UNIVERSITÉ SAVOIE MONT BLANC

THÈSE

Pour obtenir le grade de

Docteur de l'université Savoie Mont Blanc

Spécialité : **Physique Théorique**

Arrêté ministériel : 25 Mai 2016

Présentée par

Méril Reboud

Thèse dirigée par **Diego Guadagnoli**
et coencadrée par **Jean-François Marchand**

Préparée au sein du **Laboratoire d'Annecy-le-Vieux de Physique Théorique** et de l'**École doctorale de physique de Grenoble**

The discrepancies in b $\to$ s semi-leptonic decays:

A complete study from model-building aspects, to the definition of novel observables, to their measurement with the LHCb detector

Thèse soutenue publiquement le **26 Novembre 2020**,
devant le jury composé de :

Geneviève Belanger
LAPTh, Présidente
Yasmine Amhis
IJCLab, Rapporteure
Svjetlana Fajfer
IJS, Rapporteure
Danny van Dyk
TUM, Examinateur
Gino Isidori
UZH, Examinateur
Francesco Polci
LPNHE, Examinateur
Diego Guadagnoli
LAPTh, Directeur de thèse
Jean-François Marchand
LAPP, Co-Encadrant de thèse

À la mémoire de Louis Reboud.

Abstract

Although the Standard Model (SM) of particle physics gives an overall excellent description of the observations, a few results, mainly obtained by the LHCb detector at CERN, point towards deviations in the transitions from quark b to quarks s and c. If confirmed, these anomalies would give a clear signal of physics beyond the SM, as they violate lepton flavour universality (LFU). In this context, new measurements and their theoretical predictions are crucial to define possible new physics scenarios. This thesis concerns both aspects and is therefore divided into two parts. The first one is devoted to theoretical considerations on the b to s transition as well as models explaining the anomalies. The second part presents a new measurement, namely the search for the decays of B_s^0 mesons to two muons and a photon at LHCb.

Radiative leptonic decays are promising to test the SM because the additional photon not only enlarges the branching ratio by lifting the chiral suppression factor, but also offers a sensitivity to other operators. Using the language of effective field theory (EFT), the decay of the B_s^0 meson into two leptons and a photon is studied and new methods to reduce the theoretical uncertainty on its branching ratio are proposed. Besides, the behaviour of this decay at large dilepton mass gives the possibility of an indirect measurement where the total branching ratio is measured as a background of the corresponding non-radiative channel. Furthermore, if the violation LFU is experimentally confirmed, one also expects violation of lepton flavour. Measuring such violation, which would be an undeniable sign of new physics, is yet very challenging. Radiative decays can however support this search by offering additional channels with potentially larger branching ratios.

The interpretation of the B anomalies in term of shifts in the EFT coefficients put a few scenarios forward. These scenarios can then be interpreted in term of new physics models. One such model is based on the consideration of an additional symmetry group acting horizontally between the SM generations. This model is severely constrained by low energy observables, such as meson mixings and leptonic decays, but allowing for a mass degeneracy between the new group's bosons explains all $b \to s$ anomalies while passing other experimental constraints. The absence of new physics in low energy observables can also be interpreted as the presence of leptoquarks, for which the interaction between two quark- or two lepton-currents only arises at the one loop level. A model based on a vector leptoquark can for example give an explanation to both $b \to s$ and $b \to c$ anomalies. Interestingly, ultraviolet completions of these models contain natural Dark Matter candidates, hence relating two outstanding problems of particle physics.

The analysis presented in the second part is particularly challenging because the probed decay is both very rare and radiative. The main difficulty lies in the presence of a large combinatorial background due to light meson decays. Tackling it while keeping a high efficiency on the signal selection requires the use of two successive multivariate analyses. The signal is then normalised to a similar decay of the B_s^0 where no new physics is expected. Measuring a ratio of yields instead of an unique branching ratio allows for a partial cancellation of experimental uncertainties. On the other hand, this procedure requires a good knowledge of each selection step efficiencies for the two channels. These efficiencies are extracted from Monte-Carlo simulations or, when possible, directly from the data. Even if no significant excess is found, this analysis will allow to set the first limit on the total branching ratio.

Résumé

Bien que le Modèle Standard (SM) de la physique des particules permette une excellente description des observations, quelques résultats, principalement obtenus par le détecteur LHCb du CERN, montrent des signes de déviations dans les transitions du quark b aux quarks s et c. Si ces anomalies sont confirmées, elles donneront un signal clair de physique au-delà du SM car elles violent l'universalité de la saveur leptonique. Dans ce contexte, de nouvelles mesures et leurs prédictions théoriques sont cruciales pour définir de potentiels scénarios de nouvelle physique (NP). Cette thèse s'intéresse à tous ces aspects et est divisée en deux parties. La première propose des considérations théoriques sur la transition $b \to s$ ainsi que des modèles expliquant les anomalies. La seconde partie présente une nouvelle mesure, la recherche de désintégrations de mésons B_s^0 en deux muons et un photon à LHCb.

Les désintégrations leptoniques radiatives sont particulièrement prometteuses pour tester le SM, car l'ajout du photon n'entraîne pas seulement un agrandissement du rapport de branchement (BR), mais offre aussi une sensibilité à de nouveaux opérateurs. La désintégration du B_s^0 en deux leptons et un photon est alors étudiée en utilisant une théorie effective des champs, et de nouvelles méthodes permettant de réduire l'incertitude théorique sur son BR sont proposées. Le comportement de cette désintégration lorsque la masse du dilepton est grande donne par ailleurs la possibilité de réaliser une mesure indirecte, où le BR est mesuré comme bruit de fond de son équivalent non-radiatif. En outre, si la violation de l'universalité leptonique est confirmée, on s'attend aussi à une violation de la saveur leptonique. La mesure de cette dernière, signe indéniable de NP, est néanmoins très difficile. Les désintégrations radiatives permettent cependant de simplifier cette recherche en offrant des canaux supplémentaires.

L'interprétation des anomalies en termes de modification des coefficients de la théorie effective met en avant quelques scénarios. Ceux-ci peuvent alors être interprétés en terme de modèles de NP. L'un d'eux se base sur un nouveau groupe de symétrie qui agit horizontalement entre les générations du SM. Ce modèle est fortement contraint par des observables de basse énergie, mais attribuer différentes masses aux bosons du nouveau groupe permet d'expliquer les anomalies $b \to s$ tout en satisfaisant les contraintes expérimentales. L'absence de NP dans les observables de basse énergie peut en outre suggérer la présence de leptoquarks, pour lesquels l'interaction entre deux courants de quarks ou de leptons n'arrive qu'à une boucle. Un modèle basé sur un leptoquark vecteur explique par exemple à la fois les anomalies $b \to s$ et $b \to c$. Il est intéressant de noter que les complétions ultraviolettes de ces différents modèles contiennent des candidats pour la Matière Noire, reliant ainsi deux problèmes prépondérants de la physique des particules.

L'analyse présentée dans la seconde partie est rendue particulièrement difficile par la rareté et le caractère radiatif de la désintégration étudiée. La difficulté principale réside dans la présence d'un important bruit de fond combinatoire dû à la désintégration de mésons légers. Le prendre en compte tout en gardant une grande efficacité sur la sélection du signal nécessite d'utiliser deux analyses multivariées successives. Le signal est alors normalisé à une désintégration similaire du B_s^0 où aucune NP est attendue. Mesurer le rapport des efficacités au lieu d'un unique BR permet une suppression partielle des incertitudes expérimentales. En revanche, cette procédure nécessite une bonne connaissance des efficacités de chaque étape de la sélection. Celles-ci sont extraites de simulation Monte-Carlo ou, quand c'est possible, directement des données. Même si aucun excès significatif n'est observé, cette analyse permettra d'établir la première limite sur le BR total.

Contents

Acknowledgement

Je souhaite tout d'abord remercier les membres de mon jury pour avoir accepté d'en faire partie. Le contexte sanitaire de l'année 2020 ne leur auront pas permis d'assister physiquement à la soutenance, et j'espère avoir l'occasion de poursuivre les discussions que nous y avons entamées lorsque les conditions le permettront. Merci donc à Yamine Amhis et Svjetlana Fajfer d'avoir accepté d'être rapporteures, à Geneviève Bélanger d'avoir présidé ce jury et à Danny van Dyk, Gino Isidori et Francesco Polci pour leur rôle d'examinateurs.

Merci également à Diego Guadagnoli et Jean-François Marchand pour leur encadrement durant cette thèse ; ils m'ont donné l'opportunité de travailler sur des aspects de physique à la fois théorique et expérimentale. Je les remercie en particulier pour la patience dont ils ont fait preuve quand les contraintes de ces deux approches nécessitaient de se concentrer sur l'une plutôt que sur l'autre.

Je remercie ensuite chaleureusement le LAPTh, son équipe administrative et le groupe LHCb du LAPP pour leur accueil. Merci en particulier à Éric Ragoucy et Philippe Ghez pour nos nombreuses discussions scientifiques et non scientifiques. Étant resté au laboratoire pendant assez longtemps, j'ai eu le plaisir de croiser beaucoup de doctorant·es et post-doctorant·es, dans un renouvellement aussi frustrant qu'enrichissant. Merci donc aux actuel·les et aux ancien·nes[1] pour les grands moments passés ensemble. Je tiens plus particulièrement à remercier Jordan pour sa patience et son humilité, Alix pour son amitié et Anne pour toutes les discussions que nous avons eues.

Merci également à mes ami·es de prépa, Damien, Paul et Florent, et de l'ENS, Laetitia et Étienne, pour leur soutien ainsi que pour leur présence à ma soutenance.

Pour finir, j'aimerais remercier ma famille et Julie, pour beaucoup de choses en lien ou non avec cette thèse.

[1] En voici une liste non-exhaustive : Florian, Dimitri, Marc, Alexandre, Camille, Ali, Victor, Laura, Peter, Carole, Anthony, Daniel, Thibaud, Romain, Yoann, Matthieu, Mathieu, Philippe, Vincent.

Introduction

The Standard Model (SM) of particle physics was built upon a dialogue between theorists and experimentalists. Some results, sometimes unexpected, were obtained in experimental measurements and required an improvement of the theory. This has been the case for instance in the 1950s and 1960s, when particle colliders found a zoo of new hadron states. The classification of these states, the so-called "eightfold way" [1, 2], leads to the development of the quark model and paved the way to the current description of quantum chromodynamics (QCD) [3–5]. Inversely, some theoretical constructions have motivated dedicated experimental searches. The most recent example is the discovery of the Higgs boson by the ATLAS and CMS collaborations [6, 7], as predicted by the Higgs mechanism responsible for electroweak symmetry breaking (EWSB) in the SM [8–12]. Although this discovery completed the construction of the SM, the dialogue between theorists and experimentalists still occurs. Indeed, although remarkable for the precision of its predictions, the SM is known to be incomplete. Among the questions left open by the SM, some are particularly interesting in the context of this thesis.

One of the clearest evidences of physics beyond the Standard Model (BSM), comes from the observation of neutrinos. It is now well-established, from the study of different sources of these particles that they are massive and mix among flavour. Their behaviour was deciphered by several experiments that probed atmospheric neutrinos [13–18], solar neutrinos [19–27] and also neutrinos in the vicinity of nuclear reactors [28–31] and beams of neutrinos coming from particle accelators [32–34]. In the SM however, a coupling between neutrinos and the Higgs boson would require the presence of right-handed neutrinos that are singlet under the SM gauge group and therefore absent by construction. Several extensions of the SM offers mechanisms to generate neutrino masses, but the smallness of these masses ($m \lesssim 1$ eV [35]) usually requires the introduction of new physics (NP) at scales well beyond the reach of current experiments.

The discovery of neutrino masses propelled them, for a time, as potential candidates for Dark Matter (DM). The concept of DM was introduced to explain several cosmological observations. The main evidences of its existence come from the study of galactic rotation curves, gravitational lensing and large structures in the observable universe (see [36] for a review)[2]. DM triggered a lot of experimental and theoretical attentions both in astrophysics and particle physics. As far as

[2] Note that a problem of missing mass was already encountered in physics. In 1846, Le Verrier and Adams were trying to make sense of Uranus trajectory. Assuming that the observed anomaly to Newton's law was due to the presence of a massive object, they allowed Galle to discover Neptune. This discovery is proof of the effectiveness of indirect measurements.

neutrino are concerned, it appears that their small masses are hardly compatible with the scenarios of large structures formations [37]. As no other particle of the SM can be candidate for DM, the existence of the latter is a second strong hint of physics beyond the SM.

In addition to these arguments, NP is expected to arise at least at the Planck scale $M_{\rm Pl} \approx 10^{19}$ GeV, where gravitational effects cannot be neglected. The SM can therefore be considered as an effective theory, valid up to a yet unknown scale Λ. The introduction of this scale however entails a problem of naturalness known as *hierarchy problem*. Indeed the Higgs boson provides a mass to most of the SM particles, which in turn contribute to its mass *via* quantum corrections. In practice, the Higgs mass m_h is related to its bare mass $\overline{m}$ and the cutoff Λ by the relation [38, 39]:

$$m_h^2 = \overline{m}^2 + \kappa\Lambda^2 \,, \tag{1}$$

where $\kappa \sim 10^{-2}$ is a dimensionless constant [38]. The ratio $\overline{m}^2/\Lambda^2$ is a dimensionless parameter of the ultraviolet (UV) completion of the SM above the scale Λ. Using the mass of what we consider to be the Higgs boson $m_h = 125$ GeV [40] and the scale $\Lambda = M_{\rm Pl}$, we find

$$\frac{\overline{m}^2}{\Lambda^2} = -\kappa \left(1 - \frac{m_h^2}{\kappa\Lambda^2}\right) = -\kappa \left(1 - 10^{-32}\right) . \tag{2}$$

This means that in order to recover the correct Higgs mass, this ratio has to be tuned to the 32nd decimal. This extreme tuning of the UV theory seems completely unnatural and points towards NP around the TeV scale.

The Higgs mechanism and the arguments in favour of NP at the TeV scale motivated the construction of the Large Hadron Collider (LHC). So far however, the direct searches of BSM particles carried out by the LHC experiments have only be able to raise excluded NP for energies $\lesssim 1$ TeV.

Indirect searches

One well known example of the interplay between theory and experiment is the prediction of the charm quark *via* the Glashow-Iliopolus-Maiani (GIM)-mechanism [41] (1970) (see also [42] for a review). The problem at that time was that a Fermi theory based on the three quarks u, d and s, used to describe hadronic processes, was noticeably incomplete. In particular, the calculation of amplitudes with two exchanged weak bosons, *e.g.* for $K_L \to \mu^+\mu^-$ decays or for the $K^0 - \bar{K^0}$ mixing, revealed the presence of quadratic divergences. Regulating these divergences required an energy cutoff Λ that, due to the stringent experimental limits on $K^0 - \bar{K^0}$ mixing, had to be as low as 3 GeV. Adding a fourth quark, coupled to a combination of the d and s quarks, implies that each contribution of the u quark is compensated by a similar contribution, but opposite in sign. The charm quark

not only allowed to reconcile theory predictions and experimental limits by replacing the cutoff Λ by the mass difference $m_c^2 - m_u^2$, but it also paved the way for the identification of the $SU(2)_L$ structure of weak interactions. In return, the GIM-mechanism predicts a scale for the charm mass, influencing the experimental strategy towards the discovery of the J/ψ particle in 1974 [43, 44].

More generally, indirect searches offers an efficient surrogate to direct searches. Loop corrections to suitable processes allow to probe new scales often higher by several orders of magnitude with respect to those attainable by direct searches – at the price of a higher model dependence. Indirect searches have therefore been used extensively for the search of physics beyond the SM, especially in flavour physics. For instance, the very good agreement between SM predictions and experimental measurements of $K^0 - \bar{K}^0$ mixing parameters pushes the limit on the NP scale to around 10^8 GeV if one assumes flavour-universal NP couplings of order one [45].

Flavour Changing Neutral Currents

Among the indirect processes that can be probed at the LHC, those driven by flavour changing neutral currents (FCNC), such as $b \to s$, $b \to d$ or $t \to c$, are particularly interesting thanks to their suppressed amplitude in the SM. Indeed, the only part of the SM Lagrangian that induces a change of flavour is the coupling between a charged fermion current and a $W^\pm$ boson

$$\mathcal{L}_{cc} = -\frac{g'}{\sqrt{2}} V_{ij} \left(\bar{Q}_i \sigma^+ \gamma^\mu Q_j \right) W_\mu^+ + \text{h.c.}\,, \tag{3}$$

where $\sigma^\pm = (\sigma^1 \pm i\sigma^2)/2$ is a combination of Pauli matrices, g' is the weak coupling constant and V is the Cabibbo-Kobayashi-Maskawa (CKM) matrix that translates the misalignment between the up-quark and down-quark mass eigenstates. Making the quark flavours explicit, this unitary matrix reads

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \tag{4}$$

The amplitudes of each entry can be apprehended using the Wolfenstein parametrization [46], that we will use in the following

$$V = \begin{pmatrix} 1 - \lambda^2/2 & \lambda & A\lambda^3(\bar{\rho} - i\bar{\eta}) \\ -\lambda & 1 - \lambda^2/2 & A\lambda^2 \\ A\lambda^3(1 - \bar{\rho} - i\bar{\eta}) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4)\,, \tag{5}$$

where the last fit to data yields [47]

$$\begin{aligned} \lambda &= 0.224837\,^{+0.000251}_{-0.000060}\,, & A &= 0.8235\,^{+0.0056}_{-0.0145}\,, \\ \bar{\rho} &= 0.1569\,^{+0.0102}_{-0.0061}\,, & \bar{\eta} &= 0.3499\,^{+0.0079}_{-0.0065}\,. \end{aligned} \tag{6}$$

A FCNC can however be obtained by a loop process where a W boson couples twice to a quark line. With $b \to s$ for instance, the amplitude would read

$$\mathcal{A}(b \to s) = V_{ts}^* V_{tb}\, F(\hat{m}_t^2) + V_{cs}^* V_{cb}\, F(\hat{m}_t^2) + V_{us}^* V_{ub}\, F(\hat{m}_t^2)\,, \tag{7}$$

where F are loop functions that depends on the mass ratio $\hat{m}_q = m_q/m_W$. Defining $\lambda_i = V_{is}^* V_{ib}$, the unitarity of the CKM matrix implies $\lambda_c = -\lambda_t - \lambda_u$. We therefore have

$$\mathcal{A}(b \to s) = \lambda_t \left(F(\hat{m}_t^2) - F(\hat{m}_c^2)\right) + \lambda_u \left(F(\hat{m}_u^2) - F(\hat{m}_c^2)\right)\,. \tag{8}$$

This amplitude presents several suppressions. First of all, F are loop functions with two couplings to the W and will in general carry a loop factor $(g'/4\pi)^2$. Second, the amplitude is suppressed by a product of CKM factors. In the case of a $b \to s$ transition, one has using eq. (6)

$$\begin{aligned} |\lambda_t| &= A\lambda^2 + \mathcal{O}(\lambda^4) \approx 4 \times 10^{-2}\,, \\ |\lambda_u| &= A\lambda^4 |\bar{\rho} - i\bar{\eta}| + \mathcal{O}(\lambda^4) \approx 8 \times 10^{-4}\,. \end{aligned} \tag{9}$$

Finally, one recovers the GIM mechanism already described, in the fact that this amplitude go to zero for degenerate quark masses. In practice, the corresponding suppression is small for the first term of the right-hand side of eq.(8), but it is much larger for the second term which is further suppressed by the mass difference $\hat{m}_u^2 - \hat{m}_c^2 \sim 3 \times 10^{-4}$. The second term can therefore be neglected for observables not sensitive to the CKM phase.

These suppressions make the experimental measurements of these decays much harder, but are on the other hand very interesting for testing the SM. Indeed, in a SM-suppressed amplitude, a putative BSM contribution can compete with the SM one while it would be negligible in unsuppressed processes. The $b \to s\{\ell\ell, \gamma, g\}$ decays are therefore denoted as *rare decays* with branching ratios that range from *e.g.* 10^{-4} for $\mathcal{B}(B_s^0 \to X\gamma)$ down to 10^{-9} for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. These transitions have therefore been scrutinized by several experiments.

B-anomalies at the end of Run II

B-anomalies currently refer to a group of measurements of $b \to s$ and $b \to c$ transitions, initially performed by LHCb and Belle, and that shows less than perfect agreement with SM. While each of these measurements could be a first sign of NP, statistical fluctuations or underestimated theoretical uncertainties cannot be excluded for the moment. The experimental situation can be described by the following categories that involve different experimental and theoretical challenges.

LFU in $b \to s\ell\ell$ transitions The first and most striking effect was found by LHCb in the ratio known as R_K [48]

$$R_K^{\text{Run I}} = \frac{\mathcal{B}(B \to K\mu\mu)}{\mathcal{B}(B \to Kee)} = 0.75\,^{+0.090}_{-0.074} \pm 0.036\,, \qquad \text{for } 1.1\,\text{GeV}^2 < q^2 < 6\,\text{GeV}^2\,, \tag{10}$$

where the first uncertainty is statistical and the second systematic and q^2 is the squared invariant mass of the dilepton. The SM predicts unity within a few percent [49, 50] which implies a tension between theory and experiment at the level of 2.6σ. The update of the experimental value, including part of LHCb Run II, reads $R_K = 0.846\ ^{+0.060}_{-0.054}\ ^{+0.016}_{-0.014}$ [51]. The mean value is closer to the SM prediction than Run I result, but the reduced experimental uncertainties imply a comparable tension (2.5σ) to the situation before the update.

Furthermore, a similar measurement was performed by LHCb with a different final state [52]

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu\mu)}{\mathcal{B}(B \to K^* ee)} = \begin{cases} 0.66^{+0.11}_{-0.07} \pm 0.03\,, & \text{for } 0.045\,\text{GeV}^2 < q^2 < 1.1\,\text{GeV}^2\,, \\ 0.69^{+0.11}_{-0.07} \pm 0.05\,, & \text{for } 1.1\,\text{GeV}^2 < q^2 < 6\,\text{GeV}^2\,, \end{cases} \tag{11}$$

again in tension with the SM predictions ($R^{\text{SM}}_{K^*}$ approximately unity) by $\sim 2.5\sigma$ in both q^2 bins. This measurement was also performed by Belle at low and high q^2 by averaging over $B^\pm$ and B^0 decays [53]

$$R_{K^*} = \frac{\mathcal{B}(B \to K^* \mu\mu)}{\mathcal{B}(B \to K^* ee)} = \begin{cases} 0.90^{+0.27}_{-0.21} \pm 0.10\,, & \text{for } 0.1\,\text{GeV}^2 < q^2 < 8\,\text{GeV}^2\,, \\ 1.18^{+0.52}_{-0.32} \pm 0.10\,, & \text{for } 15\,\text{GeV}^2 < q^2 < 19\,\text{GeV}^2\,. \end{cases} \tag{12}$$

Given their sizable uncertainties, these values are compatible with both the SM predictions and previous results on R_{K^*} from LHCb [52].

Branching ratio suppression in $b \to s\mu\mu$ transitions Despite the fact that handling electrons is far more complex than handling muons in a hadronic collider, the departure from SM in eq. (10) is rather due to the latter. Indeed, the muon-channel measurement, expected to be experimentally more solid, yields [54, 55]

$$\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)_{[1,6]} = (1.19 \pm 0.03 \pm 0.06) \times 10^{-7}\,, \tag{13}$$

which is about 30% lower than the SM prediction, $\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)^{\text{SM}}_{[1,6]} = (1.75^{+0.60}_{-0.29}) \times 10^{-7}$ [56–58].

The very same pattern, with data lower than the SM prediction, is actually also observed in the $B_s \to \phi\mu^+\mu^-$ channel and in the same range $q^2 \in [1, 6]$ GeV^2, as initially found in 1/fb of LHCb data [59] and then confirmed by a full Run I analysis [60]. This discrepancy is estimated to be more than 3σ.

The uncertainties in these measurements are dominated by the limited knowledge of the B to light meson hadronic form factors [61–63].

Angular analysis of $b \to s\mu\mu$ transitions The study of the $B \to K^*\mu^+\mu^-$ decay, and especially its full angular analysis, also showed peculiar results. The quantity known as P'_5, designed to have reduced sensitivity to form-factor uncertainties [64], exhibits a discrepancy in two bins, again in the low-q^2 range.

This measurement triggered a lot of experimental attention by LHCb [65–67], ATLAS [68], CMS [69] and Belle [70] and the theoretical error is still debated [71–75].

LFU in $b \to c\ell\nu$ transitions Further interesting results come from the long-standing deviations from τ-μ and τ-e universality in $b \to c\ell\nu$ transitions [76–82]

$$R_{D^{(*)}} = \frac{\mathcal{B}(B \to D^{(*)}\tau\nu)}{\mathcal{B}(B \to D^{(*)}\ell\nu)_{\ell=e,\mu}} \,. \tag{14}$$

Uncertainties are dominated by statistics, with non-negligible experimental systematics but small theoretical uncertainties [83–88]. (Note that e-μ universality in $b \to c\ell\nu$ transitions is tested to hold at the percent level [89–91].)

It goes without saying that, given the experimental and theoretical uncertainties, these measurements should be taken with care. The following comments are however in order

1. Despite the fact that the tension between theory and experiment is small for each observable, we will see throughout this part and especially in section 3.1 that a simple description of the data in term of an effective field theory (EFT) leads to a consistent description of all the $b \to s$ anomalies, with a best-fit point that improves the fit to the data by more than five standard deviations compared to the SM [92–97]. Furthermore, it was shown that a simplified model with a single tree-level mediator can not only explain $b \to s$ data, but even all of the aforementionned deviations simultaneously without violating any other existing constraints [98–103].

2. Although these discrepancies triggered most of the work presented in this thesis, we emphasize that our results don't reduce to these measurements. We will see that some of the analyses and methods presented here can be adapted or generalized to other hints of new physics.

Outline

In this experimental and theoretical context, we see two natural paths to unravel the current flavour puzzle. On the one hand, any new observable sensitive or related to $b \to s$ or $b \to c$ transition can bring valuable information for the building of NP models. Performing new measurements of these transitions or precision tests of low-energy observables is thus a priority. On the other hand, the improvement of the theoretical description of these observables as well as their correlations in NP scenarios can turn existing measurements into precious constraints and therefore guide upcoming theoretical and experimental projects.

This thesis is thus divided into two parts.

The first part is devoted to theoretical studies including the definition of new observables, model building and effective theories. In Chapter 1, we will study

the rare and radiative decays $B_{(s)} \to \ell\ell^{(\prime)}\gamma$ and define observables that can be used as probes of the SM and its extensions. In Chapter 2, we will consider the implication of $b \to s$ and $b \to c$ anomalies on model building and we will examine a model based on a symmetry between the two heaviest generations of fermions. Finally, in Chapter 3 we perform fits on the coefficients of the EFTs and discuss the links between the B-anomalies and DM.

The second part presents one example of an experimental analysis, namely the search for $B_s^0 \to \mu^+\mu^-\gamma$ at LHCb. These decays have never been observed and the only existing limit was set on B^0 decays by the Babar collaboration [104]

$$\mathcal{B}(B^0 \to \mu\mu\gamma) < 1.5 \times 10^{-7} \qquad \text{at } 90\% \text{ CL}\,. \tag{15}$$

Although LHCb is probably the only current detector that can perform this analysis, we will see that this measurement is challenged by the modest photon reconstruction of the detector. LHCb and the analysis will be described in Chapter 4. In Chapter 5, we will present the first steps of the signal selection. This selection will be enriched in Chapter 6 thanks to an efficient multi-variate analysis (MVA). In Chapter 7, we will review the backgrounds that pollute the signal window as well as the fit models used to describe them and the signal. The systematic uncertainties of the analysis are reduced by the consideration of a second channel that we use to normalize the yields. A third channel is also used to probe the validity of the event simulation and to measure some selection efficiencies directly on data. These channels and methods are described in Chapter 8. The remaining uncertainties and the expected sensitivity are then given Chapter 9.

Part I

Theory and Phenomenology

1

Between theory and experiment: definition of novel observables

Contents

This chapter is devoted to an in-depth study of the $B_{(s)} \to \ell\ell^{(\prime)}\gamma$ decays and its potential for testing the SM and NP scenarios. The chapter is organized as follows: section 1.1 introduces notations and presents the computation of $B_s^0 \to \ell^+\ell^-\gamma$ branching ratio; section 1.2 discusses the main source of theoretical uncertainties and proposes methods to control this error and make $B_s^0 \to \ell^+\ell^-\gamma$ a meaningful test of NP; section 1.3 proposes a novel method to probe the $B_s^0 \to \mu^+\mu^-\gamma$ branching fraction using $B_s^0 \to \mu^+\mu^-$ decay; and section 1.4 extends this discussion to the lepton flavour violating $B_s^0 \to \ell\ell'\gamma$ decays.

1.1 The $B_s \to \ell\ell\gamma$ decay: rare and radiative

The branching ratio $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is certainly one of the cleanest probes of $b \to s$ transitions. The theoretical description of the decay is very clean because, neglecting QED corrections, all its QCD non-perturbative dynamics is described by a single number that can be precisely estimated on the lattice. The current uncertainties on the branching ratio, mainly due to this quantity and the

CKM parametrization [105], are deemed to decrease with further lattice estimates and experimental measurements. From the experimental point of view, the decay $B_s^0 \to \mu^+\mu^-$ is also particularly clean for many reasons. First of all, muons are, thanks to their penetrating power, the easiest leptons to deal with at colliders. Electrons interact early in the detector leaving therefore only small traces, and taus have to be reconstructed from their children particles which, by lepton number conservation, include invisible particles. The mass of the initial B_s^0 also shows a good compromise as it gives energetic muons that can be extracted from the background and a boosted B_s^0 in the detector frame. These characteristics are crucial for the experimental analysis of this decay both to trigger events and to deal with the large background.

An immediate question is whether these properties can also be found in similar channels. Considering $B \to K\mu\mu$ decays is one possibility. As well-known however, the theoretical description of the $B \to K$ transition is much more involved due to the estimation of form-factors and the presence of resonant contributions. The branching ratio computation, as well as the prediction of angular observables, is therefore plagued by larger theoretical uncertainties. Although the SM yield is about two orders of magnitude larger than the $B_s^0 \to \mu^+\mu^-$ one, the experimental measurement is also tarnished by systematical error due to the physics model used for the simulation of events [55]. As the B energy is shared between the K and the two leptons, the analysis also shows lower trigger and selection efficiencies, making these channels insufficient to clearly bring NP scenarios out.

Another, less studied, possibility consists in adding a photon to the final state of $B_s^0 \to \mu^+\mu^-$ decays. While keeping a purely electro-weak final state, this additional photon lifts the chiral suppression factor, therefore enhancing the total branching ratio. We will show that this enhancement comes together with a richer short-distance structure than the non radiative decay, making $B_s^0 \to \mu^+\mu^-\gamma$ decays a multipurpose tool for testing the SM, particularly interesting in the light of present data. This section is therefore devoted to the calculation of $B_s^0 \to \ell\ell\gamma$ branching ratio.

1.1.1 The Weak effective Hamiltonian

The most general dimension-six Hamiltonian relevant for the transition $b \to s\ell_1^-\ell_2^+$ with $\ell_i = e, \mu, \tau$ reads [106–109]

$$\begin{aligned}\mathcal{H}_{\text{eff}} = \frac{4G_F}{\sqrt{2}}\Bigg(&\sum_{i=1}^{2} C_i(\mu)(\lambda_u \mathcal{O}_i^u(\mu) + \lambda_c \mathcal{O}_i^c(\mu)) - \lambda_t \sum_{i=3}^{6} C_i(\mu)\mathcal{O}_i(\mu) \\ &- \lambda_t \sum_{i=7}^{10,S,P} \left(C_i(\mu)\mathcal{O}_i(\mu) + C_i'(\mu)\mathcal{O}_i'(\mu)\right)\Bigg) + \text{h.c.}\,,\end{aligned} \tag{1.1}$$

where $\lambda_i \equiv V_{is}^* V_{ib}$, with V_{ij} CKM matrix elements, and C_i are the Wilson coefficients. The operators that will be relevant in the following are defined as

$$\begin{aligned}
\mathcal{O}_1^q &= (\bar{s}_i \gamma_\mu q_{Lj})(\bar{q}_j \gamma^\mu b_{Li}) \,, & \mathcal{O}_2^q &= (\bar{s}_i \gamma_\mu q_{Li})(\bar{q}_j \gamma^\mu b_{Lj}) \,, \\
\mathcal{O}_7 &= \frac{e}{16\pi^2} m_b (\bar{s} \sigma_{\mu\nu} b_R) F^{\mu\nu} \,, & \mathcal{O}_8 &= \frac{g_s}{16\pi^2} m_b (\bar{s} \sigma_{\mu\nu} b_R) G^{\mu\nu} \,, \\
\mathcal{O}_9^{\ell_1\ell_2} &= \frac{e^2}{16\pi^2} (\bar{s} \gamma_\mu b_L)(\bar{\ell}_1 \gamma^\mu \ell_2) \,, & \mathcal{O}_{10}^{\ell_1\ell_2} &= \frac{e^2}{16\pi^2} (\bar{s} \gamma_\mu b_L)(\bar{\ell}_1 \gamma^\mu \gamma_5 \ell_2) \,, \\
\mathcal{O}_S^{\ell_1\ell_2} &= \frac{e^2}{16\pi^2} m_b (\bar{s}\, b_R)(\bar{\ell}_1 \ell_2) \,, & \mathcal{O}_P^{\ell_1\ell_2} &= \frac{e^2}{16\pi^2} m_b (\bar{s}\, b_R)(\bar{\ell}_1 \gamma_5 \ell_2) \,,
\end{aligned} \tag{1.2}$$

where i, j are color indices and $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$. The primed operators are obtained by the replacements $\{L \leftrightarrow R,\ m_b \to m_s\}$[1]. The operators associated to QCD penguin diagrams, $\mathcal{O}_{3,\ldots,6}$, are not shown as it is usual to absorb their effect by the redefinition [110, 111]

$$C_7^{\text{eff}} = C_7 - \frac{1}{3} C_3 - \frac{4}{9} C_4 - \frac{20}{3} C_5 - \frac{80}{9} C_6 \,, \tag{1.3}$$

$$C_8^{\text{eff}} = C_8 + C_3 - \frac{1}{6} C_4 + 20 C_5 - \frac{10}{3} C_6 \,, \tag{1.4}$$

$$C_9^{\text{eff}} = C_9 + Y(q^2) \,, \tag{1.5}$$

where q^2 is the squared dilepton mass and $Y(q^2)$ is given in Appendix A.1. In the following, we will split Wilson coefficients as $C_i = C_i^{\text{SM}} + \delta C_i$ so that the SM limit [106–108] for the Wilson coefficients is obtained by $\delta C_i^{(\prime)} = 0$. To lighten the notation, we will usually drop the "eff" superscript and write C_i for δC_i in the absence of possible confusion.

Before developing the computation of $\mathcal{B}(B_s^0 \to \ell\ell\gamma)$, we would like to emphasize two points.

- It should be kept in mind that the Hamiltonian (1.1) assumes that the SM is an appropriate description of the phenomenology at low scales. In particular the shifts in the Wilson coefficients can only encode new physics arising above the weak scale. In this context, the measurement of R_{K^*} in the lowest bin eq. (11) is rather surprising, as e.m.-dipole operators, which are the dominant ones in this region, are necessarily lepton-universal. It should certainly be kept in mind that, due to the proximity to the kinematical threshold, a robust error assessment may be more delicate in this region [50], but the discrepancy, if confirmed, would require light new physics not describable within our effective-theory approach (see e.g. [112–115]).

[1] The sign conventions for the electromagnetic and strong couplings of $\mathcal{O}_{7,8}$ are consistent with the covariant derivative $D_\mu = \partial_\mu + ie\, Q_f A_\mu + i g_s G_\mu$ (e.g. $Q_\mu = Q_e = -1$) and $C_{7,8}^{\text{SM}} < 0$.

- We can slightly anticipate on section 3.1 and comment the implication of experimental data on the NP shift to Wilson coefficients. It is now well known that all $b \to s$ anomalies described in the introduction can be accounted for by either of the two simple following scenarios [92–97].

  1. A shift to $C_9^{\mu\mu}$ only of the order of $30\%\, C_{9,\mathrm{SM}}^{\mu\mu}$;
  2. A shift in the direction $C_9^{\mu\mu} = -C_{10}^{\mu\mu}$ of approximatively $15\%\, C_{9,\mathrm{SM}}^{\mu\mu}$.

  These scenarios will be reappraised in section 3.1, but it is clear that such large contributions to SM processes make any new observable sensitive to these operators crucial for the understanding of this puzzle.

1.1.2 $B_s \to \ell\ell\gamma$ amplitude

As an example of our setup, let us first focus on the case $\ell_1 = \ell_2 = \ell$ and compute the $B_s^0 \to \ell^+\ell^-$ branching ratio. As every Wilson coefficients in this section refers to $\ell\ell$, we will omit this superscript. The computation of the amplitude $\langle \ell\ell | (-\mathcal{H}_{\mathrm{eff}}) | \bar{B}_s^0 \rangle$ lead to *a priori* unknown hadronic matrix elements. These matrix elements cannot be computed perturbatively, but using Lorentz and gauge invariance, one can reduce their dependence to

$$\langle 0 | \bar{s}\, \gamma^\mu \gamma_5\, b | \bar{B}_s^0(p) \rangle = i p^\mu f_{B_s^0}\,, \tag{1.6}$$

where p is the B_s^0 four-momentum and $f_{B_s^0}$ is known as the B_s^0 decay constant. This constant is computed using numerical simulations of QCD on the lattice (see Table 1.2). Using the equations of motion, one also gets

$$\langle 0 | \bar{s}\, \gamma_5\, b | \bar{B}_s^0(p) \rangle = -i \frac{M_{B_s^0} f_{B_s^0}}{m_b + m_s}\,. \tag{1.7}$$

The $B_s^0 \to \ell^+\ell^-$ branching ratio stems from these definitions, and reads [109, 116]

$$\begin{aligned} \mathcal{B}(B_s^0 \to \ell^+\ell^-) = \tau_{B_s^0} \frac{G_F^2 \alpha_{\mathrm{em}}^2 M_{B_s^0}^3 f_{B_s^0}^2}{2^6 \pi^3} \sqrt{1 - 4\hat{m}_\ell^2}\, |\lambda_t|^2 \times \\ \left[|F_P + 2\hat{m}_\ell C_{10}|^2 + (1 - 4\hat{m}_\ell^2) |F_S|^2 \right]\,, \end{aligned} \tag{1.8}$$

with $\hat{m}_\ell = m_\ell / M_{B_s^0}$ and

$$F_{S,P} = M_{B_s^0} \frac{m_b C_{S,P} - m_s C'_{S,P}}{m_b + m_s}\,. \tag{1.9}$$

We note that the above branching-ratio formula refers to 'instantaneous' B_s^0 decays. This observation is relevant for a precision calculation of B_s decay branching ratios. In this case, one should replace τ_{B_s} with τ_{B_H} (where B_H is the heaviest of the $B_s - \bar{B}_s$ mass eigenstates), to account for the large width difference in the

B_s system [117]. Numerically, one finds that the SM value is dominated by the operator $\mathcal{O}_{10}$ contribution. Including electroweak corrections, the SM prediction for $\ell = \mu$ reads [105, 116, 118][2]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\rm SM} = (3.66 \pm 0.14) \times 10^{-9}\,, \tag{1.10}$$

where the 4% uncertainty is mainly due to the uncertainty on $f_{B_s^0}$, the CKM entries and the top pole mass (see eq. (8)). Experimentally, the latest branching ratio measurements average to [120][3]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{\rm exp} = (2.69^{+0.37}_{-0.35}) \times 10^{-9}\,. \tag{1.11}$$

The tension between the experimental and the theoretical prediction will be discussed in section 3.1. It is however already clear that $\lesssim \mathcal{O}(15\%)$ new contributions to C_{10} are allowed by present errors, and actually favoured – provided they are in destructive interference with the SM contribution – by the about 25% too low central value. On the other hand, the absence of helicity suppression in the scalar and pseudoscalar terms of eq. (1.8) and the overall agreement between eqs. (1.10) and (1.11) forces the contributions of $F_{S,P}$ to negligible values [122, 123].

Let us now consider the branching fraction of $B_s^0 \to \ell^+\ell^-\gamma$. We will denote $q = p_1 + p_2$ the dilepton momentum and $k = p - q$ the photon momentum. The computation of the amplitude, defined as

$$\mathcal{A} \equiv \langle \ell^+(p_1)\ell^-(p_2)\gamma(k,\varepsilon)|(-\mathcal{H}_{\rm eff})|\bar{B}_s^0\rangle\,, \tag{1.12}$$

brings four new matrix elements that need to be estimated. Using again Lorentz and gauge invariance, and following [124], we define[4]

$$\begin{aligned}
\langle\gamma(k,\varepsilon)|\bar{s}\,\gamma^\mu(1\pm\gamma_5)\,b|\bar{B}_s^0\rangle M_{B_s^0} &= -\,e\,\{P_\perp^\mu\,V_\perp(q^2) \qquad \pm P_\parallel^\mu\,V_\parallel(q^2)\}\,,\\
\langle\gamma(k,\alpha)|\bar{s}\,iq_\nu\sigma^{\mu\nu}(1\mp\gamma_5)\,b|\bar{B}_s^0\rangle &= +\,e\,\{P_\perp^{\mu\alpha}\,T_\perp(q^2,k^2) \pm P_\parallel^{\mu\alpha}\,T_\parallel(q^2,k^2)\}\,,
\end{aligned} \tag{1.13}$$

with

$$P_\perp^{\mu\alpha} = \varepsilon^{\mu\alpha\rho\sigma}p_\rho k_\sigma\,, \quad P_\parallel^{\mu\alpha} = i\,(g^{\mu\alpha}(p\cdot k) - p^\alpha k^\mu)\,, \tag{1.14}$$

where the $\varepsilon_{0123} = 1$ convention is assumed and $P_\perp^\mu \equiv \varepsilon^*_\alpha P_\perp^{\mu\alpha}$ and analogous for the $\parallel$-direction. In practice for $T_{\perp,\parallel}(q^2,k^2)$ either q^2 or k^2 will be zero because of the

[2] The result we quote is obtained with the latest estimation of $f_{B_s^0}$ for $N_f = 2+1+1$ [119].

[3] The average eq. (1.11) is very close to the average found in [121].

[4] This notation translates into the one of Ref. [125,126] as $V_\perp = F_V$, $V_\parallel = F_A$, $T_\perp = F_{TV}$, $T_\parallel = F_{TA}$. One reason for introducing this notation is to make contact with the $B \to V\ell\ell$ literature, where V and A labels refer to the polarisation of the leptons in the effective theory language. We also note that the sign of the form-factors depends on the sign convention of the covariant derivative. Our covariant derivative convention, specified above, is consistent with all form-factors being positive, which can be inferred from Refs. [127,128] in the context of $B_u \to \ell\nu\gamma$ transitions.

on-shell photon in the final state. The difference in signs with respect to [125] is due to the ε tensor and the covariant derivative definitions.

Computing $V_{\perp,\parallel}(q^2)$ and $T_{\perp,\parallel}(q^2,k^2)$ is a complex task as these functions encapsulate the non-perturbativity of the $b\to s$ transition. On the other hand, based on exact and approximate results developed in Appendix A.2, one can assume a generic parametrization

$$\begin{aligned} V_{\perp,\parallel}(q^2) &= \beta^V_{\perp,\parallel}\,\frac{f_B M_B}{\Delta^V_{\perp,\parallel}+E_\gamma}\,,\\ T_{\perp,\parallel}(q^2,0) &= \beta^T_{\perp,\parallel}\,\frac{f_B M_B}{\Delta^T_{\perp,\parallel}+E_\gamma}\,, \end{aligned} \tag{1.15}$$

where the photon energy can be written as

$$E_\gamma = \frac{M_B}{2}(1-\frac{q^2}{M_B^2})\,. \tag{1.16}$$

The parameters β and Δ used in this manuscript are summarized in Table 1.1. The treatment of $T_{\perp,\parallel}(0,k^2)$ is complicated by the presence of meson states at

Parameter	$V_\perp$	$V_\parallel$	$T_\perp$	$T_\parallel$
β [GeV^{-1}]	0.28	0.30	0.26	0.33
Δ [GeV]	0.04	0.04	0.30	0.30

Table 1.1: $B\to\gamma$ form factor parameters for the parametrization eq. (1.15), computed in [125] and used in this manuscript.

very low-q^2. In section 1.2, we will see how these resonances can be accounted for under the narrow-width approximation.

The Hamiltonian eq. (1.1) entails two types of contributions in the amplitude $\mathcal{A}$. Most of the interactions are driven by the operators $\mathcal{O}^{(\prime)}_{7,9,10}$ and are of short distance (SD) nature. The operators $\mathcal{O}_{1-6,8}$ however give rise to q^2-dependent long distance (LD) contributions. The different SD contributions to the leading-order amplitude in weak interactions are displayed in Fig. 1.1.

The corresponding SD amplitude reads

$$\begin{aligned} \mathcal{A}_{\rm SD} = \; & -\frac{G_F\,\lambda_t}{\sqrt{2}}\frac{e\,\alpha_{\rm em}}{2\pi}\times\\ & \Big\{\frac{2m_b}{q^2}\Big((C_7+\frac{m_s}{m_b}C_7')T_\perp(q^2)P_\perp^\mu-(C_7-\frac{m_s}{m_b}C_7')T_\parallel(q^2)P_\parallel^\mu\Big)\,\bar u(p_2)\gamma_\mu v(p_1)+\\ & \frac{1}{M_{B_s^0}}\Big((C_9+C_9')V_\perp(q^2)P_\perp^\mu-(C_9-C_9')V_\parallel(q^2)P_\parallel^\mu\Big)\,\bar u(p_2)\gamma_\mu v(p_1)+ \end{aligned}$$

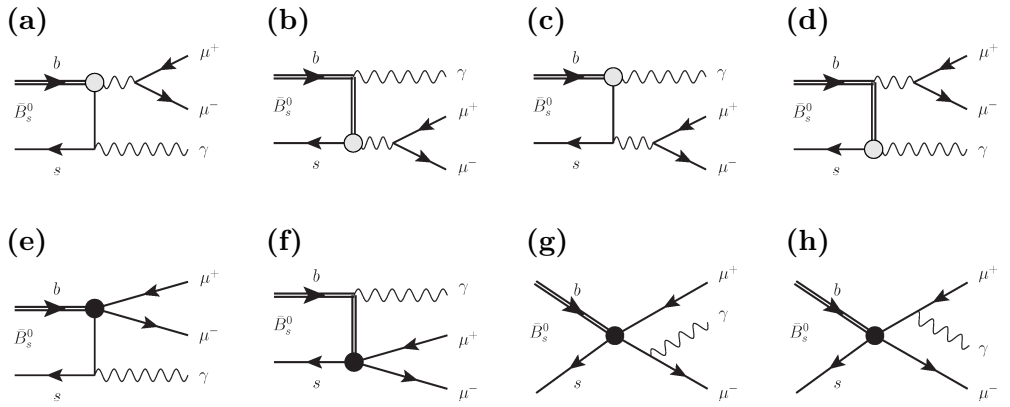

Figure 1.1: Short-distance diagrams contributing to the $\bar{B}_s^0 \to \mu^+\mu^-\gamma$ process to lowest order. The black and the grey circles denote the insertion of the four-fermion operators $\mathcal{O}_{9,10}^{(\prime)}$ and respectively of $\mathcal{O}_7^{(\prime)}$. The form factors $T_{\perp,\parallel}(q^2,0)$ and $T_{\perp,\parallel}(0,q^2)$ describe the diagrams (a,b) and (c,d) respectively. Diagrams (e) and (f) are described by the $V_{\perp,\parallel}(q^2)$ form factors and diagrams (g) and (h) encode bremsstrahlung contributions, whose hadronic matrix elements are described by the $\bar{B}_s^0$ decay constant.

$$\frac{1}{M_{B_s^0}}\left((C_{10}+C'_{10})V_\perp(q^2)P_\perp^\mu-(C_{10}-C'_{10})V_\parallel(q^2)P_\parallel^\mu\right)\bar{u}(p_2)\gamma_\mu\gamma_5 v(p_1)- \\ 2if_{B_s}m_\mu\left(C_{10}-C'_{10}\right)\bar{u}(p_2)\left(\frac{\not{\epsilon}^*\not{p}}{t-m_\mu^2}-\frac{\not{p}\not{\epsilon}^*}{u-m_\mu^2}\right)\gamma_5 v(p_1)\Big\}, \tag{1.17}$$

where

$$T_{\perp,\parallel}(q^2)\equiv T_{\perp,\parallel}(q^2,0)+T_{\perp,\parallel}(0,q^2)\,, \tag{1.18}$$

take into account diagrams (a,b) and (c,d) in Fig. (1.1).

As the LD contributions involve the same structure as the SD contribution, we parametrize the full amplitude (1.12) as

$$\mathcal{A}=\frac{G_F\,\lambda_t}{\sqrt{2}}\frac{e\,\alpha_{\rm em}}{2\pi}\times \\ \left(\frac{2m_b}{q^2}\left(\mathfrak{a}_\perp(q^2)P_\perp^\mu-\mathfrak{a}_\parallel(q^2)P_\parallel^\mu\right)\bar{u}(p_2)\gamma_\mu v(p_1)+\mathcal{O}(C_{9,10}^{(\prime)})\right), \tag{1.19}$$

such that

$$\mathfrak{a}_{\perp,\parallel}(q^2)=(C_7\pm\frac{m_s}{m_b}C'_7)T_{\perp,\parallel}(q^2)+(C_8\pm\frac{m_s}{m_b}C'_8)G_{\perp,\parallel}(q^2)+\sum_{i=1}^{6}C_iL_{i\perp,\parallel}(q^2)\,, \tag{1.20}$$

where $G_{\perp,\parallel}(q^2)$ and $L_{i\perp,\parallel}(q^2)$ stand for the chromomagnetic penguin and the four-quark operator contributions. These contributions, displayed in Fig. 1.2, are mainly relevant at low q^2 due to the $1/q^2$ enhancement from the virtual photon emitting the lepton pair.

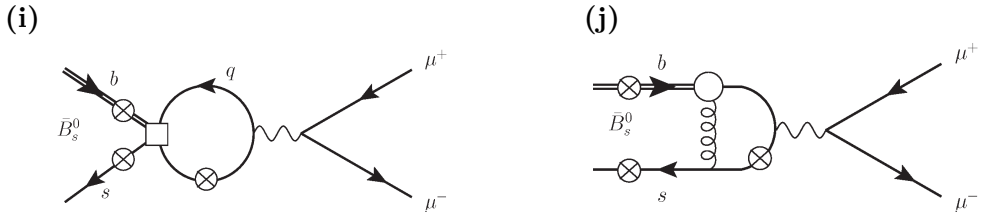

Figure 1.2: Long-distance diagrams contributing to the $\bar{B}_s^0 \to \mu^+\mu^-\gamma$ process to lowest order. The empty square or circle denote the insertion of one of the four-quark operators in eq. (1.1) and, respectively, of the operator $\mathcal{O}_8$. The symbol $\otimes$ denotes all the possible ways of attaching an on-shell photon, with exclusion of bremsstrahlung.

The four-quark contribution (diagram **(i)** in Fig. 1.2 and referred to as weak annihilation (WA) in [126]) has been computed at low q^2, at leading twist-2 and in the limit of massless up- and charm-quarks. The full quark mass dependence can however be obtained from [129] where this contribution was computed in the context of the $B \to \gamma\gamma$ decay. Parallel contributions are found to vanish, while

$$L_{1\perp} \;=\; -\frac{8}{3}\frac{f_{B_s}}{m_b}\frac{1}{\lambda_t}\left(\lambda_u g(z_u) + \lambda_c g(z_c)\right) \;, \quad L_{2\perp} = \frac{1}{3}L_{1\perp} \;, \tag{1.21}$$

with $z_i \equiv m_i^2/m_b^2$ and the $g(z)$ function defined in [129].[5] The contribution from 4-quark operators other than $\mathcal{O}_{1,2}$ ("quark loops") are neglected in the following because of their small Wilson coefficients.
Concerning the chromomagnetic penguin operator (diagram **(j)** in Fig. 1.2), only little is known, but it is expected to be small, as it is the case for $B \to V\ell\ell$ [130], in part because of, again, the small Wilson coefficient.

From eq. (1.17), it is already clear that adding a photon to the final state of $B_s^0 \to \ell^+\ell^-$ decays yields an observable not only sensitive to $\mathcal{O}_{10}$, but also to $\mathcal{O}_9$ and the electromagnetic-dipole operator $\mathcal{O}_7$, as well as to their chirality-flipped counterparts [125, 126, 131–134]. Let us now turn to the computation of the differential branching ratio.

1.1.3 $B_s \to \ell\ell\gamma$ branching ratio

The branching ratio can be decomposed into initial state radiation (ISR) – diagrams **(a)** to **(f)** in Fig. 1.1 – and final state radiations (FSR), also referred to as bremsstrahlung, that corresponds to the two last diagrams. Denoting $\Gamma^{(1)}$ (respectively $\Gamma^{(2)}$) the ISR (FSR) component and $\Gamma^{(12)}$ the interference between the two contributions, the total branching ratio reads $\Gamma = \Gamma^{(1)} + \Gamma^{(2)} + \Gamma^{(12)}$. Using $x \equiv 1 - \hat{s}$ and $\xi \equiv \hat{u} - \hat{t}$, where from now on the hat denotes that the given

[5] Note that the definitions of $\mathcal{O}_{1,2}$ in [129] are interchanged with respect to our notation in eq. (1.2), which follows [106–108].

variable has been made dimensionless by normalizing it to an appropriate power of M_B, the different contributions read

$$\frac{d^2\Gamma^{(1)}}{d\hat{s}\,d\hat{t}} = \frac{G_F^2\alpha_{\rm em}^3 M_{B_s^0}^5}{2^{10}\pi^4}|\lambda_t|^2\left[x^2 B_0 + x\,\xi B_1 + \xi^2 B_2\right] , \tag{1.22}$$

$$\frac{d^2\Gamma^{(2)}}{d\hat{s}\,d\hat{t}} = \frac{G_F^2\alpha_{\rm em}^3 M_{B_s^0}^5}{2^{10}\pi^4}|\lambda_t|^2\left(\frac{8\,f_{B_s^0}}{M_{B_s^0}}\right)^2 \hat{m}_\ell^2\,|C_{10}|^2 \times \left[\frac{\hat{s}+x^2/2}{(\hat{t}-\hat{m}_\ell^2)(\hat{u}-\hat{m}_\ell^2)} - \frac{x^2\,\hat{m}_\ell^2}{(\hat{t}-\hat{m}_\ell^2)^2(\hat{u}-\hat{m}_\ell^2)^2}\right] , \tag{1.23}$$

$$\frac{d^2\Gamma^{(12)}}{d\hat{s}\,d\hat{t}} = \frac{G_F^2\alpha_{\rm em}^3 M_{B_s^0}^5}{2^{10}\pi^4}|\lambda_t|^2\left(\frac{16\,f_{B_s^0}}{M_{B_s^0}}\right)\frac{x^2\,\hat{m}_\ell^2}{(\hat{t}-\hat{m}_\ell^2)(\hat{u}-\hat{m}_\ell^2)} \times \left[\frac{2\,x\hat{m}_b}{\hat{s}}{\rm Re}\left(C_{10}^* C_7 T_\perp(s,0)\right) + x V_\perp(s){\rm Re}\left(C_{10}^* C_9\right) + \xi V_\parallel(s)\,|C_{10}|^2\right] , \tag{1.24}$$

where

$$\hat{s} = \frac{(p-k)^2}{M_{B_s^0}^2} , \quad \hat{t} = \frac{(p-p_1)^2}{M_{B_s^0}^2} , \quad \hat{u} = \frac{(p-p_2)^2}{M_{B_s^0}^2} , \tag{1.25}$$

and the $\hat{s}, \hat{t}, \hat{u}$ variables fulfil the constraint $\hat{s}+\hat{t}+\hat{u} = 1 + 2\,\hat{m}_\ell^2$.

The B_i functions are defined as follows:

$$B_0 = (\hat{s}+4\hat{m}_\ell^2)\,(F_1(\hat{s}) + F_2(\hat{s})) - 8\,\hat{m}_\ell^2\,|C_{10}|^2\left(V_\perp^2(q^2) + V_\parallel^2(q^2)\right) ,$$

$$B_1 = 8\left[\hat{s}\,V_\perp(q^2)V_\parallel(q^2)\,{\rm Re}\left(C_{10}^* C_9\right) + \hat{m}_b\,V_\perp(q^2)\,{\rm Re}\left(C_{10}^* C_7 T_\parallel(q^2)\right) + \hat{m}_b\,V_\parallel(q^2)\,{\rm Re}\left(C_{10}^* C_7 T_\perp(q^2)\right)\right] ,$$

$$B_2 = \hat{s}\,(F_1(\hat{s}) + F_2(\hat{s})) ,$$

with

$$F_1(\hat{s}) = \left(|C_9|^2 + |C_{10}|^2\right)V_\perp^2(q^2) + \left(\frac{2\,\hat{m}_b}{\hat{s}}\right)^2\left|C_7 T_\perp(q^2)\right|^2 + \frac{4\,\hat{m}_b}{\hat{s}}V_\perp(q^2)\,{\rm Re}\left(C_9^* C_7 T_\perp(q^2)\right) , \tag{1.26}$$

$$F_2(\hat{s}) = \left(|C_9|^2 + |C_{10}|^2\right)V_\parallel^2(q^2) + \left(\frac{2\,\hat{m}_b}{\hat{s}}\right)^2\left|C_7 T_\parallel(q^2)\right|^2 + \frac{4\,\hat{m}_b}{\hat{s}}V_\parallel(q^2)\,{\rm Re}\left(C_9^* C_7 T_\parallel(q^2)\right) . \tag{1.27}$$

The q^2 spectrum of the $B_s^0 \to \mu^+\mu^-\gamma$ decay is obtained after integrating over the reduced Mandelstam variable $\hat{t}$. It is presented in Fig. 1.3, where we used the numerical values summarized in Table 1.2. One remark is in order concerning

the q^2 integration range. While $q^2_{\text{min}} = 4m_\ell^2$ is fixed by kinematics, the total branching ratio is logarithmically divergent at $E_\gamma = 0$ (namely $q^2 = M^2_{B_s^0}$). This divergence reflects the fact that the observable $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$, with one single soft photon, is not well-defined in this limit. The problem is well-known and already appears in the definition $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [135–137]. Due to bremsstrahlung emissions, the physical observable is indeed $\mathcal{B}(B_s^0 \to \mu^+\mu^- + n\gamma)$, where the n photons are undetected. These emissions can be summed analytically to all orders in the soft-photon approximation, yielding a multiplicative correction to the non-radiative rate [138]. The separation between the $B_s^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-\gamma$ spectrums in this region is discussed in section 1.3.

As concerns the measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$, the experiment cannot measure photons with energies smaller than a given threshold. In the B_s^0 rest frame, this energy reads

$$E_\gamma = \frac{M_{B_s^0}}{2}\left(1 - \frac{q^2}{M_{B_s^0}}\right) . \tag{1.28}$$

Imposing $E_\gamma \geq E_{\gamma,\text{min}} = 50$ MeV, as performed for the simulation of events in the analysis implies $q^2_{\text{max}} = M^2_{B_s^0}(1 - 2E_{\gamma,\text{min}}/M_{B_s^0})$[6]. In practice, we will see in section 5 that the experimental analysis (with a detected photon) requires a much larger photon-energy cut whose efficiency is estimated on the simulation.

Constant	Ref.	Value	Constant	Ref.	Value
$C_{1,\text{SM}}$	[106–108]	-0.278	$C_{2,\text{SM}}$	[106–108]	1.123
$C_{9,\text{SM}}$	[106–108]	4.07	$C_{10,\text{SM}}$	[106–108]	-4.31
$C_{7,\text{SM}}$	[106–108]	-0.29	G_F [GeV^{-2}]	[40]	$1.663787(6) \times 10^{-5}$
$M_{B_s^0}$ [GeV]	[40]	5.36688(14)	m_Z [GeV]	[40]	91.1876(21)
τ_{B_H} [ps]	[40]	1.620(7)	$\alpha_{\text{em}}(m_Z)$	[40]	1/127.952(9)
$f_{B_s^0}$ [GeV]	[119]	0.2303(13)	f_ϕ [GeV]	[139]	0.241(18)
$\lvert V_{tb}V_{ts}^*/V_{cb}\rvert$	[140]	0.982(1)	$\lvert V_{cb}\rvert_{\text{incl}}$	[141]	0.04200(64)

Table 1.2: Values used for the numerical evaluation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$. Wilson coefficients are estimated at $\mu_b = 5.0$ GeV. α_{em} is run from m_Z down to μ_b in the $\overline{\text{MS}}$ scheme (RGEs can be found *e.g.* in Ref. [105]).

[6] The value of 50 MeV is fixed by the q^2 resolution of LHCb at the $M_{B_s^0}$ scale. This number would be larger (~ 80 MeV) in *e.g.* CMS.

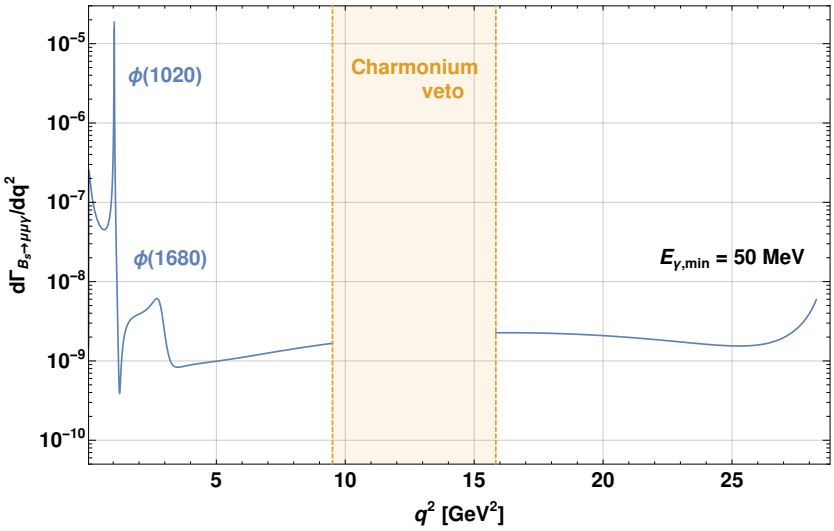

Figure 1.3: $B_s^0 \to \mu^+\mu^-\gamma$ differential branching ratio as computed in [126] and presented in the text. Details on the ϕ resonances will be given in section 1.2. The charmonium region is excluded and a minimal photon energy of $E_{\gamma,\min} = 50$ MeV is considered. Theoretical uncertainties, mainly due to the ϕ resonances and the $B \to \gamma$ form factors parametrization are discussed in section 1.2. Additional details on $E_{\gamma,\min}$ and the treatment of FSR are discussed in section 1.3

1.1.4 Scan of the branching ratio

To summarize this section, the q^2-spectrum can be described as follows:

1. At very low-q^2, near the kinematic limit, the branching ratio is enhanced by $1/q^2$ contributions (photon pole). Contributions from diquark and gluonic operators are subdominant due to small Wilson coefficients, such that the decay can be reduced to $B_s^0 \to \gamma\gamma$, with one photon converting into the dilepton pair.

2. When q^2 enters the GeV^2 region, light meson resonances appear. The subprocesses $B_s^0 \to \gamma\phi^{(\prime)}(\to \mu\mu)$ are dominated by the e.m.-dipole interactions and therefore offer a sensitive probe of the C_7 and C_7' Wilson coefficients. These resonances significantly enhance the total branching ratio. For example, the q^2 interval between 1 and 1.1 GeV^2 contributes 70% of the total branching ratio (excluding the narrow charmonium interval $[8.6, 15.8]$ GeV^2). The treatment of these resonances will be discussed in the following section 1.2.

3. Above these light-meson resonances and below the charmonium contribution, the region is dominated by $\mathcal{O}_9$ and $\mathcal{O}_{10}$. The region is theoretically clean because $E_\gamma \gg \Lambda_{\rm QCD}$ allows form factors to be computed in the framework of QCD factorization and soft-collinear effective theory (SCET). This approach

initiated in the context of $B_u \to \gamma\ell\nu$ [127, 142–144], was recently extended to $B_q \to \ell\ell\gamma$ decays [145].

4. The charmonium region is excluded both for experimental and theoretical reasons. Experimentally, this region comes with a large combinatorial background. For example $B_s^0 \to J/\psi\,\eta$ would make the search of signal in this region a tremendous task [146]. This region is also highly challenging theoretically as the branching ratio cannot be computed pertubatively nor on the lattice.[7] The size of the veto window, which is a compromise between charmonium pollution and signal yield, is discussed in Part II.

5. Above the $\psi(2S)$, the impact of broad charmonium resonances is much weaker. The region is dominated by $\mathcal{O}_9$ and $\mathcal{O}_{10}$ which defines another theoretically clean region. Form factor uncertainties are smaller and they can in principle be extracted from lattice QCD simulations [149]. The sensitivity to their parametrization can be further reduced by measuring a ratio, akin to $R_{K^{(*)}}$, of two different leptonic final states. This possibility is discussed in section 1.2.

6. At very large $q^2 > 25\ \text{GeV}^2$, the decay is dominated by bremsstrahlung radiations. These corrections are experimentally simulated and accounted for in the $B_s^0 \to \mu^+\mu^-$ measurement. Actually $B_s^0 \to \mu^+\mu^-\gamma$ ISR component can be seen as a pollution of the $B_s^0 \to \mu^+\mu^-$ sample. This viewpoint offers an alternative way of measuring the radiative branching ratio. This possibility is discussed in section 1.3.

In order to simplify the discussion, several points were left apart in the computation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$. The next section comes back on these assumptions at low-q^2, with the discussion of meson resonances, as well as at high-q^2, with the study of the impact of the form-factor parametrization.

1.2 Theoretical error

This section, adapted from [124], is devoted to a more in-depth consideration of the low- and high-q^2 regions and of the main theoretical uncertainties involved. The idea is to put forward a number of strategies to reduce these uncertainties below the level that makes $B_{(s)} \to \ell\ell\gamma$ observables a valuable new probe of the very interactions hinted at by present-day discrepancies.

As already mentioned in the previous section, the low-q^2 range includes the $\phi(1020)$ resonance. We argue that the theoretical uncertainty associated to the prediction for the *total* $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ can be drastically reduced taking into account, for low q^2, its close parenthood with the measured $B_s^0 \to \phi(\to K^+K^-)\gamma$.

[7] On the other hand, the strong phases of the resonances may be used to measure the phases of the NP Wilson coefficients via CP-dependent observables [147, 148].

Incidentally, a measurement of this ratio in the low-q^2 region would provide a cross-check of the R_{K^*} result (11) in the lowest bin.

As concerns the high-q^2 region, which is only dominated by the vector and the axial $B_s^0 \to \gamma$ form factors, theory uncertainties cancel to a large extent in the ratio between two different lepton channels. Keeping in mind that this region is by far dominated by the Wilson coefficients C_9 and C_{10}, such ratio provides a new stringent test of LUV.

1.2.1 The low-q^2 region

Modelling the ϕ and higher resonances

As already discussed, the low-q^2 region is dominated by light meson resonances. The underlying processes, mainly $B_s^0 \to (\phi \to \mu\mu)\gamma$ for B_s^0, are dominated by the e.m.-dipole interactions in the SM and therefore offers a sensitive probe of the C_7 and C_7' Wilson coefficients as well as of other interactions, mediated by light new particles, and as such beyond the effective-theory picture.

For this part of the amplitude, which is not well described by perturbative theory because of resonant behaviour, we write the reduced amplitude (1.20) in term of a n-times subtracted dispersion relation

$$\begin{aligned} \mathfrak{a}_\iota(q^2) &= \sum_{k=0}^{n-1} \frac{1}{k!} \mathfrak{a}_\iota^{(k)}(s_0)(q^2 - s_0)^k \\ &+ (q^2 - s_0)^n \frac{1}{2\pi i} \int_{\text{cut}} \frac{\text{Disc}[\mathfrak{a}_\iota(s)]ds}{(s - q^2 - i0)(s - s_0)^n} \,, \end{aligned} \tag{1.29}$$

where hereafter $\iota = \perp, \parallel$, $\text{Disc}[f(s)] \equiv f(s + i0) - f(s - i0)$ and it is assumed that the only singularity in the physical sheet of the q^2 plane runs over the real axis.

In the narrow-width approximation

$$\frac{1}{2\pi i}\text{Disc}[\mathfrak{a}_\iota(s)] = -\delta(s - M_\phi^2) M_\phi f_\phi \, \mathfrak{a}_\iota^{B_s^0 \to \phi\gamma} \; + \; \ldots \tag{1.30}$$

where $\langle \phi | \bar{s}\gamma_\mu s | 0 \rangle = f_\phi M_\phi \varepsilon_\mu^*$ is the ϕ decay constant and the dots stand for higher resonance states such as the $\phi(1680)$ and other KK-continuum, discussed later on. $\mathfrak{a}_\iota^{B_s^0 \to \phi\gamma}$ is the analogue of $\mathfrak{a}_\iota$ in eq. (1.20),

$$\mathfrak{a}_{\perp,\parallel}^{B_s^0 \to \phi\gamma} = \left(C_7 \pm \frac{m_s}{m_b} C_7'\right) T_{\perp,\parallel}^{B_s^0 \to \phi}(0) + \left(C_8 \pm \frac{m_s}{m_b} C_8'\right) G_{\perp,\parallel}^{B_s^0 \to \phi}(0) + \sum_{i=1}^{6} C_i L_{i\perp,\parallel}^{B_s^0 \to \phi} \,. \tag{1.31}$$

As already noted below eq. (1.14) for $B_s^0 \to \mu^+\mu^-\gamma$ form-factors, one has $T_\perp^{B_s^0 \to \phi}(0) = T_\parallel^{B_s^0 \to \phi}(0)$. In more standard notation, *e.g.* [61], the form factors are denoted by $T_{\perp,\parallel}^{B_s^0 \to \phi}(0) = 2T_1^{B_s^0 \to \phi}(0) = 2T_2^{B_s^0 \to \phi}(0)$, notation that we will adopt in the following.

For $s_0 = 0$ and a lower bound below M_ϕ^2, eq. (1.29) reads, beyond the narrow width approximation and for zero and one subtraction

$$\mathfrak{a}_\iota(q^2) = \begin{cases} \frac{f_\phi M_\phi \mathfrak{a}_\iota^{B_s^0 \to \phi\gamma}}{q^2 - M_\phi^2 + i M_\phi \Gamma_\phi} + \dots & [n=0] \\ \mathfrak{a}_\iota(0) + \frac{q^2}{M_\phi^2} \frac{f_\phi M_\phi \mathfrak{a}_\iota^{B_s^0 \to \phi\gamma}}{q^2 - M_\phi^2 + i M_\phi \Gamma_\phi} + \dots & [n=1] \end{cases}, \tag{1.32}$$

where Γ_ϕ is the decay width of the ϕ meson.

Concerning the numerical values, $T_1^{B_s^0 \to \phi}(0)$ is known most precisely from light-cone sum rules (LCSR), yielding [61]

$$T_1^{B_s^0 \to \phi}(0) = 0.309 \pm 0.027 \tag{1.33}$$

at twist-4 tree level and twist-3 $O(\alpha_s)$. This result is an update of the analysis [150, 151] with more recent input parameters and a twist-4 tree-level contribution. The $\mathcal{O}_8$ and four quark topologies, that enter the $O(\alpha_s)$ result, are known in the $1/m_b$-limit [152, 153] and in LCSR, relying on an $1/m_b$-expansion from Refs. [130] and [154] respectively. As already discussed, the corresponding Wilson coefficient are expected to be small, and these contributions are neglected.

With the approximation $\mathfrak{a}_{\perp,\parallel}^{B_s^0 \to \phi\gamma} = 2\, C_7\, T_1^{B_s^0 \to \phi}(0)$ and the identification $T_1(0) = -g_+(0)$, the $n = 1$ version of this expansion corresponds to the one proposed in [125].

Using $B_s^0 \to \phi\gamma$ data

It is clear that given the dominance of the ϕ resonances at low-q^2, the error attached to eq. (1.33) will directly translate into a theoretical error on the estimation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$. Hence improving the prediction (1.33) is crucial to make $B_s^0 \to \mu^+\mu^-\gamma$ an efficient probe of the SM.

An alternative strategy is to extract the amplitudes $\mathfrak{a}_{\perp,\parallel}^{B_s^0 \to \phi\gamma}$ from experiment. This approach is promising since the branching ratio [155]

$$\mathcal{B}(B_s^0 \to \phi\gamma) = (3.52 \pm 0.34) \times 10^{-5} \tag{1.34}$$

already known to 10% accuracy will be improved by ongoing analysis. In this measurement, the statistical error is about half of the global uncertainty and is expected to decrease substantially. On the other hand, as any B_s^0 modes, the branching ratio is normalized to a B_d mode, so the systematic uncertainty is limited by the uncertainty on the ratio of the B^0 and B_s^0 hadronisation fractions in pp collisions, f_s/f_d, currently of about 5% [156]. However, even with the current sensitivity, one can expect to extract the amplitudes at the 5% level, to compare

with a theory error which is above 10%.

A caveat lies however in the fact that

$$\mathcal{B}(B_s^0 \to \phi\gamma) \propto \left(|\mathfrak{a}_\perp^{B_s^0\to\phi\gamma}|^2 + |\mathfrak{a}_\parallel^{B_s^0\to\phi\gamma}|^2\right) \tag{1.35}$$

which means that the experimental measurement does not provide enough information for two complex amplitudes. Although the $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio prediction will be improved, the knowledge of the relative phases in the amplitude is crucial to use them as probes of the interference components in the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum. This situation can however be improved through theoretical knowledge and related observables in this channel. For example, further knowledge on the amplitudes may be obtained from direct and time-dependent CP asymmetries [157]. For the latter, LHCb has updated its first measurement, which now reads $\mathcal{A}_\Delta \simeq -0.67^{+0.37}_{-0.41} \pm 0.17$ [158] with a large uncertainty but also with a large deviation from the SM prediction $\mathcal{A}_\Delta \simeq 0.047(28)$ [157].

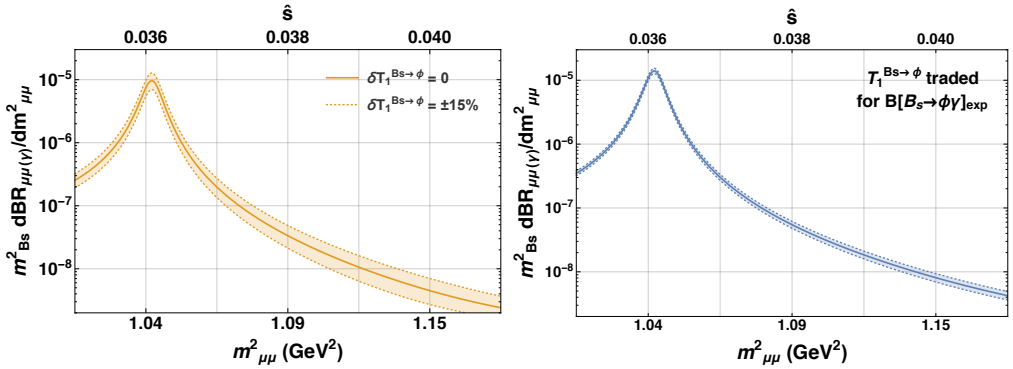

Figure 1.4: Differential branching ratio for $B_s^0 \to \mu^+\mu^-\gamma$ around the $\phi(1020)$ resonance, with (left panel) $T_1^{B_s^0\to\phi}(0)$ from eq. (1.33), and with (right panel) the ϕ-peak region constrained from the *present* experimental measurement of $\mathcal{B}(B_s^0 \to \phi\gamma)$.

In Fig. 1.4 we show the $d\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)/dq^2$ spectrum with $T_1^{B_s^0\to\phi}(0)$ from eq. (1.33) with an error of 15% (left panel), and (right panel) with $T_1^{B_s^0\to\phi}(0)$ traded for eq. (1.34). The 15% error on the left panel is indicative, but can be motivated using a theoretical or an experimental argument. On the one hand, the form-factor error in eq. (1.33) is around 10% and the error on the LD part is likewise around 10%. One can therefore understand the 15% as a Gaussian average. Alternatively, the central value in eq. (1.33) implies $\mathcal{B}(B_s^0 \to \phi\gamma)_{\text{form factor}} = 2.7 \times 10^{-5}$, which is about 30% lower than the central value in eq. (1.34), again justifying an error of about 15%. As concerns the systematic error inherent in the choice of the form-factor parametrization eq. (1.15), it can be estimated in this energy range by a comparison with the results obtained with SCET approaches [145]. Note however

that given the dominance of the ϕ resonances, this systematic uncertainty will be small compare to the one discussed here. The reduced uncertainty in the right panel shows the potential gain of the method, which has further potential for improvement with more statistics. On the systematic side, one needs to go beyond the form-factor dominance which is the underlying assumption that we made.

Additional resonances

In order to make the discussion following eq. (1.30) more transparent, we have restricted ourselves to the case of one single resonance, the $\phi(1020)$. At this stage it is very important to assess the potential impact on the prediction of the low-q^2 $B_s^0 \to \mu^+\mu^-\gamma$ spectrum of other resonances – most notably the $\phi(1680)$, to be denoted as ϕ' hereafter. The $B_s^0 \to \phi'$ form factor can be estimated by scaling by the decay constants,

$$T_1^{B_s^0\to\phi'} \simeq f_{\phi'}/f_\phi\, T_1^{B_s^0\to\phi} \,.$$

As we are unaware of an evaluation of $f_{\phi'}$ we resort to K^*-meson data, assuming $f_{\phi'}/f_\phi \simeq f_{K^{*\prime}}/f_{K^*}$. From the ratio between $\mathcal{B}(B \to K^*(1410)\gamma)$ and $\mathcal{B}(B \to K^*(892)\gamma)$ data [155], suitably corrected for the relevant kinematic factors, we get $f_{\phi'}/f_\phi \simeq 0.86$. We note that this value is encouragingly close to $f_{\rho'}/f_\rho = 0.875$ from Ref. [159] using non-local condensate sum rules.

For such a potentially large coupling, it is clear that including or not the ϕ' would considerably alter the prediction of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum at low q^2, hence of the $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio as a whole. However, the large width of the ϕ' ($\Gamma_{\phi'} \simeq 35 \times \Gamma_\phi$ [155]) turns out to suppress the ϕ' contribution to the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum to be a below-1% correction to the total branching ratio. Needless to say, our argument can put on more solid grounds with data on $\mathcal{B}(B_s^0 \to \phi'\gamma)$, to be measured in a statistically favourable ϕ' decay mode, for example $\phi' \to K\bar{K}^*(892)$. Such data are not yet available at present [155] and we would like to emphasize their interest, not only to robustly assess the systematics due to the ϕ', but also, potentially, for interference studies. We remark in fact that a large phase in the ϕ' Breit-Wigner would entirely cancel the suppression due to $\Gamma_{\phi'} \ll \Gamma_\phi$.

A final remark concerns the charmonium region. Attempting a description of this region with an approach similar to eq. (1.32) is, in principle, possible, because the J/ψ and $\psi(2S)$ resonances are sufficiently narrow. On the other hand, the required radiative branching ratios are, again, not yet measured (see *e.g.* [146] for LHCb's Run I attempt). We also remark that, at variance with the low-q^2 range, in this region the SD dynamics is dominated by the $\mathcal{O}_{9,10}^{(\prime)}$ operators, that one can more cleanly extract from the region $\hat{s} > 0.55$ to be discussed next.

1.2.2 The high-q^2 region

The ratio r_γ

We next consider the part of the spectrum above the narrow-charmonium resonances, $\hat{s} \gtrsim 0.55$. As concerns the theoretical error in this region, the first consideration to be made is that the by far largest contributions come from just two sets of terms, those proportional to $V_\perp^2$ or to $V_\parallel^2$, because of $C_{9,10}$-dominance, followed by the impact of broad charmonium resonances which we address in a later section. Terms proportional to all other form-factor combinations have an impact that numerically does not exceed a few percent. Furthermore the existing theoretical predictions of $V_\perp$ and $V_\parallel$, as well as their associated errors, are partly correlated.

From these considerations it is clear that the ratio of the $B_s^0 \to \ell^+\ell^-\gamma$ differential branching ratios between two different lepton channels offers a potentially much cleaner quantity than the two branching ratios considered separately. As such, this ratio provides a valuable test of lepton universality violation, in a channel devoid of final-state hadrons. More specifically, let us consider the following quantity

$$r_\gamma(q^2) \equiv \frac{d\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)/dq^2}{d\mathcal{B}(B_s^0 \to e^+e^-\gamma)/dq^2} , \tag{1.36}$$

as well as

$$R_\gamma(q_1^2, q_2^2) \equiv \frac{\int_{q_1^2}^{q_2^2} dq^2 \; d\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)/dq^2}{\int_{q_1^2}^{q_2^2} dq^2 \; d\mathcal{B}(B_s^0 \to e^+e^-\gamma)/dq^2} , \tag{1.37}$$

where we choose $q_1^2/M_{B_s}^2 = 0.55$ (corresponding to $q_1^2 = 15.8$ GeV^2), i.e. somewhat above the $\psi(2S)$ resonance, and $q_2^2/M_{B_s}^2 = 0.8$ ($q_2^2 = 23.0$ GeV^2) due to bremsstrahlung in the $\mu^+\mu^-$ channel, as explained below. The ratio $r_\gamma(q^2)$ has the following properties:

- Among the Wilson coefficients appearing in the Hamiltonian (1.1), the largest SM contributions are those from $C_{9,10}$, and the largest sensitivity is correspondingly to $C_{9,10}^{(\prime)}$. The ratio r_γ therefore offers a further test of the very same new-physics contributions that would be responsible for $R_{K^{(*)}}$.

- The radiative branching ratios for the $\mu^+\mu^-$ and for the e^+e^- channels appearing in r_γ are very close to each other, and not hierarchically different, as in the corresponding non-radiative decays. In fact, either of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ and $\mathcal{B}(B_s^0 \to e^+e^-\gamma)$, integrated over the whole q^2 range, are of the order of 10^{-8}. We note explicitly that such rate, in the case of the e^+e^- channel, amounts to an enhancement over the non-radiative branching ratio of about 5 orders of magnitude.

- As mentioned above, both numerator and denominator on the r.h.s. of eq. (1.36) are dominated by terms proportional to $V_\perp^2$ or $V_\parallel^2$ for $\hat{s} \in [0.55, 0.8]$.

An error of, say, $\pm 10\%$ on these form factors thus reflects in roughly twice the same error on the differential branching ratios. This is illustrated in the top panel of Fig. 1.5. Such spread, shown as a blue area, is too large to clearly resolve the $C_{9,10}$ shift required by the $b \to s$ experimental results. The effect of the latter shift is displayed by the red line in the same figure, and as shown, this line lies barely outside the blue area. On the other hand, form-factor uncertainties cancel to a large extent in r_γ. In fact, the r_γ variation due to these uncertainties is suppressed by powers of the difference $(m_\mu^2 - m_e^2)/m_{B_s}^2$. The residual theoretical uncertainty amounts to a relative error on r_γ of at most 5%,[8] well below the size of the shifts to $C_{9,10}^{(\prime)}$ required by $R_{K^{(*)}}$. This point is illustrated in the bottom panel of Fig. 1.5. In this plot the red line lies well outside the blue band of the theoretical error throughout the considered q^2 range.

- We do not consider $\hat{s}$ values above 0.8. In fact, for such values the FSR component – diagrams **(g)** and **(h)** in Fig. 1.1 – becomes, in the $\mu^+\mu^-$ channel, comparable in size with the ISR one – all the other diagrams in the same figure.[9] Being ISR diagrams proportional to just the first of the matrix elements in eq. (1.13), they steadily spoil the cancellation of form-factor uncertainties between the numerator and denominator of eq. (1.37) as q^2 increases. This can be appreciated in Fig. 1.5 (bottom panel), where from left to right r_γ gradually departs from unity and its error gets larger. Incidentally, this departure from unity distinguishes $B_s^0 \to \ell^+\ell^-\gamma$ decays from $B \to V\ell\ell$ ones.

Impact of broad charmonium

In the discussion so far we have disregarded one further source of potentially significant theory systematics, namely the contamination of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum by broad-charmonium resonances. A dedicated study in the context of $B^+ \to K^+\ell\ell$ has been performed in [73], and extended by LHCb to include low-lying $\rho, \omega, \ldots$ resonances [160]. Similar effects are possible for our decays of interest, through the subprocess $B_s^0 \to V_{c\bar{c}}(\to \ell\ell)\gamma$, with $V_{c\bar{c}}$ any of $\psi(2S)$, $\psi(3770)$, $\psi(4040)$, $\psi(4160)$ or $\psi(4415)$. We model LD effects associated with such

[8] The error depends on the degree of correlation between the form-factor errors. For example, the case of $V_\perp$ and $V_\parallel$ errors exactly anti-correlated obviously amounts to an additional cancellation – between the coefficients of the $V_\perp^2$ and $V_\parallel^2$ terms within each of the two branching ratios in R_γ. The figure displays the least favourable case, and as such the blue area represents the envelope of any realistic theoretical error on the form factors.

[9] In the e^+e^- channel the ISR component stays negligible up to q^2 very close to the endpoint, because of chiral suppression. The relative size of the ISR and FSR components will be further discussed in the next sessions (see *e.g.* Fig. 1.6).

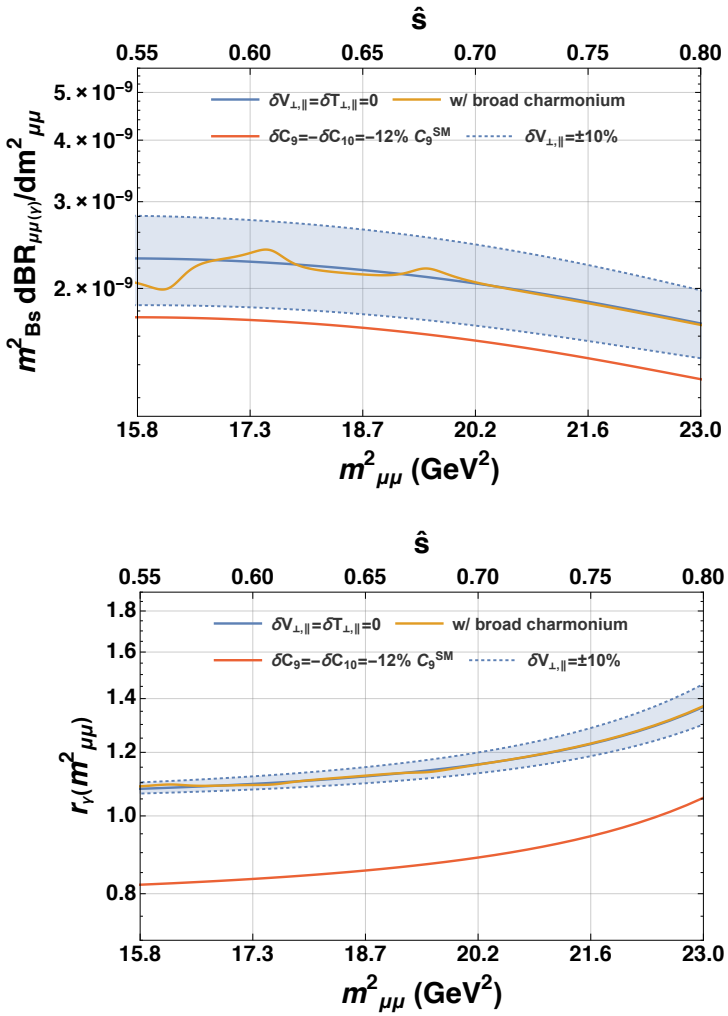

Figure 1.5: Comparison between (top panel) the theoretical error in the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum and (bottom panel) the corresponding error on r_γ as defined in eq. (1.36). The red line corresponds to a NP scenario that account for the $b \to s$ experimental results.

resonances as a sum over Breit-Wigner poles [161], through the replacement

$$C_9 \;\to\; C_9 \;-\; \frac{9\pi}{\alpha^2}\,\bar{C}\sum_V |\eta_V| e^{i\delta_V}\,\frac{\hat{m}_V\,\mathcal{B}(V\to\mu^+\mu^-)\,\hat{\Gamma}^V_{\rm tot}}{\hat{q}^2-\hat{m}_V^2+i\hat{m}_V\hat{\Gamma}^V_{\rm tot}}\;, \tag{1.38}$$

with free floating absolute value and phase [73] to measure the deviation from naive factorization ($|\eta_V| = 1$ and $\delta_V = 0$).[10] The sum runs over the five resonances mentioned above, hats indicate that the given quantity is made dimensionless by an appropriate power of M_{B_s}, and $\bar{C} = C_1 + C_2/3 + C_3 + C_4/3 + C_5 + C_6/3$. The

[10] In Ref. [73], for $B \to K\mu^+\mu^-$ it was found that $|\eta_V| \simeq 2.5$ and $\delta_V \simeq \pi$ gives a good description of the data.

relevant numerical input for all resonances but the $\psi(2S)$ is taken from the recent determination [162]. For the $\psi(2S)$ we use data from the PDG [155] and checked the stability of our results against numerical input taken from [73,161]. The effect of the shift (1.38) is shown in either panel of Fig. 1.5 as a wiggly solid line using $|\eta_V| = 1$ and $\delta_V = 0$ for illustrative purposes. The figure displays that, while the differential branching ratio (top panel) shows some sensitivity to such effect, that may partly compensate the new-physics shift required by $R_{K^{(*)}}$, this sensitivity is substantially reduced in r_γ (bottom panel), that neatly distinguishes the SM case from the new-physics one.

