.*, *Prompt K^0_S production in pp collisions at $\sqrt{s}=0.9$ TeV*, Phys. Lett. **B693** (2010) 69, `arXiv:1008.3105`.

[104] LHCb collaboration, *Framework TDR for the LHCb Upgrade: Technical Design Report*, CERN-LHCC-2012-007, 2012.

[105] LHCb collaboration, *Letter of Intent for the LHCb Upgrade*, .

[106] LHCb collaboration, R. Aaij *et al.*, *Analysis of neutral B-meson decays into two muons*, Phys. Rev. Lett. **128** (2022) 041801, `arXiv:2108.09284`.

[107] LHCb collaboration, R. Aaij *et al.*, *Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda^0_b \to J/\psi p K^-$ decays*, Phys. Rev. Lett. **115** (2015) 072001, `arXiv:1507.03414`.

[108] LHCb collaboration, R. Aaij *et al.*, *Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$*, Phys. Rev. Lett. **122** (2019) 222001, `arXiv:1904.03947`.

[109] LHCb collaboration, R. Aaij *et al.*, *Observation of the doubly charmed baryon Ξ^{++}_{cc}*, Phys. Rev. Lett. **119** (2017) 112001, `arXiv:1707.01621`.

[110] LHCb collaboration, R. Aaij *et al.*, *Evidence for CP violation in time-integrated $D^0 \to h^- h^+$ decay rates*, Phys. Rev. Lett. **108** (2012) 111602, `arXiv:1112.0938`.

[111] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP violation and the B^0_s meson decay width difference with $B^0_s \to J/\psi K^+ K^-$ and $B^0_s \to J/\psi \pi^+ \pi^-$ decays*, Phys. Rev. **D87** (2013) 112010, `arXiv:1304.2600`.

[112] LHCb collaboration, R. Aaij *et al.*, *Measurement of the CP violating phase and decay-width difference in $B^0_s \to \psi(2S)\phi$ decays*, Phys. Lett. **B762** (2016) 253, `arXiv:1608.04855`.

[113] LHCb collaboration, R. Aaij *et al.*, *Resonances and CP-violation in B^0_s and $\overline{B}^0_s \to J/\psi K^+ K^-$ decays in the mass region above the $\phi(1020)$*, JHEP **08** (2017) 037, `arXiv:1704.08217`.

[114] LHCb collaboration, R. Aaij *et al.*, *Updated measurement of time-dependent CP-violating observables in $B^0_s \to J/\psi K^+ K^-$ decays*, Eur. Phys. J. **C79** (2019) 706, Erratum ibid. **C80** (2020) 601, `arXiv:1906.08356`.

[115] LHCb collaboration, R. Aaij *et al.*, *Constraints on the $K^0_S \to \mu^+\mu^-$ branching fraction*, Phys. Rev. Lett. **125** (2020) 1231801, `arXiv:2001.10354`.

[116] LHCb collaboration, R. Aaij *et al.*, *Search for $K^0_{S(L)} \to \mu^+\mu^-\mu^+\mu^-$ decays at LHCb*, Phys. Rev. **D108** (2023) L031102, `arXiv:2212.04977`.

[117] LHCb collaboration, M. Pili, *Performance of the LHCb detector in the Run 2*, PoS **ICHEP2020** (2021) 767.

[118] R. Lindner, *LHCb layout*, LHCb Collection., 2008.

[119] LHCb collaboration, R. Aaij *et al.*, *The LHCb Upgrade I*, JINST **19** (2024) P05065, `arXiv:2305.10515`.

[120] LHCb collaboration, *LHCb VELO (VErtex LOcator): Technical Design Report*, CERN-LHCC-2001-011, 2001.

[121] R. Aaij *et al.*, *Performance of the LHCb Vertex Locator*, JINST **9** (2014) P09007, `arXiv:1405.7808`.

[122] LHCb collaboration, *LHCb VELO Upgrade Technical Design Report*, CERN-LHCC-2013-021, 2013.

