



**HAL**  
open science

## Flavor physics: Precision as an avenue to discovery

Diego Guadagnoli

► **To cite this version:**

Diego Guadagnoli. Flavor physics: Precision as an avenue to discovery. International Journal of Modern Physics A, 2020, 35 (01), pp.1930018. 10.1142/S0217751X19300187. hal-02507725

**HAL Id: hal-02507725**

**<https://hal.science/hal-02507725>**

Submitted on 9 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Auteur:** Diego Guadagnoli

**Affiliation:**

Laboratoire d'Annecy-le-Vieux de Physique Théorique (LAPTh, UMR 5108) et  
Université de Savoie Mont-Blanc

**Titre:** Flavor physics: Precision as an avenue to discovery

**Composition du Jury:**

M. Augusto CECCUCCI

M. Paolo GAMBINO (rapporteur)

M. Guido MARTINELLI

Mme. Marie-Noëlle MINARD (président, rapporteur)

M. Achille STOCCHI

**Date de soutenance:** 23 juin 2014

# Flavor physics: precision as an avenue to discovery

Diego Guadagnoli

*LAPTh, Université de Savoie et CNRS, BP110, F-74941 Annecy-le-Vieux Cedex, France*

*E-mail: [diego.guadagnoli@lapth.cnrs.fr](mailto:diego.guadagnoli@lapth.cnrs.fr)*

The present manuscript describes work pursued in the years 2008-2013 on improving the Standard-Model prediction of selected flavor-physics observables. The latter include: (1)  $\epsilon_K$ , that quantifies indirect CP violation in the  $K^0 - \bar{K}^0$  system and (2) the very rare decay  $B_s \rightarrow \mu\mu$ , recently measured at the LHC. Concerning point (1), the manuscript describes our reappraisal of the long-distance contributions to  $\epsilon_K$  (refs. [1–3]), that have permitted to unveil a potential tension between CP violation in the  $K^0$ - and in the  $B_d$ -system. Concerning point (2), the manuscript gives a detailed account of various systematic effects, pointed out in ref. [4] and affecting the Standard-Model  $B_s \rightarrow \mu\mu$  decay rate at the level of 10% – hence large enough to be potentially misinterpreted as non-standard physics, if not properly included. The manuscript further describes the multifaceted importance of the  $B_{d,s} \rightarrow \mu\mu$  decays as new-physics probes, for instance how they compare with  $Z$ -peak observables at LEP, following the effective-theory approach of ref. [5]. Both cases (1) and (2) offer clear examples in which the pursuit of precision in Standard-Model predictions offered potential avenues to discovery. Finally, the manuscript describes the impact of the above results on the literature, and what is the further progress to be expected on these and related observables.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Historical remarks . . . . .	3
1.2	The Standard Model: gauge sector . . . . .	5
1.3	The Standard Model: Higgs/Yukawa sector . . . . .	6
1.4	Flavor physics . . . . .	9
1.5	Flavor-physics bounds on beyond-SM physics . . . . .	11
1.6	Why pursuing flavor physics . . . . .	14
1.7	Content of the manuscript . . . . .	15
<b>2</b>	<b>Indirect CP violation in the <math>K^0 - \bar{K}^0</math> system</b>	<b>16</b>
2.1	Introduction to the problem . . . . .	16
2.2	CPV in $K$ physics: theory vs. experiment . . . . .	17
2.3	Derivation of the $\epsilon_K$ formula . . . . .	19
2.4	The $\kappa_\epsilon$ correction . . . . .	22
2.5	A closer look at the operator-product expansion for $\epsilon_K$ . . . . .	25
2.6	Status of the $\epsilon_K - \sin 2\beta$ tension . . . . .	32

<b>3</b>	<b>The rare decay <math>B_s \rightarrow \mu^+ \mu^-</math></b>	<b>38</b>
3.1	$B_s \rightarrow \mu\mu$ within the SM: decay structure . . . . .	38
3.2	Experimental status, and steps towards a precise prediction . . . . .	40
3.3	Renormalization-scheme dependence . . . . .	41
3.4	Soft-photon corrections . . . . .	46
3.5	$B_s - \bar{B}_s$ oscillation effects . . . . .	49
3.6	Prediction for the non-radiative decay and error budget . . . . .	51
3.7	The $B_d \rightarrow \mu^+ \mu^-$ decay . . . . .	54
3.8	$B_s \rightarrow \mu\mu$ as a probe of the Yukawa sector . . . . .	55
3.9	$B_s \rightarrow \mu\mu$ as an EW precision test . . . . .	57
<b>4</b>	<b>Outlook</b>	<b>65</b>
4.1	Paths towards improving the accuracy of the $\epsilon_K$ prediction . . . . .	65
4.2	$B_s \rightarrow \mu\mu$ : outlook and possible new directions . . . . .	74

---

## 1 Introduction

The Standard Model denotes the theoretical framework that, to the best of our knowledge, describes all observed phenomena whose nature is other than gravitational. We make this distinction between gravitational and non-gravitational interactions from the outset in order to remark two major, distinctive differences between the corresponding theories.

The first difference is epistemological, is namely related to the quality and extent of knowledge achievable ‘in principle’ in either of the two cases. The interactions described by the Standard Model have an inherently quantum character. They occur at distances, or equivalently energy-exchange scales, such that the phenomenon to be described is modified by the observer describing it. This circumstance introduces an intrinsic uncertainty in the knowledge of certain observables, summarized by Heisenberg’s Uncertainty Principle and in the resulting discrete, or quantum, nature of these observables.<sup>1</sup> On the other hand, as far as gravity is concerned, no evidence exists for an energy regime where a quantum description becomes necessary, so the theory of gravity is formulated in terms of classical fields, at variance with the Standard Model, where a ‘quantum’ field theory is necessary. This difference is such that, in spite of huge theoretical efforts, only speculations exist on how to derive Standard Model and gravitational interactions from a common principle.

The second difference is historical. The making of General Relativity has proceeded as in the dreams of every theoretical physicist: two basic, intuitively understandable principles have been worked out to their ultimate implications using only logical necessity. The beauty, consistency and solidity of the resulting construction made its proposer have little doubt about experimental confirmations, that in fact unflinching followed. Conversely, the construction of the Standard Model has been by and large a trial-and-error process. It is maybe useful to shortly summarize this process here – this excursus<sup>2</sup> should make at least clear why the final product retained the modest name of Standard Model, which sounds more like a placeholder, rather than the definitive name of a fundamental theory.

---

<sup>1</sup> From the argument of ‘intrinsic uncertainty’ one would be tempted to trace back quantum phenomena to a probabilistic theory. Such attempts would however be frustrated by quantum phenomena like interference and entanglement. As a result, we have to date no fully deterministic understanding of quantum physics, as summarized by R.P. Feynman’s words: “Nobody understands quantum mechanics”.

<sup>2</sup> I closely follow the point of view of one of the absolute protagonists of this process, S. Weinberg. See e.g. [6].

## 1.1 Historical remarks

The history of the Standard Model (SM) traces back to the 1950s, where, after the triumphs of Quantum Electrodynamics, people were stuck with a non-renormalizable four-fermion theory to explain weak-interaction phenomena like  $\beta$  decay, and with incalculable theories for strong interactions, because the strong coupling made perturbation theory unusable. What was worse is that in either case the approach was purely phenomenological, i.e. not supported by a clear rationale. To identify a viable framework for weak interactions people tried to pursue principles of symmetry, typically with frustrating results, because most of the considered symmetries – like CP, the octet structure of hadrons, etc. – were at best only approximate, not to mention those that were badly broken, like parity. A sheer symmetry-based approach was per se not sufficient to make progress.

In this situation, among the several theoretical ideas circulating, two turned out to give a major boost towards understanding weak interactions. The first is the introduction of the idea of gauge symmetry, and its application to the construction of a renormalizable theory based on a non-Abelian group,  $SU(2)$ , by Yang and Mills [7]. It should be noted, however, that people did not immediately consider applying this idea to weak interactions, for two ‘good’ reasons. First, throughout the 50s weak interactions were thought to consist not only of vector and axial-vector contributions, but also of scalar and even tensor ones, because of a number of wrong experiments. Only by the late 50s the situation was cleared up in favor of only vector and axial-vector contributions [8], that Yang-Mills (YM) gauge bosons are suitable to mediate. A second, more serious obstacle, was however the problem of how to give mass to YM gauge bosons – massless vector bosons would surely have been seen by experiment. Trying to force mass terms by hand would destroy the gauge symmetry, and make the YM theory non-renormalizable, hence it would lead to no practical advance over the initial four-fermion theory.

A second major idea came to the rescue of this difficulty, namely the idea that a Lagrangian symmetry may not be a symmetry of the vacuum, or the idea of spontaneous symmetry breaking (SSB). When applied to global symmetries, this idea implied the presence of massless and spinless bosons, by the Goldstone theorem [9]. At first sight, this conclusion seemed to prevent applicability of SSB to weak interactions (whereas in strong interactions it opened the long process that culminated in low-energy QCD, or chiral perturbation theory). But an exception to Goldstone’s ‘no-go’ argument was found by Higgs [10] as well as Englert and Brout [11] to be the case where the spontaneously-broken symmetry is local, rather than global. In this case, the Goldstone bosons provide the zero-helicity degrees of freedom of the YM vector bosons, that become massive. Consideration of  $SU(2) \times U(1)$  as the correct gauge-symmetry group was – according to Weinberg’s account [6] – largely due to the fact that, with one generation of leptons as the only fermionic ingredient, this group was basically the only possibility. (This statement makes by the way evident that quarks and generations were included afterwards.)

The main leftover conceptual issue in the establishment of the SM was its renormalizability, namely the possibility of making the theory finite by just a finite number of counterterms, irrespective of the perturbation-theory order. This requirement was *not necessary*, because at any finite order a non-renormalizable theory is just as good as a renormalizable one: once the infinities present at that order are subtracted, the theory is still usable to make predictions. Nonetheless, the requirement of renormalizability was attractive in that it would make the electroweak theory predictive even way above the scale of electroweak gauge bosons’ masses. And besides, renormalizability would provide a rationale for allowed (those with energy dimension  $\lesssim 4$ ) vs. not-allowed interactions, thus greatly narrowing model-building possibilities. As a matter of fact, when the difficult proof of renormalizability of

spontaneously-broken gauge theories was delivered [12], the theory that now is the SM grew significantly in popularity, and people started to seriously consider testing it in detail. The discovery of the  $W, Z$  bosons [13] and the measurement of their mass in accord with theory expectations crowned, and to some extent concluded, this long process.

The actual conclusion still needed establishing the existence of at least one ‘Higgs’ boson. In its minimal realization, spontaneous symmetry breaking within an  $SU(2) \times U(1)$  gauge group required an  $SU(2)$  doublet of scalars, namely four degrees of freedom. Out of them, three would be the longitudinal modes of the  $W$  and  $Z$ , and the leftover one must then be a physical spinless state. To guide experimental searches, the mass of this state can be constrained in several ways. First, the Higgs mass enters in the loop corrections to electroweak-precision observables, that have been accurately measured at LEP. By global fits, one obtains rather accurate indications on the Higgs-mass range, e.g.  $M_h = 85_{-28}^{+39}$  GeV (see [14] and references therein).<sup>3</sup> However, the overall quality of the fit is rather poor: for example leptonic asymmetries favor a light Higgs, whereas hadronic asymmetries a heavy one, and the O(100) GeV central value for the Higgs mass results only from the average between the two sets of measurements [14]. To obtain a more ‘conservative’ Higgs-mass range, one can then resort to ‘minimal’ theoretical arguments, such as triviality – namely the requirement of positivity and finiteness of the quartic self-coupling in the Higgs potential, imposing an upper bound on the Higgs mass – and stability – i.e. the requirement that the scalar potential be bounded from below, imposing a lower bound on the Higgs mass. Combinedly, these two constraints yield the range  $130 \text{ GeV} \lesssim M_h \lesssim 180 \text{ GeV}$ , if one assumes the SM to be valid up to grand-unification scales of the order of  $10^{16}$  GeV, or else  $50 \text{ GeV} \lesssim M_h \lesssim 800 \text{ GeV}$ , assuming instead TeV-scale new physics. It is clear that the most conservative range for experimental searches is the latter, apart from possibly replacing the lower bound by 114.4 GeV, obtained from direct searches at LEP [16].