1.2.3 Result

We conclude with predictions of the total branching ratio for $B_s^0 \to \mu^+\mu^-\gamma$ for low and high q^2

$$\begin{aligned} \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)_{\text{low}\,q^2} &= (8.4 \pm 1.3) \times 10^{-9} \,, \\ \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)_{\text{high}\,q^2} &= (8.90 \pm 0.98) \times 10^{-10} \,, \end{aligned} \tag{1.39}$$

where the q^2-integration windows are respectively $q^2/M_{B_s}^2 = [(2m_\mu/M_{B_s})^2, 0.30]$ and $q^2/M_{B_s}^2 = [0.55, 1 - 2E_{\gamma,\text{min}}/M_{B_s}]$, with $E_{\gamma,\text{min}} = 50$ MeV. These ranges correspond to $q^2 = [0.04, 8.64]$ GeV^2 and $q^2 = [15.84, 28.27]$ GeV^2.

For the ratio R_γ we obtain

$$R_\gamma(q_1^2, q_2^2) = 1.152 \pm 0.030 \,, \tag{1.40}$$

where, as reminder, we choose $q_1^2/M_{B_s}^2 = 0.55$ and $q_2^2/M_{B_s}^2 = 0.8$. The errors on the above predictions are obtained by assuming form factors with uncorrelated Gaussian errors of 10% and $\sqrt{(|A_\perp^{B_s^0\to\phi}|^2 + |A_\parallel^{B_s^0\to\phi}|^2)}$ traded for eq. (1.34).

In our above estimates we neglect the possible contribution from the ϕ', having estimated it to be below 1%. On the other hand, we do include possible systematic effects due to the J/ψ (low q^2) or to broad-charmonium resonances (high q^2). These effects are modelled according to eq. (1.38), where we take the resonance couplings to be uniformly distributed in the ranges $|\eta_V| \in [1, 3]$ and $\delta_V \in [0, 2\pi]$, and uncorrelated with one another. The possible pollution from the J/ψ resonance is actually the reason why we limited the low-q^2 prediction to $\hat{s} \leq 0.30$.[11]

[11] For reference, taking the low-q^2 range to extend up to $\hat{s} \leq 0.33$, as chosen elsewhere in the literature, we find

$$\begin{aligned} \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)_{\hat{s}\leq 0.33} &= (8.3 \pm 1.3) \times 10^{-9} \qquad (\text{no } J/\psi) \\ \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)_{\hat{s}\leq 0.33} &= (10.6 \pm 2.9) \times 10^{-9} \qquad (\text{with } J/\psi) \,, \end{aligned} \tag{1.41}$$

where in the first equation we take $|\eta_{J/\psi}| = 0$ in eq. (1.38). Whereas the J/ψ effect is sizeable if the integration range extends up to $\hat{s} = 0.33$, the impact for $\hat{s} \leq 0.30$ is well within our quoted uncertainties, as can be inferred from the error bands in eqs. (1.41) and (1.39).

Before concluding this section, an important remark is in order. All the above discussion concerns B_s^0 decays, however up to a few points, the results can directly be generalized to B^0. These decays would statistically suffer from the relative CKM suppression, of about 4×10^{-2}, with respect to the B_s^0 modes, *but* in principle enormously benefit from the huge statistics and detector capabilities foreseen at Belle 2, although this can be ascertained only through a dedicated study. From the theoretical point of view, many of the considerations we made for the B_s^0 modes can be extended to the B^0 ones by adjusting only a few points. Apart from the exchange of masses, decay constant, CKM factors and lifetime, these points include

- $\phi^{(\prime)}$ resonances should be replaced by ρ, $\omega\ldots$ resonances [126]. In this perspective, we emphasize the importance of $B^0 \to \rho\gamma$ and $B^0 \to \omega\gamma$ experimental measurements that can be crucial to reduce the theoretical error of the branching ratio prediction.

- Experimental data can be adapted in the parametrization of $B_s^0 \to \gamma$ and $B^0 \to \gamma$ form factors. However it appears that these two sets of form factors only differ by a few percent and we will considered them equal in the following.

In this section we proposed methods to reduce the theoretical uncertainty on $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$. We studied two q^2 regions, below and above charmonium resonances and provided predictions in these ranges. Although the uncertainty on these observables are still large, they are sufficiently under control to make $B_s^0 \to \mu^+\mu^-\gamma$ a viable probe of NP. In the next section, we will discuss the parenthood between $B_s^0 \to \mu^+\mu^-\gamma$ and $B_s^0 \to \mu^+\mu^-$ at very large q^2 and propose a novel method to experimentally probe this decay.

1.3 Novel measurement strategy

As we will see in Part II, the measurement of radiative hadron decays is usually more challenging than their non-radiative counterparts. This is due to various reasons. First of all, the detection and reconstruction efficiency of a photon is typically smaller than the one of charged tracks. Secondly, the energy being shared with the additional photon, the other children particles are softer, yielding smaller trigger and reconstruction efficiencies. Moreover, the invariant mass reconstructed in decays containing neutrals has, at these energies, a much poorer resolution than in decays containing only charged tracks. By comparing the results obtained in Part II to the one of the $B_s^0 \to \mu^+\mu^-$ analysis [163], we see *e.g.* that considering a photon enlarges the signal width by a factor 4 to 5. This in turn leads to a larger background under the signal peak. The above considerations hold in particular for hadron-collider experiments, due to the high occupancy of typical events, and

for low-energy processes such as those of interest in flavour physics. Despite these difficulties, rare radiative decays with branching ratios of order $10^{-6} - 10^{-7}$ have been observed and exploited for NP searches by several experiments (see [164] for a recent review). However, the rates just mentioned are still very 'abundant' if compared to the $B_s^0 \to \mu^+\mu^-$ decay and its radiative counterpart.

This section, adapted from [165], is devoted to an alternative method to search for $B_s^0 \to \mu^+\mu^-\gamma$ events. Basically, the method consists in measuring $B_s^0 \to \mu^+\mu^-\gamma$ as "contamination" to $B_s^0 \to \mu^+\mu^-$, by suitably enlarging downward the signal window. This possibility requires a number of qualifications, since the $B_s^0 \to \mu^+\mu^-$ measurement itself comes with some subtleties as far as photons are concerned – notably the treatment of soft final-state radiation.

1.3.1 The radiative tail of $B_s^0 \to \mu^+\mu^-$

In an idealised measurement, the $B_s^0 \to \mu^+\mu^-$ decay appears as a peak in the invariant mass squared of the two final-state muons, with negligible intrinsic width.[12] As discussed above eq. (1.28), the 'definition' of the final-state muons is complicated by the fact that they emit soft bremsstrahlung, giving rise to $B_s^0 \to \mu^+\mu^- + n\gamma$ decays, with the n photons undetected. This contribution skews downwards the peak region of the $B_s^0 \to \mu^+\mu^-$ distribution, as shown by the dotted orange curve of Fig. 1.6.

In order to compare the measured $B_s^0 \to \mu^+\mu^-$ rate with the theoretical one (1.8), the mentioned soft-radiation tail due to $B_s^0 \to \mu^+\mu^- + n\gamma$ needs to be accounted for. For example, a $B_s^0 \to \mu^+\mu^-$ signal window extending down to about 5.3 GeV is equivalent to a single-photon energy cut $E_\gamma \simeq 20 - 100$ MeV, amounting to a negative shift of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ as large as 15% [138]. Experimentally, the radiative tail is obtained and taken into account using Monte Carlo $B_s^0 \to \mu^+\mu^-$ events with full detector simulation and with bremsstrahlung photon emission modelled through the PHOTOS application [167]. The advantage of this approach over the analytic one [138] is that the correction factor is already adjusted for detector efficiencies.

For softer and softer photons (or equivalently for q^2 closer and closer to the B_s^0 peak region), the single-photon component in $\mathcal{B}(B_s^0 \to \mu^+\mu^- + n\gamma)$ is expected to match the radiative branching ratio $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$, as computed in section 1.1.3 to leading order in $\alpha_{\rm em}$. This is indeed the case, as shown by comparing the solid blue distribution with the dotted orange one in Fig. 1.6. We can actually go further in this comparison by separating the FSR and ISR contributions. This separation makes sense to the extent that we can identify two regions in q^2 where only one of the two contributions is dominant. The breakup of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum into its different components is likewise reported in Fig. 1.6. As well known, the

[12] The experimental resolution in the muon momenta gives this peak an approximately Gaussian shape, the width being for example of about 25 MeV for the LHCb experiment and ranges from 32 to 75 MeV for the CMS experiment [166].

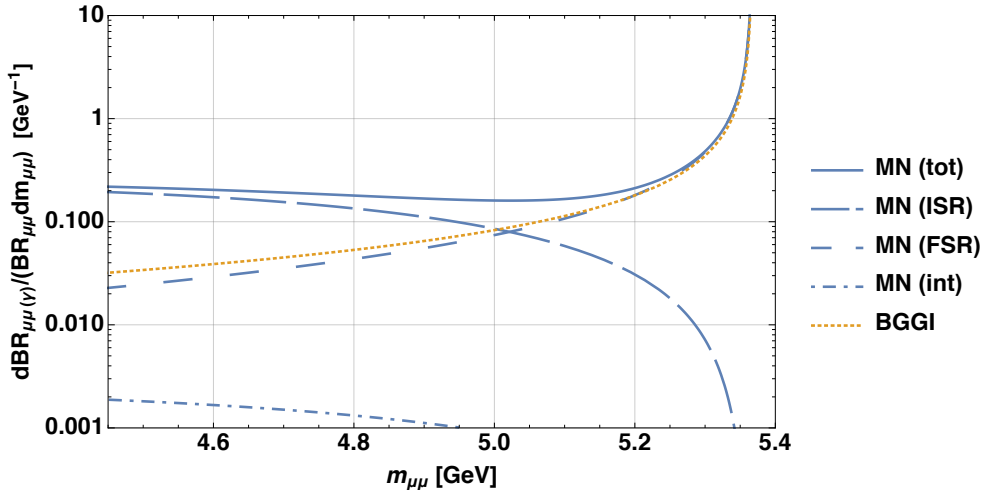

Figure 1.6: Breakup of the full $B_s^0 \to \mu^+\mu^-\gamma$ spectrum (solid blue) – calculated in section 1.1.3, denoted as MN in the legend – into its pure ISR component (long-dashed blue), FSR one (medium-dashed blue), and ISR-FSR interference (dot-dashed blue). We also report the $B_s^0 \to \mu^+\mu^- + n\gamma$ spectrum in the soft-photon approximation (dotted orange) from Ref. [138], denoted as BGGI in the legend.

FSR contribution is dominant for soft photons (or high q^2), whereas the ISR one dominates for harder and harder photons, namely as q^2 decreases from the peak region. The crossover region between the two contributions is at $\sqrt{q^2} \approx 5.0$ GeV. More importantly for our purposes, the contribution from the interference term is always below 1% of the total spectrum. This holds true fairly generally also beyond the SM. In particular, shifts in C_9 and C_{10} with opposite sign with respect to the respective SM contributions, as hinted at by the recent $b \to s$ discrepancies mentioned in the introduction, tend to decrease the interference term even further. As a consequence, the ISR and FSR contributions can be treated as two basically independent spectra.

1.3.2 $B_s^0 \to \mu^+\mu^-\gamma$ as a background of $B_s^0 \to \mu^+\mu^-$

In short, to the extent that the FSR contribution can be systematically accounted for, as it is the case for $B_s^0 \to \mu^+\mu^-$ searches, one can measure the ISR component of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum – and thereby the $B_s^0 \to \mu^+\mu^-\gamma$ differential rate – as "contamination" of $B_s^0 \to \mu^+\mu^-$ candidate events as the signal window is enlarged downwards. We note that such contamination is, in principle, already present in existing $B_s^0 \to \mu^+\mu^-$ searches. However, it is negligible in the typical signal window around the $B_s^0 \to \mu^+\mu^-$ peak, and its smooth distribution can be absorbed in other background distributions due, for example, to combinatorial background or partially reconstructed B decays. For this reason it was typically not included as *separate* component in recent $B_s^0 \to \mu^+\mu^-$ decay measurements [163, 168–170].

On the other hand, as the signal window is enlarged downwards, the ISR component of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum becomes sizable. Fig. 1.7 shows in more detail how large this contamination is expected to be. The figure displays the

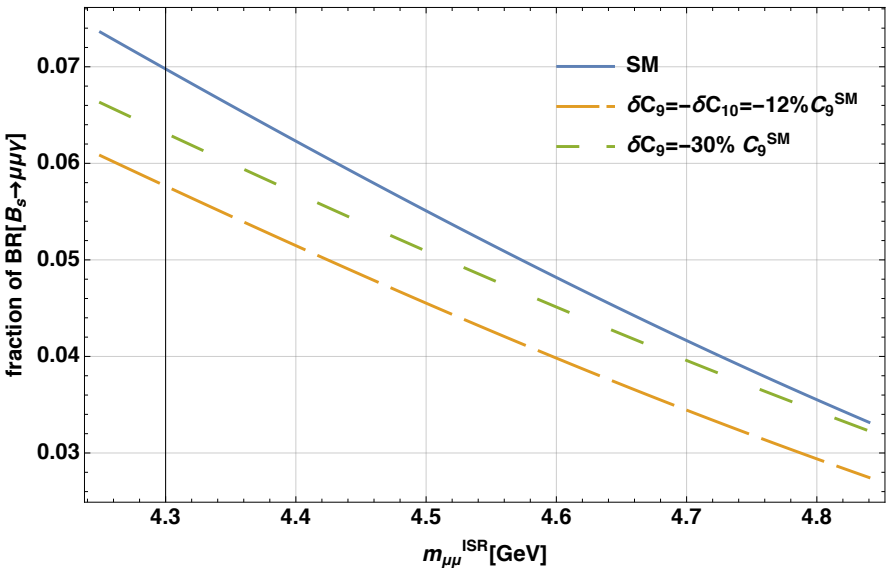

Figure 1.7: Fraction of the full $B_s^0 \to \mu^+\mu^-\gamma$ spectrum as a function of the chosen signal-region lower bound $m_{\mu\mu}^{\rm ISR} = \sqrt{q^2_{\rm min}}$, for the three scenarios discussed in the introduction and specified in the legend. See text for details.

fraction of the full $B_s^0 \to \mu^+\mu^-\gamma$ spectrum as a function of the chosen value for $q^2_{\rm min}$ for the SM case, as well as for the two scenarios that best fit the $b \to s$ anomalies: one with a $V-A$ shift to C_9 and C_{10}, and such that $\delta C_9 = -12\%\, C_9^{\rm SM}$, the other with a C_9-only shift such that $\delta C_9 = -30\%\, C_9^{\rm SM}$. The figure reveals that this fraction is larger within the SM than in the considered NP scenarios. For example, it is about 4.8% in the SM for a $B_s^0 \to \mu^+\mu^-$ signal window extending down to $\sqrt{q^2_{\rm min}} = 4.6$ GeV, whereas it is about 4% in the $V-A$ scenario.

We also note that the associated event yield is large, comparable to that for the $B_s^0 \to \mu^+\mu^-$ signal thanks to the larger branching ratio eq. (1.39). The expected size of the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum is displayed in Fig. 1.8, by superimposing this spectrum of the LHCb $B_s^0 \to \mu^+\mu^-$ analysis of Ref. [168][13]. We show the case of a SM signal as well as the NP case mentioned earlier, namely $\delta C_9 = -\delta C_{10} = -12\%\, C_9^{\rm SM}$. From the absolute size of these curves we can already infer that NP scenarios with the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum enhanced by orders of magnitude with respect to the SM are unlikely in the light of data: as shown in Fig. 1.8, a factor of 10 enhancement would result in a substantial distortion of the measured spectrum

[13] Note that this is not the latest LHCb measurement of $B_s^0 \to \mu^+\mu^-$. In LHCb's 4.4 fb^{-1} measurement [163], $B_s^0 \to \mu^+\mu^-$ events were seen with a significance of almost 8σ, making our argument even stronger.

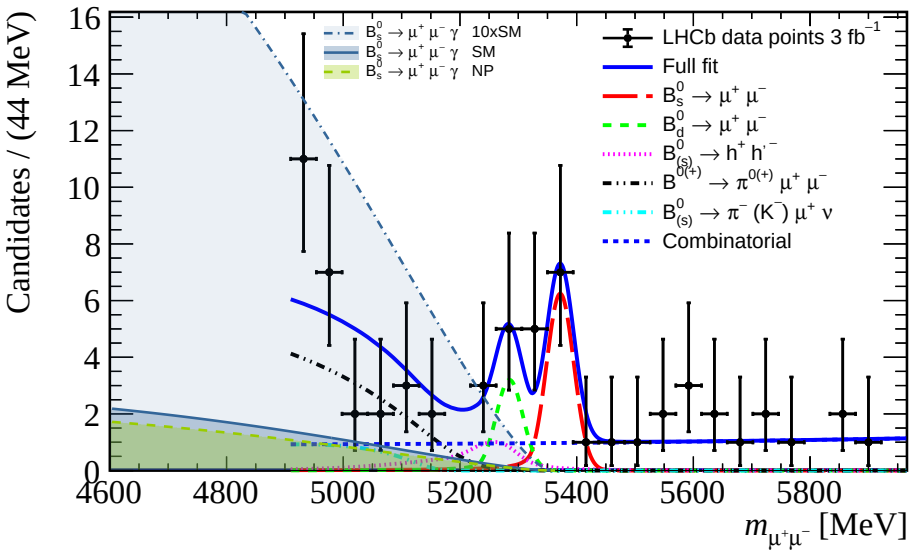

Figure 1.8: Dimuon invariant mass distribution from LHCb's measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [168] overlayed with the contribution expected from $B_s^0 \to \mu^+\mu^-\gamma$ decays (ISR only). Assumes flat efficiency versus $m_{\mu^+\mu^-}$. The line denoted as '$B_s^0 \to \mu^+\mu^-\gamma$ NP' refers to the case with $\delta C_9 = -\delta C_{10} = -12\%\, C_9^{\mathrm{SM}}$ (see also Fig. 1.7). The two filled curves are not stacked onto each other.

from $\sqrt{q^2} \simeq 5.1$ GeV downwards.

The $B_s^0 \to \mu^+\mu^-\gamma$ spectrum shown in Fig. 1.8 is obtained from our theoretical calculation, i.e. it is *not* a fit to existing $B_s^0 \to \mu^+\mu^-$ data nor on signal simulation. Specifically, the spectrum assumes that normalization and efficiency are equal to those of the $B_s^0 \to \mu^+\mu^-$ distribution itself. This is exactly true by definition at the endpoint $q^2 = m_{B_s^0}^2$, and increasingly less so for lower masses, due to the various selection criteria. For example, $B_s^0 \to \mu^+\mu^-$ analysis enforces pointing requirement with respect to the primary interaction vertex (this quantity is referred to as direction angle and will be widely used in Part II), and the latter is less satisfied when an additional undetected photon is present. The actual spectrum will therefore be distorted and will have to be extracted from Monte Carlo simulations of the considered experiment and analysis.

With enough statistics, one can go beyond the integrated $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio, and measure the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum. In fact, shifts to the differential branching ratio are roughly linear in shifts to C_9 or C_{10}. Therefore, for a C_9 or C_{10} deviation of the order of 15% (as hinted at by the global fits to $b \to s$ data), the corresponding variation in the spectrum is expected to be about 15% as well. Barring experimental uncertainties and assuming a Poissonnian distribution, a fit to the data could resolve such shift at one standard deviation for an event yield of about 50.

The above argument is of statistical nature only, *i.e.* it disregards systematic uncertainties. There are two prominent sources of such errors. The first is the theoretical error associated to the $B_s^0 \to \mu^+\mu^-\gamma$ spectrum prediction discussed in the previous sections. The dominant source of uncertainty in this respect is by far the one associated to the $B_s^0 \to \gamma$ vector and axial form factors, defined from the relations (1.13). As described in Fig. 1.7, the level of accuracy of this parametrization is currently not sufficient to clearly resolve the effects expected from new physics. However, what is needed for the proposed method are the form factors in the high-q^2 range close to the kinematic endpoint. This range is, as already discussed, the preferred one for lattice-QCD simulations.

The second potential source of systematic uncertainty for our method is of experimental nature. The impact of this uncertainty depends on the actual possibility to well constrain the other background components populating the signal window as it is enlarged towards lower values. This part of the spectrum, in addition to combinatorial background, consists mainly of the following decays [163]

- semileptonic decays in the form $B \to h^\pm\mu^\mp\nu(+X)$, where h is a pion or kaon misidentified as muon and X can be any other possible hadron (not reconstructed);
- rare decays such as $B^{0,+} \to h^{0,+}\mu^+\mu^-$, which do not need any misidentification;
- decays including charmonium resonances, *e.g.* $B_c^+ \to J/\psi\,\mu^+\nu$, where the signal is obtained with one of the two muons of the J/ψ. These events are usually tackled by vetoing muons that can form a J/ψ with any muon of the event. The mass distribution of the remaining events is quite different from the one expected for $B_s^0 \to \mu^+\mu^-\gamma$ events, so these decays should not be problematic.

The yields of semileptonic decays can be constrained from control channels directly in data so we don't expect them to spoil the fit. Rare decays are potentially more worrisome; as an example the $B^0 \to \pi^0\mu^+\mu^-$ decay is not yet observed experimentally and is currently constrained using the spectral shape measured from the $B^+ \to \pi^+\mu^+\mu^-$ decay and theoretical estimates of the ratio of the two branching fractions [166, 168]. Depending on the pollution of these backgrounds, as measured on simulated samples, some of the $B_s^0 \to \mu^+\mu^-$ analysis parameters can also be adapted. For example more stringent particle identification constraints can diminish the impact of misidentified decays.

1.3.3 Beyond $B_s^0 \to \mu^+\mu^-\gamma$ decays

We emphasize that our proposed method is potentially applicable to several other decays – in principle the radiative counterpart of any two-body decay whereby the initial-state meson mass is completely reconstructible. Straightforward examples

would be all the other $B_q \to \ell^+\ell^-\gamma$ modes, for which the only existing limits concern $B^0 \to e^+e^-\gamma$ or $\mu^+\mu^-\gamma$ with a technique based on explicit photon reconstruction [104]. We note however that for the electron mode, the branching ratio of the non-radiative decay is highly suppressed by the chiral factor, yielding a SM predictions far from current experimental sensitivity. The direct probe of $B^0_{(s)} \to e^+e^-\gamma$ is therefore certainly more promising.

Assuming that the $B^0_s \to \mu^+\mu^-\gamma$ component is under control, the $B^0_s \to \mu^+\mu^-$ tail could also be used to constrain light particles with flavour-violating couplings to b-quarks, such as the QCD axion [171].

In conclusion, we presented a novel method for the extraction of the high q^2 spectrum of the $B^0_s \to \mu^+\mu^-\gamma$ decay. The method avoids the drawbacks of explicit photon reconstruction, and takes advantage of the fact that this spectrum inevitably contaminates the $B^0_s \to \mu^+\mu^-$ event sample as the q^2 signal window is enlarged downward. Fig. 1.8 shows that order-of-magnitude enhancements of the $B^0_s \to \mu^+\mu^-\gamma$ decay rate are unlikely, already in the light of existing data below $\sqrt{q^2} \simeq 5.1$ GeV. More likely, the measurement will involve a dedicated fit by experiments, and this is where our method may make the difference. This method will first be used in LHCb's Run II update of the $B^0_s \to \mu^+\mu^-$, providing the first indirect probe of $B^0_s \to \mu^+\mu^-\gamma$. A dedicated, and more involved, analysis will follow on the same dataset and is expected to provide the first indirect measurement of this branching ratio.

As widely discussed, this method is efficient because the additional photon lifts the chiral suppression factor, hence enlarging the total branching ratio. An immediate question is whether this property can also make the measurement of other very rare decay possible. This is already the case in the electron channel, where the enhancement is as large as 10^5. In the next section, we will show that this is also the case for lepton flavour violating decays such as $B_{(s)} \to \ell\ell'\gamma$.

1.4 $B_{(s)} \to \ell\ell'\gamma$ as a test for New Physics

As already mentioned in the Introduction and developed in section 3.1, current data seem to favour a NP scenario where $C_9^{\mu\mu} = -C_{10}^{\mu\mu}$ i.e. a lepton current that is approximately of the form $V - A$. The maybe simplest model that can account for this shift is obtained assuming that the new physics part of the Hamiltonian eq. (1.1) is due to a third-generation interaction [172]

$$\mathcal{H}_{\rm NP} = G\,(\bar{b}'_L\gamma^\lambda b'_L)\,(\bar{\tau}'_L\gamma_\lambda\tau'_L)\,, \tag{1.42}$$

where $G \ll G_F$ is a new-physics Fermi constant and the primed field are the field in the interaction basis, above electroweak symmetry breaking (EWSB). This effective interaction can arise after integrating out heavy degrees of freedom (Z', leptoquarks...) of any UV theory.

Rotating the primed fields from the interaction basis to the basis of mass-eigenstates involve two rotation matrices

$$b'_L \equiv d'_{L3} = \sum_{i=1}^{3} U^d_{L3i}\, d_{Li}\,, \quad \tau'_L \equiv \ell'_{L3} = \sum_{i=1}^{3} U^\ell_{L3i}\, \ell_{Li}\,. \tag{1.43}$$

For example, using these definitions, the new-physics contribution to $b \to s\mu\mu$ transition reads

$$\mathcal{H}_{\rm NP}(b \to s\mu\mu) = G\, U^{d*}_{L33} U^d_{L32} |U^\ell_{L32}|^2\, (\bar{b}_L \gamma^\lambda s_L)\, (\bar{\mu}_L \gamma_\lambda \mu_L) \,+\, \text{h.c.}\,, \tag{1.44}$$

and the parameters, G and $U^{d,\ell}_{L3i}$ can be estimated from theory and experimental arguments. For instance, the hierarchy of the CKM matrix suggests that $|U^d_{L31}| \ll |U^d_{L32}| \ll |U^d_{L33}|$, and the same relation is favoured for the leptons by current data.

In the absence of further assumptions, it is clear from this simplified model that lepton non universality (LNU) also implies non-standard violation of the lepton flavour (LFV). Assuming the interaction (1.42), the amount of LNU pointed to by the B anomalies actually allows to quantify rather generally the expected amount of LFV [172]. In fact, anticipating again on the global fit performed in section 3.1, the ratio between the NP and the SM+NP contributions to $C^{\mu\mu}_{9,10}$ can be estimated to

$$\rho_{\rm NP} \equiv \frac{\delta C^{\mu\mu}_9}{C^{\rm SM}_9 + \delta C^{\mu\mu}_9} = -0.150 \pm 0.23 \tag{1.45}$$

Then

$$\frac{\mathcal{B}(B \to K \ell^\pm_i \ell^\mp_j)}{\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)} \;\simeq\; 2\rho^2_{\rm NP}\, \frac{|U^\ell_{L3i}|^2 |U^\ell_{L3j}|^2}{|U^\ell_{L32}|^4}\,, \tag{1.46}$$

implying

$$\begin{aligned} \mathcal{B}(B \to K \ell^\pm_i \ell^\mp_j) \;&\simeq\; 4.5\%\, \mathcal{B}(B^+ \to K^+ \mu^+ \mu^-)\, \frac{|U^\ell_{L3i}|^2 |U^\ell_{L3j}|^2}{|U^\ell_{L32}|^4} \\ &\simeq\; 1.9 \times 10^{-8}\, \frac{|U^\ell_{L3i}|^2 |U^\ell_{L3j}|^2}{|U^\ell_{L32}|^4}\,, \end{aligned} \tag{1.47}$$

where we used $\mathcal{B}(B^+ \to K^+ \mu^+ \mu^-) \simeq 4.3 \times 10^{-7}$ [55], and neglected phase-space differences across the different possible lepton final states.[14] Eq. (1.47) tells us that LFV $B \to K$ decays are expected to be in the ballpark of 10^{-8} times an unknown factor involving U^ℓ_L matrix entries. In the $\ell_i \ell_j = e\mu$ case, this ratio reads $|U^\ell_{L31}/U^\ell_{L32}| \lesssim 3.7$ [172], implying that the $B \to K\mu e$ rate may be around 10^{-8}, or much less if $|U\ell_{L31}/U^\ell_{L32}| \ll 1$. The latter possibility would suggest U^ℓ_L entries

[14] These differences are of the order of 30% [173, 174], hence they are not important for the present argument.

that decrease in magnitude with the distance from the diagonal. But then one may expect the ratio $|U^\ell_{L33}/U^\ell_{L32}| > 1$, implying a $B \to K\mu\tau$ rate of $\mathcal{O}(10^{-8})$ or above. In short, assuming the interaction (1.42), one can hope that at least one LFV $B \to K$ decay rate be in the ballpark of 10^{-8} [172], which is in principle within reach at LHCb's Run II. An analogous reasoning applies for the purely leptonic modes $B^0_s \to \ell^\pm_i \ell^\mp_j$. Similarly as eq. (1.47) one obtains

$$\mathcal{B}(B^0_s \to \ell^\pm_i \ell^\mp_j) \;\simeq\; 4.5\%\, \mathcal{B}(B^0_s \to \mu^+\mu^-)\, \frac{|U^\ell_{L3i}|^2 |U^\ell_{L3j}|^2}{|U^\ell_{L32}|^4} \,. \tag{1.48}$$

Therefore, purely leptonic LFV decays of the B^0_s may well be within the reach of LHCb Run II, if the U-matrix factor on the r.h.s. is of order unity (or larger!) for at least one LFV mode.[15]

Due to the experimental challenges in the reconstruction of decays including τ, $B^0_s \to e\mu$ is probably, despite the possible U^ℓ_L-matrix suppression, one of the easiest decay to access among the purely leptonic LFV modes. The current experimental status of LFV searches in B decays is the following.

- The full Run I search for $B^+ \to K^+\mu e$ decays now probes the 10^{-9} scale with the 90% CL limits $\mathcal{B}(B^+ \to K^+\mu^+e^-) < 6.4 \times 10^{-9}$ and $\mathcal{B}(B^+ \to K^+\mu^-e^+) < 7.0 \times 10^{-9}$ [176].

- Tauonic final states are very challenging at LHCb. The missing energy due to the unmeasured neutrino(s) makes selection strategies based on the precise reconstruction of the vertices less efficient. The kinematics can be closed by the consideration of the $\tau \to \pi\pi\pi\nu$ final state or, as proposed recently, using the $B^*_{s2} \to B^+K^-$ decay. The second method allowed to set the 90% CL limits $\mathcal{B}(B \to K\mu^-\tau^+)_{\rm exp} < 3.9 \times 10^{-5}$ [177]. This bound is not far from the best limits set by the Babar collaboration $\mathcal{B}(B \to K\mu^+\tau^-)_{\rm exp} < 4.5 \times 10^{-5}$ and $\mathcal{B}(B \to K\mu^-\tau^+)_{\rm exp} < 2.8 \times 10^{-5}$ [178].

- As concerns purely leptonic final states, the decay $B^0_s \to \tau\mu$ also suffers from the τ reconstruction efficiency. The latest 95% CL limit reads $\mathcal{B}(B^0_s \to \tau\mu) < 4.2 \times 10^{-5}$ [179], relatively far from what we can expect from eq. (1.48).

- For $B^0_s \to \mu e$, a full Run I measurement yielded the the 90% CL limit $\mathcal{B}(B^0_s \to \mu e) < 5.4 \times 10^{-9}$ [180]. Note that this limit is even better than the latest limit $\mathcal{B}(B^0_s \to ee) < 9.4 \times 10^{-9}$ [181] thanks to the presence of a muon and of a smaller background.

All these measurements will benefit from the large luminosity and trigger performances foreseen at LHCb Run III, as well as the upcoming results from Belle II.

[15] For a (rough) comparison, we should keep in mind that at Run II the LHCb is expected [175] to provide a first measurement of $\mathcal{B}(B^0 \to \mu^+\mu^-)$, which in the SM is about 3% of $\mathcal{B}(B^0_s \to \mu^+\mu^-)$.

In this context, it is very useful to search for additional decays, that can give access to the same physics, while being comparably (or, hopefully, more) accessible experimentally. In the $B_s^0 \to \mu e$ channel, one such 'proxy' decay is provided by the inclusion of an additional hard photon in the final state. In fact, as in the case of $B_s^0 \to \mu^+\mu^-\gamma$, the additional photon replaces the chiral-suppression factor, of order $\max(m_{\ell_1}, m_{\ell_2})^2/m_{B_s^0}^2$, with a factor of order $\alpha_{\rm em}/\pi$. In the case of the μe channel these two factors are respectively 4×10^{-4} and 2 per mil. The actual enhancement of $\mathcal{B}(B_s^0 \to \mu e\gamma)$ over the non-radiative counterpart needs to be worked out by explicit calculation.

In this section, adapted from [182], we thus compute the decay of a generic pseudo-scalar meson to $\ell_1^+\ell_2^-\gamma$, and study in detail the cases of B_s^0, B^0, K as initial state and of μe as final state. (To fix notation, formulas are given for B_q.) We subsequently compare the resulting radiative rates with the non-radiative ones. We find an $\mathcal{O}(1)$ factor for the B_q cases and at the percent level for the kaon case.

1.4.1 Observables

Eq. (1.42) yields a shift in the Wilson coefficient of the Hamiltonian eq. (1.1) that reads

$$\delta C_9^{\ell_1\ell_2} = -\delta C_{10}^{\ell_1\ell_2} = \frac{G}{2} \frac{U_{L33}^{d*} U_{L3q}^{d} U_{L3\ell_2}^{\ell*} U_{L3\ell_1}^{\ell}}{-\frac{4G_F}{\sqrt{2}} \lambda_t^{(q)} \frac{\alpha_{\rm em}(m_b)}{4\pi}} \,, \tag{1.49}$$

where $\lambda_t^{(q)} = V_{tq}^* V_{tb}$. The four diagrams found to contribute to the process $B_q \to \ell_1^+\ell_2^-\gamma$ are given in Fig. 1.9, where the black dot denotes the insertion of the operators $\mathcal{O}_9^{\ell_1\ell_2}$ or $\mathcal{O}_{10}^{\ell_1\ell_2}$ defined in eq. (1.2).

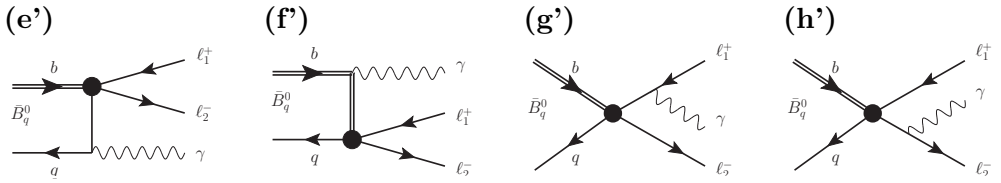

Figure 1.9: Subset of diagrams Fig. 1.1 contributing to $B_q \to \bar{\ell}_1\ell_2\gamma$, within the interaction in eq. (1.42). Here the black dot denotes the insertion of $\mathcal{O}_9^{\ell_1\ell_2}$ or $\mathcal{O}_{10}^{\ell_1\ell_2}$.

The same computation as the one done in sections 1.1.2 and 1.1.3 for the case $\ell_1 = \ell_2$ yields

$$\frac{d^2\Gamma^{(1)}}{d\hat{s}\, d\hat{t}} = \frac{G_F^2 \alpha_{\rm em}^3 M_{B_q}^5}{2^{10}\pi^4} |\lambda_t^{(q)}|^2 \left[x^2 B_0^{(1)} + x\,\xi B_1^{(1)} + \xi^2 B_2^{(1)}\right] \,, \tag{1.50}$$

$$\frac{d^2\Gamma^{(2)}}{d\hat{s}\, d\hat{t}} = \frac{G_F^2 \alpha_{\rm em}^3 M_{B_q}^5}{2^{10}\pi^4} |\lambda_t^{(q)}|^2 \left(\frac{f_{B_q}}{M_{B_q}}\right)^2 \frac{1}{(\hat{t} - \hat{m}_2^2)^2(\hat{u} - \hat{m}_1^2)^2} \times \left[x^2 B_0^{(2)} + x\,\xi B_1^{(2)} + \xi^2 B_2^{(2)}\right] \,, \tag{1.51}$$

$$\frac{d^2\Gamma^{(12)}}{d\hat{s}\,d\hat{t}} = \frac{G_F^2\alpha_{\rm em}^3 M_{B_q}^5 |\lambda_t^{(q)}|^2}{2^{10}\pi^4}\left(\frac{f_{B_q}}{M_{B_q}}\right)\frac{1}{(\hat{t}-\hat{m}_2^2)(\hat{u}-\hat{m}_1^2)}\times \left[x^2 B_0^{(12)} + x\,\xi B_1^{(12)} + \xi^2 B_2^{(12)}\right] , \tag{1.52}$$

with two slight differences with respect to 1.1.3. First, $\hat{s},\hat{t}$ and $\hat{u}$ (defined as in eq. (1.25)) now fulfil the constraint $\hat{s}+\hat{t}+\hat{u} = 1+\hat{m}_1^2+\hat{m}_2^2$. And second, we now have

$$\xi \equiv \hat{u}-\hat{t}+\frac{\hat{m}_2^2-\hat{m}_1^2}{\hat{s}} . \tag{1.53}$$

The $B_i^{(j)}$ functions are defined as follows:

$$B_0^{(1)} = \left(V_\perp^2(\hat{s})+V_\parallel^2(\hat{s})\right)\times \left[\left(\hat{s}-\frac{\hat{M}^2\hat{m}^2}{\hat{s}}\right)\left(|C_9|^2+|C_{10}|^2\right)+4\hat{m}_1\hat{m}_2\left(|C_9|^2-|C_{10}|^2\right)\right] , \tag{1.54}$$

$$B_1^{(1)} = 8\hat{s}\,V_\perp(\hat{s})V_\parallel(\hat{s})\mathrm{Re}\left(C_9 C_{10}^*\right) ,$$

$$B_2^{(1)} = \hat{s}\left(V_\perp^2(\hat{s})+V_\parallel^2(\hat{s})\right)\left(|C_9|^2+|C_{10}|^2\right) , \tag{1.55}$$

$$B_0^{(2)} = 2\hat{M}^2\left(2\hat{s}\,\rho\,(1-\hat{m}^2)+x^2\left(1-\frac{\hat{M}^2\hat{m}^2}{\hat{s}^2}\right)\right)|C_{10}|^2 + 2\hat{m}^2\left(2\hat{s}\,\rho\,(1-\hat{M}^2)+x^2\left(1-\frac{\hat{M}^2\hat{m}^2}{\hat{s}^2}\right)\right)|C_9|^2 ,$$

$$B_1^{(2)} = \frac{4x^2\hat{M}\hat{m}}{\hat{s}}\left[\hat{M}^2|C_{10}|^2+\hat{m}^2|C_9|^2\right] ,$$

$$B_2^{(2)} = 2\hat{M}^2\left(2\hat{s}(\hat{m}^2-1)-x^2\right)|C_{10}|^2 + \tag{1.56}$$

$$2\hat{m}^2\left(2\hat{s}(\hat{M}^2-1)-x^2\right)|C_9|^2 , \tag{1.57}$$

$$B_0^{(12)} = 8x\,V_\perp(\hat{s})\left(\hat{m}_1^2+\hat{m}_2^2-\frac{\hat{M}^2\hat{m}^2}{\hat{s}}\right)\mathrm{Re}\left(C_9 C_{10}^*\right) - \frac{4\hat{M}\hat{m}}{\hat{s}}\,V_\parallel(\hat{s})\left[(\hat{M}^2-\hat{s})(1-\hat{m}^2)|C_{10}|^2+(\hat{m}^2-\hat{s})(1-\hat{M}^2)|C_9|^2\right] ,$$

$$B_1^{(12)} = 8x\hat{M}\hat{m}\,V_\perp(\hat{s})\mathrm{Re}\left(C_9 C_{10}^*\right)+4xV_\parallel(\hat{s})\left[\hat{M}^2|C_{10}|^2+\hat{m}^2|C_9|^2\right] ,$$

$$B_2^{(12)} = -4\hat{s}\hat{M}\hat{m}\,V_\parallel(\hat{s})\left(|C_{10}|^2+|C_9|^2\right) , \tag{1.58}$$

where

$$\hat{M} = \hat{m}_1+\hat{m}_2 , \quad \hat{m} = \hat{m}_2-\hat{m}_1 , \quad \rho = \frac{(\hat{s}-\hat{M}^2)(\hat{s}-\hat{m}^2)}{\hat{s}^2} . \tag{1.59}$$

In the lepton-flavour conserving limit $\ell_1 = \ell_2$, these equations reproduce the $C_{9,10}$ part of the results of section 1.1.3.

For the $V_{\perp,\parallel}^{(B_q)}$ form factors we still use the parametrization eq. (1.15), recalling that within few percent, the $B_s^0 \to \gamma$ form factors coincide with the $B^0 \to \gamma$ ones. Such differences are clearly negligible in our context.

The branching ratio for the corresponding non-radiative decay $B_q \to \ell_1^+ \ell_2^-$ reads

$$\begin{aligned} \mathcal{B}(B_q \to \ell_1^+ \ell_2^-) &= \tau_{B_q} \frac{G_F^2 \alpha_{\rm em}^2 M_{B_q}^3 f_{B_q}^2}{2^6 \pi^3} |\lambda_t^{(q)}|^2 \sqrt{\rho} \times \\ &\quad \left[(1-\hat{m}^2)|F_P + \hat{M} C_{10}|^2 + (1-\hat{M}^2)|F_S - \hat{m} C_9|^2 \right] , \end{aligned} \tag{1.60}$$

with

$$F_{S,P} = M_{B_q} \frac{m_b C_{S,P} - m_q C'_{S,P}}{m_b + m_q} . \tag{1.61}$$

In the lepton-flavour conserving limit this formula reproduces exactly eq. (1.8).

1.4.2 Numerical Analysis

The BR predictions will depend on two basic parameters, the overall strength G of the interaction in eq. (1.42), and the product of four chiral rotations turning the fermion fields $(\bar{b}'b')(\bar{\tau}'\tau')$ into the fields relevant for the process, $(\bar{b}q)(\bar{\ell}_2\ell_1)$. This product of four U-matrix entries will be denoted as U_4 for brevity.

The parameters G and U_4 are completely unknown and we have at best some guiding criteria to fix them to reasonable ranges:

- Since, for a given process, G and U_4 always appear as a product, it is always possible to shuffle an arbitrary numerical factor between G and U_4. As a consequence, to fix a reasonable range for G with any confidence, one may consider predictive models for U_4, as in ref. [174] (see also [183]). One obtains new-physics scales $\Lambda_{NP} = 1/\sqrt{G}$ between 750 GeV and 5 TeV.[16] We then assume 4×10^{-8} GeV $\leq G \leq 2 \times 10^{-6}$ GeV.

- The neutrino mixing matrix (as known as Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix) anarchy suggests that the leptonic part of U_4 may be of $\mathcal{O}(1)$. As concerns the U_L^d matrix entries 33 and $3q$, one can assume them to be close in magnitude to the CKM entries V_{tb} and V_{tq} respectively[17]. We therefore consider the range $10^{-4} \leq U_4 \leq 0.05$, keeping in mind that B_s^0 would correspond to $U_4 \sim |V_{ts}| \simeq 0.04$ and B^0 to $U_4 \sim |V_{td}| \simeq 0.008$.

[16] As emphasized in [174], these mass scales may appear low for, say, a Z' as the underlying mediator of the interaction eq. (1.42). However, it must be remembered that this interaction couples primarily to the third generation.

[17] This assumption should actually hold to a good extent, provided that new interactions other than eq. (1.42) are indeed negligible, as assumed here.

To get an idea of the resulting predictions, we note that the upper limit $\mathcal{B}(B_s^0 \to \mu^\pm e^\mp) < 5.4 \times 10^{-9}$ from LHCb [180] corresponds to $G \times U_4 = 8.9 \times 10^{-9}$, which is 10 times smaller than the product of our highest allowed values for G and U_4. A general picture of the predictions for the $B_s \to \mu e\gamma$ branching ratio and its non-radiative counterpart as a function of G vs. U_4 in the above-mentioned ranges is presented in Fig. 1.10. The gray area denotes the parameter space excluded by the LHCb $B_s^0 \to \mu e$ search of ref. [180].

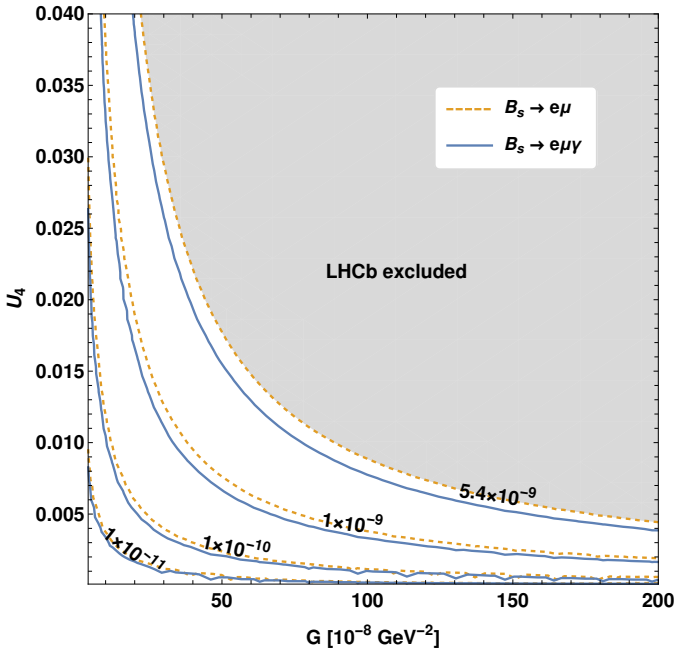

Figure 1.10: $\mathcal{B}(B_s \to e^\pm\mu^\mp\gamma)$ (blue, solid) and $\mathcal{B}(B_s^0 \to e^\pm\mu^\mp)$ (orange, dashed) as a function of G vs. the product of U-matrix entries appearing in eq. (1.49), and denoted as U_4. (See text for more details.) The gray area is excluded by the LHCb upper limit on the non-radiative decay [180].

The figure shows that the radiative mode is slightly enhanced with respect to the non-radiative counterpart. Actually, within our considered model, where the shifts to $C_9^{\ell_1\ell_2}$ and $C_{10}^{\ell_1\ell_2}$ differ only by a sign, the $|G \times U_4|^2$ dependence cancels altogether in the radiative over non-radiative ratio, and we find

$$\frac{\mathcal{B}(B_s^0 \to \mu e\gamma)}{\mathcal{B}(B_s^0 \to \mu e)} = 1.3\,, \quad \frac{\mathcal{B}(B^0 \to \mu e\gamma)}{\mathcal{B}(B^0 \to \mu e)} = 1.2\,, \quad \frac{\mathcal{B}(K \to \mu e\gamma)}{\mathcal{B}(K \to \mu e)} = 2.7{\times}10^{-2}\,. \quad (1.62)$$

These numbers can be intuitively understood as follows. First note that the Bremsstrahlung contribution $d\Gamma^{(2)}$ to the radiative decay comes with a factor of $(f/M)^2$, with M the mass of the decaying meson and f its decay constant, as well as with a chiral suppression factor. On the other hand, both of these suppression

factors are absent in the direct-emission contribution $d\Gamma^{(1)}$. Therefore, the radiative decay will be competitive with the non-radiative one – the latter also $(f/M)^2$ as well as chirally suppressed – to the extent that the direct-emission contribution can dominate, which occurs whenever f/M is small enough, that is the case for both B^0 and B_s^0, but not for kaons. In other words, the larger the ratio M/f, with M the mass of the decaying meson and f its decay constant, the larger the parametric enhancement of the radiative decay over the non-radiative counterpart.

It is clear that for the K case, the radiative mode is too suppressed to be potentially interesting, unless the $K \to \mu e$ mode is found at an unexpectedly large rate. The current experimental limit on this mode reads $\mathcal{B}(K_L \to \mu^\pm e^\mp) < 4.7 \times 10^{-12}$ [184]. Although LHCb is not designed for the study of kaon decays, simulations show that it could improve the limit on the K_L and set the first limit on K_S after the Upgrade II [185].

Assuming experimental efficiencies for radiative and non-radiative cases to be comparable, the measurement of the B_q radiative decay along with the non-radiative one offers a precious cross-check of the new-physics mechanism responsible for a possible LFV signal. It will be clear in the second part that this assumption does not hold in a hadronic collider where the large photon background requires stronger constraints on the kinematics. For $B_s^0 \to \mu^+\mu^-\gamma$ decays at LHCb, the reconstruction and selection efficiencies are found to be around one order of magnitude smaller than for the non-radiative $B_s^0 \to \mu^+\mu^-$ decay. On the other hand LFV decays benefit from much lower peaking backgrounds as the latter only comes from mis-identified decays. An additional challenge may also arise from the separation between the final state photon and material-induced bremsstrahlung of the electron. All these considerations require a specific study based on simulated signal. We also note that a method similar to the one proposed is section 1.3, namely the partial reconstruction of the radiative decay, can a priori not be applied for LFV decays as it hinges on the non-radiative dataset which in our case is obviously not present.

1.4.3 Conclusion

In this section we generalized the study of $B_s^0 \to \ell\ell\gamma$ decays to LFV decays. LFV is theoretically motivated by the hints of LNU and the fact that without additional mechanisms the latter imply the former. The search for LFV decays is however experimentally challenged by the small expected branching ratios and limited reconstruction of tau leptons. We showed that adding a photon to the final of $B \to \ell\ell'$ decays can enhance the expected branching ratio of these decays by lifting the chiral suppression factor. $B \to \ell\ell'\gamma$ therefore offers an additional channel to look for LFV at collider and particularly at Belle II.

This section closes this first chapter dedicated to radiative leptonic decays. We showed in particular that $B_{(s)} \to \ell\ell\gamma$ decays possess a very rich phenomenology

and offer stringent probes of the SM. The theoretical uncertainties attached to the prediction of their SM branching ratio are mainly due to resonances at low-q^2 and to the from-factors that describe the $B \to \gamma$ transition. These form-factors are now computed in the high energetic photon regime and should be estimated on the lattice in the other regime, where the photon is soft. The uncertainty on these form-factors however cancels in a large extent in the ratio of two different leptonic final states, making this ratio a precise test of flavour universality. Concerning the meson resonances, we showed that exploiting related experimental measurements permits a reduction of the attached uncertainty.

The discussions proposed in this chapter will serve as guide for the experimental analysis presented in Part II. The modest reconstruction of photons in LHCb motivates an alternative method to probe $B_s^0 \to \mu^+\mu^-\gamma$. This novel method consists in measuring this radiative decay on the shoulders of $B_s^0 \to \mu^+\mu^-$. A scrupulous study of the backgrounds polluting the low-mass side-band of $B_s^0 \to \mu^+\mu^-$ should allow a first measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$ already at LHCB's Run II. These measurements will in the end offer crucial cross-checks of the current B anomalies.

2

Theoretical approaches

Contents

In this chapter we will study theoretical implications of the anomalies in $b \to s$ and $b \to c$ transitions. In section 2.1, we discuss the possibility of building models that explain both sets of measurements. We will see that this approach is challenged by low-energy measurements, but can still be achieved in model involving leptoquarks. In section 2.2 we propose another approach to $b \to s$ anomalies that results in an effective model that passes all experimental constraints.

2.1 Interpreting the B-anomalies

Assuming that $b \to c$ and $b \to s$ anomalies are more than a statistical fluctuation, it is natural to look for a combined explanation to both sets of observables. The immediate difficulty of this approach is that, although the shifts implied on the amplitude are quantitatively equivalent (15% w.r.t. the SM contribution), $b \to c$ data require a correction to a tree-level process while the $b \to s$ ones imply the same correction but to a loop-induced amplitude. This qualitative difference requires a tuning of the NP contribution to achieve different size of effect in the two

transitions. Before discussing such models, let us first come back to the simplified NP model presented in section 1.4.

2.1.1 Beyond the simplest model

The main issue of eq. (1.42) is that the underlying NP scale is larger than the EWSB scale. This implies that the operators generated when one integrates out the NP must be invariant under the unbroken SM gauge group [186]. We therefore replace the left-handed fields b'_L and τ'_L with the $SU(2)_L$ doublets Q'_3 and L'_3. Two operators can then give rise to eq. (1.42)[1]

$$\mathcal{H}^{(1)}_{\rm NP} = G^{(1)}\,(\bar{Q}'_3\gamma^\lambda Q'_3)\,(\bar{L}'_3\gamma_\lambda L'_3)\,, \tag{2.1}$$
$$\mathcal{H}^{(3)}_{\rm NP} = G^{(3)}\,(\bar{Q}'_3\gamma^\lambda \sigma_a Q'_3)\,(\bar{L}'_3\gamma_\lambda \sigma_a L'_3)\,, \tag{2.2}$$

where $G^{(1)}$ and $G^{(3)}$ are both at the new-physics scale, $G^{(1,3)} \sim 1/\Lambda^2_{\rm NP}$, but not necessarily equal and σ_a are Pauli matrices.

Although this new setup limits the predicting power of the model by adding new parameters, it comes with the very interesting feature that it also involves charged current. Indeed eq. (2.2), once rotated to the basis of mass eigenstates, contains the operator $(\bar{b}_L\gamma^\lambda c_L)\,(\bar{\tau}_L\gamma_\lambda\nu_L)$, the very one responsible for $R_{D^{(*)}}$ anomalies. This means that in the simplest scenario, charged and neutral anomalies are naturally related by the SM $SU(2)_L$ symmetry.

Within this model, the new physics Hamiltonian eq. (1.44) now reads

$$\begin{aligned}\mathcal{H}_{\rm NP} = U^{\ell *}_{L3k}U^{\ell}_{L3l}\Big[&(G^{(1)}+G^{(3)})\,U^{u*}_{L3i}U^{u}_{L3j}\,(\bar{u}_{Li}\gamma^\lambda u_{Lj})\,(\bar{\nu}_{Lk}\gamma_\lambda\nu_{Ll})\\ &(G^{(1)}-G^{(3)})\,U^{u*}_{L3i}U^{u}_{L3j}\,(\bar{u}_{Li}\gamma^\lambda u_{Lj})\,(\bar{\ell}_{Lk}\gamma_\lambda\ell_{Ll})\\ &(G^{(1)}-G^{(3)})\,U^{d*}_{L3i}U^{d}_{L3j}\,(\bar{d}_{Li}\gamma^\lambda d_{Lj})\,(\bar{\nu}_{Lk}\gamma_\lambda\nu_{Ll})\\ &(G^{(1)}+G^{(3)})\,U^{d*}_{L3i}U^{d}_{L3j}\,(\bar{d}_{Li}\gamma^\lambda d_{Lj})\,(\bar{\ell}_{Lk}\gamma_\lambda\ell_{Ll})\\ &2G^{(3)}\,(U^{u*}_{L3i}U^{d}_{L3j}\,(\bar{u}_{Li}\gamma^\lambda d_{Lj})\,(\bar{\ell}_{Lk}\gamma_\lambda\nu_{Ll}) + \text{ h.c.})\Big]\,,\end{aligned} \tag{2.3}$$

where neutrino masses have been neglected. Matching this Hamiltonian on the weak basis (1.2) as we did in eq. (1.49) yields

$$\delta C_9^{\ell_1\ell_2} = -\delta C_{10}^{\ell_1\ell_2} = -\frac{\pi}{\alpha_{\rm em}}\frac{G^{(1)}+G^{(3)}}{\sqrt{2}\,G_F}\,\frac{U^{d*}_{L33}U^{d}_{L32}U^{\ell *}_{L3\ell_2}U^{\ell}_{L3\ell_1}}{\lambda_t}\,, \tag{2.4}$$

where we neglected RGE induced contributions, sub-dominant if $G^{(1)} \neq \pm G^{(3)}$

[1] Here again, the notations are kept simple to underline the physics of the processes. The notations, as well as the matching of eqs. (2.1) and (2.2) on eq. (1.42), will be developed in section 3.1.1.

[187][2]. Assuming $C_9^{ee} = 0$ the model therefore predicts

$$R_{K^{(*)}} \approx \frac{|C_9^{\rm SM} + \delta C_9^{\mu\mu}|^2}{|C_9^{\rm SM} + \delta C_9^{ee}|^2} \approx 1 - \left(\frac{G^{(1)} + G^{(3)}}{3.3\ \text{TeV}^{-2}}\right)\left(\frac{U_{L33}^{d*} U_{L32}^{d} |U_{L32}^{\ell}|^2}{10^{-3}}\right), \tag{2.5}$$

where we used the SM value $C_{9,\rm SM}^{\mu\mu} = 4.07$.

The same matching can be applied to the effective Hamiltonian describing the $b \to c$ transition

$$\mathcal{H}_{\rm eff} = \frac{4G_F}{\sqrt{2}} V_{cb}\, C_{L\,ij}\, (\bar{c}\gamma_\mu b_L)(\bar{\ell}_i \gamma^\mu \nu_j) + \text{h.c.}\,, \tag{2.6}$$

and yield [187]

$$C_{L\,ij} = \delta_{ij} - \frac{G^{(3)}}{\sqrt{2}\, G_F}\, C_3\, \frac{U_{L32}^{u*} U_{L33}^{d} U_{L3i}^{\ell *} U_{L3j}^{\ell}}{V_{cb}}\,, \tag{2.7}$$

where δ_{ij} is the SM contribution due to the exchange of a W boson. Assuming $U_{L32}^{\ell *} U_{L32}^{\ell} \ll U_{L33}^{\ell *} U_{L33}^{\ell} \sim 1$, as discussed around eq. (1.47), we found

$$\frac{R_{D^{(*)}}}{R_{D^{(*)}}^{\rm SM}} = \frac{\sum_j |C_{L\,3j}|^2}{\sum_j |C_{L\,\ell j}|^2_{\ell=1,2}} \tag{2.8}$$

$$\approx 1 - \left(\frac{G^{(3)}}{8.3\ \text{TeV}^{-2}}\right) U_{L33}^{d} \left(U_{L33}^{d*} + \frac{V_{cs}}{V_{cb}} U_{L32}^{d*}\right). \tag{2.9}$$

From eqs. (2.5) and (2.8), it is clear that an appropriate choice of $G^{(1)}$ and $G^{(3)}$ allows for a combined explanation of $R_{K^{(*)}}$ and $R_{D^{(*)}}$. However, the running of Hamiltonian (2.3) from the NP scale down to m_b gives rise to non-zero contributions in a set of well tested observables [187]. For example, one of the most constraining observables is obtained by closing the top loop in the two first lines of eq. (2.3), yielding a RGE induced contribution to the operator $(\bar{\ell}_L \gamma^\lambda \nu_L)\,(\bar{\nu}_L \gamma_\lambda \ell_L)$. This operator is responsible for LFU breaking effects in $\tau \to \ell \bar{\nu} \nu$ decays, tested at the per mil accuracy. In practice we can define the double ratio

$$R_\tau^{\ell_i} \equiv \frac{\mathcal{B}(\tau \to \ell_i \bar{\nu}\nu)|_{\rm SM+NP}}{\mathcal{B}(\tau \to \ell_i \bar{\nu}\nu)|_{\rm SM}} \frac{\mathcal{B}(\mu \to e \bar{\nu}\nu)|_{\rm SM}}{\mathcal{B}(\mu \to e \bar{\nu}\nu)|_{\rm SM+NP}}\,. \tag{2.10}$$

The experimental measurements read [188]

$$R_\tau^\mu = 1.0022(30) \qquad R_\tau^e = 1.0060(30)\,, \tag{2.11}$$

to be compared to the models prediction [187]

$$R_\tau^\mu = R_\tau^e \approx 1 + \left(\frac{G^{(3)}}{125\ \text{TeV}^{-2}}\right). \tag{2.12}$$

[2] The scenario $G^{(1)} = G^{(3)}$, discussed later in this thesis, is particularly interesting because it allows to avoid the constraints from $B \to K^{(*)} \nu \bar{\nu}$ decays.

Assuming that the U-matrices contribution is of order unity, accounting for a 15% effect in eq. (2.8) requires an energy scale larger than 1 TeV, which is challenged by eq. (2.12) and the corresponding experimental measurements.

With a more complete study, as will be performed in the next chapter, one can show that the combined constraints from τ decays and Z observables strongly disfavor an explanation of both $b \to s$ and $b \to c$ anomalies based on the Hamiltonian (2.3) [187].

2.1.2 Model builder hurdles and leptoquarks

The study of the simple third-generation dominated scenario points out the difficulty of interpreting current data, as one needs to provide

(i) a substantial correction (15-20%) to the vertex $J_q \times J_\ell$ ($J_{q(\ell)}$ denote a left-handed quark and lepton current respectively);

(ii) small shifts in $J_q \times J_q$, constrained by meson mixings;

(iii) small shifts in $J_\ell \times J_\ell$, constrained by lepton decays.

These items naturally disfavour models with an additional Z', W' or Higgs boson, as making $J_q \times J_\ell$ large in these scenarios would require either $J_q \times J_q$ or $J_\ell \times J_\ell$ to be large. This class of models is therefore severly constrained by low-energy observables (see *e.g.* [99, 189–192])[3].

On the other hand, scenarios based on leptoquarks (LQ), fields that couples simultaneously to leptons and quarks, are highly advantaged. Indeed, in these scenarios, the $J_q \times J_\ell$ vertex correction comes from a tree-level amplitude while the two other vertex corrections are loop-suppressed[4].

LQs were extensively studied in the context of unified theories based on gauge groups embedding the SM (such as $SU(5)$ [195] or $SO(10)$ in the case of Pati-Salam's model [196]). 12 scalar or vector LQs with renormalizable couplings can be written and classified by their representation under the SM gauge group. Some of these fields violate the baryon number conservation requiring very large masses to avoid limits from proton lifetime, but the others have to satisfy much weaker constraints and present a very rich phenomenology at the TeV-scale [194].

Revisiting LQs in the context of B anomalies yields the following observations:

- Several LQ models offer an explanation of $b \to s$ or $b \to c$ anomalies. These models are summarized in Table 2.1. Among them, the U_1 is the only LQ that can explain both $b \to s$ and $b \to c$ anomalies[5] [100, 189, 199–202].

[3] The difficulty of constructing a model explaining $b \to s$ and/or $b \to c$ transitions is not limited to this aspect. For instance, high-p_T searches employing $\tau^+\tau^-$ signature severely constrain NP models addressing $R_{D^{(*)}}$ [193].

[4] At least in models where diquark couplings are absent [194].

[5] Also noteworthy, thanks to loop-corrections, the S_1 model was also shown to be a viable candidate in a restricted parameter space [102] (see also [197, 198]).

- Less minimalistic models, based on two scalar LQs, can also be constructed. Recent attempts include models with S_1 and S_3 [189, 203–207] or R_2 and S_3 [208].

Model	$R_{K^{(*)}}$	$R_{D^{(*)}}$	$R_{K^{(*)}}$& $R_{D^{(*)}}$
$S_1 = (\bar{\mathbf{3}}, \mathbf{1})_{1/3}$	✗*	✓	✗*
$R_2 = (\mathbf{3}, \mathbf{2})_{7/6}$	✗*	✓	✗
$\widetilde{R}_2 = (\mathbf{3}, \mathbf{2})_{1/6}$	✗	✗	✗
$S_3 = (\bar{\mathbf{3}}, \mathbf{3})_{1/3}$	✓	✗	✗
$U_1 = (\mathbf{3}, \mathbf{1})_{2/3}$	✓	✓	✓
$U_3 = (\mathbf{3}, \mathbf{3})_{2/3}$	✓	✗	✗

Table 2.1: Summary of models based on a single leptoquark mediator explaining some of the anomalies [201]. ✗* means that the model can only partially allay the anomaly.

Although a simplified model based on the U_1 LQ is particularly attractive, one need to provide an ultraviolet completion for at least two reasons. First, this completion is required to obtain a mass term for the U_1 because it is a vector particle. Second, many contributions to loop-induced amplitudes, for instance for $b \to s\nu\nu$ decays or meson mixings, show a quadratic dependence on the energy cutoff [100]. In a renormalizable, UV complete, model, all of these amplitudes would be calculable. Before entering the discussion of a simplified model with the U_1 in section 3.1 and a UV complete model in section 3.2, let us first come back to items *(i)-(iii)* discussed above.

As we said, new bosons with flavour universal couplings to J_q and J_ℓ are highly constrained by low-energy precision observables [187, 209–211] and direct searches [193]. However more involved couplings can induce a hierarchy between $J_q \times J_\ell$ and $J_q \times J_q$ and $J_\ell \times J_\ell$ contributions. The following section is devoted to a model where this hierarchy is due to the remnant of a $SU(2)$ symmetry acting between the second a the third generation.

2.2 Alternative path: a horizontal symmetry

As we discussed, models proposing a combined explanation to $b \to s$ and $b \to c$ anomalies have to invariably withstand non-negligible constraints from low energy observables. On the other hand, experimental measurements are still affected by large statistic and systematic uncertainties. Every new measurement can result in a sizeable change of the flavour landscape. The difficulty in the interpretation of

these results may therefore be considered as the sign of too premature experimental results. In this section, adapted from [212], we will focus on $b \to s$ anomalies only and describe a model *not based on leptoquarks* that naturally presents a solution to the items *(i)-(iii)* of the previous section.

2.2.1 Model

Items *(i)-(iii)* suggest, for reasons that will be transparent shortly, the consideration of a 'horizontal' group, with $SU(2)$ being the smallest continuous group that may be at play. We therefore consider the gauge group $G_{\rm SM} \times G_h$, where $G_{\rm SM}$ is the SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$, acting 'vertically' in each generation, and $G_h = SU(2)_h$ is a horizontal group connecting the second and the third generations as defined before EWSB. More generally, we may actually assume one $SU(2)_h$ symmetry for either chirality of fermions.

We can then augment the SM Lagrangian with the following terms

$$\delta\mathscr{L} = \sum_{\mathscr{F},a} \bar{\mathscr{F}} \left(g_L \gamma_\mu P_L G_L^{\mu\, a} + g_R \gamma_\mu P_R G_R^{\mu\, a} \right) \tau^a \mathscr{F} \ , \tag{2.13}$$

where $P_{L,R}$ are the usual chirality projectors, that we henceforth include in the gamma matrices for brevity, *i.e.* $\gamma^\mu_{L,R} \equiv \gamma^\mu P_{L,R}$. Furthermore, $\tau^a = \sigma^a/2$, and $G^{\mu\, a}_{L,R}$ are the gauge bosons of the horizontal symmetry for either chirality, whose masses are assumed to be larger than the EWSB scale. The fields $\mathscr{F}$ are such that

$$\mathscr{F} \equiv \begin{pmatrix} f_2 \\ f_3 \end{pmatrix} \ , \tag{2.14}$$

where f runs over all SM fermion species (primed as they are not in their mass eigenbasis) and 2, 3 refer to generation indices.

Integrating out the G_L and G_R gauge bosons, one obtains the effective interactions

$$\begin{aligned} \delta\mathscr{L}_{\rm eff} = - \sum_{\mathscr{F}_1,\mathscr{F}_2,a} \Bigg\{ & \frac{g_L^2}{2M^2_{G_{La}}} \left(\bar{\mathscr{F}}_1 \gamma^\mu_L \tau^a \mathscr{F}_1 \right) \left(\bar{\mathscr{F}}_2 \gamma_{\mu\, L} \tau^a \mathscr{F}_2 \right) \\ & + \frac{g_R^2}{2M^2_{G_{Ra}}} \left(\bar{\mathscr{F}}_1 \gamma^\mu_R \tau^a \mathscr{F}_1 \right) \left(\bar{\mathscr{F}}_2 \gamma_{\mu\, R} \tau^a \mathscr{F}_2 \right) \Bigg\} \ , \end{aligned} \tag{2.15}$$

where both $\mathscr{F}_1$ and $\mathscr{F}_2$ are defined as in eq. (2.14). Below the EWSB scale, where SM fermions acquire masses, the fields $\mathscr{F}$ undergo unitary[6] transformations of the kind

$$\mathscr{F} = \mathcal{U}_{\mathscr{F}} \hat{\mathscr{F}} \tag{2.16}$$

[6] For two generations, as in eq. (2.14), these transformations are actually not unitary. We will make notation more precise afterwards and in particular justify how the Yukawa couplings can be compatible with the advocated symmetry.

where $\hat{\mathscr{F}}$ denotes the mass-eigenbasis fields. After such transformations, eq. (2.15) becomes

$$\delta\mathscr{L}_{\text{eff}} \propto \frac{1}{2M^2_{G_{La}}} \left(\bar{\hat{\mathscr{F}}}_1 \mathcal{U}^\dagger_{\mathscr{F}_1} \gamma^\mu_L \tau^a \mathcal{U}_{\mathscr{F}_1} \hat{\mathscr{F}}_1\right) \left(\bar{\hat{\mathscr{F}}}_2 \mathcal{U}^\dagger_{\mathscr{F}_2} \gamma_{\mu\, L} \tau^a \mathcal{U}_{\mathscr{F}_2} \hat{\mathscr{F}}_2\right) \ , \tag{2.17}$$

and analogous structures are generated by the G_{Ra} terms. Hence, in either the left- or right-handed sector, these effective interactions have the form $\sum_a J^a_{\mathscr{F}_1} \times J^a_{\mathscr{F}_2}/M^2_{Ga}$. The interesting phenomenological feature becomes apparent when considering products of currents involving the same fermions, $\mathscr{F}_1 = \mathscr{F}_2$. In the limit of mass degeneracy across the horizontal bosons of either sector, the rotations $\mathcal{U}_{\mathscr{F}}$ can be removed by the redefinition $G^{\mu\, a}_{L,R}\tau^a \to \mathcal{U}_{\mathscr{F}}\, G^{\mu\, a}_{L,R}\tau^a \mathcal{U}^\dagger_{\mathscr{F}}$, which means that *the corresponding fermion bilinears can be taken as flavour diagonal in all generality* [213]. This property brings a natural answer to the problem discussed in the previous section.

There is actually a subtlety, already mentioned in footnote 6. Although above the EWSB scale $SU(2)_h$ involves only the 2$^{\text{nd}}$ and 3$^{\text{rd}}$ generations, mixing beneath this scale involves all the three generations. As a consequence, the $\mathcal{U}_{\mathscr{F}}$ matrices in eq. (2.16) are not exactly unitary, implying that the contributions to processes such as meson mixings, as well as to decays involving only leptons, are non-zero. It is true that these contributions will be parametrically suppressed by powers of the mixing between the two heavier and the light generation. However, a non-zero mixing onto the 1$^{\text{st}}$ generation translates into contributions to light-fermion processes like $K^0 - \bar{K}^0$ mixing and $\mu \to 3e$ for example, which are well-known to be very constraining [214]. We will discuss such effects in detail in the analysis.

We first need to generalize the formalism in eqs. (2.13)-(2.17) to account for three-generation mixing. We then define

$$\mathscr{F} \equiv \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} \ . \tag{2.18}$$

The Lagrangian shift in eq. (2.13) becomes

$$\delta\mathscr{L} = \sum_{\mathscr{F}} \bar{\mathscr{F}} \left(g_L \gamma_{\mu\, L} G^{\mu\, a}_L + g_R \gamma_{\mu\, R} G^{\mu\, a}_R\right) T^a \mathscr{F} \ , \tag{2.19}$$

with namely the replacement $\tau^a \to T^a$, where

$$T^a \equiv \begin{pmatrix} 0 & & \\ & \tau^a & \end{pmatrix} \ , \tag{2.20}$$

and eqs. (2.15) and (2.17) will change accordingly. It is this Lagrangian that we will use in the analysis.

Before pursuing with the phenomenology of the model, four comments are in order.

First, one can wonder whether the enlarged gauge group may introduce anomalies. A simple horizontal gauge group G_h introduces two potentially worrisome anomaly diagrams: the one with G_h^3 and the one with $G_h^2 \times U(1)_Y$. Having chosen G_h to be an $SU(2)$ group, the first diagram vanishes – whereas it wouldn't for larger simple groups such as $SU(3)$. As concerns the second diagram, it likewise vanishes, and it does so for the same reason also at work within the SM, namely the rather magical compensation of the quark vs. lepton $U(1)_Y$ quantum numbers of either chirality. So this anomaly cancels separately for an $SU(2)_h$ coupled only to left-handed fermions or to right-handed ones.

Then, we reiterate that the argument leading to flavour-diagonal $J_q \times J_q$ and $J_\ell \times J_\ell$ amplitudes holds for horizontal bosons with degenerate masses, which needn't be the case. In fact, a departure from the hypothesis of exact mass degeneracy will be instrumental for our framework to withstand the constraints from, in particular, $D^0 - \bar{D}^0$ mixing. Effects on this and other observables will thus be parametric in the mass splittings among the horizontal bosons.

Moreover, as concerns the performance of eq. (2.19) in explaining $R_{K^{(*)}}$, we note that the introduction of a sizeable contribution to the quark right-handed bilinear is expected to upset the relation $R_{K^*} \simeq R_K$ [215]. The bulk of our analysis will therefore assume $g_R = 0$.

Finally, one can wonder how a horizontal symmetry involving the two heavier generations may be compatible with the observed fermion masses and mixing. One possibility is to a consider a scenario akin to partial compositeness [216], where the UV Yukawa terms involve the product between SM fermions, new vector-like fermions Ψ as well as suitable scalar representations Φ_i to break $SU(2)_h$ spontaneously. Similar mechanisms were also implemented in other gauge models for $b \to s$ anomalies [217–219]. For definiteness, one could consider the following field content

$$\Psi^U_{L,R} \sim (\mathbf{3},\mathbf{1};\mathbf{2})_{2/3}\ , \quad \Psi^D_{L,R} \sim (\mathbf{3},\mathbf{1};\mathbf{2})_{-1/3}\ , \quad \Phi_{1,2} \sim (\mathbf{1},\mathbf{1};\mathbf{2})_0\ , \tag{2.21}$$

where transformation properties refer to $G_{\rm SM} \times G_h$. This field content gives rise to the following renormalizable Lagrangian terms

$$\begin{aligned} \delta\mathscr{L}_Y \ \supset\ & m_U \bar{\Psi}^U_L \Psi^U_R \ +\ m_D \bar{\Psi}^D_L \Psi^D_R \\ +\ & \sum_{a=1,2} \sum_{i=1,2,3} \Big((Y_U)_{ai} \overline{\Psi^U_L} \Phi_a (u_R)_i + (Y_D)_{ai} \overline{\Psi^D_L} \Phi_a (d_R)_i \Big) \\ +\ & c_U \bar{F}_Q \,\tilde{\varphi}\, \Psi^U_R + c_D \bar{F}_Q \,\varphi\, \Psi^D_R \ +\ \text{H.c.}\,, \end{aligned} \tag{2.22}$$

where i is a flavour index, $F_Q = (Q_2, Q_3)^T$, with $_{2,3}$ generation indices, and φ is the SM Higgs doublet. SM Yukawa terms for quarks would then arise after integrating out the heavy Ψ and assigning vacuum expectation values to the Φ's. Two scalar fields are needed in eq. (2.21) in order to generate rank-3 effective Yukawa matrices.

The number of dimensionless parameters and mass scales thus involved is sufficient to accommodate quark masses and mixing. An entirely similar construction allows to also address lepton masses and mixing.

2.2.2 Scenario 0

We then start from the effective Lagrangian eq. (2.15), keeping henceforth only the left-handed interaction, and with mass-degenerate horizontal bosons. To simplify notation, we will in the rest of this section remove the L subscript from the G_{La} fields (denoted simply as G_a) as well as from the quark and lepton multiplets in generation space. We will instead keep the subscript in g_L, in order that this coupling not be confused with SM ones.

Fermionic fields $\mathscr{F}$, eq. (2.18), are rotated to the mass eigenbasis through chiral transformations $\mathcal{U}_{\mathscr{F}}$. The most general parametrization compatible with the $SU(2)_h$ symmetry would be of the form

$$\mathcal{U}_{\mathscr{F}} = \begin{pmatrix} e^{i\phi_{\mathscr{F}}} & 0_{1\times 2} \\ 0_{2\times 1} & e^{i\Phi_{\mathscr{F}}}\,\Sigma_{\mathscr{F}} \end{pmatrix} , \tag{2.23}$$

where $\Sigma_{\mathscr{F}} = \exp(i\vec{\theta}_{\mathscr{F}}\,\vec{\sigma}/2)$ parametrizes a general $SU(2)$ transformation acting on the $f_{2,3}$ components of the fermion $\mathscr{F}$. The dependence on the phases $\phi_{\mathscr{F}}$ and $\Phi_{\mathscr{F}}$ cancels in quark and lepton bilinears. As discussed, the dependence on the $SU(2)$ transformation in turn cancels in effective-Lagrangian terms involving one single fermion species ($\mathscr{F}_1 = \mathscr{F}_2$), thus preventing dangerous contributions to processes such as $B_s^0 - \bar{B}_s^0$ mixing and $\tau \to \ell\nu\nu$.

Let us focus on the interaction term involving down-type quarks and charged leptons, $\delta\mathscr{L}_{\text{eff}}^{DL}$, which is of direct interest to us. Eq. (2.15) implies

$$\delta\mathscr{L}_{\text{eff}}^{DL} = -\frac{g_L^2}{m_G^2}\left[\bar{\hat{D}}\left(\mathcal{U}_D^\dagger\,\gamma_L^\mu\,T^a\,\mathcal{U}_D\right)\hat{D}\right]\left[\bar{\hat{L}}\left(\mathcal{U}_L^\dagger\,\gamma_{\mu\,L}\,T^a\,\mathcal{U}_L\right)\hat{L}\right] , \tag{2.24}$$

where $\hat{D} = (d_L, s_L, b_L)^T$ and $\hat{L} = (e_L, \mu_L, \tau_L)^T$ denote left-handed mass-eigenstate fermions and m_G is the common mass of the G_a horizontal bosons. Using the argument below eq. (2.17), one can rewrite this contribution as

$$\delta\mathscr{L}_{\text{eff}}^{DL} = -\frac{g_L^2}{m_G^2}\left[\bar{\hat{D}}\left(\gamma_L^\mu\,T^a\right)\hat{D}\right]\left[\bar{\hat{L}}\left(\mathcal{U}_{DL}^\dagger\,\gamma_{\mu\,L}\,T^a\,\mathcal{U}_{DL}\right)\hat{L}\right] , \tag{2.25}$$

where $\mathcal{U}_{DL} \equiv \mathcal{U}_D^\dagger\mathcal{U}_L$ can be parametrized as in eq. (2.23). Again, phase terms disappear in the lepton bilinear. As concerns the $SU(2)$ transformation Σ_{DL}, it is convenient to express it in terms of Euler angles:

$$\Sigma_{DL} = e^{-i\alpha_{DL}\sigma_3}\,e^{-i\beta_{DL}\sigma_2}\,e^{-i\gamma_{DL}\sigma_3} . \tag{2.26}$$

Neglecting neutrino masses, the γ_{DL} phase can be absorbed in the definition of the charged-lepton fields. A similar rephasing of the D component fields to absorb α_{DL} would shuffle this phase to the CKM matrix, so α_{DL} and β_{DL} are physical.

We next discuss the most general effects within the parametrization in eq. (2.26), in particular whether scenario 0 explains $R_{K^{(*)}}$ and what are the relevant constraints. By matching the Lagrangian in eq. (2.25) with eq. (1.1), we obtain the following shifts of Wilson coefficients

$$\delta C_9^{\mu\mu} = -\delta C_9^{\tau\tau} = e^{-2i\alpha_{DL}} \frac{\sin 2\beta_{DL}\, g_L^2}{8 m_G^2} \cdot \frac{2\pi v^2}{\alpha_{\rm em}\lambda_t}\,, \tag{2.27}$$

$$\delta C_9^{\tau\mu} = e^{-2i\alpha_{DL}} \frac{\cos^2\beta_{DL}\, g_L^2}{4 m_G^2} \cdot \frac{2\pi v^2}{\alpha_{\rm em}\lambda_t}\,, \tag{2.28}$$

$$\delta C_9^{\mu\tau} = -e^{-2i\alpha_{DL}} \frac{\sin^2\beta_{DL}\, g_L^2}{4 m_G^2} \cdot \frac{2\pi v^2}{\alpha_{\rm em}\lambda_t}\,, \tag{2.29}$$

and one has also $\delta C_{10}^{ij} = -\delta C_9^{ij}$. The second factor on the r.h.s. of either of eqs. (2.27)-(2.29) corresponds to an effective scale ≈ 34 TeV [220]. From these equations we also see that largest (in magnitude) shifts to the considered Wilson coefficients are obtained for $\alpha_{DL} \approx 0, \pi$ because of the nearly real SM normalization in the usual CKM conventions. The 1σ bounds in the $\delta C_9^{\mu\mu} = -\delta C_{10}^{\mu\mu}$ fit to data (see Table 3.1) would then amount to

$$\frac{m_G^2}{\sin 2\beta_{DL}\, g_L^2} \in ([15, 18]\ \text{TeV})^2\,, \tag{2.30}$$

and the correlation between the rotation β_{DL} and the scale m_G/g_L is shown in the left panel of Fig. 2.1[7]. Form eqs. (2.27)-(2.29) one also learns that a non-zero value of β_{DL} is needed to successfully explain $R_{K^{(*)}}$. This requirement automatically implies a non-zero value of $\delta C_9^{\tau\mu}$ and $\delta C_9^{\mu\tau}$. This property was already advertised in section 1.4, where we showed that LFV usually arises from LNU. The $b \to s\mu^\pm\tau^\mp$ channel is therefore a very distinctive signature of this scenario and we will discuss it more extensively latter on. Interestingly, our framework predicts an asymmetry between $b \to s\mu^-\tau^+$ and $b \to s\mu^+\tau^-$, controlled by the β_{DL} value.

There is, however, an important caveat around eq. (2.23). Below the EWSB scale, the rotations $\mathcal{U}_D$ and $\mathcal{U}_U$ cannot both have the form in eq. (2.23), because $\mathcal{U}_U^\dagger \mathcal{U}_D = V_{\rm CKM}$. This implies that, although the $SU(2)_h$ symmetry prevents the occurrence of flavour-violating $J_q \times J_q$ and $J_\ell \times J_\ell$ effects for scales above the EWSB one, $J_q \times J_q$ effects will be induced for lower scales, because of $V_{\rm CKM}$-induced mixing (On the other hand, $J_\ell \times J_\ell$ effect will remain tiny because of the very small neutrino masses). The most constraining of these effects turns

[7] In the original reference [212], the fit $\delta C_9^{\mu\mu} = -\delta C_{10}^{\mu\mu}$ yielded slightly different results. The plots and predictions we propose in this section are updated with the fit that will be presented in section 3.1.1.

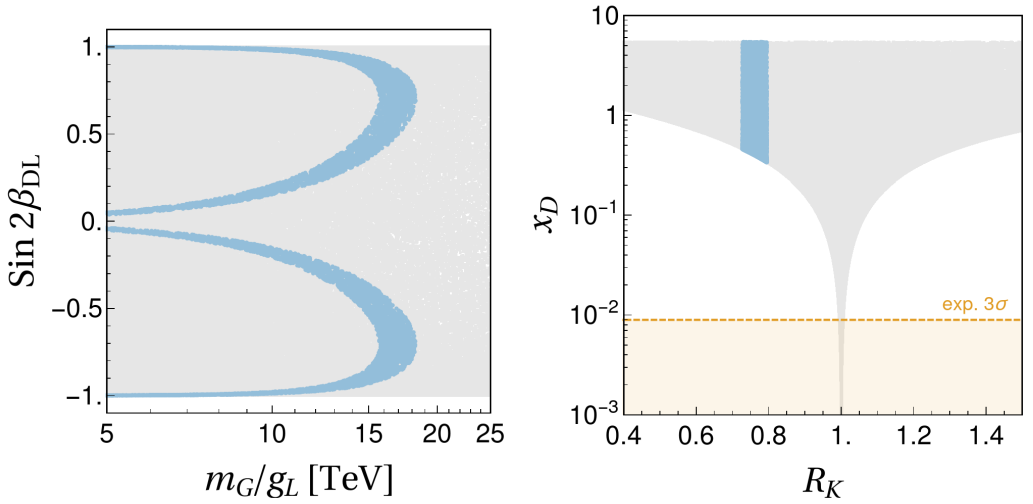

Figure 2.1: Two relevant projections of the model parameter space within scenario 0. Gray points fulfil all constraints except $C^{\mu\mu}_{9,10}$ and x_D. Light-blue points reproduce the values of $C^{\mu\mu}_{9,10}$ obtained in the fit to data at 1σ. The dashed line represents the 3σ upper bound on x_D from ref. [221].

out to be the mass difference in the $D^0 - \bar{D}^0$ system, ΔM_D. We imposed that the latest global fit to the related parameter $x_D \equiv \Delta M_D/\Gamma_D$, where Γ_D is the mean D decay width [221], be saturated by our model's short-distance prediction for the same quantity, that we estimated using ref. [222].[8] We believe that this approach is justified, given that the possible range for the SM contribution to ΔM_D encompasses several orders of magnitude [224].

Barring a tuning of order 10^{-2} between the SM and the new-physics contribution, the x_D constraint unexpectedly excludes our scenario 0. However our model is suited for straightforward generalizations. The latter fall in at least two categories: *(i)* mass splittings among the three vector bosons of the $SU(2)_h$ symmetry; *(ii)* (small) mixing terms between the first and the two heavier generations in the T^a matrices of eq. (2.20). As we will discuss in the rest of this section, the first generalization turns out to be sufficient to pass all constraints.

2.2.3 Scenario 1

The most straightforward generalization of the scenario discussed so far is to allow for non-degenerate masses for the gauge bosons of the $SU(2)_h$ symmetry. The simplest mass splitting is such that two masses stay degenerate [225].