[123] LHCb collaboration, *LHCb reoptimized detector design and performance: Technical Design Report*, CERN-LHCC-2003-030, 2003.

[124] LHCb collaboration, *LHCb Tracker Upgrade Technical Design Report*, CERN-LHCC-2014-001, 2014.

[125] LHCb collaboration, *LHCb inner tracker: Technical Design Report*, CERN-LHCC-2002-029, 2002.

[126] LHCb collaboration, *LHCb outer tracker: Technical Design Report*, CERN-LHCC-2001-024, 2001.

[127] R. Arink *et al.*, *Performance of the LHCb Outer Tracker*, JINST **9** (2014) P01002, `arXiv:1311.3893`.

[128] LHCb collaboration, *LHCb magnet: Technical Design Report*, CERN-LHCC-2000-007, 2000.

[129] LHCb collaboration, *LHCb RICH: Technical Design Report*, CERN-LHCC-2000-037, 2000.

[130] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, JINST **17** (2022) P07013, `arXiv:2205.13400`.

[131] R. Calabrese *et al.*, *Performance of the LHCb RICH detectors during LHC Run 2*, JINST **17** (2022) P07013, `arXiv:2205.13400`.

[132] LHCb collaboration, *LHCb calorimeters: Technical Design Report*, CERN-LHCC-2000-036, 2000.

[133] C. Abellan Beteta *et al.*, *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, `arXiv:2008.11556`, submitted to JINST.

[134] LHCb collaboration, E. Picatoste Olloqui, *LHCb preshower(PS) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring*, J. Phys. Conf. Ser. **160** (2009) 012046.

[135] S. Barsuk *et al.*, *Design and construction of electromagnetic calorimeter for LHCb experiment*, .

[136] J. Badier *et al.*, *Shashlik calorimetry a combined Shashlik + preshower detector for LHC: R&D proposal*, .

[137] J. Badier *et al.*, *Shashlik calorimeter: Beam test results*, Nucl. Instrum. Meth. A **348** (1994) 74.

[138] A. Arefev *et al.*, *Beam test results of the LHCb electromagnetic calorimeter*, .

[139] R. Djeliadine, O. Iouchtchenko, and V. F. Obraztsov, *LHCb hadron trigger and Hcal cell size and length optimization*, .

[140] C. Coca *et al.*, *The hadron calorimeter prototype beam-test results*, .

[141] LHCb collaboration, *LHCb muon system: Technical Design Report*, CERN-LHCC-2001-010, 2001.

[142] LHCb collaboration, E. Santovetti, *The LHCb muon detector*, Nucl. Instrum. Meth. A **462** (2001) 297.

[143] R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, `arXiv:1812.10790`.

[144] T. Head, *The LHCb trigger system*, JINST **9** (2014) C09015.

[145] R. Aaij *et al.*, *The LHCb trigger and its performance in 2011*, JINST **8** (2013) P04022, `arXiv:1211.3055`.

[146] P. d'Argent *et al.*, *Improved performance of the LHCb Outer Tracker in LHC Run 2*, JINST **12** (2017) P11016, `arXiv:1708.00819`.

[147] LHCbcollaboration, *RTA and DPA dataflow diagrams for Run 1, Run 2, and the upgraded LHCb detector* , .

[148] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039, 2014.

[149] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, `arXiv:0710.3820`.

[150] LHCb collaboration, I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, in *2010 IEEE Nuclear Science Symposium, Medical Imaging Conference, and 17th Room Temperature Semiconductor Detectors Workshop*, 1155–1161, 2010.

[151] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152.

[152] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, `arXiv:hep-ph/0506026`.

[153] J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270.

[154] GEANT4 collaboration, S. Agostinelli *et al.*, *GEANT4 - A Simulation Toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250.

[155] LHCb collaboration, M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[156] G. Corti *et al.*, *Software for the LHCb experiment*, IEEE Trans. Nucl. Sci. **53** (2006) 1323.