Within this mass range, the SM Higgs phenomenology changes dramatically: from a very narrow state, dominated by the  $h \rightarrow b\bar{b}$  mode in the low-mass range till about 130 GeV, to a broader and broader state above the two- $Z$  threshold, and dominated by  $h \rightarrow VV$  decays [17]. As a matter of fact, it took 30 more years after CERN UA1, for the Large Hadron Collider (LHC) experiments, again at CERN, to discover [18] a scalar that, to the still limited extent of data available, seems to indeed correspond to the single scalar expected within the SM, ‘the Higgs’.

To make contact with the statements at the beginning of this chapter, we observe that it is difficult to imagine that the idea of a spontaneously-broken gauge symmetry, and its specific  $SU(2) \times U(1)$  realization, could have emerged in the absence of solid and well-digested experimental data. It is fair to say that, while minimalistic in several respects, our current theory of electroweak interactions<sup>4</sup> is not nearly as compelling as General Relativity. It is instead a model whose form was progressively sculpted by data, and that, as time went by, was found to be consistent with all the further data available. This success was not to be expected on the basis of logical necessity, or model aesthetics or, for that matter, even phenomenology itself, considering e.g. the established existence of Dark Matter.

---

<sup>3</sup> This ‘indirect measurement’ may actually appear rather loose. The reason is the fact that the Higgs mass enters only logarithmically in one-loop observables. A quadratic Higgs-mass dependence is possible at two loops, but it is then ‘screened’ by two extra powers of the electroweak coupling. This argument is known as the “Veltman screening theorem” [15].

<sup>4</sup> I expressly leave the case of QCD out of this argument.

## 1.2 The Standard Model: gauge sector

It is at this point useful to introduce explicitly the electroweak part of the SM Lagrangian, in order to better identify its most ‘accidental’ components. The SM Lagrangian may be divided into a gauge part and a Higgs/Yukawa part. The gauge part is the ensemble of all the dimension-4 terms that one can write for spin-1/2 fermions and spin-1 gauge bosons (and no scalars) from the sheer requirement of gauge invariance. It is composed by one or more pieces of the following form

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\sum_j\bar{\Psi}_j\not{D}\Psi_j, \quad (1.1)$$

with  $\Psi$  a fermion field,  $D_\mu$  its gauge-covariant derivative, and  $F_{\mu\nu}$  the field strength of the gauge field.<sup>5</sup> Gauge invariance allows gauge fields only in the combination  $F_{\mu\nu}$ , and an arbitrary number of covariant derivatives  $D_\mu$  acting on any of the allowed fields. To obtain the specific gauge Lagrangian chosen by Nature one still has a twofold arbitrariness: the choice of the gauge-symmetry group – in the SM not even a simple group, but the product  $SU(3)_c \times SU(2)_L \times U(1)_Y$  – and the choice of the fermions and their representations under the gauge group.

In spite of this arbitrariness, the gauge-symmetry principle and the resulting Lagrangian can be likened in simplicity to the principle of invariance under general coordinate transformations and the resulting theory of gravity, General Relativity (GR). In fact between the two theories there is even a close analogy at the level of the formalism: gauge symmetry is the requirement to matter fields of local invariance under phase redefinitions, which parallels GR’s diffeomorphism invariance; in addition, since the Lagrangian must contain  $\partial_\mu\Psi$  terms in order to account for dynamics, gauge invariance implies the existence of gauge fields. This argument parallels the existence of GR’s metric tensor. As a matter of fact, the SM *gauge* Lagrangian is theoretically appealing in several respects. The gauge symmetry makes it natural, in the sense pointed out in ref. [20]. Because of this feature, quantum corrections remain corrections: they arise as logs of the ultraviolet scales running in the loops. As a result, the theory has a limited ultraviolet sensitivity. Finally, since the gauge symmetry is implemented identically across fermions of different generations, the gauge Lagrangian enjoys symmetry under a large group of global field transformations, the flavor group, to which I will shortly return.

We would like to further comment on the absence in eq. (1.1) of terms with mass-dimension other than 4. The absence of terms with dimension strictly less than 4 will be justified at the beginning of the next section. As regards gauge-invariant terms with energy dimensions  $E^{4+n}$ , with  $n > 0$  hence suppressed by a mass scale  $1/\Lambda^n$ , we note that, in a process characterized by a momentum scale  $p$ , these terms would scale like  $(p/\Lambda)^n$ . As such, these terms are ruled out by the requirement of renormalizability, already mentioned in the previous section, equivalent to the assumption that the scale  $\Lambda$  is much above  $p$ , implying in turn that their effect is vanishingly small. The only way to validate this assumption is to verify that data do not require Lagrangian terms with  $n > 0$ , as seems to be the case up to the present. Independently from this validation, it should be remarked that renormalizability

<sup>5</sup> In principle, in eq. (1.1) we should include also terms of the form  $F_{\mu\nu}\tilde{F}^{\mu\nu} \equiv \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}$ , that incidentally violate the CP and T symmetries. These terms can actually be rewritten as a total derivative, and as such they do not affect field equations, nor Feynman rules, hence they are irrelevant in perturbation theory. However, these terms have non-perturbative quantum-mechanical effects, arising from the existence of extended field configurations that are also stationary points of the action (see [19]). While these configurations do have physical effects both in strong and electroweak interactions, such effects are well beyond the scope of the present manuscript.

is not a fundamental physical requirement. In fact, any realistic quantum field theory is expected to contain renormalizable as well as non-renormalizable terms [21]. Concerning renormalizability, it is actually interesting to contrast once more the cases of the SM and of gravity, to appreciate how diametrically ‘opposed’ their structures are in this respect. First, GR does not contain renormalizable interactions at all; furthermore, if cancellation of divergences due to virtual gravitons were a relevant requirement in GR, then the Lagrangian would have to contain all interactions allowed by symmetries, not only those among gravitons, but also those involving *any* other particle. These considerations support the conclusion that the success of the renormalizability requirement in the SM means only that the cutoff scale  $\Lambda$  introduced above is much larger than the scales at which the model has been tested [21].

### 1.3 The Standard Model: Higgs/Yukawa sector

It turns out that, albeit elegant, a pure gauge Lagrangian as in eq. (1.1) has too much symmetry to describe the real world. In the first place, it does not allow for the presence of the gauge field  $A_\mu$ , except within the field-strength tensor  $F_{\mu\nu}$ , or within a covariant derivative. Therefore, explicit mass terms for the gauge fields are ruled out. In the case of the SM gauge group, they are likewise ruled out for fermions. As mentioned, one can remedy this situation by breaking the gauge symmetry spontaneously, namely by introducing a scalar multiplet whose interactions formally respect the gauge symmetry, but such that the resulting vacuum breaks it. Within the SM one introduces a single scalar multiplet  $\Phi$ , in the smallest representation compatible with the gauge symmetry, with Lagrangian

$$\mathcal{L}_{\text{Higgs}} = D_\mu \Phi^\dagger D^\mu \Phi - V(\Phi) , \quad (1.2)$$

where  $\Phi$  transforms as  $(\mathbf{1}, \mathbf{2}, 1/2)$  under the SM gauge group. Once the assumption of a single multiplet is made, all its gauge quantum numbers are fixed by the requirements of leaving unbroken the gauge currents of the  $SU(3)$  sector, as well as one uncharged gauge current of the  $SU(2)_L \times U(1)_Y$  sector. These quantum-number assignments rule in turn the explicit form of the covariant derivatives in eq. (1.2), namely the Higgs-gauge interactions. Concerning  $V(\Phi)$ , one may write down any potential such that the vacuum breaks the gauge symmetry. The simplest choice, assuming only renormalizable terms and the absence of  $\Phi$ -odd ones, is

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \frac{\lambda}{2} |\Phi^\dagger \Phi|^2 . \quad (1.3)$$

whereby the wrong-sign mass term causes the theory vacuum to occur at any point such that  $\langle \Phi^\dagger \Phi \rangle = \mu^2/\lambda$ , rather than at zero, whereas  $\lambda = O(1)$  causes the vacuum expectation value (vev) to be  $O(|\mu|)$ .

To summarize, spontaneous breaking of the gauge symmetry forces one to make at least two further arbitrary choices besides those inherent to the gauge Lagrangian: the choice of the number and transformation properties of the Higgs multiplets, and that of the explicit form of the SSB potential. Sticking to minimality corresponds to the SM case, and this case seems at present to be supported by the data collected at the 7-8 TeV runs of the LHC.

In the SSB discussion, we have not yet mentioned the problem that also fermions  $\Psi$  need to be given mass, which introduces further arbitrariness. As mentioned above, within the SM, one cannot simply introduce fermion mass terms as in QED, because this would break gauge invariance. Still, one can form gauge-invariant, renormalizable interactions that are bilinear in fermions and linear in Higgs fields. Among these ‘Yukawa’ interactions, terms (necessarily charge-neutral) proportional to the Higgs vev will provide the required mass



terms. It turns out that, within the SM, one can give mass to all the fermions with just the  $\Phi$  multiplet introduced above. Namely one may write

$$- \mathcal{L}_Y = \bar{Q}_L Y_u \Phi^c u_R + \bar{Q}_L Y_d \Phi d_R + \bar{L}_L Y_\ell \Phi e_R + \text{h.c.} , \quad (1.4)$$

where  $Q_L, u_R, d_R, L_L$  and  $e_R$  are the usual SM quark and lepton fields, those in capital letters denoting  $SU(2)_L$  doublets:  $Q_L \equiv (u_L, d_L)^T$  and  $L_L \equiv (\nu_L, e_L)^T$ . The subscript  $L, R$  indicates that the field is left- or right-handed, respectively. In eq. (1.4)  $\Phi^c = +i\sigma_2 \Phi^*$  (it can be thought as charge conjugation, albeit  $(\Phi^c)^c = -\Phi$ ). Finally,  $Y_{u,d,\ell}$  denote  $n_g \times n_g$  complex matrices, where  $n_g = 3$  is the number of replicas, or generations of the  $Q, u, d, L, e$  fields that it is necessary to introduce in order to account for all the quarks and leptons observed.

Several comments are in order on eq. (1.4).

1. **Neutrinos.** I stated above that Yukawa interactions can give mass to all fermions, yet those introduced in eq. (1.4) will not give mass to neutrinos. In fact, one may augment eq. (1.4) by the term  $\bar{L}_L Y_\nu \Phi^c \nu_R$ , where  $\nu_R$  denotes right-handed neutrinos, neutral with respect to all SM interactions. This solution, while minimal, is not theoretically appealing in that, to get neutrino masses in the right ballpark, one needs  $|(Y_\nu)_{ii}| \sim 10^{-12}$ ; in addition, gauge invariance would allow Majorana mass terms of the form  $M_N \nu_R^2 / 2$  (albeit they break lepton number). One notes however, that if *both* the Yukawa and the Majorana mass terms are present, and if  $M_N \gg |(Y_\nu)_{ij}|v$ , then diagonalizing the neutrino mass matrix yields three almost purely right-handed heavy neutrinos, and three almost purely left-handed neutrinos, with small masses tuned by the ratio  $(|(Y_\nu)_{ij}|v)^2 / M_N^2$ . This ‘see-saw’ mechanism [22], that can actually be implemented in two another ways depending on the extra-fields invoked, explains easily and naturally the observed pattern of neutrino masses and mixings, and also fits in SM extensions with grand unification (in this respect see also point 3 below). However, neutrino physics lies outside the topic of this manuscript, and will not be touched upon any further (for a beautiful review, see [23]).
2. **Why 3 generations.** Nobody knows why the SM fermions occur in 3 generations, and not simply one. The quip “Who ordered that?” by I.I. Rabi as the existence of a ‘heavy’ electron, the muon, had been established, remains the best way to comment on the apparent incongruousness of having around several replicas of the very same gauge representation, with namely identical quantum numbers, and differing only in mass. It is also to be noted that to describe ordinary matter it is by and large sufficient to invoke only first-generation quarks and leptons. The heavier generations are unstable states produced only in high-energy collisions or in certain astrophysical processes. At a more theoretical level, the number of generations, or rather the number of flavors  $n_f = 2n_g$ , actually enters in several places where it could affect the theory consistency, notably in the  $\beta$ -function of QCD, that governs the strength and the increase/decrease pattern of the strong-interaction coupling depending on the process energy. However, in practice, none of these known dependences makes  $n_f = 6$  more compelling than any other value between 2 and 16.
3. **Gauge anomalies.** On the other hand, within each generation, the necessity of the five SM fermion multiplets may be justified by the cancellation of all gauge anomalies. Anomalies are only possible for gauge theories including  $SU(n \geq 3)$  or  $U(1)$  factors, as is the case within the SM [6]. In this case, in order to ensure the absence of anomalies, one needs to resort to cancellations among different matter multiplets (quark and leptons within the SM). The SM  $Q_L, u_R, d_R, L_L, e_R$  multiplets, with their assigned

quantum numbers, turn out to serve this purpose. Most of the SM anomalies are actually zero for gauge-algebraic reasons; however, this is not true for the  $SU(3)^2-U(1)$ ,  $SU(2)^2-U(1)$  and  $U(1)^3$  anomalies, that nonetheless turn out to be zero, for rather fortuitous-looking cancellations. The only way, that I am aware of, to understand the occurrence of these cancellations at a deeper level is based on the observation that the SM matter multiplets, plus an additional gauge singlet, fit a complete 16-dimensional spinor representation of  $SO(10)$ , and that the SM gauge group may be embedded in  $SO(10)$  itself, which is anomaly-free. Unfortunately, this neat observation lacks experimental support to date.