With split G_a masses, we are no more allowed to bundle the two unitary transformations $\mathcal{U}_D$ and $\mathcal{U}_L$ in one single transformation, as in eq. (2.25). Our

[8] For all details on the implementation, see ref. [223].

effective interaction is thus

$$\delta\mathscr{L}_{\rm eff}^{DL} = -\sum_a \frac{g_L^2}{m_{G_a}^2}\left[\bar{\hat{D}}\left(\mathcal{U}_D^\dagger\,\gamma_L^\mu\,T^a\,\mathcal{U}_D\right)\hat{D}\right]\left[\bar{\hat{L}}\left(\mathcal{U}_L^\dagger\,\gamma_{\mu\,L}\,T^a\,\mathcal{U}_L\right)\hat{L}\right]\,, \tag{2.31}$$

i.e. akin to eq. (2.24) but for non-degenerate G_a masses. In a notation straightforwardly generalizing that in eq. (2.26), the matrices $\mathcal{U}_{D,L}$ will introduce the rotations β_D and β_L, as well as the phase parameters $\alpha_{D,L}$ and $\gamma_{D,L}$. To the extent that we do not consider CP-violating observables, non-zero values for these phase terms serve only to suppress the magnitude of the Wilson coefficients relevant to our analysis. We will therefore set $\alpha_{D,L} = \gamma_{D,L} = 0$ and focus on the rotations β_D and β_L. We note that, after EWSB, the above choice for the $\mathcal{U}_D$ matrix allows to subsequently set $\mathcal{U}_U = \mathcal{U}_D V_{\rm CKM}^\dagger$. Choosing the parametrization eq. (2.26) for $\mathcal{U}_D$ is an important assumption, based on phenomenological ground, and guarantees the absence of effects in $K^0 - \bar{K}^0$ mixing.

Quite remarkably a scenario with

$$m_{G_1} = m_{G_2} \ll m_{G_3} \qquad \& \qquad |\sin 2\beta_D| \ll |\sin 2\beta_L|\,, \tag{2.32}$$

accounts at one stroke for new effects in $b \to s\mu\mu$ as large as measured and is compatible with the SM-like results in all other collider datasets. Note however that, allowing for a mass hierarchy between the horizontal gauge bosons amounts to completely forsaking the argument made below eq. (2.17). This makes scenarios 0 and 1 completely different at the level of the underlying mechanisms. Within scenario 0 (degenerate horizontal-boson masses) the flavour diagonality of $J_q \times J_q$ and $J_\ell \times J_\ell$ currents would be the result of an underlying global symmetry coming with the postulated $SU(2)_h$ group. However, off-diagonalities are inescapable because of the CKM matrix, and the result is a too large contribution to $D^0 - \bar{D}^0$ mixing. Within scenario 1, one allows for non-degenerate masses, and phenomenological viability chooses a hierarchical pattern, *i.e.* one of $\mathcal{O}(1)$ breaking of the mentioned *global* symmetry.

The basic mechanism at work can be straightforwardly understood by inspection of the model's prediction of $D^0 - \bar{D}^0$ mixing and R_K, that scenario 0 fell short to describe simultaneously. We will see that, with these two phenomenological requirements fulfilled, all other constraints fall in place, either because of the pattern in eq. (2.32), or because of the underlying $SU(2)_h$ symmetry. We will next discuss all these requirements in turn.

Within scenario 1, the contribution to $D^0 - \bar{D}^0$ is due to

$$\delta\mathscr{L}_{\rm eff}^{UU} = -\sum_a \frac{g_L^2}{2m_{G_a}^2}\left(\bar{U}'\gamma_L^\mu T^a U'\right)^2 \supset C_{(uc)^2}\left(\bar{u}\gamma_L^\mu c\right)^2\,, \tag{2.33}$$

where, exploiting CKM hierarchies, we can write

$$C_{(uc)^2} = -(V_{us}V_{cs}^*)^2\frac{g_L^2}{8}\left(\frac{\sin^2(2\beta_D)}{m_{G_1}^2} + \frac{\cos^2(2\beta_D)}{m_{G_3}^2}\right) + O(\lambda^4)\,. \tag{2.34}$$

with λ the Wolfenstein parameter. While the approximate formula in eq. (2.34) is very convenient to exhibit the mechanism at work, the numerical analysis includes the exact CKM dependence. In turn, the model's contribution to $\delta C_{9,10}^{\mu\mu}$ reads

$$\delta C_9^{\mu\mu} = -\delta C_{10}^{\mu\mu} = \frac{g_L^2}{4}\frac{v^2\pi}{\alpha_{\rm em}\lambda_t}\left[\frac{\sin(2\beta_D)\cos(2\beta_L)}{m_{G_3}^2} - \frac{\sin(2\beta_L)\cos(2\beta_D)}{m_{G_1}^2}\right] . \quad (2.35)$$

Clearly, with the advocated pattern of G_a masses and rotation angles, the contribution in eq. (2.34) will be parametrically suppressed by the decoupling of G_3 plus the smallness of β_D, and this very pattern ensures a sizeable contribution to $\delta C_9^{\mu\mu}$ from the second, negative term in eq. (2.35). These features are displayed more quantitatively in Fig. 2.2. In particular, the effective scale for the lighter among the G_a bosons is shown versus R_K in the left panel, where dark blue denotes points that fulfil all other constraints to be described later, and including x_D. The effective mass scale pointed to by the x_D constraint, ruled by m_{G_3}/g_L,[9] is displayed in the right panel of Fig. 2.2.

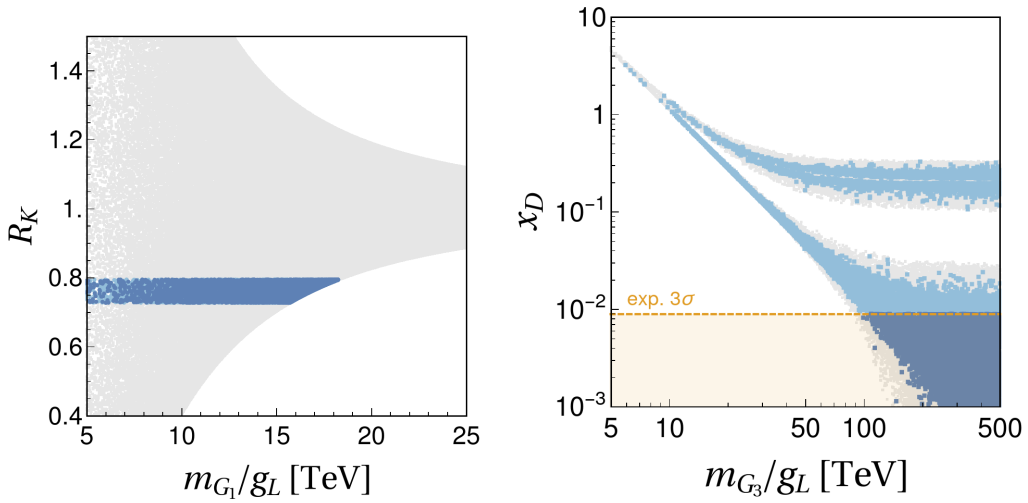

Figure 2.2: Colour code henceforth: grey points fulfil all constraints except $C_{9,10}^{\mu\mu}$ and x_D; light-blue points also fulfil $C_{9,10}^{\mu\mu}$, but not x_D; dark-blue points fulfil all constraints. Left panel: R_K vs. m_{G_1}/g_L as implied by eq. (2.35). Right panel: effective mass scale for the heavier among the G_a bosons, as required by the x_D constraint. The horizontal dashed line denotes the 3σ upper bound on x_D [221].

So far we focused on $D^0 - \bar{D}^0$ mixing because it turned out to be the most constraining observable within scenario 0. However, the pattern of parameters that we advocated in eq. (2.32) may generate large effects in other observables, to be discussed now on. With the exception of $\tau \to \mu\gamma$, all of the considered observables

[9] It is clear that the requirement that x_D saturate the experimental result entails a strong correlation between m_{G_1} and β_D in eq. (2.34).

depend on the ratios m_{G_a}/g_L, rather than on masses and g_L separately. Our main numerical scan will then assumes the ranges to follow

$$m_{G_a}/g_L \in [1, 500] \text{ TeV} \qquad \& \qquad \beta_D, \beta_L \in [0, 2\pi) \ . \tag{2.36}$$

The upper limit of 500 TeV in general corresponds to m_{G_a} values below this mass scale, because g_L is in general below unity. As concerns the hierarchies in eq. (2.32), we followed two alternative procedures: on the one side, we performed scans imposing such hierarchies from the outset; on the other side, we let the constraints choose them. It is quite remarkable that we found no appreciable difference in the results obtained with these two procedures.

$b \to s\ell\ell'$ and leptonic LFV A large β_L, combined with a value for m_{G_1} as low as required by $R_{K^{(*)}}$, may lead to troublesome effects in particular in $b \to s\tau^\pm\mu^\mp$ as well as in leptonic LFV decays such as $\tau \to 3\mu$. Quite interestingly, the effects are indeed sizeable, but below existing limits. Besides, the small number of parameters involved establishes clear-cut correlations between LNU and LFV observables. These correlations represent a prominent feature of this model, as can be qualitatively understood, again, from the basic formula for the relevant Wilson coefficients. A first comment concerns $b \to s\tau\tau$. Since

$$\delta C^{\tau\tau}_{9,10} = -\delta C^{\mu\mu}_{9,10} \ , \tag{2.37}$$

the departure of $\mathcal{B}(B \to K\tau\tau)$ from its SM prediction can be written as a function of the departure of R_K from unity. As a consequence, modifications of branching ratios for $B \to K\tau\tau$ as well as $B_s^0 \to \tau\tau$ will be of the order of 20% with respect to the respective SM expectations, which are sizeably below existing limits [226,227]. Note that this enhancement is relatively small compare to other models, like the one discussed in section 3.1.2, where these branching ratios are enlarged by several orders of magnitude.

We next turn to the predictions for $b \to s\tau\mu$ decays. The relevant Wilson coefficients read

$$\delta C_9^{\mu\tau} = -\frac{g_L^2}{4}\frac{v^2\pi}{\alpha_{\rm em}\lambda_t}\left(\frac{\cos 2\beta_D \cos 2\beta_L}{m_{G_1}^2} + \frac{\sin 2\beta_D \sin 2\beta_L}{m_{G_3}^2} - \frac{1}{m_{G_2}^2}\right) \ , \tag{2.38}$$

$$\delta C_9^{\tau\mu} = -\frac{g_L^2}{4}\frac{v^2\pi}{\alpha_{\rm em}\lambda_t}\left(\frac{\cos 2\beta_D \cos 2\beta_L}{m_{G_1}^2} + \frac{\sin 2\beta_D \sin 2\beta_L}{m_{G_3}^2} + \frac{1}{m_{G_2}^2}\right) \ . \tag{2.39}$$

Keeping in mind the main assumptions defining our scenario 1, eq. (2.32), it is clear that the dominant dependence is on $|\cos 2\beta_L \pm 1|/m^2_{G_{1,2}}$. Hence, a rather distinctive feature of this scenario is that $\mathcal{B}(B \to K\tau^+\mu^-) \neq \mathcal{B}(B \to K\tau^-\mu^+)$, although either can be larger than the other, depending on the choice of the β_L phase. The correlation between these two modes is displayed in Fig. 2.3. The dominant parametric dependence highlighted above translates into an approximate reflection symmetry of the plot around the diagonal.

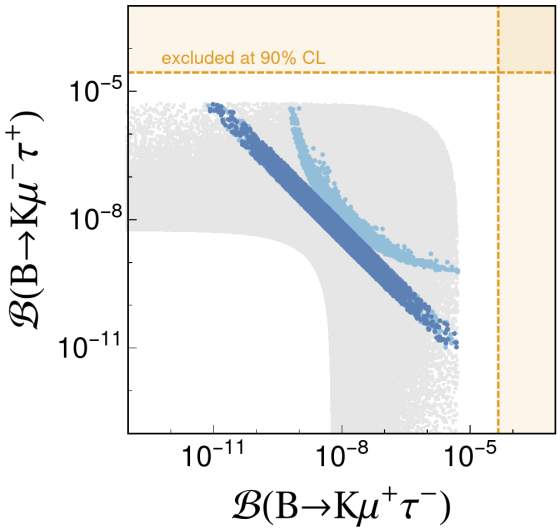

Figure 2.3: Correlation between the two $\mathcal{B}(B \to K\mu\tau)$ modes within scenario 1. Colour code as in Fig. 2.2. Dashed lines denote the existing bounds on the respective modes [178], see text for details.

Most importantly, our model predicts not only an upper bound, but also a lower bound on the LFV rates[10]. We obtain (see also left panel of Fig. 2.4)

$$1.0 \times 10^{-8} \lesssim \mathcal{B}(B \to K\mu^+\tau^-) + \mathcal{B}(B \to K\mu^-\tau^+) \lesssim 5.0 \times 10^{-6} \,. \tag{2.40}$$

Interestingly, the maximal rate predicted by our scenario lies just one order of magnitude below the existing limits discussed in section 1.4.

As anticipated, the above LFV predictions are in turn correlated with purely leptonic LFV, in particular in the processes $\tau \to 3\mu$ and $\tau \to \phi\mu$. From

$$\delta\mathscr{L}_{\text{eff}}^{LL} \supset +\frac{g_L^2}{8}\sin(4\beta_L)\left(\frac{1}{m_{G_3}^2} - \frac{1}{m_{G_1}^2}\right)(\bar{\tau}\gamma_L^\mu\mu)(\bar{\mu}\gamma_{\mu L}\mu) \,, \tag{2.41}$$

one gets [230]

$$\mathcal{B}(\tau \to 3\mu) = \frac{m_\tau^5}{3072\,\pi^3\Gamma_\tau}\frac{g_L^4}{64}\sin^2(4\beta_L)\left(\frac{1}{m_{G_3}^2} - \frac{1}{m_{G_1}^2}\right)^2 \,. \tag{2.42}$$

Similarly to the transition $b \to s\mu^+\mu^-$, this observable is only modified for non-zero values of β_L. Besides, from

$$\delta\mathscr{L}_{\text{eff}}^{LD} \supset C_{ss}^{\mu\tau}\,(\bar{\mu}\gamma_L^\mu\tau)\,(\bar{s}\gamma_{\mu\,L}s) \tag{2.43}$$

[10] Interestingly, the lower bound is in good accord with the predictions obtained within approaches motivated by completely different considerations [174, 183, 197, 228, 229].

with

$$C_{ss}^{\mu\tau} = -\frac{g_L^2}{4}\left[-\frac{\cos(2\beta_D)\sin(2\beta_L)}{m_{G_3}^2} + \frac{\sin(2\beta_D)\cos(2\beta_L)}{m_{G_1}^2}\right] , \tag{2.44}$$

one likewise arrives at

$$\mathcal{B}(\tau \to \mu\phi) \simeq |C_{ss}^{\mu\tau}|^2 \, \frac{f_\phi^2 \, m_\phi^4}{64\,\pi\, m_\tau \Gamma_\tau}\left(1 - \frac{m_\phi^2}{m_\tau^2}\right)\left(-1 + \frac{m_\tau^2}{2m_\phi^2} + \frac{m_\tau^4}{2m_\phi^4}\right) , \tag{2.45}$$

where f_ϕ is the ϕ-meson decay constant given in table 1.2, and we have neglected the m_μ mass dependence, which amounts to an approximation of few percent.

From the current experimental limit [231,232], and keeping in mind eq. (2.32), we obtain

$$\frac{m_{G_{1,2}}^2}{g_L^2 \sin 2\beta_D \cos 2\beta_L} \geq (3.7 \text{ TeV})^2 \qquad (90\% \text{ CL}), \tag{2.46}$$

which is once again consistent with the constraint derived in eq. (2.30) from $R_{K^{(*)}}$. From the requirement that the $b \to s$ discrepancies are reproduced at 1σ, we obtain $\mathcal{B}(\tau \to \mu\phi)$ as large as 5×10^{-11} and, in general, model points mostly populating the range between 10^{-15} and 10^{-11} (see right panel of Fig. 2.4). It is worth mentioning that the projected Belle-II sensitivity to this decay is around 10^{-9} [233].

In the parameter space of eq. (2.32), the above formula translate into a triple correlation between $\mathcal{B}(B \to K\mu^\pm\tau^\mp)$, $\mathcal{B}(\tau \to 3\mu)$ and $\mathcal{B}(\tau \to \mu\phi)$, illustrated in the two plots of Fig. 2.4.

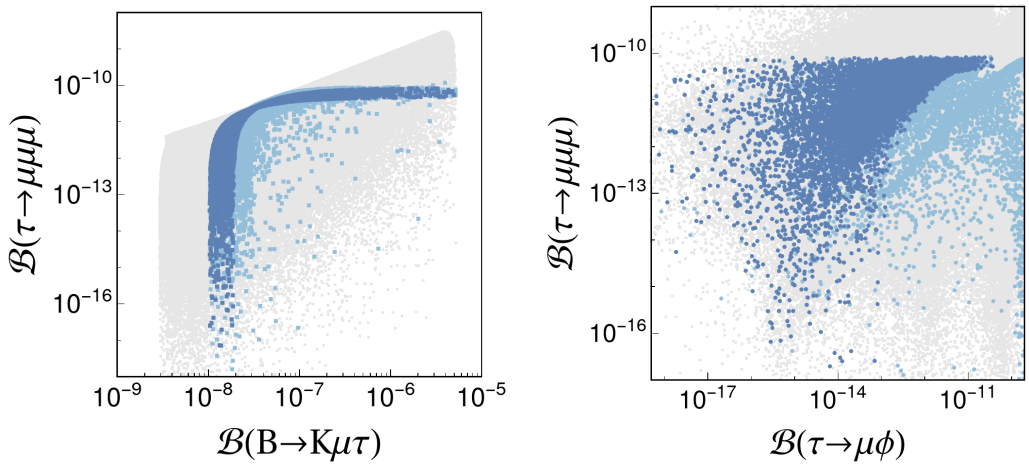

Figure 2.4: Left panel: correlation between $\mathcal{B}(B \to K\tau^\pm\mu^\mp)$ and $\tau \to 3\mu$. Right panel: correlation between $\tau \to 3\mu$ and $\tau \to \mu\phi$. Colour code as in Fig. 2.2.

$\Delta M_s/\Delta M_d$ Mixings in the $B_{d,s}$ sector, in particular the ratio $\Delta M_s/\Delta M_d$, represent a strong constraint (see ref. [234] for a recent discussion). Actually, it is

mainly this constraint that selects the mass hierarchy in eq. (2.32), the angles' hierarchy being instead mostly the result of $C_{9,10}^{\mu\mu}$ and x_D. The effective Hamiltonian relevant for these observables reads

$$\mathcal{H}_{\text{eff}}^{\Delta M_q} = (C_{\Delta M_q}^{\text{SM}} + \delta C_{\Delta M_q}^{q})(\bar{q}\gamma_L^\mu b)(\bar{q}\gamma_{\mu,L} b) + \text{h.c.}\,, \tag{2.47}$$

with q either d or s.[11] Using [106] to get $C_{\Delta M_q}^{\text{SM}}$ in our normalization, we obtain

$$\delta C_{\Delta M_q}^{s} = \frac{g_L^2}{8}\left(\frac{\cos^2 2\beta_D}{m_{G_1}^2} + \frac{\sin^2 2\beta_D}{m_{G_3}^2} - \frac{1}{m_{G_2}^2}\right)\,, \qquad \delta C_{\Delta M_q}^{d} = 0\,. \tag{2.48}$$

We can then write

$$\frac{\Delta M_s}{\Delta M_d} = \left|\frac{V_{ts}}{V_{td}}\right|^2 \xi^2 \frac{m_{B_s}}{m_{B_d}}\left|1 + \frac{\delta C_{\Delta M_q}^{s}}{C_{\Delta M_q}^{\text{SM}}}\right|\,. \tag{2.49}$$

Taking $\xi = 1.239(46)$ [235] as well as $|V_{ts}/V_{td}| = 4.58(24)$ from a Unitarity-Triangle fit using only quantities not affected by new physics [236],[12] one would obtain the SM prediction $(\Delta M_s/\Delta M_d)_{\text{SM}} = 33(4)$, perfectly consistent with the value $(\Delta M_s/\Delta M_d)_{\text{exp}} = 35.06(14)$ obtained from the mass differences reported in Ref. [238]. Our model's prediction for this ratio, normalized to the SM result, reads

$$\frac{\Delta M_s}{\Delta M_d} \Big/ \left(\frac{\Delta M_s}{\Delta M_d}\right)_{\text{SM}} \in [0.8, 1]\,. \tag{2.50}$$

In short, the model tends to predict a suppression of the order of 10-20%. However, such shift comes with an error of comparable size, about 12%, dominated by the CKM input, followed by the ξ input. This error at present prevents this observable from providing a stringent test of our framework. Such test will however be possible with improvements on fits to the unitarity triangle using only observables realistically unaffected by new physics, such as γ from $B \to DK^{(*)}$ (see in particular [239–241]). This highlights the well-known importance of improvements in such 'standard-candle' measurements.

$B \to K\nu\bar{\nu}$ Within our model, an explanation of $R_{K^{(*)}}$ implies new contributions to the $B \to K^{(*)}\nu\bar{\nu}$ decays (this is not specific to our model, see [242] for more general conclusions). The only part of the Hamiltonian eq. (1.1) relevant for $b \to s\nu\bar{\nu}$ is

$$\mathcal{O}_L^{\nu_i\nu_j} = \frac{e^2}{16\pi^2}(\bar{s}\gamma_L^\mu b)(\bar{\nu}_i\gamma_\mu(1-\gamma_5)\nu_j)\,. \tag{2.51}$$

[11] The Wilson coefficient $C_{\Delta M_q}$ is usually written C_1 in the literature, but differs from our C_1 given in eq. (1.2).

[12] A similar prediction may be obtained using the CKMfitter code [237].

We define the ratio $R^{(*)}_{\nu\nu} \equiv \mathcal{B}(B \to K^{(*)}\nu\nu)/\mathcal{B}(B \to K^{(*)}\nu\nu)^{\rm SM}$, where, as usual, a sum over the (undetected) neutrino species is understood. In our scenario this ratio is modified as follows

$$R^{(*)}_{\nu\nu} = \frac{\sum_{ij} |\delta_{ij} C_L^{\rm SM} + \delta C_L^{\nu_i\nu_j}|^2}{3|C_L^{\rm SM}|^2} = 1 + \frac{2C_L^{\rm SM}\sum_i \delta C_L^{\nu_i\nu_i} + \sum_{ij} |\delta C_L^{\nu_i\nu_j}|^2}{3|C_L^{\rm SM}|^2} \,, \tag{2.52}$$

where $C_L^{\rm SM} = -6.38(6)$ is the SM Wilson coefficient corresponding to the operator (2.51) [243]. The contributions to $R^{(*)}_{\nu\nu}$ induced by new physics are encoded in the Wilson coefficients $\delta C_L^{\nu_i\nu_j}$, which satisfy $\delta C_L^{\nu_i\nu_j} = 2\,\delta C_9^{\ell_i\ell_j}$ due to the different normalization between eqs. (1.2) and (2.51). Note that $R_{\nu\nu} = R^*_{\nu\nu}$ in our framework because of the absence of contributions to the right-handed counterpart of the operator $\mathcal{O}_L^{\nu_i\nu_j}$ [215]. By replacing eqs. (2.27)–(2.29) in eq. (2.52), we obtain

$$R^{(*)}_{\nu\nu} = 1 + \frac{1}{3|C_L^{\rm SM}|^2}\frac{v^4\pi^2}{\alpha_{\rm em}^2|\lambda_t|^2}\frac{g_L^4}{8}\left(\frac{\cos^2 2\beta_D}{m_{G_1}^4} + \frac{\sin^2 2\beta_D}{m_{G_3}^4} + \frac{1}{m_{G_2}^4}\right) . \tag{2.53}$$

Interestingly, the flavour-diagonal contributions satisfy $\delta C_L^{\nu_\mu\nu_\mu} = -\delta C_L^{\nu_\tau\nu_\tau}$ (whereas $C_L^{\nu_e\nu_e}$ is not modified), so that interference terms between the SM and NP *vanish*, and NP contributions only enter at second order in the small ratio $\delta C_L^{\nu_i\nu_j}/C_L^{\rm SM}$. Because of this feature, which is a consequence of the underlying $SU(2)_h$ symmetry, the strong experimental constraints $R^*_{\nu\nu} < 2.7$ [243, 244] do not pose, within our model, a challenge in the description of $b \to s$ anomalies. More quantitatively, using eq. (2.32) the $R^{\rm exp}_{\nu\nu}$ constraint can be translated into the bound

$$\frac{m_{G_{1,2}}^2}{g_L^2\sqrt{1+\cos^2 2\beta_D}} \geq (3.8\ {\rm TeV})^2 \,, \tag{2.54}$$

much weaker than the constraint derived in eq. (2.30). Note that the dependence on β_L disappears because of the sum over all neutrino species.

$\tau \to \mu\nu\bar{\nu}$ In spite of the horizontal gauge bosons G_a being electrically neutral, they can also contribute to processes that in the SM are generated by charged currents, for example $\ell \to \ell'\nu\bar{\nu}$ decays. The corresponding constraint will however be noticeably weaker than in the previous section were this contribution was induced by the charged currents, *cf.* eq. (2.12). More precisely, one can show that the following term appears in $\delta\mathscr{L}_{\rm eff}$

$$\delta\mathscr{L}_{\rm eff} \supset -\frac{g_L^2}{4}\left(\frac{\sin^2 2\beta_L}{m_{G_3}^2} + \frac{\cos^2 2\beta_L}{m_{G_1}^2} + \frac{1}{m_{G_2}^2}\right)(\bar{\nu}_\tau\gamma_L^\mu\nu_\mu)(\bar{\mu}\gamma_{\mu\,L}\tau) \,. \tag{2.55}$$

This contribution entails the following modification of $\mathcal{B}(\tau \to \mu\nu\nu)$ with respect to its SM prediction

$$\frac{\mathcal{B}(\tau \to \mu\nu\nu)}{\mathcal{B}(\tau \to \mu\nu\nu)^{\rm SM}} = 1 + \frac{g_L^2 v^2}{4}\left(\frac{\cos^2 2\beta_L}{m_{G_1}^2} + \frac{\sin^2 2\beta_L}{m_{G_3}^2} + \frac{1}{m_{G_2}^2}\right) + \ldots \,, \tag{2.56}$$

where, for simplicity, we only show the dominant term coming from the interference, whereas in the numerics we include also the order 2 sub-leading contributions. By using the experimental average $\mathcal{B}(\tau \to \mu\bar{\nu}_\mu\nu_\tau)_{\rm exp} = 17.33(5)\%$ [238] and the SM prediction $\mathcal{B}(\tau \to \mu\bar{\nu}_\mu\nu_\tau)^{\rm SM} = 17.29(3)\%$ [228], we obtain the following 2σ bound

$$\frac{m^2_{G_{1,2}}}{g_L^2\sqrt{1+\cos^2 2\beta_L}} \geq (1.6\ \text{TeV})^2\ , \tag{2.57}$$

which is weaker than the one derived in eq. (2.54) from the experimental limit on $\mathcal{B}(B \to K\nu\nu)$.

$D^0 \to \mu\mu$ Similarly to $D^0 - \bar{D}^0$ mixing, the CKM matrix induces a non-zero contribution to other charm-physics observables, most notably $D^0 \to \mu\mu$. We will show however that the induced modifications are too small to be observed with the sensitivity of the current experiments. The piece of eq. (2.17) describing four-fermion interactions of up-type quarks U' and charged leptons L', the primes denoting as usual the 'gauge' basis, reads

$$\delta\mathscr{L}^{UL}_{\rm eff} = -\sum_a \frac{g_L^2}{m^2_{G_a}}(\bar{U}'\gamma_L^\mu T^a U')(\bar{L}'\gamma_{\mu\,L}T^a L') \supset C^{\mu\mu}_{uc}\,(\bar{u}\gamma_L^\mu c)\,(\bar{\mu}\gamma_{\mu\,L}\mu)\ . \tag{2.58}$$

After manipulations entirely analogous to those leading to eq. (2.34), we obtain

$$C^{\mu\mu}_{uc} = -V^*_{cs}V_{us}\frac{g_L^2}{4}\left(\frac{\cos 2\beta_D\cos 2\beta_L}{m^2_{G_3}} + \frac{\sin 2\beta_D\sin 2\beta_L}{m^2_{G_1}}\right) + O(\lambda^3)\ . \tag{2.59}$$

We note again that the expansion in λ is only for illustrative purposes, and that in the numerics we use exact expressions. The corresponding branching ratio is then given by

$$\begin{aligned}\mathcal{B}(D^0 \to \mu^+\mu^-) = \frac{f_D^2 m_\mu^2 m_D}{32\pi\Gamma_D}\sqrt{1-\frac{4m_\mu^2}{m^2_{D^0}}} \times \Bigg[\frac{g_L^2}{4}\,|V^*_{cs}V_{us}| \\ \left(\frac{\cos 2\beta_D\cos 2\beta_L}{m^2_{G_3}} + \frac{\sin 2\beta_D\sin 2\beta_L}{m^2_{G_1}}\right)\Bigg]^2 + O(\lambda^4)\ .\end{aligned} \tag{2.60}$$

where we have neglected the small SM contribution. This expression should be confronted with the current experimental limit $\mathcal{B}(D^0 \to \mu^+\mu^-)_{\rm exp} < 6.2\times 10^{-9}$ [245]. However, by inspection of eq. (2.60) one sees that the relevant masses being bounded are a combination of $m^2_{G_1}/\sin 2\beta_D$ and $m^2_{G_3}$. Keeping in mind the parameter space in eq. (2.32), one concludes that eq. (2.60) provides a bound on the heavier scale, not the lighter one, and is thus irrelevant.

Further constraints Here we collect comments on further, potentially constraining, experimental information not discussed so far. A first comment deserves the decay $\tau \to \mu\gamma$, whose current experimental limit reads $\mathcal{B}(\tau \to \mu\gamma) \leq 4.4 \times 10^{-8}$ [246]. By construction, our model induces the required dipole interaction only at one loop. Therefore this decay is, at present, not very constraining within our model, as it suffers from a further loop and $\alpha_{\rm em}$ suppression with respect to the other τ LFV decays discussed before, whose predictions are summarized in the right panel of Fig. 2.4.

Further consideration deserve possible bounds coming from direct searches. To our knowledge, the most relevant analysis for our case is Ref. [247]. In particular, our model induces a contribution to $pp(s\bar{s}) \to \mu\mu$, which can be tested at the LHC by looking at the tails of dilepton distributions. Assuming that such distortions be the result of contact interactions of the kind

$$\mathscr{L}_{\rm eff} \supset \frac{C_{ij}^{D\mu}}{v^2}(\bar{d}^i \gamma_L^\mu d^j)(\bar{\mu}\gamma_{\mu L}\mu) \ , \tag{2.61}$$

with an effective scale well above the typical momentum exchange in the process, Ref. [247] quotes a present-day limit on $|C_{ss}^{D\mu}|$ of around 1×10^{-2}. In our model, this coefficient is of order $g_L^2 \cdot v^2/\min(m_{G_a}^2)$, that we can bound with $1 \times (0.246/5)^2 \simeq 2 \times 10^{-3}$, see e.g. y-axis scale on the left panel of Fig. 2.2. It is true that in some parts of our parameter space – with very low $\min(m_{G_a})$ and sizeable g_L – there may be distortions with respect to the effective-theory description in Ref. [247]. While this aspect may warrant further investigation, we believe that the above argument provides a robust order-of-magnitude assessment of the constraint.

2.2.4 Conclusions

Explaining $b \to c$ or $b \to s$ anomaly requires a mechanism that provides a large coupling between lepton and quark currents while avoiding accordingly large contributions in coupling between quark currents and between lepton currents. This is naturally achieved in model based on leptoquarks where the undesirable contributions are loop suppressed. In this section we take a different approach. The aforementioned mechanism can be achieved through the consideration of a 'horizontal' group, $SU(2)$ being the smallest one that may be at play. We accordingly invoke the possibility of a gauged such symmetry, $SU(2)_h$, with all the left-handed $2^{\rm nd}$- and $3^{\rm rd}$-generation fermions universally charged under the corresponding group – in the 'gauge' basis.

After integrating out the heavy $SU(2)_h$ bosons, one generates all sorts of $J_{q,\ell} \times J_{q,\ell}$ amplitudes. However, assuming degenerate masses for the horizontal bosons, and in the absence of mixing between the two heavier generations and the lighter one, the assumed symmetry would make $J_q \times J_q$ and $J_\ell \times J_\ell$ amplitudes exactly flavour-diagonal, in the fermion mass eigenstate basis. This property prevents dangerous tree-level contributions to processes such as meson mixings and

purely leptonic flavour-violating transitions. In reality, such contributions are not exactly zero because of CKM-induced mixing across all the generations. The most constraining of these effects turns out to be the mass difference in the $D^0 - \bar{D}^0$ system, ΔM_D.

However, one can accomplish a successful description of $b \to s$ deviations as well as of all constraints, by advocating a splitting of the horizontal-boson masses – per se a plausible possibility – in particular a configuration with two mass-degenerate gauge bosons hierarchically lighter than the third one.

This scenario has, by construction, distinctive signatures in $b \to s\ell\ell'$ decays. In particular, $B \to K\tau^\pm\mu^\mp$ is predicted in the range

$$1.0 \times 10^{-8} \lesssim \mathcal{B}(B \to K\mu^+\tau^-) + \mathcal{B}(B \to K\mu^-\tau^+) \lesssim 5.0 \times 10^{-6} \,, \tag{2.62}$$

with the $\tau^+\mu^-$ and $\tau^-\mu^+$ modes in general differing by a sizeable amount that could have either sign. Besides, the small number of parameters involved establishes clear-cut correlations between semi-leptonic LFV decays of B mesons and LFV decays involving only leptons, in particular a triple correlation between $B(B \to K\mu^\pm\tau^\mp)$, $B(\tau \to 3\mu)$ and $B(\tau \to \mu\phi)$.

At this point, our framework still contains several open questions.

First, in order to be fully calculable beyond tree level, the model still requires specification of the scalar sector that accomplishes the spontaneous breaking of the $SU(2)_h$ symmetry. While we already commented on this completion to address the issue of Yukawa matrices, answering the question in full would introduce a degree of model dependence, and probably requires more data. We therefore limited our study to the effective model, assuming that such scalar a sector simply exists.

As we also restricted our study to $b \to s$ anomalies, an interesting follow-up would be to consider an appropriate variation of the mechanism that may also explain charged-current discrepancies. We do not see however how such an extension could avoid introducing relations between up-type and down-type fermion chiral rotations, hardly compatible with our original argument.

Finally, another interesting question is whether a suitable extension of our framework may include a candidate for thermal Dark Matter. This possibility will be discussed in section 3.2.

3

Theory paths forward

Contents

This chapter brings together two lines of research that can serve as a guide for future theoretical models and experimental studies. In section 3.1, we will perform fits of the current experimental data, both in the weak effective theory and in the SM effective field theory. These fits will allow to scrutinize the viable scenarios that account for the B-anomalies and justify the study performed in the previous chapters. We will show in particular that a simplified model based on the U_1 leptoquark can provide an excellent description of data. In section 3.2, we will embed this model in a UV-complete theory and show that it can contain a candidate for Dark Matter. Such candidate can also be obtained in link with the model discussed in chapter 2 in a construction discussed in the last section.

3.1 A global analysis after Moriond 2019

In this section we present the implication of the current experimental status using effective field theories at the weak scale and at higher scale. Although we already

discussed the impact of $R_{K^{(*)}}$ measurements on C_9 and C_{10} as well as constraints due to low energy observables in previous sections, a more systematic study will allow us to identify the scenarios that lead to a good description of data and discuss a possible realization in term of a simplified new-physics model, while evaluating the impact experimental measurements.

To perform this study, we use a global likelihood function in the space of Wilson coefficients both in the weak effective theory (WET) eq. (1.2) and in the SM effective theory (SMEFT) (see *e.g.* [248] for a review). To achieve a fully model-independent analysis, all Wilson coefficients should be floated at a time. This approach, computationnaly challenging and difficult to interprete, has been discussed in the context of a Bayesian approach [96, 249, 250] and in earlier literature (see e.g. [123, 251, 252]). Here we will focus on theoretically motivated scenarios based on one or two varying Wilson coefficients.

This analysis, based on [121], proceeds in two steps:

1. We first investigate the Wilson coefficients of the WET eq. (1.2). This analysis can be seen as an update of earlier analyses (see *e.g.* [92–97]) where the considered scenarios have already been discussed. Such analysis is completely general within the assumed one- or two-coefficient assumption, and, as already mentioned, barring new particles lighter than the b quark.

2. Next, we embed these results into the SMEFT at a scale Λ above the electroweak scale. This is based on the additional assumptions that there are no new particles beneath Λ and that EW symmetry breaking is approximately linear (see e.g. [253]). This allows us to correlate NP effects in $b \to s\ell\ell$ with other sectors like EW precision tests or $b \to c$ transitions within a general and common formalism (cf. [254–257]). This formalism will also permit to put on more solid ground the conclusions of sections 1.4 and 2.1.1.

An update of the analyses already performed in these two basis was made necessary by the publication of several results, mainly presented during the Moriond-2019 conference. These results, hereafter referred to as Moriond-2019 results, includes

- LHCb's update of the R_K measurement [51];
- The measurement of R_{K^*} by Belle [53]. Averaged over $B^\pm$ and B^0 decays, the measured R_{K^*} values at low and high q^2 are

$$R_{K^*} = \begin{cases} 0.90^{+0.27}_{-0.21} \pm 0.10\,, & \text{for } 0.1\,\text{GeV}^2 < q^2 < 8\,\text{GeV}^2\,, \\ 1.18^{+0.52}_{-0.32} \pm 0.10\,, & \text{for } 15\,\text{GeV}^2 < q^2 < 19\,\text{GeV}^2\,. \end{cases} \tag{3.1}$$

  Given their sizable uncertainties, these values are compatible with both the SM predictions and previous results on R_{K^*} from LHCb [52].

- The average of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurements eq. (1.34).

The numerical analysis is entirely based on open-source software, notably the global likelihood in Wilson coefficient space provided by the `smelli` package [256], built on `flavio` [258] and `wilson` [259].

Comment on uncertainties and observables Before discussing the scenarios, two comments are in order.

The first one concerns the treatment of theoretical uncertainties. When they are sizable, these uncertainties are accounted for by computing their covariance matrix within the SM and combining it with the experimental uncertainties (approximated as symmetrized Gaussians see *e.g.* [92,93,123,251,260,261]). The main assumption in this approach is that the sizes of theoretical uncertainties are weakly dependent on NP, which we checked for the observables included. This approach was first applied to $b \to s\ell\ell$ transitions in [123].

As discussed in the first chapter, the theoretical uncertainties in exclusive B-decay observables stem mainly from hadronic form factors. In this study, B to light vector meson form factors are taken from [61] and from [63] for $B \to K$. Other unknown non-factorizable effects are parametrized as in [61, 123, 258] (and are compatible with more sophisticated approaches [71, 262]). Additional parametric uncertainties (*e.g.* from CKM matrix elements) are based on `flavio` v1.3 with default settings [258]. More details on the statistical approach and the list of observables and measurements included are given in [256].

The second comment concerns the observables that are included. The following study is based on the global likelihood described in [256], where as many observables sensitive to the Wilson coefficients as possible are included (other approachs, like [92], limit the observables to the one sensitive to the $b \to s\ell\ell$ transition). This means *e.g.* that we also include all the observables sensitive to the $b \to s\gamma, g$ dipole transitions studied in [263]. In addition, the global likelihood also includes observables that do not directly depend on the Wilson coefficients of interest but whose theory uncertainties are strongly correlated with those of the directly dependent observables. This is in particular relevant for the $b \to s\mu\mu$ observables. In our figures, we indicate the set of observables consisting of $b \to s\mu\mu$, $b \to s\gamma, g$, and other correlated observables as "$b \to s\mu\mu$ & corr. obs.".

3.1.1 Effective-theory analysis

WET

The weak basis operators are defined by the effective Hamiltonian eq. (1.1). For this analysis, we have omitted however the chromomagnetic and the four-quark operators. The latter can contribute via one-loop matrix elements to $b \to s\ell\ell$ processes, but their dominant effects typically stem from renormalization group evolution (RGE) above the scale μ_b, and we will discuss these effects in the SMEFT framework in the next section. For the same reason, we have constrained the sum

over lepton flavours to e and μ: semi-tauonic WET operators can contribute via QED RG mixing, but their direct matrix elements are subleading [264].

Scenarios with a single Wilson coefficient Let us first focus on scenarios where only one NP Wilson coefficient (or a single linear combination motivated by UV scenarios) is non-zero. The best-fit values, 1 and 2σ ranges, and pulls for several such scenarios are listed in Table 3.1.[1]

Coeff.	1σ	2σ	pull
$C_9^{\mu\mu}$	$-0.97\,^{+0.16}_{-0.15}$	$[-1.27, -0.65]$	5.9σ
$C_9^{\prime\mu\mu}$	$0.14\,^{+0.18}_{-0.17}$	$[-0.20, +0.51]$	0.8σ
$C_{10}^{\mu\mu}$	$0.75\,^{+0.14}_{-0.13}$	$[+0.48, +1.03]$	5.7σ
$C_{10}^{\prime\mu\mu}$	$-0.24\,^{+0.12}_{-0.12}$	$[-0.49, +0.00]$	2.0σ
$C_9^{\mu\mu} = C_{10}^{\mu\mu}$	$0.20\,^{+0.16}_{-0.14}$	$[-0.09, +0.52]$	1.4σ
$C_9^{\mu\mu} = -C_{10}^{\mu\mu}$	$-0.53\,^{+0.08}_{-0.08}$	$[-0.69, -0.37]$	6.6σ
C_9^{ee}	$0.93\,^{+0.24}_{-0.27}$	$[+0.40, +1.42]$	3.5σ
$C_9^{\prime ee}$	$0.39\,^{+0.26}_{-0.34}$	$[-0.27, +0.95]$	1.2σ
C_{10}^{ee}	$-0.83\,^{+0.23}_{-0.22}$	$[-1.28, -0.37]$	3.6σ
$C_{10}^{\prime ee}$	$-0.27\,^{+0.25}_{-0.30}$	$[-0.84, +0.26]$	1.1σ
$C_9^{ee} = C_{10}^{ee}$	$-1.49\,^{+0.31}_{-0.30}$	$[-2.05, -0.79]$	3.2σ
$C_9^{ee} = -C_{10}^{ee}$	$0.47\,^{+0.12}_{-0.14}$	$[+0.20, +0.73]$	3.5σ
$(C_S^{\mu\mu} = -C_P^{\mu\mu}) \times \text{GeV}$	$-0.006\,^{+0.003}_{-0.003}$	$[-0.014, -0.001]$	2.8σ
$\left(C_S^{\prime\mu\mu} = C_P^{\prime\mu\mu}\right) \times \text{GeV}$	$-0.006\,^{+0.003}_{-0.003}$	$[-0.014, -0.001]$	2.8σ

Table 3.1: Best-fit values, 1 and 2σ ranges, and pulls between the best-fit point and the SM point for scenarios with NP in a single Wilson coefficient (or Wilson coefficient combination). For the scalar Wilson coefficients, we show the SM-like solution, while also a sign-flipped solution is allowed, see [265].

The result of this first approach can be interpreted as follows:

- As already known from other analysis and discussed in the previous chapters, two scenarios stand out, namely a shift to $C_9^{\mu\mu}$ by approximately -25% of

[1] For the 1D scenarios, the pull in σ is defined as pull $= \sqrt{\Delta\chi^2}$, with $-\frac{1}{2}\Delta\chi^2 = \ln L(\vec{0}) - \ln L(\vec{C}_{\text{best fit}})$, where $L(\vec{C})$ is the global likelihood in the space of NP Wilson coefficients.

its SM value, or a shift to the combination $C_9^{\mu\mu} = -C_{10}^{\mu\mu}$ by approximately -15% of its SM value. However, due to the most recent measurements of $\mathcal{B}(B_s^0 \to \mu\mu)$ eq. (1.11), the second scenario is preferred to the first one. Indeed, as visible in eq. (1.8), C_9 alone cannot explain the $\sim 2\sigma$ tension in this channel. This suppression can be accounted for by the scalar Wilson coefficients $C_{S/P}^{\mu\mu}$ and $C_{S/P}^{\prime\mu\mu}$, but the two combinations possible in theories that have SMEFT as their EW-scale limit, $C_S^{\mu\mu} = -C_P^{\mu\mu}$ and $C_S^{\prime\mu\mu} = C_P^{\prime\mu\mu}$ [122], remain severely constrained.

- New physics in $C_{10}^{\mu\mu}$ alone also improves the agreement between theory and data considerably. However, tensions in $B \to K^*\mu\mu$ angular observables remain in this scenario.

- Muonic scenarios with right-handed currents on the quark side, $C_9^{\prime\mu\mu}$ and $C_{10}^{\prime\mu\mu}$, or the lepton side, $C_9^{\mu\mu} = C_{10}^{\mu\mu}$, do not lead to a good description of the data.

- Scenarios with NP in ee Wilson coefficients only, while able to accommodate the discrepancies in $R_{K^{(*)}}$, do not help for the rest of the data.

Scenarios with a pair of Wilson coefficients The results in Table 3.1 suggest that NP in both $C_9^{\mu\mu}$ and $C_{10}^{\mu\mu}$ ought to give an excellent fit to the data. The left

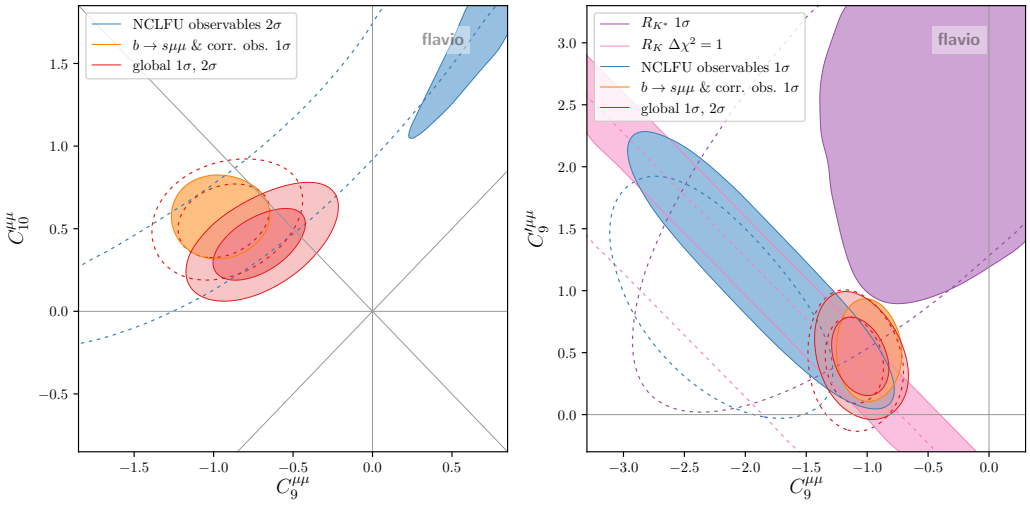

Figure 3.1: Likelihood contours of the global fit and several fits to subsets of observables (see text for details) in the plane of the WET Wilson coefficients $C_9^{\mu\mu}$ and $C_{10}^{\mu\mu}$ (left), and $C_9^{\mu\mu}$ and $C_9^{\prime\mu\mu}$ (right). Solid (dashed) contours include (exclude) the Moriond-2019 results for R_K and R_{K^*}. As R_K only constrains a single combination of Wilson coefficients in the right plot, its 1σ contour corresponds to $\Delta\chi^2 = 1$. For the other fits, 1 and 2σ contours correspond to $\Delta\chi^2 \approx 2.3$ and 6.2, respectively.

plot of Fig. 3.1 shows the best fit regions in the $C_9^{\mu\mu}$ - $C_{10}^{\mu\mu}$ plane. The orange regions correspond to the 1σ constraints from $b \to s\mu\mu$ observables (including $B_s \to \mu^+\mu^-$) and observables whose uncertainties are correlated with those of the $b \to s\mu\mu$ observables (as discussed earlier in the second general comment). In blue we show regions corresponding to the 1σ (right plot) and 2σ (left plot) constraints from the *neutral-current LFU* (NCLFU) observables R_K, R_{K^*}, $D_{P_4'}$, and $D_{P_5'}$[2]. In the right plot, the 1σ constraints from only R_K (purple) and only R_{K^*} (pink) are shown. The combined 1 and 2σ region is shown in red. The dotted contours indicate the situation without the Moriond-2019 results for R_K and R_{K^*}. The best fit point $C_9^{\mu\mu} \simeq -0.73$ and $C_{10}^{\mu\mu} \simeq 0.40$ has a $\sqrt{\Delta\chi^2} = 6.6$, which, corrected for the two degrees of freedom, corresponds to a pull of 6.3σ. In this scenario a slight tension between R_K and R_{K^*} remains, as it predicts $R_K \simeq R_{K^*}$ while the data seems to indicate $R_K > R_{K^*}$. In addition, there is also a slight tension between the fit to NCLFU observables and the fit to $b \to s\mu\mu$ ones, especially in the $C_9^{\mu\mu}$ direction.

Overall, we find a similarly good fit of the data in a scenario with NP in $C_9^{\mu\mu}$ and $C_9^{\prime\mu\mu}$. The scenario is shown in the right plot of Fig. 3.1. The best fit values for the Wilson coefficients are $C_9^{\mu\mu} \simeq -1.06$ and $C_9^{\prime\mu\mu} \simeq 0.47$. The $\sqrt{\Delta\chi^2} = 6.4$ corresponds to a pull of 6.0σ. Interestingly, in this scenario a non-zero $C_9^{\prime\mu\mu}$ is preferred at the 2σ level. The right-handed quark current allows one to accommodate the current experimental results for the LFU ratios, $R_K > R_{K^*}$. This scenario cannot address the tension in $\mathcal{B}(B_s \to \mu^+\mu^-)$ as it predicts $\mathcal{B}(B_s \to \mu^+\mu^-) = \mathcal{B}(B_s \to \mu^+\mu^-)_{\rm SM}$.

Other two-coefficient scenarios (including dipole coefficients, scalar coefficients, and electron specific semileptonic coefficients) show a weaker agreement with data. They are discussed in the appendix of Ref. [121].

Universal *vs.* non-universal Wilson coefficients The fit in $C_9^{\mu\mu}$ and $C_{10}^{\mu\mu}$ (Fig. 3.1 left) shows a tension between NCLFU observables and the fit to $b \to s\mu\mu$ ones. This tension is mainly due to the updated of the $R_{K^{(*)}}$ measurements, which are closer to the SM prediction than the Run I results. This tension, which is mainly in the $C_9^{\mu\mu}$ direction, suggests the presence of a lepton flavour universal new physics contribution that affects $b \to s\mu\mu$ observables but not NCLFU ones. We therefore adopt the following decomposition[3]

$$C_9^{\mu\mu} = \Delta C_9^{\mu\mu} + C_9^{\rm univ.}\,, \tag{3.2}$$

$$C_9^{ee} = C_9^{\tau\tau} = C_9^{\rm univ.}\,, \tag{3.3}$$

[2] These quantities are defined as $D_{P_{4,5}'} = P_{4,5}'(B \to K^*\mu^+\mu^-) - P_{4,5}'(B \to K^*e^+e^-)$ (the observables $P_{4,5}'$ are defined in [64]).

[3] Such decomposition was adopted for the first time in [266], to which we refer the reader for additional scenarios beyond the one we consider. We note that a shift in $C_{10}^{\rm univ.}$ would not produce a good overall fit. This may be appreciated from Fig. 3.1 (left). A $C_{10}^{\rm univ.}$ shift would only move the (yellow) $b \to s\mu\mu$ region vertically, hence it would not help reach better agreement with the (blue) NCLFU region. We therefore set non-muonic C_{10} contributions to zero for simplicity.

$$C_{10}^{\mu\mu} = -\Delta C_9^{\mu\mu}\,, \tag{3.4}$$
$$C_{10}^{ee} = C_{10}^{\tau\tau} = 0\,. \tag{3.5}$$

The corresponding fit is displayed in Fig. 3.2, where the best fit values are

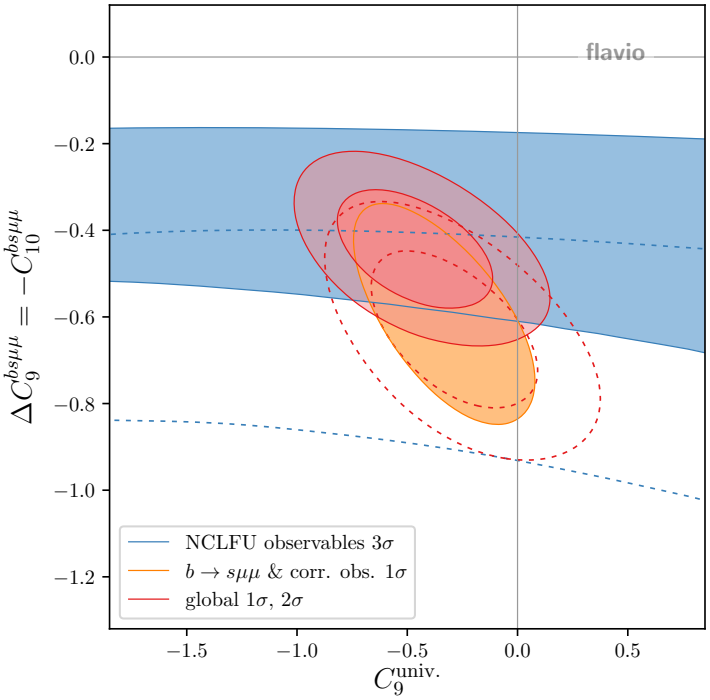

Figure 3.2: Likelihood contours from NCLFU observables ($R_{K^{(*)}}$ and $D_{P'_{4,5}}$), $b \to s\mu\mu$ observables, and the global fit in the plane of a lepton flavour universal contribution to $C_9^{\text{univ.}} \equiv C_9^{\ell\ell}, \forall \ell$, and a muon-specific contribution to the linear combination $C_9 = -C_{10}$ (see text for details). Solid (dashed) contours include (exclude) the Moriond-2019 results for R_K and R_{K^*}.

$C_9^{\text{univ.}} = -0.49$ and $\Delta C_9^{\mu\mu} = -0.44$ with a $\sqrt{\Delta\chi^2} = 6.8$ that corresponds to a pull of 6.5σ. From this figure, it is clear that the updated values of $R_{K^{(*)}}$ favour a non-zero lepton flavour universal contribution to C_9 in this scenario.

Before studying how such contributions can arise in the SMEFT, one qualification is in order. The new effects discussed so far, and especially the $C_9^{\text{univ.}}$ contribution, can possibly be due to a hadronic SM effect that couples to the lepton current via a virtual photon, for example charm-loop effects at low q^2 and resonance effects at high q^2, see *e.g.* [73,75,267]. In this analysis, this is accounted for by the uncertainty attached to the relevant observables that contribute to the (yellow) $b \to s\mu\mu$ region in Fig. 3.2 (see [123] for a description). Doubling non-form-factor hadronic uncertainties is for example enough to accommodate an effect in $C_9^{\text{univ.}}$ in the range $-[0.1, 0.6]$ [92]. With the above qualification in mind, in the

following sections we will address the question whether the above scenario, with a muon-specific contribution to $\Delta C_9^{\mu\mu} = -C_{10}^{\mu\mu}$ plus a universal contribution to $C_9^{\text{univ.}}$, can be justified from the UV point of view, assuming that both contributions are due to new physics.

The global picture in the SMEFT

We now turn to the discussion of above results within the SMEFT. In contrast to the discussion in WET at the b-quark scale, more Wilson coefficients become relevant in SMEFT due to RG mixing above [268–270] and below [271, 272] the EW scale. Due to the pattern of Wilson coefficients preferred by the global fit, we focus on SMEFT Wilson coefficients that either contribute to the semi-muonic Wilson coefficients in the form $C_9^{\mu\mu} = -C_{10}^{\mu\mu}$ or induce a LFU effect in $C_9^{\ell\ell}$.

The direct matching contributions to $C_{9,10}$ at the EW scale are well known [243, 273][4]

$$C_9^{\ell_i\ell_i} = \frac{\pi}{\sqrt{2}\, G_F \lambda_t\, \alpha_{\text{em}}} \left([C_{Qe}]_{23ii} + [C_{LQ}^{(1)}]_{ii23} + [C_{LQ}^{(3)}]_{ii23} - \zeta c_Z \right) , \tag{3.6}$$

$$C_{10}^{\ell_i\ell_i} = \frac{\pi}{\sqrt{2}\, G_F \lambda_t\, \alpha_{\text{em}}} \left([C_{Qe}]_{23ii} - [C_{LQ}^{(1)}]_{ii23} - [C_{LQ}^{(3)}]_{ii23} + c_Z \right) , \tag{3.7}$$

where the Z penguin coefficient c_Z is

$$c_Z = [C_{\varphi Q}^{(1)}]_{23} + [C_{\varphi Q}^{(3)}]_{23} , \tag{3.8}$$

and $\zeta = 1 - 4s_w^2 \approx 0.08$ is the accidentally suppressed vector coupling of the Z to charged leptons. The corresponding operators are given by [275]

$$[O_{Qe}]_{23ii} = (\bar{Q}_2 \gamma_\mu Q_3)(\bar{e}_i \gamma^\mu e_i) , \tag{3.9}$$

$$[O_{LQ}^{(1)}]_{ii23} = (\bar{L}_i \gamma_\mu L_i)(\bar{Q}_2 \gamma^\mu Q_3) , \quad [O_{LQ}^{(3)}]_{ii23} = (\bar{L}_i \gamma_\mu \sigma^a L_i)(\bar{Q}_2 \gamma^\mu \sigma^a Q_3) , \tag{3.10}$$

$$[O_{\varphi Q}^{(1)}]_{23} = (\varphi^\dagger i \overleftrightarrow{D}_\mu \varphi)(\bar{Q}_2 \gamma^\mu Q_3) , \quad [O_{\varphi Q}^{(3)}]_{23} = (\varphi^\dagger i \overleftrightarrow{D}^a_\mu \varphi)(\bar{Q}_2 \gamma^\mu \sigma^a Q_3) , \tag{3.11}$$

where Q_i, and L_i are the left-handed $SU(2)_L$ doublet quarks and leptons and e_i are the right-handed lepton singlets, φ is the SM Higgs doublet, and σ^a are the Pauli matrices. The operators discussed in sections 1.4 and 2.1.1 therefore correspond to a combination of $[O_{LQ}^{(1,3)}]_{ijkl}$. Eqs. (3.6) and (3.7) are the lepton-universal generalizations of eq. (2.4), the minus sign being due to the definition of the operators.

Equations (3.6) and (3.7) highlight the well-known fact that a LFU contribution to $C_{9,10}$ induced by the SMEFT coefficients $[C_{\varphi Q}^{(1,3)}]_{23}$ (yielding a flavour-changing

[4] To standardise notations, we will keep L and Q for the fields above the EW scale and ℓ and q below. This notation differs from [121], and from the usual SMEFT notation, where l is used for the leptons above the EW scale. Furthermore, we will work in a basis where generation indices for RH quarks are taken to coincide with the mass basis [274], which can be done without loss of generality.

$\bar{s}bZ$ coupling) is not preferred by the data since it leads to $|C_{10}^{\ell_i\ell_i}| \gg |C_9^{\ell_i\ell_i}|$. Likewise, the coefficient $[C_{Qe}]_{23ii}$ alone leads to $C_9^{\ell_i\ell_i} = C_{10}^{\ell_i\ell_i}$ that is in poor agreement with the data as well. Thus, if the dominant NP effect in $C_{9,10}^{\mu\mu}$ does *not* stem from an RG effect but a direct matching contribution, it must involve one of the SMEFT Wilson coefficients $[C_{LQ}^{(1,3)}]_{2223}$. In addition to the dominant $[C_{LQ}^{(1,3)}]_{2223}$ contribution, there can also be contributions from other Wilson coefficients. However, such additional contributions need, in general, be small (see *e.g.* [249]).[5]

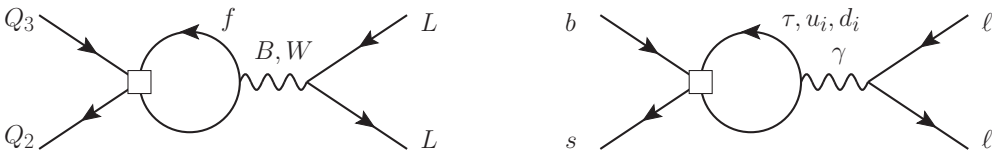

Figure 3.3: Diagrams inducing a contribution to C_9 through RG running above (left panel) and below (right panel) the EW scale. A sizeable contribution to C_9 is obtained when $f = u_{1,2}, d_{1,2,3}$ or L_3, see text for details.

Apart from the direct matching contributions, additional SMEFT Wilson coefficients can induce a contribution to C_9 at the b mass scale through RG evolution above or below the EW scale, as pictured in Fig. 3.3. In view of the size of the effect preferred by the data, we can identify three qualitatively different effects that can play a role:

- Wilson coefficients $[C_{eu}]_{2233}$ and $[C_{Lu}]_{2233}$ of the ditop-dimuon operators $[O_{eu}]_{2233} = (\bar{e}_2\gamma_\mu e_2)(\bar{u}_3\gamma^\mu u_3)$ and $[O_{Lu}]_{2233} = (\bar{L}_2\gamma_\mu L_2)(\bar{u}_3\gamma^\mu u_3)$ that induce a contribution to $C_9^{\mu\mu}$ from electroweak running above the EW scale. However, this solution is seriously challenged by EW precision tests [255].[6]

- Semitauonic operators $[O_{LQ}^{(1)}]_{3323}$, $[O_{LQ}^{(3)}]_{3323}$, or $[O_{Qe}]_{2333}$ defined in (3.9) and (3.10), that induce a LFU contribution to $C_9^{\ell\ell}$ from gauge-induced running both above and below the EW scale [264, 276].

- Four-quark operators ($[O_{QQ}^{(1,3)}]_{23ii}$, $[O_{Qu}^{(1)}]_{23ii}$ and $[O_{Qd}^{(1)}]_{23ii}$) that also induce a LFU contribution to $C_9^{\ell\ell}$ analogously to the semitauonic ones [277]. These operators, discussed in [121], are almost completely excluded by several observables, especially meson mixings and direct CP violation in K decays. Although a C_9 effect from $[O_{Qu}^{(1)}]_{2322}$ and $[O_{Qd}^{(1)}]_{2333}$ is not strongly constrained,

[5] Exceptions can occur if one admits cancellations. An example is a scenario with $[C_{LQ}^{(1,3)}]_{2223}$ and $[C_{Qe}]_{2322}$ of comparable size and like sign [249] thus cancelling each other in $C_{10}^{\mu\mu}$ and adding up in $C_9^{\mu\mu}$ (see Eqs. (3.6)-(3.7)). We note that this cancellation would need to occur at the m_b scale, not at the matching scale. Such scenario gives effectively our $C_9^{\mu\mu}$-only solution in Table 3.1.

[6] This conclusion holds barring the possibility of additional tree-level contributions from other Wilson coefficients that happen to cancel, at the EW scale, the RGE-induced contributions to EW precision observables. We are not aware of literature where such a possibility has been invoked on theoretical grounds, and we do not consider it further.

simplified tree-level NP models are tightly challenged by di-jet resonance searches at the LHC due to a TeV-scale mass mediator with large couplings to quarks.

As emphasize in section 2.1.1, semitauonic operators are particularly interesting as they potentially allow for a simultaneous explanation of the anomalies in neutral-current $b \to s$ transitions and in charged-current $b \to c$ transitions involving taus [256, 276]. We focus on this case in the following.

LFU C_9 from running of semi-tauonic operators

As we will see, a large value for $[C_{LQ}^{(3)}]_{3323}$, that can explain the hints for LFU violation in charged-current $b \to c$ transitions, can indeed induce a LFU effect in C_9 that goes in the right direction to solve the $b \to s\mu\mu$ anomalies in branching ratios and angular observables. An additional contribution to $[C_{LQ}^{(1,3)}]_{2223}$ of similar size can accommodate the deviations in R_K and R_{K^*}. However, the linear combination $[C_{LQ}^{(1)}]_{3323} - [C_{LQ}^{(3)}]_{3323}$ generates a sizable contribution to $B \to K^{(*)}\nu\bar{\nu}$ decays [243] that is constrained by B-factory searches for these modes, such models are only viable if the semi-tauonic singlet and triplet Wilson coefficients are approximately equal[7].

Fig. 3.4 shows the likelihood contributions from $R_{D^{(*)}}$, NCLFU observables, $b \to s\mu\mu$ observables, and the global likelihood in the space of the two Wilson coefficients $[C_{LQ}^{(1)}]_{3323} = [C_{LQ}^{(3)}]_{3323}$ and $[C_{LQ}^{(1)}]_{2223} = [C_{LQ}^{(3)}]_{2223}$ at the renormalization scale $\mu = 2\,\text{TeV}$. It is interesting to note that before the Moriond 2019 updates (indicated by the dashed contours), for a purely muonic solution with $[C_{LQ}^{(1,3)}]_{3323} = 0$ (corresponding to the vertical axis), the best-fit values for NCLFU and $b \to s\mu\mu$ data were in perfect agreement with each other (even though $R_{D^{(*)}}$ cannot be explained in this case). Including the $R_{K^{(*)}}$ updates, the best-fit point of the NCLFU and $b \to s\mu\mu$ data instead lies in the region with non-zero semitauonic Wilson coefficients, just as required to explain the $R_{D^{(*)}}$ anomalies. In fact, the agreement between the 1σ regions for $R_{K^{(*)}}$ & $D_{P'_{4,5}}$, $R_{D^{(*)}}$, and $b \to s\mu\mu$ *improves* compared to the case without the $R_{K^{(*)}}$ updates. We note that a further improvement of the fit is achieved by taking into account the Moriond 2019 update of $R_{D^{(*)}}$ by Belle [278, 279], which moves the 1σ region for $R_{D^{(*)}}$ slightly closer to the SM value, exactly to the region where the contours of NCLFU and $b \to s\mu\mu$ observables overlap. The best fit values in this scenario are $[C_{LQ}^{(1,3)}]_{3323} = -5.0 \times 10^{-2}$ TeV^{-2} and $[C_{LQ}^{(1,3)}]_{2223} = 3.9 \times 10^{-4}$ TeV^{-2} with a $\sqrt{\Delta\chi^2} = 8.1$ that corresponds to a pull of 7.8σ. The pull is considerably larger in the present scenario than in those discussed before since it can also explain discrepancies in $b \to c$ transitions.

It is interesting to use the global fit in this scenario as the basis for predictions of several observables that are sensitive to the Wilson coefficients $[C_{LQ}^{(1,3)}]_{3323}$ and

[7] Note that exact equality is not preserved by the RG evolution in SMEFT.

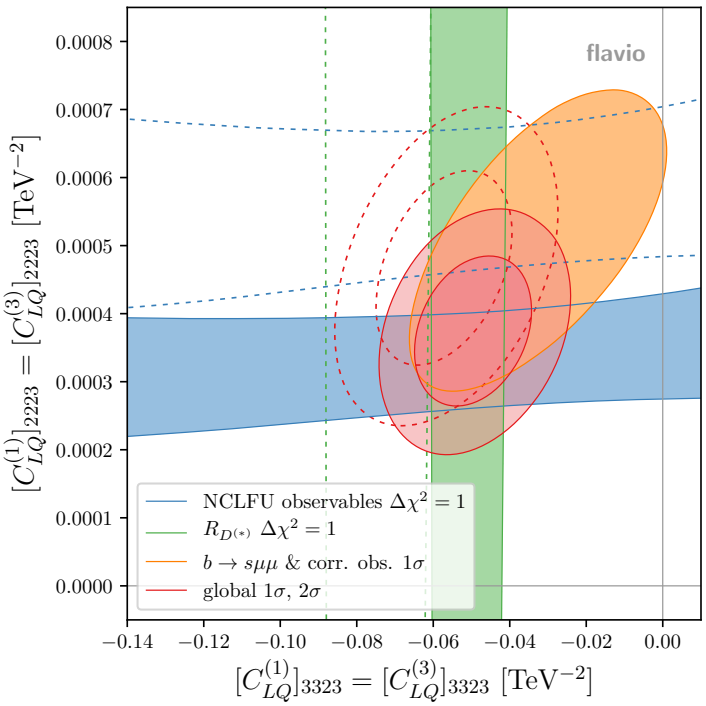

Figure 3.4: Likelihood contours from $R_{D^{(*)}}$, NCLFU observables ($R_{K^{(*)}}$ and $D_{P'_{4,5}}$), and $b \to s\mu\mu$ observables in the space of the two SMEFT Wilson coefficients $[C_{LQ}^{(1)}]_{3323} = [C_{LQ}^{(3)}]_{3323}$ and $[C_{LQ}^{(1)}]_{2223} = [C_{LQ}^{(3)}]_{2223}$ at 2 TeV. All other Wilson coefficients are assumed to vanish at 2 TeV. Solid (dashed) contours include (exclude) the Moriond-2019 results for R_K, R_{K^*}, R_D, and R_{D^*}. For sets of data that effectively constrain only a single Wilson coefficient (namely $R_{D^{(*)}}$ and NCLFU observables), 1σ contours correspond to $\Delta\chi^2 = 1$. For the other data ($b \to s\mu\mu$ and the global likelihood), 1 and 2σ contours correspond to $\Delta\chi^2 \approx 2.3$ and 6.2, respectively.

$[C_{LQ}^{(1,3)}]_{2223}$ and are supposed to be measured with higher precision in the near future. We collect predictions for LFU ratios, angular observables, and branching ratios in B and B_s decays in Table 3.2. This table confirms the well-known discovery potential of LFU ratios [215] but also underlines the interest of decay with taus. We note that the very large enhancement of $\mathcal{B}(B_s \to \tau^+\tau^-)$ put this decay at the edge of LHCb's discovery potential. The search for this decay was performed with Run I data and set the only existing limit $\mathcal{B}(B_s \to \tau^+\tau^-) < 6.8 \times 10^{-3}$ at 95% CL [227]. Concerning the $B \to K\tau^+\tau^-$ decay, a method akin to the one applied to the LFV $B \to K\mu^+\tau^-$ decay discussed in section 1.4 may also allow LHCb to improve the existing limit set by the Babar Collaboration. Needless to say, the upcoming results from Belle-II will provide a stringent test of this scenario in the near future.

The natural question that emerges from this SMEFT picture is whether a

Observable	1σ	2σ	SM
$R_{K^*}^{[0.045,1.1]}$	$0.88\,^{+0.01}_{-0.01}$	$[0.86, 0.90]$	0.926 ± 0.004
$R_{K^*}^{[1.1,6.0]}$	$0.81\,^{+0.04}_{-0.04}$	$[0.73, 0.89]$	0.9964 ± 0.0006
$R_{K^*}^{[0.1,8.0]}$	$0.83\,^{+0.04}_{-0.03}$	$[0.77, 0.90]$	0.995 ± 0.002
$R_{K^*}^{[15,19]}$	$0.79\,^{+0.04}_{-0.04}$	$[0.71, 0.88]$	0.99807 ± 0.00004
$R_{K}^{[1.0,6.0]}$	$0.80\,^{+0.04}_{-0.04}$	$[0.71, 0.88]$	1.0008 ± 0.0003
$R_{\phi}^{[1.0,6.0]}$	$0.81\,^{+0.04}_{-0.04}$	$[0.73, 0.89]$	0.9970 ± 0.0003
$\langle P_5' \rangle^{[4.0,6.0]}$	$-0.58\,^{+0.13}_{-0.12}$	$[-0.82, -0.33]$	-0.763 ± 0.072
R_D	$0.34\,^{+0.01}_{-0.01}$	$[0.32, 0.37]$	0.303 ± 0.006
R_{D^*}	$0.29\,^{+0.01}_{-0.01}$	$[0.27, 0.31]$	0.255 ± 0.004
$\mathcal{B}(B_s \to \mu\mu)$	$2.98\,^{+0.20}_{-0.19} \times 10^{-9}$	$[2.60, 3.38] \times 10^{-9}$	$(3.67 \pm 0.16) \times 10^{-9}$
$\mathcal{B}(B \to K\tau\tau)$	$3.05\,^{+1.78}_{-1.06} \times 10^{-5}$	$[1.01, 6.47] \times 10^{-5}$	$(1.66 \pm 0.19) \times 10^{-7}$
$\mathcal{B}(B_s \to \tau\tau)$	$1.41\,^{+0.80}_{-0.47} \times 10^{-4}$	$[0.52, 2.94] \times 10^{-4}$	$(7.78 \pm 0.33) \times 10^{-7}$

Table 3.2: Predictions for LFU ratios, angular observables, and branching ratios in B and B_s decays from the global likelihood in the space of SMEFT Wilson coefficients $[C_{lq}^{(1)}]_{3323} = [C_{lq}^{(3)}]_{3323}$ and $[C_{lq}^{(1)}]_{2223} = [C_{lq}^{(3)}]_{2223}$ (cf. Fig. 3.4) and corresponding SM predictions.

model (either simplified or UV complete) can account for these shifts.

3.1.2 Explaining the data by a single mediator: the U_1 leptoquark solution

As anticipated in section 2.1.2, the only single mediator that can yield non-zero values for $[C_{LQ}^{(1)}]_{3323} = [C_{LQ}^{(3)}]_{3323}$ and $[C_{LQ}^{(1)}]_{2223} = [C_{LQ}^{(3)}]_{2223}$ is the U_1 vector leptoquark, which transforms as $(\mathbf{3}, \mathbf{1})_{2/3}$ under the SM gauge group [98, 101, 203, 280–286]. In this section, we consider a simplified model where this LQ is added to the SM and derive the mass and coupling constraints imposed by low-energy observables.

We define the U_1 couplings to the left-handed SM fermion doublets Q and L as

$$\mathcal{L}_{U_1} \supset g_{LQ}^{ji} \left(\bar{Q}^i \gamma^\mu L^j\right) U_\mu + \text{h.c.} \tag{3.12}$$

From the tree-level matching at the scale $\Lambda = M_U$, one finds

$$[C_{LQ}^{(1)}]_{ijkl} = [C_{LQ}^{(3)}]_{ijkl} = -\frac{g_{LQ}^{jk}\, g_{LQ}^{il*}}{2M_U^2}. \tag{3.13}$$

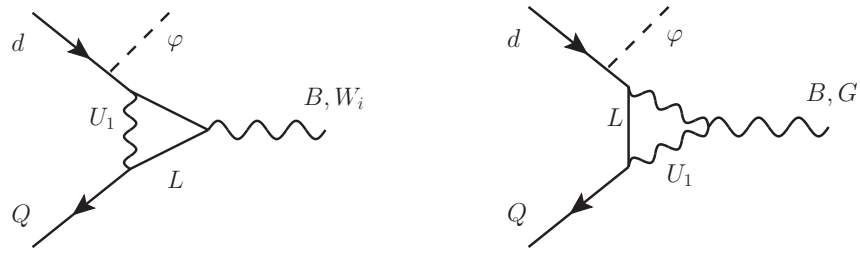

Figure 3.5: Diagrams contributing to the matching of the U_1 leptoquark model onto the SMEFT operators O_{dB}, O_{dW} and O_{dG}.

Consequently, for a given leptoquark mass, a τ-channel contribution to $R_{D^{(*)}}$ depends only on g_{LQ}^{32} and g_{LQ}^{33}, while a μ-channel contribution to $R_{K^{(*)}}$ depends only on g_{LQ}^{22} and g_{LQ}^{23}. The NLO corrections to such semileptonic operators are known to be of the order 13% [287] and will be neglected in the following.

As has been shown in [276], the U_1 leptoquark model generates one-loop matching contributions to the electric and chromomagnetic dipole operators in the WET. They can lead to relevant shifts in the Wilson coefficient C_7 at the b-quark scale, which are constrained by measurements of $b \to s\gamma$ observables (cf. [263]). In order to be sensitive to this possibly important effect, we will take the one-loop matching contributions to the SMEFT quark-dipole operators into account. These operators are defined as [275]

$$[O_{dB}]_{ij} = (\bar{Q}_i \sigma^{\mu\nu} d_j)\varphi B_{\mu\nu}\,, \qquad [O_{dW}]_{ij} = (\bar{Q}_i \sigma^{\mu\nu} d_j)\sigma^a \varphi W^a_{\mu\nu}\,, \tag{3.14}$$

$$[O_{dG}]_{ij} = (\bar{Q}_i \sigma^{\mu\nu} T^A d_j)\varphi G^A_{\mu\nu}\,. \tag{3.15}$$

The matching result depends on the couplings of the U_1 vector leptoquark to the SM gauge bosons, which can be written as

$$\mathcal{L}_{U_1} \supset -\frac{1}{2} U^\dagger_{\mu\nu} U^{\mu\nu} + i g_s k_s U^\dagger_\mu T^A U_\nu G^{A,\mu\nu} + i g' \frac{2}{3} k_Y U^\dagger_\mu U_\nu B^{\mu\nu}\,, \tag{3.16}$$

where

$$U_{\mu\nu} = D_\mu U_\nu - D_\nu U_\mu \quad \text{with} \quad D_\mu = \partial_\mu + i g_s T^A G^A_\mu + i g' \frac{2}{3} B_\mu. \tag{3.17}$$

These couplings are determined by SM gauge invariance except for the two parameters k_s and k_Y. In the following, we make the choice $k_s = k_Y = 1$, which leads to a cancellation of divergent tree-level diagrams in U_1-gluon and U_1-B-boson scattering [288] and further avoids logarithmically divergent contributions to the dipole operators [100], making them finite and gauge independent. We note that $k_s = k_Y = 1$ is automatically satisfied in any model in which the U_1 leptoquark stems from the spontaneous breaking of a gauge symmetry but can also be realized for a composite U_1 [289].

We perform the one-loop matching at the scale $\Lambda = M_U$ by computing the diagrams shown in Fig. 3.5. Working in the basis in which the down-type Yukawa matrix is diagonal, and using the conventions mentioned above, we find the Wilson coefficients of the EW dipole operators

$$[C_{dW}]_{23} = Y_b \frac{g}{16\pi^2}\left(\frac{1}{6}\right)\frac{g_{LQ}^{i2}\, g_{LQ}^{i3*}}{M_U^2}\,, \qquad [C_{dW}]_{32} = Y_s \frac{g}{16\pi^2}\left(\frac{1}{6}\right)\frac{g_{LQ}^{i3}\, g_{LQ}^{i2*}}{M_U^2}\,, \tag{3.18}$$

$$[C_{dB}]_{23} = Y_b \frac{g'}{16\pi^2}\left(-\frac{4}{9}\right)\frac{g_{LQ}^{i2}\, g_{LQ}^{i3*}}{M_U^2}\,, \qquad [C_{dB}]_{32} = Y_s \frac{g'}{16\pi^2}\left(-\frac{4}{9}\right)\frac{g_{LQ}^{i3}\, g_{LQ}^{i2*}}{M_U^2}\,, \tag{3.19}$$

where Y_b and Y_s denote the Yukawa couplings of the b and s quark respectively and a summation over the lepton index is implied. The Wilson coefficients of the chromomagnetic dipole operators at the matching scale read

$$[C_{dG}]_{23} = Y_b \frac{g_s}{16\pi^2}\left(-\frac{5}{12}\right)\frac{g_{LQ}^{i2}\, g_{LQ}^{i3*}}{M_U^2}\,, \quad [C_{dG}]_{32} = Y_s \frac{g_s}{16\pi^2}\left(-\frac{5}{12}\right)\frac{g_{LQ}^{i3}\, g_{LQ}^{i2*}}{M_U^2}\,. \tag{3.20}$$

Using the tree-level matching conditions from SMEFT onto WET [290, 291], we have checked that these results are consistent with the findings in [276].

$R_{D^{(*)}}$ and indirect constraints Within the defined framework we now search for a region of the leptoquark parameter space that explains all the B anomalies while at the same time avoids indirect low-energy constraints. We perform a fit with fixed $M_U = 2$ TeV in the space of tauonic couplings g_{LQ}^{32} and g_{LQ}^{33}, which we take to be real for simplicity. This allows us to determine the region in which $R_{D^{(*)}}$ can be explained by the semi-tauonic operators discussed in section 3.1.1. The results of the fit are shown in Fig. 3.6 and our findings are as follows:

- The strongest constraints are due to
  - leptonic tau decays $\tau \to \ell\nu\nu$, which receive a contribution due to RG running,[8]
  - $\mathcal{B}(B \to X_s\gamma)$, which receives a contribution from the one-loop matching onto dipole operators in SMEFT as discussed above.[9]

[8] Our analysis includes RG-induced logarithms. Note that the interactions in eq. (3.12) and (3.16) provide a simplified model and not a complete UV theory of the U_1-leptoquark. In such a UV theory, it could in principle be possible that the RG-induced logarithms are (partially) cancelled by finite terms, which are not present in the simplified model. Barring cancellations, and in view of the renormalization-scale independence of the full result – logarithms plus analytic terms – the contributions from the RG-induced logarithms usually provide a realistic estimate of the size of the effects.

[9] Such contributions are, however, model-dependent. For example, they will be quite different in models with additional vector-like fermions running in the loops [282, 283], as shown explicitly in Ref. [292].

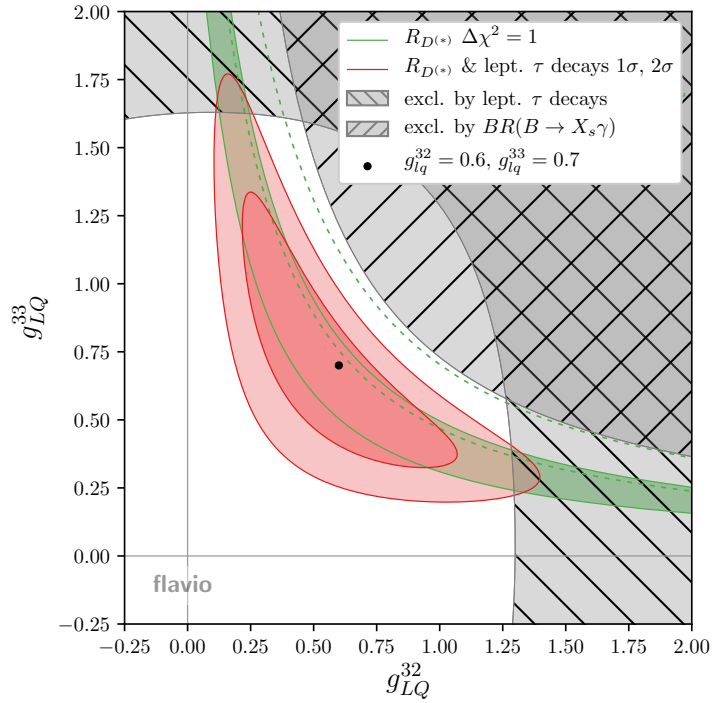

Figure 3.6: Likelihood contours from different observables in the space of the tauonic U_1 leptoquark couplings g_{LQ}^{32} and g_{LQ}^{33} at 2 TeV. The grey areas are excluded at the 2σ level. $R_{D^{(*)}}$data and leptonic τ decays select a well-defined region in the g_{LQ}^{32} versus g_{LQ}^{33} plane. For $R_{D^{(*)}}$, which only constrain one degree of freedom, 1σ contours correspond to $\Delta\chi^2 = 1$, while for others (the global likelihood, leptonic τ decays, $\mathcal{B}(B \to X_s\gamma)$), 1 and 2σ contours correspond to $\Delta\chi^2 \approx 2.3$ and 6.2, respectively. The dashed contour refers to pre-Moriond data of the corresponding solid contour.

We are recovering here the results of section 2.1.1. Indeed, in a third-generation dominated scenario, one would have $g_{LQ}^{32} \ll g_{LQ}^{33}$ on top of the $R_{D^{(*)}}$ constraint. This region of the plot is clearly excluded by leptonic tau decays, as foreseen around eq. (2.12). This underlines the importance of taking into account loop effects, both in the RG running and in the matching, as emphasized in [187, 209, 276].

- A combined fit to $R_{D^{(*)}}$and leptonic tau decays selects a well-defined region in the space of g_{LQ}^{32} and g_{LQ}^{33} in which $R_{D^{(*)}}$ can be explained while satisfying all constraints.

- In order to explain $R_{D^{(*)}}$ while at the same time avoiding exclusion at the 2σ level from leptonic tau decays, a minimal ratio of tauonic couplings $\frac{g_{LQ}^{32}}{g_{LQ}^{33}} \gtrsim 0.1$ is required (assuming vanishing right-handed couplings), which is compatible with findings in [189]. This puts some tension on models based on a $U(2)_Q$

flavor symmetry [100, 189, 283, 285], where the natural expectation for the size of $\frac{g_{LQ}^{32}}{g_{LQ}^{33}}$ is $|V_{cb}| \approx 0.04$.

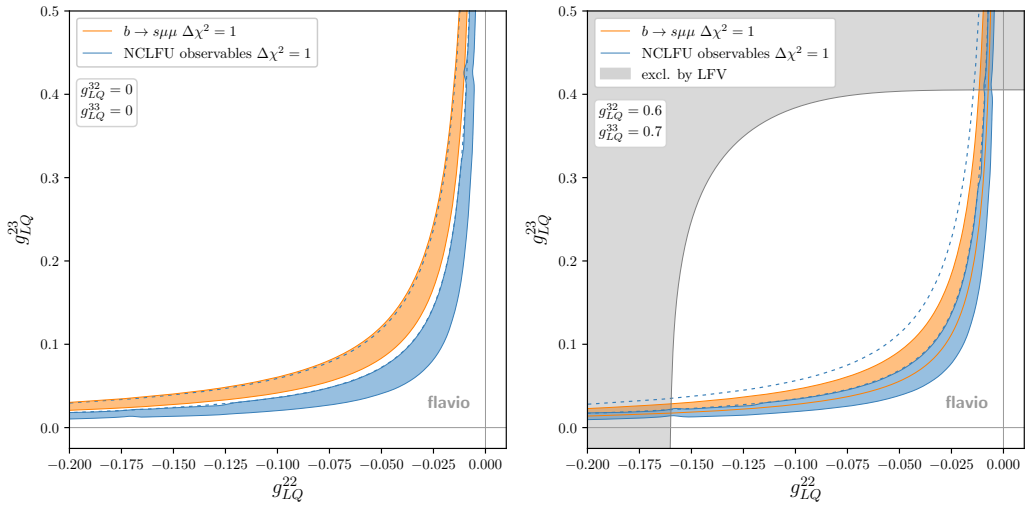

Figure 3.7: Likelihood contours from different observables in the space of the muonic U_1 leptoquark couplings g_{LQ}^{22} and g_{LQ}^{23} at 2 TeV. Fits are shown for vanishing tauonic couplings $g_{LQ}^{32} = 0$, $g_{LQ}^{33} = 0$ (left) and at the benchmark point $g_{LQ}^{32} = 0.6$, $g_{LQ}^{33} = 0.7$ (right). The grey area is excluded at the 2σ level. For observables that only constrain one degree of freedom (here NCLFU and $b \to s\mu\mu$ observables), 1σ contours correspond to $\Delta\chi^2 = 1$, while for the lepton flavour violating observables, the 2σ contour corresponds to $\Delta\chi^2 \approx 6.2$. The dashed contours refers to pre-Moriond data of the corresponding solid contour.

Based on the above results, we select a benchmark point from the best-fit region in the fit to tauonic couplings,

$$g_{LQ}^{32} = 0.6\,, \qquad g_{LQ}^{33} = 0.7\,, \tag{3.21}$$

which is also shown in Fig. 3.6. We then perform two fits in the space of muonic couplings g_{LQ}^{22} and g_{LQ}^{23} shown in Fig. 3.7: one for vanishing tauonic couplings (left panel) and one at the benchmark point $g_{LQ}^{32} = 0.6$, $g_{LQ}^{33} = 0.7$ (right panel). Our findings are as follows:

- For vanishing tauonic couplings (left panel of Fig. 3.7), the data available before Moriond 2019 leads to a very good agreement between the fits to $b \to s\mu\mu$ (orange contour) and NCLFU observables (dashed blue contour), while the $R_{D^{(*)}}$ measurements cannot be explained in this scenario. Taking into account the updated and new measurements of $R_{K^{(*)}}$ presented at Moriond 2019, one finds a slight tension between the fits to $b \to s\mu\mu$ (orange contour) and NCLFU observables (solid blue contour). This is analogous to the tension mentioned in fit in the WET basis.