[157] T. L. Collaboration, *The BOOLE Project*, 2024. `https://lhcbdoc.web.cern.ch/lhcbdoc/boole/`.

[158] LHCb collaboration, *LHCb Upgrade Software and Computing*, CERN-LHCC-2018-007, 2018.

[159] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, J. Fluids Eng. **82** (1960) 35.

[160] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. A **262** (1987) 444.

[161] M. De Cian, S. Farry, P. Seyfert, and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, .

[162] LHCb collaboration, R. Aaij *et al.*, *Measurement of the track reconstruction efficiency at LHCb*, JINST **10** (2015) P02007, `arXiv:1408.1251`.

[163] G. Lanfranchi *et al.*, *The Muon Identification Procedure of the LHCb Experiment for the First Data*, LHCb-PUB-2009-013, 2009.

[164] O. Lupton, L. Anderlini, B. Sciascia, and V. Gligorov, *Calibration samples for particle identification at LHCb in Run 2*, .

[165] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, `arXiv:physics/0402083`.

[166] L. Anderlini *et al.*, *The PIDCalib package*, LHCb-PUB-2016-021, 2016.

[167] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, `arXiv:1803.00824`.

[168] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Meth. A **552** (2005) 566, `arXiv:physics/0503191`.

[169] J. Amoraal *et al.*, *Application of vertex and mass constraints in track-based alignment*, Nucl. Instrum. Meth. A **712** (2013) 48, `arXiv:1207.4756`.

[170] A. Pais, *Some Remarks on the V-Particles*, Phys. Rev. **86** (1952) 663.

[171] M. Gell-Mann, *Isotopic Spin and New Unstable Particles*, Phys. Rev. **92** (1953) 833.

[172] P. Langacker, *The standard model and beyond*, 2010.

[173] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the* 2π *Decay of the* K_2^0 *Meson*, Phys. Rev. Lett. **13** (1964) 138.

[174] C. Bobeth and A. J. Buras, *Leptoquarks meet* ε'/ε *and rare Kaon processes*, JHEP **02** (2018) 101, `arXiv:1712.01295`.

[175] V. Chobanova *et al.*, *Probing SUSY effects in* $K_S^0 \to \mu^+\mu^-$, JHEP **05** (2018) 024, `arXiv:1711.11030`.

[176] A. J. Buras, *Weak Hamiltonian, CP violation and rare decays*, `arXiv:hep-ph/9806471`.

[177] A. J. Buras and M. Lindner, eds., *Heavy flavours II*, WSP, Singapore, 1998.

[178] A. Pich, *Effective field theory: Course*, `arXiv:hep-ph/9806303`.

[179] G. D'Ambrosio, G. Isidori, and J. Portoles, *Can we extract short distance information from* $B(K_{\rm L}^0 \to \mu^+\mu^-)$*?*, Phys. Lett. B **423** (1998) 385, `arXiv:hep-ph/9708326`.

[180] G. D'Ambrosio and D. Espriu, *Rare Decay Modes of the K Mesons in the Chiral Lagrangian*, Phys. Lett. B **175** (1986) 237.

[181] G. D'Ambrosio and T. Kitahara, *Direct CP Violation in* $K \to \mu^+\mu^-$, Phys. Rev. Lett. **119** (2017) 201802, `arXiv:1707.06999`.

[182] A. A. Alves Junior *et al.*, *Prospects for Measurements with Strange Hadrons at LHCb*, JHEP **05** (2019) 048, `arXiv:1808.03477`.

[183] G. D'Ambrosio, D. Greynat, and G. Vulvert, *Standard Model and New Physics contributions to* K_L *and* K_S *into four leptons*, Eur. Phys. J. C **73** (2013) 2678, `arXiv:1309.5736`.

[184] J. Podolanski and R. Armenteros, *Iii. analysis of v-events*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **45** (1954) 13, `arXiv:https://doi.org/10.1080/14786440108520416`.

[185] M. Alexander *et al.*, *Mapping the material in the LHCb vertex locator using secondary hadronic interactions*, JINST **13** (2018) P06008, `arXiv:1803.07466`.