4. **Mass-eigenstate basis and flavor/CPV physics.** Kinetic vs. mass terms for fermions appear in eq. (1.1) and respectively (1.4). There is no a priori reason why the two sets of terms should be flavor-diagonal in the same field basis, since  $Y_{u,d,\ell}$  are as mentioned generic complex matrices. A generic complex matrix  $m$  can be put in diagonal and real form by the transformation  $AmB$ , with  $A$  and  $B$  unitary matrices (see e.g. [19]). These unitary matrices can be introduced as fermion-field redefinitions. One would need four such unitary matrices in order to diagonalize both  $Y_u$  and  $Y_d$ , whereas in eq. (1.4) these two couplings are multiplied on the left by the same field  $Q_L = (u_L, d_L)^T$ . The  $Y_u$  and  $Y_d$  diagonalization is therefore performed after breaking the  $SU(2)_L$  part of the EW symmetry via the field redefinitions  $u_{L,R} = \mathcal{U}_{L,R} \hat{u}_{L,R}$ ,  $d_{L,R} = \mathcal{D}_{L,R} \hat{d}_{L,R}$ , where the hatted fields represent the mass-eigenstate basis, and  $\mathcal{U}, \mathcal{D}$  unitary matrices. Since the  $u_L$  and  $d_L$  components of  $Q_L$  are ‘rotated’ differently, charged currents within the kinetic term  $i\bar{Q}_L \not{D} Q_L$  will not be flavor-diagonal in the mass-eigenstate basis. In particular

$$\mathcal{L}_{\text{gauge}} \supset i\bar{Q}_L \not{D} Q_L \supset \frac{g}{\sqrt{2}} \bar{u}_L \not{W} d_L + \text{h.c.} = \frac{g}{\sqrt{2}} \bar{\hat{u}}_L \mathcal{U}_L^\dagger \mathcal{D}_L \not{W} \hat{d}_L + \text{h.c.} . \quad (1.5)$$

The unitary matrix  $\mathcal{U}_L^\dagger \mathcal{D}_L \equiv V_{\text{CKM}}$  is the so-called CKM matrix [24]. The fact that  $V_{\text{CKM}}$  can be, and in fact is, different than the identity matrix implies that the  $W$ -mediated currents in eq. (1.5) will mix quarks of different generations, i.e. give rise to flavor-violating interactions.

These interactions turn out to also violate the discrete CP symmetry. The latter cannot be violated if, by a suitable field basis, all the complex phases in  $Y_u$  and  $Y_d$  can be made to disappear. This possibility depends on the CKM dimensionality. Being a unitary matrix, the CKM for  $n_g$  generations can be parameterized in terms of  $n_g^2$  parameters. Out of them  $n_g(n_g - 1)/2$  are Euler angles, as in the orthogonal group  $O(n_g)$ , and the rest are phases. Out of these phases,  $2n_g - 1$  can be absorbed as relative phases between the quark fields (the overall quark-field phase can not, because it remains a symmetry of the SM Lagrangian). As a consequence, for  $n_g = 3$  the CKM matrix will contain one non-trivial phase.

Therefore, flavor and CP violations are, within the SM, inextricably intertwined. There is to date no consensus on why the amount of flavor and CP violation within the SM is what it is and not otherwise, not even an anthropic argument.<sup>6</sup> It is nonetheless noteworthy that the interaction in eq. (1.5) gives rise to a tremendous amount of flavor and/or CPV phenomena, to a few of which is dedicated the rest of this manuscript.

---

<sup>6</sup> For example, the SM amount of CP violation seems insufficient to account for baryogenesis, albeit there is no consensus on this matter either (see e.g. [25]).

## 1.4 Flavor physics

The above points are meant to demonstrate on the one hand the amount of arbitrariness introduced in particular by the Lagrangian terms in eq. (1.4) – as a matter of fact, and at variance with gauge interactions, little is said in textbooks about Yukawa interactions at the level of first-principle arguments. At the very same time, the above items offer the opportunity to emphasize again the enormous richness brought about in particle-physics phenomenology by the interactions in eq. (1.4). This richness can be traced back to the specific parametrics of the Yukawa couplings, namely:

- **The peculiar values of quark masses.** Concerning, first of all, the Yukawa eigenvalues, it is to be noted that those in the quark sector are hierarchically separated from one another. The  $u, d, s, c, b, t$  quark masses are  $1.3 \cdot 10^{-5}, 2.8 \cdot 10^{-5}, 5.5 \cdot 10^{-4}, 7.3 \cdot 10^{-3}, 2.4 \cdot 10^{-2}, 1$  [26] in units of the reduced Higgs vev of  $246/\sqrt{2}$  GeV.<sup>7</sup> The fact that quark masses are hierarchically different has far-reaching implications, for example it implies a very different phenomenology than neutrinos: quark wave functions are immediately separated after production, hence quarks will not oscillate into one another, at variance with neutrinos (see e.g. [23]). In turn, well-defined quark flavors imply the possibility of a whole zoo of flavored hadrons, with different phenomenology and even different theoretical treatments according to their masses and widths.
- **The peculiar form of the CKM matrix.** Also the CKM matrix, quantifying the ‘mismatch’ in the flavor-group space between the  $Y_u$  and  $Y_d$  matrices, turns out to be highly non-generic, very close to the unit matrix, and with entries becoming hierarchically smaller the higher the distance from the diagonal. This circumstance is best visualized in the so-called Wolfenstein parameterization [27] of the CKM matrix, whereby the small deviations from the identity matrix are described by a small parameter,  $\lambda \simeq \sin \theta_C \simeq 0.23$ , where  $\theta_C$  is the Cabibbo angle.<sup>8</sup> In this parameterization, CKM entries have the following magnitudes

$$|V_{CKM}| \approx \begin{pmatrix} 1 & \lambda & \text{O}(\lambda^3) \\ -\lambda & 1 & \text{O}(\lambda^2) \\ \text{O}(\lambda^3) & \text{O}(\lambda^2) & 1 \end{pmatrix}. \quad (1.6)$$

Among the CKM entries, the off-diagonal ones dial the rates of occurrence of flavor-violating processes. Since these entries are smaller than 1, especially those involving the third together with one of the light generations, flavor-violating processes come with a CKM suppression with respect to flavor-conserving ones. Besides, it is to be noted that, within the SM, it is possible to build flavor-violating, but charge-neutral interactions only at the loop level. These processes, so-called flavor-changing neutral currents (FCNCs) come therefore with an additional loop suppression. Finally, since flavor violation occurs only via weak interactions, flavor-violating rates are further suppressed by factors of  $p^2/M_W^2$ , where  $p$  equals the typical masses of the external flavored mesons,  $p \lesssim 5$  GeV. Because of this triple suppression mechanism, FCNC decays are usually rare, sometimes very rare, processes. As such, they are ideal probes of possible UV effects not described by SM interactions and manifesting themselves in loops.

<sup>7</sup> It is to be noted that the top-mass eigenvalue is compatible with one. None of the many efforts attempted to make sense of this circumstance has found support from further data.

<sup>8</sup> As emphasized in [28], the Cabibbo theory of hadronic currents had extended to the weak decays of strange particles the idea of universality pioneered by E. Fermi and later developed and fully established in [8].

The two items above highlight the very non-generic structure of the SM flavor sector. Besides the absence of a fundamental understanding of this structure, there is no apparent reason why the same flavor structure should also hold for physics beyond the SM. This simple fact motivates the following discovery strategy:

*look for however small, but clear-cut, deviations from the SM pattern of predictions in selected processes, that should include in particular:*

(a) *very rare, very clean decays and* (b) *CP violation.*

Concerning CP violation, one may raise two objections to the opportunity of pursuing this kind of measurements: first, that it is by now established that the CKM phase explains the bulk of low-energy measured CP violation. Second, that anyway the CKM phase is, probably, not sufficient to explain baryogenesis within the SM, as already remarked. Both these objections are cogent. At the same time, two important facts should however be kept in mind. First, that we do not measure the CKM phase better and better for its own sake, but in order to find, sooner or later, deviations from the CKM-phase-only scenario. Even small deviations from this scenario may provide major hints on CP-violating dynamics at the high scale, and give us precious model-building information on the SM UV completion.<sup>9</sup> Second, that usually CP-violating observables are constructed from asymmetries between dimensioned quantities, e.g. decay rates. In these asymmetries, many SM uncertainties cancel. Hence CP-violating observables usually offer very clean tests. One may even quote the example of charm physics, where CP violation is the *only* clean sector.

This remark on CP violation being made, several further arguments may be raised in support of the overall strategy stated above, the first being historical. It is a fact that, in recent high-energy physics history, direct discoveries have typically been anticipated by indirect effects in loops. This ‘double’ discovery may actually be seen as one of the most spectacular, compelling aspects of HEP progress, testifying a healthy interdependence between theory and experiment, and the robustness of their respective methods. The story of the so-called GIM mechanism [29], that reconciled the stringent experimental limit on the  $K^0 \rightarrow \bar{K}^0$  transition with theory by the introduction of a new quark, the charm, provides in this respect a very illustrative example, and deserves to be shortly recalled explicitly (I will closely follow [28]).

As already mentioned above, the Fermi theory was plagued by divergences, and people were working on the problem of understanding this theory in terms of a better-behaved, if possible renormalizable theory. This task was not easy, because different sets of quantities pointed to different missing pieces of the puzzle towards the SM. A clear identification of the problem behind the need for the charm quark was provided by the calculation [30] of several amplitudes with exchange of two weak bosons, like the amplitudes for  $K_L \rightarrow \mu^+ \mu^-$  and  $K^0 - \bar{K}^0$  mixing, in the context of the then-existing theory with three quarks, the  $u, d$  and  $s$ . This calculation revealed quadratically divergent amplitudes,<sup>10</sup> that therefore had to be regulated by a cutoff  $\Lambda$ . The stringent limit on the  $K^0 \rightarrow \bar{K}^0$  transition provided a *peculiarly low cutoff*:  $\Lambda_{K^0-\bar{K}^0} \simeq 3$  GeV. Several solutions to this issue were put forward. The authors of [29] came out with the proposal of a fourth quark, the charm,<sup>11</sup> coupled through charged

**The GIM  
mechanism**

<sup>9</sup> As an example, an even small new phase in the low-energy Wilson coefficient of the SM magnetic operator, entering several flavor observables, can be easily related via renormalization-group running to the Wilson coefficient of a more exotic operator such as the chromomagnetic one, that however is very common e.g. in supersymmetry.

<sup>10</sup> It is easy to reproduce these divergences by calculating the underlying box diagrams without the charm.

<sup>11</sup> Note that this proposal was at the time less obvious, and more bold, than it may appear to be today, because for one thing a large part of the community was skeptic about the reality of quarks, until the dual (confining/asymptotically free) behavior of strong interactions was finally established.

weak currents to a combination of the  $d$  and  $s$  quarks orthogonal with respect to the one the up quark couples to. This implied that for every contribution from the up quark there was a corresponding contribution from the charm with opposite sign. (This can easily be seen from eq. (1.6): in  $K^0 - \bar{K}^0$  mixing the external quarks are  $d$  and  $s$ . An intermediate up quark means a contribution proportional to the entries 11 and 12 of the CKM matrix, yielding  $1 \cdot \lambda$ , whereas an intermediate charm quark means a contribution from entries 21 and 22, which just differs in sign.) For exactly equal up and charm masses the two contributions would exactly cancel, whereas for different masses the amplitude would be proportional to the difference  $m_u^2 - m_c^2$ , which then replaced the quadratic cutoff dependence. The  $\Lambda_{K^0-\bar{K}^0}$  estimate then turned into a prediction for the charm mass. On the theoretical side, the introduction of the charm quark substantially helped in the identification of the  $SU(2)_L$  structure of weak interactions, as now quarks could be accommodated in two doublets, thus restoring the quark-lepton symmetry, in turn helping towards the cancellation of gauge anomalies. On the experimental side, the GIM paper provided a mass scale for the charm quark, along with its expectedly rich decay phenomenology, which in turn gave guidance to experimental searches. After various partial pieces of evidence, the first unequivocal evidence for a  $c\bar{c}$  state took place with the discovery of the  $J/\psi$  particle [31].