- The tension disappears if one considers non-zero tauonic couplings that can also explain $R_{D^{(*)}}$, which is exemplified in the right panel of Fig. 3.7 for the benchmark point $g_{LQ}^{32} = 0.6$, $g_{LQ}^{33} = 0.7$. As discussed in section 3.1.1, the semi-tauonic operators obtained from the tree-level matching (*cf.* eq. (3.13)) induce a lepton-flavour universal contribution to C_9, which affects the predictions of $b \to s\mu\mu$ observables and makes the fits to $b \to s\mu\mu$ and NCLFU observables again compatible with each other at the 1σ level. Consequently, the deviations in neutral current and charged current B-decays can all be explained at once. This very well agrees with our findings in the SMEFT scenario in section 3.1.1.

- Given the presence of non-zero values for the tauonic couplings at the benchmark point, the strongest constraint on the muonic couplings g_{LQ}^{22} and g_{LQ}^{23} is due to LFV observables, in particular $\tau \to \phi\mu$ and $B \to K\tau\mu$. The region in the space of muonic couplings that is excluded at the 2σ level by these observables is shown in gray in the right panel of Fig. 3.7.

Having identified a viable benchmark point, we conclude that the U_1 vector leptoquark can still provide an excellent description of the B anomalies while satisfying all indirect constraints.

Comparison between indirect and direct constraints In addition to indirect constraints, high-p_T signatures of models containing a U_1 leptoquark have been discussed in detail considering current and future LHC searches [193, 201, 247, 293–295]. In this section, we compare direct constraints found in the latest study, [295], to the strong indirect constraints discussed in the previous section. To this end, we adopt the notation of [295] and use the parameters β_L^{ij} and g_U, which are related to our notation by

$$g_{LQ}^{ij} = \frac{\beta_L^{ji}\, g_U}{\sqrt{2}}. \tag{3.22}$$

We perform a fit with fixed $g_U = 3$, $\beta_L^{33} = 1$ (i.e. $g_{LQ}^{33} \approx 2$) in the space of M_U and β_L^{23}. These values are chosen to allow for a direct comparison with the constraint from $pp \to \tau\tau$ shown in Fig. 1 of [295] and $pp \to \tau\nu$ shown in Fig. 6 of [295]. We include both of these direct constraints in our Fig. 3.8 as hatched areas. In addition, we show the results from our fit, namely the constraint from leptonic τ decays and the region preferred by the $R_{D^{(*)}}$ measurements. Our findings are as follows:

- The indirect constraint from leptonic τ decays is stronger than the direct constraints in nearly all of the parameter space shown in Fig. 3.8, except for a small region at large $\beta_L^{23} \gtrsim 0.75$, where the constraint from $pp \to \tau\nu$ is the strongest one.

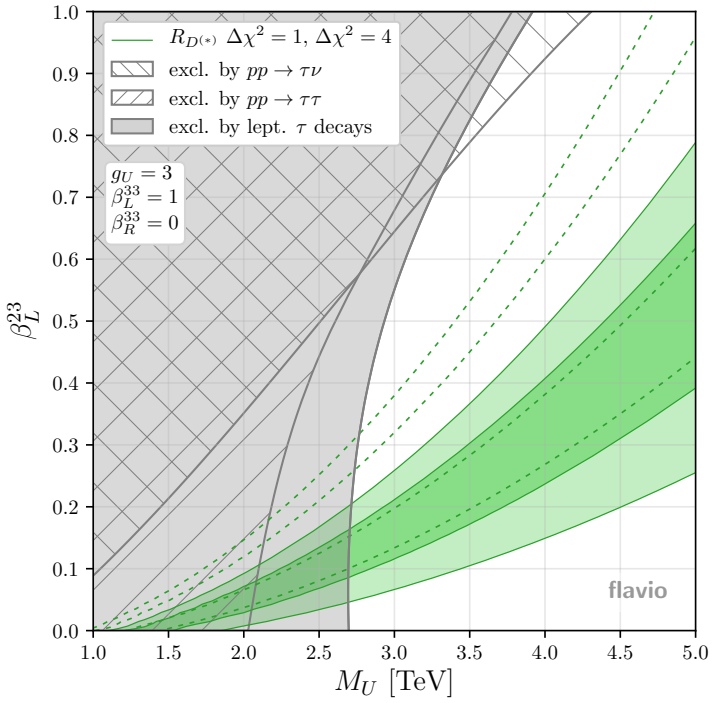

Figure 3.8: Best fit region in the space of U_1 leptoquark mass M_U and coupling β_L^{23} (*cf.* eq. (3.22)). The green region is the preferred region from $R_{D^{(*)}}$, while the gray shaded area is excluded by leptonic tau decays at the 2σ level. The hatched areas are excluded by LHC searches recasted in [295].

- In the region where $R_{D^{(*)}}$ can be explained, the indirect constraint from leptonic τ decays is considerably stronger than the direct ones.
- Small values for $\frac{\beta_L^{23}}{\beta_L^{33}}$ as naturally expected in models based on a $U(2)_Q$ flavour symmetry require a relatively small mass M_U to explain $R_{D^{(*)}}$. Thus, as also pointed out in [189, 193], there is already some tension between this natural expectation and the direct searches.

We note that the direct constraints shown in Fig. 3.8 depend on the coupling strength g_U. While the assumptions $g_U = 3$, $\beta_L^{33} = 1$ lead to a lower bound on the leptoquark mass $M_U \gtrsim 2.7$ TeV, this bound does not apply to the scenario discussed in the previous section, which features considerably smaller couplings[10]. Latest direct constraints from U_1 pair production that are independent of the coupling strength g_U only exclude masses $M_U \lesssim 1.5$ TeV [201, 295]. Therefore, the scenario discussed in the previous section is currently not constrained by direct searches.

[10] The partonic cross section relevant for the direct constraints in Fig. 3.8 scales as $\sigma \sim (g_U/M_U)^4$ [295].

3.1.3 Conclusions

A new-physics effect in the semi-muonic Wilson coefficient $C_9^{\mu\mu}$ continues to give a much improved fit to the data compared to the SM. However, compared to previous global analyses, we find that there is now also a preference for a non-zero value of the semi-muonic Wilson coefficient $C_{10}^{\mu\mu}$, mostly driven by the global combination of the $B_s \to \mu^+\mu^-$ branching ratio including the ATLAS measurement. The single-coefficient scenario giving the best fit to the data is the one where $C_9^{\mu\mu} = -C_{10}^{\mu\mu}$, which is known to be well suited to UV-complete interpretations, and indeed is predicted in several new-physics models with tree-level mediators coupling dominantly to left-handed fermions. We have also studied the possibility of a simultaneous interpretation of the $b \to s\ell\ell$ data and the discrepancies in $b \to c\tau\nu$ transitions in the framework of a global likelihood in SMEFT Wilson coefficient space. We find one especially compelling scenario, characterised by new physics in all-left-handed semi-tauonic four-fermion operators. These operators can explain directly the discrepancies in $b \to c\tau\nu$ transitions, and, at the same time, radiatively induce a lepton flavour universal contribution to the $b \to s\ell\ell$ Wilson coefficients. An additional non-zero semi-muonic Wilson coefficient then allows accommodating the $R_{K^{(*)}}$ discrepancies. Such picture can be quantitatively realized in the context of the U_1 leptoquark simplified model, and we find that indeed an excellent description of the data can be obtained, including the deviations in $b \to c\tau\nu$ transitions. This scenario results in experimental predictions that we summarized in Table 3.2.

In the next section we will briefly present a possible UV completion for the U_1 leptoquark simplified model and show that it can be enhanced by DM fermionic multiplet that leads to a model in natural agreement with the relic-density observation and with the most severe direct detection bounds.

3.2 B-anomalies and Dark Matter

As discussed in the introduction, DM is one of the strongest observational arguments in favour of physics beyond the Standard Model.

Despite its cosmological origins, DM triggered a lot of development in particle physics. One of the reasons of this concern is the so-called weakly interactive massive particle (WIMP) 'miracle', namely that a massive particle whose couplings to the SM are of the order of the weak scale reproduces the observed DM density in the Universe[11]. Intriguingly, the mass scale preferred by WIMP scenarios ranges from a few tens of GeV to a few tens of TeV [296], similar to the scales hinted at by the B anomalies. This resemblance suggests a combined explanation to both

[11] Two other well-known reasons are the strong CP problem, where scenarios based on axions propose natural DM candidates, and sterile neutrinos that can also account for neutrino masses.

puzzles, already proposed in several publications [112, 218, 297–307].

The next two sections investigate extensions of the models proposed in sections 3.1.2 and 2.2 and their possible links with DM.

3.2.1 Dark Matter in the 4321 model(s)

Let us first consider possible UV completions for the simplified U_1 leptoquark model discussed in section 3.1.2. In this section, adapted from [308], we would like to address the possibility that the vector bosons appearing in these completions serve as mediators between SM fermionic currents and a DM current. We therefore consider a fermionic DM candidate χ_0 that fulfils the following assumptions (*cf. e.g.* [309, 310])

(i) it is a thermal relic,

(ii) it is colourless and electrically neutral,

(iii) it has zero hypercharge to avoid direct-detection bounds (as discussed later in this section).

In our model, scalar DM can annihilate to SM particles via a Higgs portal, making DM phenomenology independent of the effects of the new vector bosons. For this reason, we restrict our study to fermionic DM candidates.

4321 models in a nutshell

The U_1 LQ appears in Pati-Salam models [196] where the SM $SU(3)_c$ is embedded in a larger gauge group $SU(4)$. As well known, these models cannot reproduce the flavour structure needed to explain the anomalies. Variants of the Pati-Salam model can however be built based on the '4321' gauge group [229, 280, 283, 285, 286, 292, 293, 311–313]

$$SU(4) \times SU(3)' \times SU(2)_L \times U(1)_X \,. \tag{3.23}$$

The SM group is recovered below a scale $v_{\rm LQ}$ where the spontaneous breaking

$$SU(4) \times SU(3)' \times U(1)_X \to SU(3)_c \times U(1)_Y \tag{3.24}$$

takes place. We denote by H_μ^α, C_μ^a, and B'_μ the gauge bosons of $SU(4)$, $SU(3)'$, and $U(1)_X$ respectively, where $\alpha = 1, \ldots, 15$ and $a = 1, \ldots, 8$. The breaking eq. (3.24) yields a massive $U_1 \sim (\mathbf{3}, \mathbf{1})_{2/3}$ leptoquark

$$U_\mu^{\pm 1,2,3} = \tfrac{1}{\sqrt{2}}(H_\mu^{9,11,13} \mp i H_\mu^{10,12,14}) \,, \tag{3.25}$$

as well as a massive $Z'_\mu \sim (\mathbf{1}, \mathbf{1})_0$ and a massive gluon-like 'colouron' fields $G'^a_\mu \sim (\mathbf{8}, \mathbf{1})_0$, given by the linear combinations

$$Z'_\mu = H_\mu^{15} \cos\theta_{41} - B'_\mu \sin\theta_{41} \,, \qquad G'^a_\mu = H_\mu^a \cos\theta_{43} - C_\mu^a \sin\theta_{43} \,. \tag{3.26}$$

The angle θ_{41} and θ_{43} are related to the coupling constants g_4, g_3, g_1, g_s, and g_Y (for $SU(4)$, $SU(3)'$, $U(1)_X$, $SU(3)_c$, and $U(1)_Y$ respectively), analogously to the weak-mixing angle, by

$$\cos\theta_{41} = \frac{g_4}{\sqrt{g_4^2 + \frac{2}{3}g_1^2}} = \frac{g_Y}{g_1}\,, \qquad \cos\theta_{43} = \frac{g_4}{\sqrt{g_4^2 + g_3^2}} = \frac{g_s}{g_3}\,. \tag{3.27}$$

The masses of the new bosons are then given by

$$\begin{aligned} M_U^2 &= \tfrac{1}{4}\, g_4^2\, v_{\mathrm{LQ}}^2\,, \\ M_{Z'}^2 &= \tfrac{1}{4}\, \left(g_4^2 + \tfrac{2}{3}\, g_1^2\right)\, v_{\mathrm{LQ}}^2\,, \\ M_{G'}^2 &= \tfrac{1}{4}\, \left(g_4^2 + g_3^2\right)\, v_{\mathrm{LQ}}^2\,, \end{aligned} \tag{3.28}$$

such that the U_1 has the smallest tree-level mass.

The derivation of the couplings between U_1, Z', and G' and the SM fermions is given in Appendix A.3. These couplings read

$$\begin{aligned} \mathcal{L}_{Z'} \supset\; & \frac{g_4}{2\sqrt{6}\cos\theta_{41}}\, Z'_\mu \Big(\xi_Q^i\, \bar{Q}^i \gamma^\mu Q^i + \xi_u^i\, \bar{u}^i \gamma^\mu u^i + \xi_d^i\, \bar{d}^i \gamma^\mu d^i \\ & \qquad - 3\left(\xi_L^i\, \bar{L}^i \gamma^\mu L^i + \xi_e^i\, \bar{e}^i \gamma^\mu e^i + \xi_\nu\, \bar{\nu}^3 \gamma^\mu \nu^3 \right) \Big), \\ \mathcal{L}_{G'} \supset\; & \frac{g_4}{\cos\theta_{43}}\, G'^a_\mu \left(\kappa_Q^i\, \bar{Q}^i \gamma^\mu T^a Q^i + \kappa_u^i\, \bar{u}^i \gamma^\mu T^a u^i + \kappa_d^i\, \bar{d}^i \gamma^\mu T^a d^i \right), \\ \mathcal{L}_{U_1} \supset\; & \frac{g_4}{\sqrt{2}}\, U_\mu^+ \left(\beta_L^{ij}\, \bar{Q}^i \gamma^\mu L^j + \beta_{de}\, \bar{d}^3 \gamma^\mu e^3 + \beta_{u\nu}\, \bar{u}^3 \gamma^\mu \nu^3 \right) + \text{h.c.}\,, \end{aligned} \tag{3.29}$$

where the constants κ and ξ that appear in the G' and Z' couplings are collected in table 3.3, ν^i stand for right-handed neutrinos and the constants β that appear in the U_1 couplings are given by

$$\beta_L = \begin{pmatrix} \sin\theta_{Q_1}\, \sin\theta_{L_1} & 0 & 0 \\ 0 & \sin\theta_{Q_2}\, \sin\theta_{L_2}\, \cos\theta_{LQ} & \sin\theta_{Q_2}\, \sin\theta_{LQ} \\ 0 & -\sin\theta_{L_2}\, \sin\theta_{LQ} & \cos\theta_{LQ} \end{pmatrix}, \tag{3.30}$$

$$\beta_{de} = e^{i\phi_e}\,, \qquad \beta_{u\nu} = e^{i\phi_\nu}\,.$$

The effective model of the previous section is recovered with the identification $g_4 = g_U$ in eq. (3.22). While the parametrization described above allows explaining the B-meson anomalies and avoids strong constraints from large flavour violating effects, the number of parameters can be further reduced by the following phenomenologically motivated assumptions:

- To maximize the agreement with the B-decay measurements that deviate from the SM, one can take $\beta_{de} = -1$ [292], which fixes the phase $\phi_e = \pi$.

Field	κ		ξ	
	$i = 1, 2$	$i = 3$	$i = 1, 2$	$i = 3$
Q^i	$\sin^2\theta_{Q_{1,2}} - \sin^2\theta_{43}$	$\cos^2\theta_{43}$	$\sin^2\theta_{Q_{1,2}} - \sin^2\theta_{41}$	$\cos^2\theta_{41}$
u^i	$-\sin^2\theta_{43}$	$\cos^2\theta_{43}$	$-4\sin^2\theta_{41}$	$1 - 4\sin^2\theta_{41}$
d^i	$-\sin^2\theta_{43}$	$\cos^2\theta_{43}$	$2\sin^2\theta_{41}$	$1 + 2\sin^2\theta_{41}$
L^i			$\sin^2\theta_{L_{1,2}} - \sin^2\theta_{41}$	$\cos^2\theta_{41}$
e^i			$-2\sin^2\theta_{41}$	$1 - 2\sin^2\theta_{41}$
ν^i				1
ψ	$\cos^2\theta_{43}$		$1 - 4\sin^2\theta_{41}$	
χ			1	

Table 3.3: Constants κ and ξ entering the couplings of fermions to G' and Z'.

- Since the phase ϕ_ν is currently not constrained by any measurement one can use $\phi_\nu = 0$ for simplicity.
- An approximate $U(2)$ symmetry in the quark sector, i.e. $\theta_{Q_1} \approx \theta_{Q_2}$, can be employed to suppress tree-level FCNCs in the up-quark sector that are mediated by the Z' and G' [312]. Without such a $U(2)$ protection, excessive contributions to $\Delta C = 2$ observables would be possible.
- The first-generation lepton doublet can be taken to be purely a singlet of $SU(4)$, i.e. $\theta_{L_1} = 0$, to be safe from LFV due to U_1 couplings involving the electron.

Making all of the above assumptions and defining $\theta_{Q_{12}} = \theta_{Q_1} = \theta_{Q_2}$, the only remaining free parameters in the fermion sector are $\theta_{Q_{12}}$, θ_{L_2}, and θ_{LQ}.

In addition to the assumptions *(i-iii)* made on the DM candidate, we assume the two following points.

(iv) The DM candidate is the component of a massive vector-like fermion multiplet Ψ_{DM},

(v) (co-)annihilation proceeds via $2 \to 2$ processes induced at tree level through the new vector bosons U_1, G', and Z'.

Assumptions *(ii-iv)* restricts the possible representation of the DM multiplet under the 4321 group. In particular, *(ii)* and *(iii)* require that Ψ_{DM} transforms under an odd representation of $SU(2)_L$ and fixes the $U(1)_X$ hypercharge for each $SU(4)$ representation. Guided by minimality, we restrict our discussion to singlets of

$SU(3)'$ and to the smallest non-trivial representation of $SU(4)$, the fundamental **4** representation, which leads to[12]

$$\Psi_{\rm DM} \sim (\mathbf{4},\mathbf{1},\mathbf{N})_{1/2}, \qquad \mathbf{N} \in \{\mathbf{1},\mathbf{3},\mathbf{5},\dots\} \tag{3.31}$$

under the 4321 gauge group. After the 4321 symmetry breaking, the $\Psi_{\rm DM}$ multiplet splits into the two components χ and ψ, which transform under the SM gauge group as

$$\chi \sim (\mathbf{1},\mathbf{N})_0, \qquad \psi \sim (\mathbf{3},\mathbf{N})_{2/3}, \qquad \mathbf{N} \in \{\mathbf{1},\mathbf{3},\mathbf{5},\dots\}. \tag{3.32}$$

The dark matter candidate χ_0 is then identified with the electrically neutral component of the $SU(2)_L$ N-plet χ. The remaining components are refered to as co-annihilation partners [309]. Eq. (3.29) is then augmented with the following terms

$$\begin{aligned} \mathcal{L}_{Z'} &\supset \frac{g_4}{2\sqrt{6}\cos\theta_{41}}\, Z'_\mu \Big(\xi_\psi\, \bar\psi\gamma^\mu\psi - 3\xi_\chi\, \bar\chi\gamma^\mu\chi\Big)\,, \\ \mathcal{L}_{G'} &\supset \frac{g_4}{\cos\theta_{43}}\, G'^a_\mu \Big(\kappa_\psi\, \bar\psi\gamma^\mu T^a\psi\Big)\,, \qquad \mathcal{L}_{U_1} \supset \frac{g_4}{\sqrt{2}}\, U^+_\mu \Big(\bar\psi\gamma^\mu\chi\Big) + \text{h.c.} \end{aligned} \tag{3.33}$$

where $\xi_{\psi,\chi}$ and κ_ψ are given in Table 3.3.

Note that in the present setup, renormalizable couplings between the DM candidates and SM particles can be written for $\mathbf{N}=\mathbf{1}$ and $\mathbf{N}=\mathbf{3}$. Indeed, for $\mathbf{N}=\mathbf{1}$, ψ has the same quantum numbers as a right-handed up-type quark and even a small mixing between ψ and u^i can make χ unstable on cosmological time-scales. For $\mathbf{N}=\mathbf{3}$, a coupling between χ, the Higgs and a lepton doublet would open the decay channel of DM to a Higgs and a neutrino. For these two cases we therefore advocate an additional symmetry that protects the DM candidate and make it a viable relic. For $\mathbf{N}=\mathbf{5}$ no such term can be written and these additional arguments are not needed [309].

Although no tree-level term introduces a mass-splitting between χ and ψ, these fermions get mass corrections at one-loop due to the gauge bosons associated to the spontaneously broken symmetries. Such correction is crucial for the phenomenology because for a small splitting, comparable or smaller than the freeze-out temperature, the co-annihilation partners become as kinematically accessible as the DM candidate. The mass splitting due to the non-trivial transformation under $SU(2)_L$ between components ξ and η of a VL multiplet of hypercharge Y and mass $\hat M$, $\Delta_{\xi\eta} \equiv (M_\xi - M_\eta)/\hat M$, is well known (*cf. e.g.* [309])

$$\begin{aligned} \Delta^{\rm EW}_{\xi\eta} = \frac{g^2}{16\,\pi^2}\Bigg\{ &\Big((Q_\xi - Y)^2 - (Q_\eta - Y)^2\Big)\left[f\left(\tfrac{M_W}{\hat M}\right) - f\left(\tfrac{M_Z}{\hat M}\right)\right] \\ &+ s_W^2\,(Q_\xi^2 - Q_\eta^2)\, f\left(\tfrac{M_Z}{\hat M}\right)\Bigg\}, \end{aligned} \tag{3.34}$$

[12] If $\Psi_{\rm DM}$ is a singlet of $SU(4)$, then its 4321 quantum numbers are fixed to $(\mathbf{1},\mathbf{1},\mathbf{N})_0$, and couplings to U_1, G', and Z' are absent. This corresponds to "Minimal Dark Matter", discussed in [309].

where Q_ξ and Q_η are the electric charges of ξ and η, and $f(r)$ is the finite loop function

$$f(r) = r^4 \ln r - r^2 + \frac{r}{2}\sqrt{r^2-4}\left(r^2+2\right)\ln\left(\frac{r^2}{2} - \frac{r}{2}\sqrt{r^2-4} - 1\right). \quad (3.35)$$

For reference, the relative mass splitting within the ψ and χ $SU(2)_L$ multiplets of our DM sector is between $\mathcal{O}(10^{-3})$ and $\mathcal{O}(10^{-4})$ for $\hat{M} = \mathcal{O}(1\text{ TeV})$. This result can be generalized to the one-loop mass splitting induced by the vector bosons associated with the 43(2)1 symmetry breaking. Applying the generic result of Ref. [308] to our DM candidate χ and its coloured co-annihilation partner ψ, the mass splitting is given by

$$\Delta^{4321}_{\psi\chi} = \frac{g_4^2}{16\,\pi^2}\left\{ f\left(\frac{M_U}{\hat{M}}\right) + \frac{1}{3}(2\,\sin^2\theta_{41}+1)\,f\left(\frac{M_{Z'}}{\hat{M}}\right) + \frac{4}{3}(\sin^2\theta_{43}-1)\,f\left(\frac{M_{G'}}{\hat{M}}\right)\right\}. \quad (3.36)$$

The value of $\Delta^{4321}_{\psi\chi}$ is around $8-15\%$ for the parameter region of interest (see fig. 3.9) and its impact on the phenomenology will be discussed in the following. Note that additional mass-splittings arise from the non-zero temperature at

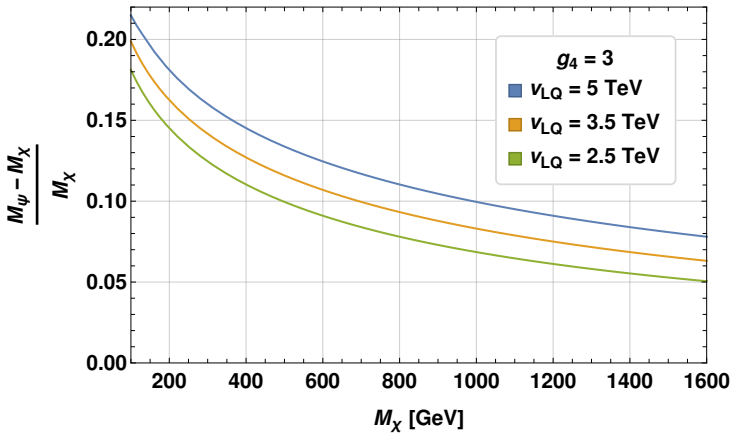

Figure 3.9: Mass splitting between ψ and χ induced by the 4321 gauge bosons.

which the freeze-out process takes place. These additional corrections are however negligible with respect to the one given in eq. (3.36) [308].

Dark matter relic abundance

The first constraint on our DM candidate is that it reproduces the relic abundance observed today, $\Omega_0 h^2$. To obtain this quantity, the first step consists in the estimation of the freeze-out temperature T_f at which the DM candidate decouple from thermal equilibrium. Defining $x \equiv M_{\chi_0}/T$ and $x_f = x|_{T_f}$, this temperature

is constrained by the relation [314]

$$x_f \;=\; \ln \frac{0.038\, g_{\rm eff}\, M_{\rm Pl}\, M_{\chi_0}\, \langle \sigma_{\rm eff}\, v\rangle}{g_*^{1/2}\, x_f^{1/2}}\,, \tag{3.37}$$

where $M_{\rm Pl} = 1.22 \times 10^{19}$ GeV and g_* is the total number of effectively relativistic d.o.f. at freeze-out. Eq. (3.37) also contains two effective quantities, $g_{\rm eff}$ that denotes the number of effective d.o.f. within the DM sector and the thermally averaged annihilation cross-section $\langle \sigma_{\rm eff}\, v\rangle$. After freeze-out, the relic abundance is subject to post-freeze-out annihilation processes. The efficiency of this post-freeze-out annihilation is given by [314, 315]

$$J \equiv \int_{x_f}^{\infty} \frac{\langle \sigma_{\rm eff}\, v\rangle}{x^2} dx\,, \tag{3.38}$$

which yield to the present-day DM abundance[13]

$$\Omega_0 h^2 = \sqrt{\frac{45}{\pi}} \frac{s_0}{\rho_c} \frac{1}{g_*^{1/2}\, M_{\rm Pl}\, J} \simeq \frac{1.07 \times 10^9\, {\rm GeV}^{-1}}{g_*^{1/2}\, M_{\rm Pl}\, J}\,. \tag{3.39}$$

The main dynamical quantity entering this estimation is the thermally averaged cross-section $\langle \sigma_{\rm eff}\, v\rangle$. In the presence of co-annihilation partners, the latter and the effective number of degree of freedom are given by [314]

$$\begin{aligned} g_{\rm eff} &= \sum_i \left(g_\chi\, (1+\Delta_{\chi_i})^{3/2} e^{-x\,\Delta_{\chi_i}} + g_\psi\, (1+\Delta_{\psi_i})^{3/2} e^{-x\,\Delta_{\psi_i}} \right)\,, \\ \sigma_{\rm eff} &= \frac{1}{g_{\rm eff}^2} \sum_{i,j} \Big(\sigma_{\chi_i\chi_j}\, g_\chi^2\, (1+\Delta_{\chi_i})^{3/2}\, (1+\Delta_{\chi_j})^{3/2}\, e^{-x(\Delta_{\chi_i}+\Delta_{\chi_j})} \\ &\qquad + \sigma_{\psi_i\psi_j}\, g_\psi^2\, (1+\Delta_{\psi_i})^{3/2}\, (1+\Delta_{\psi_j})^{3/2}\, e^{-x(\Delta_{\psi_i}+\Delta_{\psi_j})} \\ &\qquad + 2\,\sigma_{\chi_i\psi_j}\, g_\chi\, g_\psi\, (1+\Delta_{\chi_i})^{3/2}\, (1+\Delta_{\psi_j})^{3/2}\, e^{-x(\Delta_{\chi_i}+\Delta_{\psi_j})} \Big)\,, \end{aligned} \tag{3.40}$$

where the indices i and j run over the N components of the $SU(2)_L$ multiplets χ and ψ and e.g.

$$\sigma_{\chi_i\psi_j} \equiv \sigma(\chi_i\psi_j \to XX')\,, \tag{3.41}$$

X, X' denoting any particles other than χ and ψ. Besides $g_\chi = 4$ and $g_\psi = 12$ denote the internal (spin, colour, ...) d.o.f. of the components of these multiplets. For the relative mass splitting $\Delta_{\chi_i} = (M_{\chi_i} - M_{\chi_0})/M_{\chi_0}$ and $\Delta_{\psi_i} = (M_{\psi_i} - M_{\chi_0})/M_{\chi_0}$, the discussion around eq. (3.36) allows the following simplification

$$M_{\chi_i} \to M_\chi\,, \qquad M_{\psi_i} \to M_\psi\,, \qquad \Delta_{\chi_i} \to 0\,, \qquad \Delta_{\psi_i} \to \Delta_\psi\,. \tag{3.42}$$

[13] One can derive this relation by using $H(T) = \sqrt{8\pi^3 g_*/90}\; T^2/M_{\rm Pl}$, $s = 2\pi^2 g_* T^3/45$ and $\rho_c = 3H_0^2/(8\pi G_N)$, with $H(T_0) \equiv H_0 = 100h\frac{\rm km}{\rm s\ Mpc}$.

Employing this replacement, eq. (3.40) simplifies to

$$\begin{aligned} g_{\text{eff}} &= N\left(g_\chi + g_\psi\,(1+\Delta_\psi)^{3/2} e^{-x\,\Delta_\psi}\right)\,, \\ \sigma_{\text{eff}} &= \frac{1}{g_{\text{eff}}^2}\sum_{i,j}\Big(\sigma_{\chi_i\chi_j}\,g_\chi^2 + 2\,\sigma_{\chi_i\psi_j}\,g_\chi\,g_\psi\,(1+\Delta_\psi)^{3/2}\,e^{-x\Delta_\psi} \\ &\qquad + \sigma_{\psi_i\psi_j}\,g_\psi^2\,(1+\Delta_\psi)^3\,e^{-2x\Delta_\psi}\Big)\,. \end{aligned} \tag{3.43}$$

Numerically, g_{eff} differs considerably from $N(g_\chi+g_\psi)$ for $x = x_f$. Indeed, plugging $\Delta_\psi \approx 0.1$ as computed in eq. (3.36) yields the suppression factor

$$\varepsilon_\psi \equiv (1+\Delta_\psi)^{3/2} e^{-x\,\Delta_\psi} \approx 0.06\,. \tag{3.44}$$

The cross sections $\sigma_{\chi_i\chi_j}$, $\sigma_{\chi_i\psi_j}$, and $\sigma_{\psi_i\psi_j}$ are due to the exchange of either SM bosons or the new heavy gauge bosons U_1, Z', and G'. Because of the dependence of the cross sections on the fourth power of the couplings, contributions due to the electroweak sector are negligible compared to those involving the relatively strongly coupled new heavy gauge bosons or gluons. We find that all the cross sections mediated by U_1, Z', G', and gluons are of comparable size. However, from eq. (3.43), one sees that $\sigma_{\chi_i\psi_j}$ and $\sigma_{\psi_i\psi_j}$ are suppressed by one and two power of ε_ψ respectively, which means that *the freeze-out process is mainly driven by Z' exchange*.

Although all cross-sections are kept in the numerical analysis, the phenomenology can be understood assuming only Z' exchange. Neglecting the masses of the annihilation products, we define [316]

$$\sigma_0(y) \;=\; 2\sqrt{y^2-y}\,\sigma_{\text{eff}}(y) \tag{3.45}$$

where we substituted s by the dimensionless variable $y = s/(4\,M_\chi^2)$. Using σ_{eff} as in eq. (3.43), eq. (3.45) yields

$$\sigma_0(y) \approx \frac{1}{128\,\pi}\left(\frac{g_4}{\cos\theta_{41}}\right)^4 \frac{M_\chi^2\,(2\,y^2+y)}{(4\,y\,M_\chi^2 - M_{Z'}^2)^2}\,\frac{1}{N}\,f(\{\xi^i\})\,, \tag{3.46}$$

where for brevity we introduced the flavour function

$$f(\{\xi^i\}) \equiv \sum_{i=1}^{3}\Big(2|\xi_Q^i|^2 + |\xi_u^i|^2 + |\xi_d^i|^2 + 3(2|\xi_L^i|^2 + |\xi_e^i|^2 + |\xi_\nu^i|^2)\Big)\,. \tag{3.47}$$

This function has values of typical order $\mathcal{O}(10)$. For example, from Table 3.3 with $g_4 = 3$, $\sin\theta_{Q_{12}} = 0.2$, one has $f(\{\xi^i\}) \approx 16$.

In the large x_f limit one has $\langle\sigma_{\text{eff}}\,v\rangle \approx \sigma_0(1)$ [316]. Plugging eq. (3.46) into eqs. (3.38) and (3.39), and taking the representative value $g_4 = 3$, we find in the limit $4M_\chi^2 \ll M_{Z'}^2$

$$\Omega_0 h^2 \;\approx\; 0.06\frac{N}{f(\{\xi^i\})}\left(\frac{v_{LQ}}{5\text{ TeV}}\right)^2\left(\frac{v_{LQ}}{M_\chi}\right)^2\,. \tag{3.48}$$

From this equation, it is clear that the experimental measurement $\Omega_0 h^2 = 0.120 \pm 0.001$ [317] can be reproduced for DM masses of a few hundreds of GeV.

Considering also U_1 exchanges and the $1/x_f$ correction to $\langle \sigma_{\text{eff}}\, v \rangle$ leads to the constraints displayed on Fig. 3.10. In this figure we excluded the region $2M_\chi >$

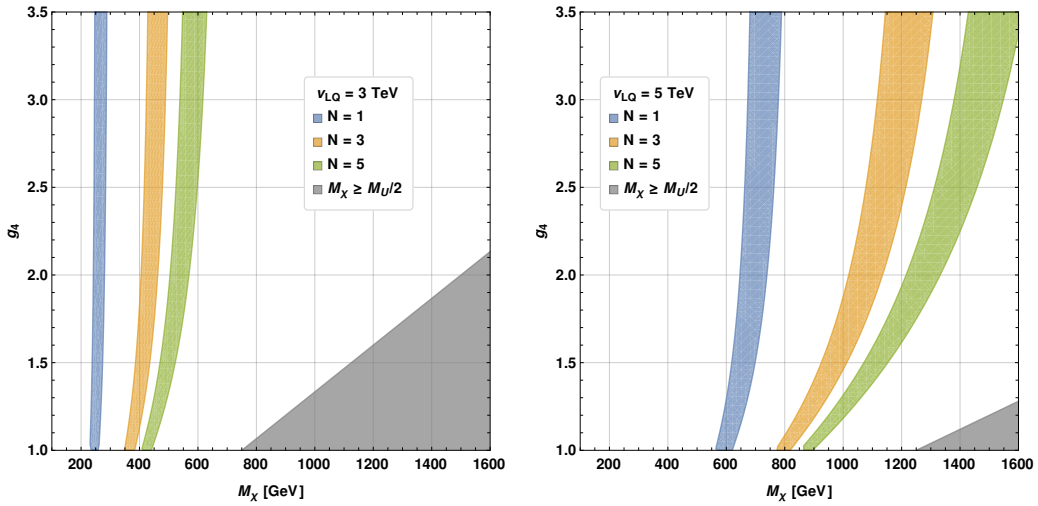

Figure 3.10: Coloured rays fulfil the $\Omega_0 h^2$ constraint within $\pm 15\%$. The gray region corresponds to an on-shell intermediate U_1 leptoquark (see text for details).

M_U where the production of on-shell leptoquarks challenges our assumption (v), namely that phenomenology takes place only through $2 \to 2$ processes. The small dependence on g_4, more constrained by other observables to follow, allows $\Omega_0 h^2$ to be comfortably reproduced for the following ranges

$$M_\chi\,[\text{GeV}] \in \begin{array}{ll} [260, 720] & (N = 1) \\ [450, 1190] & (N = 3) \\ [570, 1460] & (N = 5) \end{array} . \tag{3.49}$$

Direct and Indirect Detection

Scattering of the DM candidate on nuclei is one of the most constraining signal for DM. The experimental bounds, set by numerous experiments [318–320], already forced us to odd values of N to have a zero-hypercharge DM candidate and hence to avoid a direct coupling to the SM matter. On the other hand, our model still contains a tree-level contribution mediated by the Z'. Since DM is non-relativistic, and since $M_{Z'}$ is also much larger than the relevant momentum transfer, the scattering process with the nucleon constituents may be accounted by a local Lagrangian

$$\mathcal{L}_{\chi q} \;=\; \frac{g_{Z'}^2}{12\, M_{Z'}^2} \xi_\chi \left(\bar{\chi}\gamma^\mu \chi\right) \left(\xi_Q^i\, \bar{Q}^i \gamma_\mu Q^i + \xi_u^i\, \bar{u}^i \gamma_\mu u^i + \xi_d^i\, \bar{d}^i \gamma_\mu d^i\right), \tag{3.50}$$

where $g_{Z'} = \sqrt{\frac{3}{2}}\frac{g_4}{\cos\theta_{41}}$. Furthermore, if we are able to neglect corrections due to the finite momentum transfer between the DM and the nucleons, we may parametrize the matrix elements between vector or axial-vector quark q currents and the external-state nucleons N as

$$\begin{aligned}\langle N(p')|\bar{q}\gamma^\mu q|N(p)\rangle_{\vec{p}=\vec{p}'=0} &= F_1^{q/N}(0)\ \bar{u}_N(p')\gamma^\mu u_N(p)\ ,\\ \langle N(p')|\bar{q}\gamma^\mu\gamma_5 q|N(p)\rangle_{\vec{p}=\vec{p}'=0} &= F_A^{q/N}(0)\ \bar{u}_N(p')\gamma^\mu\gamma_5 u_N(p)\ .\end{aligned} \tag{3.51}$$

For the two form factors at zero momentum transfer we follow conventions common in the literature. In particular, $F_1^{q/N}(0)$ counts the number of valence quarks q in the nucleon N, *e.g.* $F_1^{d/n}(0) = 2$. Eqs. (3.50)-(3.51) yield the following spin-independent cross-section for elastic scattering between DM and a single nucleon $N = p$ or n

$$\sigma_{\rm SI}^N \ = \ \frac{g_{Z'}^4\,\xi_\chi^2\,M_N^2}{144\pi\,M_{Z'}^4}|C_V^N|^2\ , \tag{3.52}$$

where

$$C_V^p = \frac{3\xi_Q^1+2\xi_u^1+\xi_d^1}{2}\ , \qquad C_V^n = \frac{3\xi_Q^1+2\xi_d^1+\xi_u^1}{2}\ . \tag{3.53}$$

Starting from eq. (3.52), in order to estimate the matrix element on a nucleus $\mathcal{N}$ with mass number A and atomic number Z, one may assume (see e.g. [321, 322]) that DM scatters coherently on the A nucleons of the target. In the static limit, the DM - nucleon cross section measured by experiments operating with nuclei $\mathcal{N}$ as target material can thus be estimated from eq. (3.52) with the replacement

$$|C_V^N|^2 \to \frac{|ZC_V^p+(A-Z)C_V^n|^2}{A^2}\ . \tag{3.54}$$

This approach is approximate for several reasons (amply discussed in the literature, see *e.g.* [323] for a recent review). First, eqs. (3.50) and (3.51) are defined at different scales and renormalization-group effects should be taken into account. Second, the underlying assumption of a point-like nucleon in (3.51) is challenged by the potentially large DM momentum. However, from a numerical study based on public codes, it appears that for the DM mass ranges of eq. (3.49), this crude approach give a quite realistic result [308].

The spin-independent DM-nucleon cross-section is displayed in Fig. 3.11. Due to a somewhat magical compensation in the $\xi_{Q,u,d}$ terms, the cross-section presents $\sin\theta_{Q_{12}}$-dependent holes. It is quite remarkable that experimental constraints, that can be as strong as $\sigma_{\rm SI}^N < 10^{-45}{\rm cm}^2$, selects region of the parameter space fully compatible with the requirement of the B anomalies. As can be seen on this figure, this limit can be comfortably satisfied with the choice $\sin\theta_{Q_{12}} \lesssim 0.2$ and $g_4 \gtrsim 3$.

As concerns indirect detection, our DM candidate may also be constrained by the production of cosmic rays due to DM annihilations in the Milky Way and

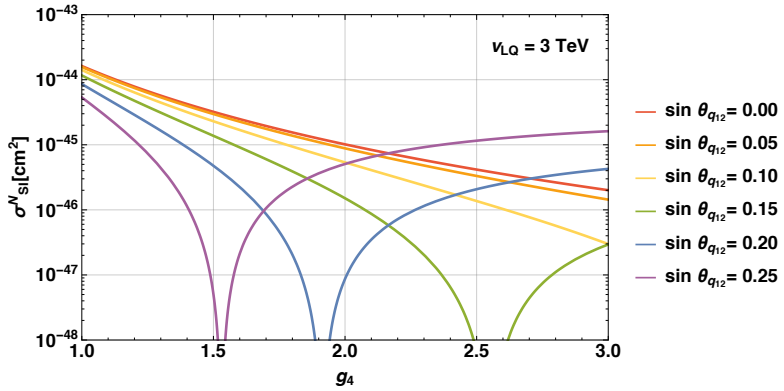

Figure 3.11: The DM - nucleon cross section (see eq. (3.52)) as a function of g_4 for different choices of the mixing angle $\sin\theta_{Q_{12}}$ and $v_{LQ} = 3$ TeV. See text for details.

elsewhere. Out of the numerous channels probed by experiments, three are relevant and potentially constraining for our model, namely $\chi\chi \to \gamma\gamma$, $\tau^+\tau^-$ and W^+W^- (for a recent review see [324]). As already noted below eq. (3.43), the W^+W^- channel has a smaller cross-section than the $\tau^+\tau^-$ one. As the experimental bound is also weaker for the W flux [325], we focus hereafter on the two other channels.

The present-day $\chi_0\chi_0 \to \tau\tau$ cross-section, velocity-averaged within the Milky Way, may be estimated using eq. (3.46) where the limit $x \to \infty$ is amply justified by the present-day temperature. In this respect, we will denote such cross-section simply as σv, following a similar notation used throughout the literature. We obtain

$$\sigma(\chi_0\chi_0 \to \tau\tau)v \simeq N\,\sigma_0(1)|_{f(\{\xi^i\})=6|\xi^3_\ell|^2+3|\xi^3_e|^2}\,, \tag{3.55}$$

where the choice of $f(\{\xi^i\})$ specializes $\sigma_0(1)$ to the $\chi_0\chi_0 \to \tau\tau$ case.

This cross-section is displayed in Fig. 3.12 (left). The width of the prediction corresponds to the range of v_{LQ} values compatible with the $\Omega_0 h^2$ constraint for a given M_χ and taking into account the accuracy of 15% we attach to our analytic $\Omega_0 h^2$ calculation. The dashed line shows the value required to obtain the full relic density, if this were the only annihilation channel. Our prediction is below the current HESS bound [325], but close to it[14]. An improvement of the limit by a factor of a few would offer a valuable probe of our scenario.

The discussion of the $\chi_0\chi_0 \to \gamma\gamma$ channel, induced at one loop via $W-\chi$ or $U_1-\psi$ exchange, is much more involved. The $W-\chi$ contribution is well-known in the large-M_χ limit and yields (*cf. e.g.* [309, 329])[15]

$$\sigma v \simeq (N^2-1)^2\frac{\pi\alpha_{\rm em}^2\alpha_2^2}{16M_W^2} \simeq (N^2-1)^2\cdot 2\times 10^{-29}\ {\rm cm}^3\ {\rm s}^{-1}\,. \tag{3.56}$$

[14] The HESS bound we show is the one obtained with the Einasto profile [328], producing the most constraining bound among the different DM-distribution profiles considered by Ref. [325].
[15] This one-loop result undergoes enhancements of order $2-3$ due to non-perturbative effects [308].

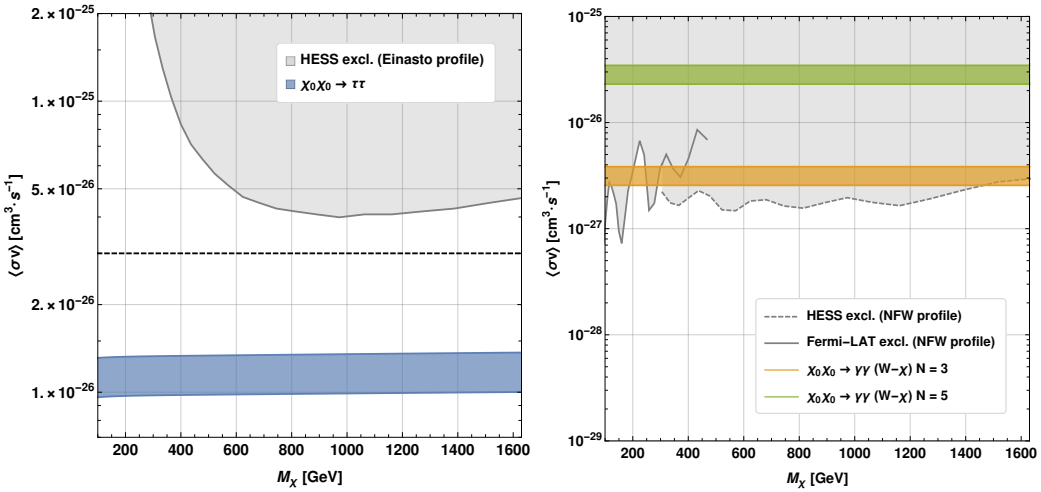

Figure 3.12: Left: Velocity-averaged DM-annihilation cross section to $\tau\tau$ at present time. The grey area denotes the region excluded by HESS [325], whereas the blue band represents the prediction in our model. The dashed line is the value required for $\langle\sigma v\rangle_0$ to saturate the Ωh^2 constraint. Right: $W-\chi$ contribution to $\sigma(\chi_0\chi_0 \to \gamma\gamma)$ for $N = 3, 5$ and corresponding bounds from HESS [326] and Fermi-LAT [327]. See text for further details.

This cross-section may be compared with the 95% CL experimental bound in Refs. [326, 327] as shown in Fig. 3.12 (right). Assuming that interference with $U_1 - \psi$ exchange diagrams does not significantly reduce the cross-section, already in the case $N = 3$ we obtain a velocity-averaged cross-section in excess of $10^{-27}\text{cm}^3 \cdot \text{s}^{-1}$ for M_χ in the range of interest to us (*cf.* eq. (3.49)). The figure also shows the observed exclusion lines from HESS [326] and from Fermi-LAT [327] in the respective M_χ ranges, and assuming a NFW [330] distribution for DM. We see that at face value this constraint favours $N = 1$, and strongly disfavours $N \geq 5$.

For $N = 1$, the $W - \chi$ contribution is zero and a detailed analysis of the $U_1 - \psi$ amplitude is required. Such a calculation is also interesting for $N = 3, 5$ in order to determine the interference terms. An accurate estimate of the $\chi_0\chi_0 \to \gamma\gamma$ cross-section in our model along these lines will be interesting to further test its different scenarios against data. Such comparison will also depend in an important way on the model assumed for the distribution of DM in the Milky Way – keeping in mind that different such models imply different 'best' regions of interest within the Celestial dataset at each given M_χ.

Conclusion

We adopted the well-motivated 4321 completion of the simplified U_1-based model and added a minimal DM sector composed of a fermionic multiplet sitting in the fundamental **4** of $SU(4)$. We showed that the DM phenomenology was mainly

driven by the Z' boson and that a large range of parameters allowed for a combine explanation of B-anomalies and relic and direct detection constraints. These conclusions are presented in Fig. 3.13. Once a complete calculation and an accurate

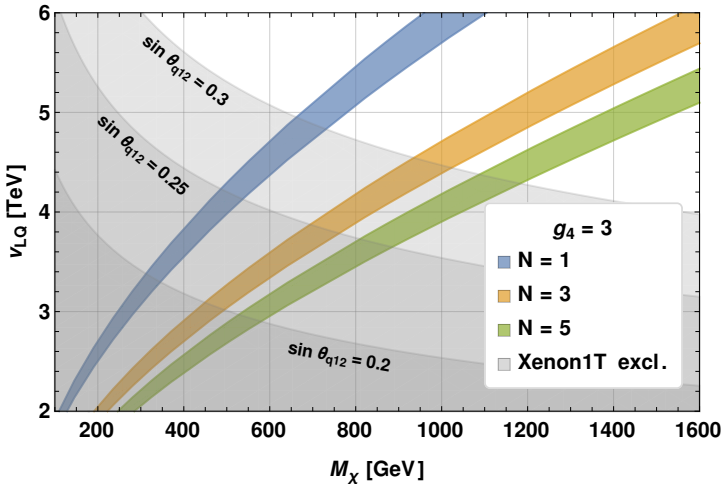

Figure 3.13: DM constraints in the M_χ vs. v_{LQ} plane. Coloured rays fulfil the $\Omega_0 h^2$ constraint within $\pm 15\%$. Grey regions, with $\sin\theta_{Q_{12}}$ fixed at the displayed value, are excluded by Xenon1T [320].

modelling of the astrophysical aspects are available, the constraints from indirect measurements, and in particular the DM annihilation into two photons, could offer a powerful set of tests of this model.

3.2.2 Composite DM related to the horizontal symmetry

The maybe most constraining property of WIMPs is the need to be stable on cosmological time scales. This property often implies the presence of an additional, usually *ad hoc*, symmetry that forbid the decay of the DM candidate. We advocated for instance the presence of such symmetries in the $\mathbf{N = 1}$ and $\mathbf{3}$ cases of the previous section. In the SM, the only global symmetry that provides stability to heavy particles is the $U(1)_B$ baryon symmetry. The conservation of the baryon number makes indeed the mean lifetime of the proton more than 10^{29} years [40].

An immediate question is whether a similar symmetry can arise in an extension of the SM. A new QCD-like confining "hypercolour" (HC) sector would for instance provide an accidental $U(1)_{HB}$ hyperbaryon symmetry under which hyperquarks are charged. The bound states formed by these hyperquarks are then rendered stable by carrying a non-zero $U(1)_{HB}$ charge and are thus potential DM candidates. Several models with a composite DM made stable by a similar symmetry mechanism have been proposed (see [331] for a review). We follow here the discussion presented in Ref. [332].

Apart from the hyperbaryons, the hyperquarks will also form hypermesons

that are *a priori* unstable due to their zero charge under $U(1)_{HB}$. Without an explicit breaking of the hyperquarks' chiral symmetry, the lightest hypermesons are furthermore massless Nambu-Goldstone bosons (NGBs), which yields to phenomenologically unacceptable massless scalars. This breaking can be engineered in two different ways

1. The breaking can be explicit, which would generate a potential for the NGBs, turning them into massive pseudo NGBs (pNGBs).

2. The breaking can be obtained by the gauging of an appropriate subgroup of the chiral symmetry. In this case, the corresponding gauge bosons become massive due to the Higgs mechanism and the would-be NGBs become the gauge bosons' longitudinal degrees of freedom.

These possibilities are both realized in the SM, as exemplified by two-flavour QCD. In the absence of explicit mass terms, the left- and right-handed u and d quarks transform as doublets, $(u_L\,,\,d_L)^T$ and $(u_R\,,\,d_R)^T$, under the global symmetry groups $SU(2)_L$ and $SU(2)_R$, respectively. The global chiral symmetry group is therefore $\mathcal{G}_\chi = SU(2)_L \times SU(2)_R$. This symmetry is spontaneously broken to $\mathcal{H}_V = SU(2)_V$ by the quark condensate $\langle \bar{u}_L u_R + \bar{d}_L d_R + \text{H.c.}\rangle \neq 0$ at the scale $f_\pi \simeq 93$ MeV. The consequence are Goldstone bosons spanning the coset space $\mathcal{G}_\chi/\mathcal{H}_V$, the known pions. As well known, the $SU(2)_L$ group is actually gauged, and this gauge symmetry is spontaneously broken *also* by the Higgs vev $v = 246$ GeV. Because of the hierarchy $f_\pi \ll v$, the pions contribute negligibly to the longitudinal degrees of freedom of the $SU(2)_L$ gauge bosons.

The second breaking possibility is particularly interesting in the context of a DM model since the gauging of a subgroup of the chiral symmetry implies that the corresponding gauge bosons, which we denote by H, couple to the hyperbaryon DM candidates. If these H gauge bosons also couple to some of the SM particles, they then not only solve the problem of massless NGBs, but they could also serve as mediators to a dark sector naturally endowed with a stability mechanism.

This model becomes particularly promising when we consider the involved mass scales. By studying the SM, where $m_p \approx 10\, f_\pi$, we infer that $m_\chi \approx 10\, v_{\text{HC}}$ with v_{HC} the decay constant associated to the dynamical breaking of the chiral symmetry by a hyperquark condensate. Unitarity arguments suggest that the mass of a thermal relic DM candidate cannot exceed 340 TeV [333], yielding a upper bound on the hypercolour scale $v_{\text{HC}} \lesssim 34$ TeV. This limit means that the H bosons are in a range accessible by flavour experiments.

From the previous argument, it is tempting to identify the gauged subgroup of the new chiral symmetry with the $SU(2)_h$ horizontal symmetry introduced in section 2.2. Such a construction shows several assets [332]

- A mass term for the $SU(2)_h$ gauge bosons is generated by the chiral symmetry breaking in the HC sector and the corresponding NGBs become the longitudinal polarizations of these gauge bosons.

- The $SU(2)_h$ gauge bosons bring lepton universality violation in $b \to s$ data.
- The $SU(2)_h$ gauge bosons connect the baryon-like DM candidates in the HC sector to the SM fields, thereby playing the role of the DM mediator.

The DM constraints can then be obtained by pursuing the analogy to SM. Direct detection processes *via* tree-level exchange of $SU(2)_h$ bosons. The Lagrangian that describes such interactions is analogous to the one describing pion-nucleon interactions and can be computed from chiral perturbation theory [334, 335].

On the other hand, the dominant processes contributing to $\chi\bar{\chi}$ annihilation take place in the strongly coupled regime such that it is not straightforward to calculate the cross section from first principles. As it is the case in nucleon-antinucleon annihilation, the main contribution to the cross-section is due to a rearrangement of quark and anti-quark into mesons (for reviews, see *e.g.* [336–339]). An approximation of the $\chi\bar{\chi}$ annihilation cross-section can however be obtained by scaling up the experimental measurements of nucleon-antinucleon annihilation cross-section at low momentum. Several groups provide fits to experimental data on the annihilation cross-section σ_{ann} using the parametrisation

$$\sigma_{\text{ann}} = \frac{1}{m_p^2}\left(A + B\,\frac{m_p}{p_{\text{lab}}} + C\,\frac{m_p^2}{p_{\text{lab}}^2}\right), \tag{3.57}$$

where p_{lab} is the momentum of the antinucleon in the rest frame of the nucleon and m_p is the proton mass. Fit results are shown in table 3.4. In most fits, the coefficient C is set to 0. A non-zero C can provide a slightly better fit at p_{lab} around 75 MeV [340], corresponding to $v \sim 10^{-1}$.

Fit	Ref.	$N\bar{N}$	p_{lab} [GeV]	A	B	C
1	[341]	$p\bar{p}$	$[0.26, 0.47]$	86	84	0
2	[342]	$p\bar{p}$	$[0.40, 0.60]$	66.5 ± 4.1	77.1 ± 2.2	0
3	[343]	$p\bar{p}$	$[1.90, 1.96]$	19.2 ± 5.7	98.1 ± 3.1	0
4	[341, 344]	$n\bar{p}$	$[0.26, 0.47]$	63	63	0
5	[340]	$N\bar{n}$	$[0.05, 0.40]$	150.4 ± 6.8	48.0 ± 2.2	0
6	[340]	$N\bar{n}$	$[0.05, 0.40]$	199.9 ± 10.6	23.9 ± 4.1	2.5 ± 0.4

Table 3.4: Fit parameters describing experimental data on nucleon-antinucleon annihilation. All dimensionful parameters in [340–343] have been expressed in units of m_p such that the coefficients A, B, and C used here are dimensionless.

Results Our model basically have only one parameter $v_{\rm HC}$. Our findings are as follows

- The prediction of the DM relic density can be obtained under the approximation $\sigma_{\rm eff} = \sigma_{\rm ann}|_{m_p \to m_\chi}$, whose validity is discussed in Ref [332] and using the analytic results of the previous section. The result is shown in Fig. 3.14 (left panel) for the different fit of Table 3.4. Remarkably the scale we find for all scenarios falls in the range hinted at by the $b \to s$ anomalies obtained in section 2.2, which corresponds to a DM candidate mass $m_\chi \sim 200$ TeV.

- For such a high mass, the direct-detection cross-section are found to be safely below the experimental bounds.

- The cross-sections relevant for indirect detection signals are bounded by the total annihilation cross-section and can therefore also be estimated from nucleon-antinucleon annihilation data. Extending the fits of Table 3.4 to the region of momentum relevant for today's DM annihilation, namely $p_{\rm lab} \sim 10^{-3} m_\chi$ (see *e.g.* [345]), we obtained the predictions shown in Fig. 3.14 (right panel). The predicted cross-sections are safely below indirect bound, with the exception of fit 6. However, none of the fits takes into account data for $p_{\rm lab} \sim 1$ MeV, which corresponds to relative nucleon velocities as low as 10^{-3}. This result underlines the importance of obtaining nucleon-antinucleon data at low momentum, even though this might be very challenging.

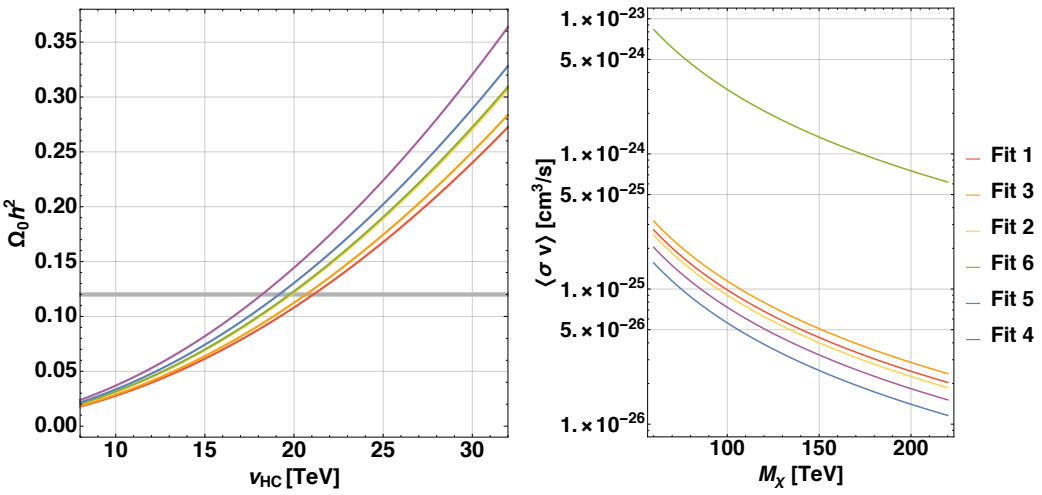

Figure 3.14: Left panel: Present DM relic density as a function of $v_{\rm HC}$ for the different fits in Table 3.4. The thick horizontal gray line represents the measured value $\Omega_0 h^2 = 0.12$. Right panel: estimate (see text for details) of the velocity-averaged DM-DM annihilation cross section into anything else, $\langle \sigma_{\rm ann} v \rangle$, as a function of the DM mass, for the same fits as the left panel.

To conclude, we studied the phenomenology of the hyperbaryon DM, in particular its relic density, expected to be produced by a mechanism of thermal freeze-out, and its direct-detection signals. Dark Matter predictions are fixed by the single scale of our model, which is found to be in agreement with the results of sec. 2.2.

This section concludes our theoretical part. The main results and implications of these chapters will be summarized in chapter 9.3. It is however already clear that new experimental results will be crucial to interpret the current anomalies within a broader coherent theory. The focus is therefore now on the search for $B_s^0 \to \mu^+\mu^-\gamma$ events within LHCb.

Part II

Radiative leptonic decays at LHCb - Experimental measurement

4

The LHCb detector and the $B_s^0 \to \mu^+\mu^-\gamma$ analysis

Contents

The description of the LHCb detector and its performances can be found in several reviews (see *e.g.* [346,347]). For brevity, this chapter only details the properties of the detector relevant for the analysis of $B_s^0 \to \mu^+\mu^-\gamma$ decays. The chapter is organized as follows: LHCb is presented in section 4.1; section 4.2 describes how the trigger and the stripping operates and how events are simulated; section 4.3 gives an overlook of the $B_s^0 \to \mu^+\mu^-\gamma$ analysis and the data and simulated samples we used.

4.1 LHCb, muons and photons

The LHCb detector was designed to look for indirect evidence of new physics in the decays of beauty and charm hadrons. $b\bar{b}$ pairs being produced close to the beam, the detector was conceived as a single-arm spectrometer with an angular coverage of approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. The layout of the detector is sketched in Fig. 4.1. This section is devoted to the description of the parts of the detector relevant for the measurement of muons and photons.

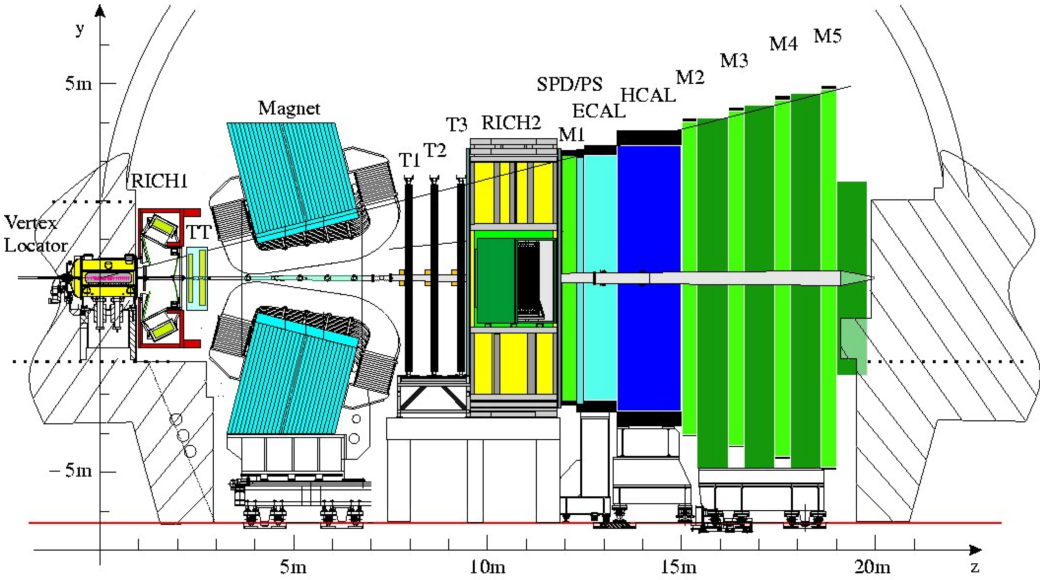

Figure 4.1: LHCb during Run I and II.

4.1.1 Charged tracks and muons in LHCb

LHCb tracking system is composed of several trackers (the Vertex Locator and the four stations TT, T1-T3) that aim at reconstructing charged-particle tracks. The different trajectories and hits measured in the detector define categories of tracks (as pictured in Fig. 4.2). The identification of tracks (PID) as kaons, pions or protons is mainly performed thanks to the ring imaging Cherenkov (RICH) counters.

The muon detection system is crucial in this analysis both for the triggering of events and for the off-line particle identification. The muon detector is composed of five stations (M1-M5) interleaved with iron walls. Muons are the only detected particles that can penetrate these filtering walls, ensuring a good identification of the tracks. These tracks are reconstructed from the linear interpolation of hits in the different sub-detectors. This interpolation defines fields of interests (FoI) and

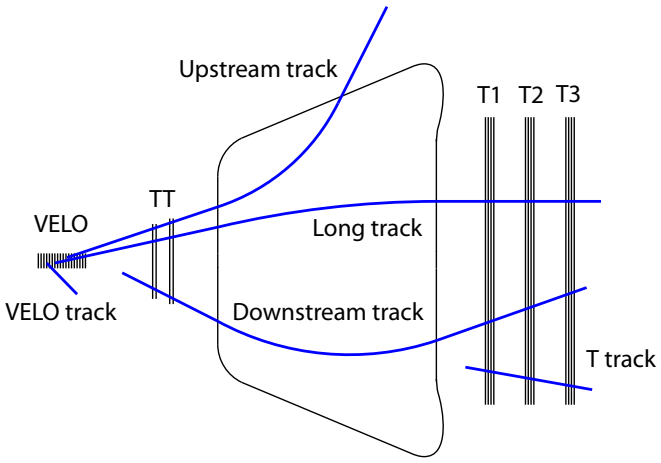

Figure 4.2: The different types of tracks in LHCb. The shape in the middle represents the magnet.

the presence of hits in all the FoIs gives a first criteria for the identification of muons. This criteria, known as `IsMuon` [348] (see also [349, 350] for performance studies) is described in Table 4.1. The quality of the tracks fit is also used for fur-

Muon energy	Required Station
$E_\mu < 3$ GeV	Always false
3 GeV $\leq E_\mu < 6$ GeV	M2 & M3
6 GeV $\leq E_\mu < 10$ GeV	M2 & M3 & (M4 || M5)
10 GeV $\leq E_\mu$	M2 & M3 & M4 & M5

Table 4.1: Required stations with hits in the FoI for the `IsMuon` criteria depending on the tracks momentum.

ther selection (track χ^2). The sub-detectors response to a muon track is measured with calibration samples. A global likelihood is then constructed from those of the sub-detectors (possibly hinging on Neural Network techniques) and allows a further identification of the track ($\Delta \log \mathcal{L}$ and `ProbNN` variables). Another neural-network based algorithm is trained to identify fake tracks and assigns each track a so-called *ghost probability* [351].

The number of tracks in the events (occupancy) has a large impact on the reconstruction and identification performances. To ensure maximal performances,

LHC beams are shifted to keep a constant and contained number of collision per-event (pile-up). Indeed, unlike the other LHC experiments that have a 4π coverage of the interaction point, LHCb has to reconstruct the pp interactions only based on forward tracks[1]. A large pile-up would therefore decrease the resolution at the interaction point and turn the rest of the pp collisions into irreducible backgrounds.

Vertices are fitted from the reconstructed tracks. The interaction point of the pp collision defines the primary vertex (PV), while all other vertices are referred to as secondary vertices (SV). The fact that a track origins from any vertex is hinted by its impact parameter (IP), defined as the minimal distance between the tracks best fit and the vertex. The common origin of two tracks can also be tested with their distance of closest approach (doca) defined likewise as the closest distance between the two tracks best fits.

4.1.2 The photons in LHCb

As already emphasized in Part I, photons are much more difficult to deal with at LHCb than muons or hadrons. As neutral particles, they are not bent by the magnet, and their penetrating power is much weaker than the muon's [40].

The calorimeter system comprises a scintillation pad detector (SPD), a pre-shower (PS) an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). These components are composed of cells and divided into regions of decreasing cell size, as sketched in Fig. 4.3. Neutral particles are reconstructed from clusters that are not associated to reconstructed tracks. The clusterisation is performed using a cellular automaton by growing the clusters iteratively around local maxima [352] (see also [353] for a performance study).

Photons are reconstructed from 3×3 cells clusters in the ECAL. The energy associated to the photon is a linear combination of the energy measured in the cluster and that measured in the corresponding hit in the PS. The associated coefficients are extracted from on-line and off-line calibration samples.

Neutral pions can be reconstructed from two separated photon clusters (resolved π^0) or from a single cluster (merged π^0). Clusters can therefore be caused by photons, electrons, merged π^0 or non-electromagnetic deposits associated to hadrons. Multivariate analyses are trained to separate photons from hadrons, electrons or merged π^0 based on the shape of the electromagnetic shower associated to the cluster. The corresponding variables are referred to as IsNotH (also known as CL[2]), IsNotE and IsPhoton [353, 354].

[1] This is not entirely true because the Vertex Locator covers the interaction point (*cf.* Fig. 4.2). A few backward tracks can therefore be reconstructed.
[2] For Run II CL and IsNotH are the same variable, this was not the case in Run I.

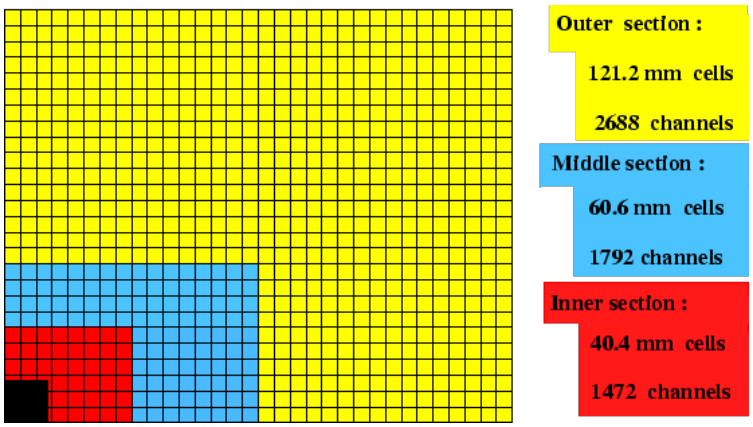

Figure 4.3: Lateral segmentation of the SPD/PS and ECAL (only one quarter of the detector front face is shown). The size of the cells in the Inner, Middle and Outer region are given for the ECAL. The black square corresponds to the empty region of the proton beam.

4.2 Data and simulation paths

The following paragraphs give a brief description of the trigger, stripping and simulation processes.

4.2.1 The event trigger

Given the huge crossing rate (40 MHz) taking place at the interaction point, a stringent selection has to be performed to reduce the amount of events to an operating frequency. The trigger is decomposed in several levels, as sketched in Fig. 4.4. The L0 level is a hardware trigger for which the event is only partially reconstructed. It brings down to rate to 1 MHz. The second and third level are software triggers, known as high-level triggers (Hlt), which allow for more sophisticated selections. The event storage rate is around 12.5 kHz.

Each level is composed of trigger lines that can be specified in the analyses. The status of the Trigger (the available lines and their definition) is stored in a *Trigger Configuration Key* (TCK), that can be specified for the simulation of data. An event can be stored for several reasons:

- If the signal of interest is enough to trigger, the event is referred to as Triggered On Signal (TOS);
- If the event is triggered by the rest of the event (either coming from the decay of the other b quark, or from another collision), it is called Triggered Independent of Signal (TIS);

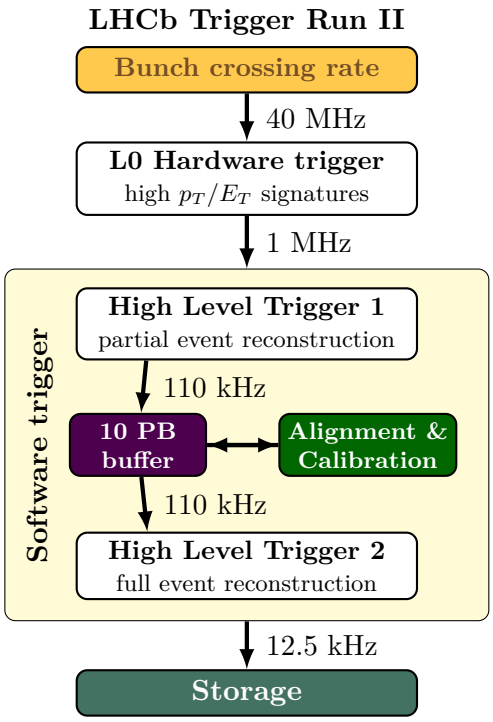

Figure 4.4: Overview of the LHCb trigger system [355] in Run II.

.

- If a combination of the signal and the rest of the event is needed to trigger, the event is Triggered On Both (TOB).

Note that an event can be both TIS and TOS if the signal and the rest of the event are enough to trigger (TISTOS). All other combinations are impossible by construction.

4.2.2 Stripping

Triggered events are classified in streams during a first and flexible selection step called stripping. The main goal of the stripping is to centralize and optimize the search for signal decay trees in the large amount of triggered events. As it is the case for the trigger, the stripping is decomposed in lines specific to several analyses. The lines contain the reconstruction procedure and a minimal set of cut to efficiency select the signal of interest.

4.2.3 Simulated data

Most of the analysis looking for very rare events rely on signal simulation for the computation of selection efficiencies. Indeed, methods to measure these efficiencies directly on data exist, but they are usually based on signal events. The use of proxy

channels or efficiency maps allow to bypass this issue, but they still require a sketch of the signal kinematics.

The simulation of events is based on a Monte-Carlo (MC) method and divided into several steps. pp collisions are simulated using PYTHIA and a specific LHCb configuration [356–358]. The decay of hadronic particles is performed with EVTGEN [359]. We already mentioned the PHOTOS algorithm that simulates final-state radiation [167]. The interaction with LHCb material is implemented in GEANT4 [360–362]. The main simulation steps are stored in a file specific for each decay, called DecFile, which is used during the MC production.

Once events are simulated, they are digitalized, and the very same reconstruction, trigger and stripping algorithms as applied to data are applied. This similarity ensures a minimal bias between data and MC samples. Measuring and accounting for the resulting discrepancies between the two samples is one of the main challenges of the analyses.

4.3 The $B_s^0 \to \mu^+\mu^-\gamma$ analysis

As described in Part I, two complementary methods can be used to measure $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio. The indirect method, detailed in section 1.3 consists in probing this decay as a background of $B_s^0 \to \mu^+\mu^-$, namely without reconstructing the photon. Although very promising, this method only probes the high-q^2 region and may miss new physics contributions. The second method is the classical (direct) analysis, with full final state reconstruction, and is presented here.

LHCb's sensitivity to this measurement was first studied in [363]. From preliminary studies on simulated signal and background samples, it was shown that, assuming no signal is found, the 90% C.L. upper limit can be set to $6.02^{+2.37}_{-1.39} \times 10^{-9}$ with 2/fb.

4.3.1 Analysis overlook

The search for $B_s^0 \to \mu^+\mu^-\gamma$ decays is performed using data samples of proton-proton collisions corresponding to an integrated luminosity of 5.57/fb collected by the LHCb experiment from 2016 to 2018 at a centre-of-mass energy of 13 TeV. Table 4.2 lists the integrated luminosities collected each year and the versions of the reconstruction software used. 2015 dataset is not used for several reasons, mainly related to different trigger conditions. Run I is not used either, because the dedicated stripping line was not yet available. Without this stripping line, whose introduction in the software was the first step of the present analysis, the stripping efficiency is too small to yield an interesting sensitivity.

This analysis is a blinded analysis with full final state reconstruction. The number of events is normalized using $\boldsymbol{B_s^0 \to (J/\psi \to \mu\mu)(\eta \to \gamma\gamma)}$ decays and

Year	Integrated Luminosity (fb^{-1})	Reco Version	Stripping Version
2016	1.67	16	28r1p1
2017	1.71	17	29r2
2018	2.19	18	34

Table 4.2: Integrated luminosities and versions of the reconstruction software and the stripping used for each year.

controlled via $\boldsymbol{B_s^0 \to (\phi \to KK)\gamma}$. The analysis is performed in 3 dimuon mass bins:

Bin I The low-q^2 region, $m(\mu\mu) < 1700$ MeV;

Bin II The middle-q^2 region, 1700 MeV$< m(\mu\mu) <$ 2880 MeV;

Bin III The high-q^2 region, $m(\mu\mu) > 3920$ MeV.

The value of 1700 MeV is chosen to be above the ϕ and ϕ' resonances, see Figure 1.3. Using the theoretical differential branching ratio derived in section 1.1.3 allows to compute the fraction of signal in each bin ; results are given in Table 4.3.

Bin	I	II	III
$10^{10} \times \mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma)$	82 ± 15	2.54 ± 0.34	9.1 ± 1.1
Fraction of $B_s^0 \to \mu^+\mu^-\gamma$ events	87%	2.7%	9.8%

Table 4.3: Branching fraction and fraction of events in the different q^2 bins. These numbers are obtained using the spectrum discussed in section 1.1.3.

A LHCb event usually contains a lot of low energy photons that mainly come from π^0 decays of the underlying events. As the calorimeter only measures an energy deposit, the photon direction is only known approximatively [3]. Picking the right photon from all these candidates is therefore the most difficult part of this analysis. Some analyses avoid this problem by using converted photons, *i.e.* photons that decay into an electron-positron pair in the detector material before the magnet. By measuring the electron momentum, we can infer information on the photon and improve the selection based on the B_s^0 vertex. However the

[3] By default, the photon is considered as emitted from the origin of the reconstruction frame, in the Vertex Locator. The decay tree can be kinetically adjusted to attach the photon to the secondary vertex. This improves the mass peak width, but mixes signal and backgrounds, see section 5.5.2.

probability of conversion is small (around 20% of the photons get converted before the magnet [364]) and the bremsstrahlung emissions of the electrons have to be properly handled. We therefore decided to keep only calorimetric photons.

The selection strategy is similar to the one applied for the measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [365]. The event are triggered on signal (TOS) with the muons or the photon and independently of signal (TIS). A stripping line has been written for this decay and selects events with a high efficiency. The background is then removed in two steps:

- A first MVA, mainly based on geometrical variables, efficiently removes a lot of combinatorial background.
- A second MVA, based on isolation and geometrical variables, allows for a further reduction of the background.

4.3.2 Monte-Carlo simulations

A large number of simulated samples is used in the analysis. Table 4.4 gives the characteristics of all the MC that were used. Concerning $B_s^0 \to J/\psi\,\eta$ no simulation was performed for the years 2017 and 2018. Indeed no changes were expected in the decay geometry with respect to 2016, as confirmed by the signal and control samples, and trigger and PID efficiencies are extracted from data.

$B_s^0 \to \mu^+\mu^-\gamma$ events are simulated using the differential branching ratio computed in section 1.1.3 as theory input, as detailed in [366]. In particular, the $B \to \gamma$ form factors are implemented as in eq. (1.15) with the numerical inputs gathered in Table 1.1. The ϕ resonance is implemented as in $n = 1$ version of eq. (1.32) while the ϕ' resonance, considered as negligible, is not implemented.

The following chapters are organized as follows, section 5 describes the selection (trigger, stripping and particle identification); section 6 contains the global multivariate analysis used to classify the events; section 7 summarizes the backgrounds and their treatments; section 8 presents the normalization and control channels as well as the expected sensitivity of the analysis.

Decay	DecFile	Year	Sim	Reco	TCK	Stripping	Events
$B_s^0 \to \mu^+\mu^-\gamma$	13112203	2016	09c	16	0x6138160f	28r1p1	600k*
		2017	09h	17	0x62661709	29r2	1200k*
		2018	09h	18	0x617d18a4	34	1500k*
$B_s^0 \to J/\psi\,\eta\ (\mu\mu)(\gamma\gamma)$	13142401	2016	09b	16	0x6138160f	28	2M
$B_s^0 \to \phi\gamma\ (K\pi)$	13102202	2016	09c	16	0x6138160f	28r1	4M
		2017	09f	17	0x62661709	29r2	4M
		2018	09f	18	0x617d18a4	34	4M
$B^0 \to \mu^+\mu^-\pi^0\ (\gamma\gamma)$	11112401	2016	09b	16	0x6138160f	26	2M
		2017	09h	17	0x62661709	29r2	2M
		2018	09h	18	0x617d18a4	34	2M
$B^0 \to K^{*0}\gamma\ (K\pi)$	11102201	2016	09b	16	0x6138160f	28	2M
	11102202	2017	09f	17	0x62661709	29r2	4M
		2018	09f	18	0x617d18a4	34	4M
$B^+ \to J/\psi\eta K^+\ (\mu\mu)(\gamma\gamma)$	12243402	2016	09c	16	0x6138160f	28r1	1M
$B^+ \to \phi\gamma K^+\,(K\pi)$	12103202	2016	09c	16	0x6138160f	28r1	1M

Table 4.4: Samples simulated for the analysis. All candidates are required to have their children particles in the LHCb angular acceptance. The star means that the MC was stripped, which means that only the candidates that passed the stripping line are saved.

5 Signal Selection

Contents

The reconstruction of $B_s^0 \to \mu^+\mu^-\gamma$ events is challenging at LHCb. The energy of the B_s^0 is shared between the photon and the muon pair implying that events will either have a low-energy photon, that needs to be selected from the large electromagnetic background, or a low energy dimuon pair, which is disfavoured by the muon trigger. This chapter describes the successive selections applied on signal, normalization and control samples. The selection is performed in 3 steps: a specific stripping line, a general and mainly geometrical multivariate analysis and a second, efficient, multivariate analysis selection. The global selection efficiency is written as the product of partial efficiencies:

$$\varepsilon^{\text{Tot}} = \varepsilon^{\text{Acc}} \times \varepsilon^{\text{Strip \& Reco}} \times \varepsilon^{\text{Sel}} \times \varepsilon^{\text{Trig \& PID}} \times \varepsilon^{\text{MLP}}, \tag{5.1}$$

where each term is estimated either on simulation or using data:

$\boldsymbol{\varepsilon^{\text{Acc}}}$ is the acceptance efficiency and is measured during the MC simulations production. See section 5.1.

$\varepsilon^{\text{Strip \& Reco}}$ is the efficiency of event reconstruction and stripping. It is measured on MC simulations corrected for their mis-simulated variables. See section 5.3.

ε^{Sel} is the selection efficiency. The selection is based on a multivariate analysis and the efficiency is measured on simulated events. See section 5.5.

$\varepsilon^{\text{Trig \& PID}}$ is the trigger and particle identification efficiency. Although it appears chronologically before the previous steps, it is measured for selected events only. It is estimated on data using efficiency tables. See sections 5.2 and 5.4 for the description of these steps and 8.4 for the computation of the trigger efficiency.

ε^{MLP} is the efficiency of the final multivariate analysis classifier. This selection and its efficiency are presented in section 6.3.

5.1 Geometrical acceptance

The acceptance efficiency is defined as the ratio of decays with children particles in the $[10, 400]$ mrad range to the total number of decays. This range is larger than the fiducial acceptance, in order to recover particles bent by the magnet. The loss of efficiency due to particles in acceptance bent out by the magnet is taken into account in the reconstruction efficiency.

Acceptance efficiency is measured on simulated samples during MC production by simply counting the number of generated events in acceptance. The efficiencies for the channels of interest are summarized in Table 5.1. The efficiency is also computed for $B_s^0 \to \mu^+\mu^-\gamma$ events in each q^2 bin (defined in section 4.3.1) and given in Table 5.2. As this information cannot be extracted from the simulated events based on the DecFile 13311203 (no q^2 bins), we generated small samples of events (10000 events per year) by adding the q^2 cuts in the DecFiles.

Systematic errors can arise from discrepancies between data and simulation. We consider however that, as the acceptance is a purely geometrical property, these discrepancies will cancel in the ratio between signal and normalization.

5.2 Trigger requirements

The trigger strategy is based on the fact that the energy in $B_s^0 \to \mu^+\mu^-\gamma$ events is shared between the muon pair and the photon. Events with a soft photon will mainly be triggered by the muon trigger, while events with a soft muon pair will rather rely on `L0Photon` as depicted in Fig. 5.1. To maximize the statistics, we also trigger independently of signal (TIS).

Decay	Year	$\varepsilon^{\text{Acc}}_{\text{Down}}$ [%]	$\varepsilon^{\text{Acc}}_{\text{Up}}$ [%]	ε^{Acc} [%]
$B^0_s \to \mu^+\mu^-\gamma$	2016	22.676 ± 0.068	22.562 ± 0.068	22.619 ± 0.093
	2017	22.550 ± 0.067	22.580 ± 0.067	22.565 ± 0.093
	2018	22.782 ± 0.068	22.703 ± 0.068	22.742 ± 0.093
$B^0_s \to J/\psi\,\eta$	2016	16.344 ± 0.035	16.244 ± 0.037	16.294 ± 0.051
$B^0_s \to \phi\gamma$	2016	26.885 ± 0.084	27.129 ± 0.085	27.01 ± 0.12
	2017	27.05 ± 0.11	27.01 ± 0.11	27.03 ± 0.16
	2018	27.00 ± 0.11	26.95 ± 0.11	26.97 ± 0.16

Table 5.1: Efficiency of the acceptance cuts measured on MC production (Simulation Table) for up and down magnet polarity. The quoted error is purely statistical.

Decay	Year	$\varepsilon^{\text{Acc}}_{\text{Bin I}}$ [%]	$\varepsilon^{\text{Acc}}_{\text{Bin II}}$ [%]	$\varepsilon^{\text{Acc}}_{\text{Bin III}}$ [%]
$B^0_s \to \mu^+\mu^-\gamma$	2016	22.81 ± 0.32	20.06 ± 0.29	20.13 ± 0.28
	2017	22.73 ± 0.32	20.31 ± 0.29	19.56 ± 0.28
	2018	22.20 ± 0.32	20.13 ± 0.29	19.79 ± 0.28

Table 5.2: Efficiency of $B^0_s \to \mu^+\mu^-\gamma$ acceptance cuts per q^2 bins, measured on local MC production. The quoted error is purely statistical and corresponds to the generated number of events (10000 for each year). Averaging these results using the fraction of event per bin, one recover the signal efficiency quoted in Table 5.1.

5.2.1 L0 Trigger On Signal (TOS)

We use[1]:

$$\text{L0} \equiv \texttt{L0Muon}||\texttt{L0Photon}\,. \tag{5.2}$$

The `L0Photon` line exclusively triggers 10% of the simulated events before selection. Due to their energetic photon, these events are preferred by the multivariate analysis in the rest of the selection, and the 10% reflects a larger part of the selected events ($\sim 17\%$ after the selection cuts). `L0DiMuon` is not used because its cuts are partially superseded by further cuts in the stripping line. These extra events have quite high detector occupancy ($450 <$ `nSPDHits` < 900) and therefore mainly add backgrounds. `L0Electron` is not used either as it showed no improvement (no event is selected exclusively by this line in the stripped and selected sample) and again decreases the purity of the sample. Adding it would improve the efficiency of the trigger by 3% *before* the selection cuts, but these events are not selected by the further cuts and the efficiency on selected events is sub-percent.

[1] The trigger lines requirements are given in Appendix B.1.

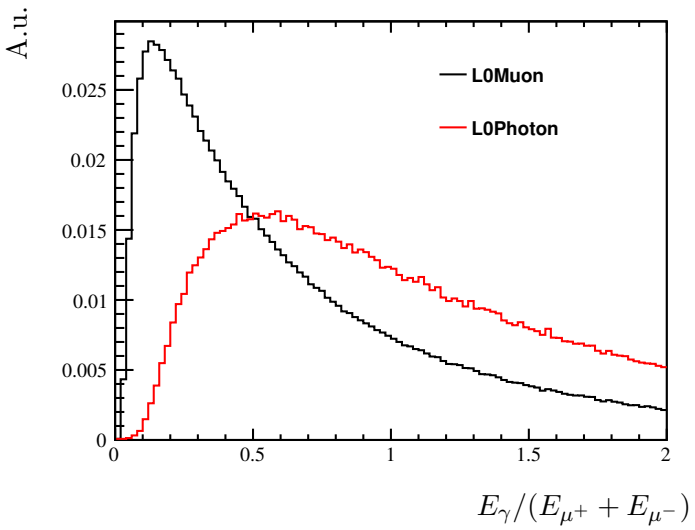

Figure 5.1: Normalized distributions of the ratio between the photon and the muons energy for signal events passing `L0Photon` or `L0Muon`. The selection of events is shared between the trigger lines depending on the energy of the photon and the muon pair. (A.u. stands for arbitrary units and will be used throughout this part.)

5.2.2 Hlt1 Trigger On Signal (TOS)

To choose the high level trigger lines, we applied the following algorithm on simulated events:

1. Choose the Hlt physical line that selects the largest number of events.
2. While the selection efficiency of the remaining lines is larger than 1%, remove the events selected by this line and restart from 1.

When the efficiency of the remaining lines is too small (sub-percent), we considered the set of lines as optimal. At Hlt1 level, events are efficiently triggered by multivariate analysis-based lines and muonic lines. Following the previous algorithm, we used the following selection criteria:

$$\text{Hlt1} \equiv \texttt{Hlt1TrackMVA}||\texttt{Hlt1TwoTrackMVA}||\texttt{Hlt1TrackMuon}||\texttt{Hlt1DiMuonLowMass}\,. \tag{5.3}$$

5.2.3 Hlt2 Trigger On Signal (TOS)

No specific Hlt2 line exists for $B_s^0 \to \mu^+\mu^-\gamma$ decays. The same algorithm than for Hlt1 is applied and the Hlt2 decision is fixed to

$$\text{Hlt2} \equiv \texttt{Hlt2SingleMuon}||\texttt{Hlt2SingleMuonLowPT}||\texttt{Hlt2TopoMuMu2Body}\,. \tag{5.4}$$

The trigger efficiency is computed on events passing the complete selection and discussed in section 8.4. Radiative Hlt lines are poorly efficient in this analysis due to their large cut on the photon transverse energy and are not used. One could thus expect a limited improvement from the L0Photon line. Events selected exclusively by this line (*i.e.* failing L0Muon) are however recovered by the more refined selection of the Hlt muon lines.