[186] G. E. Forsythe, *Von neumann's comparison method for random sampling from the normal and other distributions*, Mathematics of Computation **26** (1972) 817.

[187] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer Publishing Company, Incorporated, 2008.

[188] F. A. Maurice Pierre Morgenthaler, Giulia Frau, *TrackCalib2*, 2024. `https://gitlab.cern.ch/farchill/trackcalib2`.

[189] A. Hoecker *et al.*, *TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide.*, `arXiv:physics/0703039`.

[190] T. L. Collaboration, *The URANIA Project*, 2021. `https://lhcbdoc.web.cern.ch/lhcbdoc/urania/`.

[191] G. D'Ambrosio and T. Kitahara, *Direct CP Violation in $K \to \mu^+\mu^-$*, Phys. Rev. Lett. **119** (2017) 201802, `arXiv:1707.06999`.

[192] LHCb collaboration, *LHCb Upgrade II Scoping Document*, CERN-LHCC-2024-010, 2024, under LHCC review.

[193] C. E. Thomas, *Composition of the Pseudoscalar Eta and Eta' Mesons*, JHEP **10** (2007) 026, `arXiv:0705.1500`.

[194] R. Escribano, S. Gonzàlez-Solís, R. Jora, and E. Royo, *Theoretical analysis of the doubly radiative decays $\eta^{(\prime)} \to \pi^0\gamma\gamma$ and $\eta' \to \eta\gamma\gamma$*, Phys. Rev. D **102** (2020) 034026, `arXiv:1812.08454`.

[195] Y. Balytskyi, *Leptophobic dark photon interpretation of the $\eta^{(\prime)} \to \pi^0\,(\eta)\,\gamma\gamma$ puzzle*, `arXiv:2112.02769`.

[196] B.-l. Young, *Rare Decay Modes of the eta Meson*, Phys. Rev. **161** (1967) 1620.

[197] C. Quigg and J. D. Jackson, *Decays of neutral pseudoscalar mesons into lepton pairs*, UCRL–18487 (1968).

[198] K. S. Babu and E. Ma, *Pseudoscalar Electromagnetic Form-factors: Vector Dominance and Quantum Chromodynamics*, Phys. Lett. **119B** (1982) 449.

[199] L. Ametller, L. Bergstrom, A. Bramon, and E. Masso, *The Quark Triangle: Application to Pion and η Decays*, Nucl. Phys. **B228** (1983) 301.

[200] M. D. Scadron and M. Visinescu, *The Soft Momentum Quark Model Predictions for the Pseudoscalar Meson Decays Into Lepton Pairs*, Phys. Rev. **D29** (1984) 911.

[201] B. Margolis, J. Ng, M. Phipps, and H. D. Trottier, *Quark model calculation of $\eta \to$ lepton+ lepton- to all orders in the bound state relative momentum*, Phys. Rev. **D47** (1993) 1942, `arXiv:hep-ph/9210259`.

[202] M. J. Savage, M. E. Luke, and M. B. Wise, *The Rare decays $\pi^0 \to e^+e^-$, $\eta \to e^+e^-$ and $\eta \to \mu^+\mu^-$ in chiral perturbation theory*, Phys. Lett. **B291** (1992) 481, `arXiv:hep-ph/9207233`.

[203] D. Gomez Dumm and A. Pich, *Long distance contributions to the $K(L) \to \mu^+\mu^-$ decay width*, Phys. Rev. Lett. **80** (1998) 4633, `arXiv:hep-ph/9801298`.

[204] L. Ametller, A. Bramon, and E. Masso, *The $\pi^0 \to e^+e^-$ and $\eta \to \mu^+\mu^-$ decays revisited*, Phys. Rev. **D48** (1993) 3388, `arXiv:hep-ph/9302304`.

[205] Z. K. Silagadze, *On the two-photon contributions to $e^+e^- \to \eta\gamma$ and $e^+e^- \to \eta'\gamma$*, Phys. Rev. **D74** (2006) 054003, `arXiv:hep-ph/0606284`.