It is hard to find, even in particle physics, a more neat example of theory-experiment feedback than the one initiated by the GIM paper.

## 1.5 Flavor-physics bounds on beyond-SM physics

The GIM-mechanism/charm-discovery example is so often quoted because it is literally a prototype of the theory-experiment interplay mentioned above. Its physics line of argument starts from the observation that one measurable quantity has too strong a sensitivity to the UV cutoff, which suggests that a symmetry is being missed. This symmetry calls for a new particle, whose mass replaces the cutoff sensitivity. This argument applies identically to very actual questions, such as supersymmetry as a solution to the Higgs-sector ultraviolet sensitivity. Analogous examples, albeit perhaps less ‘didactic’ in their historical development, may be provided for the top quark, the  $W, Z$  bosons and even the Higgs scalar.

All the mentioned examples show that, by and large, new particles are typically announced by indirect effects, in FCNC loops or elsewhere, before they are directly produced in collisions performed at their mass scale. Indirect observables are thereby able to probe scales much larger (sometimes by several orders of magnitude) than the energy scale at which they are measured. In this respect, indirect observables have a competitive advantage over direct searches. The constraining power of indirect observables on beyond-SM interactions can be appreciated most generally by taking an effective-theory approach, in which the SM Lagrangian is augmented by a tower of effective interactions with energy dimension  $d = 4 + p$ , with  $p > 0$  and suppressed by a power  $-p$  of the respective UV cutoff. One namely assumes

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda_i^p} \mathcal{O}_i^{[4+p]}. \quad (1.7)$$

In this expansion,  $i$  simply labels all the effective operators, whereas the superscript  $[4 + p]$  specifies that the dimension of the given operator exceeds 4 by  $p$ . Note that the cutoffs of the various operators do not need to be the same, hence they are also labelled by  $i$ . One usually constructs the  $\mathcal{O}_i$  out of SM fields only. This approach is justified, since the lower bounds on the  $\Lambda_i$  that one obtains correspond to scales typically much larger than the mass of the heaviest among the d.o.f. building up the  $\mathcal{O}_i$ . The  $c_i$  are unknown coefficients, that depend on unspecified UV dynamics that has been integrated out. In this sense, the

non-renormalizable part of eq. (1.7) is just a generalization of the Fermi theory.<sup>12</sup>

This effective-theory approach has the basic advantage of being able to capture any beyond-SM effect (provided it is consistent with the symmetries assumed for the operators  $\mathcal{O}_i$ ) in terms of a small number of new couplings. The main drawback is that, as argued, these couplings are unknown. This lack of knowledge implies that what one actually constrains in this approach is not  $\Lambda_i$ , but rather the ratio  $\Lambda_i/c_i^{1/p}$ . Correlations among different  $c_i$  are likewise unknown. As a consequence, the different  $c_i$  are typically treated as independent quantities, or else switched on one by one in phenomenological analyses. While this is the best one can do, in general neither of these two choices provides necessarily an accurate approximation of reality.

With these caveats in mind, it is useful to get a more quantitative idea of the severity of FCNC constraints on the  $\Lambda_i$ . To this end, the Lagrangian (1.7) may be applied to transitions where a given flavor changes by two units between the initial and the final state, the so-called  $\Delta F = 2$  processes. Within the SM, these processes are mediated by box diagrams like the one depicted in fig. 1, and are loop-, CKM-, and  $p^2/M_W^2$ -suppressed. In principle there is also a GIM suppression, because two different quarks  $q_i$  and  $q_j$  in a given internal line amount to a dependence of the kind  $m_i^2 - m_j^2$ , as already noted, hence to a partial cancellation. However, this argument holds mostly for  $K^0 - \bar{K}^0$  mixing, whereas in the cases of  $B_{d,s} - \bar{B}_{d,s}$  mixings, the amplitude is totally dominated by the top-top contribution, hence there is basically no GIM cancellation at all.

Accurate data exist to constrain  $\Delta F = 2$  amplitudes: the measurements of mass splittings between the mass eigenstates of the  $K^0 - \bar{K}^0$  and the  $B_{d,s} - \bar{B}_{d,s}$  systems, as well as CP violation, quantified by  $\epsilon_K$  for the  $K^0 - \bar{K}^0$  system, and by  $\sin 2\beta_{(s)}$  for the  $B_{d(s)} - \bar{B}_{d(s)}$  systems. Full details on the analyses as well as updated bounds can be found in [32–34]. To make an explicit example, one may consider the case of a shift to just the  $\Delta F = 2$  operator allowed within the SM, namely<sup>13</sup>

$$\Delta\mathcal{L}^{\Delta F=2} = \sum_{i \neq j} \frac{c_{ij}}{\Lambda^2} (\bar{Q}_{Li} \gamma^\mu Q_{Lj})^2, \quad (1.8)$$

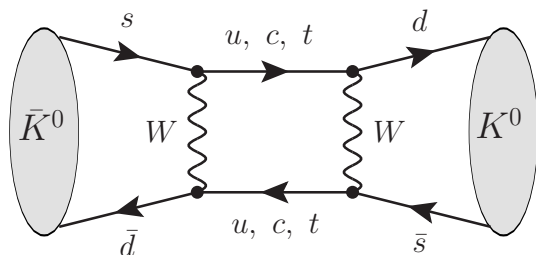
where the indices  $i, j$  label the external states:  $sd, bd$  and  $bs$  for  $K^0 - \bar{K}^0$ ,  $B_d - \bar{B}_d$  and  $B_s - \bar{B}_s$  respectively. From the (rather conservative) requirement that the new-physics amplitude does not exceed in magnitude the SM one,  $|\mathcal{A}_{\text{NP}}^{\Delta F=2}| < |\mathcal{A}_{\text{SM}}^{\Delta F=2}|$ , one arrives at constraints as severe as [33]

$$\Lambda > \frac{4.4 \text{ TeV}}{|V_{ti}^* V_{tj}|/|c_{ij}|^{1/2}} \sim \begin{cases} 1.3 \times 10^4 \text{ TeV} \times |c_{sd}|^{1/2} \\ 5.1 \times 10^2 \text{ TeV} \times |c_{bd}|^{1/2} \\ 1.1 \times 10^2 \text{ TeV} \times |c_{bs}|^{1/2} \end{cases}. \quad (1.9)$$

The by far strongest bounds are those from the  $K^0 - \bar{K}^0$  system, that therefore is generically regarded as the most sensitive probe of non-SM contributions. As anticipated above, these

<sup>12</sup> An alternative approach consists in taking an explicit extension of the SM, whereby the extra d.o.f. and their dynamics are completely specified. This approach is generally more predictive, because the short-distance couplings  $c_i$  are in principle computable, but conclusions are model-specific.

<sup>13</sup> Beyond the SM, one can in general write down a total of 8 operators with the SM field content.



**Figure 1:** Example of a box diagram responsible for  $\bar{K}^0 - K^0$  oscillations. One can obtain the case of  $\bar{B}_d - B_d$  oscillations by the replacement  $s \rightarrow b$ . From the latter, one can obtain the case of  $\bar{B}_s - B_s$  oscillations by the further replacement  $d \rightarrow s$ .

bounds depend on the magnitude assumed for the Wilson coefficients. Given the very non-generic structure of the SM flavor violation, there is no reason to assume that the same structure holds also beyond the SM. Therefore, a natural choice seems to be to take the new-physics Wilson coefficients of  $O(1)$ . In this case however, according to the bounds in eq. (1.9), flavor-violating new physics is allowed only way above the TeV scale – in particular, from the  $K^0 - \bar{K}^0$  system it is allowed at scales not lower than 10000 TeV. This circumstance is known as the flavor problem: if we insist that beyond-SM flavor effects should emerge at the TeV scale, then we have to conclude that the underlying new physics possesses a highly non-generic flavor structure [33].

Several ways have been put forward to circumvent the flavor problem. The most drastic one is perhaps that of invoking minimal flavor violation (MFV).<sup>14</sup> The latter is the assumption that the SM sources of flavor breaking, in the quark sector the  $Y_u$  and  $Y_d$  matrices, are the only sources of flavor breaking also beyond the SM.<sup>15</sup> This assumption can be implemented in the effective Lagrangian in eq. (1.7) via the following steps:

**MFV**

1. Identify the ‘flavor group’, namely the group of global transformations that the Yukawa couplings break. In the absence of Yukawa interactions (1.4), the gauge Lagrangian enjoys symmetry under a group of unitary transformations as large as  $G_F \equiv U(3)_{Q_L} \times U(3)_{u_R} \times U(3)_{d_R} \times U(3)_{L_L} \times U(3)_{e_R}$ , where the subscript indicates the fermion field on which the symmetry acts. This group is broken completely by the Yukawa interactions, apart from three  $U(1)$ ’s, that can be identified with baryon number, lepton number, and hypercharge [35]. The part of this group relevant to quark flavor physics is  $G_q \equiv SU(3)_{Q_L} \times SU(3)_{u_R} \times SU(3)_{d_R}$ .
2. Attribute to the SM Yukawa couplings formal transformation properties under  $G_q$ , so as to recover invariance of Yukawa interactions under this group. By looking at interactions (1.4), one immediately sees that formal  $G_q$  invariance is recovered for  $Y_{u,d}$  transforming as  $Y_u \sim (3, \bar{3}, 1)$  and  $Y_d \sim (3, 1, \bar{3})$ .
3. Accordingly identify the transformation properties required to any beyond-SM flavor structure in eq. (1.7), for the corresponding interaction to recover formal invariance under  $G_q$ . Then express this flavor structure in terms of  $Y_u$  and  $Y_d$ , as mentioned the only  $G_q$ -breaking sources within MFV.

This procedure is justified to the extent that the scale of flavor breaking is well above the largest of the new-physics cutoffs in eq. (1.7). In these circumstances, expansions of beyond-SM flavor structures in terms of  $Y_{u,d}$  are stable under RGE evolution. From the above definition, it is clear that MFV is not a theory of flavor, since it takes the Yukawa couplings as given, without attempting any dynamical explanation of their structure. It is rather a criterion to achieve maximal protection of FCNC processes from beyond-SM contributions. The maximality is due to the fact that any new-physics flavor structure is assumed to inherit completely from the SM flavor structures, the Yukawa couplings. In this way, both the SM flavor-suppressing mechanisms, namely that of the CKM hierarchies and that of the quark mass patterns, are carried over to the new-physics interactions. This explains why the MFV solution to the flavor problem was denoted above as a drastic one. As a matter of fact, assuming that the  $c_{ij}$  couplings in eq. (1.9) are MFV, the new-physics scale goes down to about 4 TeV.

<sup>14</sup> I will follow here the effective-theory approach to MFV of [35]. The notion of MFV is also used to denote CKM-like flavor violation in [36]. This definition is more restrictive than the one in [35], because, as also discussed in sec. 1.4, the peculiar form of the CKM matrix is only one of the mechanisms behind the pattern of SM flavor breaking, the other mechanism being the peculiar values of quark masses.

<sup>15</sup> Note that this assumption leaves still freedom on the amount of allowed CP violation beyond the SM.

To summarize, MFV should not be regarded as ‘the expected pattern of beyond-SM flavor effects’, but rather as a useful way to ‘parameterize’ the question whether or not there are other sources of flavor-symmetry breaking at testable energies, beside the Yukawa couplings and the neutrino mass matrix. It should be stated clearly that, at present, the accuracy of this test does not exceed about 15% for MFV beyond-SM effects in  $K$  physics and about 25% in the case of  $B$  physics [32]. Therefore, we are still far from being able to claim that this hypothesis has been tested with high accuracy. Data on new FCNC observables and higher accuracies on existing observables, typically in the 1% – 10% range [33], are necessary to obtain bounds on beyond-SM effects able to exceed in severity those from electroweak precision observables, or else to finally uncover new effects.

## 1.6 Why pursuing flavor physics

Several of the points made in secs. 1.4 and 1.5 provide as many motivations for pursuing flavor physics. These motivations can be summarized by the following statements. Flavor observables provide an unequalled probe into higher energies and into the structure of interactions at these energies. Unequalled because beyond-SM flavor structures are plausibly different than the SM ones, hence it is correspondingly plausible that the SM pattern of flavor effects should be distorted by new physics, however high the scale at which it sets in. Furthermore several flavor observables can be measured with high accuracy. In short, the flavor-physics chances to unveil the unexpected rely:

- on the plausibility of new-physics effects showing up there;
- on the richness of observables;
- and on the accuracy achievable in their measurement.