For the reader convenience, the most efficient lines are summarized in Table 5.3 for signal events and in Table 5.5 for $B_s^0 \to J/\psi\,\eta$ events. In these tables, only lines that show a non-negligible efficiency on the sample after the full selection are shown. This explains why the efficiency of the logical OR of the chosen lines differs from the efficiency obtained when keeping all lines (referred to as L0Global, Hlt1Phys and Hlt2Phys).

Line	2016	2017	2018
L0Muon	68.3%	75.3%	69.5%
L0DiMuon	38.6%	40.5%	34.4%
L0Photon	29.8%	29.2%	24.1%
|| of these lines	81.2%	85.2%	79.6%
L0Global	89.5%	91.7%	87.2%
Hlt1TrackMVA	74.6%	74.0%	75.1%
Hlt1TwoTrackMVA	59.4%	59.1%	59.8%
Hlt1TrackMuon	71.0%	74.9%	72.7%
Hlt1DiMuonLowMass	63.6%	67.2%	65.2%
|| of these lines	85.8%	87.2%	86.2%
Hlt1Phys	98.7%	98.8%	98.7%
Hlt2SingleMuon	57.1%	60.8%	58.5%
Hlt2TopoMuMu2Body	60.8%	60.4%	62.4%
Hlt2SingleMuonLowPT	21.6%	21.2%	21.9%
|| of these lines	75.7%	77.5%	77.0%
Hlt2Phys	85.7%	91.3%	92.2%

Table 5.3: Most efficient trigger lines for $B_s^0 \to \mu^+\mu^-\gamma$ events. The efficiencies correspond to the number of events triggered on signal (TOS) compared to the number of stripped events for L0 lines, to the number of events selected at the L0 level for Hlt1 lines and to the number of events selected at the Hlt1 level for Hlt2 lines. These efficiencies are given for presentation purpose, the trigger efficiency is precisely measured in section 8.4, where the large differences in L0 efficiencies are explained by changes of the transverse momentum thresholds.

5.2.4 Trigger Independent of Signal (TIS)

To maximise the signal sample, events triggered by the rest of the event are also considered. The efficiency of this trigger is only slightly correlated with the signal kinematics due to the common origin of the $b\bar{b}$ pair in the collision, which improves the efficiency in kinematic regions not reachable by the TOS trigger. The TIS requirement is chosen to be fully generic

$$\text{TIS} \equiv \texttt{L0Global\_TIS} \;\&\&\; \texttt{Hlt1Phys\_TIS} \;\&\&\; \texttt{Hlt2Phys\_TIS}\,, \tag{5.5}$$

and its efficiency is computed in section 8.4.

5.3 Reconstruction and Stripping

5.3.1 Signal

$B_s^0 \to \mu^+\mu^-\gamma$ decays required a specific stripping line as none of the previously existing lines selected them with a fair efficiency. The main difficulty is that, as for the trigger, none of the other lines provided a proper coverage of the q^2 range. A stripping line (`Bs2MuMuGammaLine`) has therefore been written and its requirements are described in Table 5.4.

The difficulty of the stripping selection lies in the choice of the photon transverse momentum. The simplest way to get rid of the large photon background is to increase the cut on the photon energy. However, as pictured in Fig. 5.2 (left), the number of events decreases rapidly with the cut. For example, a cut at 2000 MeV on the photon transverse momentum reduces the number of signal events by 20%. On the other hand, cutting the photon transverse momentum at 500 MeV increases the retention rate of the stripping line by a factor of almost 5 which exceeds the allocated bandwidth. This is mainly due to the π^0 background. Indeed, as depicted in Fig. 5.2 (right), the reconstruction efficiency of $B^0 \to \mu^+\mu^-\pi^0$ is 60% smaller at 2000 MeV.

The choice of the cut is therefore a compromise between the signal efficiency and the π^0 rejection efficiency that impacts the retention rate. As the photon transverse momentum is a variable of the multivariate analysis, we chose to apply the loosest possible cut, $\text{PT}_\gamma > 1000$ MeV, to maximize the signal efficiency. This value of 1000 MeV is the smallest that keeps the retention rate of the line within the allocated range.

The charmonium resonances are excluded by a dimuon mass veto as they come together with a lot of peaking backgrounds (*e.g.* $B_s^0 \to J/\psi\,\eta$). As discussed in Part I, they also present larger theoretical uncertainties. The veto regions is chosen to be $[2880\text{ MeV}, 3920\text{ MeV}]$. These numbers and the remaining background from J/ψ decays is discussed in section 7.

For computing reasons, the MC is stripped, which means that only events passing the stripping line are fully reconstructed and stored. As the stripping line

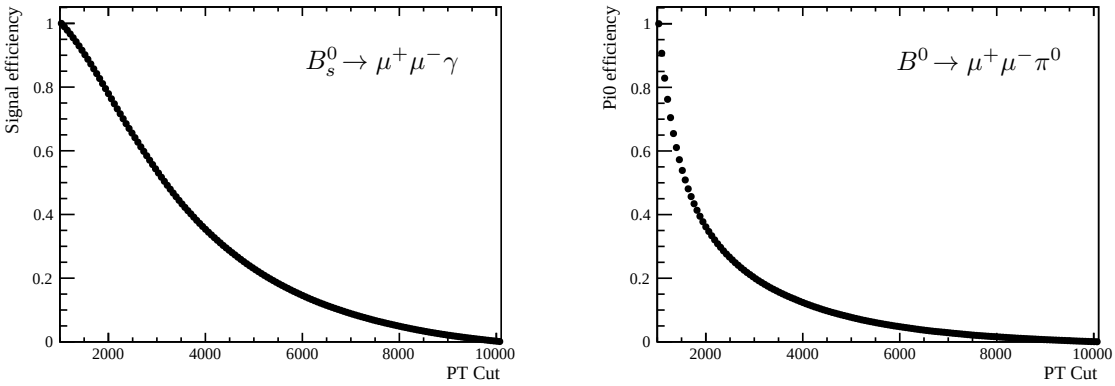

Figure 5.2: Loss of efficiency for a cut on the photon transverse energy for $B_s^0 \to \mu^+\mu^-\gamma$ simulated events (left) and $B^0 \to \mu^+\mu^-\pi^0$. (Efficiency is normalized to 1 for the chosen cut $P_{\mathrm{T}} > 1000$ MeV. For example, half of the photons of the signal have a transverse energy between 1000 and 3000 MeV.)

Particle	Cut	Value
γ	p_{T}	> 1000 MeV
	CL	> 0.2
μ's	p_{T}	> 250 MeV
	track χ^2/ndf	< 3
	track IP χ^2	> 9
	track Ghost Prob	< 0.4
	doca	< 0.5 mm
	$m(\mu\mu)$	$\notin [2880, 3920]$ MeV
B_s^0	p_{T}	> 500 MeV
	dira	> 0.995
	Δm	< 1500 MeV
	PV IP χ^2	< 20
	End Vertex χ^2	< 15

Table 5.4: Requirement of the stripping line `Bs2MuMuGammaLine`. Δm stands for the absolute value of the difference between the mass of the candidate and the B_s^0 PDG mass [367]. A veto is applied on the dimuon mass, $m(\mu\mu)$, to exclude charmonium resonances.

uses part of the reconstruction, the stripping and reconstruction efficiencies cannot be easily disentangled and are computed together. The ratio of truth-matched MC

events to the number of simulated events before reconstruction gives the efficiency of Stripping and Reconstruction for events in acceptance: $\varepsilon^{\text{Strip \& Reco|Acc}}$. These efficiencies are summarized in Table 5.8.

5.3.2 Normalization channel $B_s^0 \to J/\psi\,\eta$

The normalization strategy is described in section 8. $B_s^0 \to J/\psi\,\eta$ events are reconstructed from a muon and a photon pair. The same L0 and Hlt1 trigger requirements as for the signal are applied. For Hlt2, we use the line dedicated to $J/\psi \to \mu\mu$ decays, `Hlt2DiMuonDetachedJPsi`. The efficiency of these lines are sketched in Table 5.5 and precisely measured in section 8.4.

Line	2016
L0Muon	81.4%
L0DiMuon	52.4%
L0Photon	21.1%
|| of these lines	86.4%
L0Global	88.3%
Hlt1TrackMVA	87.4%
Hlt1TwoTrackMVA	69.0%
Hlt1TrackMuon	85.3%
Hlt1DiMuonLowMass	82.5%
|| of these lines	94.8%
Hlt1Phys	99.5%
Hlt2DiMuonDetachedJPsi	99.8%
Hlt2Phys	100%

Table 5.5: Most efficient trigger lines for $B_s^0 \to J/\psi\,\eta$ events. The efficiencies correspond to the number of events triggered on signal (TOS) compared to the number of stripped events for L0 lines, to the number of events selected at the L0 level for Hlt1 lines and to the number of events selected at the Hlt1 level for Hlt2 lines. These efficiencies are given for presentation purpose, the trigger efficiency is precisely measured in section 8.4.

The stripping line `FullDSTDiMuonJpsi2MuMuDetachedLine` is used to get the J/ψ and the η is taken in the `StdLooseEta2gg` container[2]. The selection requirements are chosen to be as close as possible as for the signal, as depicted in Table 5.6. The efficiencies of the stripping and reconstruction processes are computed on MC simulation and summarized in Table 5.8.

[2] Containers are set of particles produced during the reconstruction process and that can be used as basic units for the stripping lines.

Particle	Cut	Value
γ's	p_T	> 1000 MeV
	CL	> 0.2
η	p_T	> 1000 MeV
	Δm	< 105 MeV
μ's	p_T	> 500 MeV
	track χ^2/ndf	< 3
	doca χ^2	< 30
	doca	< 0.5 mm
	track IP χ^2	> 9
J/ψ	Δm	< 100 MeV
	Vertex χ^2/ndf	< 20
	DLS	> 3
B_s^0	p_T	> 500 MeV
	dira	> 0.995
	Δm	< 1500 MeV
	PV IP χ^2	< 20
	End Vertex χ^2	< 15

Table 5.6: Selection requirements for the normalization channel $B_s^0 \to J/\psi\eta$. The cuts are a combination of those of the stripping line `FullDSTDiMuonJpsi2MuMuDetachedLine`, the container `StdLooseEta2gg` and the ones applied to signal. DLS stands for decay length significance and corresponds to the ratio between the decay length and its uncertainty.

5.3.3 Control channel $B_s^0 \to \phi\gamma$

The control channel is used to measure potential discrepancies between data and MC, as described in section 8. $B_s^0 \to \phi\gamma$ events are triggered by `L0Photon`, `Hlt1TrackMVA` and `Hlt1TwoTrackMVA` lines. The exclusive Hlt2 line `Hlt2Bs2PhiGamma` is used. Kaons and photons are then extracted from the `Beauty2XGammaExclusiveBs2PhiGammaLine` stripping line, and the decay is reconstructed as a 3-body decay. This procedure ensures that no bias from reconstruction modifies the geometrical variables distributions. The selection cuts are again chosen to approach the one used for signal. The stripping and selection requirements are described in Table 5.7.

The efficiencies of the stripping and reconstruction processes are computed on MC simulation and summarized in Tables 5.8 and 5.9.

Particle	Cut	Value
γ	p_T	> 2500 MeV
	CL	> 0.2
K's	p_T	> 500 MeV
	E	> 3000 MeV
	track χ^2/ndf	< 3
	track IP χ^2	> 16
	track Ghost Prob	< 0.4
	doca	< 0.5 mm
ϕ	Vertex χ^2/ndf	< 9
	Δm	< 15 MeV
	$\sum_{\text{tracks}} p_T$	> 1500 MeV
B_s^0	p_T	> 2000 MeV
	$\sum_{\text{tracks}} p_T$	> 3000 MeV
	dira	> 0.995
	Δm	< 1500 MeV
	PV IP χ^2	< 9
	End Vertex χ^2	< 9

Table 5.7: Selection requirements for the control channel $B_s^0 \to \phi\gamma$. The cuts are a combination of those of the stripping line `Beauty2XGammaExclusiveBs2PhiGammaLine` and the ones applied to signal.

5.4 Particle identification

5.4.1 Muon identification

A correct identification of the two muons is required to avoid the peaking backgrounds $B_s^0 \to \phi\gamma$ and $B^0 \to K^{*0}\gamma$ and to suppress the combinatorial background coming from hadrons. The identification procedure is similar to the one applied in the $B_s^0 \to \mu^+\mu^-$ analysis, namely a two steps procedure: the muons need to fulfil the `IsMuon` requirement Table 4.1 and then a further PID selection is applied. The different PID variables ($\Delta \log \mathcal{L}$, `ProbNN`) are known to give similar results, so we have chosen to use the `ProbNN` variables, in line with the $B_s^0 \to \mu^+\mu^-$ analysis. These variables are the output of a multivariate classifier that combines PID information (from RICH, calorimeter and trackers) into a probability for each particle hypothesis. The latest version of these variables, known as `MC15TuneV1`, is used. Following the study performed for $B_s^0 \to \mu^+\mu^-$ [365], a global PID variable is com-

Decay	Year	$\varepsilon^{\text{Strip \& Reco\|Acc}}_{\text{Down}}$ [%]	$\varepsilon^{\text{Strip \& Reco\|Acc}}_{\text{Up}}$ [%]	$\varepsilon^{\text{Strip \& Reco\|Acc}}$ [%]
$B^0_s \to \mu^+\mu^-\gamma$	2016	$11.21 \pm 0.02 \pm 0.13$	$11.18 \pm 0.02 \pm 0.09$	$11.19 \pm 0.03 \pm 0.16$
	2017	$11.23 \pm 0.02 \pm 0.05$	$11.21 \pm 0.02 \pm 0.06$	$11.22 \pm 0.03 \pm 0.08$
	2018	$11.21 \pm 0.02 \pm 0.06$	$11.20 \pm 0.02 \pm 0.06$	$11.20 \pm 0.03 \pm 0.08$
$B^0_s \to J/\psi\,\eta$	2016	$3.54 \pm 0.02 \pm 0.06$	$3.58 \pm 0.02 \pm 0.06$	$3.56 \pm 0.03 \pm 0.08$
$B^0_s \to \phi\gamma$	2016	$5.68 \pm 0.02 \pm 0.18$	$5.67 \pm 0.02 \pm 0.19$	$5.67 \pm 0.03 \pm 0.25$
	2017	$5.57 \pm 0.02 \pm 0.17$	$5.59 \pm 0.02 \pm 0.18$	$5.58 \pm 0.03 \pm 0.24$
	2018	$5.56 \pm 0.02 \pm 0.18$	$5.57 \pm 0.02 \pm 0.18$	$5.56 \pm 0.03 \pm 0.25$

Table 5.8: Stripping and reconstruction efficiency computed as the ratio of truth-matched events to simulated events in acceptance. The first error is a statistical error (Poisson), the second is a systematic error due to the truth matching algorithm, as described in section 9.1.1. The lower efficiency in $B^0_s \to \phi\gamma$ and $B^0_s \to J/\psi\,\eta$ is due to the fact that the cuts are chosen to be as similar as possible to the cuts used for the signal and the fact that $B^0_s \to J/\psi\,\eta$ is a 2×2 body decay. The selection is therefore not optimized for these channels.

Decay	Year	$\varepsilon^{\text{Strip \& Reco\|Acc}}_{\text{Bin I}}$ [%]	$\varepsilon^{\text{Strip \& Reco\|Acc}}_{\text{Bin II}}$ [%]	$\varepsilon^{\text{Strip \& Reco\|Acc}}_{\text{Bin III}}$ [%]
$B^0_s \to \mu^+\mu^-\gamma$	2016	$12.21 \pm 0.03 \pm 0.14$	$9.94 \pm 0.13 \pm 0.004$	$3.45 \pm 0.04 \pm 0.02$
	2017	$12.24 \pm 0.03 \pm 0.07$	$9.93 \pm 0.10 \pm 0.002$	$3.49 \pm 0.03 \pm 0.01$
	2018	$12.22 \pm 0.03 \pm 0.07$	$9.89 \pm 0.08 \pm 0.002$	$3.47 \pm 0.03 \pm 0.01$

Table 5.9: $B^0_s \to \mu^+\mu^-\gamma$ stripping and reconstruction efficiency for each q^2 bin. The first error is a statistical error (Poisson), the second is a systematic error due to the truth matching algorithm, as described in section 4.3.2.

puted by multiplying the `ProbNN` variables

$$\text{PID}_\mu \equiv \texttt{ProbNN}_\mu \times (1 - \texttt{ProbNN}_p) \times (1 - \texttt{ProbNN}_K)\,. \tag{5.6}$$

The $(1 - \texttt{ProbNN}_K)$ is introduced to reduce the $B^0_s \to \phi\gamma$ and $B^0 \to K^{*0}\gamma$, while $(1 - \texttt{ProbNN}_p)$ mainly impacts Λ_b decays (see section 7.3.2).
The cut on PID_μ is chosen to keep $B^0_s \to \phi\gamma$ and $B^0 \to K^{*0}\gamma$ contributions at least one order of magnitude smaller than the signals. This requires a cut $\text{PID}_\mu > 0.4$, weaker than the cut used in $B^0_s \to \mu^+\mu^-$ analysis (where 0.8 is used for Run II). This is explained by the fact that hadronic backgrounds are relatively larger for $B^0_s \to \mu^+\mu^-$. Indeed [40]

$$\frac{\mathcal{B}(B^0_s \to KK)}{\mathcal{B}(B^0_s \to \mu^+\mu^-)} \approx 9000\,, \text{ while } \frac{\mathcal{B}(B^0_s \to \phi\gamma) \times \mathcal{B}(\phi \to KK)}{\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma)} \approx 1800\,. \tag{5.7}$$

This cut is applied to the signal and the normalization channels. For the control

channel $B_s^0 \to \phi\gamma$, the variable PID_μ manifestly presents poor interest. The cut $\texttt{ProbNN}_K > 0.2$ was shown to present good performances [368] and is instead used.

The efficiencies of the PID cut on signal and peaking backgrounds are estimated using the *PIDCalib* package [369] (see next paragraph for the signal and section 7.3.2 for backgrounds).

Cut efficiency

The PID cut efficiency is computed on simulation by convoluting, for each event, the single particle efficiencies computed on data samples. As they rely on the RICHs, the `ProbNN` variables are expected to depend on the momentum and the pseudo-rapidity of the particle. A dependency on the event occupancy is also expected due to the tracking. We therefore estimate the PID efficiency in bins of energy **P**, pseudo-rapidity $\boldsymbol{\eta}$ and number of tracks **nTracks** from the data samples provided in *PIDCalib*[3]. The number of tracks being approximately reproduced by the simulation, the efficiencies are estimated on weighted MC (see section 8.3 for the weighting procedure), and a systematic uncertainty is associated to the weighting procedure. Muon data samples are selected from $J/\psi \to \mu\mu$ decays, imposing that the muons are in the detector acceptance. To avoid any bias from muon selections, the events are triggered independently of signal,

$$\texttt{L0Global\_TIS \&\& Hlt1Phys\_TIS \&\& Hlt2Phys\_TIS}\,, \tag{5.8}$$

and a *tag and probe* method is used.

The choice of the binning is studied using the `BinningOptimizer` algorithm[4] which aims at capturing changes in the PID efficiency. The algorithm populates the space of PID variables by bins with almost constant PID efficiency and enough events to be statistically relevant. For the energy, the binning obtained with the algorithm[5] shows similar performance than the regular binning already used in $B_s^0 \to \mu^+\mu^-$ analysis and we kept the latter for simplicity.
We therefore used the following binning:

- **P** (GeV) $\quad \leq 5 : 10 : 15 : 20 : 25 : 30 : 35 : 40 : 50 : \geq 60$
- $\boldsymbol{\eta}$ $\quad 1.5 : 2.58 : 3.05 : 3.57 : 5.5$
- **nTracks** $\quad \leq 24 : 31 : 37 : 43 : 51 : 61 : 70 : 76 : 84 : \geq 100$

The efficiencies obtained in energy bins for different pseudo-rapidity region are sketched in Figure 5.3.

[3] The variables used in *PIDCalib* are `Brunel_P`, `Brunel_ETA` and `nTracks_Brunel`.
[4] The code of this algorithm can be found here.
[5] **P** (MeV) $\quad \leq 3575 : 6900 : 10225 : 13550 : 16875 : 20200 : 33500 : 46800 : \geq 83375$

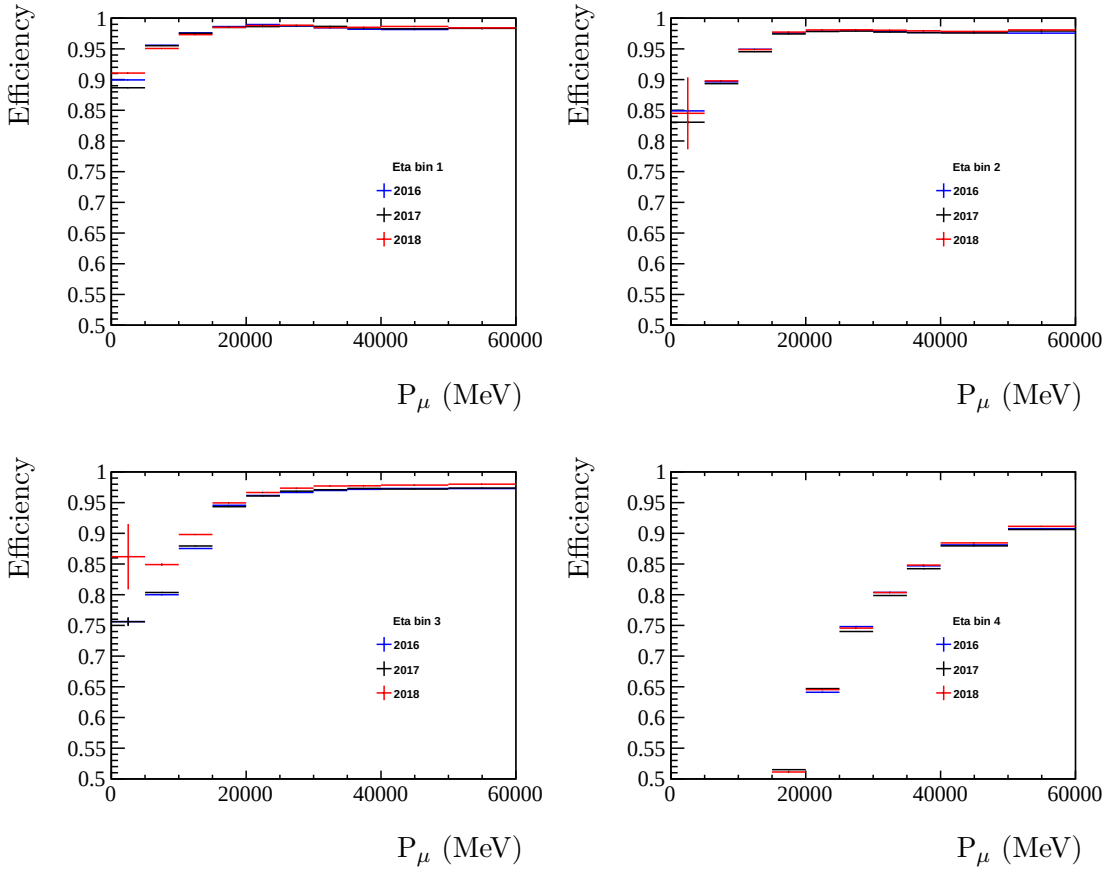

Figure 5.3: Efficiency of the $\text{PID}_\mu \times$ `IsMuon` selection for muons of different energy (P), in different $\boldsymbol{\eta}$ region. The region are defined by the binning $\{1.5, 2.58, 3.05, 3.57, 5.5\}$. The bins are averaged over the number of tracks and the magnet polarity.

Statistical uncertainties

A first statistical uncertainty directly comes from the finite number of events in the simulated signal sample and is considered as normally distributed. A second statistical uncertainty is due to the limited statistics of the calibration samples. To estimate it, a resampling method is used. Random efficiency histograms are generated by drawing random values following the mean and uncertainty specified in each bin of the calibration histograms provided by *PIDCalib*. The global PID cut efficiency eq. 5.6 is computed per simulated event for each of these random histograms. The variance obtained over the different histograms is due to the calibration sample statistics and can be averaged over the simulated events to get the global statistical uncertainty of the calibration procedure.

The results are presented in table 5.10.

Decay	Year	$\varepsilon^{\text{PID\|Strip}}$(PIDCalib)	$\varepsilon^{\text{PID\|Strip}}$(MC)
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$(85.84 \pm 0.15_{\text{stat}} \pm 0.15_{\text{syst}})\%$	$(88.84 \pm 0.14_{\text{stat}})\%$
	2017	$(85.72 \pm 0.12_{\text{stat}} \pm 0.15_{\text{syst}})\%$	$(89.37 \pm 0.10_{\text{stat}})\%$
	2018	$(88.83 \pm 0.23_{\text{stat}} \pm 0.15_{\text{syst}})\%$	$(89.42 \pm 0.09_{\text{stat}})\%$
$B_s^0 \to J/\psi\,\eta$	2016	$(90.37 \pm 0.41_{\text{stat}} \pm 0.15_{\text{syst}})\%$	$(96.08 \pm 0.41_{\text{stat}})\%$
	2017	$(90.30 \pm 0.41_{\text{stat}} \pm 0.15_{\text{syst}})\%$	–
	2018	$(91.73 \pm 0.41_{\text{stat}} \pm 0.15_{\text{syst}})\%$	–

Table 5.10: Efficiency of the PID_μ × `IsMuon` requirement as computed from weighted MC kinematics and single particle efficiencies estimated on data (PIDCalib) and using MC simulated `ProbNN` variables (MC). The quoted error contains the statistical uncertainty from signal and calibration samples and the systematic uncertainty from binning, background subtraction and weighting procedures. For the normalization channel, only 2016 MC is used. Although the global statistics of the 2018 sample is larger, some bins are less populated than for 2017 and 2016 yielding a larger statistical error.

Mis-identification rates

The $\pi, K, p \to \mu$ mis-identification rates are also estimated using the *PIDCalib* package. Pions and kaons are obtained from $D^0 \to K\pi$ calibration lines, while protons come from $\Lambda \to p\pi$ decays. The same procedure as for the muons is applied and the same binning is used. The mis-identification rates for pions, kaons and protons are sketched in Fig. 5.4, 5.5 and 5.6. In these figures, some unphysical negative efficiencies were obtained due to the *sPlot* method [370] used for background subtraction. Negative efficiencies are compatible with zero within 1 sigma so we kept them for the background mis-identification estimations. The mis-identification efficiencies are then convoluted in simulated samples to get the expected yields of each background. The results are presented in section 7.

5.4.2 Photon identification

As already discussed, a good photon identification is crucial to reduce the large combinatorial background. This identification is performed in two steps, the separation between photons and charged tracks and between photons and neutrals. Specific variables are then used to treat photons coming from π^0 and η decays.

Separation between photons and charged tracks

The separation between photons and charged hadrons is performed by the photon Confidence Level (`gamma_CL`). This variable is based on a multivariate analysis trained on MC and calibrated on data samples and aims at separating photon

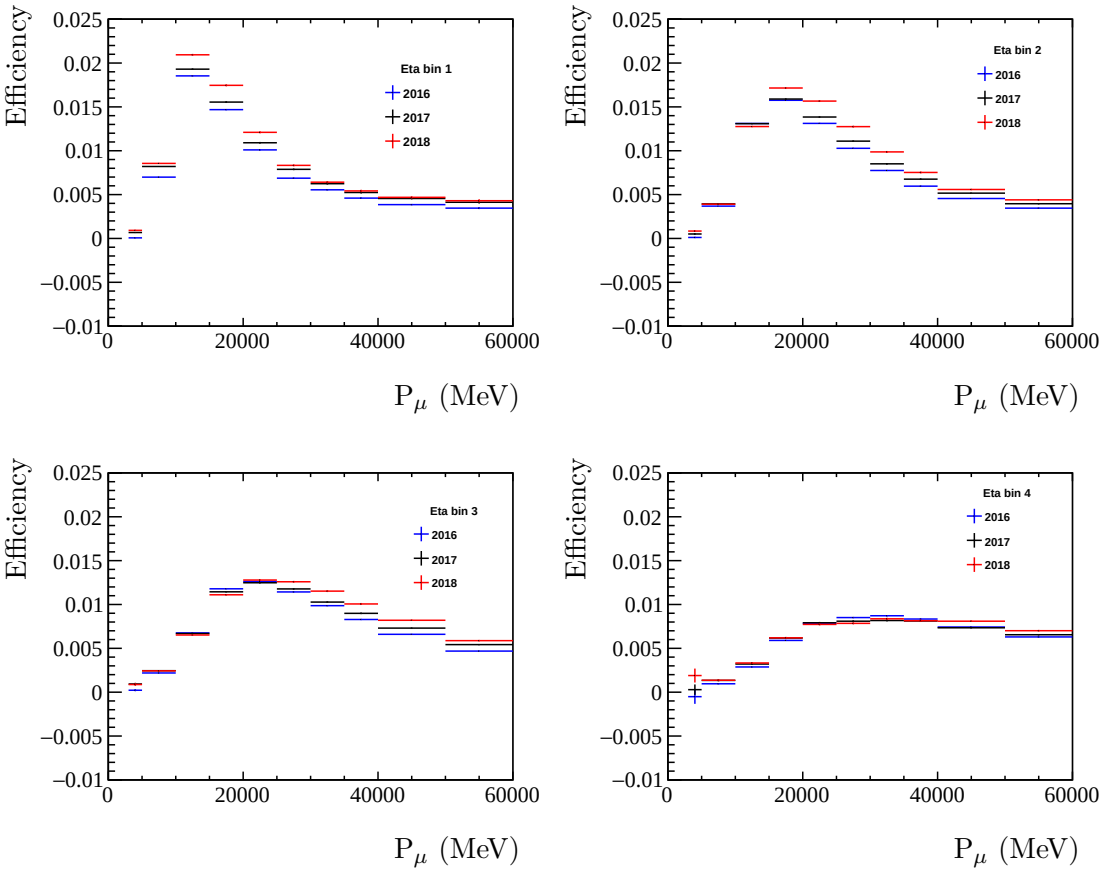

Figure 5.4: Efficiency of the PID$_\mu$ × `IsMuon` selection for pions of different energy (P), in different η regions. This corresponds to the $\pi \to \mu$ mis-identification rate. The region are defined by the binning $\{1.5, 2.58, 3.05, 3.57, 5.5\}$. The bins are averaged over the number of tracks and the magnet polarity.

clusters from non-electromagnetic charged clusters. The cut

$$\text{CL} \equiv \texttt{gamma\_CL} > 0.2 \tag{5.9}$$

was shown to be optimal for the $B_s^0 \to \phi\gamma$ events selection [368], so we apply this cut at the stripping level for all samples. The efficiency of this cut is accounted for in the stripping efficiency estimation. On the other hand, as this variable is not well reproduced by the simulation, the efficiency is *a priori* biased. To account for this effect, we measured the cut efficiency directly on data, as we did for charged particles. This procedure is developed in the next paragraph.

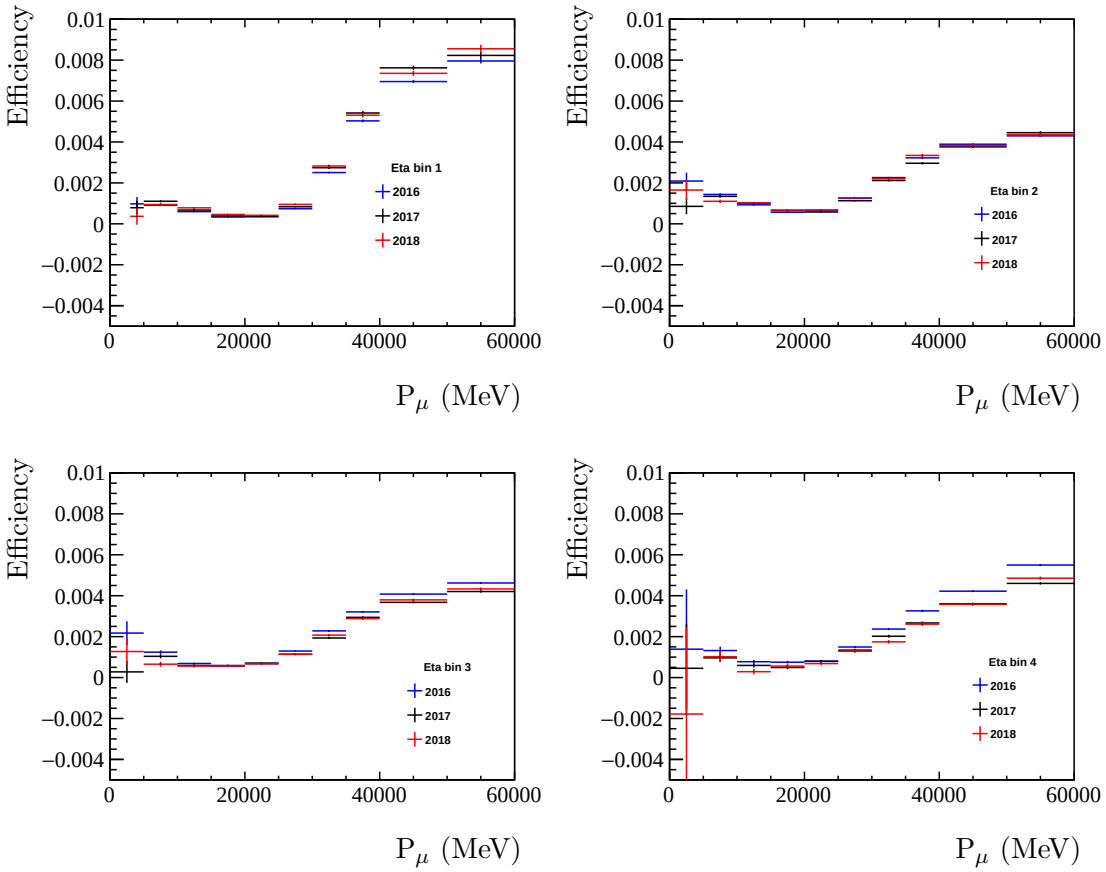

Figure 5.5: Efficiency of the PID$_\mu \times$ `IsMuon` selection for kaons of different energy (P), in different η regions. This corresponds to the $K \to \mu$ mis-identification rate. The region are defined by the binning $\{1.5, 2.58, 3.05, 3.57, 5.5\}$. The bins are averaged over the number of tracks and the magnet polarity. Negative efficiencies are due to *sPlot* weights but are compatible with zero.

Separation between photons and neutrals

Most of the photons detected in the calorimeter come from π^0 decays (around 88% for B decays [354]). These decays can be resolved (two separated reconstructed photons), merged (two photons reconstructed as a single cluster due to the detector granularity) or a photon can be lost (*e.g.* because it is out of LHCb acceptance). The latter category is sub-dominant, less than 1% of such mis-reconstructions were measured on the $B^0 \to \mu^+\mu^-\pi^0$ MC sample. However, 70% of the merged π^0 are also reconstructed as a single photon, which provides a large background for radiative analysis. The separation between photons and merged π^0 is performed by the variable `IsPhoton` [354]. This variable is the output of a multivariate classifier that is trained to separate photon and merged π^0, based on the shape of the elec-

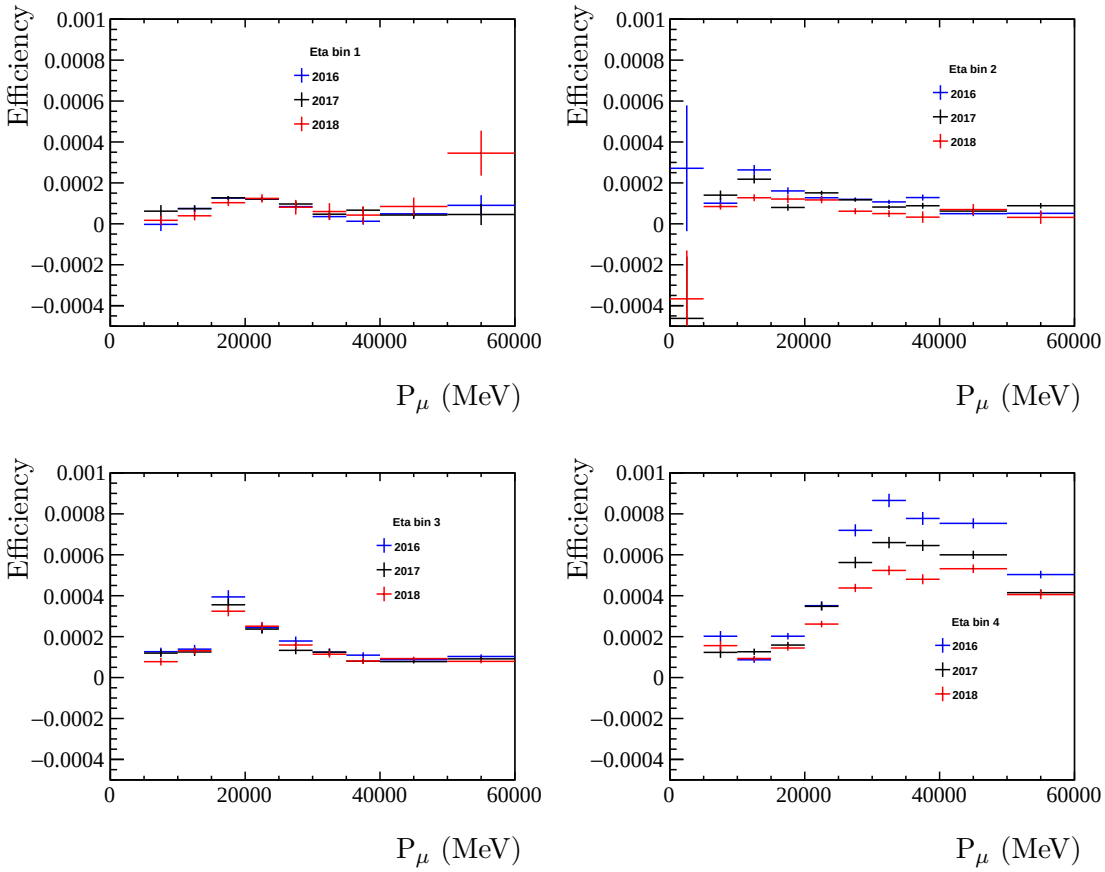

Figure 5.6: Efficiency of the PID$_\mu$ × `IsMuon` selection for protons of different energy (P), in different $\boldsymbol{\eta}$ regions. This corresponds to the $p \to \mu$ mis-identification rate. The region are defined by the binning $\{1.5, 2.58, 3.05, 3.57, 5.5\}$. The bins are averaged over the number of tracks and the magnet polarity. Negative efficiencies are due to *sPlot* weights but are compatible with zero.

tromagnetic cluster in the ECAL, the PS and the SPD. The tool was trained on simulation, using $B^0 \to K^{*0}\gamma$ as signal and several B decays as background. It is trained on photons with transverse momentum larger than 2000 MeV, which are the most concerned by this mis-identification. Indeed, π^0 with transverse momentum smaller than 2000 MeV are mostly resolved because the distance between the clusters increases when the transverse energy decreases [352].

Following the study performed for $B^0_s \to \phi\gamma$ decays [368], we apply the following cut on photons:

$$\text{IsPhoton} \equiv \texttt{IsPhoton} > 0.6 \;||\; \texttt{gamma\_PT} < 2000 \text{ MeV}\,. \tag{5.10}$$

The `gamma_CL` and `IsPhoton` variables are not perfectly reproduced in simulated samples. Although the weighting of the MC improves the matching with data, this

consistency could only be checked for events passing the CL and IsPhoton cuts. Hence, the efficiency of the CL and the IsPhoton cuts need to be measured directly on data. As CL is already applied at the stripping level, its efficiency, measured on MC, is already accounted for in the stripping efficiency. The ratio of CL and IsPhoton efficiencies on MC and data is therefore considered as a correction factor.

To compute this correction factor, we apply a method similar to the one applied to charged tracks. We first compute the efficiency of the cuts in several photon momentum **P**, photon pseudo-rapidity $\boldsymbol{\eta}$ and **nTracks** bins. The binning is chosen by taking the quantiles of the simulated signal distribution:

- **P** (MeV) $\leq 14850 : 24030 : 35580 : 51950 : \geq 82880$
- $\boldsymbol{\eta}$ $\leq 2.53 : 3.03 : \geq 3.53$
- **nTracks** $\leq 95 : 130 : 160 : 188 : \geq 233$

The efficiency is extracted from a control sample via the *sPlot* method . As the photons in $B_s^0 \to \phi\gamma$ don't cover the full transverse momentum spectrum of the signal (due to the cut $\mathrm{PT}_\gamma > 2500$ MeV), we rather used $\eta \to \mu\mu\gamma$ as control sample. As a cross check, we also tested the procedure with $B_s^0 \to \phi\gamma$ samples, which have a lower statistic, and got a compatible efficiency. *E.g.* for 2017, $B_s^0 \to \phi\gamma$ sample yield an efficiency of $(91.6 \pm 6.2_{\mathrm{stat}})\%$, to be compare with the $\eta \to \mu\mu\gamma$ result, $(88.92 \pm 0.60_{\mathrm{stat}})\%$.

2017 and 2018 dataset are selected by dedicated Turcal lines[6]. In 2016, these lines were not available, and we had to reconstruct the events from Turbo lines. The differences between the two approaches are:

- Using the raw-banks, the ECAL can be calibrated offline in Turcal data but not in Turbo data. The width of the η peak is therefore ~ 2 MeV smaller in 2017 and 2018 data. As ECAL correction are small (at the percent level) we disregard the impact on the PID.
- The PID cut on the muons is weaker in Turcal data ($\texttt{ProbNN}_\mu > 0.80$ *vs.* 0.95 in Turbo). As neutral PID and ECAL variables are a priori not correlated *in each* **P**-$\boldsymbol{\eta}$-**nTracks** bins, we consider that the impact of this change is also negligible.
- The use of the **nTracks** variable is however not safe for Turbo data, where the **nTracks** variable is computed differently online and offline [7]. For 2016 we therefore used bins of **nLongTracks**.

[6] Turbo and Turcal lines are specific Hlt lines whose output can be used directly for analysis or calibration purposes. The events passing the Turcal lines are saved with their raw data, allowing dedicated study of the detector response. These so-called raw-banks are not saved in events passing Turbo lines to save place.

[7] We refer to the *PIDCalib* webpage for a detailed discussion on these aspects.

Efficiency tables are also needed for the simulation because the CL requirement being applied at the stripping level and the MC being stripped, signal simulated samples only contain events passing this cut. For the simulation, tables are produced from $B_s^0 \to \phi\gamma$ events. To do so, the samples are re-stripped without the CL requirement. As the difference in transverse momentum coverage did not impact the resulting efficiency in data, we considered that this is the same for MC. To check this assumption, we also generated a table from re-stripped 2016 $B_s^0 \to J/\psi\,\eta$ events. The difference obtained in the efficiency was around 1%, *i.e.* within the statistical error. Furthermore, as the result is expected to be correlated with **nTracks** which shows discrepancies even after the MC weighting is applied, a second weighting is computed in 2017 and 2018 where **nLongTracks** is replaced by **nTracks** (see section 8.3).

The binned efficiency is finally convoluted using the corrected signal kinematics for signal and normalization channels. The statistical uncertainty due to the finite number of events in the calibration and simulation samples are estimated as for charged tracks, *i.e.* by a re-sampling of the calibration tables.

Results

The efficiencies of the overall CL×IsPhoton cut are presented in Tables 5.11 and 5.12.

Decay	Year	$\varepsilon^{\text{NeutralPID\|Strip}}$ [%]		Ratio
		Calib	MC	
$B_s^0 \to \mu^+\mu^-\gamma$	2016	88.06 ± 1.62	89.75 ± 1.13	$0.981 \pm 0.022_{\text{stat}}$
	2017	88.92 ± 0.60	89.93 ± 0.72	$0.989 \pm 0.010_{\text{stat}}$
	2018	88.81 ± 0.39	89.78 ± 0.61	$0.989 \pm 0.008_{\text{stat}}$
$B_s^0 \to J/\psi\,\eta$	2016	81.66 ± 1.79	84.79 ± 0.66	$0.963 \pm 0.022_{\text{stat}}$
	2017	83.81 ± 0.55	–	$0.988 \pm 0.010_{\text{stat}}$
	2018	83.63 ± 0.36	–	$0.986 \pm 0.009_{\text{stat}}$

Table 5.11: Efficiency of the neutral PID requirements CL eq. (5.9) and IsPhoton eq. (5.10) as computed from single particle efficiencies estimated on $\eta \to \mu\mu\gamma$ data (Calib) and $B_s^0 \to \phi\gamma$ MC (MC) and convoluted using signal kinematics. The quoted error is only statistical and comes from the limited signal and calibration samples. The systematic uncertainties coming from the binning, the subtraction method and the weighting procedure are computed on the ratio between signal and normalization and added in Table 9.4, 9.5 and 9.6. For the normalization channel, the kinematics is extracted from 2016 MC for each year.

For completeness, here is the efficiency of the CL cut only efficiency measured

Decay	Year	Bin	$\varepsilon^{\text{NeutralPID}\|\text{Strip}}$ [%] Calib	MC	Ratio
$B_s^0 \to \mu^+\mu^-\gamma$	2016	I	87.93 ± 1.64	89.78 ± 1.12	$0.979 \pm 0.022_{\text{stat}}$
		II	88.49 ± 1.52	89.74 ± 1.17	$0.986 \pm 0.021_{\text{stat}}$
		III	91.33 ± 1.07	88.63 ± 1.44	$1.030 \pm 0.021_{\text{stat}}$
	2017	I	88.77 ± 0.61	89.95 ± 0.71	$0.987 \pm 0.010_{\text{stat}}$
		II	89.43 ± 0.54	89.94 ± 0.82	$0.994 \pm 0.011_{\text{stat}}$
		III	92.66 ± 0.66	89.09 ± 0.92	$1.040 \pm 0.013_{\text{stat}}$
	2018	I	88.68 ± 0.40	89.83 ± 0.61	$0.987 \pm 0.007_{\text{stat}}$
		II	89.29 ± 0.36	89.78 ± 0.88	$0.995 \pm 0.011_{\text{stat}}$
		III	92.53 ± 0.19	88.84 ± 1.02	$1.042 \pm 0.012_{\text{stat}}$

Table 5.12: Binned efficiency of the neutral PID requirements CL eq. (5.9) and IsPhoton eq. (5.10) as computed from single particle efficiencies estimated on $\eta \to \mu\mu\gamma$ data (Calib) and convoluted using signal kinematics and on local $B_s^0 \to \mu^+\mu^-\gamma$ MC (MC). The quoted error is only statistical and comes from the limited signal and calibration samples. The systematic uncertainties coming from the binning, the subtraction method and the weighting procedure are computed on the ratio between signal and normalization.

on 2017 signal samples:

$$\varepsilon^{\text{CL}>0.2|\text{Strip}}(\text{Calib}) = (94.27 \pm 0.48_{\text{stat}})\% \tag{5.11}$$

$$\varepsilon^{\text{CL}>0.2|\text{Strip}}(\text{MC}) = (93.82 \pm 0.69_{\text{stat}})\% \,. \tag{5.12}$$

Light mesons veto

Resolved π^0 also need to be tackled as they represent the main source of background. The usual method to limit their number is to reconstruct photon pairs in the event and veto the π^0 mass region. This method is poorly efficient due to the large uncertainty on the photon energy and direction. In this analysis we improved this veto by implementing new variables.

The main issue with the π^0 veto is the definition of the mass window for the photon pair. Increasing the mass window improves the π^0 rejection but decreases the signal efficiency due to false positive decisions. The default choice is $|m_{\gamma\gamma} - m_{\pi^0}^{\text{PDG}}| < 60$ MeV and the algorithm is just checking whether the candidate's photon is present in the children particles of the objects of the `StdLoosePi02gg` container. To improve the method, we trade this Boolean variable for two variables:

`gamma_npi0`: the number of π^0 in `StdLoosePi02gg` sharing a photon with the candidate,

gamma_dmpi0: the smallest mass difference between the π^0 candidates and the π^0 PDG mass. If no π^0 can be reconstructed, gamma_npi0 = 0 and gamma_dmpi0 is set to 60 MeV (the size of the mass window in the StdLoosePi02gg container).

The same procedure can also be applied for η decays, and gives two similar variables (gamma_neta and gamma_dmeta). The corresponding container is StdLooseEta2gg, whose mass window is set to $|m_{\gamma\gamma} - m_\eta^{\text{PDG}}| < 105$ MeV. These variables are of course highly correlated with the photon's transverse momentum so we added them in the multivariate analysis instead of performing fixed cuts. They are also correlated to the calorimeter occupancy, which means that they may suffer from discrepancies between data and simulation. These discrepancies are studied in Section 8.3 and found to be small.

Note that the choice of setting gamma_dmpi0 and gamma_dmeta at the mass limit (60 and 105 MeV respectively) in the absence of π^0 and η yields peaks in the distribution of these variables (these peaks can be seen in the last bins in Fig. 5.7). This feature however does not impact the efficiency of the multivariate analysis that uses this variable. Indeed, these peaks are present if, and only if, gamma_npi0 and gamma_neta are zero and this correlation can be easily understood by a multivariate analysis. We suggest therefore not to use these variables separately.

To test the efficiency of this improved veto, we trained a multivariate analysis based on these 4 variables on signal MC and data side-bands (to be precisely defined in the next section). The ROC curves[8] of this MVA can be compared to the classical veto as shown on Fig. 5.7. At fixed signal efficiency, the gain in background rejection is as high as 13%. This number should however be taken with care as the new variables show larger correlation with the transverse momentum of the signal's photon. The precise gain in efficiency therefore depends on the kinematic of the studied signal.

5.5 The selection MVA discriminant

To further reduce the background, a first multivariate classifier (MLPS, where S stands for Selection) is trained. A loose cut is applied on the output of this classifier to keep a good selection efficiency for signal events. The choice of this complex tool instead of a simple set of cuts is driven by the complexity of the 3-body geometry of the decay and the presence of the soft photon.

[8] ROC stands for Receiver Operating Characteristic. The ROC curve of a classifier is defined as the signal efficiency *vs.* background rejection curve. The integral of this curve is a good test of the efficiency of a multivariate analysis.

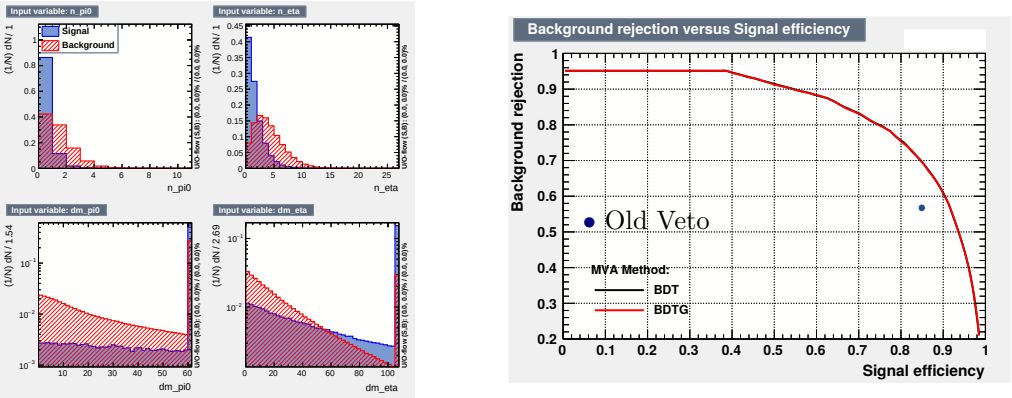

Figure 5.7: Comparison of the distribution of the new veto variables for signal MC and data side-bands (left) and ROC curves of multivariate analysis trained between this two samples (right). Boosted (BDT) and gradient-boosted (BDTG) decision trees behave equally well. The blue dot on the ROC plot corresponds to the classical π^0 veto width for a mass window of 60 MeV. The plateau is due to events with no reconstructed π^0 nor η (here, 38% of events have `n_pi0` = `n_eta` = 0, so the plateau extends from 0 to 0.38. These 38% are also visible on the last bins of the bottom plots on the left panel). These events cannot be tackled with these variables.

5.5.1 Definition

The MLPS mainly aims at reducing the combinatorial background and is based on geometrical variables and on the photon variables defined in section 5.4. The MLPS is trained on signal MC and data side-bands using TMVA [371]. The side-bands are defined as the region below ([3867 MeV, 5167 MeV]) and above ([5567 MeV, 6867 MeV]) the B_s^0 mass[9].

The choice of the discriminating variables has been made with the following algorithm:

1. Select all variables that show a sizeable difference between signal and background (the difference can be measured *via* a Kolmogorov-Smirnov test).
2. Train a multivariate analysis with these N variables
3. Remove one of these variables and train a multivariate analysis with $N-1$ remaining ones
4. Keep the $N-1$ variables that give the best ROC integral, and repeat from step 3 until the ROC integral is maximal.

The ROC is increasing in step 4 because of the overtraining due to the limited MC sample. This algorithm therefore aims at balancing overtraining *vs.* performances.

[9] These ranges correspond to $[-1500\text{ MeV}, -200\text{ MeV}]$ and $[200\text{ MeV}, 1500\text{ MeV}]$ around the B_s^0 PDG mass.

For example, using the photon transverse momentum, the photon energy and the photon pseudo-rapidity lead to redundancy that decreases the multivariate analysis efficiency. Sometimes, the improvement is not significant enough to remove a specific variable and the one that shows the best MC-data agreement is kept. This was *e.g.* the case for the B transverse momentum that showed no significant improvement with respect to the photon transverse momentum. Using the plot in Appendix B.5 (Fig. B.8 and B.9), we kept the photon transverse momentum. We note finally that combining variables that show small individual separation power may also lead to interesting new variables. Although the above procedure should in principle find these combinations, the set of variables obtained is certainly not optimal. Yet, we consider that given the discrepancies that exist between data and MC, a more refined study will not bring a sizeable improvement of the multivariate analysis efficiency.

Applying this procedure yields a set of 12 input variables:

- The B_s^0 candidate's impact parameter (`B_ip`) and impact parameter significance (`B_ips`) [10];
- The B_s^0 candidate's vertex chi2 (`B_vchi2`);
- The B_s^0 candidate's direction angle (`B_dira`)[11];
- The minimal distance between the muon tracks (distance of closest approach, `mu_doca`);
- The muon smallest impact parameter (`mu_minips`);
- The photon energy (`gamma_PE`) and transverse momentum (`gamma_PT`);
- The 4 photon isolation variables defined in section 5.4.2.

The distributions of these variables are plotted for MC signal and data side-bands in App. B.2.

The multivariate analysis is trained separately for each year by merging the magnet polarities (no sizeable difference between Up and Down samples were observed). To ensure that a cut on the multivariate analysis output has the same efficiency on signal for each year, a transformation is applied on the output variable, such as it is flat on $[0, 1]$ for signal MC. This is done by trading the multivariate analysis output for its p-value:

$$\mathrm{MLPS}_{\mathrm{flat}}(x) = \int_{-\infty}^{x} f(t)dt$$

[10] Significance is defined as the square root of the χ^2, `B_ips = sqrt(B_ipchi2)`.

[11] This angle corresponds to the angle between the reconstructed B_s^0 momentum and the vector defined by the primary and the secondary vertices.

$$\approx \sum_{i=0}^{i_x} N_i \,, \tag{5.13}$$

where x is the output of the MLPS, f is the distribution function of x on signal MC. For the binned version of the equation, i_x is the number of the bin that contains x and N_i is the number of events in the i^{th} bin of the MLPS binned output. In the following we will call MLPS the variable $\text{MLPS}_{\text{flat}}$. A cut at *e.g.* MLPS > 0.25 will thus have an efficiency of 75% on the signal for all the years.

5.5.2 Note on Decay Tree Fitter (DTF)

The origin of the photon is unknown and set by default to $(0, 0, 0)$, the origin of the reconstruction frame. This alters the estimation of geometrical variables such as the dira, the impact parameters, etc. The decay can however be fitted by imposing a common vertex for all particles and, when appropriate, by fixing the masses of the decaying children particles. This improves the quality of part of the variables but shows a limited improvement on the candidates mass resolution in this analysis (3% on 2018 $B_s^0 \to \mu^+\mu^-\gamma$ MC). In fact, this constraint also alters the backgrounds and make them harder to distinguish from signal. Given the huge number of low energetic photons at this step of the analysis, the use of DTF [372] decreases the performance of the MLPS. This can be seen in Fig. 5.8 where the ROC curve (MLP DTF) is below the other curves in a large range of efficiencies.

DTF is therefore *not* used for the training of the multivariate analysis, but it is used for the final mass fits.

5.5.3 Performance and choice of the cut

The ROC curve of the MLPS is shown on Fig. 5.8, where several classifiers have been tested. In this case, the Multi-Layer Perceptron (MLP) showed slightly better performances than other classifiers[12]. Fig. 5.8 also shows the correlation between the input variables and the absence of overtraining for the MLP. As expected the new veto variables are highly correlated among them and with the other photon variables. A non-negligible correlation is also found with the B_s^0 vertex χ^2 due to the event occupancy. This effect will be discussed in the section on MC weighting 8.3.

The training was performed on upper data side-bands, lower data side-bands and both. To compare the efficiency of these different setups, one cannot rely on the ROC curves given by TMVA as they are only computed on the testing samples (*i.e.* on upper, lower and both side-bands respectively). We therefore trained the three MLPS on data, applied them to the full sample and compared the background rejection. The result is shown on Fig. 5.8 (bottom-right panel). As the MLPS output is flat for signal, the curve that shows the best background

[12] The TMVA configuration of the MLP is `NeuronType=tanh:VarTransform=N:NCycles=500:HiddenLayers=N+5:TestRate=5`

rejection also shows the best performances. We conclude that training on the upper, lower or on both side-bands shows similar performances for the MLPS, and choose to train on both side-bands because the overtraining was smaller in this case.

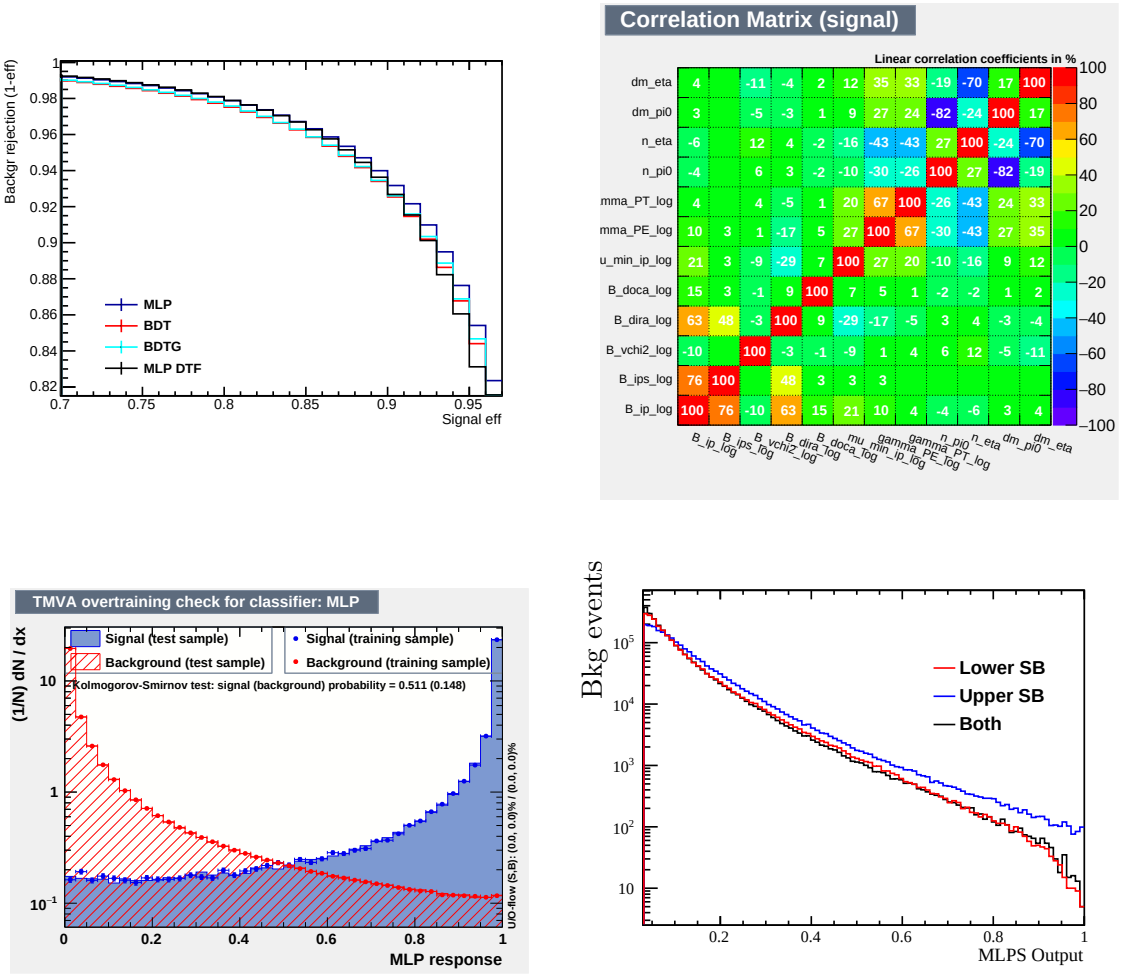

Figure 5.8: Results of the MLPS training for the 2016 sample. Similar results are obtained for 2017 and 2018. The best ROC curve (upper-left plot) was obtained with a Multi-Layer Perceptron (MLP). The upper-right plot on shows the correlations between the input variables for signal events. As shown in the lower-left plot, no overtraining is found in the MLPS output. The last plot shows the distribution of the multivariate analysis output for 2016 data side-bands for a MLPS trained on upper side-bands (blue curve), lower side-bands (red curve) and both side-bands (black curve) but applied to the whole sample. Training on both side-bands shows the best background rejection.

A loose cut is applied on the output of the MLPS. This cut is chosen in order to optimize the full selection process, namely the efficiency of the MLPS *and* the

second multivariate analysis. To do so, we trained the global MLP (defined in section 6.3) imposing different MLPS cuts.

The MLP performance is evaluated using the Punzi figure of merit (FoM) [373]:

$$\mathrm{FoM} = \frac{\varepsilon_{\mathrm{S}}}{\sigma/2 + \sqrt{N_B}}\,, \tag{5.14}$$

where ε_{S} is the signal efficiency, $\sigma = 5$ is the target statistical significance and N_B is the number of events in the signal region. For each MLPS cut, the MLP defined in section 6.3 is trained and the FoM is computed along the MLP ROC curve. A difficulty arises from the fact that N_B being the number of background events in the signal region, one needs to know the mass distribution of background events. Now, computing and applying the MLP for each MLPS is very time and CPU consuming and the analysis is performed blinded so N_B is anyway not precisely known. To avoid this problem, the number of background events is assumed to decrease linearly with the mass in the entire mass range. This allows to estimate N_B directly from the output of the MLP training without fitting the background mass p.d.f because by symmetry, N_B is independent of the slope of the linear fit. This assumption is done for simplicity and will be improved in section 6.3 by using an exponentially decreasing fit for the estimation of the optimal MLP cut. The maximal FoMs are plotted in Fig. 5.9. As can be seen, the dependence on the MLPS cut is small, which can be understood as a compensation between the MLP and MLPS efficiencies. Decreasing the MLPS efficiency by lowering the cut makes the MLP more efficient. Indeed, if the MLPS cut is smaller, there are more background events in the selected sample. The MLP removes these background events easily, and looks therefore more efficient.

The cut MLPS > 0.25 is finally chosen for all samples.

For the normalization channel, $B_s^0 \to J/\psi\,(\to \mu_1\mu_2)\,\eta(\to \gamma_1\gamma_2)$, we impose that the cut is passed by either $\{\mu_1, \mu_2, \gamma_1\}$ or $\{\mu_1, \mu_2, \gamma_2\}$. This increases the efficiency of the cut and does not affect the cancellation of uncertainties in the ratio of efficiencies between signal and normalization because the most energetic photon will play the same role as the photon in $B_s^0 \to \mu^+\mu^-\gamma$. Imposing that the cut is satisfied by both sets, apart from drastically lowering the efficiency, would imply constraints on the kinematics that do not reflect the signal constraints. For the control channel, $B_s^0 \to \phi(\to K_1K_2)\,\gamma)$, the cut is applied on the set $\{K_1, K_2, \gamma\}$.

The distribution of the output of the MLPS for the three channels are presented in Fig. 5.10. Note first that the MLPS selects $B_s^0 \to \phi\gamma$ events more efficiently than signal events (see the peak at MLPS ~ 1 in the upper right plot). This is mainly due to the cut on the photon transverse momentum $\mathrm{P_T} > 2500$ MeV performed at the stripping level. Indeed, as shown in Fig. 5.2, the photon background coming from π^0 is 80% smaller for a photon transverse momentum cut of 2500 MeV. The higher efficiency of the MLPS for the control channel compensates the poor stripping efficiency in Table. 5.8 and ensures a large enough statistics for the control studies. On the other hand, for the normalization mode, high MLPS values are

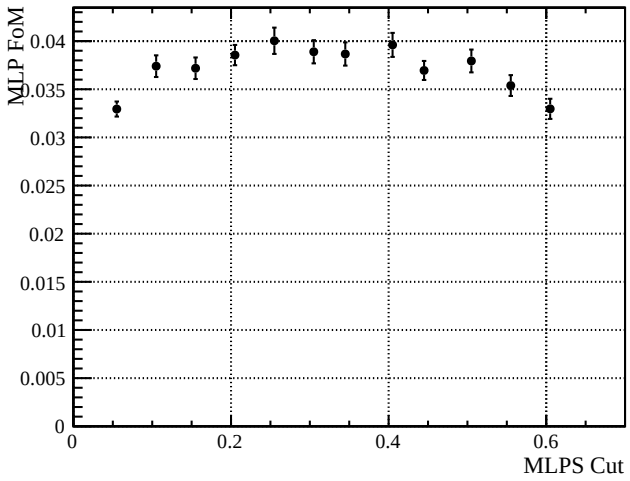

Figure 5.9: Figure of merit of the full selection process for different cut on the MLPS output. A cut of 0.25 on the MLPS output (*i.e.* an efficiency of 75% on signal) showed slightly better performances for the full classification procedure. Note that the overall normalization is arbitrary, as for this plot, the signal efficiency ε_S in eq. (5.14) is restricted to the MLPS signal efficiency.

disfavoured by the less energetic photons as depicted in Fig. 5.10 (bottom left). However, as the MLPS output is the maximum of the output for the two photons, and thanks to the cut on the η transverse momentum, the efficiency remains in the ballpark of the signal one.

The cut efficiency for signal, normalization and control channels are summarized in Tables 5.13 and 5.14.

Decay	Year	$\varepsilon^{\text{MLPS\|Strip\&Reco}}$ [%]
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$75.00 \pm 0.13_{\text{stat}}$
	2017	$75.00 \pm 0.10_{\text{stat}}$
	2018	$75.00 \pm 0.09_{\text{stat}}$
$B_s^0 \to J/\psi\,\eta$	2016	$70.96 \pm 0.36_{\text{stat}}$
$B_s^0 \to \phi\gamma$	2016	$96.31 \pm 0.22_{\text{stat}}$
	2017	$97.36 \pm 0.21_{\text{stat}}$
	2018	$96.43 \pm 0.22_{\text{stat}}$

Table 5.13: Efficiency of the MLPS cut on signal, normalization and control channels. The low efficiency for the normalization channel is due to the fact that the energy of the η is shared between the two photons.

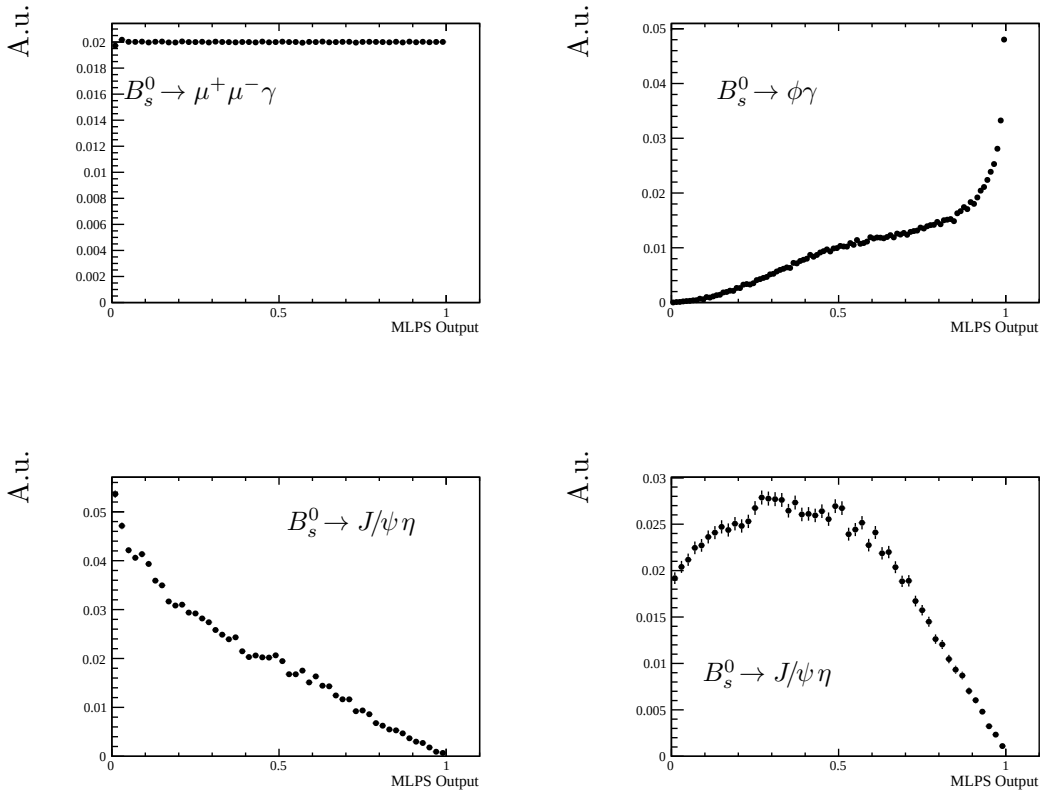

Figure 5.10: Distribution of the output of the MLPS for signal (upper left), $B_s^0 \to \phi\gamma$ (upper right), $B_s^0 \to J/\psi\,\eta$ (bottom left and right) simulations. The signal is flat by construction, $B_s^0 \to \phi\gamma$ are favoured by the MLPS (peak at 1) thanks to their energetic photon. For $B_s^0 \to J/\psi\,\eta$, the first plots shows the MLPS output as computed on the muons and one photon, while the second plots shows the maximum of the MLPS computed on each photon (see text for details).

Decay	Year	$\varepsilon^{\text{MLPS\|Strip\&Reco}}$ [%]		
		Bin I	Bin II	Bin III
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$76.29 \pm 0.16_{\text{stat}}$	$70.04 \pm 0.79_{\text{stat}}$	$50.00 \pm 0.61_{\text{stat}}$
	2017	$76.41 \pm 0.11_{\text{stat}}$	$69.13 \pm 0.66_{\text{stat}}$	$42.14 \pm 0.43_{\text{stat}}$
	2018	$76.43 \pm 0.11_{\text{stat}}$	$70.05 \pm 0.62_{\text{stat}}$	$42.60 \pm 0.39_{\text{stat}}$

Table 5.14: Efficiency of the MLPS cut MLPS>0.25 on signal events for each q^2 bin.

6 Signal classification

Contents

The selection based on the MLPS reduces the background using geometrical variables. Due to the huge combinatorial background, the sensitivity at the end of this selection process is still very small and other variables are investigated to construct a multivariate analysis classifier. The most efficient variables are based on the property that particles coming from a $B_s^0 \to \mu^+\mu^-\gamma$ decay are relatively isolated with respect to the other tracks of the event. Muons or photons coming from combinatorial background are usually surrounded by other particles produced from the same vertex (*e.g.* to conserve leptonic number).

6.1 Photon isolation

Specific variables have been developed and are presented in section 5.4.2 to reduce the photon background due to π^0 and η decays. Photons can also be produced

by charged tracks, so we also measure the isolation of the photon with respect to charged particles. To do so, we define an $\eta - \phi$ cone around the photon direction and add the transverse momentum of all the tracks in this cone (ignoring the two muon candidates). We then define the following variable

$$\mathrm{I}_\gamma = \frac{\gamma_{\mathrm{PT}}}{\gamma_{\mathrm{PT}} + \sum\limits_{t \in \mathrm{cone}} t_{\mathrm{PT}}} \in [0, 1] \,, \tag{6.1}$$

where 0 means a poorly isolated photon, while 1 is obtained only if the photon is alone in its cone[1]. Different cone angles have been tested and an angle of $\sqrt{\Delta\phi^2 + \Delta\eta^2} = 1$ rad was found to give the best separation between signal and data side-bands. Fig. 6.1 shows the intersection of such cones with the ECAL plane for the 3 different regions. Note that the cone is defined by the reconstructed photon momentum. This means that the cone starts from the origin of the detector frame and not from the actual photon production vertex. Given the size of the cone angle, the effect on the photon isolation is expected to be small.

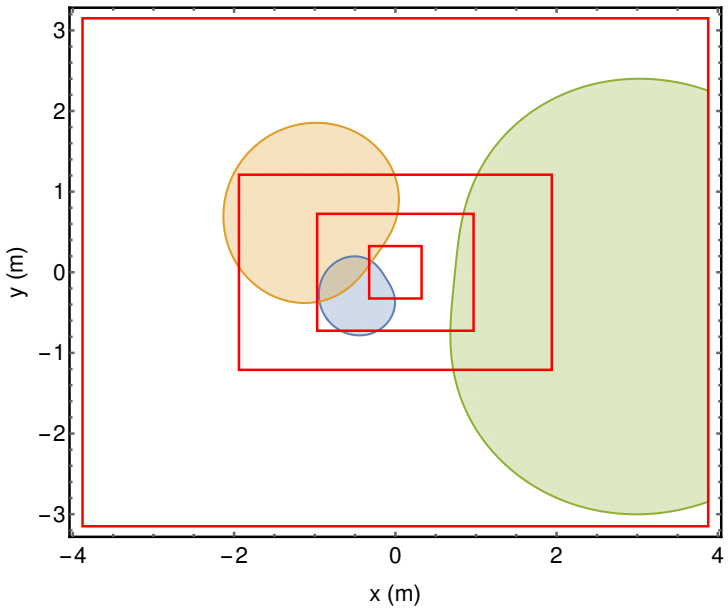

Figure 6.1: Projection of $\eta - \phi$ cones of angle 1 rad on the ECAL. Three cones are shown in the three different regions of the ECAL (sketched by the red rectangles, see Fig. 4.3).

6.2 Tracks isolations

The leptonic combinatorial background can be efficiently reduced by cutting on the muon isolation. The naive estimation of the muons isolation consists in counting

[1] This isolation was implemented in `TupleToolConeIsolation`, but we implemented it at the stripping level.

the number of tracks inside an $\eta - \phi$ cone around the muon track. In Section 6.2.1 we define an alternative set of variables developed in the $B_s^0 \to \mu^+\mu^-$ analysis and present its performance on data in Section 6.2.2.

6.2.1 Tracks isolation variables in the $B_s^0 \to \mu^+\mu^-$ analysis

We summarize here the isolation variables developed in the $B_s^0 \to \mu^+\mu^-$ analysis. New variables are needed to estimate whether the tracks surrounding a muon indicate that the event comes from background or not. The complete description of these variables can be found in [374] and their efficiency in the context of the $B_s^0 \to \mu^+\mu^-$ analysis is described in [365]. The idea is to train a multivariate analysis on MC and data events to separate tracks that share a common ancestor with the muon (so-called "non-isolating tracks") from the other tracks of the events. The variables used for the multivariate analysis were updated in 2014. The sets are as follows:

Set 1 (variables already present in the 2013 $B_s^0 \to \mu^+\mu^-$ analysis):

- `trk_ips`: minimum of the square root of the impact parameter χ^2 of the track with respect to any primary vertex (PV) of the event
- `trk_pvdis`: signed distance between the (track, muon) vertex and the PV
- `trk_svdis`: signed distance between the (track, muon) vertex and the B_s^0 end-vertex
- `trk_doca`: distance of closest approach between the track and the muon
- `trk_angle`: angle between the track and the muon
- `fc`: geometrical variable defined in [375] that goes to zero when the (track, muon) system originates from the PV

Set 2 (variables added in 2014):

- `trk_dphi`: absolute value of the difference between the azimuthal angles of the track and the muon
- `trk_deta`: absolute value of the difference between the pseudo-rapidities of the track and the muon
- `trk_pt`: transverse momentum of the track

These two sets of variables can be defined for Long tracks and for Velo tracks (cf Fig. 4.2). Other sets of variables have been tested but the best isolation performances were obtained with Set 1 + Set 2 for long tracks and Set 1 only for Velo tracks. The long (respectively Velo) isolation variable is defined as the mean of the two largest output values obtained when applying the multivariate analysis on each long (resp. Velo) track of the event. These variables have been computed directly during the stripping.

6.2.2 Performance of the track isolation variables on data

The isolation variables are expected to depend on the number of tracks in the event and the kinematics of the muons. They should therefore be affected by MC weighting. As these variables cannot be tested on the control channel (that involves kaons instead of muons), we tested them on the normalization channel. To avoid any bias, this comparison is performed without MLP cut, *i.e.* on samples where these variables are not used for background rejection. Fig. 6.2 shows the isolation variables for MC and data events. Considering that the shapes are correctly reproduced in data, the remaining difference (measured with the means of the distributions) is considered as a systematic uncertainties and propagated to the global multivariate analysis.

6.3 Global multivariate analysis

The events are finally classified with a second multivariate analysis trained on weighted signal simulation and data side-bands (as defined in section 5.5). The variables used for this analysis are inspired from other analysis and chosen using the same algorithm as for the first multivariate analysis.

6.3.1 Definition

The 11 selected variables are:

- The cosine of the angle between the μ^+ momentum in the B rest frame and the vector perpendicular to the B momentum and the beam axis (`CosP`). The *angle* distribution should be flat for the signal (due to rotational invariance) while the distribution of the *cosine* is expected to be flat for combinatorial background that mainly comes from the beam.
- The B_s^0 candidate's direction angle (`B_dira`)
- The B_s^0 candidate's impact parameter significance (`B_ips`)
- The B_s^0 candidate's transverse momentum (`B_PT`)
- The photon energy (`gamma_PE`) and transverse momentum (`gamma_PT`)
- The photon isolation variable I_γ defined in 6.1 (`gamma_isolation`);
- The muon smallest impact parameter significance (`mu_minips`);
- The muon isolation variables defined in 6.2;
- The distance between the two muons in the $\eta - \phi$ plane: $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ (`DeltaRmumu`). Note that $\Delta\phi$ is not always equal to |`muplus_phi`−`muminus_phi`| due to the invariance under a 2π rotation.

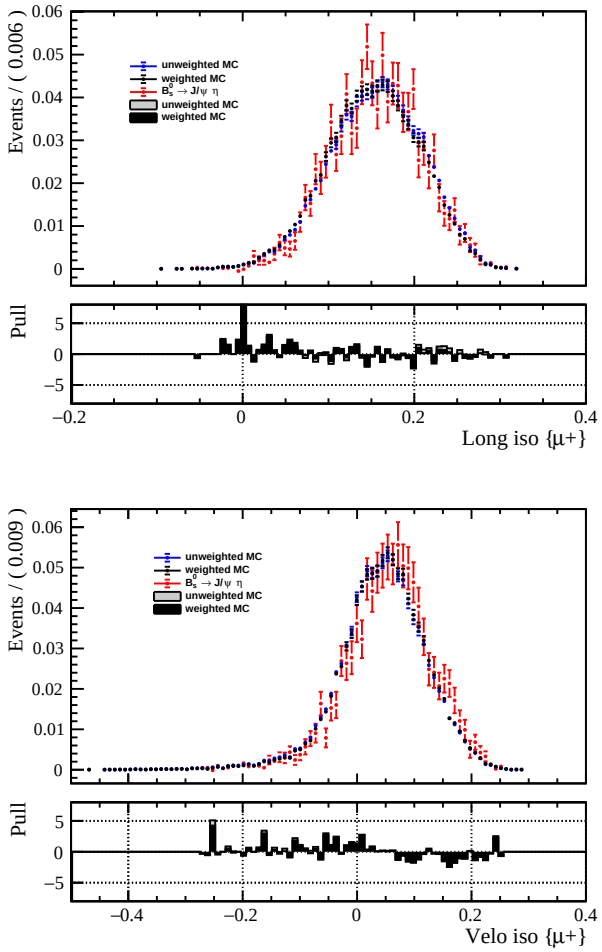

Figure 6.2: Muon isolation variables based on Long (top) and Velo (bottom) tracks for $B_s^0 \to J/\psi\eta$ MC and data events. Note that these variables come from the output of a multi-variable classifier, hence negative values should not be interpreted as a "negative isolation".

The distribution of these variables is shown in App. B.2. Note that the MLP shares variables with the MLPS. This could have been a problem for the propagation of systematic uncertainties if the MLPS was used as a variable of the MLP. Here, the cut MLPS > 0.25 is applied and the systematic uncertainties can be safely propagated first through the MLPS and then through the MLP. As for the MLPS, the transformation Eq. (5.13) is applied on the output of this classifier to obtain a flat output for signal.

6.3.2 Performances and choice of the cut

As for the MLPS, the best classifier was found to be a MLP and no overtraining was found, as pictured on Fig. 6.3. The training is performed separately on each year. The training was again performed on upper and lower data side-bands (defined in 5.5) but training on both showed the best background rejection. This can be seen on the same figure (bottom-right plot), where the output of MLPs trained on upper, lower and both side-bands are plotted. As the MLP output is flat for signal, the curve that shows the best background rejection also shows the best performances. Fig. 6.3 also presents the correlations among input variables.

The best MLP cuts are chosen from the FoM of the signal as defined in Eq. (5.14), where N_B is computed by fitting the side-bands with an exponential distribution (see *e.g.* Fig. **??**) and the signal mass-window is defined as $[5267, 5467]$ MeV. The FoM for signal is depicted in Fig. 6.4 for the full sample. The optimal cut is found to be MLP>0.7. The same method is applied to find the optimal cuts in each q^2 bins. The cut MLP>0.7 is optimal for Bin I, for Bin II the best value is MLP>0.1 and MLP>0.075 is found for Bin III. The FoM are displayed in Appendix, see Fig. B.3.

The efficiency of the MLP cut is measured on weighted simulation. The result is presented in Tables 6.1 and 6.2.

Decay	Year	$\varepsilon^{\text{MLP\|Sel \& Trig \& PID}}$ [%]
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$30.00 \pm 0.10_{\text{stat}}$
	2017	$30.00 \pm 0.07_{\text{stat}}$
	2018	$30.00 \pm 0.06_{\text{stat}}$
$B_s^0 \to J/\psi\,\eta$	2016	$92.52 \pm 0.48_{\text{stat}}$

Table 6.1: Efficiency of the optimal cut MLP>0.7 on signal and the loose cut MLP>0.02 on the normalization channel.

Decay	Year	$\varepsilon^{\text{MLP\|Sel \& Trig \& PID}}$ [%]		
		Bin I	Bin II	Bin III
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$31.40 \pm 0.11_{\text{stat}}$	$38.34 \pm 0.81_{\text{stat}}$	$50.50 \pm 0.96_{\text{stat}}$
	2017	$31.34 \pm 0.08_{\text{stat}}$	$40.39 \pm 0.61_{\text{stat}}$	$50.75 \pm 0.73_{\text{stat}}$
	2018	$31.38 \pm 0.08_{\text{stat}}$	$36.90 \pm 0.53_{\text{stat}}$	$51.52 \pm 0.73_{\text{stat}}$

Table 6.2: Efficiency of the MLP cut on signal events for each q^2 bin. Note that the MLP cut differs in each bin, 0.7 in Bin I, 0.1 in Bin II and 0.075 in Bin III.

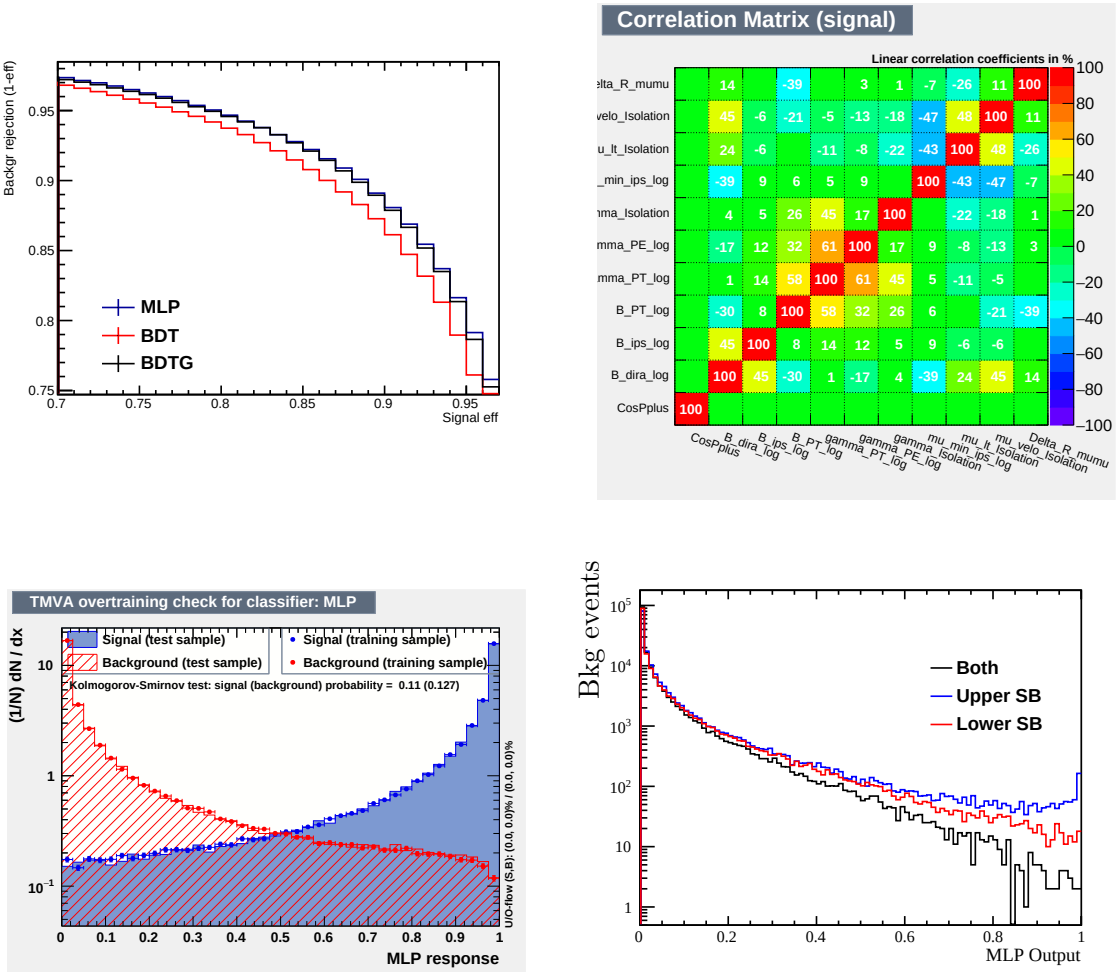

Figure 6.3: Results of the MLP training for the 2016 sample. Similar results are obtained for 2017 and 2018. The best ROC curve (upper-left plot) was obtained with a Multi-Layer Perceptron (MLP). The upper-right plot on shows the correlations between the input variables for signal events. As shown in the lower-left plot, no overtraining is found in the MLP output. The last plot shows the distribution of the multivariate analysis output for 2016 data side-bands for a MLP trained on upper side-bands (blue curve), lower side-bands (red curve) and both side-bands (black curve) but applied to the whole sample. Training on both side-bands shows the best background rejection.