[206] M. Knecht, S. Peris, M. Perrottet, and E. de Rafael, *Decay of pseudoscalars into lepton pairs and large N(c) QCD*, Phys. Rev. Lett. **83** (1999) 5230, `arXiv:hep-ph/9908283`.

[207] A. E. Dorokhov and M. A. Ivanov, *Rare decay $\pi^0 \to e^+e^-$: Theory confronts KTeV data*, Phys. Rev. **D75** (2007) 114007, `arXiv:0704.3498`.

[208] A. E. Dorokhov, *How the recent BABAR data for $P \to \gamma\gamma*$ affect the Standard Model predictions for the rare decays $P \to l^+l^-$*, JETP Lett. **91** (2010) 163, `arXiv:0912.5278`.

[209] P. Masjuan and P. Sanchez-Puertas, *η and η' decays into lepton pairs*, JHEP **08** (2016) 108, `arXiv:1512.09292`.

[210] H. Davoudiasl, H.-S. Lee, and W. J. Marciano, *'Dark' Z implications for Parity Violation, Rare Meson Decays, and Higgs Physics*, Phys. Rev. **D85** (2012) 115019, `arXiv:1203.2947`.

[211] H. Davoudiasl, H.-S. Lee, and W. J. Marciano, *Muon Anomaly and Dark Parity Violation*, Phys. Rev. Lett. **109** (2012) 031802, `arXiv:1205.2709`.

[212] Q. Chang and Y.-D. Yang, *Rare decay $\pi^0 \to e^+e^-$ as a sensitive probe of light CP-odd Higgs in NMSSM*, Phys. Lett. **B676** (2009) 88, `arXiv:0808.2933`.

[213] C. E. Carlson and B. C. Rislow, *New Physics and the Proton Radius Problem*, Phys. Rev. **D86** (2012) 035013, `arXiv:1206.3587`.

[214] S. G. Karshenboim, D. McKeen, and M. Pospelov, *Constraints on muon-specific dark forces*, Phys. Rev. **D90** (2014) 073004, `arXiv:1401.6154`, [Addendum: Phys. Rev.D90,no.7,079905(2014)].

[215] T. Petri, *Anomalous decays of pseudoscalar mesons*, `arXiv:1010.2378`.

[216] B. Borasoy and R. Nissler, *η, $\eta' \to \pi^+\pi^-l^+l^-$ in a chiral unitary approach*, Eur. Phys. J. A **33** (2007) 95, `arXiv:0705.0954`.

[217] A. Faessler, C. Fuchs, and M. I. Krivoruchenko, *Dilepton spectra from decays of light unflavored mesons*, Phys. Rev. C **61** (2000) 035206, `arXiv:nucl-th/9904024`.

[218] Z.-Y. Zhang, L.-Q. Qin, and S.-S. Fang, *Event generators for η/η' rare decays into $\pi^+\pi^-l^+l^-$ and $e^+e^-\mu^+\mu^-$*, Chin. Phys. C **36** (2012) 926.

[219] A. Luchinskii, *New EvtGen model for eta' -¿ mu mu pi pi, eta' -¿ e e pi pi decays* , 2020. `https://gitlab.cern.ch/lhcb/Gauss/-/merge_requests/626`.

[220] F. Jegerlehner and A. Nyffeler, *The Muon g-2*, Phys. Rept. **477** (2009) 1, `arXiv:0902.3360`.

[221] L. Gan, B. Kubis, E. Passemar, and S. Tulin, *Precision tests of fundamental physics with η and η' mesons*, `arXiv:2007.00664`.

[222] N. T. Huong, E. Kou, and B. Viaud, *Novel approach to measure the leptonic $\eta(')\rightarrow \mu^+\mu^-$ decays via charmed meson decays*, Phys. Rev. D **94** (2016) 054040, `arXiv:1606.08195`.