As an example, the race towards an ever more accurate knowledge of the CKM parameters, started with  $B$ -factories and still ongoing with the Belle upgrade, is not intended for precision’s sake, but rather because it is a very delicate SM consistency check, that will fail very easily in presence of new flavor-breaking interactions, even if their energy scale is not-so-nearby.

Given the progressive shrinkage of arguments about new physics having to show up at the EW scale, the possibility of new discoveries will depend more and more on experimental probes being able to test higher and higher scales. In this situation, flavor observables will have more and more weight in defining future strategies, bearing in mind also the definitely lower cost of measuring low-energy, high-intensity observables that probe a certain energy scale, with respect to producing collisions at that energy scale, assuming that this is doable at all.

Discoveries at stake are as formidable as the effort needed to hopefully make them reality. While ‘hopefully’ understates the possibility of failure, it should be stressed that drawing definite negative conclusions based on 20% tests is overhasty. This point can be made more explicit by one last example from history.<sup>16</sup> The early 60s saw, as discussed above, a flourishing of experimental tests of discrete symmetries, among the others the CP symmetry in the  $K^0 - \bar{K}^0$  system. The latter system consists of two mass eigenstates, denoted  $K_S$  (shorter-lived) and  $K_L$  (longer-lived). If CP is a good symmetry, then the  $K_S$  is exactly CP even, and cannot decay into  $3\pi$ , whereas the  $K_L$  is exactly CP odd, and cannot decay into  $2\pi$ . A dedicated search was carried out in Dubna by the group of E. Okonov, collecting about 600

<sup>16</sup> The example to follow is quoted from the talk “Spacetime and vacuum as seen from Moscow”, given by L.B. Okun in 2001 [37], and of which I became aware thanks to talks by A. Soni and T. Browder.



decays of  $K_L$  to charged particles, and *not* finding a single  $K_L \rightarrow \pi^+\pi^-$  candidate [38]. At that stage, the lab administration decided to put an end to the search. Shortly afterwards, in 1964, this CP-violating decay was discovered at the level of 1/350 by the famous Brookhaven experiment of Christensen, Cronin, Fitch and Turlay [39], that was awarded the Nobel prize. This bold result decreed the breaking of a ‘sacred’<sup>17</sup> discrete symmetry, breaking that would soon be recognised by Sakharov [41] as one of the conditions for the observed imbalance between matter and antimatter in the universe.

## 1.7 Content of the manuscript

The present manuscript is based on work I pursued in the years 2008 - 2013, on improving the Standard Model prediction of two among the flavor-physics observables that are deemed as most promising for beyond-SM effects. The latter include: (1)  $\epsilon_K$ , that quantifies indirect CP violation in the  $K^0 - \bar{K}^0$  system, and (2) the very rare decay  $B_s \rightarrow \mu\mu$ , recently measured at the LHC. Item (1) will be dealt with in chapter 2, where I will describe our reconsideration of the long-distance contributions to  $\epsilon_K$  [1-3] and their impact on the follow-up literature. Item (2) will instead be the topic of chapter 3. Here I will give an account of various systematic effects, discussed in [42], and affecting the SM branching ratio for  $B_s \rightarrow \mu\mu$  at the 10% level. Chapter 3 further describes the several aspects of the  $B_{d,s} \rightarrow \mu\mu$  constraining power, including in particular an effective-theory comparison between  $B_s \rightarrow \mu\mu$  and  $Z$ -peak observables [5]. Finally, the outlook part, chapter 4, describes some further developments on the above subjects, and what is the progress to be expected on these and related observables.

---

<sup>17</sup> L.D. Landau put forward in 1956 the idea of absolute CP invariance, namely the idea that the observed ‘sin’ of P-violation was committed by particles and antiparticles in a way to globally leave CP as a good symmetry [40]. This idea was widely accepted.

## 2 Indirect CP violation in the $K^0 - \bar{K}^0$ system

Aim of the present chapter is to present a reappraisal of the long-distance contributions to  $\epsilon_K$ , that quantifies indirect CP violation in the  $K^0 - \bar{K}^0$  system. This reappraisal was undertaken in refs. [1–3]. The initial aim of ref. [1] was actually a reconsideration of the consistency between indirect CP violation in the  $K^0$ - vs. the  $B_d$ -meson system, at the time motivated by various tensions between data and theory predictions relevant to CKM-matrix fits. By a closer look at the level of precision required for this test to be meaningful, we almost stumbled on the necessity to reliably calculate and include the long-distance corrections to  $\epsilon_K$ , in the following referred to as the multiplicative factor  $\kappa_\epsilon$ . At the time of its proposal, the  $\kappa_\epsilon$  factor made the above mentioned tensions worse, and indicated a clear pattern of beyond-SM effects. This circumstance spurred further work on CKM fits, beyond-SM flavor effects, and associated model building. The latter aspect has lately lost momentum, especially because, at present accuracies, the initial tensions have by and large disappeared (and direct searches have likewise reported negative results so far). Mature data from LHCb and new data from the Belle upgrade will put a definite word on this issue, or maybe uncover effects at present swamped by too large errors. Independently of these tensions, a new, effective-theory calculation of the mentioned long-distance contributions to  $\epsilon_K$ , and a systematic discussion of this calculation and its limitations, is the lasting result of [1–3]. This result has motivated a new campaign of first-principle, non-perturbative calculations of the  $\kappa_\epsilon$  corrections within lattice QCD, and with  $\kappa_\epsilon$  of a number of historically challenging quantities related to  $K \rightarrow \pi\pi$  matrix elements.

### 2.1 Introduction to the problem

The discussion in this chapter is concerned with CP violation within the SM. More specifically, it is concerned with *indirect* CP violation, that namely does not arise directly from a decay, and is correspondingly denoted as *direct* CP violation [26]. Among the observables that give access to indirect CP violation (CPV), three are especially well measured and theoretically controlled, namely  $\epsilon_K$ , quantifying CPV in  $K^0 - \bar{K}^0$  mixing, and  $\sin 2\beta_{(s)}$ , quantifying CPV in the interference between decays with and without mixing, for the  $B_{d(s)} - \bar{B}_{d(s)}$  systems, respectively. (Note that this kind of CPV is mixing- and not decay-induced, hence it still qualifies as indirect CPV.) All these quantities will be properly defined in due course. The main point to make here is that, within the SM, there is one single source of CPV, namely the single phase  $\delta$  of the CKM matrix. Therefore, within the SM, these three observables are correlated to one another. Said otherwise, once one of them is measured, the other ones can be univocally predicted. Therefore, these three quantities offer a stringent test of the SM mechanism of CPV.

To make the discussion more concrete, let us focus on  $\epsilon_K$  vs.  $\sin 2\beta$ . As we will see later on in the detailed derivation, the short-distance contributions to  $\epsilon_K$  (from top-top exchange) can be approximately expressed as

$$|\epsilon_K| \simeq C \cdot \hat{B}_K \cdot \sin 2\beta . \quad (2.1)$$

Here  $C$  denotes a calculable coefficient and  $\hat{B}_K$  a non-perturbative matrix element, input from lattice QCD. The precise definition of these two quantities is not relevant here, and it will be dwelled upon in the detailed discussion. Eq. (2.1) neglects the subdominant short-distance contributions from top-charm and charm-charm exchange, that are likewise irrelevant for the point to be made here. This point is that there is a correlation, in particular a direct proportionality, as displayed in eq. (2.1), between indirect CPV in the  $K^0$ -system

and the one in the  $B_d$ -system. This correlation can be understood intuitively by just noting that, by its very definition – an observable quantifying CPV in  $K^0 - \bar{K}^0$  mixing –  $\epsilon_K$  must be of the form

$$\epsilon_K \sim \text{Im} \left( \begin{array}{c} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \\ \bar{K}^0 \quad \quad \quad K^0 \\ \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \end{array} \right). \quad (2.2)$$

The imaginary part of the CKM couplings at the four vertices yields the  $\sin 2\beta$  factor; the effective-operator structure arising from the diagram, calculated between the external  $\bar{K}^0, K^0$  states, yields  $\hat{B}_K$ ; the rest is a calculable coefficient.

The novelty of [1] was, as mentioned, the reconsideration of various additional contributions, especially long-distance ones, neglected in the previous literature in view of what used to be a large error on  $\hat{B}_K$ . These contributions can be simplistically visualized as a negative correction, of order  $-8\%$ , to the r.h.s. of eq. (2.1), that would then become

$$|\epsilon_K| \simeq C \cdot (1 - \text{O}(8\%)) \cdot \hat{B}_K \cdot \sin 2\beta. \quad (2.3)$$

Now, taking  $\sin 2\beta = S_{\psi K_S} \simeq 0.68$  (values refer to experimental averages at the time of [1]), one obtained  $|\epsilon_K| = 1.78 \times 10^{-3}$ , to be compared with the experimental figure of  $|\epsilon_K|_{\text{exp}} = 2.232(7) \times 10^{-3}$ . Conversely, identifying the l.h.s. of eq. (2.3) with its experimental value results in  $\sin 2\beta \sim 0.8$ , too large with respect to the experimental average. Leaving aside errors for the moment being, in either case the discrepancy between central values is as large as 20%.

The above considerations will be made more quantitative in the next sections.

## 2.2 CPV in $K$ physics: theory vs. experiment

In this section we introduce the minimal necessary formalism, and establish a contact between what experiment measures and what theory calculates. The  $K^0 - \bar{K}^0$  system consists of two flavor eigenstates, with flavor content

$$|K^0\rangle \sim \begin{pmatrix} d \\ \bar{s} \end{pmatrix}, \quad |\bar{K}^0\rangle \sim \begin{pmatrix} \bar{d} \\ s \end{pmatrix}. \quad (2.4)$$

The CP symmetry acts on these states as follows

$$\begin{aligned} \text{CP}|K^0\rangle &= e^{i\xi}|\bar{K}^0\rangle, \\ \text{CP}|\bar{K}^0\rangle &= e^{-i\xi}|K^0\rangle, \end{aligned} \quad (2.5)$$

namely it connects them to one another, up to a phase that remains arbitrary, because the two states do not communicate via strong interactions. In the CKM conventions, where all CPV arises from the CKM phase  $\delta$ , one can take

$$\begin{aligned} \text{CP}|K^0\rangle &= |\bar{K}^0\rangle, \\ \text{CP}|\bar{K}^0\rangle &= |K^0\rangle, \end{aligned} \quad (2.6)$$

whence the two CP eigenstates are found to be

$$|K_{\pm}\rangle = \frac{|K^0\rangle \pm |\bar{K}^0\rangle}{\sqrt{2}}, \quad (2.7)$$

with namely CP-eigenvalues  $\pm 1$ , respectively. If CP were a good symmetry of the weak Hamiltonian  $\mathcal{H}_W$ , i.e.  $[\text{CP}, \mathcal{H}_W] = 0$ , then  $|K_{\pm}\rangle$  would also be good physical eigenstates. As

we know after ref. [39], CP is slightly, but surely violated by weak interactions. Therefore the physical eigenstates are mostly the CP eigenstates, but for a small admixture with the opposite-CP state, namely

$$|K_{S,L}\rangle \propto |K_{\pm}\rangle + \bar{\epsilon}|K_{\mp}\rangle \quad (2.8)$$

with  $|\bar{\epsilon}| \ll 1$ . Therefore, the shorter-lived  $|K_S\rangle$  state decays mostly into  $2\pi$ , and occasionally into  $3\pi$ , and the other way around occurs for the longer-lived  $|K_L\rangle$  state.

To quantify the amount of CPV we need to connect what theory calculates with what experiment measures. Experiment can access the ratios

$$\eta^{+-} \equiv \frac{\langle \pi^+ \pi^- | K_L \rangle}{\langle \pi^+ \pi^- | K_S \rangle}, \quad \eta^{00} \equiv \frac{\langle \pi^0 \pi^0 | K_L \rangle}{\langle \pi^0 \pi^0 | K_S \rangle}. \quad (2.9)$$

These ratios would be zero if CP were conserved. From eqs. (2.8)-(2.9), CPV can occur either because of the small  $|K_{-(+)}\rangle$  component in the  $|K_{S(L)}\rangle$ , that then decays into  $3\pi$  ( $2\pi$ ), or because the  $|K_{\pm}\rangle$  components of either physical eigenstate decay directly into the wrong-CP final state,  $3\pi$  or  $2\pi$  respectively. In the former case one has indirect CPV, through mixing; in the latter case CPV occurs directly in the decay. The two CPV components are therefore intertwined within the observables  $\eta^{+-}, \eta^{00}$ .