6.3.3 Correlation between $m(\mu\mu\gamma)$ and the MLP output

A strong correlation between the B_s^0 candidate mass and the MLP output could bias the final fit. To check that this is not the case, the correlation coefficient between this two variables is measured on upper and lower side-bands. The result, presented in Fig. 6.5, shows no peaking structures and small correlation coeffi-

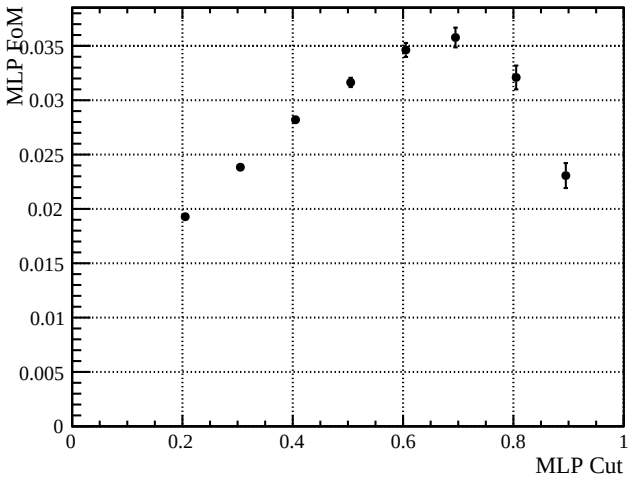

Figure 6.4: Figure of merit for the signal as defined in Eq. (5.14) for different cut on the MLP output. The uncertainties on the MLP FoM are due to the uncertainties in the exponential fit of the background. Note that the overall normalization is arbitrary, as for this plot, the signal efficiency ε_S in Eq. (5.14) is restricted to the MLP signal efficiency. The cut MLP > 0.7 gives the best results.

cients.

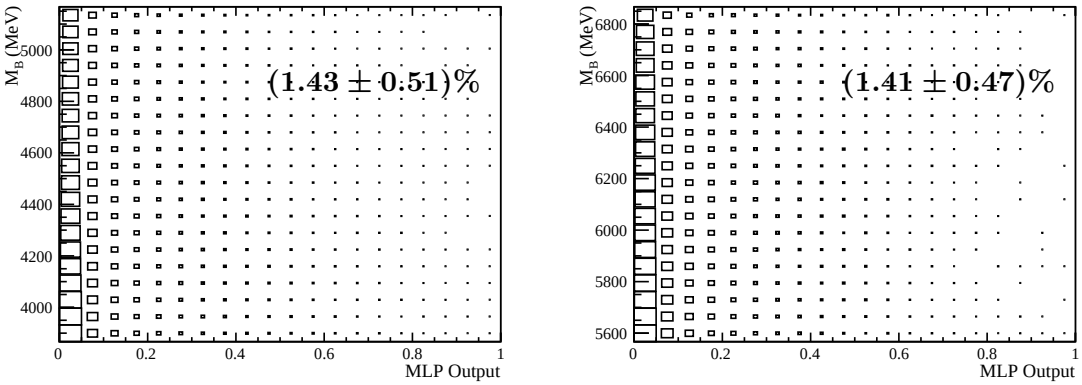

Figure 6.5: Correlation between the B_s^0 candidate mass and the MLP output for lower side band (left) and upper side band (right).

To further ensure the absence of peaking structures in the signal region, fake combinatorial background is produced from data following two methods

1. The candidate photon is traded for the candidate of another event. Taking the next event or ten events after showed no difference (as one could expect

due to multiple candidates).

2. The candidate photon is flipped around the Z (beam) axis, *i.e.* $(E_\gamma, \mathrm{PX}, \mathrm{PY}, \mathrm{PZ})$ is traded for $(E_\gamma, -\mathrm{PX}, -\mathrm{PY}, \mathrm{PZ})$.

In both cases, the B_s^0 is reconstructed with the new photon and the MLP is recomputed with the updated geometrical variables. This produces large samples of purely combinatorial background. Note that this background is not fully representative of the true combinatorial background as the other selection cuts are not recomputed. Figure 6.6 shows the evolution of the mean MLP with the B_s^0 candidate mass for the two procedures. The absence of peaking structure for both methods ensures that the MLP will not bias the final fit.

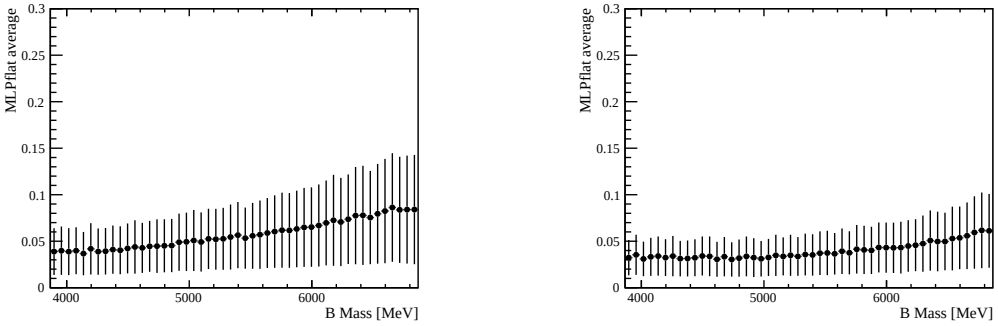

Figure 6.6: Evolution of the mean MLP output computed on faked combinatorial background with respect to the B_s^0 mass. The faked combinatorial background is obtained by flipping (left) or trading (right) the photon of the event. The absence of peak ensures that the MLP cannot reconstruct the B_s^0 mass from its input variables.

6.3.4 Treatment of multiple candidates

As described in [376], the treatment of events with multiple candidates can lead to biases in the branching ratio estimate. Keeping only the candidate with the largest MLP output can indeed produce artificial peaks in the signal region. The mean number of candidates passing all the selection requirements per event is 1.07 for a MLP cut of 0.7. Given this small number, we keep all candidates, which avoids potential biases.

7 Fits and Backgrounds

Contents

This chapter provides a description of the backgrounds that pollute the signal mass window and the fit models used to describe their mass distribution as well as that of the signal.

7.1 Signal Fit

The search for signal events is only performed in the sample passing the optimal MLP cut.

7.1.1 Fit Model

The signal invariant mass distribution is extracted from MC samples and described with a double sided Crystal-Ball p.d.f defined as [368, 377]:

$$\mathrm{CB}_2(m;\mu,\sigma,\alpha_i,n_i) = \begin{cases} a_L\left(b_L - \dfrac{m-\mu}{\sigma}\right)^{-n_L} & \text{for } \dfrac{m-\mu}{\sigma} \leq -\alpha_L, \\ \exp\left\{-\dfrac{(m-\mu)^2}{2\sigma^2}\right\} & \text{for } -\alpha_L < \dfrac{m-\mu}{\sigma} \leq \alpha_R, \\ a_R\left(b_R + \dfrac{m-\mu}{\sigma}\right)^{-n_R} & \text{for } \alpha_R < \dfrac{m-\mu}{\sigma}, \end{cases} \quad (7.1)$$

where $\alpha_{L,R} > 0$ and

$$\begin{aligned} a_i &= \left(\frac{n_i}{\alpha_i}\right)^{n_i} \exp\left\{-\frac{\alpha_i^2}{2}\right\}, \\ b_i &= \frac{n_i}{\alpha_i} - \alpha_i \,. \end{aligned} \quad (7.2)$$

7.1.2 Fit Strategy

Given the rareness of the process, and unless the branching ratio is much larger than predicted in the SM, the signal yield can be estimated to be at most $\mathcal{O}(10)$ (see section 9.3). Leaving all the models parameters free in the fit would hence results in large uncertainties or no convergence at all. n_i, α_i $(i = L, R)$ and σ are therefore fixed to the simulation. The mean μ and the global yield are floated in the fit to data. Fixing the width of the signal to that obtained in the simulation fit results in a systematic error which is studied in section 9.1, based on the control sample.

Tables 7.1 and 7.2 give the values of the parameters for each year as measured on MC samples[1]. The values used for the final fit are the mean parameters weighted by the MC samples size and per-year integrated luminosity. The fits are displayed in Fig. 7.1 (and in Appendix B.3 for the q^2-binned fits).

The behaviour of the signal width σ in the different bins can be understood from the following estimation:

$$\begin{aligned} m(\mu\mu\gamma)^2 &= m(\mu\mu)^2 + 2\,\mathbf{p}_\gamma \cdot (\mathbf{p}_{\mu^+} + \mathbf{p}_{\mu^-}) \\ &= m(\mu\mu)^2 + 2\,E_\gamma\,(E_{\mu^+} + E_{\mu^-} - \frac{\vec{p}_\gamma}{E_\gamma} \cdot (\vec{p}_{\mu^+} + \vec{p}_{\mu^-}))\,. \end{aligned}$$

[1] All the fits presented in this thesis are performed using RooFit [378] and Minuit [379]

Year	n_L	α_L	n_R	α_R	σ [MeV]	μ [MeV]
2016	1.04 ± 0.02	2.22 ± 0.01	6 ± 1	1.62 ± 0.02	92.9 ± 0.1	5347.4 ± 0.7
2017	1.07 ± 0.05	2.21 ± 0.04	5.88 ± 0.8	1.78 ± 0.06	91.6 ± 0.8	5347.9 ± 0.3
2018	1.09 ± 0.04	2.20 ± 0.02	5.99 ± 0.3	1.79 ± 0.02	91.5 ± 0.1	5347.9 ± 0.3
Mean	1.07 ± 0.04	2.21 ± 0.02	6.0 ± 0.6	1.74 ± 0.09	92.0 ± 0.3	–

Table 7.1: Fit parameters for signal MC for each year and for a MVA cut of 0.7 (*i.e.* 30% efficiency on signal). n_i, α_i and σ will be used in the fit on data while μ will be left floating. The last row gives the parameters, fixed in the final fit, weighted by the luminosity.

Mean	n_L	α_L	n_R	α_R	σ [MeV]	μ [MeV]
Bin I	1.08 ± 0.01	2.21 ± 0.01	6.0 ± 0.3	1.74 ± 0.02	91.8 ± 0.1	5347.7 ± 0.2
Bin II	1.6 ± 0.4	2.0 ± 0.1	5 ± 2	1.7 ± 0.1	75 ± 2	5353 ± 2
Bin III	3.0 ± 0.9	1.72 ± 0.04	5.9 ± 0.7	1.42 ± 0.04	40.6 ± 0.5	5364.9 ± 0.5

Table 7.2: Averaged fit parameters for signal MC in bins of q^2. The MLP cuts are set to 0.7, 0.1 and 0.075 for Bin I, II and III respectively. n_i, α_i and σ will be used in the fit on data while μ will be left floating.

The uncertainty of the r.h.s of the previous equation is mainly due to the uncertainty on the photon energy E_γ. Therefore

$$\sigma \approx \sigma_\gamma \, F(\vec{p}_{\mu^+}, \vec{p}_{\mu^-}, \frac{\vec{p}_\gamma}{E_\gamma}) \qquad \text{with } F = \frac{E_{\mu^+} + E_{\mu^-} - \frac{\vec{p}_\gamma}{E_\gamma} \cdot (\vec{p}_{\mu^+} + \vec{p}_{\mu^-})}{M_{B_s^0}}, \tag{7.3}$$

and the uncertainty on the parameters of the F factor are small with respect to the uncertainty σ_γ on the photon energy. From MC simulation we get that σ_γ is almost constant among the q^2 bins[2], while we have $F_{\text{Bin I}} \approx 10\%$, $F_{\text{Bin II}} \approx 8\%$ and $F_{\text{Bin III}} \approx 5\%$, consistent with the width obtained in Table 7.2.

7.2 Combinatorial background

Due to the modest resolution on the photon momentum and the large occupancy in the ECAL, the main background is combinatorial. A relevant pair of muons is randomly associated with the large number of low-energy photons of the event,

[2] Although E_γ is completely fixed by q^2 in the B_s^0 rest frame, these two quantities are almost independent due to the B_s^0 boost in the lab frame.

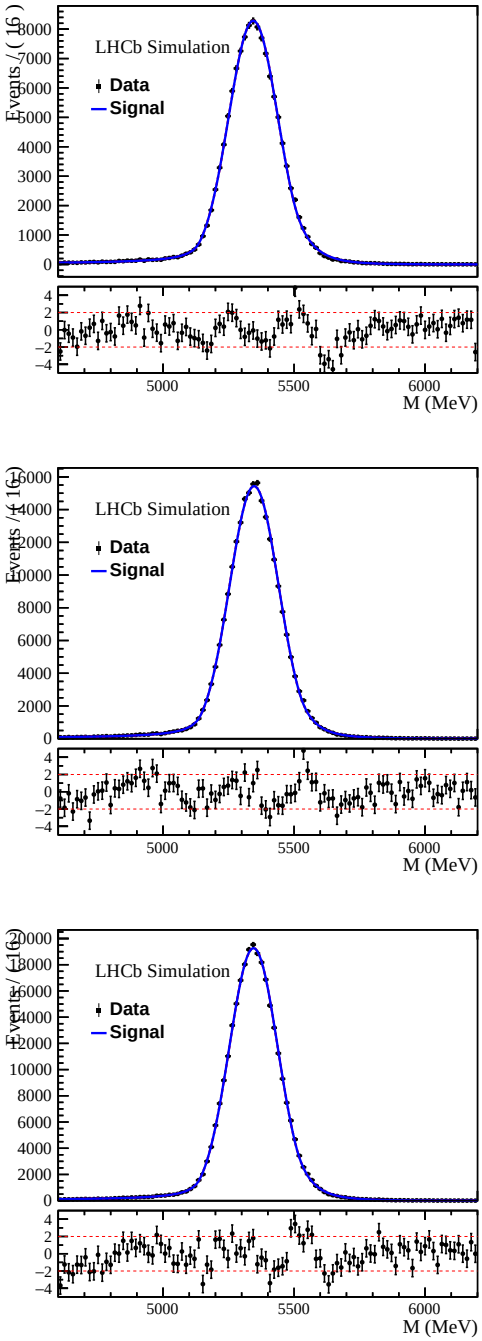

Figure 7.1: Double sided Crystal Ball fit of the B_s^0 candidate mass for signal simulations in 2016 (top), 2017 (middle) and 2018 (bottom). Results of the fits are displayed on Table 7.1.

most of them coming from light mesons decays. This background is usually tackled by increasing the cut on the photon's transverse momentum. As detailed in section 5.4.2, new photon veto variables yield a slightly better handling of this background.

The combinatorial background is described with an exponential distribution; the slope and the yield are floated in the final fit as depicted in Fig. **??**.

7.3 Physical backgrounds

Apart from this combinatorial background, a few peaking backgrounds have been studied:

- decays including muons coming from the charmonium resonances;
- the doubly mis-identified $B_s^0 \to \phi\gamma$ and $B^0 \to K^{*0}\gamma$;
- $B^0 \to \mu^+\mu^-\gamma$ and the partially reconstructed $B^0 \to \mu^+\mu^-\pi^0$ and $B^0 \to \mu^+\mu^-\eta$;
- baryonic decays, such as $\Lambda_b^0 \to pK\gamma$ or $\Lambda_b^0 \to \Lambda\gamma$.

7.3.1 Muons from J/ψ and $\psi(2S)$

The decay $B_s^0 \to J/\psi\,\gamma$ has the same initial and final states as the signal but has never been observed yet. An upper limit has been set on the branching ratio by the LHCb Collaboration $\mathcal{B}(B_s^0 \to J/\psi\,\gamma) < 7.3 \times 10^{-6}$ [146]. This decay is largely suppressed by the veto applied on the dimuon mass at the stripping level. The size of the veto range, $[2880\ \text{MeV}, 3920\ \text{MeV}]$, is chosen to suppress $J/\psi \to \mu\mu$ events by a factor 10^3.

The main backgrounds in the search for $B_s^0 \to J/\psi\,\gamma$ events are $B_s^0 \to J/\psi\,\eta$ ($\mathcal{B} = (4.0 \pm 0.7) \times 10^{-4}$) and $B^0 \to J/\psi\,\pi^0$ ($\mathcal{B} = (1.66 \pm 0.10) \times 10^{-5}$) [146, 367] decays. To estimate the pollution from these decays, $B_s^0 \to \mu^+\mu^-\gamma$ events are reconstructed on simulated $B_s^0 \to J/\psi\,\eta$ samples without the q^2 veto. As depicted in Fig. 7.2, the dimuon mass shows an exponential leak toward smaller masses that makes the charmonium veto less efficient. Fitting the slope of this distribution and integrating below 2880 MeV, we estimate that $(1.89 \pm 0.01)\%$ of the $J/\psi \to \mu\mu$ events survive the charmonium veto.

Taking into account that $\mathcal{B}(J/\psi \to \mu\mu) = (5.961 \pm 0.033)\%$ [367], we get that $J/\psi \to \mu\mu$ events are suppressed by a factor $(1.12 \pm 0.01) \times 10^{-3}$. Given this small number we don't expect J/ψ events from $B_s^0 \to J/\psi\,\eta$ or $B^0 \to J/\psi\,\pi^0$, but residual components would in any case (1) be accounted for by $B^0 \to \mu\mu\pi^0$ and $B_s^0 \to \mu\mu\eta$ backgrounds, discussed later and (2) peak outside the signal region due to the energy loss. As the mass of the $\psi(2S)$ resonance in almost 600 MeV larger than the J/ψ mass, the veto efficiency is even better for decays coming from the $\psi(2S)$, and no background event is expected either.

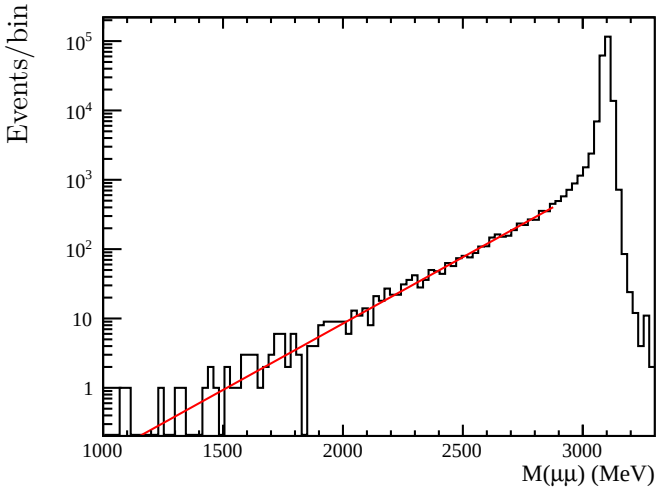

Figure 7.2: Dimuon mass shape of simulated $B_s^0 \to J/\psi\,\eta$ events reconstructed as $B_s^0 \to \mu^+\mu^-\gamma$. The exponential fit (red line) is used to estimate the pollution of charmonium resonances in the signal region.

7.3.2 Hadronic $B \to hh'\gamma$ decays

Peaking backgrounds can emerge from doubly mis-identified hadronic decays. The main ones are $B_s^0 \to \phi\gamma$ where the ϕ decays into two charged kaons, and $B^0 \to K^{*0}\gamma$ where the K^{*0} decays into a pion and a kaon. The mis-identification of the hadrons as muons was accurately studied in the $B_s^0 \to \mu^+\mu^-$ analysis [365]. To estimate the expected yields of these backgrounds, the mis-identification probabilities measured with *PIDCalib* (see section 5.4) are convoluted using the kinematics of weighted signal simulations. The results are presented in Table 7.3 for each year of data-taking.

Year	$B_s^0 \to \phi\gamma,\ \phi \to KK$	$B^0 \to K^{*0}\gamma,\ K^{*0} \to K\pi$
2016	2.7 ± 0.2	4.5 ± 0.7
2017	2.7 ± 0.2	4.2 ± 0.9
2018	2.6 ± 0.2	4.7 ± 0.5

Table 7.3: Effective PID efficiencies for double mis-identified decays in units of 10^{-6}. The meson branching ratio is included in these numbers, *e.g.* for $B_s^0 \to \phi\gamma$ the quoted number is $(\varepsilon(K \to \mu) \otimes \varepsilon(K \to \mu)) \times \mathcal{B}(\phi \to KK)$. The quoted error is the statistical error due to the limited size of the MC samples and the PID tables.

The main source of hadron mis-identification is due to their decay in flight. By

lepton number conservation, the muon is created with an additional neutrino before it is observed by the detector. The mass distributions of these backgrounds are therefore distorted by the missing neutrinos. To estimate this effect, we smeared the momentum of the hadrons in $B \to hh'\gamma$ MC. The smearing parameters are measured on $B_s^0 \to \phi\gamma$ (respectively $B^0 \to K^{*0}\gamma$) simulation imposing that the kaon (resp. the pion) passes the IsMuon algorithm. The difference between the reconstructed and the true momenta are fitted with two Gaussian distributions (see Fig. 7.3 for the distributions and Table 7.4 for the parameters).

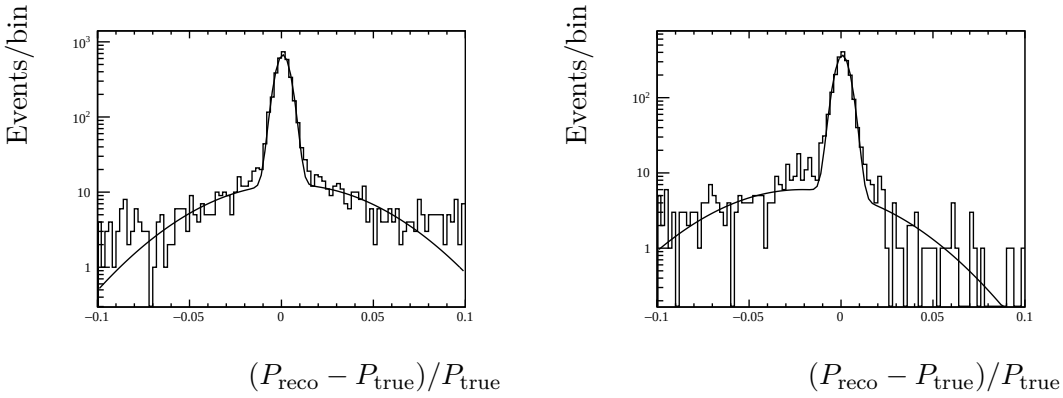

Figure 7.3: Relative error in the reconstruction of the hadrons momentum for kaons (left) and pions (right) passing the IsMuon algorithm. The distributions are fitted with the sum of two Gaussian distributions.

Parameters	Kaons	Pions
F	$(98 \pm 1)\%$	$(98 \pm 1)\%$
μ_1	$(7.9 \pm 0.7) \times 10^{-4}$	$(7.5 \pm 0.9) \times 10^{-4}$
σ_1	$(3.78 \pm 0.01) \times 10^{-3}$	$(3.9 \pm 0.09) \times 10^{-3}$
μ_2	$(4.5 \pm 1.8) \times 10^{-3}$	$-(2.1 \pm 0.3) \times 10^{-2}$
σ_2	$(4.1 \pm 0.3) \times 10^{-2}$	$(4.1 \pm 0.4) \times 10^{-2}$

Table 7.4: Parameters of the double Gaussian fit of the hadron smearing. F is the fraction of events in the main Gaussian (with parameters (μ_1, σ_1)). (μ_2, σ_2) define the second Gaussian and are used to smear the hadrons momentum.

The smeared mass distributions are finally fitted with double-sided Crystal Balls (see eq. 7.1). The result of these fits are shown in Fig. 7.4 and the parameters of the Crystal Balls are given in Table 7.5.

The expected yields, shown in Table 7.7 are small enough to safely neglect

other hh' resonances such as ϕ', f_2', and higher K^* resonances.

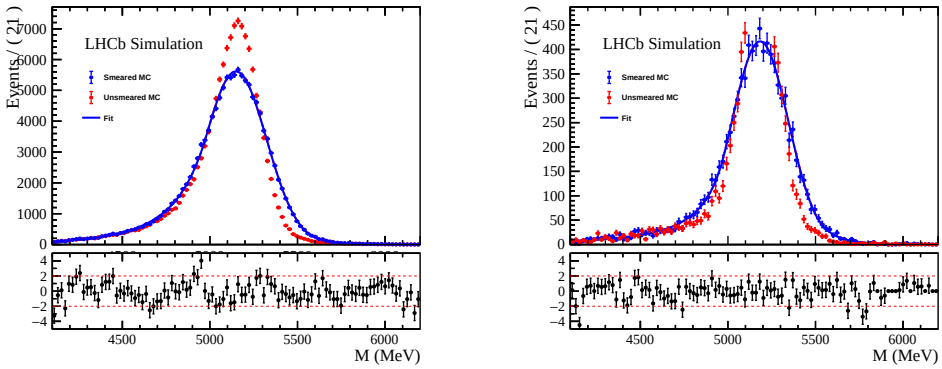

Figure 7.4: Fit of the mass distributions of the two $B \to hh'\gamma$ backgrounds ($B_s^0 \to \phi\gamma$ on the left panel and $B^0 \to K^{*0}\gamma$ on the right panel). The event are reconstructed as $B_s^0 \to \mu^+\mu^-\gamma$ decays by changing the mass hypothesis (red points) and the hadron momenta are smeared using the method described in the text (blue points). The smeared distributions (blue curves) are used to model the backgrounds in the final fit.

Parameters	$B_s^0 \to \phi\gamma$, $\phi \to KK$	$B^0 \to K^{*0}\gamma$, $K^{*0} \to K\pi$
Mean [MeV]	5148 ± 1	5183 ± 2
Width [MeV]	176 ± 1	161 ± 3
n_L	6.1 ± 0.6	2.3 ± 0.3
α_L	0.96 ± 0.03	1.27 ± 0.06
n_R	5.4 ± 0.6	8 ± 7
α_R	1.9 ± 0.3	2.19 ± 0.04

Table 7.5: Results of the fit of the smeared mass distribution for $B_s^0 \to \phi\gamma$ and $B^0 \to K^{*0}\gamma$ backgrounds. Although the initial hadron is heavier, $B_s^0 \to \phi\gamma$ events peak below $B^0 \to K^{*0}\gamma$ events because a K is mis-identified in the former while it is a π for the latter. Note that the convergence of the fit is less good for $B^0 \to K^{*0}\gamma$ due to a smaller statistic in the sample.

7.3.3 $B^0 \to \mu^+\mu^-\gamma$

Apart from a slightly smaller B mass and the CKM suppression, $B^0 \to \mu^+\mu^-\gamma$ decays differ from their B_s^0 counterpart by their light meson resonances. As discussed in section 1.2.3, $B^0 \to \gamma$ form-factors involve the ρ ($M_\rho = 775.26 \pm 0.25$

MeV) and the ω ($M_\omega = 782.65 \pm 0.12$ MeV), while $B_s^0 \to \gamma$ form-factors involve the ϕ ($M_\phi = 1019.461 \pm 0.019$ MeV) [367]. The sensitivity of this analysis is therefore expected to be the same for B_s^0 and B^0 in Bin II and III. Assuming that the branching ratio in Bin I is dominated by the resonances, one can approximate the initial number of events under the narrow-width approximation [124, 367]:

$$\left.\frac{N_{B^0\to\mu\mu\gamma}}{N_{B_s^0\to\mu\mu\gamma}}\right|_{\text{Bin I}} \approx \frac{f_d}{f_s}\frac{\mathcal{B}(B^0\to\rho\gamma)\mathcal{B}(\rho\to\mu\mu)+\mathcal{B}(B^0\to\omega\gamma)\mathcal{B}(\omega\to\mu\mu)}{\mathcal{B}(B_s^0\to\phi\gamma)\mathcal{B}(\phi\to\mu\mu)} \approx 0.03\,. \tag{7.4}$$

In the two other bins, the relative factor is larger [367]:

$$\left.\frac{N_{B^0\to\mu\mu\gamma}}{N_{B_s^0\to\mu\mu\gamma}}\right|_{\text{Bin II}} \approx \left.\frac{N_{B^0\to\mu\mu\gamma}}{N_{B_s^0\to\mu\mu\gamma}}\right|_{\text{Bin III}} \approx \frac{f_d}{f_s}\left|\frac{V_{td}}{V_{ts}}\right|^2 \approx 0.17\,, \tag{7.5}$$

where V is the CKM matrix. $B^0\to\mu^+\mu^-\gamma$ events are therefore considered as negligible in Bin I, and added to the fit only in Bin II and III. In these bins the signal p.d.f is used with a scaling of $M_{B^0}/M_{B_s^0}$ for the mass and the width.

7.3.4 $B^0 \to \mu^+\mu^-\pi^0$

The most difficult background in this analysis is due to the non-resonant $B^0 \to \mu^+\mu^-\pi^0$ decays, although they have never been observed yet. The current upper limit on the branching ratio is set by BaBar and reads $\mathcal{B}(B^0\to\mu^+\mu^-\pi^0) < 6.9\times 10^{-8}$ at 90% CL [380] [3]. The branching ratio can however be estimated from the LHCb measurement of $B^\pm\to\mu^+\mu^-\pi^\pm$ [381]

$$\mathcal{B}(B^\pm\to\mu^+\mu^-\pi^\pm) = (1.83\pm0.25)\times10^{-8}\,, \tag{7.6}$$

and the theoretical estimation [382]

$$\frac{\mathcal{B}(B^0\to\mu^+\mu^-\pi^0)}{\mathcal{B}(B^\pm\to\mu^+\mu^-\pi^\pm)} = 0.47^{+0.22}_{-0.18}\,. \tag{7.7}$$

Combining these two measurements, one gets

$$\mathcal{B}(B^0\to\mu^+\mu^-\pi^0) = (8.6\pm3.6)\times10^{-9}\,. \tag{7.8}$$

The pollution in $B_s^0\to\mu^+\mu^-\gamma$ events is obtained either when

- the π^0 decays into two resolved photons (resolved π^0) ;
- the π^0 is mis-identified as a photon by the calorimeter (merged π^0).

[3] In this measurement the charmonium region is also excluded and used as control channel.

For the merged π^0, most of the energy of the initial B^0 is measured, so a Gaussian shape, centred around the B^0 mass, is expected. The width of this Gaussian is due to the uncertainty on the π^0 momentum and is expected to be comparable to that of $B_s^0 \to \mu^+\mu^-\gamma$ events. However, as the size of the π^0 shower may exceed the 3×3 cells cluster of the ECAL, some energy can be missed in the reconstruction, implying a spread of the distribution to lower mass. This effect can be accounted for by using a simple Crystal Ball distribution instead of a Gaussian.
Concerning the resolved π^0, one photon is not reconstructed and the invariant mass of the remaining particles can be fitted by an Argus distribution with a kinematic limit close to the B^0 mass[4]. To account for the uncertainties on the energy, this Argus distribution is convoluted with a Gaussian, centred on zero (this convoluted distribution will be denoted as Argus⊗Gaussian distribution in the following).
The reconstruction efficiency is different in each category due to the different kinematics and because some cuts have different impacts on the merged and resolved π^0. For instance, merged π^0 are mainly affected by the `IsPhoton` algorithm, while resolved π^0 are tackled by the 'improved' veto (section 5.4.2) and kinematic variables.

Now, it appears that the convergence of the fit is challenged by (a) the larger suppression of resolved events due to the kinematics, (b) the overlap between the two distributions and (c) the limited statistics of the simulated sample. In practice, the fit converges *on MC*, as shown in Fig. 7.5, left panel. However, generating an amount of event similar to the one expected in the final fit shows that it is unable to resolve Merged and Resolved π^0.
On the other hand, fitting both Resolved and Merged events with one single double-sided Crystal Ball (eq. 7.1) shows stable performances and this simpler shape is kept for the final fit. The fit, performed on MC, is shown on Fig. 7.5, right panel, and the results of the fit are given in Table 7.6.

7.3.5 $B^0 \to \mu^+\mu^-\eta$

In the same vein as $B^0 \to \mu^+\mu^-\pi^0$, other light meson resonances, such as $B^0 \to \mu^+\mu^-\eta$, could pollute the signal mass region. This decay has never been observed outside the charmonium peak, and an upper limit was set by the BaBar collaboration [380]

$$\mathcal{B}(B^0 \to \mu^+\mu^-\eta) < 1.12 \times 10^{-7}\,. \tag{7.9}$$

The SM branching ratio is expected to lie in the $[2.5, 3.7] \times 10^{-8}$ range, with an uncertainty mainly due to the lack of knowledge on the $B^0 \to \eta$ form-factors [383]. A small set of events (20k) is produced using the DecFile 11412200 and the 2016 setup. The contribution of merged η is negligible so we fit the mass distribution with the same distribution as for resolved $B^0 \to \mu^+\mu^-\pi^0$ decays, namely an Argus⊗Gaussian distribution. The result of the fit is shown on Fig. 7.6. As the mass of the η is larger than the mass of the π^0, the energy of the photons is also

[4] The distributions used in this part are defined in Appendix B.4

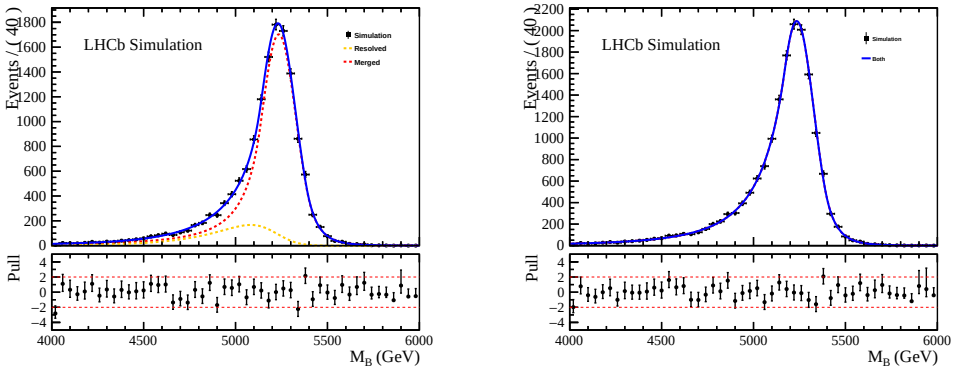

Figure 7.5: Mass distribution of the $B^0 \to \mu^+\mu^-\pi^0$ events (reconstructed as $B_s^0 \to \mu^+\mu^-\gamma$) with two fitting procedures. On the left panel, the fit has two components, the merged π^0 (dashed red) are fitted with a Crystal Ball distribution while the resolved π^0 (dashed orange) are fitted by an Argus distribution. On the right panel, both contributions are fitted with a double-sided Crystal Ball. This second method is kept for the final fit.

larger, yielding a broader peak in the mass distribution, with lower pollution than for $B^0 \to \mu^+\mu^-\pi^0$ decays.

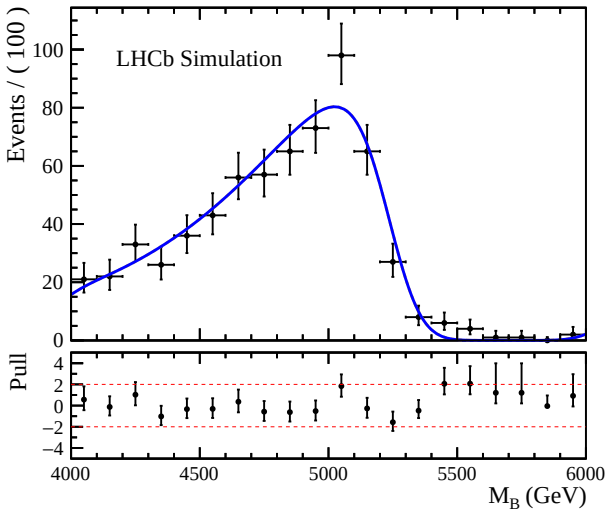

Figure 7.6: Mass distribution of the $B^0 \to \mu^+\mu^-\eta$ events (reconstructed as $B_s^0 \to \mu^+\mu^-\gamma$). The fit is an Argus distribution convoluted with a Gaussian distribution.

Parameters	$B^0 \to \mu^+\mu^-\pi^0$	$B^0 \to \mu^+\mu^-\eta$
Effective mass [MeV]	5249 ± 2	5267 ± 2
Resolved χ	–	-6.3 ± 0.3
Merged width [MeV]	91 ± 5	–
Merged α_r	1.602 ± 0.001	–
Merged n_r	9 ± 8	–
Merged α_l	0.703 ± 0.001	–
Merged n_l	3.71 ± 0.03	–

Table 7.6: Fit parameters obtain on simulated samples, see text for details. Effective mass refers to the μ parameter of the Crystal Ball distribution and to the upper bound of the Argus distribution. χ refers to the Argus parameter.

7.3.6 Λ_b decays

One could expect background events from the $\Lambda_b \to pK\gamma$ decays, where the proton and the kaon are misidentify as muons. The branching ratio $\mathcal{B}(\Lambda_b^0 \to pK\gamma)$ has never been measured yet, but the estimate

$$\mathcal{B}(\Lambda_b \to pK\gamma) = (3.39 \pm 0.48) \times 10^{-5} \tag{7.10}$$

was obtained in [384] for the resonant and non-resonant contributions for $m(pK)$ in $[1400, 2600]$ MeV. Using conservative values for the mis-identification probabilities $\varepsilon(p \to \mu) < 10^{-3}$ and $\varepsilon(K \to \mu) < 10^{-2}$ and the hadronisation fraction $f_{\Lambda_b^0}/f_{B_s^0} \approx 2.1$ [156], one get an upper limit on the effective branching fraction $\mathcal{B}_{\text{eff}} < 7.1 \times 10^{-10}$. Assuming that the other efficiencies are comparable to the signal efficiency, this background is, at most, one order of magnitude smaller than the signal. As $m(\Lambda_b^0) = 5619.60 \pm 0.17$ [367], the peak would furthermore be in the upper sideband and we neglect this background in the following.

The same procedure can be applied to $\Lambda_b^0 \to \Lambda\gamma$ decays, whose branching fraction has recently be measured by the LHCb collaboration [385]

$$\mathcal{B}(\Lambda_b^0 \to \Lambda\gamma) = (7.1 \pm 1.8) \times 10^{-6}. \tag{7.11}$$

The two branching fractions $\mathcal{B}(\Lambda \to p\pi) = (63.9 \pm 0.5)\%$ and $\mathcal{B}(\Lambda \to p\mu\nu) = (1.57 \pm 0.35) \times 10^{-4}$ [367] yield effective branching ratios of the order 10^{-11}, far below the current sensitivity.

7.3.7 Yields

The background yields expected in the final fit can be roughly estimated from the efficiencies previously measured and the integrated luminosity. These yields are

summarized in Table 7.7. In this table we estimated global uncertainties assuming that the errors computed for each separate efficiency are uncorrelated. This is not true, *e.g.* the statistical error for the PID and the multivariate analysis efficiency are both due to the limited size of the simulation sample.

We did not restrict the yields to the signal region $[5167, 5567]$ MeV, but used the larger range $[3867, 6867]$ MeV. Hence, the backgrounds with a large yield but a small pollution will not be a problem for the final fit. For example, the $B^0 \to \mu^+\mu^-\eta$ yield is large, especially in Bin II and III, but its mass distribution (*cf.* Fig. 7.6) is such that the impact on the signal region is small.

Background	Expected BR	Global efficiency	Yield
$B_s^0 \to \phi\gamma$	$(1.7 \pm 0.2) \times 10^{-5}$ [367]	$(1.6 \pm 0.1) \times 10^{-8}$	0.03 ± 0.01
$B^0 \to K^{*0}\gamma$	$(4.18 \pm 0.25) \times 10^{-5}$ [367]	$(1.8 \pm 0.4) \times 10^{-8}$	0.4 ± 0.1
$B^0 \to \mu^+\mu^-\pi^0$	$(8.6 \pm 3.6) \times 10^{-9}$ Eq. (7.8)	$(6.81 \pm 0.05) \times 10^{-4}$	3 ± 1
$B^0 \to \mu^+\mu^-\eta$	$[2.5, 3.7] \times 10^{-8}$ [383]	$(2.8 \pm 0.4) \times 10^{-4}$	4 ± 1

Table 7.7: Expected yields for the main backgrounds for an integrated luminosity of 5.57/fb and the optimal cut MLP>0.7. Note that no other bounds than the one applied at the stripping level are applied on the B_s^0 candidate mass, so the background events can be far from the signal region (esp. for $B^0 \to \mu^+\mu^-\eta$ events, cf. Fig. 7.6).

Background	Bin I	Bin II	Bin III
$B_s^0 \to \phi\gamma$	0.06 ± 0.01	–	–
$B^0 \to \mu^+\mu^-\gamma$	0.10 ± 0.02	–	0.26 ± 0.03
$B^0 \to K^{*0}\gamma$	0.8 ± 0.2	–	–
$B^0 \to \mu^+\mu^-\pi^0$	3 ± 1	10 ± 4	9 ± 4
$B^0 \to \mu^+\mu^-\eta$	3 ± 1	28 ± 6	16 ± 4

Table 7.8: Expected yields in q^2 bins for the main backgrounds for an integrated luminosity of 5.57/fb. – means that the yield is too small to be computed and the background is neglected.

8

Normalization and control channels, weighting and trigger efficiencies

Contents

8.1 Normalization

Due to the uncertainty on the estimation of the integrated luminosity and on the $b\bar{b}$ production cross-section, the signal branching ratio is normalized to $B_s^0 \to J/\psi\,\eta$ (where the J/ψ decays into two muons and the η into two photons). This also allows a partial cancellation in the ratio between the different efficiencies which lowers the global uncertainty on the final result. The branching ratio is therefore expressed as

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-\gamma) = \frac{\mathcal{B}_{\rm norm}}{N_{\rm norm}} \times f_{\rm norm} \times N_{B_s^0\to\mu\mu\gamma}\,, \tag{8.1}$$

where the normalization constant is divided in several ratios

$$f_{\text{norm}} = \frac{\varepsilon^{\text{Acc}}_{\text{norm}}}{\varepsilon^{\text{Acc}}_{B^0_s\to\mu\mu\gamma}} \times \frac{\varepsilon^{\text{Strip \& Reco}}_{\text{norm}}}{\varepsilon^{\text{Strip \& Reco}}_{B^0_s\to\mu\mu\gamma}} \times \frac{\varepsilon^{\text{Sel}}_{\text{norm}}}{\varepsilon^{\text{Sel}}_{B^0_s\to\mu\mu\gamma}} \times \frac{\varepsilon^{\text{Trig}}_{\text{norm}}}{\varepsilon^{\text{Trig}}_{B^0_s\to\mu\mu\gamma}} \times \frac{\varepsilon^{\text{PID}}_{\text{norm}}}{\varepsilon^{\text{PID}}_{B^0_s\to\mu\mu\gamma}} \times \frac{\varepsilon^{\text{MLP}}_{\text{norm}}}{\varepsilon^{\text{MLP}}_{B^0_s\to\mu\mu\gamma}} \,. \tag{8.2}$$

The normalization channel is usually chosen to have a similar kinematics as the signal and, as far as possible, similar final-state particles. The choice of $B^0_s \to J/\psi\,\eta$ was made because it presents muons and low energetic photons (compared *e.g.* to $B^0_s \to \phi\gamma$). On the other hand, its statistic is not large, which triggers larger statistical errors.

For $B^0_s \to J/\psi\,\eta$ branching ratio, the latest CP-averaged values read [367]

$$\mathcal{B}(B^0_s \to J/\psi\,\eta) = (4.0 \pm 0.7) \times 10^{-4}\,, \tag{8.3}$$
$$\mathcal{B}(J/\psi \to \mu\mu) = (5.961 \pm 0.033)\%\,, \tag{8.4}$$
$$\mathcal{B}(\eta \to \gamma\gamma) = (39.41 \pm 0.20)\%\,, \tag{8.5}$$
$$\mathcal{B}_{\text{norm}} = (9.3 \pm 1.6) \times 10^{-6}\,. \tag{8.6}$$

Note the two important following points:

1. The 17% error on the $B^0_s \to J/\psi\,\eta$ branching ratio directly impacts the final $B^0_s \to \mu^+\mu^-\gamma$ branching ratio measurement.

2. Any new measurement of $B^0_s \to J/\psi\,\eta$ will directly improve this measurement, so the final result will be presented in such a way that it can be easily updated.

8.1.1 Backgrounds

The following backgrounds are considered for the normalization channel:

- $B^0 \to J/\psi\,\eta$: the B^0 counterpart of the B^0_s decay. The ratio of branching ratios experimentally averages to [40]

$$\frac{\mathcal{B}(B^0 \to J/\psi\,\eta)}{\mathcal{B}(B^0_s \to J/\psi\,\eta)} = 0.027 \pm 0.008\,, \tag{8.7}$$

  in agreement with LHCb's measurement $0.0185 \pm 0.0061 \pm 0.0014$ [386]. These events are fitted with the same shape as for the signal, but the relative contribution are floated in the fit to take into account the difference of efficiency. The final yield is finally cross-checked with the theoretical prediction eq. (8.7).

- $B \to J/\psi\,\eta\,\pi$ and $B \to J/\psi\,\eta\,K$, where the hadron is missed, give peaks in the left side band. The shape of these backgrounds is extracted from simulation and described with Argus⊗Gaussian distributions. The upper bound of the Argus distribution is fixed to $(m_B - m_h)$, with $h = K, \pi$.
- $B \to J/\psi\,\eta'$ and $B \to \psi(2S)\eta$ can both lead to a $J/\psi\,\eta\pi\pi$ final state which pollutes the low B-mass region. These backgrounds are well described by Gaussian distributions (centred around 4750 MeV and with a width of less than 100 MeV), but they peak below the mass-window chosen for the fit and their pollution is negligible.
- The combinatorial background is fitted with an exponential distribution.

8.1.2 Fit and yields

$B_s^0 \to J/\psi\,\eta$ events are fitted with a double-sided Crystal Ball distribution with tail parameters fixed to simulation. The yields are given in Table 8.1 for the full dataset 2016-2017-2018 and the fit is shown in Fig. 8.1. Note that a decay tree fitter is applied on the B_s^0 decays with constraints on the η and J/ψ masses to reduce the B_s^0 width.

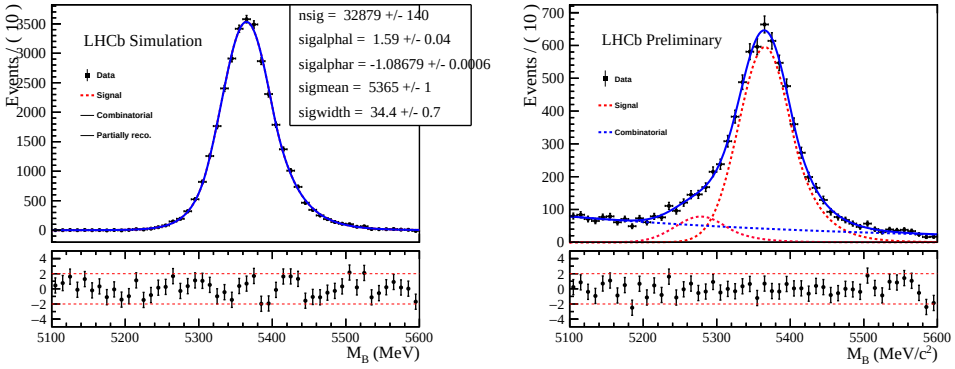

Figure 8.1: Fit of the normalization channel. The fit is performed on MC (left) to extract tail parameters for the double-sided Crystal Ball. The fit is then performed on the B_s^0 candidate mass using 2016, 2017 and 2018 data (5.57/fb, right). The yields and parameters are given in Table 8.1.

We obtained the following yields:

$$\begin{aligned} N^{2016}_{B_s^0 \to J/\psi\,\eta} &= 1712 \pm 42\,, \\ N^{2017}_{B_s^0 \to J/\psi\,\eta} &= 1723 \pm 45\,, \\ N^{2018}_{B_s^0 \to J/\psi\,\eta} &= 2153 \pm 50\,, \end{aligned} \tag{8.8}$$

$$N^{2016-2017-2018}_{B_s^0 \to J/\psi\,\eta} = 5600 \pm 100\,. \tag{8.9}$$

8.1.3 Cross-checks

The expected number of $B_s^0 \to J/\psi\,\eta$ event can also be evaluated using simulation and the measured integrated luminosity. We have

$$N_{\rm norm}^{\rm exp} = \int \mathcal{L}dt \times (2 f_s\, \sigma(pp \to b\bar{b} + X)) \times \mathcal{B}_{\rm norm} \times \varepsilon_{\rm norm}\,, \tag{8.10}$$

$$\varepsilon_{\rm norm} = \varepsilon_{\rm norm}^{\rm Acc} \times \varepsilon_{\rm norm}^{\rm Strip\ \&\ Reco} \times \varepsilon_{\rm norm}^{\rm PID} \times \varepsilon_{\rm norm}^{\rm MLPS}\,, \tag{8.11}$$

where $2 f_s\, \sigma(pp \to b\bar{b} + X)$ gives the number of B_s^0 produced in LHCb[1] and $\mathcal{B}_{\rm norm}$ is defined in eq. (8.6).

We use $f_s = 0.103 \pm 0.005$ [389] and $\sigma(pp \to b\bar{b}+X)_{13\ \rm TeV} = 495 \pm 2 \pm 52\ \mu$b [390]. Injecting the efficiencies presented in Table 9.4 in eq. (8.10), the expected number of $B_s^0 \to J/\psi\,\eta$ events reads

$$N_{\rm norm}^{\rm exp} = 7800 \pm 1400 \pm 800\,. \tag{8.12}$$

Where the first is statistical (mainly due to the uncertainty on the normalization branching ratio), and the second is systematic. This number is encouragingly close to the observed yield and the ratio of expected/observed yield is year independent within errors. The remaining discrepancy can be associated to an overestimation of the luminosity (due to failed jobs, missing data...) and to the remaining data/MC differences. For both categories, a cancellation of the errors is expected in the ratio between the signal and normalization yields. The remaining data/MC differences are already accounted for as systematic uncertainties on the multivariate analysis cuts efficiencies.

As a second test, we checked that the number of $B^0 \to J/\psi\,\eta$ is in the expected ballpark to ensure that the mixing between B_s^0 and B^0 events is not large. Using eq. (8.7) and the ratio f_s/f_d [387] one get an effective ratio of 0.097 ± 0.035, which is in good agreement with the fit output in Table 8.1 that yields 0.130 ± 0.013. The uncertainty due to this overlap is already present in the fit output.

As a final cross-check, the fitting procedure is performed in two steps. First, the η mass is fitted with a double-sided Crystal Ball distribution with tails parameters fixed to simulation and a linear background. Then the *sPlot* method [370] is used to subtract the η background from the B_s^0 mass plot. The final fit is performed on this decay-tree-fitted, background-subtracted mass spectrum. This procedure ensures a very clean peak with low remaining background, see Fig. 8.2. The results of this method are compatible with the direct fit.

[1] Note that the measurement $\sigma(pp \to B_s^0 + X)$ also exists [387, 388] and eq. (8.10) could be written as

$$N_{\rm norm}^{\rm exp} = \int \mathcal{L}dt \times \sigma(pp \to B_s^0 + X) \times \mathcal{B}_{\rm norm} \times \varepsilon_{\rm norm}\,.$$

However, $\sigma(pp \to B_s^0 + X)$ is measured in these references with an acceptance cut already accounted for in $\varepsilon_{\rm norm}$, which would result in a double-counting.

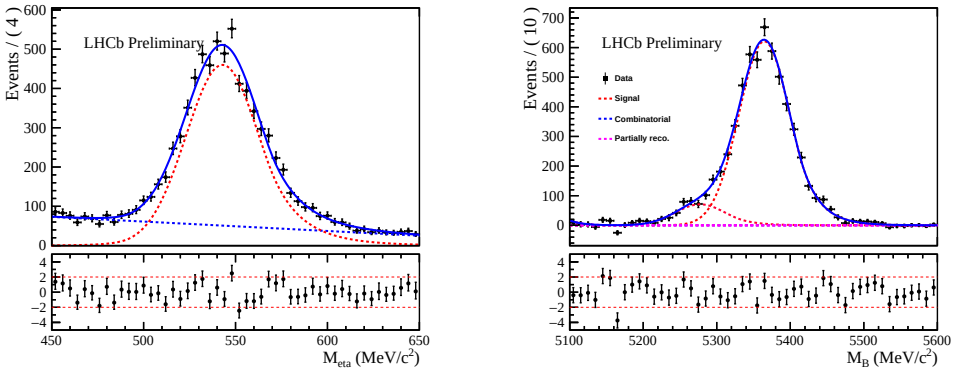

Figure 8.2: Cross-check fit for the normalization channel $B_s^0 \to J/\psi\,\eta$. The fit is performed in two steps (see the text for details) on 2016, 2017 and 2018 data (5.57/fb). The yields and parameters are given in Table 8.1.

Channel	Variable	Direct Fit	*sWeighted* Fit
$\eta \to \gamma\gamma$	μ_η [MeV]	–	542.9 ± 0.4
	σ_η [MeV]	–	20.3 ± 0.4
	$N_{\eta\to\gamma\gamma}$	–	6180 ± 120
$B_s^0 \to J/\psi\,\eta$	$\mu_{B_s^0}$ [MeV]	5365.4 ± 0.8	5364.7 ± 0.7
	$\sigma_{B_s^0}$ [MeV]	35.4 ± 0.9	34.2 ± 0.6
	$\boldsymbol{N_{B_s^0\to J/\psi\,\eta}}$	$\mathbf{5600 \pm 100}$	5530 ± 80
	$N_{B^0\to J/\psi\,\eta}$	727 ± 70	620 ± 40
	$N_{B^0\to\psi(2S)\eta}$	< 100	28 ± 5

Table 8.1: Results of the normalization channel fit on 2016, 2017 and 2018 data (5.57/fb). Two methods are used, the direct fit of the B_s^0 candidate mass and the two-step procedure described in the text as a cross-check. The MLP cut is fixed at 0.02. μ, σ and N respectively stand for the mean, the width and the yield of the distributions.

8.2 Control channel

Signal simulation is not perfect and one needs to estimate the change of efficiency due to the entailed discrepancies. The impact of these imperfections on the analysis is twofold. First, the efficiencies computed on simulation may be biased. Second, multivariate analysis are trained on imperfect MC, so the selection will loose efficiency by mis-classifying events. Discrepancies can be found in global variables (number of tracks, number of hits in the SPD...), in the reconstruction (tracks and vertices χ^2...) or in the theoretical inputs (resonances, form-factors...). Global variables can be corrected for by an appropriate weighting, estimated by

comparing any MC and data samples. Reconstruction variables require samples with similar particles, kinematics and selections. Finally, the impact of theoretical inputs can be checked by weighting the MC using updated parameters.

As the normalization channel has a different kinematics from the signal, a second control channel is needed. $B_s^0 \to \phi\gamma$ decays, where the ϕ decays into two kaons, is therefore used to control the global variables and some of the reconstruction variables. The branching fraction of this decay, $(3.4 \pm 0.4) \times 10^{-5}$ [40], is more than three orders of magnitude larger than the signal one, which yields a large enough statistics to perform the comparisons.

8.2.1 Event selection and yields

To improve the similarity with $B_s^0 \to \mu^+\mu^-\gamma$ events, $B_s^0 \to \phi\gamma$ events are reconstructed as 3-body decays, namely without imposing additional mass or vertex requirement on the two K's on top of those present in the stripping line. For example, constraining the reconstructed ϕ mass to its PDG value would reduce the B_s^0 width by a couple of MeV as can be seen by comparing Fig. 8.2 ($\sigma = 90.2 \pm 0.9$ MeV/c^2) to the $B_s^0 \to \phi\gamma$ analysis [368] where this method is applied (Table 3.6, $\sigma = 86.31 \pm 1.97$ MeV/c^2). On the other hand, imposing this ϕ nominal mass may alter other variables distribution and blur the comparison. The absence of further constraints therefore avoids potential bias in the variables distributions that could arise from the decay reconstruction procedure. Furthermore, the comparison is performed on TOS events only to simplify the comparison. Indeed, TIS events can have different cuts especially on the ϕ transverse momentum which impacts the variables distributions.

The selection of events is performed by training a multivariate analysis based on the same variables as the signal MLPS (see section 5.5) on $B_s^0 \to \phi\gamma$ MC and data side-bands. To avoid any bias in the data-MC comparison, the cut on this multivariate analysis has to be kept as low as possible. A large cut would enforce data to mimic MC distributions and would blur the comparison. A cut of 0.1 on the flatten MLPS output was found to be a good compromise. This approach implies that the `sPlot` [370] has to be used to deal with the large combinatorial background in order to extract the distribution of variables that are then compared to the simulation.

The B_s^0 candidate mass distribution is fitted with a double-sided Crystal Ball with tails parameters fixed on MC, the background distributions (described in the next paragraph) and an exponentially decreasing component for the remaining combinatorial background (the resulting fit is shown on Fig. 8.3 and Table 8.2.). Background subtracted variables can then be extracted using *sPlot* and compared with MC. To avoid large weights from events coming from the tails of the signal distribution (where the ratio between signal and background is large), the mass window is reduced to $[5000, 5700]$ MeV. The results of this comparison are discussed in the following section.

8.2.2 Backgrounds

The backgrounds found in the dedicated $B_s^0 \to \phi\gamma$ analysis are [368]:

- $B \to \phi\gamma K$, the MC was available and fitted with an Argus⊗Gaussian distribution;
- $B_s^0 \to \phi\gamma\pi^0$, was found to be negligible;
- $\Lambda_b^0 \to \Lambda^*(pK)\gamma$ and the non-resonant $B \to KK\pi^0$ have contamination of the order of 2%. These contaminations are small enough not to alter the variables distribution and these backgrounds are neglected.

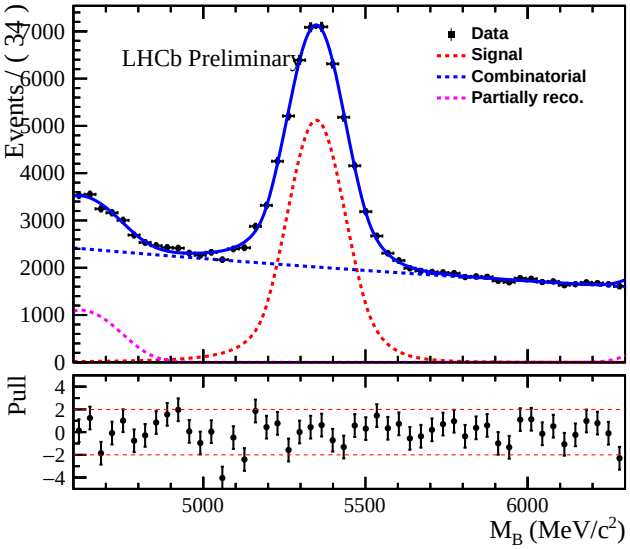

Figure 8.3: Fit of $B_s^0 \to \phi\gamma$ events using 5.57/fb of data. The parameter of the fit are displayed in Table 8.2. *SWeights* are extracted from the fit and used to compare data and MC. The partially reconstructed background is due to $B \to \phi\gamma K$ decays.

8.3 MC-data discrepancies and weighting

MC-data discrepancies are studied by comparing background-subtracted data and simulation distributions. As neither $B_s^0 \to \phi\gamma$ nor $B_s^0 \to J/\psi\,\eta$ completely reproduces $B_s^0 \to \mu^+\mu^-\gamma$ kinematics, we use both samples to study the discrepancies. This double-check yields the following conclusions, all the comparison plots being shown in Appendix B.5:

- As expected, *global* variables discrepancies are visible and consistent for $B_s^0 \to \phi\gamma$ and $B_s^0 \to J/\psi\,\eta$. The number of tracks, long tracks and hits in

Channel	Variable	Direct Fit	
		Data	MC
$B_s^0 \to \phi\gamma$	$\mu_{B_s^0}$ [MeV]	5348.5 ± 0.8	5352.9 ± 0.1
	$\sigma_{B_s^0}$ [MeV]	90.2 ± 0.9	94.0 ± 0.1
	$N_{B_s^0 \to \phi\gamma}$	36000 ± 300	$(4.25 \pm 0.01) \times 10^5$
	$N_{B^+ \to \phi\gamma K^+}$	5400 ± 200	–

Table 8.2: Results of the control channel fit on 2016, 2017 and 2018 data (5.57/fb) and MC simulation. The MVA cut is fixed at 0.1.

the SPD are not well reproduced. These variables being correlated with other variables, these discrepancies impact other distributions.

- The B_s^0 vertex χ^2 shows large discrepancies. They are due first to the mis-modelling of the number of tracks but also to the tracking of the kaons in $B_s^0 \to \phi\gamma$. As can be checked on $B_s^0 \to J/\psi\,\eta$ events, correcting for the number of tracks is enough to improve the distribution of this variable (see later in text and Table 8.3).

- The distribution of the B transverse momentum is also distorted in both $B_s^0 \to \phi\gamma$ and $B_s^0 \to J/\psi\,\eta$ samples.

- Despite its correlation with the number of tracks and SPD hits, the photon cone isolation is correctly reproduced.

- Muon isolation variables are also correctly reproduced.

The strategy used to account for these discrepancies is twofold. In a first step, weights are computed from the control channel and applied to the simulated samples. This limits the mis-training of the multivariate analysis. Then, the remaining discrepancies are measured on the control and the normalization channels and propagated through the multivariate analysis. The change of efficiency of the multivariate analysis cuts is considered as a systematic error.

8.3.1 Weighting

The weighting of the MC samples is performed on three variables, the number of long tracks (`nLongTracks`), the energy of the B (`B_PE`) and the pseudo-rapidity of the B (`B_eta`). Weighting on `nLongTracks` allows to correct the variables correlated to the event occupancy such as isolations and vertex χ^2. Correcting for the discrepancies in `B_PE` and `B_eta` allows to correct for `B_PT`, on which the MLP output showed a large correlation, and for the children particles kinematics. (The weighting cannot be done directly on `B_PT` because this variable has different cut in

the signal, normalization and control samples.) The distribution of `nLongTracks`, `B_PE` and `B_eta` are presented in Fig. 8.4.

Two methods have been tested for the weighting procedure. The simplest method consists in an uniform binning of the weighting variables distributions. The weights are then computed by taking the ratio of events between data and simulations in each of these bins. This method gives unsatisfactory results, mainly because bins with few events lead to non-physical weights. To improve this method, the binning is not chosen uniformly, but by a multivariate analysis. Regression trees are trained to find the best possible weights and bins. This method, known as *GradBoost* weighting [391], is far more stable than the "naive" approach. To increase stability and avoid a training bias, the control sample is randomly divided into two parts, and a classifier is trained on each part and tested on the other. The predicted weight is obtained as the average of the two classifiers output.

The post-weighting discrepancies visible on Fig. 8.4 are mainly due to the fact that, to compute weights, the data sample needs to be pure enough. Although the selection applied on data is kept as light as possible, the distributions are slightly biased. This bias is accounted for by comparing all the remaining discrepancies.

8.3.2 Remaining discrepancies

The remaining discrepancies are evaluated on the normalization and the control channels. As the shapes of the variable distributions are usually the same for MC and data, the difference of the mean of the distributions is considered as a good approximation of the error. The shape of the distributions are shown in Appendix B.5, and the values obtained by this method are given in Table 8.3. The values used to estimate the multivariate analysis systematic errors are given in the third column. For most of the variables, the physical error is the relative error. However, for muon isolations and the `CosP` variable, the mean of the distribution is close to zero and the absolute difference is used.

For `gamma_dm_eta`, the normalization channel gives unexpected results. This is due to the method used to extract the variables, the *sWeights* are largely correlated with the eta mass which is fully correlated with `gamma_dm_eta`. The error measured on the control channel is therefore used.

8.4 Ratios of trigger efficiency

The trigger efficiency can be estimated on weighted simulations or using data and the TISTOS method [392]. Both approaches are presented here.

8.4.1 MC estimation

The main difficulty in the estimation of the trigger efficiency is that the trigger configuration changed many times during the years, each configuration yielding

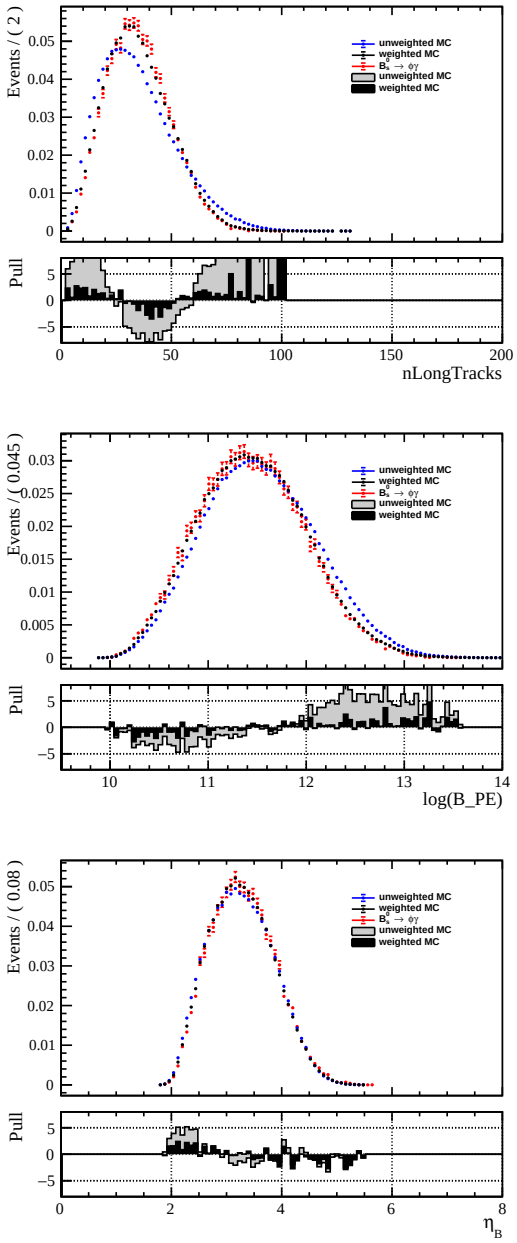

Figure 8.4: Distribution of the variables used for the MC weighting and extracted from $B_s^0 \to \phi\gamma$ events. The correction on the number of long tracks improves the distribution of variables correlated to the multiplicity in the event. Correcting on the B candidate energy and pseudo-rapidity mainly improves those of other geometrical variables. The weighting is performed using a *GradBoost* weighting, see text for details.

Variable	$B_s^0 \to \phi\gamma$	$B_s^0 \to J/\psi\,\eta$	Used error
log `B_ip`	$(1.8 \pm 0.1)\%$	$(1.2 \pm 0.3)\%$	1.2%
log `B_ips`	$(9.6 \pm 0.7)\%$	$(8 \pm 2)\%$	8%
`B_PT`	$(-3.0 \pm 0.3)\%$	$(1.1 \pm 0.8)\%$	1.1%
`B_vchi2`	$(-8.1 \pm 0.3)\%$	$(-0.01 \pm 0.7)\%$	−0.01%
log `B_dira`	$(2.6 \pm 0.1)\%$	$(1.1 \pm 0.3)\%$	1.1%
`mu_doca`	$(-1.1 \pm 0.2)\%$	$(-0.8 \pm 0.4)\%$	−0.8%
log `mu_minip`	$(-4 \pm 1)\%$	$(-0.4 \pm 3)\%$	−0.4%
log `mu_minips`	$(-1.3 \pm 0.2)\%$	$(-0.6 \pm 0.5)\%$	−0.6%
`DeltaRmumu`	$(2.1 \pm 0.3)\%$	$(-1.9 \pm 0.8)\%$	−1.9%
log `gamma_PT`	$(-0.06 \pm 0.03)\%$	$(0.2 \pm 0.09)\%$	0.2%
log `gamma_PE`	$(0.14 \pm 0.04)\%$	$(0.06 \pm 0.1)\%$	0.06%
`gamma_isolation`	$(2.7 \pm 0.3)\%$	$(0.9 \pm 0.5)\%$	0.9%
`gamma_npi0`	$(5 \pm 2)\%$	$(13 \pm 4)\%$	13%
`gamma_neta`	$(1.0 \pm 0.8)\%$	$(2.7 \pm 1.0)\%$	1%
`gamma_dmpi0`	$(-3 \pm 1)\%$	$(-5 \pm 3)\%$	−5%
`gamma_dmeta`	$(-1.3 \pm 0.6)\%$	$(40 \pm 2)\%$	−1.3%
muons Long isolation	$(-3.1 \pm 0.2)\ 10^{-3}$	$(-1.4 \pm 0.9)\ 10^{-3}$	$-1.4\ 10^{-3}$
muons Velo isolation	$(-2.1 \pm 0.4)\ 10^{-2}$	$-(0.8 \pm 0.1)\ 10^{-2}$	$-0.8\ 10^{-2}$
`CosP`	0.006 ± 0.005	0.01 ± 0.01	0.01

Table 8.3: Variables uncertainties computed by comparing *sPlotted* data and weighted MC distributions for the control and the normalization channels. As the distributions are similar, the values correspond to the relative difference of the mean of the distributions $((\mu_{\text{data}} - \mu_{\text{MC}})/\mu_{\text{MC}})$ and the quoted error is due to the limited statistics of the data sample. For `CosP` and the muon isolations, the absolute error is used as the distributions are centred on zero and the relative error has no physical interpretation.

a different efficiency. For the present analysis, three lines are impacted by these trigger changes, `L0Photon`, `L0Muon` and `Hlt1TrackMVA`.

At the L0 level, these changes concern the L0 transverse energy threshold of the photon and the L0 transverse momentum threshold of the muons. These changes were necessary to keep a constant trigger rate, especially in 2016 where the calibration of the photon energy was not fully efficient.

The `Hlt1TrackMVA` line selects tracks with a high quality ($\chi^2 < 2.5$) and large P_{T} and χ^2_{IP}. For energetic tracks ($P_{\text{T}} > 25000$ MeV) the constraint is given by

$\chi^2_{\text{IP}} > 7.4$, while the other tracks are constrained by an ellipsoidal cut:

$$\ln(\chi^2_{\text{IP}}) > \ln(7.4) + \frac{1}{(P_{\text{T}}/1000 - 1)^2} + \frac{A}{25000}(25000 - P_{\text{T}})\,, \tag{8.13}$$

where the transverse momentum P_{T} is expressed in MeV. The constant A changed during 2016 data taking between 1.1, 1.6 and 2.3, hereafter referred to as Loose, Tight and Tighter configurations. Note that `Hlt1TwoTrackMVA` properties did not change.

As the MC samples only simulate one TCK per year, one needs to simulate the different configurations offline to compute an average efficiency. The trigger configurations are summarized in Table 8.4 for 2016, tables 8.5 for 2017 and 8.6 for 2018[2]. As apparent in these tables, the MC does not simulate the TCK with the smallest thresholds. To avoid any bias and to correctly compute the trigger efficiency, the minimal thresholds are fixed to the simulated one. In other words, the L0 requirement eq. (5.2) is kept for 2018 and modified for 2016 and 2017 to

$$\begin{aligned} \text{L02016} \;=\; & (\texttt{L0Muon \&\& L0MuonPt} > 1850\ \text{MeV}) \\ & \|\,(\texttt{L0Photon \&\& L0PhotonEt} > 2784\ \text{MeV})\,; \end{aligned} \tag{8.14}$$

$$\begin{aligned} \text{L02017} \;=\; & (\texttt{L0Muon \&\& L0MuonPt} > 1450\ \text{MeV}) \\ & \|\,(\texttt{L0Photon \&\& L0PhotonEt} > 2472\ \text{MeV})\,. \end{aligned} \tag{8.15}$$

To compute the efficiency of the photon trigger, we therefore estimated the efficiency of the different thresholds in Tables 8.4 and 8.5 by simulating the trigger cut offline and weighted the efficiencies with the corresponding luminosities. Note that, as described in [393], the transverse momenta (`L0PhotonEt` and `L0MuonPt`) used for the L0 decision does not correspond to the offline reconstructed `gamma_PT` and `muplus_PT` or `muminus_PT`. For the photon, this is due to the fact that at the L0 level, the energy is computed from 2×2 cells clusters while 3×3 clusters are used offline. The transverse energy is therefore larger offline (except for a negligible amount of events). To properly simulate the trigger cuts, we used the L0 values used for the trigger decision. As the global TIS decision also depends on these thresholds, TIS is decomposed as photon TIS, muon TIS and the rest. The thresholds can the be applied as described above. Note that the hadron trigger thresholds also varied during data taking, but the impact on the TIS decision efficiency is small.

The efficiency is estimated step by step:

1. L0M $\equiv$ `L0Muon`

[2] In the TCKs, P_{T} thresholds are stored in ADC counts. The conversion factor between ADC counts and MeV changed between Run I and Run II. Here we used 24 MeV/ADC counts for the photons and 50 MeV/ADC counts for the muons. For the latter, the cuts were directly performed on the ADC counts so this conversion factor does not impact our results.

TCK	Luminosity [pb^{-1}]	Thresholds [MeV]		HLT1
		`L0Photon` E_T	`L0Muon` P_T	
`0x11291600`	12.26	2112	750	Loose
`0x11291603`	34.99	2304	1150	Loose
`0x11291604`	24.74	2785	1350	Loose
`0x11291605`	77.12	2976	1550	Loose
`0x11321609`	104.96	2800	1350	Loose
`0x11341609`	116.36	2800	1350	Tight
`0x11351609`	22.24	2800	1350	Tighter
`0x11361609`	400.93	2800	1350	Tighter
`0x11371609`	68.13	2800	1350	Tighter
`0x1137160e`	22.42	2976	1550	Tighter
`0x1138160f`	**563.66**	**2784**	**1850**	**Loose**
`0x11381611`	43.61	2976	1550	Loose
`0x11381612`	89.65	2976	1650	Loose
`0x11381609`	6.85	2800	1350	Loose
`0x1138160e`	31.10	2976	1550	Loose
TOTAL	1619.03			

Table 8.4: L0 and Hlt1 configurations used during 2016 data taking and the corresponding luminosities. Signal MC was simulated using TCK `0x6138160f`, equivalent to `0x1138160f`, in bold in the table.

2. L0P $\equiv$!L0M `&& L0Photon`

3. L0TIS $\equiv$!L0M `&&` !L0P `&& L0Global_TIS`,

where muons and photons are required to stay above the thresholds defined in eqs. (8.14) and (8.15).

The resulting efficiencies are given in Tables. 8.7, 8.8 and 8.9 and the global L0 efficiency is given in Table 8.10.

8.4.2 L0 Data driven estimation

The main issue of measuring the efficiency on MC, apart from changes in the thresholds, is that the calibration and the ageing of the ECAL changed the *effective* thresholds of the trigger lines. Furthermore, at the beginning of 2016 data taking, the calibration constants were assigned wrong values. This error affects the efficiency of the trigger lines that use information from the ECAL. Although

TCK	Luminosity [pb^{-1}]	Thresholds [MeV]		HLT1
		`L0Photon` E_T	`L0Muon` P_T	
`0x114e1702`	2.69	2112	750	Loose
`0x114e1703`	1.66	2304	1150	Loose
`0x11501703`	22.37	2304	1150	Loose
`0x11501704`	25.38	2784	1350	Loose
`0x11501705`	123.72	2976	1550	Loose
`0x11501706`	27.76	3072	1950	Loose
`0x11541707`	89.30	2712	1750	Loose
`0x115417a7`	0.51	2712	1750	Loose
`0x11561707`	327.57	2712	1750	Loose
`0x11601707`	125.62	2712	1750	Loose
`0x11601708`	99.07	2304	1150	Loose
`0x11611707`	126.90	2712	1750	Loose
`0x11611708`	132.11	2304	1150	Loose
`0x11611709`	**577.24**	**2472**	**1450**	**Loose**
TOTAL	1681.89			

Table 8.5: L0 and Hlt1 configurations used during 2017 data taking and the corresponding luminosities. Signal MC was simulated using TCK `0x62661709`, equivalent to `0x11611709`, in bold in the table.

this error is also simulated in the MC, a cross-check of the trigger efficiencies is needed.

The trigger efficiency can be measured directly on data using the TISTOS method [392]. The idea is basically to measure the TOS probability on the TIS sample (TIS and TOS were defined in section 4.2). Under the assumption that TIS and TOS efficiency are uncorrelated (which is true once the distribution of the decaying meson is unfolded), the TOS efficiency on the TIS sample is indeed equal to the global TOS efficiency.

To apply this method, trigger efficiency tables are extracted from data and convoluted using the simulated signal kinematics. $B_s^0 \to \phi\gamma$ events are used to produce photon efficiency tables. For muon tables, the statistics of $B_s^0 \to J/\psi\,\eta$ is unfortunately too small. Following the $B_s^0 \to \mu^+\mu^-$ analysis, $B^+ \to J/\psi\,K^+$ events are therefore used for muon and TIS estimations. Events are extracted from the stripping line `Bs2MuMuLinesBu2JPsiKLine`, the B^+ mass is fitted with a double sided Crystal Ball and an exponentially decreasing background and efficiencies are extracted using the *sPlot* method [370].

To produce the photon and muon tables, events are required to pass the general

TCK	Luminosity [pb^{-1}]	Thresholds [MeV]		HLT1
		L0Photon E_T	L0Muon P_T	
0x11671801	5.04	2952	1800	Loose
0x11711801	5.58	2952	1800	Loose
0x11731801	125.42	2952	1800	Loose
0x11741801	276.52	2952	1800	Loose
0x11751801	289.94	2952	1800	Loose
0x11771801	211.02	2952	1800	Loose
0x117718a1	0.49	2952	1800	Loose
0x117a18a2	851.50	2952	1800	Loose
0x117a18a4	**365.09**	**2952**	**1800**	**Loose**
TOTAL	2130.61			

Table 8.6: L0 and Hlt1 configurations used during 2018 data taking and the corresponding luminosities. All the configurations used the same thresholds. Signal MC was simulated using TCK 0x617d18a4, equivalent to 0x117a18a4, in bold in the table.

TIS requirement eq. (5.5). As discussed in [392], the B kinematics is binned to ensure that TIS and TOS are uncorrelated. The binning for the B kinematics is extracted from the quantiles of the signal simulated distributions

- $\boldsymbol{B}$ **PZ** (MeV) $\leq 64742 : 97751 : \geq 152591$;
- $\boldsymbol{B}$ **PT** (MeV) $\leq 4380 : 6413 : 8570 : \geq 11820$.

The choice of 4×5 bins is motivated by the analysis of the efficiency of the TISTOS method and is evaluated in [392] to yield a $(0.5 \pm 0.4)\%$ relative bias.

8.4.3 LO TOS efficiency

The photon trigger efficiency is estimated in bins of photon transverse momentum, and the muon efficiency is also divided in bins of impact parameter. The tracks binning is adapted from the quantiles of the signal distributions:

- $\boldsymbol{\gamma}$ **PT** (MeV) $\leq 3230 : 3960 : 5000 : \geq 6973$;
- $\boldsymbol{\mu}$ **PT** (MeV) $\leq 1500 : 2000 : 3000 : \geq 5000$;
- $\boldsymbol{\mu}$ **IP** (mm) $\leq 0.1 : 0.2 : \geq 0.5$.

For the muon, the efficiency tables are filled with the muon of largest transverse momentum, as this is the one that triggered *a priori*. The muon and photon efficiency tables, averaged over the B kinematics, are displayed in Fig. 8.5 and 8.6. Non-averaged plots are also shown for L0Muon in Appendix B.6.

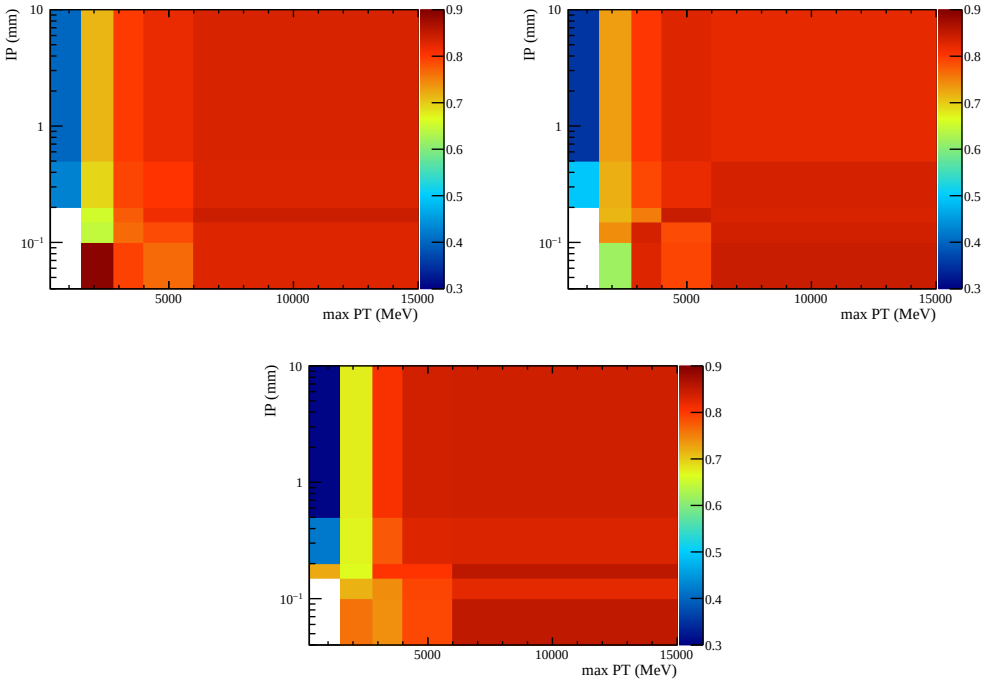

Figure 8.5: `L0Muon` trigger efficiency tables extracted from the muon with the largest transverse momentum in $B^+ \to J/\psi K^+$ 2016 (top left), 2017 (top right) and 2018 (bottom) events. An average is taken over the B kinematics.

8.4.4 TIS efficiency

For the TIS efficiency estimation, $B^+ \to J/\psi K^+$ events are required to pass the global TOS requirement

$$\texttt{L0Global\_TOS} \;\&\&\; \texttt{Hlt1Phys\_TOS} \;\&\&\; \texttt{Hlt2Phys\_TOS}\,. \tag{8.16}$$

The 4×5 binning is then used to produce TIS efficiency tables. We are implicitly assuming here that the correlation between TIS and TOS are only due to the B kinematics. This is actually *not* the case, especially due to the reconstruction in the calorimeters. For example, a constraint on the number of hits in the SPD impacts both TIS and TOS decisions. On the other hand, as events are mainly triggered via the muons for which the absence of correlation has been tested [392], we neglect these additional correlations. The overall agreement between the data-driven and the simulated efficiency ensures that this approach is sensible and additional correlations are accounted for in the systemic uncertainty.

The efficiency tables are shown in Fig. 8.7. The results of this L0 TIS efficiency estimation for the signal and normalization channels are given in Table 8.7.

The results of this TIS efficiency estimation for the signal and normalization channels are given in Table 8.9.

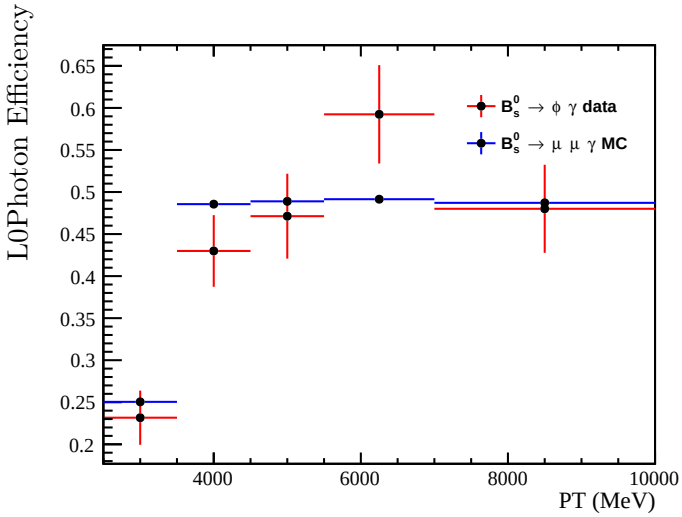

Figure 8.6: L0Photon Trigger efficiency tables extracted from $B_s^0 \to \phi\gamma$ and measured on $B_s^0 \to \mu^+\mu^-\gamma$ MC (2016, 2017 and 2018 averaged). An average is also taken over the B meson kinematics.

Decay	Year	L0M Efficiency [%]	
		Data	Weighted MC
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$65.94 \pm 0.14 \pm 3.22$	64.95 ± 0.14
	2017	$68.19 \pm 0.11 \pm 1.66$	70.60 ± 0.11
	2018	$65.81 \pm 0.09 \pm 1.90$	66.86 ± 0.09
$B_s^0 \to J/\psi\,\eta$	2016	$76.65 \pm 0.46 \pm 1.26$	77.32 ± 0.44
	2017	$77.76 \pm 0.46 \pm 1.00$	–
	2018	$76.30 \pm 0.46 \pm 1.07$	–

Table 8.7: L0M efficiency computed on efficiency tables (first column). The first error is due to the MC statistics while the second one is due to the trigger tables statistics. The second column is the simulated efficiency. For $B_s^0 \to J/\psi\,\eta$, 2017 and 2018 efficiencies are obtained using the 2016 MC kinematics. The large differences between the years are due to changes of the L0Muon line threshold.

Despite the large statistical errors, adding the L0 efficiencies measured with the two methods hint a slight disagreement, that may come from many sources. The main one may be the mis-modelling of the B transverse momentum, on which the trigger efficiency (especially for TIS) is largely dependent. As the signal and the normalization kinematics differ, as shown in Fig. 8.8, the cancellation of errors is only partial. The unfolding of the B kinematics may also be not fully efficient,

Decay	Year	L0P Efficiency [%]	
		Data	Weighted MC
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$11.09 \pm 0.06 \pm 1.65$	13.11 ± 0.06
	2017	$10.42 \pm 0.04 \pm 1.12$	10.25 ± 0.04
	2018	$11.07 \pm 0.04 \pm 1.24$	10.12 ± 0.03
$B_s^0 \to J/\psi\,\eta$	2016	$5.50 \pm 0.08 \pm 0.66$	5.14 ± 0.11
	2017	$5.26 \pm 0.08 \pm 0.60$	–
	2018	$5.47 \pm 0.08 \pm 0.63$	–

Table 8.8: L0P efficiency computed on efficiency tables (first column). The first error is due to the MC statistics while the second one is due to the trigger tables statistics. The second column is the simulated efficiency. The events are required to pass the cut MLPS > 0.25. For $B_s^0 \to J/\psi\,\eta$, the MC efficiency is computed on 2016 MC only.

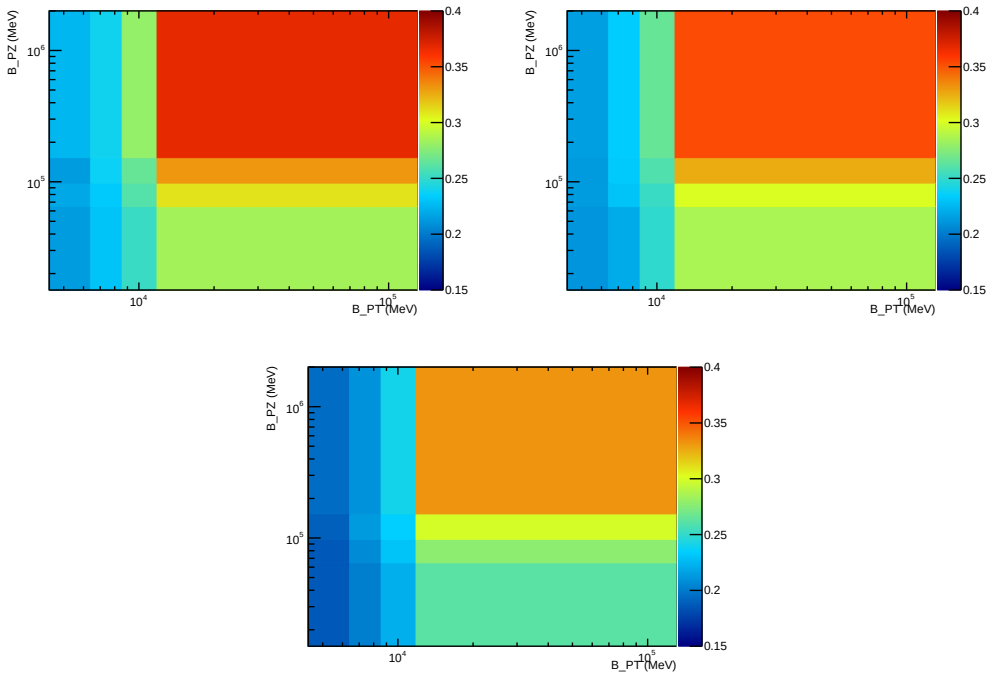

Figure 8.7: L0Global TIS efficiency tables extracted from $B^+ \to J/\psi\,K^+$ 2016 (top left), 2017 (top right) and 2018 (bottom) events.

explaining the larger disagreement in the TIS efficiency.