[223] M. J. Ramsey-Musolf and M. B. Wise, *Hadronic light by light contribution to muon g-2 in chiral perturbation theory*, Phys. Rev. Lett. **89** (2002) 041601, `arXiv:hep-ph/0201297`.

[224] B. D. Hyams *et al.*, *Observation of the muon-minus muon-plus decay mode of the eta-neutral meson*, Phys. Lett. B **29** (1969) 128.

[225] R. I. Dzhelyadin *et al.*, *Study of $\eta \rightarrow \mu^+\mu^-$ Decay*, Sov. J. Nucl. Phys. **32** (1980) 518.

[226] R. S. Kessler *et al.*, *Measurement of the branching ratio for the decay $\eta \rightarrow \mu^+\mu^-$*, Phys. Rev. Lett. **70** (1993) 892.

[227] R. Abegg *et al.*, *Measurement of the branching ratio for the decay $\eta \rightarrow \mu^+\mu^-$*, Phys. Rev. D **50** (1994) 92.

[228] BESIII collaboration, M. Ablikim *et al.*, *Observation of $\eta' \rightarrow \pi^+\pi^-\mu^+\mu^-$*, `arXiv:2012.04257`.

[229] BESIII collaboration, M. Ablikim *et al.*, *Measurement of $\eta' \rightarrow \pi^+\pi^-e^+e^-$ and $\eta' \rightarrow \pi^+\pi^-\mu^+\mu^-$*, Phys. Rev. D **87** (2013) 092011, `arXiv:1303.7360`.

[230] CLEO collaboration, P. Naik *et al.*, *Observation of η' decays to $\pi^+\pi^-\pi^0$ and $\pi^+\pi^-e^+e^-$*, Phys. Rev. Lett. **102** (2009) 061801, `arXiv:0809.2587`.

[231] BESIII collaboration, M. Ablikim *et al.*, *Study of $\eta \rightarrow \pi^+\pi^-l^+l^-$*, `arXiv:2501.10130`.

[232] M. Reece and L.-T. Wang, *Searching for the light dark gauge boson in GeV-scale experiments*, JHEP **07** (2009) 051, `arXiv:0904.1743`.

[233] D. Aloni, Y. Soreq, and M. Williams, *Coupling QCD-Scale Axionlike Particles to Gluons*, Phys. Rev. Lett. **123** (2019) 031803, `arXiv:1811.03474`.

[234] G. Landini and E. Meggiolaro, *Study of the interactions of the axion with mesons and photons using a chiral effective Lagrangian model*, Eur. Phys. J. C **80** (2020) 302, `arXiv:1906.03104`.

[235] BESIII collaboration, M. Ablikim *et al.*, *Measurement of absolute branching fractions of D_s^+ hadronic decays*, `arXiv:2403.19256`.

[236] LHCb collaboration, R. Aaij *et al.*, *Measurements of the branching fraction ratio $\mathcal{B}(\phi \rightarrow \mu^+\mu^-)/\mathcal{B}(\phi \rightarrow e^+e^-)$ with charm meson decays*, `arXiv:2402.01336`, submitted to JHEP.

[237] M. Williams *et al.*, *The HLT2 Topological Lines* , CERN, Geneva, 2011.

[238] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), 785–794, ACM, 2016.

[239] T. Akiba *et al.*, *Optuna: A next-generation hyperparameter optimization framework*, in *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631, 2019.

List of Figures

List of Tables

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

CAMPUS VIDA
CAMPUS DE EXCELENCIA INTERNACIONAL

This thesis presents the first-ever study of the decay of a neutral kaon into four muons. Using data collected by the LHCb experiment between 2016 and 2018, no signal was found, and stringent upper limits on the branching fractions of both the short- and long-lived modes were set.

Additionally, the study of four decays of eta and eta prime mesons are presented. The thesis introduces the first-ever study of the eta prime decaying into a pair of muons, as well as significant improvements on the precision of the branching fractions of the di-muon decay of the eta, as well as the decays of both the eta and eta prime mesons into two pairs of muons and two pairs of pions.