In order to separate direct vs. indirect CPV components in the  $\eta^{+-}, \eta^{00}$  observables, we need to introduce final states of definite isospin  $SU(2)_I$ . In fact, there is no direct CPV in  $K \rightarrow (\pi\pi)_I$ .<sup>18</sup> To obtain final states with definite isospin, one starts from recalling that the  $u, d$  quarks are  $+, -$  states, respectively, under isospin, and that, because of their quark content [26], the  $|\pi^+\rangle, |\pi^0\rangle, |\pi^-\rangle$  states form an  $I = 1$  triplet.  $|\pi\pi\rangle$  isospin eigenstates are accordingly obtained by composing symmetrically two  $I = 1$  representations. In the PDG phase conventions one obtains

$$\begin{aligned} |\pi^+ \pi^-\rangle &= \sqrt{\frac{2}{3}} |(\pi\pi)_0\rangle + \sqrt{\frac{1}{3}} |(\pi\pi)_2\rangle, \\ |\pi^0 \pi^0\rangle &= \sqrt{\frac{1}{3}} |(\pi\pi)_0\rangle - \sqrt{\frac{2}{3}} |(\pi\pi)_2\rangle, \end{aligned} \quad (2.10)$$

which are the relations we need in order to express measured states in terms of states of definite isospin. Plugging the first of these relations into eq. (2.9) yields

$$\eta^{+-} = \frac{a_{L,0} + \frac{1}{\sqrt{2}} a_{L,2}}{a_{S,0} + \frac{1}{\sqrt{2}} a_{S,2}} = \frac{\epsilon + \frac{1}{\sqrt{2}} \epsilon_2}{1 + \frac{1}{\sqrt{2}} \omega}, \quad (2.11)$$

where

$$a_{i,I} \equiv \langle (\pi\pi)_I | \mathcal{H}_W | K_i \rangle, \quad (2.12)$$

and in the last member we have introduced the three amplitude ratios (all three of them small in magnitude)

$$\epsilon_K \equiv \frac{a_{L,0}}{a_{S,0}}, \quad \epsilon_2 \equiv \frac{a_{L,2}}{a_{S,0}}, \quad \omega \equiv \frac{a_{S,2}}{a_{S,0}}. \quad (2.13)$$

The ratio  $\omega = 0.045$  is very well measured, and quantifies what is known as the  $\Delta I = 1/2$  rule [44], namely the unexpected fact that in  $K \rightarrow \pi\pi$  decays the final state is  $\sim 1/\omega^2$  times more likely to be an  $I = 0$  eigenstate than an  $I = 2$  one. There is to date no simply

<sup>18</sup> A beautiful discussion of this matter can be found in [43]. The main point is however that the  $I = 0, 2$  amplitudes are fully described by just two strong phases. This in turn follows from the fact that, in QCD, a  $\pi\pi$  state can only rescatter into itself (due to CP conservation, or else to energy conservation), and the isospin symmetry is almost exact.

understandable dynamical explanation of this fact, that remains one of the longest-standing puzzles in particle physics phenomenology. A numerical understanding starts to emerge from lattice QCD simulations (see in particular [45]) – this undertaking being per se extremely challenging, because, for one thing, the possibility to simulate the physical region of matrix elements with more than one hadron in the final state is limited by a no-go theorem [46], from which one has to find very clever workarounds.

The amplitude ratio in eq. (2.13) denoted as  $\epsilon_K$  quantifies, as mentioned, *indirect* CPV in  $K$  decays, and is our main quantity of interest. (It is the only ratio involving final states with the same isospin, so there cannot be *direct* CPV, as argued above.) Another combination of  $K \rightarrow \pi\pi$  amplitudes, that we do not need to introduce here, accounts for  $\epsilon'$ , the parameter quantifying *direct* CPV in  $K$  decays. Through the definitions in eq. (2.13) the direct- vs. indirect-CPV contributions to  $\eta^{+-}, \eta^{00}$  are thereby disentangled.

### 2.3 Derivation of the $\epsilon_K$ formula

Let us now focus on

$$\epsilon_K \equiv \frac{\langle (\pi\pi)_0 | \mathcal{H}_W | K_L \rangle}{\langle (\pi\pi)_0 | \mathcal{H}_W | K_S \rangle}, \quad (2.14)$$

as in eq. (2.13). We can express the  $|K_{S,L}\rangle$  states in terms of the  $|K^0\rangle$  ones by combining eqs. (2.7) and (2.8), yielding

$$|K_{S,L}\rangle = N_{\bar{\epsilon}} \left[ (1 + \bar{\epsilon}) |K^0\rangle \pm (1 - \bar{\epsilon}) |\bar{K}^0\rangle \right], \quad (2.15)$$

where  $N_{\bar{\epsilon}}$  is a normalization factor. We finally need symbols for weak matrix elements between  $|K^0\rangle$  states and  $\langle (\pi\pi)_I |$  ones:

$$\begin{aligned} \langle (\pi\pi)_I | \mathcal{H}_W | K^0 \rangle &\equiv a_I e^{i\delta_I}, \\ \langle (\pi\pi)_I | \mathcal{H}_W | \bar{K}^0 \rangle &\equiv a_I^* e^{i\delta_I}, \end{aligned} \quad (2.16)$$

where  $a_I$  denotes a weak, complex amplitude, and  $\delta_I$  a strong phase, accounting for final-state interactions.<sup>19</sup> Plugging eqs. (2.15) and (2.16) into eq. (2.14) gives

$$\epsilon_K = \frac{\bar{\epsilon} + i\xi}{1 + i\bar{\epsilon}\xi} \simeq \bar{\epsilon} + i\xi, \quad \text{with } \xi \equiv \frac{\text{Im } a_0}{\text{Re } a_0} \quad (2.17)$$

where the approximation on the r.h.s. of the  $\epsilon_K$  equation is well justified, given that both of  $\bar{\epsilon}$  and  $\xi$  are small in magnitude. Eq. (2.17) is a first important formula on  $\epsilon_K$ , and it is worthwhile to pause on its physics content. This relation shows that  $\epsilon_K$  is the sum of two contributions: the first,  $\bar{\epsilon}$ , is due to the fact that the weak Hamiltonian mixes different CP eigenstates (see eq. (2.8)), it represents the main contribution, and is fully calculable in perturbation theory;<sup>20</sup> the second,  $i\xi$ , is due to the small but non-zero weak phase of the  $K^0 \rightarrow (\pi\pi)_0$  amplitude, and represents a subdominant correction. Its non-negligibility in view of the improved errors on the rest of the  $\epsilon_K$  parametrics was one of the points made originally in [1].

The  $\xi$  correction is not easy to estimate accurately. A first strategy was proposed in [1] (see also related work by U. Nierste in the Fermilab report [43]), whereby this correction is extracted from data on  $\epsilon'/\epsilon_K$ ; a more refined effective-field theory treatment was performed

<sup>19</sup> That this final-state interaction can only be elastic and is thus fully described by just two scattering phases follows from kinematics and QCD conservation laws, and is known as Watson's theorem [47].

<sup>20</sup> With qualifications to be made in sec. 2.5.

in ref. [3], using Chiral Perturbation Theory. Both these calculations will be described in more detail below. As already mentioned, the calculation in ref. [3] has spurred non-negligible activity across the lattice community to calculate the  $\xi$  correction in lattice QCD [48] (for related work, see [45, 49]; for an excellent review, see talk by A. Jüttner [50]).

To make the calculability of the  $\bar{\epsilon}$  term more apparent, we next express this contribution in terms of weak-Hamiltonian matrix elements. For this purpose, we need to introduce and solve the  $K^0 - \bar{K}^0$  eigenvalue problem. As we saw, the  $K^0$  and  $\bar{K}^0$  states mix because of weak interactions, and  $\bar{\epsilon}$  parameterizes this mixing. A generic admixture between  $K^0$  and  $\bar{K}^0$  at time  $t$  is parameterized by the wave-function  $\vec{\Phi}(t)$ , with<sup>21</sup>

$$\vec{\Phi}(t) \equiv \begin{pmatrix} \phi_K(t) \\ \phi_{\bar{K}}(t) \end{pmatrix} = \phi_K(t)|K^0\rangle + \phi_{\bar{K}}(t)|\bar{K}^0\rangle, \quad (2.18)$$

whereby the  $|K^0\rangle$  states have been taken to be  $(1, 0)^T$  and  $(0, 1)^T$  respectively.  $\vec{\Phi}$  obeys the Schrödinger time-evolution equation

$$i \frac{d\vec{\Phi}(t)}{dt} = \mathcal{H} \vec{\Phi}(t) \quad (2.19)$$

with

$$\mathcal{H} = \begin{pmatrix} M_0 - i\frac{\Gamma_0}{2} & M_{12} - i\frac{\Gamma_{12}}{2} \\ M_{12}^* - i\frac{\Gamma_{12}^*}{2} & M_0 - i\frac{\Gamma_0}{2} \end{pmatrix}, \quad (2.20)$$

where the  $M_i, \Gamma_j$  entries are real numbers. The diagonal elements of the  $\mathcal{H}$  matrix are constrained to be the same by CPT invariance, whereas Hermiticity (applied separately to the real and imaginary parts) requires  $M_{21} = M_{12}^*$  and  $\Gamma_{21} = \Gamma_{12}^*$ , whence the relation between the 12 and the 21 element in eq. (2.20) follows. Before proceeding to the solution, it is, in fact, worth remarking explicitly that the equations of motion (2.19)-(2.20) are far from obvious, and are the result of the so-called Weisskopf-Wigner ansatz on a generic decay problem [52]. This ansatz assumes: (i) that the time dependence is like the one of an harmonic-oscillator system, namely  $e^{-i\omega t}$ ; (ii) that the  $M$  and  $\Gamma$  entries in the  $\mathcal{H}$  matrix are time-independent in the interval corresponding to the energy scale of the matrix eigenvalues, parametrized by the mass and width differences  $\Delta m_K$  and  $\Delta \Gamma_K$  respectively. As an a-posteriori consistency check, one verifies that  $\Delta m_K, \Delta \Gamma_K$  are much smaller than the energy scale of the system, which is of order  $M_K$ .

Let us now solve the eigenvalue problem in eq. (2.20), recalling that our aim is just to express  $\bar{\epsilon}$  in terms of  $\mathcal{H}$  entries. This task can be easily accomplished by simply noting that, labeling the two eigenvalues as  $\lambda_+$  and  $\lambda_-$ , their difference must be

$$\lambda_+ - \lambda_- = \Delta m_K - i \frac{\Delta \Gamma_K}{2}, \quad (2.21)$$

which implies

$$\lambda_+ = M_0 - i\frac{\Gamma_0}{2} + \frac{1}{2} \left( \Delta m_K - i \frac{\Delta \Gamma_K}{2} \right). \quad (2.22)$$

Let us now apply the eigenvalue equations  $(\mathcal{H} - \lambda \mathbb{1})|\text{eigenvec.}\rangle = 0$  to  $|K_S\rangle$ , whose eigenvalue is  $\lambda_+$ . Taking in particular the second of these equations, one arrives at

$$\frac{1 - \bar{\epsilon}}{1 + \bar{\epsilon}} = \frac{2M_{12}^* - i\Gamma_{12}^*}{\Delta m_K - i\frac{\Delta \Gamma_K}{2}} \simeq 1 - 2\bar{\epsilon}, \quad (2.23)$$

<sup>21</sup> I will use a notation very close to [51] in this part.

where the r.h.s. follows again from the fact that  $\bar{\epsilon}$  is small in magnitude. From the r.h.s. of eq. (2.23) it is clear that the middle member of this equation is almost unity, but for a small correction, that we have to isolate. To this end, let us expand eq. (2.23) explicitly in its real and imaginary components (denoted briefly by  $R$  and  $I$  superscripts)

$$\frac{2M_{12}^R - 2iM_{12}^I - i\Gamma_{12}^R - \Gamma_{12}^I}{\Delta m_K - i\frac{\Delta\Gamma_K}{2}} \simeq 1 - 2\bar{\epsilon}. \quad (2.24)$$

Consistency demands [51] that  $M_{12}^I \ll M_{12}^R$  and  $\Gamma_{12}^I \ll \Gamma_{12}^R$ , and the validity of these relations can also be checked a posteriori. Then eq. (2.24) leads us to the identifications

$$\Delta m_K \simeq 2M_{12}^R, \quad \Delta\Gamma_K \simeq 2\Gamma_{12}^R, \quad (2.25)$$

where, experimentally

$$\begin{aligned} \Delta m_K &= M_{K_L} - M_{K_S} \simeq 3.5 \times 10^{-15} \text{ GeV}, \\ \Delta\Gamma_K &= \Gamma_{K_L} - \Gamma_{K_S} \simeq -7.4 \times 10^{-15} \text{ GeV}. \end{aligned} \quad (2.26)$$

Since experimentally  $\Delta m_K$  and  $\Delta\Gamma_K/2$  are accidentally almost identical in magnitude, it is useful to also introduce the parameter

$$\tan \phi_\epsilon \equiv -\frac{\Delta m_K}{\Delta\Gamma_K/2}. \quad (2.27)$$

Eqs. (2.25) and (2.27) allow us to rewrite eq. (2.24) as

$$\bar{\epsilon} = \frac{iM_{12}^I + \frac{1}{2}\Gamma_{12}^I}{\Delta m_K (1 + i \cot \phi_\epsilon)}, \quad (2.28)$$

which is what we were looking for: a calculable expression of  $\bar{\epsilon}$ , in terms of  $\mathcal{H}$  entries.