However, given the overall consistency between the MC estimation with simulated thresholds and the data-driven TISTOS method for the L0 efficiency, we

Decay	Year	L0TIS Efficiency [%]	
		Data	Weighted MC
$B_s^0 \to \mu^+\mu^-\gamma$	2016	$5.11 \pm 0.04 \pm 0.57$	5.88 ± 0.04
	2017	$4.68 \pm 0.03 \pm 0.35$	5.60 ± 0.03
	2018	$4.49 \pm 0.03 \pm 0.35$	5.65 ± 0.03
$B_s^0 \to J/\psi\,\eta$	2016	$4.53 \pm 0.08 \pm 0.32$	4.07 ± 0.10
	2017	$4.21 \pm 0.08 \pm 0.27$	–
	2018	$4.04 \pm 0.08 \pm 0.25$	–

Table 8.9: L0TIS efficiency as defined in eq. (5.5) computed on efficiency tables (first column). The first error is due to the MC statistics while the second one is due to the trigger tables statistics. The second column is the simulated efficiency which is higher due to the absence of trigger prescales in the simulations. For $B_s^0 \to J/\psi\,\eta$, 2017 and 2018 efficiencies are obtained using the 2016 MC kinematics.

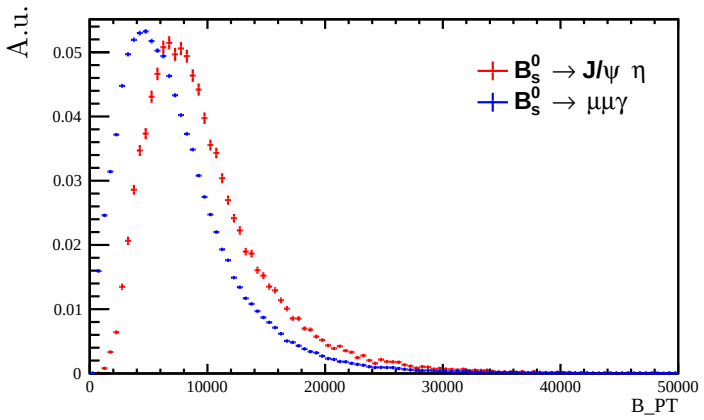

Figure 8.8: Normalized B_s^0 transverse momentum distributions for $B_s^0 \to J/\psi\,\eta$ and $B_s^0 \to \mu^+\mu^-\gamma$ simulated events. The difference in the shape is due to the cuts on the photon and the η in $B_s^0 \to J/\psi\,\eta$ events.

used the former in the estimation of the normalization constants. We also considered that given the fair agreement for the L0 trigger, Hlt1 and Hlt2 efficiency can be computed on simulated samples (correcting again the thresholds).

The efficiencies that result from this study are given in table 8.10. The impact of the different L0 thresholds is visible in the global efficiency where a 3% change is observed between 2017 and 2018. A Hlt2 line, dedicated to the selection of $B_s^0 \to \mu^+\mu^-\gamma$ events, will be added in LHCb Run III. Being based on a selection close to the Stripping one, it should improve the Hlt2 efficiency by $\sim 10\%$. The

binned efficiency are measured on simulation using the same method, the results are shown in Tables 8.11 and 8.12.

Decay	Year	$\varepsilon^{\text{L0\|MLPS}}$ [%]	$\varepsilon^{\text{Hlt1\|L0}}$ [%]	$\varepsilon^{\text{Hlt2\|Hlt1}}$ [%]	$\varepsilon^{\text{Trigger\|MLPS}}$ [%]
$B_s^0 \to \mu^+\mu^-\gamma$	2016	83.94 ± 0.16	99.89 ± 0.08	87.70 ± 0.18	73.53 ± 0.15
	2017	86.45 ± 0.12	99.91 ± 0.07	86.36 ± 0.13	74.59 ± 0.11
	2018	82.63 ± 0.11	99.92 ± 0.07	86.56 ± 0.12	71.46 ± 0.10
$B_s^0 \to J/\psi\,\eta$	2016	86.53 ± 0.47	99.97 ± 0.02	100	86.50 ± 0.47

Table 8.10: Efficiency of the trigger requirements eqs. (5.2), (5.3) and (5.4), measured on weighted simulation and average on the TCK configurations. Note that the final statistical error is not the sum in quadrature of each trigger level error because the latter are fully correlated. The 100% Hlt2 efficiency for the normalization channel comes from the fact that Hlt2 constraints are weaker than the constraints we applied at stripping level.

Decay	Year	Bin	L0M [%]	L0P [%]	L0TIS [%]
$B_s^0 \to \mu^+\mu^-\gamma$	2016	I	65.32 ± 0.14	13.42 ± 0.06	5.98 ± 0.04
		II	71.04 ± 0.98	9.42 ± 0.36	4.56 ± 0.25
		III	91.82 ± 1.28	0.65 ± 0.11	1.90 ± 0.18
	2017	I	70.07 ± 0.11	10.50 ± 0.04	5.69 ± 0.03
		II	76.59 ± 0.76	6.47 ± 0.22	4.66 ± 0.18
		III	93.40 ± 0.98	0.57 ± 0.07	1.43 ± 0.12
	2018	I	66.27 ± 0.10	10.37 ± 0.04	5.75 ± 0.03
		II	73.26 ± 0.67	6.74 ± 0.20	4.54 ± 0.17
		III	92.51 ± 0.87	0.33 ± 0.05	1.55 ± 0.11

Table 8.11: Signal efficiencies of the trigger requirement L0M, L0P and L0TIS per q^2 bin, measured on weighted simulation and averaged on the TCK configurations.

Decay	Year	Bin	Efficiency [%] $\varepsilon^{\text{L0\|MLPS}}$	$\varepsilon^{\text{Hlt1\|L0}}$	$\varepsilon^{\text{Hlt2\|Hlt1}}$	$\varepsilon^{\text{Trigger\|MLPS}}$
$B_s^0 \to \mu^+\mu^-\gamma$	2016	I	84.72 ± 0.16	99.90 ± 0.09	87.42 ± 0.18	73.99 ± 0.15
		II	85.02 ± 1.07	99.94 ± 0.05	89.91 ± 1.19	76.40 ± 1.02
		III	94.37 ± 1.30	99.98 ± 0.02	99.07 ± 0.92	93.47 ± 1.12
	2017	I	86.26 ± 0.12	99.91 ± 0.08	86.06 ± 0.13	74.17 ± 0.11
		II	87.72 ± 0.81	99.96 ± 0.04	89.21 ± 0.86	78.22 ± 0.77
		III	95.40 ± 0.99	99.97 ± 0.03	99.03 ± 0.96	94.45 ± 0.82
	2018	I	82.39 ± 0.11	99.92 ± 0.08	86.26 ± 0.12	71.01 ± 0.10
		II	84.54 ± 0.72	99.96 ± 0.04	89.00 ± 0.79	75.21 ± 0.67
		III	94.39 ± 0.88	99.96 ± 0.04	98.84 ± 0.93	93.26 ± 0.71

Table 8.12: Per bin efficiency of the trigger requirements eqs. (5.2), (5.3) and (5.4), measured on weighted simulation and average on the TCK configurations. Note that the final statistical error is not the sum in quadrature of each trigger level error because the latter are fully correlated.

9 Systematic uncertainties and expected sensitivity

The tools developed in the previous chapters allows us to estimate the expected sensitivity of this analysis. These tools are however subject to systematics uncertainties that diminish the effectiveness of the overall analysis. In section 9.1, the sources of such uncertainties are scrutinize. This allows us to accurately estimate the normalization factor in section 9.2 and the sensitivity to signal events in section 9.3.

9.1 Systematic uncertainties

The methods used throughout this analysis are subject to systematic uncertainties. Although a partial cancellation of these uncertainties is expected in the ratio between signal and normalization yields, the imperfect resemblance between these two channels forces us to study these uncertainties carefully.

9.1.1 Stripping & Reconstruction

Simulated events are reconstructed with the same algorithm as for the reconstruction of data. This implies that background events appear in the simulated samples due to events that are wrongly reconstructed. An algorithm is used to match reconstructed events with true events.

For each sample the following matching condition is required [394]:

1. The reconstructed objects ($\mu, \gamma, B_s^0, K, \dots$) are correctly matched to the simulated ones (the Monte-Carlo identifier `TRUEID` corresponds to the particle identity).

2. The parents of the particles, if specified in the decay chain, are correctly matched.

3. If two particles come from the same grand-parent, their grand-parent should indeed be the same.

These conditions are strong enough to ensure that no background event remains in the simulated samples. On the other hand, some events are correctly reconstructed but wrongly matched because they fail any of the previous items. This can lead to an underestimation of the selection efficiencies. To estimate this effect, the mass distribution of events failing the truth matching requirement is fitted and the resulting number of events is considered as a systematic error on the number of true events in the MC sample. The fits are presented in Fig. 9.1 for 2016 events and the corresponding systematic uncertainties, computed separately for each year, are accounted for in the reconstruction efficiency (summarized in Table 5.8). This approach is conservative because peaking backgrounds can also be part of the Gaussian shape in Fig. 9.1 (as hinted by the difference between the width of the Gaussian fit and the signal width). As the main source of mismatched events is the tracking, we considered systematic effects for up and down polarities and summed them in quadrature.

A second error is due to the discrepancies between data and MC. As the signal MC sample is stripped, the weights computed in section 8.3, cannot be directly applied. We therefore generated a sample of 30000 unstripped signal events per year and measured the reconstruction and stripping efficiency on weighted events by comparing the weights of generated and reconstructed and stripped events. The difference obtained in the efficiency partially cancels in the ratio between signal and normalization and is therefore added as a systematic uncertainty on the ratio of efficiencies in section 9.2. The values obtained, 0.04 in 2016, 0.02 in 2017 and 0.02 in 2018, are found to be independent of the q^2 bin and small with respect to the first discussed systematic uncertainty.

9.1.2 Charged PID

A first systematic uncertainty is associated to the background subtraction method used in the calibration samples. As shown in Fig. 9.2 for $J/\psi \to \mu\mu$, the large statistics makes the fit robust, but discrepancies are visible on the tails and the peak of the distributions. Following the study performed for the $B_s^0 \to \mu^+\mu^-$ analysis [365], an uncertainty of 0.1% is considered.

A second systematic uncertainty is associated to the choice of the binning by varying the number and limits of the bins. After applying the whole procedure for 10, 15 and 20 bins for **P** and 5, 10 and 15 bins for **nTracks**, a conservative systematic uncertainty of 0.05% is considered.

9.1.3 Neutral PID

A first systematic uncertainty is associated to the fit and the *sPlot* method used to extract tables from $\eta \to \mu\mu\gamma$ decays. To do so we compared the efficiency of

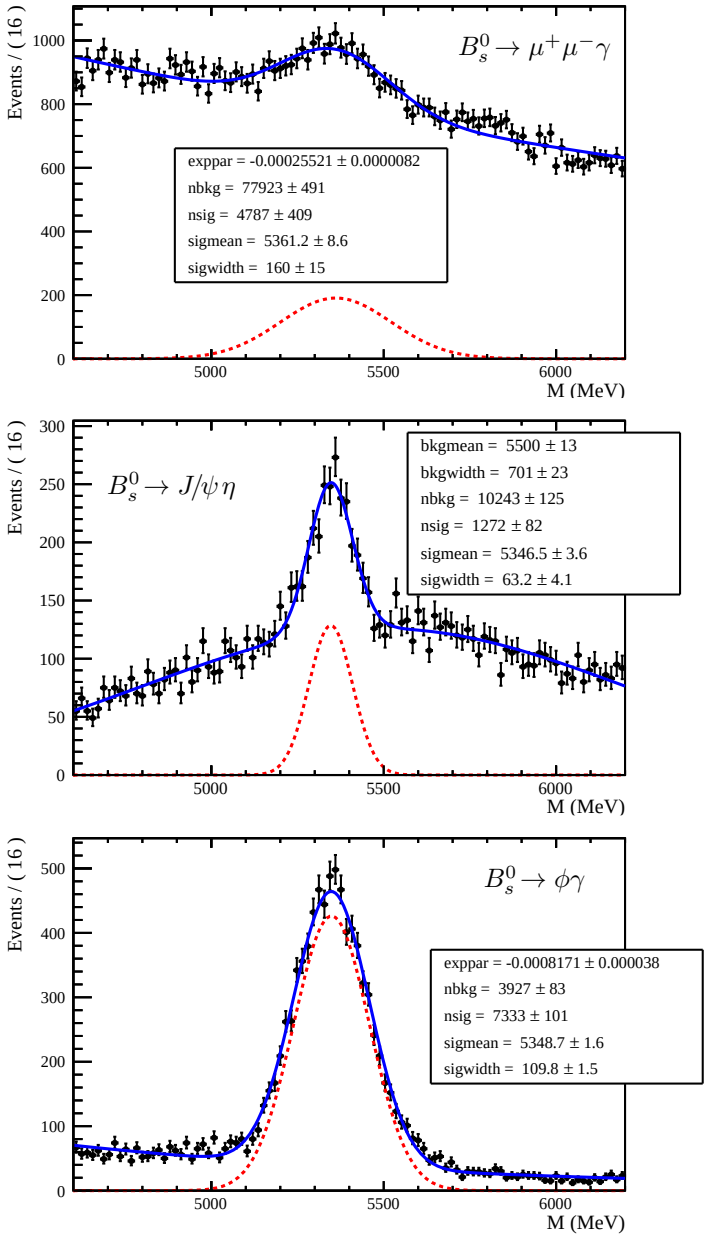

Figure 9.1: Invariant mass distribution of events failing the `MCTruth` requirement for $B_s^0 \to \mu^+\mu^-\gamma$, $B_s^0 \to J/\psi\eta$ and $B_s^0 \to \phi\gamma$ 2016 simulated samples. The signal events are fitted with a Gaussian and the background is model by a exponential or a polynomial distribution.

the PID cut on $\eta \to \mu\mu\gamma$ events by fitting the η mass before and after the cut, to

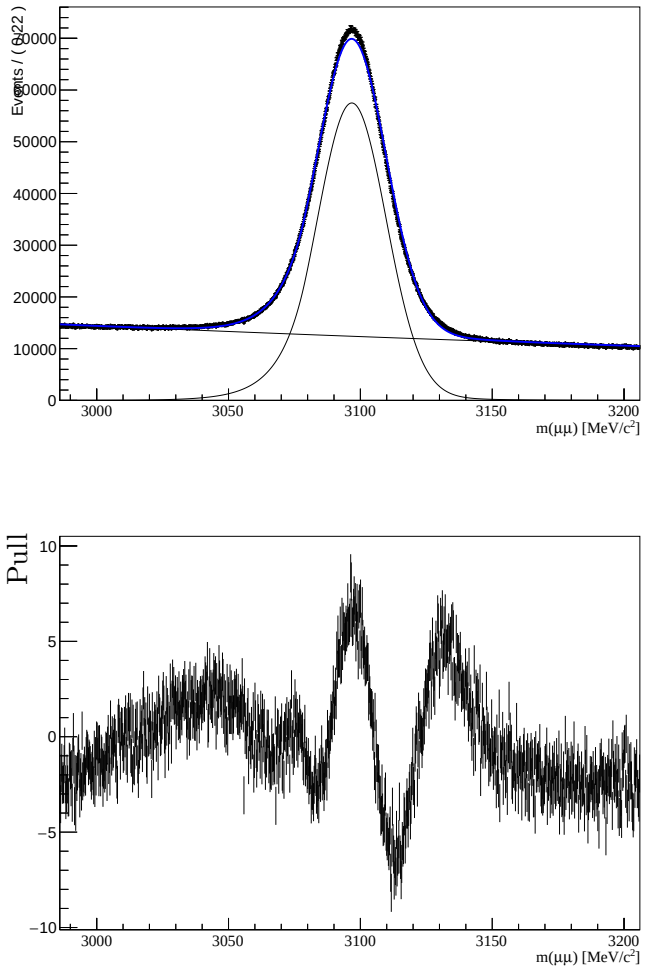

Figure 9.2: Output of the 2016 fit of $J/\psi \to \mu\mu$ events from the PIDCalib package. Almost 9×10^6 events are selected with a global fit $\chi^2 = 7.4$. The bottom plot shows the pull of the fit, the double peak is due to the choice of the fitting function.

the efficiency obtained with the sWeights ("cut and fit" method). The resulting error was found to be 5.2% in 2016, 0.9% in 2017 and 1.1% in 2018. The large difference between 2016 and 2017-18 is due to a less accurate fit in 2016 data.

Concerning the uncertainty due to MC weights, the final ratio was found to be rather dependent on the weighting (mainly through the **n(Long)Tracks** variable. The remaining uncertainty on **nLongTracks** after weighting is around 1%. By applying the full procedure on events where the number of tracks is 1% larger we get a $(0.6 \pm 0.4)\%$ change on the final ratio, independently of the year. We therefore considered a systematic uncertainty of 0.6% due the weighting.

Finally, the uncertainty due to the binning is estimated by varying the number of **P**, η and **n(Long)Tracks** bins. The uncertainty partially cancels in the ratio for $B_s^0 \to \mu^+\mu^-\gamma$ events yielding a 0.05% uncertainty in 2016, and 0.03% in 2017 and 2018. The uncertainties are added in quadrature in Tables 9.4, 9.5 and 9.6. They are also considered independent of the q^2 bin.

9.1.4 MLPS and MLP

Discrepancies between data and simulation have two impacts on multivariate analysis. On one hand, training a multivariate analysis on an approximate simulation will decrease its capacity to recognize signal events. This lowers the multivariate analysis efficiency, but as far as the input variables are not correlated with the candidate mass distribution, this will not bias the final distribution. On the other hand, the multivariate analysis propagates the uncertainties attached to the input variables. This gives rise to a global systematic uncertainty on the efficiency of the multivariate analysis cut.

To evaluate the uncertainty on the cut efficiency of the MLPS and the MLP, the uncertainties computed in Table 8.3 are used. The MLP are then applied on MC where all variables have been shifted by their uncertainties, and the efficiency of the cuts are again estimated for the signal and the normalization channel. The difference of the ratio obtained with the shifted and the non-shifted MLPS are considered as (conservative) uncertainties. For the MLPS, the dependence of these uncertainties on the q^2 bin was found to be small so the same uncertainty is considered for every bins. The results are given in Table 9.1

Year	$\varepsilon^{\text{shifted MLPS}\vert\text{Strip\&Reco}}$		Uncertainty on the ratio
	$B_s^0 \to \mu^+\mu^-\gamma$	$B_s^0 \to J/\psi\,\eta$	
2016	$(73.11 \pm 0.15_{\text{stat}})\%$	$(68.03 \pm 0.35_{\text{stat}})\%$	0.018
2017	$(72.97 \pm 0.11_{\text{stat}})\%$	–	0.016
2018	$(72.94 \pm 0.10_{\text{stat}})\%$	–	0.015

Table 9.1: Systematic uncertainty on the ratio of MLPS efficiencies due to discrepancies in the simulation. The uncertainty is the difference between the ratio of efficiencies of the cut on the shifted and the non-shifted MLPS.

For the MLP, the main source of uncertainty is due to the sensitivity of the output on the B transverse momentum. Although the B energy and pseudo-rapidity are used for the MC weighting, the different cuts applied on this variable for signal, normalization and control events prevents a better treatment of the transverse momentum. The 1.1% uncertainty measured on this variable *via* the comparison of its distribution in $B_s^0 \to J/\psi\,\eta$ events is the main responsible for the shift of the MLP cut efficiency. Furthermore, as different cuts on the MLP output are considered for signal (0.7) and normalization (0.02), the cancellation of uncertainties in the ratio is less efficient than for the MLPS.

The shifts obtained in the ratio between signal efficiency and normalization channel efficiency are summarized in Tables 9.2 and 9.3.

Year	$\varepsilon^{\text{shifted MLP}\|\text{Sel\&Trig\&PID}}$		Uncertainty on the ratio
	$B_s^0 \to \mu^+\mu^-\gamma$	$B_s^0 \to J/\psi\,\eta$	
2016	$(27.35 \pm 0.10_{\text{stat}})\%$	$(90.29 \pm 0.47_{\text{stat}})\%$	0.021
2017	$(26.00 \pm 0.07_{\text{stat}})\%$	–	0.036
2018	$(26.34 \pm 0.07_{\text{stat}})\%$	–	0.032

Table 9.2: Systematic uncertainties on the ratio of MLP efficiencies due to discrepancies in the simulation. The uncertainty is the difference between the ratio of efficiencies of the cut on the shifted and the non-shifted MLP.

Year	Uncertainty on the ratio		
	Bin I (MLP>0.7)	Bin II (MLP>0.1)	Bin III (MLP>0.075)
2016	0.022	0.018	0.014
2017	0.037	0.031	0.018
2018	0.034	0.024	0.017

Table 9.3: Systematic uncertainties on the ratio of MLP efficiencies due to discrepancies in the simulation for each q^2 bin. The uncertainty is smaller in Bin II and III as the MLP cut is closer to the normalization one, ensuring a better cancellation of the uncertainties in the ratio.

9.1.5 Trigger

The two methods used to estimate the trigger efficiency (corrected simulation and TISTOS) showed slight disagreement. Although this disagreement is enhanced by limited control samples, we considered the difference in L0 efficiency as a systematic uncertainty. This difference (2.3%) is measured on 2016 MC and applied to each year and to all the q^2 bins. On the other hand, as the impact of the discrepancies between data and MC were already accounted for in the multivariate analysis efficiencies, we did not add an additional systematic uncertainty for the estimation of Hlt1 and Hlt2 efficiencies.

9.1.6 Fit model

The signal fit procedure is described in section 7.1.2. It is based on the assumption that the signal width is correctly reproduced in the MC. As this width is mainly due to the uncertainty on the photon energy (multiplied by a kinematic function), we can estimate the validity of this hypothesis on the control channel. From Table 8.2, we read that the relative difference in the width between data and MC

is $(4.0 \pm 0.9)\%$. We therefore varied the width in the fits following this uncertainty and considered the variation of the signal yield as a systematic uncertainty.

9.2 Normalization factor

Collecting all the efficiencies measured in the previous chapter and sections, the normalization factor can be extracted. The statistical uncertainties computed on a single sample by successive cuts are only counted once (for example the statistical uncertainties of Reconstruction and Stripping and of the MLPS cut don't appear in the total except for binned efficiencies where the Reconstruction and Stripping efficiency is computed on a separate sample). The efficiencies are given in Tables 9.4, 9.5 and 9.6, and the normalization factors defined in Eq. (8.1) read for each year

$$\begin{aligned} f_{\text{norm}}^{2016} &= 1.19 \pm 0.04_{\text{stat}} \pm 0.11_{\text{syst}} \,, \\ f_{\text{norm}}^{2017} &= 1.18 \pm 0.02_{\text{stat}} \pm 0.14_{\text{syst}} \,, \\ f_{\text{norm}}^{2018} &= 1.17 \pm 0.02_{\text{stat}} \pm 0.13_{\text{syst}} \,. \end{aligned} \tag{9.1}$$

Selection Step	Efficiency [%]		Ratio of Efficiencies
	$B_s^0 \to \mu^+\mu^-\gamma$	$B_s^0 \to J/\psi\,\eta$	
Acceptance	22.619 ± 0.093	16.294 ± 0.051	$1.388 \pm 0.007_{\text{stat}}$
Reco & Strip	$11.19 \pm 0.03 \pm 0.16$	$3.56 \pm 0.03 \pm 0.08$	$3.14 \pm 0.03_{\text{stat}} \pm 0.09_{\text{syst}}$
Charged PID	$85.84 \pm 0.15 \pm 0.15$	$90.37 \pm 0.41 \pm 0.15$	$0.950 \pm 0.005_{\text{stat}} \pm 0.003_{\text{syst}}$
Neutral PID	98.1 ± 2.2	96.3 ± 2.2	$1.02 \pm 0.03_{\text{stat}} \pm 0.05_{\text{syst}}$
MLPS	75.00 ± 0.13	70.96 ± 0.36	$1.057 \pm 0.006_{\text{stat}} \pm 0018_{\text{syst}}$
Trigger	73.53 ± 0.15	86.50 ± 0.47	$0.850 \pm 0.005_{\text{stat}} \pm 0.023_{\text{syst}}$
MLP	30.00 ± 0.10	95.84 ± 0.49	$0.313 \pm 0.002_{\text{stat}} \pm 0021_{\text{syst}}$
Total			$1.19 \pm 0.04_{\text{stat}} \pm 0.11_{\text{syst}}$

Table 9.4: Summary of the cuts efficiencies in 2016. The Neutral PID ratio is the fraction of MC and data-driven efficiencies, see section 5.4 for details.

We also computed the normalization factor in each q^2 bins:

$$\begin{aligned} f_{\text{norm, Bin I}}^{2016} &= 1.40 \pm 0.05_{\text{stat}} \pm 0.11_{\text{syst}} \,, \\ f_{\text{norm, Bin I}}^{2017} &= 1.37 \pm 0.03_{\text{stat}} \pm 0.16_{\text{syst}} \,, \\ f_{\text{norm, Bin I}}^{2018} &= 1.31 \pm 0.03_{\text{stat}} \pm 0.15_{\text{syst}} \,, \end{aligned} \tag{9.2}$$

Selection Step	Efficiency [%]		Ratio of Efficiencies
	$B_s^0 \to \mu^+\mu^-\gamma$	$B_s^0 \to J/\psi\,\eta$	
Acceptance	22.565 ± 0.093	16.294 ± 0.051	$1.384 \pm 0.007_{\text{stat}}$
Reco & Strip	$11.22 \pm 0.03 \pm 0.08$	$3.56 \pm 0.03 \pm 0.08$	$3.15 \pm 0.03_{\text{stat}} \pm 0.08_{\text{syst}}$
Charged PID	$85.72 \pm 0.12 \pm 0.15$	$90.30 \pm 0.41 \pm 0.15$	$0.949 \pm 0.005_{\text{stat}} \pm 0.002_{\text{syst}}$
Neutral PID	98.9 ± 1.0	98.8 ± 1.0	$1.00 \pm 0.01_{\text{stat}} \pm 0.01_{\text{syst}}$
MLPS	75.00 ± 0.10	70.96 ± 0.36	$1.057 \pm 0.006_{\text{stat}} \pm 0016_{\text{syst}}$
Trigger	74.59 ± 0.11	86.50 ± 0.47	$0.862 \pm 0.005_{\text{stat}} \pm 0.023_{\text{syst}}$
MLP	30.00 ± 0.07	95.84 ± 0.49	$0.313 \pm 0.002_{\text{stat}} \pm 0036_{\text{syst}}$
Total			$1.18 \pm 0.02_{\text{stat}} \pm 0.14_{\text{syst}}$

Table 9.5: Summary of the cuts efficiencies in 2017. The Neutral PID ratio is the fraction of MC and data-driven efficiencies, see section 5.4 for details.

Selection Step	Efficiency [%]		Ratio of Efficiencies
	$B_s^0 \to \mu^+\mu^-\gamma$	$B_s^0 \to J/\psi\,\eta$	
Acceptance	22.742 ± 0.093	16.294 ± 0.051	$1.396 \pm 0.007_{\text{stat}}$
Reco & Strip	$11.20 \pm 0.03 \pm 0.08$	$3.56 \pm 0.03 \pm 0.08$	$3.15 \pm 0.03_{\text{stat}} \pm 0.08_{\text{syst}}$
Charged PID	$88.83 \pm 0.23 \pm 0.15$	$91.73 \pm 0.41 \pm 0.15$	$0.968 \pm 0.005_{\text{stat}} \pm 0.002_{\text{syst}}$
Neutral PID	98.9 ± 0.8	98.6 ± 0.9	$1.00 \pm 0.01_{\text{stat}} \pm 0.01_{\text{syst}}$
MLPS	75.00 ± 0.09	70.96 ± 0.36	$1.057 \pm 0.006_{\text{stat}} \pm 0015_{\text{syst}}$
Trigger	71.46 ± 0.10	86.50 ± 0.47	$0.826 \pm 0.005_{\text{stat}} \pm 0.023_{\text{syst}}$
MLP	30.00 ± 0.06	95.84 ± 0.49	$0.313 \pm 0.002_{\text{stat}} \pm 0032_{\text{syst}}$
Total			$1.17 \pm 0.02_{\text{stat}} \pm 0.13_{\text{syst}}$

Table 9.6: Summary of the cuts efficiencies in 2018. The Neutral PID ratio is the fraction of MC and data-driven efficiencies, see section 5.4 for details.

$$
\begin{aligned}
f^{2016}_{\text{norm, Bin II}} &= 1.17 \pm 0.05_{\text{stat}} \pm 0.09_{\text{syst}}\,, \\
f^{2017}_{\text{norm, Bin II}} &= 1.23 \pm 0.04_{\text{stat}} \pm 0.10_{\text{syst}}\,, \\
f^{2018}_{\text{norm, Bin II}} &= 1.10 \pm 0.03_{\text{stat}} \pm 0.08_{\text{syst}}\,, \\
f^{2016}_{\text{norm, Bin III}} &= 0.49 \pm 0.02_{\text{stat}} \pm 0.03_{\text{syst}}\,,
\end{aligned}
\tag{9.3}
$$

$$
\begin{aligned}
f^{2017}_{\text{norm, Bin III}} &= 0.40 \pm 0.01_{\text{stat}} \pm 0.02_{\text{syst}} \,, \\
f^{2018}_{\text{norm, Bin III}} &= 0.42 \pm 0.01_{\text{stat}} \pm 0.02_{\text{syst}} \,.
\end{aligned} \tag{9.4}
$$

9.3 Expected sensitivity

Using Eq. 8.1, and the normalization factors given in Eq. 9.1, the expected sensitivity can be computed for each year. Defining the single event sensitivity as

$$
\alpha = \frac{\mathcal{B}_{\text{norm}}}{N_{\text{norm}}} \times f_{\text{norm}} \tag{9.5}
$$

we found

$$
\alpha^{2016} = (6.45 \pm 1.14_{\text{stat}} \pm 0.61_{\text{syst}}) \times 10^{-9} \,, \tag{9.6}
$$

$$
\alpha^{2017} = (6.39 \pm 1.12_{\text{stat}} \pm 0.78_{\text{syst}}) \times 10^{-9} \,, \tag{9.7}
$$

$$
\alpha^{2018} = (5.04 \pm 0.88_{\text{stat}} \pm 0.56_{\text{syst}}) \times 10^{-9} \,, \tag{9.8}
$$

where the changes in efficiency are mainly due to the luminosity of each year. By weighting the normalization factor with the luminosities, we get the following global and binned sensitivities

$$
\alpha = (1.96 \pm 0.34_{\text{stat}} \pm 0.21_{\text{syst}}) \times 10^{-9} \,, \tag{9.9}
$$

$$
\alpha_{\text{Bin I}} = (2.25 \pm 0.39_{\text{stat}} \pm 0.25_{\text{syst}}) \times 10^{-9} \,, \tag{9.10}
$$

$$
\alpha_{\text{Bin II}} = (1.93 \pm 0.34_{\text{stat}} \pm 0.15_{\text{syst}}) \times 10^{-9} \,, \tag{9.11}
$$

$$
\alpha_{\text{Bin III}} = (0.72 \pm 0.13_{\text{stat}} \pm 0.04_{\text{syst}}) \times 10^{-9} \,. \tag{9.12}
$$

Given these numbers, no significant excess is expected. The use of the CLs method will however allows, in a future work, to put a limit on $\mathcal{B}(B^0_s \to \mu^+\mu^-\gamma)$.

Summary and Outlook

In spite of the arguments pointing toward NP at the TeV scale, direct searches, mainly held at the LHC, have only succeeded in pushing back the limits of validity of the SM. On the other hand, indirect measures offer some interesting hints of BSM physics. In particular, flavour physics is one of the most convenient fields to test a wide range of NP models.

The first part of this thesis was devoted to theoretical aspects. Motivated by the current status of measurements in $b \to s$ and $b \to c$ transitions, we performed several studies aiming at providing new observables or new methods, interpreting the current experimental measurements in term of effective theories and at building NP models. Our findings can be summarize as follows.

In Chapter 1, we scrutinized the rare and radiative decays $B_{(s)} \to \ell\ell^{(\prime)}\gamma$. The role of the photon in the final state was found to be manifold. By lifting the helicity factor, this additional particle enlarges the branching ratio of these decays. It also provides sensitivity to several operators that don't play role in the corresponding non-radiative decays. $B_{(s)} \to \ell\ell^{(\prime)}\gamma$ are however challenging to predict and to measure. The main theoretical uncertainty on the decay branching ratio is due to the $B \to \gamma$ form-factors that encapsulate the non-perturbativity of the transition. Our conclusions were as follows:

1. The $B_{(s)} \to \ell\ell\gamma$ branching ratios are dominated by light meson resonances, that are currently known experimentally and theoretically at 15% accuracy.

2. The uncertainty on the form-factor results in a 10% uncertainty on the high-q^2 branching ratio. This uncertainty partially cancels in the ratio between two different leptonic final states which provide a stringent cross-check of $R_{K^{(*)}}$.

We then generalised this study to two suggestions for experimental measurements. An indirect measurement of the $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio can be performed by considering the pollution of this decay in the $B_s^0 \to \mu^+\mu^-$ sample. This method will offer a precious cross-check of the direct analysis by probing a kinematic range hardly accessible by the full reconstruction. Finally, the interest of adding a photon in the final state also extends to LFV decays where it leads to an enhancement of several branching ratios.

In Chapter 2, we discussed the implication of $b \to s$ and $b \to c$ anomalies on model building. Although both sets of anomalies are related by the $SU(2)$ symmetry of the SM, combined explanations are challenged by precision measurement of low-energy processes. Models based on leptoquarks are particularly appealing solutions to this issue because these shifts to low-energy observables are loop

suppressed. Focusing on $b \to s$ anomalies, we showed that an additional 'horizontal' symmetry acting on the second and third generations of fermions give another mechanism to achieve this suppression. Although the simplest model based on this symmetry is ruled out by charm meson mixing, a straight forward generalization accounts for all $b \to s$ anomalies while passing all other experimental constraints.

Chapter 3 addresses the interpretation of current experimental data in term of effective theories. These studies are performed below and above the electroweak scale and to take into account the running and mixing of the operators of the effective theories. Important conclusions are:

1. $b \to s$ transitions are still overall well explained in two scenarios, one with a shift to C_9 and one with $C_9 = -C_{10}$. Due to the latest measurements of $B_s^0 \to \mu^+\mu^-$, the latter scenario is now favoured by the data. These scenarios however predict $R_K \approx R_{K^*}$, resulting in a small tension in data where $R_K > R_{K^*}$ is now measured.

2. This tension suggests the presence of a lepton-universal contribution to C_9, that, if added to the fit, leads to a scenario with a pull of 6.5σ with respect to the SM.

3. This scenario gets a natural explanation above the EWSB. $R_{K^{(*)}}$ can be accounted for by the operators $[O_{LQ}^{(1,3)}]_{2223}$ while the universal $\mathcal{C}_9$ shift is obtained by the running of the effective operators $[O_{LQ}^{(1,3)}]_{3323}$ between the NP and the weak scale. As the latter also contributes to $b \to c$ anomalies, this scenario allows a combine explanation of both sets of discrepancies.

We then discussed a possible realisation of this scenario, namely a simplified model based on the U_1 leptoquark. To the best of our knowledge, this model satisfies all indirect constraints as well as direct searches. We also showed that usual UV completions of this model contains a Z' that can play the role of DM portal. The link between B anomalies and DM was not unexpected because both puzzles have similar energy scales. Indeed we also showed that a composite DM model based on a hypercolour symmetry could also be linked to the model of the previous chapter.

From these studies we also conclude that the multitude of new analyses using the full Run II data set, as well as the Belle-II data set, will bring very valuable information in the near future. These measurements will not only allow to clarify the picture of flavour in the SM but should also impact much larger topics, ranging from high energy searches to DM phenomenology. Several research avenues, suggested in these chapters are also worth investigating. High-p_T searches, indirect constraints or neutrino physics could for instance bring other valuable information for future models.

In the second part of this thesis, we presented the search for $B_s^0 \to \mu^+\mu^-\gamma$ decays performed at LHCb. This analysis, being both rare and radiative, is not straightforward for LHCb, for several reasons.

1. The average transverse momentum of the photon is small compare to 'classical' radiative analysis (such as those of $B_s^0 \to \phi\gamma$, $B^0 \to K^*\gamma$, $\Lambda_b^0 \to \Lambda\gamma$ decays). This implies lower reconstruction and trigger efficiencies, larger backgrounds and larger signal widths.

2. The signal window is polluted by hardly reducible peaking backgrounds due to $B^0 \to \mu^+\mu^-\pi^0$ decays.

3. The choice of a normalization channel is limited by the presence of muons and a photon in the final state. The choice of the $B_s^0 \to J/\psi\,\eta$ decay leads to a difference in kinematics which imposes particular care in the cancellation of systematic uncertainties.

The analysis steps are guided by these challenges. In Chapter 5 we defined a dedicated stripping line constrained by the cut on the photon transverse momentum. We defined and tested the charged and neutral particles identification requirements. We developed a new method to diminish the background of resolved π^0 and η, and finally built a first selection based on the multivariate analysis of geometric variables.

Chapter 6 is devoted to a more subtle selection of events. This second selection is mainly based on the isolation of final state particles with respect to the rest of the event. The cuts on the two multivariate analysis selections are optimized together to maximize the efficiency of the whole selection process.

In Chapter 7, we reviewed the backgrounds that pollute the signal window and presented the fit procedure applied to the final sample. The stringent selection reduces the background to its combinatorial component and to channels containing merged π^0.

Chapter 8 discussed the normalization and the control channels. These two channels allowed us to study the discrepancies between the data and the simulated samples. These discrepancies are partially corrected for by a weighting of the MC and the reminding deviations are accounted for as systematic uncertainties. The control channels also permitted a precise study of the trigger efficiency both by emulating the trigger configurations and with data-driven methods.

Finally, Chapter 9 collected all the systematic uncertainties of the analysis steps to predict the expected global sensitivity. The final fits and the corresponding limits on the $B_s^0 \to \mu^+\mu^-\gamma$ branching ratio will appear in future works of the LHCb collaboration.

Apart from the indirect measurement discussed in the first part, this analysis can be improved in a few directions. The consideration of Run I dataset and the addition of a specific trigger line as planned for Run III, will certainly increase the sensitivity. Normalizing the signal yield to $B_s^0 \to \phi\gamma$ or $B^0 \to K^{*0}\gamma$ events in addition to $B_s^0 \to J/\psi\,\eta$ ones will also drastically reduce the systematic uncertainty by lifting the uncertainty on the normalization branching ratio and by improving the cancellation of the other systematic uncertainties. Finally, a more refine study

of the discrepancies between data and MC may lead to a more efficient weighting and result in a better simulation of signal events.

References

[1] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, .

[2] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance*, *Nucl. Phys.* **26** (1961) 222–229.

[3] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 1*, .

[4] G. Zweig, *An SU(3) model for strong interaction symmetry and its breaking. Version 2*, pp. 22–101. 2, 1964.

[5] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, *Phys. Lett.* **8** (1964) 214–215.

[6] 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) 30–61, [1207.7235].

[7] 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) 1–29, [1207.7214].

[8] P. W. Anderson, *Plasmons, Gauge Invariance, and Mass*, *Phys. Rev.* **130** (1963) 439–442.

[9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323.

[10] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* **12** (1964) 132–133.

[11] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509.

[12] G. Guralnik, C. Hagen and T. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585–587.

[13] KAMIOKANDE-II collaboration, K. Hirata et al., *Experimental Study of the Atmospheric Neutrino Flux*, .

[14] R. Becker-Szendy et al., *The Electron-neutrino and muon-neutrino content of the atmospheric flux*, *Phys. Rev. D* **46** (1992) 3720–3724.

[15] KAMIOKANDE collaboration, Y. Fukuda et al., *Atmospheric muon-neutrino / electron-neutrino ratio in the multiGeV energy range*, *Phys. Lett. B* **335** (1994) 237–245.

[16] W. Allison et al., *Measurement of the atmospheric neutrino flavor composition in Soudan-2*, *Phys. Lett. B* **391** (1997) 491–500, [hep-ex/9611007].

[17] SUPER-KAMIOKANDE collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567, [hep-ex/9807003].

[18] SUPER-KAMIOKANDE collaboration, K. Abe et al., *A Measurement of atmospheric neutrino flux consistent with tau neutrino appearance*, *Phys. Rev. Lett.* **97** (2006) 171801, [hep-ex/0607059].

[19] KAMIOKANDE collaboration, Y. Fukuda et al., *Solar neutrino data covering solar cycle 22*, *Phys. Rev. Lett.* **77** (1996) 1683–1686.

[20] B. Cleveland, T. Daily, J. Davis, Raymond, J. R. Distel, K. Lande, C. Lee et al., *Measurement of the solar electron neutrino flux with the Homestake chlorine detector*, *Astrophys. J.* **496** (1998) 505–526.

[21] SAGE collaboration, J. Abdurashitov et al., *Measurement of the solar neutrino capture rate with gallium metal*, *Phys. Rev. C* **60** (1999) 055801, [astro-ph/9907113].

[22] GNO collaboration, M. Altmann et al., *GNO solar neutrino observations: Results for GNO I*, *Phys. Lett. B* **490** (2000) 16–26, [hep-ex/0006034].

[23] SNO collaboration, Q. Ahmad et al., *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001) 071301, [nucl-ex/0106015].

[24] SAGE collaboration, V. Gavrin, *Solar neutrino results from SAGE*, *Nucl. Phys. B Proc. Suppl.* **91** (2001) 36–43.

[25] SNO collaboration, S. Ahmed et al., *Measurement of the total active B-8 solar neutrino flux at the Sudbury Neutrino Observatory with enhanced neutral current sensitivity*, *Phys. Rev. Lett.* **92** (2004) 181301, [nucl-ex/0309004].

[26] SUPER-KAMIOKANDE collaboration, J. Cravens et al., *Solar neutrino measurements in Super-Kamiokande-II*, *Phys. Rev. D* **78** (2008) 032002, [0803.4312].

[27] Borexino collaboration, G. Bellini et al., *Measurement of the solar 8B neutrino rate with a liquid scintillator target and 3 MeV energy threshold in the Borexino detector*, *Phys. Rev. D* **82** (2010) 033006, [0808.2868].

[28] KamLAND collaboration, K. Eguchi et al., *First results from KamLAND: Evidence for reactor anti-neutrino disappearance*, *Phys. Rev. Lett.* **90** (2003) 021802, [hep-ex/0212021].

[29] Double Chooz collaboration, Y. Abe et al., *Indication of Reactor $\bar{\nu}_e$ Disappearance in the Double Chooz Experiment*, *Phys. Rev. Lett.* **108** (2012) 131801, [1112.6353].

[30] Daya Bay collaboration, R. Leitner, *Recent results of Daya Bay reactor neutrino experiment*, *Nucl. Part. Phys. Proc.* **285-286** (2017) 32–37.

[31] RENO collaboration, J. Ahn et al., *Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment*, *Phys. Rev. Lett.* **108** (2012) 191802, [1204.0626].

[32] K2K collaboration, S. Ahn et al., *Detection of accelerator produced neutrinos at a distance of 250-km*, *Phys. Lett. B* **511** (2001) 178–184, [hep-ex/0103001].

[33] MINOS collaboration, D. Michael et al., *Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam*, *Phys. Rev. Lett.* **97** (2006) 191801, [hep-ex/0607088].

[34] T2K collaboration, K. Abe et al., *Indication of Electron Neutrino Appearance from an Accelerator-produced Off-axis Muon Neutrino Beam*, *Phys. Rev. Lett.* **107** (2011) 041801, [1106.2822].

[35] E. Otten and C. Weinheimer, *Neutrino mass limit from tritium beta decay*, *Rept. Prog. Phys.* **71** (2008) 086201, [0909.2104].

[36] G. Bertone and D. Hooper, *History of dark matter*, *Rev. Mod. Phys.* **90** (2018) 045002, [1605.04909].

[37] A. Del Popolo, *Dark matter and structure formation a review*, *Astron. Rep.* **51** (2007) 169–196, [0801.1091].

[38] L. Susskind, *Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory*, *Phys. Rev. D* **20** (1979) 2619–2625.

[39] G. F. Giudice, *Naturally Speaking: The Naturalness Criterion and Physics at the LHC*, 0801.2562.

[40] Particle Data Group collaboration, P. Zyla et al., *Review of Particle Physics*, *Prog. Theor. Exp. Phys.* **2020** (2020) .

[41] S. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, *Phys. Rev. D* **2** (1970) 1285–1292.

[42] L. Maiani, *The GIM Mechanism: origin, predictions and recent uses*, in *48th Rencontres de Moriond on Electroweak Interactions and Unified Theories*, pp. 3–16, 2013, 1303.6154.

[43] E598 collaboration, J. Aubert et al., *Experimental Observation of a Heavy Particle J*, *Phys. Rev. Lett.* **33** (1974) 1404–1406.

[44] SLAC-SP-017 collaboration, J. Augustin et al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, *Phys. Rev. Lett.* **33** (1974) 1406–1408.

[45] UTfit collaboration, M. Bona et al., *Model-independent constraints on $\Delta F = 2$ operators and the scale of new physics*, *JHEP* **03** (2008) 049, [0707.0636].

[46] L. Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, *Phys. Rev. Lett.* **51** (1983) 1945.

[47] CKMfitter group collaboration, J. Charles et al., *Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ new physics*, *Phys. Rev.* **D91** (2015) 073007, [1501.05013].

[48] LHCb collaboration, R. Aaij et al., *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **113** (2014) 151601 LHCb-PAPER-2014-024, CERN-PH-EP-2014-140, [1406.6482].

[49] G. Isidori, S. Nabeebaccus and R. Zwicky, *QED Corrections in $\bar{B} \to \bar{K}\ell^+\ell^-$ at the Double-Differential Level*, 2009.00929.

[50] M. Bordone, G. Isidori and A. Pattori, *On the Standard Model predictions for R_K and R_{K^*}*, *Eur. Phys. J.* **C76** (2016) 440, [1605.07633].

[51] LHCb collaboration, R. Aaij et al., *Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays*, *Phys. Rev. Lett.* **122** (2019) 191801 LHCb-PAPER-2019-009 CERN-EP-2019-043, [1903.09252].

[52] LHCb collaboration, R. Aaij et al., *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, *JHEP* **08** (2017) 055 LHCb-PAPER-2017-013, CERN-EP-2017-100, [1705.05802].

[53] Belle collaboration, A. Abdesselam et al., *Test of lepton flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle*, 1904.02440.

[54] LHCb collaboration, R. Aaij et al., *Differential branching fraction and angular analysis of the $B^+ \to K^+\mu^+\mu^-$ decay*, *JHEP* **02** (2013) 105 LHCb-PAPER-2012-024, CERN-PH-EP-2012-263, [1209.4284].

[55] LHCb collaboration, R. Aaij et al., *Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays*, *JHEP* **06** (2014) 133 LHCb-PAPER-2014-006, CERN-PH-EP-2014-055, [1403.8044].

[56] C. Bobeth, G. Hiller and D. van Dyk, *More Benefits of Semileptonic Rare B Decays at Low Recoil: CP Violation*, *JHEP* **07** (2011) 067, [1105.0376].

[57] C. Bobeth, G. Hiller, D. van Dyk and C. Wacker, *The Decay $B \to K\ell^+\ell^-$ at Low Hadronic Recoil and Model-Independent $\Delta B = 1$ Constraints*, *JHEP* **01** (2012) 107, [1111.2558].

[58] C. Bobeth, G. Hiller and D. van Dyk, *General analysis of $\bar{B} \to \bar{K}^{(*)}\ell^+\ell^-$ decays at low recoil*, *Phys. Rev.* **D87** (2013) 034016, [1212.2321].

[59] LHCb collaboration, R. Aaij et al., *Differential branching fraction and angular analysis of the decay $B_s^0 \to \phi\mu^+\mu^-$*, *JHEP* **07** (2013) 084 CERN-PH-EP-2013-078, LHCb-PAPER-2013-017, [1305.2168].

[60] LHCb collaboration, R. Aaij et al., *Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$*, *JHEP* **09** (2015) 179 LHCb-PAPER-2015-023, CERN-PH-EP-2015-145, [1506.08777].

[61] A. Bharucha, D. M. Straub and R. Zwicky, *$B \to V\ell^+\ell^-$ in the Standard Model from Light-Cone Sum Rules*, *JHEP* **08** (2015) 098, [1503.05534].

[62] R. R. Horgan, Z. Liu, S. Meinel and M. Wingate, *Rare B decays using lattice QCD form factors*, *PoS* **LATTICE2014** (2015) 372, [1501.00367].

[63] N. Gubernari, A. Kokulu and D. van Dyk, *$B \to P$ and $B \to V$ Form Factors from B-Meson Light-Cone Sum Rules beyond Leading Twist*, *JHEP* **01** (2019) 150, [1811.00983].

[64] S. Descotes-Genon, T. Hurth, J. Matias and J. Virto, *Optimizing the basis of $B \to K^*\ell\ell$ observables in the full kinematic range*, *JHEP* **05** (2013) 137, [1303.5794].

[65] LHCb collaboration, R. Aaij et al., *Measurement of form-factor-independent observables in the decay $B^0 \to K^{*0}\mu^+\mu^-$*, *Phys. Rev. Lett.* **111** (2013) 191801 LHCb-PAPER-2013-037, CERN-PH-EP-2013-146, [1308.1707].

[66] LHCb collaboration, R. Aaij et al., *Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using $3\,fb^{-1}$ of integrated luminosity*, *JHEP* **02** (2016) 104 LHCb-PAPER-2015-051, CERN-PH-EP-2015-314, [1512.04442].

[67] 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 LHCb-PAPER-2020-002, CERN-EP-2020-027, [2003.04831].

[68] ATLAS collaboration, A. collaboration, *Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2017-023, CERN, Geneva, Apr, 2017.

[69] CMS collaboration, C. collaboration, *Measurement of the P_1 and P_5' angular parameters of the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV*, Tech. Rep. CMS-PAS-BPH-15-008, CERN, Geneva, 2017.

[70] Belle collaboration, A. Abdesselam et al., *Angular analysis of $B^0 \to K^*(892)^0\ell^+\ell^-$*, in *Proceedings, LHCSki 2016 - A First Discussion of 13 TeV Results: Obergurgl, Austria, April 10-15, 2016*, 2016, 1604.04042, http://inspirehep.net/record/1446979/files/arXiv:1604.04042.pdf.

[71] A. Khodjamirian, T. Mannel, A. A. Pivovarov and Y. M. Wang, *Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$*, *JHEP* **09** (2010) 089, [1006.4945].

[72] S. Descotes-Genon, J. Matias and J. Virto, *Understanding the $B \to K^*\mu^+\mu^-$ Anomaly*, *Phys. Rev.* **D88** (2013) 074002, [1307.5683].

[73] J. Lyon and R. Zwicky, *Resonances gone topsy turvy - the charm of QCD or new physics in $b \to s\ell^+\ell^-$?*, 1406.0566.

[74] S. Jäger and J. Martin Camalich, *Reassessing the discovery potential of the $B \to K^*\ell^+\ell^-$ decays in the large-recoil region: SM challenges and BSM opportunities*, *Phys. Rev.* **D93** (2016) 014028, [1412.3183].

[75] M. Ciuchini, M. Fedele, E. Franco, S. Mishima, A. Paul, L. Silvestrini et al., *$B \to K^*\ell^+\ell^-$ decays at large recoil in the Standard Model: a theoretical reappraisal*, *JHEP* **06** (2015) 116, [1512.07157].

[76] BaBar collaboration, J. P. Lees et al., *Evidence for an excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ decays*, *Phys. Rev. Lett.* **109** (2012) 101802, [1205.5442].

[77] BaBar collaboration, J. P. Lees et al., *Measurement of an Excess of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons*, *Phys. Rev.* **D88** (2013) 072012, [1303.0571].

[78] Belle collaboration, M. Huschle et al., *Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, *Phys. Rev.* **D92** (2015) 072014, [1507.03233].

[79] Belle collaboration, Y. Sato et al., *Measurement of the branching ratio of $\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau$ relative to $\bar{B}^0 \to D^{*+}\ell^-\bar{\nu}_\ell$ decays with a semileptonic tagging method*, *Phys. Rev.* **D94** (2016) 072007, [1607.07923].

[80] Belle collaboration, S. Hirose et al., *Measurement of the τ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^* \tau^- \bar{\nu}_\tau$*, *Phys. Rev. Lett.* **118** (2017) 211801, [1612.00529].

[81] LHCb collaboration, R. Aaij et al., *Measurement of the ratio of branching fractions $\mathcal{B}(\overline{B}^0 \to D^{*+} \tau^- \overline{\nu}_\tau)/\mathcal{B}(\overline{B}^0 \to D^{*+} \mu^- \overline{\nu}_\mu)$*, *Phys. Rev. Lett.* **115** (2015) 111803 LHCb-PAPER-2015-025, CERN-PH-EP-2015-150, [1506.08614].

[82] LHCb collaboration, R. Aaij et al., *Measurement of the ratio of the $\mathcal{B}(B^0 \to D^{*-} \tau^+ \nu_\tau)$ and $\mathcal{B}(B^0 \to D^{*-} \mu^+ \nu_\mu)$ branching fractions using three-prong τ-lepton decays*, *Phys. Rev. Lett.* **120** (2018) 171802 LHCb-PAPER-2017-017, CERN-EP-2017-212, [1708.08856].

[83] S. Fajfer, J. F. Kamenik and I. Nisandzic, *On the $B \to D^* \tau \bar{\nu}_\tau$ Sensitivity to New Physics*, *Phys. Rev.* **D85** (2012) 094025, [1203.2654].

[84] J. F. Kamenik and F. Mescia, *$B \to D$ tau nu Branching Ratios: Opportunity for Lattice QCD and Hadron Colliders*, *Phys. Rev.* **D78** (2008) 014003, [0802.3790].

[85] MILC collaboration, J. A. Bailey et al., *$B \to D \ell \nu$ form factors at nonzero recoil and $|V_{cb}|$ from 2+1-flavor lattice QCD*, *Phys. Rev.* **D92** (2015) 034506, [1503.07237].

[86] HPQCD collaboration, H. Na, C. M. Bouchard, G. P. Lepage, C. Monahan and J. Shigemitsu, *$B \to D l \nu$ form factors at nonzero recoil and extraction of $|V_{cb}|$*, *Phys. Rev.* **D92** (2015) 054510, [1505.03925].

[87] F. U. Bernlochner, Z. Ligeti, M. Papucci and D. J. Robinson, *Combined analysis of semileptonic B decays to D and D^*: $R(D^{(*)})$, $|V_{cb}|$, and new physics*, *Phys. Rev.* **D95** (2017) 115008, [1703.05330].

[88] D. Bigi, P. Gambino and S. Schacht, *$R(D^*)$, $|V_{cb}|$, and the Heavy Quark Symmetry relations between form factors*, *JHEP* **11** (2017) 061, [1707.09509].

[89] Belle collaboration, A. Abdesselam et al., *Precise determination of the CKM matrix element $|V_{cb}|$ with $\bar{B}^0 \to D^{*+} \ell^- \bar{\nu}_\ell$ decays with hadronic tagging at Belle*, 1702.01521.

[90] Belle collaboration, A. Abdesselam et al., *Measurement of CKM Matrix Element $|V_{cb}|$ from $\bar{B} \to D^{*+} \ell^- \bar{\nu}_\ell$*, 1809.03290.

[91] M. Jung and D. M. Straub, *Constraining new physics in $b \to c \ell \nu$ transitions*, *JHEP* **01** (2019) 009, [1801.01112].

[92] W. Altmannshofer, C. Niehoff, P. Stangl and D. M. Straub, *Status of the $B \to K^* \mu^+ \mu^-$ anomaly after Moriond 2017*, *Eur. Phys. J.* **C77** (2017) 377, [1703.09189].

[93] W. Altmannshofer, P. Stangl and D. M. Straub, *Interpreting Hints for Lepton Flavor Universality Violation*, *Phys. Rev.* **D96** (2017) 055008, [1704.05435].

[94] B. Capdevila, A. Crivellin, S. Descotes-Genon, J. Matias and J. Virto, *Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data*, *JHEP* **01** (2018) 093, [1704.05340].

[95] L.-S. Geng, B. Grinstein, S. Jäger, J. Martin Camalich, X.-L. Ren and R.-X. Shi, *Towards the discovery of new physics with lepton-universality ratios of $b \to s\ell\ell$ decays*, *Phys. Rev.* **D96** (2017) 093006, [1704.05446].

[96] M. Ciuchini, A. M. Coutinho, M. Fedele, E. Franco, A. Paul, L. Silvestrini et al., *On Flavourful Easter eggs for New Physics hunger and Lepton Flavour Universality violation*, *Eur. Phys. J.* **C77** (2017) 688, [1704.05447].

[97] T. Hurth, F. Mahmoudi, D. Martinez Santos and S. Neshatpour, *Lepton nonuniversality in exclusive $b \to s\ell\ell$ decays*, *Phys. Rev.* **D96** (2017) 095034, [1705.06274].

[98] R. Alonso, B. Grinstein and J. Martin Camalich, *Lepton universality violation and lepton flavor conservation in B-meson decays*, *JHEP* **10** (2015) 184, [1505.05164].

[99] A. Greljo, G. Isidori and D. Marzocca, *On the breaking of Lepton Flavor Universality in B decays*, *JHEP* **07** (2015) 142, [1506.01705].

[100] R. Barbieri, G. Isidori, A. Pattori and F. Senia, *Anomalies in B-decays and $U(2)$ flavour symmetry*, *Eur. Phys. J.* **C76** (2016) 67, [1512.01560].

[101] L. Calibbi, A. Crivellin and T. Ota, *Effective Field Theory Approach to $b \to s\ell\ell^{(\prime)}$, $B \to K^{(*)}\nu\bar{\nu}$ and $B \to D^{(*)}\tau\nu$ with Third Generation Couplings*, *Phys. Rev. Lett.* **115** (2015) 181801, [1506.02661].

[102] M. Bauer and M. Neubert, *One Leptoquark to Rule Them All: A Minimal Explanation for $R_{D^{(*)}}$, R_K and $(g-2)_\mu$*, *Phys. Rev. Lett.* **116** (2016) 141802, [1511.01900].

[103] S. Fajfer and N. Košnik, *Vector leptoquark resolution of R_K and $R_{D^{(*)}}$ puzzles*, *Phys. Lett.* **B755** (2016) 270–274, [1511.06024].

[104] BaBar collaboration, B. Aubert et al., *Search for the decays $B^0 \to e^+e^-\gamma$ and $B^0 \to \mu^+\mu^-\gamma$*, *Phys. Rev.* **D77** (2008) 011104, [0706.2870].

[105] M. Beneke, C. Bobeth and R. Szafron, *Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+\mu^-$*, *JHEP* **10** (2019) 232, [1908.07011].

[106] G. Buchalla, A. J. Buras and M. E. Lautenbacher, *Weak decays beyond leading logarithms*, *Rev. Mod. Phys.* **68** (1996) 1125–1144, [hep-ph/9512380].

[107] A. J. Buras and M. Munz, *Effective Hamiltonian for $B \to X(s)e^+e^-$ beyond leading logarithms in the NDR and HV schemes*, *Phys. Rev.* **D52** (1995) 186–195, [hep-ph/9501281].

[108] M. Misiak, *The $b \to se^+e^-$ and $b \to s\gamma$ decays with next-to-leading logarithmic QCD corrections*, *Nucl. Phys.* **B393** (1993) 23–45.

[109] C. Bobeth, T. Ewerth, F. Krüger and J. Urban, *Analysis of neutral Higgs boson contributions to the decays $\bar{B}(\ s^{)} \to \ell^+\ell^-$ and $\bar{B} \to K\ell^+\ell^-$*, *Phys. Rev.* **D64** (2001) 074014, [hep-ph/0104284].

[110] K. G. Chetyrkin, M. Misiak and M. Munz, *Weak radiative B meson decay beyond leading logarithms*, *Phys. Lett. B* **400** (1997) 206–219, [hep-ph/9612313].

[111] C. Bobeth, M. Misiak and J. Urban, *Photonic penguins at two loops and m_t dependence of $BR[B \to X_s l^+ l^-]$*, *Nucl. Phys. B* **574** (2000) 291–330, [hep-ph/9910220].

[112] F. Sala and D. M. Straub, *A New Light Particle in B Decays?*, *Phys. Lett.* **B774** (2017) 205–209, [1704.06188].

[113] D. Ghosh, *Explaining the R_K and R_{K^*} anomalies*, *Eur. Phys. J.* **C77** (2017) 694, [1704.06240].

[114] A. Datta, J. Kumar, J. Liao and D. Marfatia, *New light mediators for the R_K and R_{K^*} puzzles*, *Phys. Rev.* **D97** (2018) 115038, [1705.08423].

[115] W. Altmannshofer, M. J. Baker, S. Gori, R. Harnik, M. Pospelov, E. Stamou et al., *Light resonances and the low-q^2 bin of R_{K^*}*, *JHEP* **03** (2018) 188, [1711.07494].

[116] C. Bobeth, M. Gorbahn, T. Hermann, M. Misiak, E. Stamou and M. Steinhauser, *$B_{s,d} \to \ell^+\ell^-$ in the Standard Model with Reduced Theoretical Uncertainty*, *Phys. Rev. Lett.* **112** (2014) 101801, [1311.0903].

[117] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, A. Pellegrino et al., *Probing New Physics via the $B_s^0 \to \mu^+\mu^-$ Effective Lifetime*, *Phys. Rev. Lett.* **109** (2012) 041801, [1204.1737].

[118] M. Beneke, C. Bobeth and R. Szafron, *Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \to \mu^+\mu^-$*, *Phys. Rev. Lett.* **120** (2018) 011801, [1708.09152].

[119] Flavour Lattice Averaging Group collaboration, S. Aoki et al., *FLAG Review 2019: Flavour Lattice Averaging Group (FLAG)*, *Eur. Phys. J. C* **80** (2020) 113, [1902.08191].

[120] LHCb collaboration, *Combination of the ATLAS, CMS and LHCb results on the $B^0_{(s)} \to \mu^+\mu^-$ decays*, .

[121] J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl and D. M. Straub, *B-decay discrepancies after Moriond 2019*, *Eur. Phys. J. C* **80** (2020) 252, [1903.10434].

[122] R. Alonso, B. Grinstein and J. Martin Camalich, *$SU(2) \times U(1)$ gauge invariance and the shape of new physics in rare B decays*, *Phys. Rev. Lett.* **113** (2014) 241802, [1407.7044].

[123] W. Altmannshofer and D. M. Straub, *New physics in $b \to s$ transitions after LHC run 1*, *Eur. Phys. J.* **C75** (2015) 382, [1411.3161].

[124] D. Guadagnoli, M. Reboud and R. Zwicky, *$B_s^0 \to \ell^+ \ell^- \gamma$ as a test of lepton flavor universality*, *JHEP* **11** (2017) 184, [1708.02649].

[125] F. Kruger and D. Melikhov, *Gauge invariance and form-factors for the decay $B \to \gamma \ell^+ \ell^-$*, *Phys. Rev.* **D67** (2003) 034002, [hep-ph/0208256].

[126] D. Melikhov and N. Nikitin, *Rare radiative leptonic decays $B^0_{d,s} \to \ell^+\ell^-\gamma$*, *Phys. Rev.* **D70** (2004) 114028, [hep-ph/0410146].

[127] S. Descotes-Genon and C. T. Sachrajda, *Factorization, the light cone distribution amplitude of the B meson and the radiative decay $B \to \gamma \ell \nu_\ell$*, *Nucl. Phys.* **B650** (2003) 356–390, [hep-ph/0209216].

[128] P. Ball and E. Kou, *$B \to \gamma e \nu_e$ transitions from QCD sum rules on the light cone*, *JHEP* **04** (2003) 029, [hep-ph/0301135].

[129] S. W. Bosch and G. Buchalla, *The Double radiative decays $B \to \gamma\gamma$ in the heavy quark limit*, *JHEP* **08** (2002) 054, [hep-ph/0208202].

[130] M. Dimou, J. Lyon and R. Zwicky, *Exclusive Chromomagnetism in heavy-to-light FCNCs*, *Phys. Rev.* **D87** (2013) 074008, [1212.2242].

[131] C. Q. Geng, C. C. Lih and W.-M. Zhang, *Study of $B_{s,d} \to \ell^+\ell^-\gamma$ decays*, *Phys. Rev.* **D62** (2000) 074017, [hep-ph/0007252].

[132] Y. Dincer and L. M. Sehgal, *Charge asymmetry and photon energy spectrum in the decay $B_s \to \ell^+\ell^-\gamma$*, *Phys. Lett.* **B521** (2001) 7–14, [hep-ph/0108144].

[133] S. Descotes-Genon and C. T. Sachrajda, *Universality of nonperturbative QCD effects in radiative B decays*, *Phys. Lett.* **B557** (2003) 213–223, [hep-ph/0212162].

[134] T. M. Aliev, A. Ozpineci and M. Savci, *$B_q \to \ell^+\ell^-\gamma$ decays in light cone QCD*, *Phys. Rev.* **D55** (1997) 7059–7066, [hep-ph/9611393].

[135] D. Yennie, S. C. Frautschi and H. Suura, *The infrared divergence phenomena and high-energy processes*, *Annals Phys.* **13** (1961) 379–452.

[136] S. Weinberg, *Infrared photons and gravitons*, *Phys. Rev.* **140** (1965) B516–B524.

[137] G. Isidori, *Soft-photon corrections in multi-body meson decays*, *Eur. Phys. J.* **C53** (2008) 567–571, [0709.2439].

[138] A. J. Buras, J. Girrbach, D. Guadagnoli and G. Isidori, *On the Standard Model prediction for $\mathcal{B}(B_{s,d} \to \mu^+\mu^-)$*, *Eur. Phys. J.* **C72** (2012) 2172, [1208.0934].

[139] HPQCD collaboration, G. C. Donald, C. T. H. Davies, J. Koponen and G. P. Lepage, *V_{cs} from $D_s \to \phi\ell\nu$ semileptonic decay and full lattice QCD*, *Phys. Rev.* **D90** (2014) 074506, [1311.6669].

[140] M. Bona, C. Alpigiani, A. Bevan, M. Ciuchini, D. Derkach, E. Franco et al., *Unitarity Triangle analysis in the Standard Model from the UTfit collaboration*, *PoS* **ICHEP2016** (2017) 554.

[141] P. Gambino, K. J. Healey and S. Turczyk, *Taming the higher power corrections in semileptonic B decays*, *Phys. Lett. B* **763** (2016) 60–65, [1606.06174].

[142] G. P. Korchemsky, D. Pirjol and T.-M. Yan, *Radiative leptonic decays of B mesons in QCD*, *Phys. Rev. D* **61** (2000) 114510, [hep-ph/9911427].

[143] E. Lunghi, D. Pirjol and D. Wyler, *Factorization in leptonic radiative B $\to \gamma e\nu$ decays*, *Nucl. Phys. B* **649** (2003) 349–364, [hep-ph/0210091].

[144] S. Bosch, R. Hill, B. Lange and M. Neubert, *Factorization and Sudakov resummation in leptonic radiative B decay*, *Phys. Rev. D* **67** (2003) 094014, [hep-ph/0301123].

[145] M. Beneke, C. Bobeth and Y.-M. Wang, *$B_{d,s} \to \gamma\ell\bar{\ell}$ decay with an energetic photon*, 2008.12494.

[146] LHCb collaboration, R. Aaij et al., *Search for the decays $B^0 \to J/\psi\gamma$ and $B_s^0 \to J/\psi\gamma$*, *Phys. Rev.* **D92** (2015) 112002 LHCb-PAPER-2015-044, CERN-PH-EP-2015-276, [1510.04866].

[147] D. Bečirević, S. Fajfer, N. Košnik and A. Smolkovič, *Enhanced CP asymmetries in $B \to K\mu^+\mu^-$*, 2008.09064.

[148] L. Calibbi, A. Crivellin, F. Kirk, C. A. Manzari and L. Vernazza, *Z' models with less-minimal flavour violation*, *Phys. Rev. D* **101** (2020) 095003, [1910.00014].

[149] C. Kane, C. Lehner, S. Meinel and A. Soni, *Radiative leptonic decays on the lattice*, *PoS* **LATTICE2019** (2019) 134, [1907.00279].

[150] P. Ball and R. Zwicky, *$B_{d,s} \to \rho, \omega, K^*, \phi$ decay form-factors from light-cone sum rules revisited*, *Phys. Rev.* **D71** (2005) 014029, [hep-ph/0412079].

[151] P. Ball, G. W. Jones and R. Zwicky, *$B \to V\gamma$ beyond QCD factorisation*, *Phys. Rev.* **D75** (2007) 054004, [hep-ph/0612081].

[152] M. Beneke, T. Feldmann and D. Seidel, *Systematic approach to exclusive $B \to V\ell^+\ell^-$, $V\gamma$ decays*, *Nucl. Phys.* **B612** (2001) 25–58, [hep-ph/0106067].

[153] S. W. Bosch and G. Buchalla, *The Radiative decays $B \to V\gamma$ at next-to-leading order in QCD*, *Nucl. Phys.* **B621** (2002) 459–478, [hep-ph/0106081].

[154] J. Lyon and R. Zwicky, *Isospin asymmetries in $B \to (K^*, \rho)\gamma/\ell^+\ell^-$ and $B \to K\ell^+\ell^-$ in and beyond the standard model*, *Phys. Rev.* **D88** (2013) 094004, [1305.4797].

[155] Particle Data Group collaboration, C. Patrignani et al., *Review of Particle Physics*, *Chin. Phys.* **C40** (2016) 100001.

[156] LHCb collaboration, R. Aaij et al., *Measurement of b-hadron fractions in 13 TeV pp collisions*, LHCb-PAPER-2018-050.

[157] F. Muheim, Y. Xie and R. Zwicky, *Exploiting the width difference in $B_s \to \phi\gamma$*, *Phys. Lett.* **B664** (2008) 174–179, [0802.0876].

[158] LHCb collaboration, R. Aaij et al., *Measurement of CP-violating and mixing-induced observables in $B_s^0 \to \phi\gamma$ decays*, *Phys. Rev. Lett.* **123** (2019) 081802 LHCb-PAPER-2019-015; CERN-EP-2019-077, [1905.06284].

[159] A. P. Bakulev and S. V. Mikhailov, *The rho meson and related meson wave functions in QCD sum rules with nonlocal condensates*, *Phys. Lett.* **B436** (1998) 351–362, [hep-ph/9803298].

[160] LHCb collaboration, R. Aaij et al., *Measurement of the phase difference between the short- and long-distance amplitudes in the $B^+ \to K^+\mu^+\mu^-$ decay*, *Eur. Phys. J.* **C77** (2017) 161 LHCb-PAPER-2016-045, CERN-EP-2016-301, [1612.06764].

[161] F. Kruger and L. M. Sehgal, *Lepton polarization in the decays $b \to X(s)\mu^+\mu^-$ and $B \to X(s)\tau^+\tau^-$*, *Phys. Lett.* **B380** (1996) 199–204, [hep-ph/9603237].

[162] BES collaboration, M. Ablikim et al., *Determination of the $\psi(3770)$, $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$ resonance parameters*, *eConf* **C070805** (2007) 02, [0705.4500].

[163] LHCb collaboration, R. Aaij et al., *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays*, *Phys. Rev. Lett.* **118** (2017) 191801 LHCb-PAPER-2017-001, CERN-EP-2017-041, [1703.05747].

[164] T. Blake, G. Lanfranchi and D. M. Straub, *Rare B Decays as Tests of the Standard Model*, *Prog. Part. Nucl. Phys.* **92** (2017) 50–91, [1606.00916].

[165] F. Dettori, D. Guadagnoli and M. Reboud, *$B^0_s \to \mu^+\mu^-\gamma$ from $B^0_s \to \mu^+\mu^-$*, *Phys. Lett. B* **768** (2017) 163–167, [1610.00629].

[166] CMS, LHCb collaboration, V. Khachatryan et al., *Observation of the rare $B^0_s \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data*, *Nature* **522** (2015) 68–72, [1411.4413].

[167] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, *Eur. Phys. J.* **C45** (2006) 97–107, [hep-ph/0506026].

[168] LHCb collaboration, R. Aaij et al., *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment*, *Phys. Rev. Lett.* **111** (2013) 101805 CERN-PH-EP-2013-128, LHCb-PAPER-2013-046, [1307.5024].

[169] CMS collaboration, S. Chatrchyan et al., *Measurement of the $B^0_s \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ with the CMS Experiment*, *Phys. Rev. Lett.* **111** (2013) 101804, [1307.5025].

[170] ATLAS collaboration, M. Aaboud et al., *Study of the rare decays of B^0_s and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 513, [1604.04263].

[171] J. Albrecht, E. Stamou, R. Ziegler and R. Zwicky, *Probing flavoured Axions in the Tail of $B_q \to \mu^+\mu^-$*, 1911.05018.

[172] S. L. Glashow, D. Guadagnoli and K. Lane, *Lepton Flavor Violation in B Decays?*, *Phys. Rev. Lett.* **114** (2015) 091801, [1411.0565].

[173] A. Crivellin, L. Hofer, J. Matias, U. Nierste, S. Pokorski and J. Rosiek, *Lepton-flavour violating B decays in generic Z′ models*, *Phys. Rev. D* **92** (2015) 054013, [1504.07928].

[174] D. Guadagnoli and K. Lane, *Charged-Lepton Mixing and Lepton Flavor Violation*, *Phys. Lett.* **B751** (2015) 54–58, [1507.01412].

[175] LHCb collaboration, R. Aaij *et al.*, and A. Bharucha et al., *Implications of LHCb measurements and future prospects*, *Eur. Phys. J.* **C73** (2013) 2373 LHCb-PUB-2012-006, LHCb-PAPER-2012-031, CERN-PH-EP-2012-334, [1208.3355].

[176] LHCb collaboration, R. Aaij et al., *Search for Lepton-Flavor Violating Decays $B^+ \to K^+\mu^\pm e^\mp$*, *Phys. Rev. Lett.* **123** (2019) 241802 CERN-EP-2019-172, LHCb-PAPER-2019-022, [1909.01010].

[177] LHCb collaboration, R. Aaij et al., *Search for the lepton flavour violating decay $B^+ \to K^+\mu^-\tau^+$ using B^{*0}_{s2} decays*, *JHEP* **06** (2020) 129, [2003.04352].

[178] BaBar collaboration, J. P. Lees et al., *A search for the decay modes $B^{+-} \to h^{+-}\tau^{+-}l$*, *Phys. Rev.* **D86** (2012) 012004, [1204.2852].

[179] LHCb collaboration, R. Aaij et al., *Search for the lepton-flavour-violating decays $B^0_s \to \tau^\pm\mu^\mp$ and $B^0 \to \tau^\pm\mu^\mp$*, *Phys. Rev. Lett.* **123** (2019) 211801 CERN-EP-2019-076, LHCb-PAPER-2019-016, [1905.06614].

[180] LHCb collaboration, R. Aaij et al., *Search for the lepton-flavour violating decays $B^0_{(s)} \to e^\pm\mu^\mp$*, *JHEP* **03** (2018) 078 LHCb-PAPER-2017-031, CERN-EP-2017-242, [1710.04111].

[181] LHCb collaboration, R. Aaij et al., *Search for the rare decays $B^0_s \to e^+e^-$ and $B^0 \to e^+e^-$*, *Phys. Rev. Lett.* **124** (2020) 211802 LHCb-PAPER-2020-001, CERN-EP-2020-023, [2003.03999].

[182] D. Guadagnoli, D. Melikhov and M. Reboud, *More Lepton Flavor Violating Observables for LHCb's Run 2*, *Phys. Lett.* **B760** (2016) 442–447, [1605.05718].

[183] S. M. Boucenna, J. W. F. Valle and A. Vicente, *Are the B decay anomalies related to neutrino oscillations?*, *Phys. Lett. B* **750** (2015) 367–371, [1503.07099].

[184] BNL collaboration, D. Ambrose et al., *New limit on muon and electron lepton number violation from $K_L \to \mu^{\pm} e^{\mp}$ decay*, *Phys. Rev. Lett.* **81** (1998) 5734–5737, [hep-ex/9811038].

[185] M. Borsato, V. V. Gligorov, D. Guadagnoli, D. Martinez Santos and O. Sumensari, *Effective-field-theory arguments for pursuing lepton-flavor-violating K decays at LHCb*, *Phys. Rev. D* **99** (2019) 055017, [1808.02006].

[186] B. Bhattacharya, A. Datta, D. London and S. Shivashankara, *Simultaneous Explanation of the R_K and $R_{D^{(*)}}$ Puzzles*, *Phys. Lett. B* **742** (2015) 370–374, [1412.7164].

[187] F. Feruglio, P. Paradisi and A. Pattori, *Revisiting Lepton Flavor Universality in B Decays*, *Phys. Rev. Lett.* **118** (2017) 011801, [1606.00524].

[188] A. Pich, *Precision Tau Physics*, *Prog. Part. Nucl. Phys.* **75** (2014) 41–85, [1310.7922].

[189] D. Buttazzo, A. Greljo, G. Isidori and D. Marzocca, *B-physics anomalies: a guide to combined explanations*, *JHEP* **11** (2017) 044, [1706.07808].

[190] S. M. Boucenna, A. Celis, J. Fuentes-Martin, A. Vicente and J. Virto, *Non-abelian gauge extensions for B-decay anomalies*, *Phys. Lett. B* **760** (2016) 214–219, [1604.03088].

[191] A. J. Buras, F. De Fazio and J. Girrbach, *331 models facing new $b \to s\mu^+\mu^-$ data*, *JHEP* **02** (2014) 112, [1311.6729].

[192] S. Descotes-Genon, M. Moscati and G. Ricciardi, *Nonminimal 331 model for lepton flavor universality violation in $b \to s\ell\ell$ decays*, *Phys. Rev. D* **98** (2018) 115030, [1711.03101].

[193] D. A. Faroughy, A. Greljo and J. F. Kamenik, *Confronting lepton flavor universality violation in B decays with high-p_T tau lepton searches at LHC*, *Phys. Lett.* **B764** (2017) 126–134, [1609.07138].

[194] I. Dorsner, S. Fajfer, A. Greljo, J. F. Kamenik and N. Kosnik, *Physics of leptoquarks in precision experiments and at particle colliders*, *Phys. Rept.* **641** (2016) 1–68, [1603.04993].

[195] H. Georgi and S. Glashow, *Unity of All Elementary Particle Forces*, *Phys. Rev. Lett.* **32** (1974) 438–441.

[196] J. C. Pati and A. Salam, *Lepton Number as the Fourth Color*, *Phys. Rev. D* **10** (1974) 275–289.

[197] D. Bečirević, N. Kosnik, O. Sumensari and R. Zukanovich Funchal, *Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to s\ell_1\ell_2$ modes*, *JHEP* **11** (2016) 035, [1608.07583].

[198] Y. Cai, J. Gargalionis, M. A. Schmidt and R. R. Volkas, *Reconsidering the One Leptoquark solution: flavor anomalies and neutrino mass*, *JHEP* **10** (2017) 047, [1704.05849].

[199] G. Hiller, D. Loose and K. Schönwald, *Leptoquark Flavor Patterns & B Decay Anomalies*, *JHEP* **12** (2016) 027, [1609.08895].

[200] B. Bhattacharya, A. Datta, J.-P. Guévin, D. London and R. Watanabe, *Simultaneous Explanation of the R_K and $R_{D^{(*)}}$ Puzzles: a Model Analysis*, *JHEP* **01** (2017) 015, [1609.09078].

[201] A. Angelescu, D. Bečirević, D. A. Faroughy and O. Sumensari, *Closing the window on single leptoquark solutions to the B-physics anomalies*, *JHEP* **10** (2018) 183, [1808.08179].

[202] J. Kumar, D. London and R. Watanabe, *Combined Explanations of the $b \to s\mu^+\mu^-$ and $b \to c\tau^-\bar{\nu}$ Anomalies: a General Model Analysis*, *Phys. Rev.* **D99** (2019) 015007, [1806.07403].

[203] B. Fornal, S. A. Gadam and B. Grinstein, *Left-Right SU(4) Vector Leptoquark Model for Flavor Anomalies*, *Phys. Rev.* **D99** (2019) 055025, [1812.01603].

[204] D. Marzocca, *Addressing the B-physics anomalies in a fundamental Composite Higgs Model*, *JHEP* **07** (2018) 121, [1803.10972].

[205] A. Crivellin, D. Müller and T. Ota, *Simultaneous explanation of $R_{D^{(*)}}$ and $b \to s\mu\mu$: the last scalar leptoquarks standing*, *JHEP* **09** (2017) 040, [1703.09226].

[206] A. Crivellin, D. Müller and F. Saturnino, *Flavor Phenomenology of the Leptoquark Singlet-Triplet Model*, *JHEP* **06** (2020) 020, [1912.04224].

[207] V. Gherardi, D. Marzocca and E. Venturini, *Low-energy phenomenology of scalar leptoquarks at one-loop accuracy*, 2008.09548.

[208] D. Bečirević, I. Doršner, S. Fajfer, N. Košnik, D. A. Faroughy and O. Sumensari, *Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies*, *Phys. Rev. D* **98** (2018) 055003, [1806.05689].

[209] F. Feruglio, P. Paradisi and A. Pattori, *On the Importance of Electroweak Corrections for B Anomalies*, *JHEP* **09** (2017) 061, [1705.00929].

[210] C. Cornella, F. Feruglio and P. Paradisi, *Low-energy Effects of Lepton Flavour Universality Violation*, *JHEP* **11** (2018) 012, [1803.00945].

[211] F. Feruglio, P. Paradisi and O. Sumensari, *Implications of scalar and tensor explanations of $R_{D^{(*)}}$*, *JHEP* **11** (2018) 191, [1806.10155].

[212] D. Guadagnoli, M. Reboud and O. Sumensari, *A gauged horizontal $SU(2)$ symmetry and $R_{K^{(*)}}$*, *JHEP* **11** (2018) 163, [1807.03285].

[213] R. N. Cahn and H. Harari, *Bounds on the Masses of Neutral Generation Changing Gauge Bosons*, *Nucl. Phys.* **B176** (1980) 135–152.

[214] G. Isidori, Y. Nir and G. Perez, *Flavor Physics Constraints for Physics Beyond the Standard Model*, *Ann. Rev. Nucl. Part. Sci.* **60** (2010) 355, [1002.0900].

[215] G. Hiller and M. Schmaltz, *Diagnosing lepton-nonuniversality in $b \to s\ell\ell$*, *JHEP* **02** (2015) 055, [1411.4773].

[216] D. B. Kaplan, *Flavor at SSC energies: A New mechanism for dynamically generated fermion masses*, *Nucl. Phys. B* **365** (1991) 259–278.

[217] W. Altmannshofer, S. Gori, M. Pospelov and I. Yavin, *Quark flavor transitions in $L_\mu - L_\tau$ models*, *Phys.Rev.* **D89** (2014) 095033, [1403.1269].

[218] S. D. Aristizabal, F. Staub and A. Vicente, *Shedding light on the $b \to s$ anomalies with a dark sector*, 1503.06077.

[219] A. Crivellin, G. D'Ambrosio and J. Heeck, *Explaining $h \to \mu^\pm\tau^\mp$, $B \to K^*\mu^+\mu^-$ and $B \to K\mu^+\mu^-/B \to Ke^+e^-$ in a two-Higgs-doublet model with gauged $L_\mu - L_\tau$*, *Phys.Rev.Lett.* **114** (2015) 151801, [1501.00993].

[220] G. D'Amico, M. Nardecchia, P. Panci, F. Sannino, A. Strumia, R. Torre et al., *Flavour anomalies after the R_{K^*} measurement*, *JHEP* **09** (2017) 010, [1704.05438].

[221] HFLAV collaboration, Y. S. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018*, 1909.12524.

[222] E. Golowich, J. Hewett, S. Pakvasa and A. A. Petrov, *Implications of D^0 - $\bar{D}^0$ Mixing for New Physics*, *Phys. Rev.* **D76** (2007) 095009, [0705.3650].

[223] D. Guadagnoli and R. N. Mohapatra, *TeV Scale Left Right Symmetry and Flavor Changing Neutral Higgs Effects*, *Phys. Lett.* **B694** (2011) 386–392, [1008.1074].

[224] A. F. Falk, Y. Grossman, Z. Ligeti, Y. Nir and A. A. Petrov, *The D0 - anti-D0 mass difference from a dispersion relation*, *Phys. Rev.* **D69** (2004) 114021, [hep-ph/0402204].

[225] V. A. Monich, B. V. Struminsky and G. G. Volkov, *Oscillation and CP Violation in the 'Horizontal' Superweak Gauge Scheme*, *Phys. Lett.* **104B** (1981) 382.

[226] BaBar collaboration, J. P. Lees et al., *Search for $B^+ \to K^+\tau^+\tau^-$ at the BaBar experiment*, *Phys. Rev. Lett.* **118** (2017) 031802, [1605.09637].

[227] LHCb collaboration, R. Aaij et al., *Search for the decays $B_s^0 \to \tau^+\tau^-$ and $B^0 \to \tau^+\tau^-$*, *Phys. Rev. Lett.* **118** (2017) 251802 LHCb-PAPER-2017-003, CERN-EP-2017-034, [1703.02508].

[228] D. Bečirević, O. Sumensari and R. Zukanovich Funchal, *Lepton flavor violation in exclusive $b \to s$ decays*, *Eur. Phys. J.* **C76** (2016) 134, [1602.00881].

[229] M. Bordone, C. Cornella, J. Fuentes-Martín and G. Isidori, *Low-energy signatures of the PS^3 model: from B-physics anomalies to LFV*, *JHEP* **10** (2018) 148, [1805.09328].

[230] E. Bertuzzo, Y. F. Perez G., O. Sumensari and R. Zukanovich Funchal, *Limits on Neutrinophilic Two-Higgs-Doublet Models from Flavor Physics*, *JHEP* **01** (2016) 018, [1510.04284].

[231] Belle collaboration, Y. Miyazaki et al., *Search for Lepton-Flavor-Violating tau Decays into a Lepton and a Vector Meson*, *Phys. Lett.* **B699** (2011) 251–257, [1101.0755].

[232] Belle-II collaboration, C. Schwanda, *Charged Lepton Flavour Violation at Belle and Belle II*, *Nucl. Phys. Proc. Suppl.* **248-250** (2014) 67–72.

[233] B. Golob, K. Trabelsi and P. Urquijo, *Impact of Belle II on flavour physics*, tech. rep., BELLE2-NOTE-0021, 2015.

[234] L. Di Luzio, M. Kirk and A. Lenz, *Updated B_s-mixing constraints on new physics models for $b \to s\ell^+\ell^-$ anomalies*, *Phys. Rev.* **D97** (2018) 095035, [1712.06572].

[235] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, *Eur. Phys. J.* **C77** (2017) 112, [1607.00299].

[236] UTfit collaboration, M. Bona et al., *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters from Lattice QCD: A Reappraisal after the Measurements of ΔM_s and $\mathcal{B}(B \to \tau\nu_\tau)$*, *JHEP* **10** (2006) 081, [hep-ph/0606167].

[237] CKMFitter Group collaboration, J. Charles, A. Hocker, H. Lacker, S. Laplace, F. Le Diberder, J. Malcles et al., *CP violation and the CKM matrix: Assessing the impact of the asymmetric B factories*, *Eur. Phys. J. C* **41** (2005) 1–131, [hep-ph/0406184].