We may now plug eq. (2.28) into our first  $\epsilon_K$  formula, eq. (2.17). Since, however, the  $\Gamma_{12}^I$  piece turns out to contribute to the  $\xi$  term, we should make this dependence explicit. Let us recall that  $\Gamma_{12}$  represents the absorptive part of the weak matrix element between an initial  $|\bar{K}^0\rangle$  and a final  $\langle K^0|$  state, namely

$$\Gamma_{12} \propto \sum_n \langle K^0 | \mathcal{H}_W | n \rangle \langle n | \mathcal{H}_W | \bar{K}^0 \rangle, \quad (2.29)$$

with  $n$  all the intermediate states with appropriate quantum numbers. This sum is dominated by the  $(\pi\pi)_0$  intermediate state [51]. Using definitions (2.16), one then has

$$\begin{aligned} \Gamma_{12} &\propto (a_0 e^{i\delta_0})^* (a_0^* e^{i\delta_0}) = (a_0^*)^2 \\ \Rightarrow \frac{\Gamma_{12}^I}{\Gamma_{12}^R} &= -2 \frac{\text{Im} a_0}{\text{Re} a_0} = -2\xi. \end{aligned} \quad (2.30)$$

It is now a trivial  $c$ -number algebra exercise to show that the  $\epsilon_K$  relation in eq. (2.17), with  $\bar{\epsilon}$  as in eq. (2.28), and with  $\Gamma_{12}^I$  expressed as in eq. (2.30), becomes

$$\epsilon_K = \sin \phi_\epsilon e^{i\phi_\epsilon} \left( \frac{M_{12}^I}{\Delta m_K} + \xi \right), \quad (2.31)$$

which is our final  $\epsilon_K$  relation, and opens the way to a more precise statement of the point originally made in [1].

## 2.4 The $\kappa_\epsilon$ correction

In eq. (2.31) the only complex quantity is the phase factor. Therefore, for the  $\epsilon_K$  magnitude this relation trivially implies

$$|\epsilon_K| = \sin \phi_\epsilon \left( \frac{M_{12}^I}{\Delta m_K} + \xi \right). \quad (2.32)$$

In usual applications before ref. [1] it was customary to assume

- $\xi = 0$ , whereas it is actually *smaller* than zero;
- $\phi_\epsilon = 45^\circ$ , whereas it actually equals  $43.5(7)^\circ$ ,<sup>22</sup> it is namely *smaller* than  $45^\circ$ .

In other words, both the  $\xi \neq 0$  and the  $\phi_\epsilon < 45^\circ$  corrections go in the same, *negative* direction, and their sum builds up a total correction of  $O(-10\%)$  [1]. (This figure has slightly decreased in magnitude since the analysis in ref. [3], that forms the topic of the next section.) To appreciate quantitatively the effect of this correction, let us rewrite eq. (2.32) by parameterizing explicitly the above mentioned approximations, namely as

$$|\epsilon_K| = \underbrace{\frac{1}{\sqrt{2}} \frac{M_{12}^I}{\Delta m_K}}_{|\epsilon_K|^{\text{approx}}} \underbrace{\frac{\sin \phi_\epsilon}{1/\sqrt{2}} \left( 1 + \frac{\xi}{\sqrt{2}|\epsilon_K|} \right)}_{\kappa_\epsilon}, \quad (2.33)$$

where the difference between  $|\epsilon_K|$  and  $|\epsilon_K|^{\text{approx}}$  has been neglected on the r.h.s. parenthesis. According to this relation, the difference between the approximate and the more accurate  $\epsilon_K$  formula can be accounted for by the multiplicative factor  $\kappa_\epsilon$  [1], defined as

$$\kappa_\epsilon \equiv \frac{\sin \phi_\epsilon}{1/\sqrt{2}} \times \bar{\kappa}_\epsilon, \quad \text{with } \bar{\kappa}_\epsilon \equiv 1 + \frac{\xi}{\sqrt{2}|\epsilon_K|} \equiv 1 + \Delta_\epsilon, \quad (2.34)$$

where namely the first factor parameterizes the effect of  $\phi_\epsilon \neq \pi/4$ , and  $\bar{\kappa}_\epsilon$  that of  $\xi \neq 0$ . The factor proportional to  $\xi$  has been denoted as  $\Delta_\epsilon$  for later convenience. Plugging in the PDG value of  $\phi_\epsilon$ , the  $\kappa_\epsilon$  factor proportional to  $\sin \phi_\epsilon$  amounts to a correction of  $-2.6\%$ . Furthermore, using the estimate of  $\xi$  of ref. [1], to be shortly discussed next (before the more refined treatment in sec. 2.5) the  $\bar{\kappa}_\epsilon$  factor amounts to a correction of  $-5.4\%$ . Hence the *like sign of the two corrections* implies  $\kappa_\epsilon - 1$  as large as  $-8\%$  [1], i.e. not negligible given the recent progress made on the rest of the  $\epsilon_K$  parametrics, to which I will comment in sec. 2.6.

### Extracting $\xi$ from $\epsilon'/\epsilon_K$

The main obscure point in the above argument is the  $\bar{\kappa}_\epsilon$  estimate. In principle, any  $K \rightarrow \pi\pi$  amplitude is computable in terms of matrix elements times Wilson coefficients of the  $|\Delta S| = 1$  effective Hamiltonian (see [53] for a review), whose operator basis consists of two current-current operators  $\mathcal{Q}_{1,2}$ , four so-called QCD penguins  $\mathcal{Q}_{3-6}$ , and four EW penguins  $\mathcal{Q}_{7-10}$ . However, in practice most of the  $\langle \pi\pi | \mathcal{Q}_i | K \rangle$  matrix elements are extremely challenging to calculate – remember in this respect the considerations made at the end of sec. 2.2. This is especially true for the case of QCD-penguin operators. It turns out that the main unknown in the  $\bar{\kappa}_\epsilon$  correction, namely  $\xi$ , is dominated by contributions from QCD penguins, in particular

<sup>22</sup> The  $\phi_\epsilon$  and  $\epsilon'/\epsilon_K$  values used in this section are, for consistency, those in the PDG as of the time of completion of ref. [1]. However, the latest PDG averages [26] are basically identical:  $\phi_\epsilon = 43.5(5)^\circ$ ,  $|\epsilon'/\epsilon_K|_{\text{exp}} = 1.66(23) \cdot 10^{-3}$ .



$\mathcal{Q}_6$ . At present, a *direct* estimate of this matrix element is subject to very large hadronic uncertainties. A much more reliable strategy consists in evaluating the  $\xi$  correction indirectly, using the experimental value of  $\epsilon'/\epsilon_K$  [43, 54]. To make this strategy more transparent, we need to explicitly introduce the  $\epsilon'$  definition in terms of  $K \rightarrow \pi\pi$  amplitudes (see e.g. [53]):

$$\epsilon' \equiv \frac{1}{\sqrt{2}} \text{Im} \left( \frac{a_2}{a_0} \right) e^{i\Phi} , \quad (2.35)$$

with  $a_I$  defined in eq. (2.16) and  $\Phi$  measured to be close to  $\pi/4$ . This definition can be understood by recalling that  $\epsilon'$  quantifies *direct* CPV in  $K \rightarrow \pi\pi$  decays, and that there is no direct CPV in  $K$  decays to  $\pi\pi$  states of definite isospin, therefore one needs to invoke the amplitude ratio in eq. (2.35) to have direct CPV. The above definition can be rewritten as

$$\epsilon' \simeq -\frac{\omega\xi}{\sqrt{2}} (1 - \Omega) e^{i\Phi} , \quad \text{with } \Omega \equiv \frac{1 \text{Im } a_2}{\omega \text{Im } a_0} \quad (2.36)$$

and  $\omega, \xi$  already introduced in eqs. (2.13) and (2.17) respectively. To obtain eq. (2.36) from eq. (2.35) one takes in the denominator  $|a_i| \simeq \text{Re } a_I$ , which is valid to very good approximation. From eq. (2.36) it should be clear why we introduced  $\epsilon'$  at all: the first factor on the r.h.s. is, aside from a (known) rescaling by  $\omega/|\epsilon_K|$ , exactly the factor in the parenthesis on the r.h.s. of eq. (2.33). One may then estimate this factor from the known experimental value of  $\epsilon'/\epsilon_K$ , provided one can reliably estimate  $\Omega$ . This quantity parameterizes the EW-penguin contributions to  $\epsilon'$ , or rather their ratio with respect to QCD-penguin contributions – note in fact that  $\Omega$  is a correction to unity in eq. (2.36), and that the other factor accounts, as already said, for QCD penguins.  $\Omega$  can be calculated in a much more reliable way than  $\xi$  [55], and we find the range

$$\Omega = 0.40 \pm 0.08 \quad (2.37)$$

with namely a very acceptable 20% error. Plugging this relation into

$$\left| \frac{\epsilon'}{\epsilon_K} \right| = -\frac{\omega\xi}{\sqrt{2}|\epsilon_K|} (1 - \Omega) , \quad (2.38)$$

and using for the l.h.s.  $|\epsilon'/\epsilon_K|_{\text{exp}} = 1.66(26) \cdot 10^{-3}$ ,<sup>23</sup> one arrives easily at the announced correction  $\Delta_\epsilon = -5.4\%$ , with a 25% error, which is the sum in quadrature of the 20% error on  $\Omega$  and the 15% error on  $|\epsilon'/\epsilon_K|_{\text{exp}}$ . This procedure finally yields [1]

$$\kappa_\epsilon = 0.92 \pm 0.02 \quad (2.39)$$

for the multiplicative correction to  $\epsilon_K$ .

\*\*\*

While the above argument is self contained and has the advantage of requiring very little notation, the actual  $\kappa_\epsilon$  calculation from  $\epsilon'/\epsilon_K$  involves the explicit introduction of the  $\epsilon'/\epsilon_K$  operator-product expansion (OPE). Albeit not necessary for the rest of the present manuscript, we shortly present this more rigorous procedure here for the sake of completeness. This procedure is mandatory whenever  $\kappa_\epsilon$  needs to be recalculated altogether, because e.g. of a major change in input, or because  $\epsilon'/\epsilon_K$ , assumed to be SM-like in the  $\kappa_\epsilon$  estimation, is actually affected by non-SM contributions, as discussed in ref. [2].

<sup>23</sup> See footnote 22.