[238] Particle Data Group collaboration, C. Patrignani et al., *Review of Particle Physics*, *Chin. Phys.* **C40** (2016) 100001.

[239] J. Brod and J. Zupan, *The ultimate theoretical error on γ from $B \to DK$ decays*, *JHEP* **01** (2014) 051, [1308.5663].

[240] LHCb collaboration, R. Aaij et al., *Measurement of the CKM angle γ using $B^{\pm} \to DK^{\pm}$ with $D \to K^0_S\pi^+\pi^-$, $K^0_S K^+K^-$ decays*, *JHEP* **08** (2018) 176 LHCb-PAPER-2018-017. CERN-EP-2018-135, [1806.01202].

[241] D. Craik, T. Gershon and A. Poluektov, *Optimising sensitivity to γ with $B^0 \to DK^+\pi^-$, $D \to K^0_S\pi^+\pi^-$ double Dalitz plot analysis*, *Phys. Rev.* **D97** (2018) 056002, [1712.07853].

[242] S. Descotes-Genon, M. Novoa-Brunet, S. Fajfer and J. F. Kamenik, *Implications of $b \to s\mu\mu$ Anomalies for Future Measurements of $B \to K^{(*)}\nu\bar{\nu}$ and $K \to \pi\nu\bar{\nu}$*, 2005.03734.

[243] A. J. Buras, J. Girrbach-Noe, C. Niehoff and D. M. Straub, *$B \to K^{(*)}\nu\overline{\nu}$ decays in the Standard Model and beyond*, *JHEP* **02** (2015) 184, [1409.4557].

[244] Belle collaboration, J. Grygier et al., *Search for $\boldsymbol{B \to h\nu\bar{\nu}}$ decays with semileptonic tagging at Belle*, *Phys. Rev.* **D96** (2017) 091101, [1702.03224].

[245] LHCb collaboration, R. Aaij et al., *Search for the rare decay $D^0 \to \mu^+\mu^-$*, *Phys. Lett.* **B725** (2013) 15 CERN-PH-EP-2013-083, LHCb-PAPER-2013-013, [1305.5059].

[246] BaBar collaboration, B. Aubert et al., *Searches for Lepton Flavor Violation in the Decays $\tau^{\pm} \to e^{\pm}\gamma$ and $\tau^{\pm} \to \mu^{\pm}\gamma$*, *Phys. Rev. Lett.* **104** (2010) 021802, [0908.2381].

[247] A. Greljo and D. Marzocca, *High-p_T dilepton tails and flavor physics*, *Eur. Phys. J.* **C77** (2017) 548, [1704.09015].

[248] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1–98, [1706.08945].

[249] M. Ciuchini, A. M. Coutinho, M. Fedele, E. Franco, A. Paul, L. Silvestrini et al., *New Physics in $b \to s\ell^+\ell^-$ confronts new data on Lepton Universality*, *Eur. Phys. J. C* **79** (2019) 719, [1903.09632].

[250] J. De Blas et al., *HEPfit: a code for the combination of indirect and direct constraints on high energy physics models*, *Eur. Phys. J. C* **80** (2020) 456, [1910.14012].

[251] W. Altmannshofer and D. M. Straub, *Cornering New Physics in $b \to s$ Transitions*, *JHEP* **08** (2012) 121, [1206.0273].

[252] F. Beaujean, C. Bobeth and D. van Dyk, *Comprehensive Bayesian analysis of rare (semi)leptonic and radiative B decays*, *Eur. Phys. J.* **C74** (2014) 2897, [1310.2478].

[253] O. Catà and M. Jung, *Signatures of a nonstandard Higgs boson from flavor physics*, *Phys. Rev.* **D92** (2015) 055018, [1505.05804].

[254] A. Celis, J. Fuentes-Martin, A. Vicente and J. Virto, *Gauge-invariant implications of the LHCb measurements on lepton-flavor nonuniversality*, *Phys. Rev.* **D96** (2017) 035026, [1704.05672].

[255] J. E. Camargo-Molina, A. Celis and D. A. Faroughy, *Anomalies in Bottom from new physics in Top*, *Phys. Lett.* **B784** (2018) 284–293, [1805.04917].

[256] J. Aebischer, J. Kumar, P. Stangl and D. M. Straub, *A Global Likelihood for Precision Constraints and Flavour Anomalies*, *Eur. Phys. J. C* **79** (2019) 509, [1810.07698].

[257] T. Hurth, S. Renner and W. Shepherd, *Matching for FCNC effects in the flavour-symmetric SMEFT*, *JHEP* **06** (2019) 029, [1903.00500].

[258] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond*, 1810.08132.

[259] J. Aebischer, J. Kumar and D. M. Straub, *Wilson: a Python package for the running and matching of Wilson coefficients above and below the electroweak scale*, *Eur. Phys. J.* **C78** (2018) 1026, [1804.05033].

[260] W. Altmannshofer, P. Paradisi and D. M. Straub, *Model-Independent Constraints on New Physics in $b \to s$ Transitions*, *JHEP* **04** (2012) 008, [1111.1257].

[261] W. Altmannshofer and D. M. Straub, *New physics in $B \to K^* \mu\mu$?*, *Eur. Phys. J.* **C73** (2013) 2646, [1308.1501].

[262] C. Bobeth, M. Chrzaszcz, D. van Dyk and J. Virto, *Long-distance effects in $B \to K^* \ell\ell$ from analyticity*, *Eur. Phys. J.* **C78** (2018) 451, [1707.07305].

[263] A. Paul and D. M. Straub, *Constraints on new physics from radiative B decays*, *JHEP* **04** (2017) 027, [1608.02556].

[264] C. Bobeth and U. Haisch, *New Physics in Γ_{12}^s: $(\bar{s}b)(\bar{\tau}\tau)$ Operators*, *Acta Phys. Polon.* **B44** (2013) 127–176, [1109.1826].

[265] W. Altmannshofer, C. Niehoff and D. M. Straub, *$B_s \to \mu^+\mu^-$ as current and future probe of new physics*, *JHEP* **05** (2017) 076, [1702.05498].

[266] M. Algueró, B. Capdevila, S. Descotes-Genon, P. Masjuan and J. Matias, *Are we overlooking lepton flavour universal new physics in $b \to s\ell\ell$?*, *Phys. Rev. D* **99** (2019) 075017, [1809.08447].

[267] S. Jäger and J. Martin Camalich, *On $B \to V\ell\ell$ at small dilepton invariant mass, power corrections, and new physics*, *JHEP* **05** (2013) 043, [1212.2263].

[268] E. E. Jenkins, A. V. Manohar and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators I: Formalism and lambda Dependence*, *JHEP* **10** (2013) 087, [1308.2627].

[269] E. E. Jenkins, A. V. Manohar and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators II: Yukawa Dependence*, *JHEP* **01** (2014) 035, [1310.4838].

[270] R. Alonso, E. E. Jenkins, A. V. Manohar and M. Trott, *Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology*, *JHEP* **04** (2014) 159, [1312.2014].

[271] J. Aebischer, M. Fael, C. Greub and J. Virto, *B physics Beyond the Standard Model at One Loop: Complete Renormalization Group Evolution below the Electroweak Scale*, *JHEP* **09** (2017) 158, [1704.06639].

[272] E. E. Jenkins, A. V. Manohar and P. Stoffer, *Low-Energy Effective Field Theory below the Electroweak Scale: Anomalous Dimensions*, *JHEP* **01** (2018) 084, [1711.05270].

[273] G. D'Ambrosio, G. F. Giudice, G. Isidori and A. Strumia, *Minimal flavor violation: An Effective field theory approach*, *Nucl. Phys.* **B645** (2002) 155–187, [hep-ph/0207036].

[274] J. Aebischer et al., *WCxf: an exchange format for Wilson coefficients beyond the Standard Model*, *Comput. Phys. Commun.* **232** (2018) 71–83, [1712.05298].

[275] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, *JHEP* **10** (2010) 085, [1008.4884].

[276] A. Crivellin, C. Greub, D. Müller and F. Saturnino, *Importance of Loop Effects in Explaining the Accumulated Evidence for New Physics in B Decays with a Vector Leptoquark*, *Phys. Rev. Lett.* **122** (2019) 011805, [1807.02068].

[277] S. Jäger, M. Kirk, A. Lenz and K. Leslie, *Charming new physics in rare B-decays and mixing?*, *Phys. Rev.* **D97** (2018) 015021, [1701.09183].

[278] Belle collaboration, A. Abdesselam et al., *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method*, 1904.08794.

[279] Belle collaboration, G. Caria et al., *Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method*, *Phys. Rev. Lett.* **124** (2020) 161803, [1910.05864].

[280] L. Di Luzio, A. Greljo and M. Nardecchia, *Gauge leptoquark as the origin of B-physics anomalies*, *Phys. Rev.* **D96** (2017) 115011, [1708.08450].

[281] N. Assad, B. Fornal and B. Grinstein, *Baryon Number and Lepton Universality Violation in Leptoquark and Diquark Models*, *Phys. Lett.* **B777** (2018) 324–331, [1708.06350].

[282] L. Calibbi, A. Crivellin and T. Li, *Model of vector leptoquarks in view of the B-physics anomalies*, *Phys. Rev.* **D98** (2018) 115002, [1709.00692].

[283] M. Bordone, C. Cornella, J. Fuentes-Martin and G. Isidori, *A three-site gauge model for flavor hierarchies and flavor anomalies*, *Phys. Lett.* **B779** (2018) 317–323, [1712.01368].

[284] R. Barbieri and A. Tesi, *B-decay anomalies in Pati-Salam SU(4)*, *Eur. Phys. J.* **C78** (2018) 193, [1712.06844].

[285] A. Greljo and B. A. Stefanek, *Third family quark–lepton unification at the TeV scale*, *Phys. Lett.* **B782** (2018) 131–138, [1802.04274].

[286] M. Blanke and A. Crivellin, *B Meson Anomalies in a Pati-Salam Model within the Randall-Sundrum Background*, *Phys. Rev. Lett.* **121** (2018) 011801, [1801.07256].

[287] J. Aebischer, A. Crivellin and C. Greub, *QCD improved matching for semileptonic B decays with leptoquarks*, *Phys. Rev.* **D99** (2019) 055002, [1811.08907].

[288] S. Ferrara, M. Porrati and V. L. Telegdi, *$g = 2$ as the natural value of the tree-level gyromagnetic ratio of elementary particles*, *Phys. Rev.* **D46** (1992) 3529–3537.

[289] R. Barbieri, C. W. Murphy and F. Senia, *B-decay Anomalies in a Composite Leptoquark Model*, *Eur. Phys. J.* **C77** (2017) 8, [1611.04930].

[290] J. Aebischer, A. Crivellin, M. Fael and C. Greub, *Matching of gauge invariant dimension-six operators for $b \to s$ and $b \to c$ transitions*, *JHEP* **05** (2016) 037, [1512.02830].

[291] E. E. Jenkins, A. V. Manohar and P. Stoffer, *Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching*, *JHEP* **03** (2018) 016, [1709.04486].

[292] C. Cornella, J. Fuentes-Martin and G. Isidori, *Revisiting the vector leptoquark explanation of the B-physics anomalies*, *JHEP* **07** (2019) 168, [1903.11517].

[293] B. Diaz, M. Schmaltz and Y.-M. Zhong, *The leptoquark Hunter's guide: Pair production*, *JHEP* **10** (2017) 097, [1706.05033].

[294] A. Greljo, J. Martin Camalich and J. D. Ruiz-Álvarez, *Mono-τ Signatures at the LHC Constrain Explanations of B-decay Anomalies*, *Phys. Rev. Lett.* **122** (2019) 131803, [1811.07920].

[295] M. J. Baker, J. Fuentes-Martín, G. Isidori and M. König, *High- p_T signatures in vector–leptoquark models*, *Eur. Phys. J. C* **79** (2019) 334, [1901.10480].

[296] L. Roszkowski, E. M. Sessolo and S. Trojanowski, *WIMP dark matter candidates and searches – current status and future prospects*, *Rept. Prog. Phys.* **81** (2018) 066201, [1707.06277].

[297] A. Celis, W.-Z. Feng and M. Vollmann, *Dirac dark matter and $b \to s\ell^+\ell^-$ with* U(1) *gauge symmetry*, *Phys. Rev. D* **95** (2017) 035018, [1608.03894].

[298] W. Altmannshofer, S. Gori, S. Profumo and F. S. Queiroz, *Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model*, *JHEP* **12** (2016) 106, [1609.04026].

[299] P. Ko, T. Nomura and H. Okada, *Explaining $B \to K^{(*)}\ell^+\ell^-$ anomaly by radiatively induced coupling in $U(1)_{\mu-\tau}$ gauge symmetry*, *Phys. Rev. D* **95** (2017) 111701, [1702.02699].

[300] J. M. Cline, J. M. Cornell, D. London and R. Watanabe, *Hidden sector explanation of B-decay and cosmic ray anomalies*, *Phys. Rev.* **D95** (2017) 095015, [1702.00395].

[301] J. Ellis, M. Fairbairn and P. Tunney, *Anomaly-Free Models for Flavour Anomalies*, *Eur. Phys. J. C* **78** (2018) 238, [1705.03447].

[302] J. Kawamura, S. Okawa and Y. Omura, *Interplay between the $b\to s\ell\ell$ anomalies and dark matter physics*, *Phys. Rev. D* **96** (2017) 075041, [1706.04344].

[303] C. Hati, G. Kumar, J. Orloff and A. M. Teixeira, *Reconciling B-meson decay anomalies with neutrino masses, dark matter and constraints from flavour violation*, *JHEP* **11** (2018) 011, [1806.10146].

[304] J. M. Cline and J. M. Cornell, *$R(K^{(*)})$ from dark matter exchange*, *Phys. Lett. B* **782** (2018) 232–237, [1711.10770].

[305] S. Baek, *Dark matter contribution to $b \to s\mu^+\mu^-$ anomaly in local $U(1)_{L_\mu - L_\tau}$ model*, *Phys. Lett. B* **781** (2018) 376–382, [1707.04573].

[306] A. Vicente, *Anomalies in $b \to s$ transitions and dark matter*, *Adv. High Energy Phys.* **2018** (2018) 3905848, [1803.04703].

[307] D. Cerdeño, A. Cheek, P. Martín-Ramiro and J. Moreno, *B anomalies and dark matter: a complex connection*, *Eur. Phys. J. C* **79** (2019) 517, [1902.01789].

[308] D. Guadagnoli, M. Reboud and P. Stangl, *The Dark Side of 4321*, 2005.10117.

[309] M. Cirelli, N. Fornengo and A. Strumia, *Minimal dark matter*, *Nucl. Phys. B* **753** (2006) 178–194, [hep-ph/0512090].

[310] M. J. Baker et al., *The Coannihilation Codex*, *JHEP* **12** (2015) 120, [1510.03434].

[311] H. Georgi and Y. Nakai, *Diphoton resonance from a new strong force*, *Phys. Rev. D* **94** (2016) 075005, [1606.05865].

[312] L. Di Luzio, J. Fuentes-Martin, A. Greljo, M. Nardecchia and S. Renner, *Maximal Flavour Violation: a Cabibbo mechanism for leptoquarks*, *JHEP* **11** (2018) 081, [1808.00942].

[313] J. Fuentes-Martín and P. Stangl, *Third-family quark-lepton unification with a fundamental composite Higgs*, 2004.11376.

[314] K. Griest and D. Seckel, *Three exceptions in the calculation of relic abundances*, *Phys. Rev. D* **43** (1991) 3191–3203.

[315] G. Jungman, M. Kamionkowski and K. Griest, *Supersymmetric dark matter*, *Phys. Rept.* **267** (1996) 195–373, [hep-ph/9506380].

[316] M. Cannoni, *Relativistic $\langle\sigma v_{rel}\rangle$ in the calculation of relics abundances: a closer look*, *Phys. Rev.* **D89** (2014) 103533, [1311.4508].

[317] Planck collaboration, N. Aghanim et al., *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6, [1807.06209].

[318] LUX collaboration, D. Akerib et al., *Results from a search for dark matter in the complete LUX exposure*, *Phys. Rev. Lett.* **118** (2017) 021303, [1608.07648].

[319] PandaX-II collaboration, X. Cui et al., *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, *Phys. Rev. Lett.* **119** (2017) 181302, [1708.06917].

[320] XENON collaboration, E. Aprile et al., *Dark Matter Search Results from a One Ton-Year Exposure of XENON1T*, *Phys. Rev. Lett.* **121** (2018) 111302, [1805.12562].

[321] P. Salati, *Indirect and direct dark matter detection*, *PoS* **CARGESE2007** (2007) 009.

[322] N. Anand, A. L. Fitzpatrick and W. Haxton, *Weakly interacting massive particle-nucleus elastic scattering response*, *Phys. Rev. C* **89** (2014) 065501, [1308.6288].

[323] F. Bishara, J. Brod, B. Grinstein and J. Zupan, *From quarks to nucleons in dark matter direct detection*, *JHEP* **11** (2017) 059, [1707.06998].

[324] R. K. Leane, *Indirect Detection of Dark Matter in the Galaxy*, in *3rd World Summit on Exploring the Dark Side of the Universe*, 5, 2020, 2006.00513.

[325] H.E.S.S. collaboration, V. Lefranc and E. Moulin, *Dark matter search in the inner galactic center halo with H.E.S.S*, in *51st Rencontres de Moriond on Cosmology*, pp. 149–152, 8, 2016, 1608.08453.

[326] HESS collaboration, H. Abdallah et al., *Search for γ-Ray Line Signals from Dark Matter Annihilations in the Inner Galactic Halo from 10 Years of Observations with H.E.S.S.*, *Phys. Rev. Lett.* **120** (2018) 201101, [1805.05741].

[327] Fermi-LAT collaboration, M. Ackermann et al., *Updated search for spectral lines from Galactic dark matter interactions with pass 8 data from the Fermi Large Area Telescope*, *Phys. Rev. D* **91** (2015) 122002, [1506.00013].

[328] J. F. Navarro, A. Ludlow, V. Springel, J. Wang, M. Vogelsberger, S. D. White et al., *The Diversity and Similarity of Cold Dark Matter Halos*, *Mon. Not. Roy. Astron. Soc.* **402** (2010) 21, [0810.1522].

[329] J. Hisano, S. Matsumoto, M. M. Nojiri and O. Saito, *Non-perturbative effect on dark matter annihilation and gamma ray signature from galactic center*, *Phys. Rev. D* **71** (2005) 063528, [hep-ph/0412403].

[330] J. F. Navarro, C. S. Frenk and S. D. White, *The Structure of cold dark matter halos*, *Astrophys. J.* **462** (1996) 563–575, [astro-ph/9508025].

[331] G. D. Kribs and E. T. Neil, *Review of strongly-coupled composite dark matter models and lattice simulations*, *Int. J. Mod. Phys. A* **31** (2016) 1643004, [1604.04627].

[332] A. Carvunis, D. Guadagnoli, M. Reboud and P. Stangl, *Composite Dark Matter and a horizontal symmetry*, 2007.11931.

[333] K. Griest and M. Kamionkowski, *Unitarity Limits on the Mass and Radius of Dark Matter Particles*, *Phys. Rev. Lett.* **64** (1990) 615.

[334] H. Georgi, *Weak Interactions and Modern Particle Theory*. 1984.

[335] J. Gasser, M. Sainio and A. Svarc, *Nucleons with Chiral Loops*, *Nucl. Phys. B* **307** (1988) 779–853.

[336] A. M. Green and J. Niskanen, *Low-energy Anti-proton Physics in the Early LEAR Era*, *Prog. Part. Nucl. Phys.* **18** (1987) 93–182.

[337] C. B. Dover, T. Gutsche, M. Maruyama and A. Faessler, *The Physics of nucleon - anti-nucleon annihilation*, *Prog. Part. Nucl. Phys.* **29** (1992) 87–174.

[338] C. Amsler, *Nucleon-antinucleon annihilation at LEAR*, 8, 2019, 1908.08455.

[339] J.-M. Richard, *Antiproton physics*, *Front. in Phys.* **8** (2020) 6, [1912.07385].

[340] M. Astrua, E. Botta, T. Bressani, D. Calvo, C. Casalegno, A. Feliciello et al., *Antineutron nucleus annihilation cross-sections below 400-MeV/c*, *Nucl. Phys. A* **697** (2002) 209–224.

[341] T. Kalogeropoulos and G. Tzanakos, *TOTAL AND PARTIAL ANTI-P D CROSS-SECTIONS FROM 0.26-GEV/C TO 0.47-GEV/C*, *Phys. Rev. D* **22** (1980) 2585–2594.

[342] W. Brückner et al., *Search for a Narrow Resonance in Anti-proton - Proton Annihilation Cross-sections in the Beam Momentum Range Between 400-{MeV}/c and 600-{MeV}/c*, *Phys. Lett. B* **197** (1987) 463.

[343] D. Lowenstein et al., *Search for narrow structure in proton-antiproton annihilation cross sections from 1900 to 1960 MeV*, *AIP Conf. Proc.* **132** (1985) 137–143.

[344] R. Bizzarri, P. Guidoni, F. Marcelja, F. Marzano, E. Castelli and M. Sessa, *Anti-proton - deuteron low-energy cross-section*, *Nuovo Cim. A* **22** (1974) 225–250.

[345] M. Cirelli, G. Corcella, A. Hektor, G. Hutsi, M. Kadastik, P. Panci et al., *PPPC 4 DM ID: A Poor Particle Physicist Cookbook for Dark Matter Indirect Detection*, *JCAP* **03** (2011) 051, [1012.4515].

[346] LHCb collaboration, A. A. Alves Jr. et al., *The LHCb detector at the LHC*, *JINST* **3** (2008) S08005.

[347] LHCb collaboration, R. Aaij et al., *LHCb Detector Performance*, *Int. J. Mod. Phys. A* **30** (2015) 1530022, [1412.6352].

[348] G. Lanfranchi, X. Cid Vidal, S. Furcas, M. Gandelman, J. A. Hernando, J. H. Lopez et al., *The Muon Identification Procedure of the LHCb Experiment for the First Data*, LHCb-PUB-2009-013. CERN-LHCb-PUB-2009-013.

[349] F. Archilli et al., *Performance of the Muon Identification at LHCb*, *JINST* **8** (2013) P10020, [1306.0249].

[350] R. Aaij, J. Albrecht, F. Dettori, K. Dungs, H. Lopes, D. Martinez Santos et al., *Optimization of the muon reconstruction algorithms for LHCb Run 2*, Tech. Rep. LHCb-PUB-2017-007. CERN-LHCb-PUB-2017-007, CERN, Geneva, Feb, 2017.

[351] M. De Cian, S. Farry, P. Seyfert and S. Stahl, *Fast neural-net based fake track rejection in the LHCb reconstruction*, Tech. Rep. LHCb-PUB-2017-011. CERN-LHCb-PUB-2017-011, CERN, Geneva, Mar, 2017.

[352] O. Deschamps, F. P. Machefert, M. H. Schune, G. Pakhlova and I. Belyaev, *Photon and neutral pion reconstruction*, LHCb-2003-091.

[353] C. A. Beteta et al., *Calibration and performance of the LHCb calorimeters in Run 1 and 2 at the LHC*, 2008.11556.

[354] M. Calvo Gomez, E. Cogneras, O. Deschamps, M. Hoballah, R. Lefevre, S. Monteil et al., *A tool for γ/π^0 separation at high energies*, Tech. Rep. LHCb-PUB-2015-016. CERN-LHCb-PUB-2015-016, CERN, Geneva, Aug, 2015.

[355] LHCb collaboration, 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, [1812.10790].

[356] T. Sjöstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175].

[357] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [0710.3820].

[358] I. Belyaev et al., *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, *J. Phys. Conf. Ser.* **331** (2011) 032047.

[359] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[360] Geant4 collaboration collaboration, S. Agostinelli et al., *Geant4: A simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250.

[361] Geant4 collaboration collaboration, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois et al., *Geant4 developments and applications*, *IEEE Trans.Nucl.Sci.* **53** (2006) 270.

[362] M. Clemencic et al., *The LHCb simulation application, Gauss: Design, evolution and experience*, *J. Phys. Conf. Ser.* **331** (2011) 032023.

[363] W. Bonivento, F. Dettori and N. Serra, *LHCb sensitivity to the* $B_s^0 \to \mu^\pm\mu^\mp\gamma$ *decay*, LHCb-PUB-2010-008, CERN-LHCb-PUB-2010-008.

[364] L. Beaucourt, *Study of* $B^0 \to K^{*0}\gamma$, $B_s^0 \to \phi\gamma$ *and* $B_s^0 \to K^{*0}\gamma$ *decays using converted photons with the LHCb detector.*, Oct, 2016.

[365] F. Archilli, M. O. Bettler, X. Cid Vidal, H. V. Cliff, F. Dettori, H. M. Evans et al., *Search for the* $B^0 \to \mu^+\mu^-$ *decay and measurement of the* $B_s^0 \to \mu^+\mu^-$ *branching fraction and effective lifetime*, LHCb-ANA-2016-038.

[366] N. V. Nikitin, A. V. Popov and D. V. Savrina, *The EvtGen based Monte-Carlo generator model for the rare radiative leptonic B-mesons decays*, Tech. Rep. LHCb-INT-2011-011. CERN-LHCb-INT-2011-011, CERN, Geneva, Apr, 2011.

[367] Particle Data Group collaboration, M. Tanabashi et al., *Review of particle physics*, *Phys. Rev.* **D98** (2018) 030001.

[368] M. Calvo, O. Deschamps, M. Hoballah, R. Lefèvre, A. Oyanguren, A. Puig et al., *Photon polarization in* $B_s^0 \to \phi\gamma$, LHCb-ANA-2014-102.

[369] L. Anderlini, A. Contu, C. R. Jones, S. S. Malde, D. Muller, S. Ogilvy et al., *The PIDCalib package*, July, 2016.

[370] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, *Nucl. Instrum. Meth.* **A555** (2005) 356–369, [physics/0402083].

[371] A. Hoecker et al., *TMVA 4 — Toolkit for Multivariate Data Analysis. Users Guide.*, physics/0703039.

[372] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, *Nucl. Instrum. Meth.* **A552** (2005) 566–575, [physics/0503191].

[373] G. Punzi, *Sensitivity of searches for new signals and its optimization*, *eConf* **C030908** (2003) MODT002, [physics/0308063].

[374] F. Archilli, M. Palutan, M. Rama, A. Sarti, B. Sciascia and R. Vazquez Gomez, *Background studies for $B^0 \to \mu^+\mu^-$ analysis optimization*, LHCb-INT-2014-047. CERN-LHCb-INT-2014-047.

[375] G. Mancinelli and J. Serrano, *Study of Muon Isolation in the $B_s^0 \to \mu^+\mu^-$ Channel*, LHCb-INT-2010-011. CERN-LHCb-INT-2010-011.

[376] P. Koppenburg, *Statistical biases in measurements with multiple candidates*, 1703.01128.

[377] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, Ph.D. thesis, Institute of Nuclear Physics, Krakow, 1986.

[378] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, *eConf* **C0303241** (2003) MOLT007, [physics/0306116].

[379] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, *Comput. Phys. Commun.* **10** (1975) 343–367.

[380] BaBar collaboration, J. P. Lees et al., *Search for the rare decays $B \to \pi\ell^+\ell^-$ and $B^0 \to \eta\ell^+\ell^-$*, *Phys. Rev.* **D88** (2013) 032012, [1303.6010].

[381] LHCb collaboration, R. Aaij et al., *First measurement of the differential branching fraction and CP asymmetry of the $B^+ \to \pi^+\mu^+\mu^-$ decay*, *JHEP* **10** (2015) 034 LHCb-PAPER-2015-035, CERN-PH-EP-2015-219, [1509.00414].

[382] W.-F. Wang and Z.-J. Xiao, *The semileptonic decays $B/B_s \to (\pi, K)(\ell^+\ell^-, \ell\nu, \nu\bar{\nu})$ in the perturbative QCD approach beyond the leading-order*, *Phys. Rev.* **D86** (2012) 114025, [1207.0265].

[383] G. Erkol and G. Turan, *The Analysis of $B^0 \to (\eta, \eta')\ell^+\ell^-$ decays in the standard model*, *Eur. Phys. J.* **C28** (2003) 243–248, [hep-ph/0212079].

[384] V. J. Rives Molina, *Study of b-hadron decays into two hadrons and a photon at LHCb and first observation of b-baryon radiative decays*, Ph.D. thesis, Barcelona U., 7, 2016.

[385] LHCb collaboration, R. Aaij et al., *First Observation of the Radiative Decay $\Lambda_b^0 \to \Lambda\gamma$*, *Phys. Rev. Lett.* **123** (2019) 031801, [1904.06697].

[386] LHCb collaboration, R. Aaij et al., *Study of η–η' mixing from measurement of $B^0_{(s)} \to J/\psi\,\eta^{(\prime)}$ decay rates*, *JHEP* **01** (2015) 024 CERN-PH-EP-2014-266, LHCb-PAPER-2014-056, [1411.0943].

[387] LHCb collaboration, R. Aaij et al., *Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **08** (2013) 117 CERN-PH-EP-2013-095, LHCb-PAPER-2013-004, [1306.3663].

[388] LHCb collaboration, R. Aaij et al., *Measurement of the $B^\pm$ production cross-section in pp collisions at $\sqrt{s} = 7$ and 13 TeV*, *JHEP* **12** (2017) 026 LHCb-PAPER-2017-037, CERN-EP-2017-254, [1710.04921].

[389] HFLAV collaboration, Y. Amhis et al., *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016*, *Eur. Phys. J. C* **77** (2017) 895, [1612.07233].

[390] LHCb collaboration, R. Aaij et al., *Measurement of forward J/ψ production cross-sections in pp collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **10** (2015) 172 LHCb-PAPER-2015-037, CERN-PH-EP-2015-222, [1509.00771].

[391] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, *J. Phys. Conf. Ser.* **762** (2016) , [1608.05806].

[392] S. Tolk, J. Albrecht, F. Dettori and A. Pellegrino, *Data driven trigger efficiency determination at LHCb*, LHCb-PUB-2014-039. CERN-LHCb-PUB-2014-039.

[393] P. Robbe, *Handling of the wrong L0 Calorimeter calibration constants in 2016 data and simulation*, LHCb-INT-2018-027. CERN-LHCb-INT-2018-027.

[394] I. Belyaev, *LoKi: Smart & Friendly C++ Physics Analysis Toolkit*, Tech. Rep. LHCb-2004-023, CERN, Geneva, Mar, 2004.

[395] W. Altmannshofer, P. Ball, A. Bharucha, A. J. Buras, D. M. Straub and M. Wick, *Symmetries and Asymmetries of $B \to K^*\mu^+\mu^-$ Decays in the Standard Model and Beyond*, *JHEP* **01** (2009) 019, [0811.1214].

[396] D. Melikhov, *Dispersion approach to quark binding effects in weak decays of heavy mesons*, *Eur. Phys. J.direct* **C4** (2002) 2, [hep-ph/0110087].

[397] D. Fakirov and B. Stech, *F and D Decays*, *Nucl. Phys. B* **133** (1978) 315–326.

[398] ARGUS collaboration, H. Albrecht et al., *Search for Hadronic $b \to u$ Decays*, *Phys. Lett. B* **241** (1990) 278–282.

A Miscellanies

A.1 The effective C_9 coefficient

The effects of $\mathcal{O}_{1-6}$ can be accounted for by a q^2-dependent shift to C_9 [395]

$$C_9^{\text{eff}} = C_9 + Y(q^2)\,,$$

where at next-to-leading logarithm (NLL)

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

and

$$h(q^2, m_q) = -\frac{4}{9}\left(\ln\frac{m_q^2}{\mu^2} - \frac{2}{3} - z\right) - \frac{4}{9}\,(2+z)\sqrt{|z-1|}\,\times \begin{cases} \arctan\dfrac{1}{\sqrt{z-1}} & z > 1 \\ \ln\dfrac{1+\sqrt{1-z}}{\sqrt{z}} - \dfrac{i\pi}{2} & z \le 1 \end{cases} \tag{A.2}$$

where $z = 4m_q^2/q^2$.

A.2 $B \to \gamma$ form-factors

This appendix discusses the parametrization of the form-factor of the $B \to \gamma^*$ transition. These results are discussed in details in [125].

The most general Lorentz and gauge invariant structure of the *axial-vector current* involves three k-dependent form-factors [396]

$$\begin{aligned}\langle\gamma^*(k,\varepsilon)|\bar{s}\,\gamma^\mu\,\gamma_5\,b|\bar{B}_s^0\rangle = ie\,\varepsilon_\alpha^*\Bigg[\left(g_{\mu\alpha}-\frac{k_\alpha k_\mu}{k^2}\right)f + p_\mu\left(p_\alpha-(k\cdot p)\,\frac{k_\alpha}{k^2}\right)a_1 \\ + k_\mu\left(p_\alpha-(k\cdot p)\,\frac{k_\alpha}{k^2}\right)a_2\Bigg]\,.\end{aligned} \tag{A.3}$$

The regularity of the amplitude at $k^2=0$ (on-shell photon) imposes the relations

$$f+(k\cdot p)\,a_2 = 0 \qquad \text{and} \qquad a_1 = 0\,. \tag{A.4}$$

Using this relation, eq. (A.3) can be written as

$$\langle\gamma^*(k,\varepsilon)|\bar{s}\,\gamma^\mu\,\gamma_5\,b|\bar{B}_s^0\rangle = -ie\,\varepsilon_\alpha^*\,[g_{\mu\alpha}(p\cdot k)-k_\mu p_\alpha]\,a_2\,. \tag{A.5}$$

The *vector current* only involves one form-factor

$$\langle\gamma^*(k,\varepsilon)|\bar{s}\,\gamma^\mu\,b|\bar{B}_s^0\rangle = 2e\,g\,\varepsilon_\alpha^*(k)\,\varepsilon^{\mu\alpha\rho\sigma}p_\rho k_\sigma\,. \tag{A.6}$$

The *pseudo-tensor current* also requires three form-factors

$$\begin{aligned}\langle\gamma^*(k,\alpha)|\bar{s}\,\sigma^{\mu\nu}\gamma_5\,b|\bar{B}_s^0\rangle = e\,\varepsilon_\alpha^*\Bigg[\left(p_\mu\left(g_{\alpha\nu}-\frac{k_\alpha k_\nu}{k^2}\right)+(\mu\leftrightarrow\nu)\right)g_1 \\ (g_{\alpha\nu}k_\mu-g_{\alpha\mu}k_\nu)\,g_2+(k_\mu p_\nu-k_\nu p_\mu)\left(p_\alpha-(k\cdot p)\,\frac{k_\alpha}{k^2}\right)g_0\,.\end{aligned} \tag{A.7}$$

Here again, the absence of pole at $k^2=0$ imposes the relation between the form-factors

$$g_1+(k\cdot p)\,g_0 = 0\,. \tag{A.8}$$

One can finally use the relation $\sigma_{\mu\nu}\gamma_5 = -\frac{i}{2}\varepsilon_{\mu\nu\alpha\beta}\sigma^{\alpha\beta}$ to infer the *tensor current* amplitude

$$\begin{aligned}\langle\gamma^*(k,\alpha)|\bar{s}\,\sigma^{\mu\nu}\gamma_5\,b|\bar{B}_s^0\rangle = ie\,\varepsilon_\alpha^*\Bigg[\left(\varepsilon_{\mu\nu\alpha\rho}\,p^\rho-\frac{\varepsilon_{\mu\nu\sigma\rho}\,k_\alpha k^\sigma p^\rho}{k^2}\right)g_1 \\ \varepsilon_{\mu\nu\alpha\rho}\,k^\rho\,g_2+\varepsilon_{\mu\nu\rho\sigma}\,p^\rho k^\sigma\left(p_\alpha-(k\cdot p)\,\frac{k_\alpha}{k^2}\right)g_0\,.\end{aligned} \tag{A.9}$$

Defining the dimensionless form-factors $V_{\perp,\parallel}, T_{\perp,\parallel}$ as

$$V_\parallel = M_{B_s^0}\,a_2\,, \qquad V_\perp = -2M_{B_s^0}\,g\,, \tag{A.10}$$

$$T_\parallel = g_2+(p^2-p\cdot k)\,g_0\,, \qquad T_\perp = g_2+(p\cdot k)\,g_0\,, \tag{A.11}$$

we get the parametrization used in eq. (1.13),

$$\begin{aligned}
\langle\gamma(k,\varepsilon)|\bar{s}\,\gamma_\mu\,\gamma_5\,b|\bar{B}_s^0(p)\rangle M_{B_s^0} &= -ie\,\varepsilon^{\alpha*}(g_{\mu\alpha}(p\cdot k)-p_\alpha k_\mu)\,V_\parallel(q^2)\,,\\
\langle\gamma(k,\varepsilon)|\bar{s}\,\gamma_\mu\,b|\bar{B}_s^0(p)\rangle M_{B_s^0} &= -e\,\varepsilon^{\alpha*}\varepsilon_{\mu\alpha\rho\sigma}p^\rho k^\sigma\,V_\perp(q^2)\,,\\
\langle\gamma^*(k,\alpha)|\bar{s}\,\sigma_{\mu\nu}\,\gamma_5\,b|\bar{B}_s^0(p)\rangle(p-k)^\nu &= -e\,\varepsilon^{\alpha*}(g_{\mu\alpha}(p\cdot k)-p_\alpha k_\mu)\,T_\parallel(q^2,k^2)\,,\\
\langle\gamma^*(k,\alpha)|\bar{s}\,\sigma_{\mu\nu}\,b|\bar{B}_s^0(p)\rangle(p-k)^\nu &= -ie\,\varepsilon^{\alpha*}\varepsilon_{\mu\alpha\rho\sigma}p^\rho k^\sigma\,T_\perp(q^2,k^2)\,.
\end{aligned}$$

We note that eq. (A.11) implies $T_\parallel = T_\perp + (p^2 - 2(p\cdot k))\,g_0$, which yields the exact well-known relation $T_\parallel(0,0) = T_\perp(0,0)$.

At our best knowledge, no calculation of these form-factors, based on first principles, exists for the full q^2 range. On the other hand several regimes of the photon energy permit a general parametrization of these functions.

(i) In the region where the photon is hard, $E_\gamma \gg \Lambda_{QCD}$, the use of QCD factorization and soft collinear effective field theory (SCET) allows for a systematic treatment of non-perturbative effects [145].

(ii) At large q^2, one can assume the nearest pole dominance [397]. These poles correspond to the excited B_s mesons, namely B_s^* for $V_\perp$ and $T_\perp$ and B_s^{**} for $V_\parallel$ and $T_\parallel$.

The parametrization eq. (1.15) associated to the parameters in Table 1.1 are consistent with these two regimes [125].

A.3 The fermions in 4321 models

Among the different possibilities for implementing the SM fermions in a 4321 model, a well-motivated and phenomenologically successful variant corresponds to a unification of third-family quarks and leptons [229, 283, 285, 292, 313]. In this case, the first and second families of SM-like fermions transform under the $SU(3)' \times SU(2)_L \times U(1)_X$ subgroup of the 4321 symmetry like the usual SM fermions, whereas the third-family quarks and leptons are unified into $\Psi_L'^3 \equiv (Q'^3\ \ L'^3)^\intercal$, $\Psi_R'^{+3} \equiv (u'^3\ \ \nu'^3)^\intercal$, and $\Psi_R'^{-3} \equiv (d'^3\ \ e'^3)^\intercal$, which transform under the 4321 symmetry as shown in table A.1. Due to their quantum numbers, the light SM fermions cannot directly couple to the U_1. However, small but non-vanishing couplings between the U_1 and light SM fermions are required to explain the B-meson anomalies. To realize this, we introduce two massive fermions that couple to the U_1 and mix with the left-handed first and second generation SM-like fermions. In addition to couplings between light fermions and the U_1, whose sizes are controlled by the mixing, this construction also generates the 2-3 entries in the CKM matrix. The new heavy fermions transform in the same way as $\Psi_L'^3$ (cf.

Field	$SU(4)$	$SU(3)'$	$SU(2)_L$	$U(1)_X$
$L'^{1,2}$	**1**	**1**	**2**	$-1/2$
$e'^{1,2}$	**1**	**1**	**1**	-1
$Q'^{1,2}$	**1**	**3**	**2**	$+1/6$
$u'^{1,2}$	**1**	**3**	**1**	$+2/3$
$d'^{1,2}$	**1**	**3**	**1**	$-1/3$
Ψ'^{3}_{L}	**4**	**1**	**2**	0
$\Psi'^{+\,3}_{R}$	**4**	**1**	**1**	$+1/2$
$\Psi'^{-\,3}_{R}$	**4**	**1**	**1**	$-1/2$
Ψ_{DM}	**4**	**1**	**N**	$+1/2$

Table A.1: Quantum numbers of SM-like fermions (upper block) and the vector-like DM multiplet Ψ_{DM} (last row). First and second generation fermions transform under $SU(3)' \times SU(2)_L \times U(1)_X$ like the usual SM fermions; the third generation quarks and leptons are unified into $\Psi'^{3}_{L} \equiv (Q'^{3}\ L'^{3})^{\intercal}$, $\Psi'^{+\,3}_{R} \equiv (u'^{3}\ \nu'^{3})^{\intercal}$, and $\Psi'^{-\,3}_{R} \equiv (d'^{3}\ e'^{3})^{\intercal}$.

table A.1) and we denote their left-handed components by $\Psi'^{1,2}_{L}$. While the mixing is important for the couplings of the SM fermions to U_1, Z', and G', we do not further discuss the new heavy fermion mass eigenstates since they are not relevant for the DM dynamics as long as their masses are larger than $M_\chi + M_\psi$, which we assume in the following. Due to the mixing, the first and second generation SM $SU(2)_L$ doublets are in general linear combinations of the SM-like fields $Q'^{1,2}$, $L'^{1,2}$ and the new heavy fields $\Psi'^{1,2}_{L}$. To avoid large flavor violating effects, we align the mixings between SM fermions and new heavy fermions in the basis in which the down-quark mass matrix is diagonal (cf. e.g. [312]) such that the mixings are flavor-diagonal for the fields

$$Q^i = \begin{pmatrix} V^*_{ji}\, u^j_L \\ d^i_L \end{pmatrix}, \qquad L^j = \begin{pmatrix} \nu^j_L \\ e^j_L \end{pmatrix}, \qquad u^i_R\,, \qquad d^i_R\,, \qquad e^i_R\,, \qquad \nu^i_R\,, \tag{A.12}$$

where V is the CKM matrix and u^i, d^i, e^i, and ν^i are mass eigenstates. A possible misalignment between the quark and lepton components of the fields Ψ'^{i}_{L} is parameterized by embedding the quark and lepton components $\Psi'^{1,2}_{Q}$ and $\Psi'^{1,2}_{\ell\,L}$ that have a flavor-diagonal mixing with $Q'^{1,2}$ and $L'^{1,2}$, respectively, as

$$\Psi'^{i}_{L} = \begin{pmatrix} \Psi'^{i}_{q\,L} \\ W_{ij}\, \Psi'^{j}_{\ell\,L} \end{pmatrix}, \tag{A.13}$$

where W is a unitary matrix parameterizing the misalignment. This matrix is usually chosen to be CP-conserving and to mix only the second and third generation,

i.e. we use

$$W = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_{LQ} & \sin\theta_{LQ} \\ 0 & -\sin\theta_{LQ} & \cos\theta_{LQ} \end{pmatrix}. \tag{A.14}$$

In the absence of additional new heavy fermions that mix with right-handed SM fermions, a possible quark-lepton misalignment in $\Psi_R'^{+\,3}$ and $\Psi_R'^{-\,3}$ corresponds to only a phase difference, which we parameterize as

$$\Psi_R'^{+\,3} = \begin{pmatrix} \Psi'^{\,3}_{u\,R} \\ e^{i\phi_\nu}\,\Psi'^{\,3}_{\nu\,R} \end{pmatrix}, \qquad \Psi_R'^{-\,3} = \begin{pmatrix} \Psi'^{\,3}_{d\,R} \\ e^{i\phi_e}\,\Psi'^{\,3}_{e\,R} \end{pmatrix}. \tag{A.15}$$

Consequently, the SM fields in the basis where the down-quark mass matrix is diagonal can be expressed as

$$\begin{aligned}
q_L^{1,2} &= q_L'^{1,2}\,\cos\theta_{q_{1,2}} + \Psi_{q\,L}'^{1,2}\,\sin\theta_{q_{1,2}}\,, & q_L^3 &= \Psi_{q\,L}'^{\,3}\,, \\
\ell_L^{1,2} &= \ell_L'^{1,2}\,\cos\theta_{\ell_{1,2}} + \Psi_{\ell\,L}'^{1,2}\,\sin\theta_{\ell_{1,2}}\,, & \ell_L^3 &= \Psi_{\ell\,L}'^{\,3}\,, \\
u_R^{1,2} &= u_R'^{1,2}\,, & u_R^3 &= \Psi_{u\,R}'^{\,3}\,, \\
d_R^{1,2} &= d_R'^{1,2}\,, & d_R^3 &= \Psi_{d\,R}'^{\,3}\,, \\
e_R^{1,2} &= e_R'^{1,2}\,, & e_R^3 &= \Psi_{e\,R}'^{\,3}\,, \\
& & \nu_R^3 &= \Psi_{\nu\,R}'^{\,3}\,.
\end{aligned} \tag{A.16}$$

B
Additional information

B.1 Trigger lines

We detail here the requirements imposed by the trigger lines used in the analysis. Transverse momentum thresholds can be found in Tables 8.4, 8.5 and 8.6. Trigger prescales are set to 1 unless specified (a prescale of 0.1 means that 10% of the events passing the line are kept).

`L0Muon`: $P_T > P_{T,\text{threshold}}$, `nSPDHits` < 450, $\Sigma E_T|_{\text{previous event}} < 1$ GeV

`L0Photon`: $P_T > P_{T,\text{threshold}}$, `nSPDHits` < 450, $\Sigma E_T|_{\text{previous event}} < 1$ GeV

`Hlt1TrackMVA`: track $\chi^2/\text{ndf} < 2.5$, χ^2_{IP} constrained by eq. (8.13)

`Hlt1TwoTrackMVA`: $P > 5$ GeV, $P_T > 0.6$ GeV, track $\chi^2/\text{ndf} < 2.5$, $\chi^2_{\text{vtx}} < 10$, $2 < \eta_{\text{combination}} < 5$, MVA > 0.95

`Hlt1TrackMuon`: $P > 3$ GeV, $P_T > 0.8$ GeV, $\chi^2_{\text{IP}} > 9$, track $\chi^2/\text{ndf} < 3$, `IsMuon`

`Hlt1DiMuonLowMass`: doca < 0.2 mm, $\chi^2_{\text{vtx}} < 25$, $P(\mu^\pm) > 3$ GeV, $P_T(\mu^\pm) > 0.2$ GeV, $\chi^2_{\text{IP}}(\mu^\pm) > 6$, track $\chi^2(\mu^\pm)/\text{ndf} < 3$, `IsMuon`

`Hlt2SingleMuon`: `Hlt1TrackMuon`, $P_T > 1.3$ GeV, $\chi^2_{\text{IP}} > 16$, **prescale** $= 0.5$

`Hlt2SingleMuonLowPT`: $P_T > 4.8$ GeV, track $\chi^2/\text{ndf} < 3$, `IsMuon`, **prescale** $= 0.1$

`Hlt2TopoMuMu2Body`: $P(\mu^\pm) > 3$ GeV, $P_T(\mu^\pm) > 0.2$ GeV, $\chi^2_{\text{IP}}(\mu^\pm) > 4$, $\chi^2_{\text{IP}}(\mu^\mp) > 16$, track $\chi^2(\mu^\pm)/\text{ndf} < 3$, $\mu^+\mu^-$ vertex distance $\chi^2 > 16$, $2 < \eta(\mu^+\mu^-) < 5$, $P_T(\mu^+ + \mu^-) > 2$ GeV, 1 GeV $< m(\mu\mu) < 10$ GeV, BBDT > 0.993

B.2 Variables used in the multivariate analysis

The signal and data side-bands distribution of the variables used for the MLPS and the MLP training are presented in Figs. B.1 and B.2.

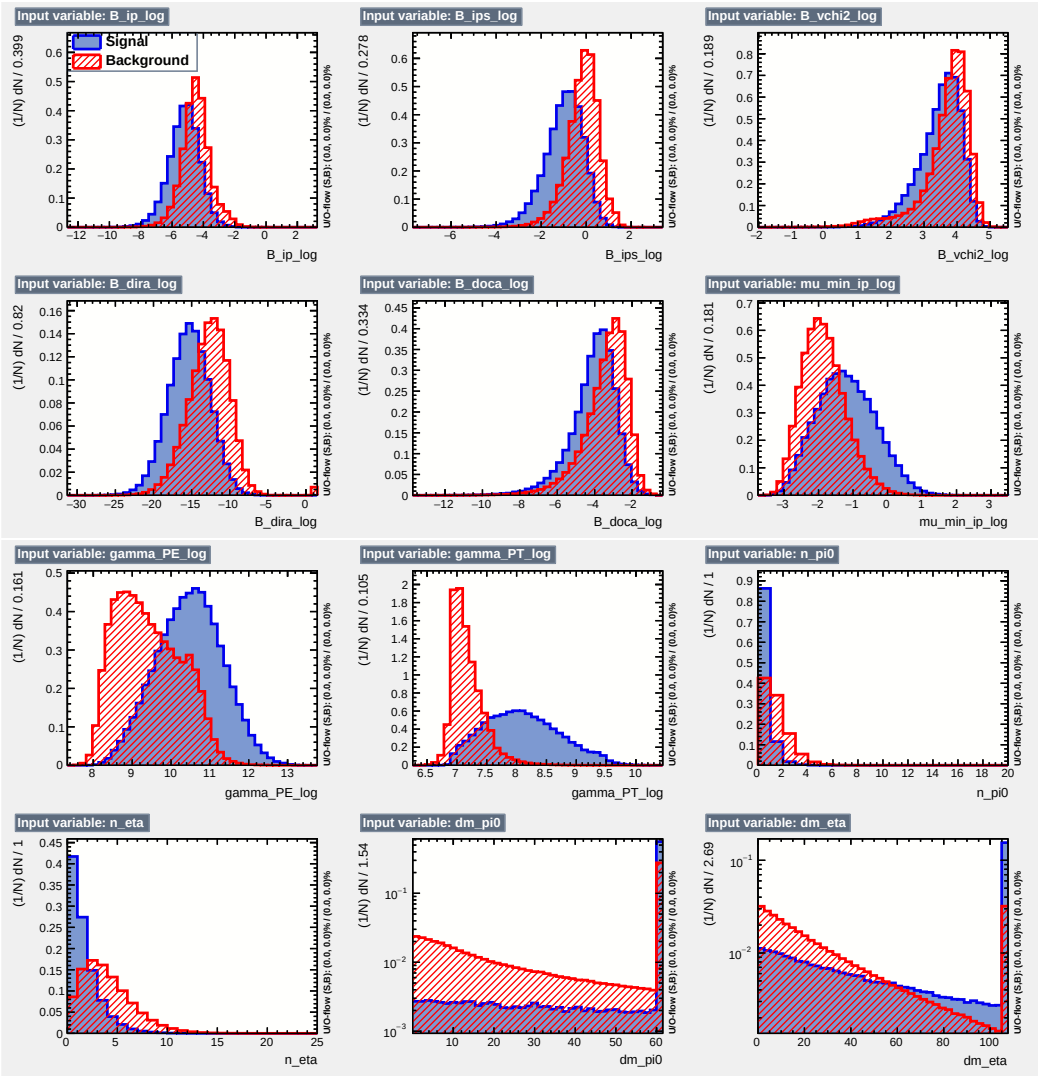

Figure B.1: Distribution of the variables used in the selection multivariate analysis. $B_s^0 \to \mu^+\mu^-\gamma$ events are plotted in blue and data side-bands in red.

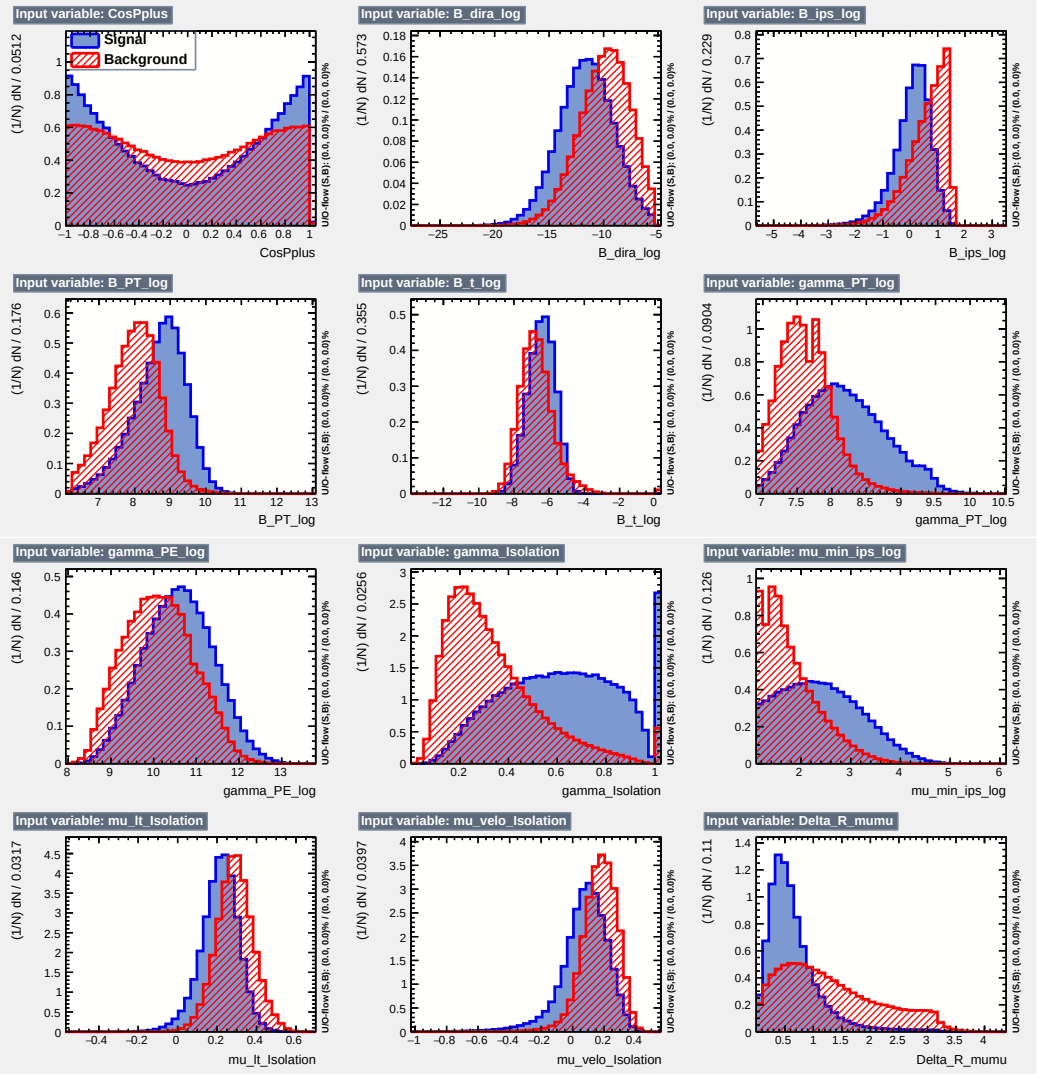

Figure B.2: Distribution of the variables used in the multivariate analysis classifier. $B_s^0 \to \mu^+\mu^-\gamma$ events are plotted in blue and data side-bands in red.

B.3 Additional data for the q^2 bins

B.4 Fit function

We give here the definition of the distributions used in the fits.

Argus distribution: This function corresponds to the distribution of the invari-

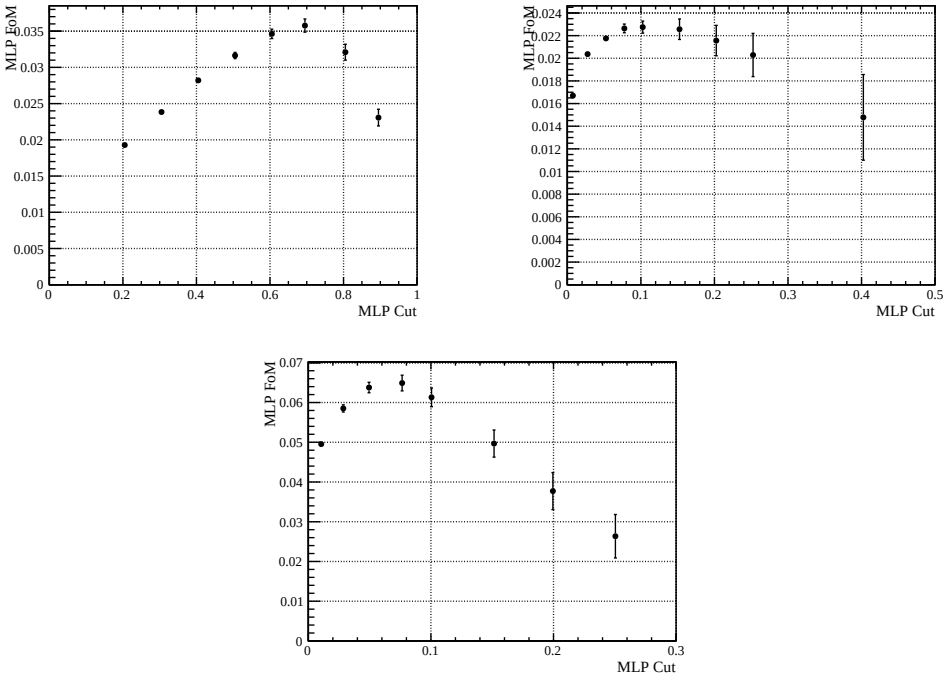

Figure B.3: Figure of merit of $B_s^0 \to \mu^+\mu^-\gamma$ events as defined in Eq. (5.14) for different cuts on the MLP output and for the three different bins (up left for Bin I, up right for Bin II and bottom for Bin III). The error on the MLP FoM are mainly due to the uncertainties in the exponential fit of the background. Note that the overall normalization is arbitrary, as for this plot, the signal efficiency ε_S in Eq. (5.14) is restricted to the MLP signal efficiency.

ant mass of a partially reconstructed background [398]

$$\mathrm{A}(m;\chi,M) = \mathcal{N}(\chi)\,\frac{m}{M^2}\sqrt{1-\frac{m^2}{M^2}}\,\exp\left\{-\frac{1}{2}\chi^2\left(1-\frac{m^2}{M^2}\right)\right\}, \tag{B.1}$$

with

$$\mathcal{N}(\chi) = \frac{\chi^3}{\sqrt{2\pi}(G(\chi)-\chi\,g(\chi)-\frac{1}{2})}$$

and g (G) the p.d.f (respectively cumulative) of the standard normal distribution.

Crystal Ball distribution: This function is defined by a Gaussian core and a power-law left tail [377]

$$\mathrm{CB}(m;\mu,\sigma,\alpha,n) = \begin{cases} a\left(b-\dfrac{m-\mu}{\sigma}\right)^{-n} & \text{for } \dfrac{m-\mu}{\sigma} \le -\alpha, \\ \exp\left\{-\dfrac{(m-\mu)^2}{2\sigma^2}\right\} & \text{for } -\alpha < \dfrac{m-\mu}{\sigma}, \end{cases} \tag{B.2}$$

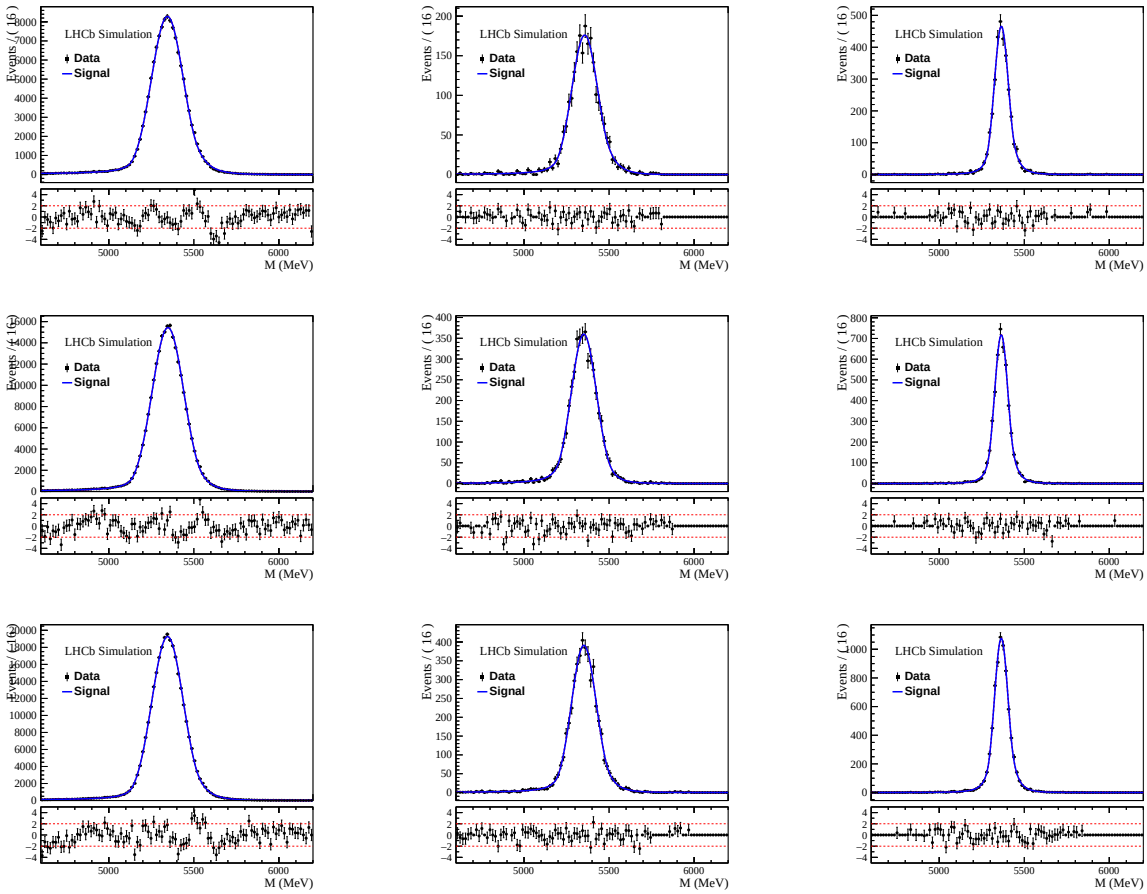

Figure B.4: Double sided Crystal Ball fit of the B_s^0 candidate mass for signal simulations in 2016 (top), 2017 (middle) and 2018 (bottom). The fit are performed in the q^2 bins, Bin I (left), Bin II (middle) and Bin III (right). Results of the fits are displayed on Table 7.2.

where $\alpha > 0$ and

$$a = \left(\frac{n_i}{\alpha_i}\right)^{n_i} \exp\left\{-\frac{\alpha_i^2}{2}\right\},$$
$$b = \frac{n}{\alpha} - \alpha\,.$$

B.5 Data-MC distributions

The distribution of the variables used in the analysis are shown in Figs. B.5, B.6, B.7 for the control channel $B_s^0 \to \phi\gamma$ and in Figs. B.8, B.9, B.10 for the normalization channel $B_s^0 \to J/\psi\,\eta$.

B.6 L0Muon Trigger efficiency plots

The L0Muon efficiency tables used in for the estimation of the L0Muon efficiency are displayed in Fig. B.11. Due to the correlation between the B transverse momentum and the maximal transverse momentum of the muon, some bins have a large error. On the other hand $B_s^0 \to \mu^+\mu^-\gamma$ and $B \to J/\psi K$ kinematics are closed, so only few signal events fall in these bins.

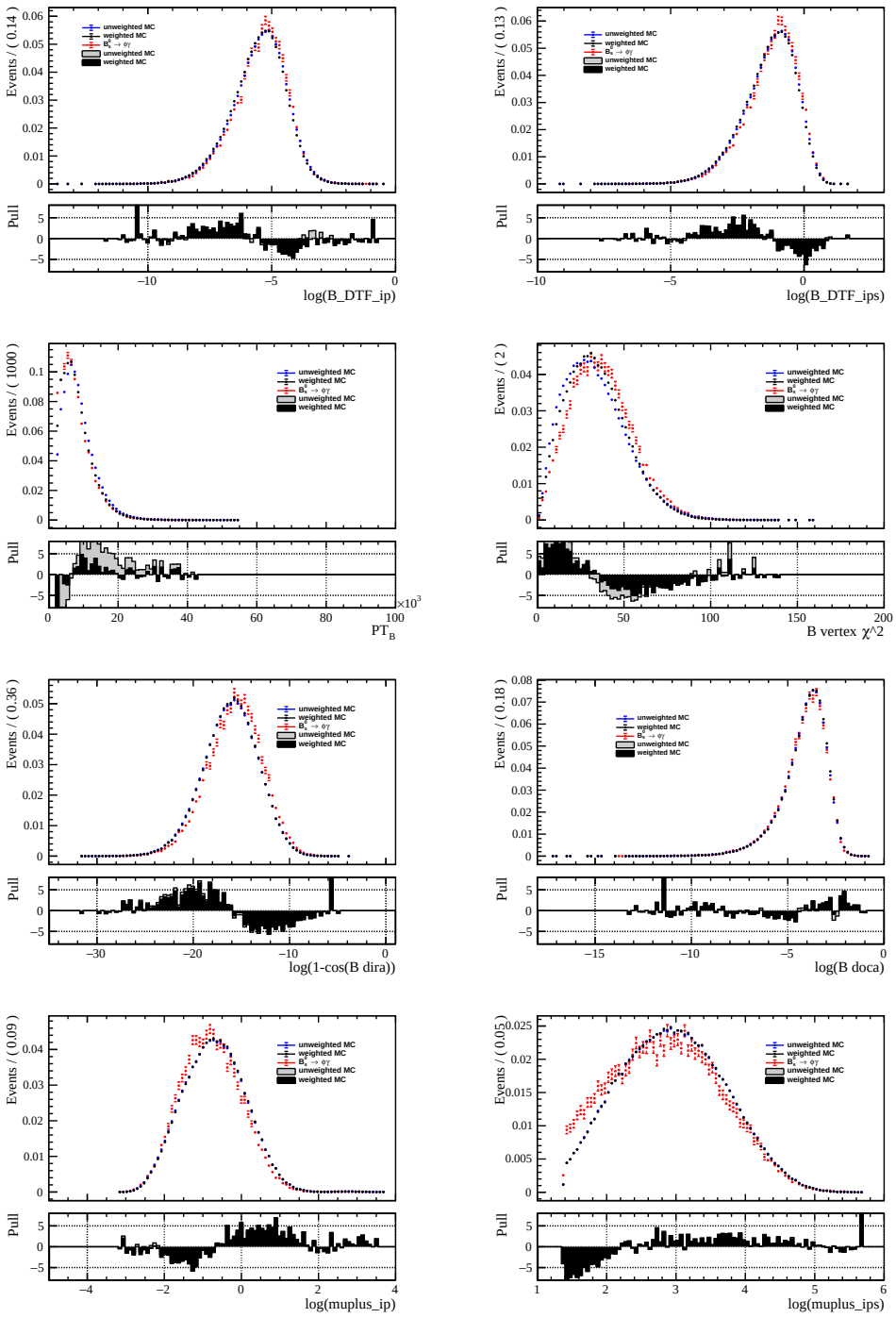

Figure B.5: Data *vs.* simulated distributions for $B_s^0 \to \phi\gamma$ events. The main source of discrepancies is due to the tracking of the kaons.

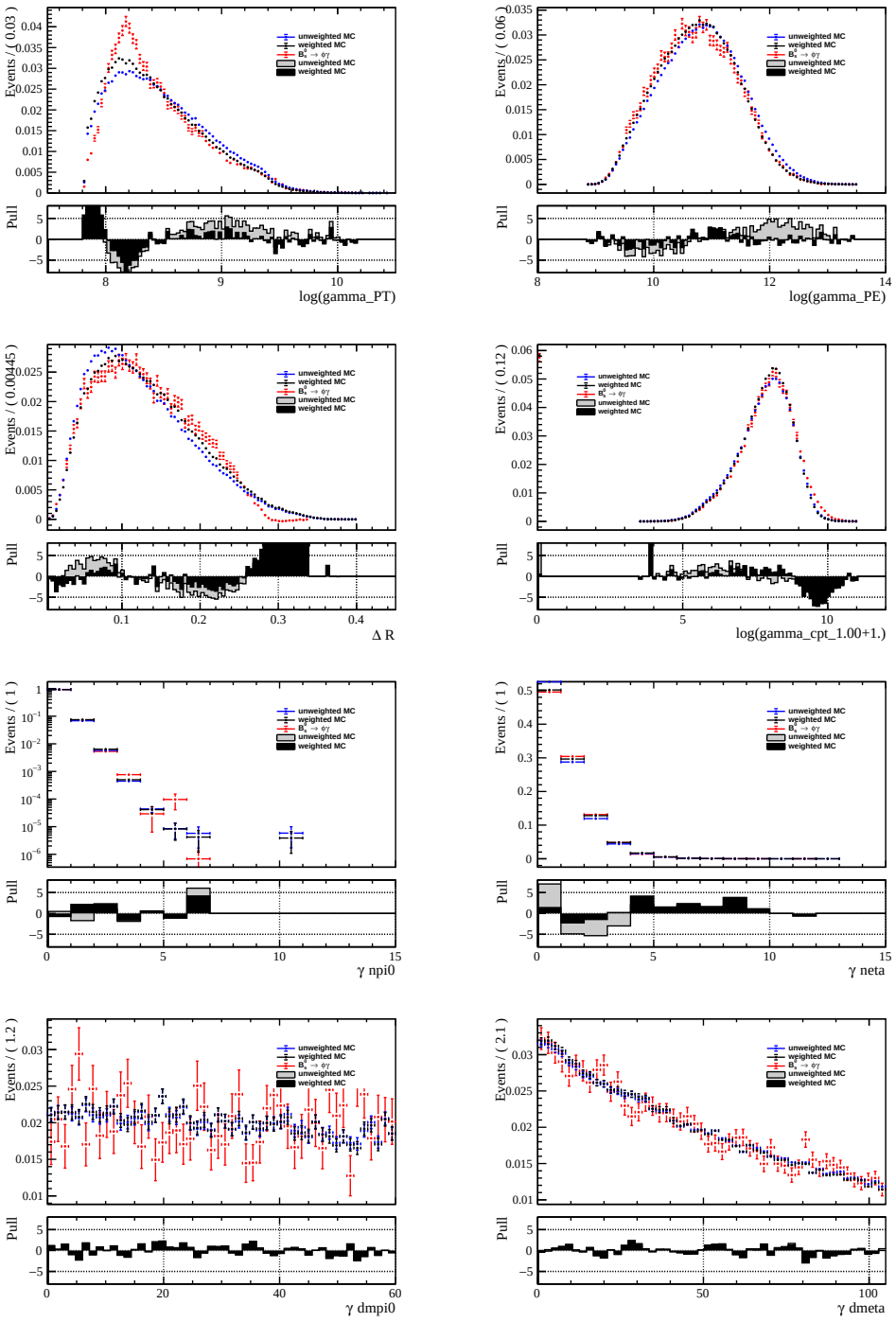

Figure B.6: Data *vs.* simulated distributions for $B_s^0 \to \phi\gamma$ events. The main source of discrepancies is due to the tracking of the kaons.

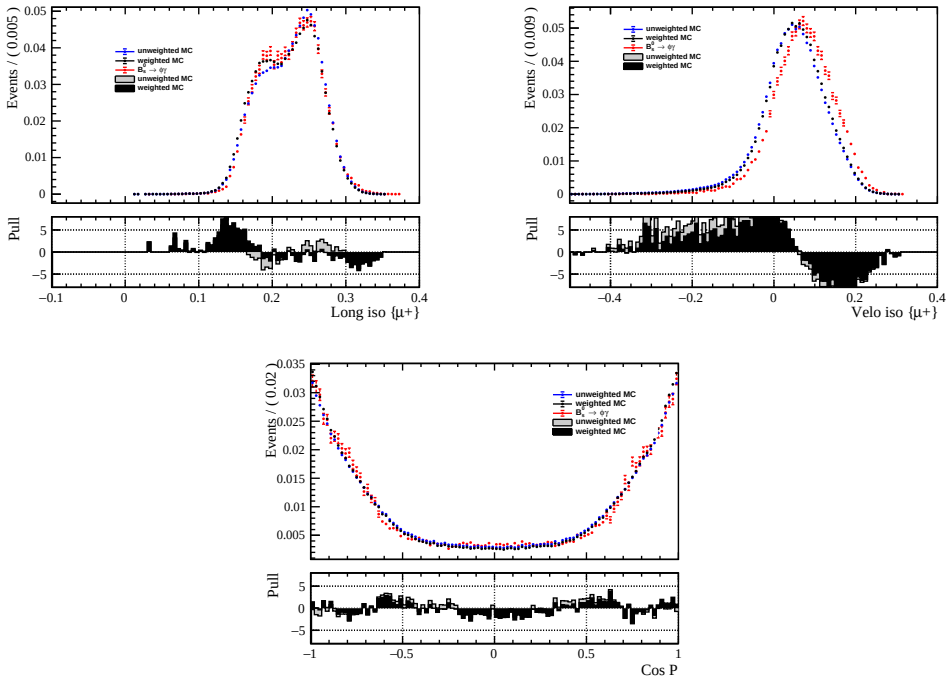

Figure B.7: Data *vs.* simulated distributions for $B_s^0 \to \phi\gamma$ events. The main source of discrepancies is due to the tracking of the kaons.

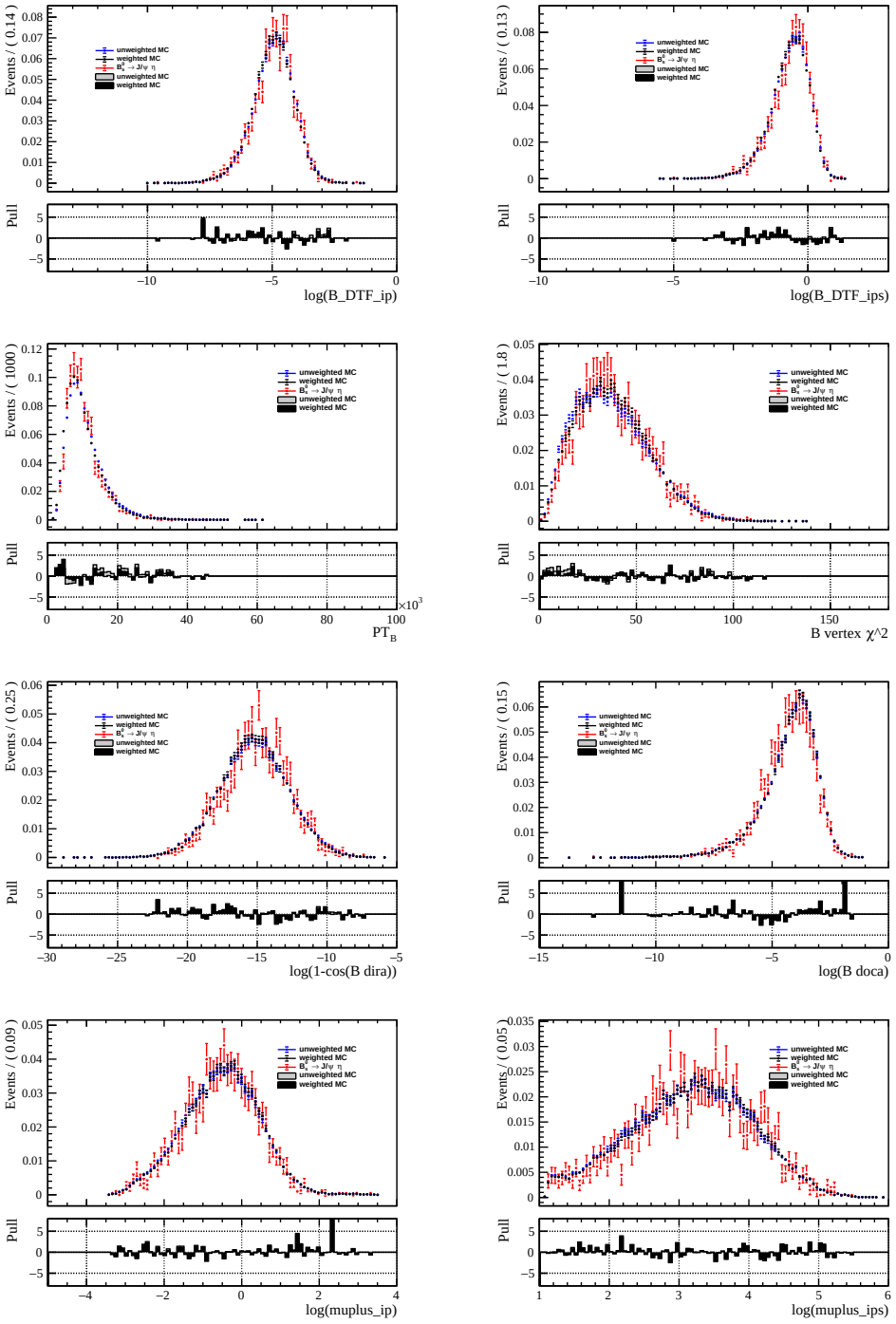

Figure B.8: Data *vs.* simulated distributions for $B_s^0 \to J/\psi\,\eta$ events.

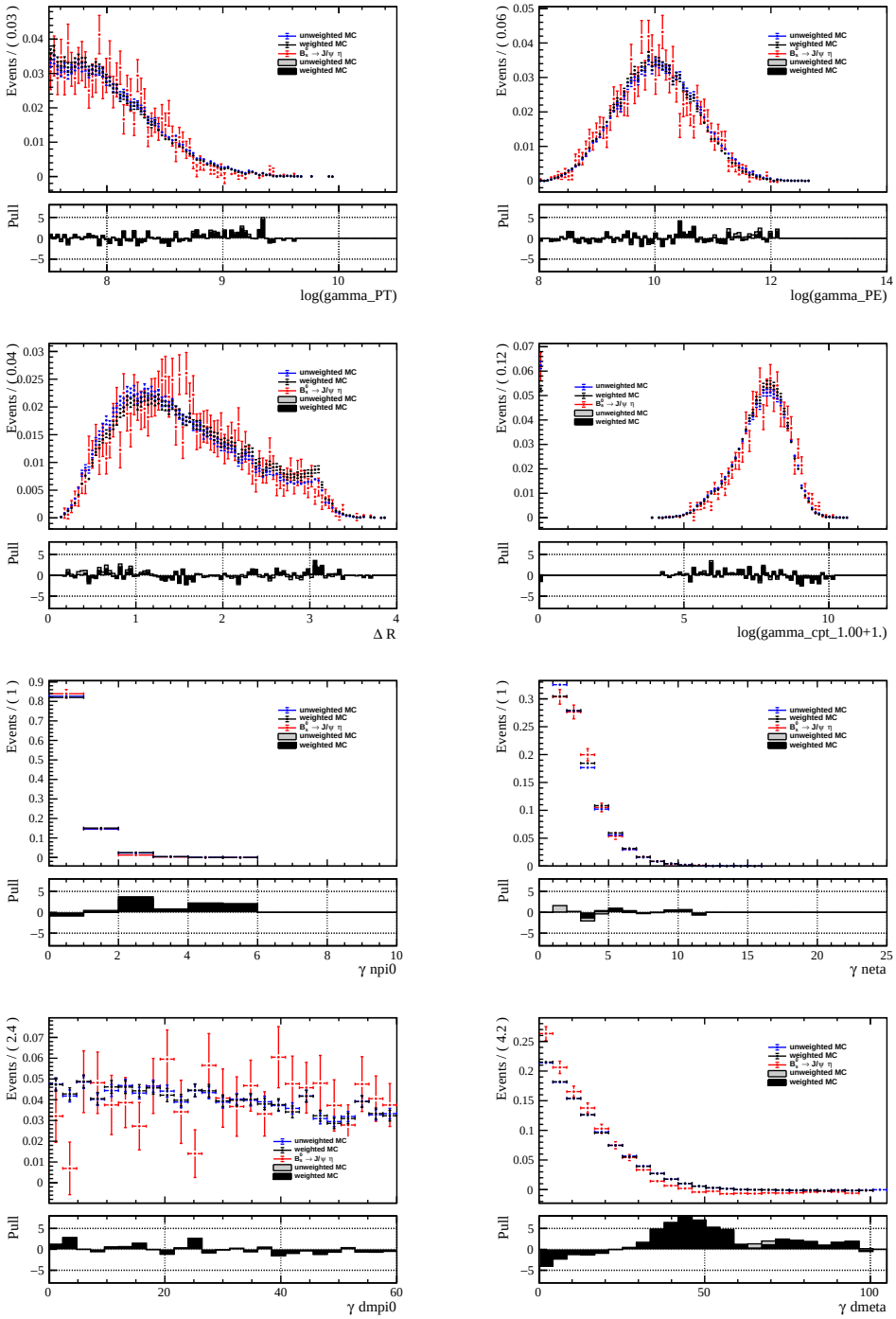

Figure B.9: Data *vs.* simulated distributions for $B_s^0 \to J/\psi\,\eta$ events.

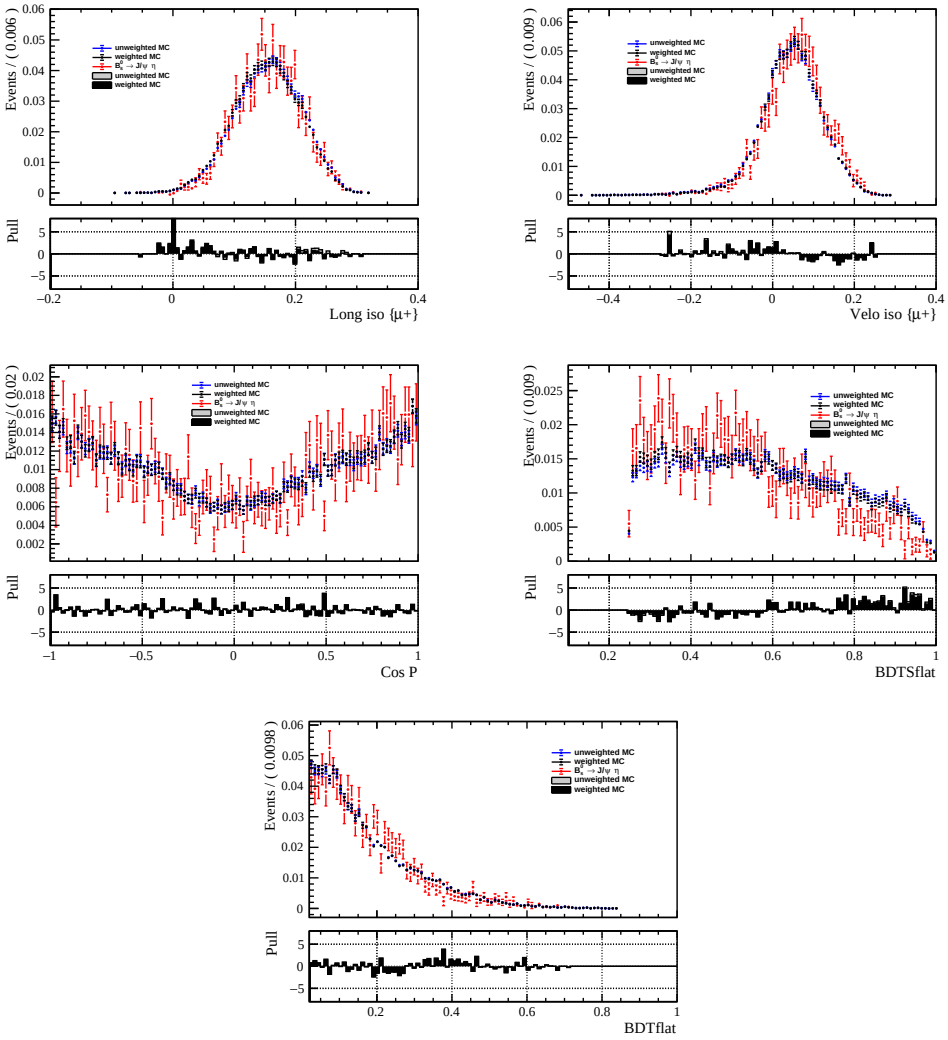

Figure B.10: Data *vs.* simulated distributions for $B_s^0 \to J/\psi\,\eta$ events.

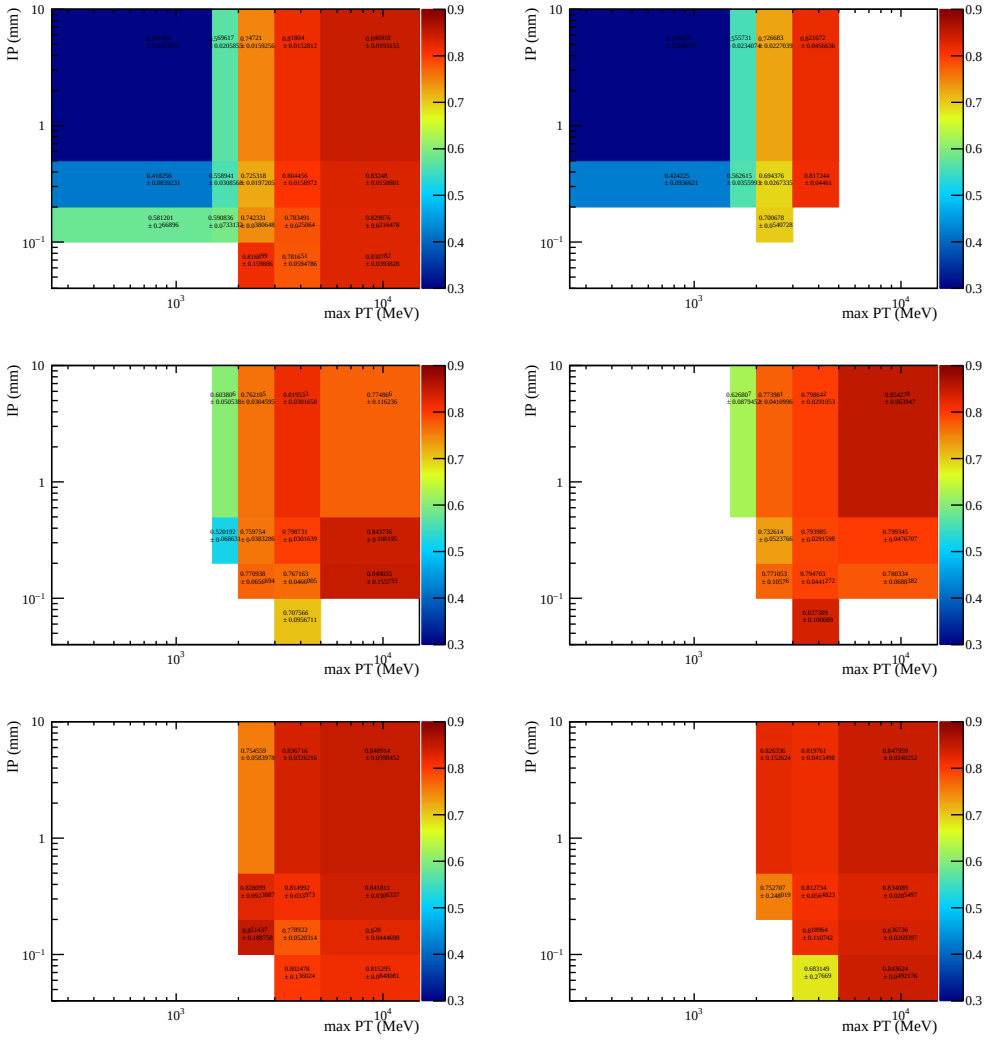

Figure B.11: L0Muon trigger efficiency tables computed on $B \to J/\psi K$ events using the TISTOS method. The upper left plot is averaged over the B kinematics. The five following plots are computed on B_{PT} bins (see 8.4 for details and the edge of the bins.) Bins with low statistics (compatible with zero at 2 sigma) are removed for readability.