The OPE for  $\epsilon'/\epsilon_K$  can be very conveniently recast in the so-called ‘‘penguin-box expansion’’ [56, 57], whereby  $\epsilon'/\epsilon_K$  has the form

$$\frac{\epsilon'}{\epsilon_K} = \text{Im}\lambda_t \cdot F_{\epsilon'}(x_t) , \quad (2.40)$$

with  $\lambda_t = V_{ts}^* V_{td}$ ,  $x_t = \bar{m}_t^2(m_t)/M_W^2$  (the bar denoting a mass evaluated in the  $\overline{\text{MS}}$  scheme with respect to QCD corrections) and  $F_{\epsilon'}$  given by

$$F_{\epsilon'}(x_t) = P_0 + P_X X_0(x_t) + P_Y Y_0(x_t) + P_Z Z_0(x_t) + P_E E_0(x_t) , \quad (2.41)$$

where  $X_0$ ,  $Y_0$ ,  $Z_0$  and  $E_0$  are combinations of Inami-Lim functions [58]. The coefficients  $P_i$  in eq. (2.41) are defined as [55, 57]

$$P_i = r_i^{(0)} + r_i^{(6)} R_6 + r_i^{(8)} R_8 . \quad (2.42)$$

Here  $r_i^{(0)}$ ,  $r_i^{(6)}$  and  $r_i^{(8)}$  encode the information on the Wilson-coefficient functions of the  $|\Delta S| = 1$  effective Hamiltonian at the next-to-leading order [59]. Their numerical values for different choices of the QCD scale  $\Lambda_{\overline{\text{MS}}}^{(4)}$  at  $\mu = m_c$  can be found in table 1 of [2] (this scale is evaluated in  $\overline{\text{MS}}$ , with 4 active flavors). On the other hand,  $R_{6,8}$ , defined as

$$R_6 \equiv B_6^{(1/2)} \left[ \frac{121 \text{ MeV}}{m_s(m_c) + m_d(m_c)} \right]^2 , \quad R_8 \equiv B_8^{(3/2)} \left[ \frac{121 \text{ MeV}}{m_s(m_c) + m_d(m_c)} \right]^2 , \quad (2.43)$$

encode, through the ‘ $B$ -parameters’  $B_6^{(1/2)}$  ( $B_8^{(3/2)}$ ), the information on the operator matrix elements  $\langle \mathcal{Q}_6 \rangle_0$  ( $\langle \mathcal{Q}_8 \rangle_2$ ) between a  $K$ -meson and a  $\pi\pi$ -state with isospin  $I=0$  ( $I=2$ ). Eqs. (2.40)-(2.42) assume the  $|\Delta S| = 1$  operator basis  $Q_{1-10}$  already mentioned above<sup>24</sup> wherein  $Q_6$  ( $Q_8$ ) represents the QCD-penguin (EW-penguin) operator giving the largest contributions. Concerning  $R_8$ , we assume the reasonable range

$$R_8 = 1.0 \pm 0.2 , \quad (2.44)$$

that encompasses various estimates reviewed in [55]. This assumption has a one-to-one correspondence with the  $\Omega$  range assumed in eq. (2.37). On the other hand, we make no assumption on  $R_6$ , in view of its huge theoretical uncertainties. Its range, necessary for the estimation of  $\xi$ , hence  $\kappa_\epsilon$ , is extracted indirectly by demanding equality of the theoretical  $\epsilon'/\epsilon_K$  formula with  $|\epsilon'/\epsilon_K|_{\text{exp}}$  within its  $1\sigma$  range.

More explicitly, once the  $R_6$  range has been estimated, the entailed range for the correction  $\Delta_\epsilon$  can be obtained from the following approximate, but quite accurate formula

$$\Delta_\epsilon \approx -\frac{1}{\omega} \text{Im}\lambda_t \cdot F_{\epsilon'}(x_t)|_{R_8 \rightarrow 0} . \quad (2.45)$$

We note that the r.h.s. of eq. (2.45) includes in the  $\Delta_\epsilon$  estimate the contributions from the coefficients  $r_i^{(0)}$  (see eq. (2.42)), that consist also of a  $\Delta I = 3/2$  component along with the  $\Delta I = 1/2$  one. The former component is not separated away in eq. (2.45). Using the results of ref. [56], one can however convince oneself that this approximation amounts to overestimating  $|\Delta_\epsilon|$  by less than 10%, even for substantial new-physics contributions to  $Z$ -penguins [2]. Therefore, effectively, the limit  $R_8 \rightarrow 0$  in the  $P_i$  coefficients (2.42) corresponds to  $\Omega \rightarrow 0$  in (2.38), hence the possibility to estimate  $\Delta_\epsilon$  from the simple relation (2.45).

<sup>24</sup> For the operator definitions, unnecessary here, see [56].

## 2.5 A closer look at the operator-product expansion for $\epsilon_K$

Most of the discussion so far has been centered around eq. (2.32), and the fact that  $\phi_\epsilon \neq \pi/4$  as well as  $\xi \neq 0$  imply systematic corrections that cannot be neglected in modern applications in view of the overall improvements in the input parameters to  $\epsilon_K$ . This line of argument calls for a more in-depth discussion of the  $\epsilon_K$  theory error, in particular the systematic one, in order to understand whether further corrections may be missing. (On the parametric error we will pause in detail in sec. 2.6.)

In order to address the above question, we need to look in detail into the OPE for  $\epsilon_K$ . To this end, let us restart from eq. (2.32), that we rewrite here

$$|\epsilon_K| = \sin \phi_\epsilon \left( \frac{\text{Im}M_{12}}{\Delta m_K} + \xi \right). \quad (2.46)$$

### Short-distance contributions to $\text{Im}M_{12}$

Thus far we have not defined in detail how  $\text{Im}M_{12}$  is calculated. Recalling eq. (2.20),  $M_{12}$  is defined by the matrix element

$$M_{12} = \langle K^0 | \mathcal{H}_{|\Delta S|=2} | \bar{K}^0 \rangle, \quad (2.47)$$

where  $\mathcal{H}_{|\Delta S|=2}$  is the effective Hamiltonian for transitions where the strangeness quantum number changes by two units between the final and the initial state (recall that the  $s$  quark has strangeness  $-1$ ). At leading order in EW interactions,  $|\Delta S| = 2$  transitions are induced by ‘box’ diagrams, in particular the one in fig. 1, and the one obtained by a  $90^\circ$  rotation of this diagram. After inclusion of QCD corrections [60], this leading, dimension-6 contribution to  $\mathcal{H}_{|\Delta S|=2}$  reads (see [61])

$$\begin{aligned} \mathcal{H}_{|\Delta S|=2}^{(6)} &= \frac{G_F^2}{16\pi^2} M_W^2 \left[ \lambda_c^2 \eta_{cc} S_0(x_c) + \lambda_t^2 \eta_{tt} S_0(x_t) + 2\lambda_c \lambda_t \eta_{ct} S_0(x_c, x_t) \right] \times \\ &\quad [\alpha_s(\mu)]^{-2/9} \left[ 1 + \frac{\alpha_s(\mu)}{4\pi} J_3 \right] \mathcal{Q}_1 + \text{h.c.} . \end{aligned} \quad (2.48)$$

Here  $G_F$  denotes the effective Fermi coupling, such that  $G_F/\sqrt{2} = g^2/(8M_W^2)$ : since the box diagram in fig. 1 consists of four interaction vertices, the contribution in eq. (2.48) is effectively of order  $G_F^2$ . The Inami-Lim functions [58], denoted as  $S_0$ , are short-distance functions encoding the dynamics of  $c - c$ ,  $t - t$  and respectively  $c - t$  exchange in the box, with  $x_i = m_i^2/M_W^2$ . Their explicit form can be found e.g. in [61]. These functions appear multiplied by the respective CKM couplings, with  $\lambda_i \equiv V_{is}V_{id}^*$  (note that this is the complex conjugate of the corresponding definition in [61]). The contributions proportional to  $\lambda_u$  are eliminated using CKM unitarity, and one further sets  $x_u = 0$ , which holds to excellent approximation. The short-distance functions in eq. (2.48) are furthermore multiplied by  $\eta$  factors, encoding the effect of QCD corrections at next-to-leading order. These factors are likewise reported in [61]. The  $\alpha_s$ -dependent parentheses in the second line of eq. (2.48) account for the renormalization-group evolution between the UV scale of the process, of order  $M_Z$ , and the scale  $\mu$ . The latter scale is eventually fixed at 2 GeV, because at the mass scale of the external states  $\alpha_s$  is already in the non-perturbative regime. The  $J_3$  symbol in eq. (2.48) denotes a function depending on the QCD  $\beta$ -function as well as on the anomalous dimension of the effective operator  $\mathcal{Q}_1$ : the index 3 specifies the number of active flavors at the scale  $\mu$ . Its explicit form is again to be found in [61].<sup>25</sup> Finally the effective operator  $\mathcal{Q}_1$

<sup>25</sup> As well known, this RGE evolution is extremely important in order to sum to all orders the effects of large logs of the ratio  $M_Z/\mu$ . A calculation at order  $N^m$  LO is effectively of order  $\alpha_s^n(M_Z) \times (\alpha_s(M_Z) \log(M_Z/\mu))^m$ , with  $m$  arbitrary.

is defined as

$$\mathcal{Q}_1 \equiv (\bar{d}\gamma_L^\mu s) (\bar{d}\gamma_{\mu L} s) , \quad (2.49)$$

with the subscript  $L$  denoting a left-handed projector. This  $(V - A) \times (V - A)$  structure is the only one allowed within the SM. Plugging eq. (2.48) into the  $M_{12}$  definition, eq. (2.47), one sees that within the SM the only matrix element to calculate is

$$\langle K^0 | \mathcal{Q}_1 | \bar{K}^0 \rangle \equiv \frac{1}{3} m_K f_K^2 B_K(\mu) , \quad (2.50)$$

where  $f_K$  is the  $K$ -meson decay constant. The above relation defines the ‘bag’ parameter  $B_K(\mu)$ ,<sup>26</sup> that needs to be calculated non-perturbatively. Its  $\mu$  dependence cancels the one in the coefficient multiplying  $\mathcal{Q}_1$  in eq. (2.48). In phenomenological applications, it is customary and practical to define the renormalization-group-invariant parameter  $\hat{B}_K$  as

$$\hat{B}_K \equiv B_K(\mu) [\alpha_s(\mu)]^{-2/9} \left[ 1 + \frac{\alpha_s(\mu)}{4\pi} J_3 \right] , \quad (2.51)$$

namely in such a way to incorporate the QCD RGE factors in eq. (2.48) into the definition of the bag parameter itself.

\*\*\*

In view of the improvements on the long-distance corrections to  $\epsilon_K$ , also short-distance contributions have recently undergone a major reappraisal, thanks to work by Brod and Gorbahn, and it is worthwhile to open a parenthesis on this topic. Both the NNLO corrections (requiring the calculation of three-loop anomalous dimensions and two-loop matching conditions) to the charm-top factor  $\eta_{ct}$  and to the charm-charm factor  $\eta_{cc}$  in eq. (2.48) have been calculated, and presented in refs. [62] and [63], respectively. As a reference, let us first quote the  $\epsilon_K$  prediction at NLO, and with inclusion of the  $\kappa_\epsilon$  correction in eq. (2.76). It reads

$$|\epsilon_K|_{\text{SM}} = (1.82 \pm 0.26) \times 10^{-3} \quad [\text{NLO}] . \quad (2.52)$$

Inclusion of NNLO corrections to the charm-top term yields [62]

$$|\epsilon_K|_{\text{SM}} = (1.90 \pm 0.26) \times 10^{-3} \quad [\text{NNLO } \eta_{ct}] , \quad (2.53)$$

whereas inclusion of NNLO corrections to both charm-top and charm-charm terms results in [62]

$$|\epsilon_K|_{\text{SM}} = (1.81 \pm 0.28) \times 10^{-3} \quad [\text{NNLO } \eta_{ct} \text{ and } \eta_{cc}] , \quad (2.54)$$

which barely differs from the NLO prediction in eq. (2.52). This fact is actually not surprising, if one keeps in mind that the terms proportional to  $\eta_{tt}$ ,  $\eta_{ct}$  and  $\eta_{cc}$  contribute respectively +74%, +40% and -14% of the total  $|\epsilon_K|$  central value. Therefore, charm-top and charm-charm contributions, hence also their NNLO corrections, tend to partially compensate each other. An interesting outcome of ref. [63] is a large NNLO  $\eta_{cc}$  shift, and a considerable residual scale dependence, in view of which sizable corrections beyond NNLO may be expected.

**$\eta_{ct}$  &  $\eta_{cc}$   
at NNLO**

<sup>26</sup> In phenomenological applications that consider the full  $\Delta S = 2$  operator basis (consisting of eight operators), the parameter  $B_K$  is usually indicated with  $B_1$ .

## Subleading contributions to $\text{Im}M_{12}$

We are now ready to address the initial question of this section, namely whether contributions to  $\text{Im}M_{12}$  subleading in the OPE may be competitive with the  $\xi$  correction. Besides the leading, dimension-6  $O(G_F^2)$  contribution (2.48), subleading contributions to  $\text{Im}M_{12}$  include

- non-local contributions generated by the double-insertion of dimension-6  $O(G_F)$   $|\Delta S| = 1$  operators;
- local contributions generated by dimension-8  $O(G_F^2)$   $|\Delta S| = 2$  operators.

In terms of Feynman diagrams with effective-operator insertions, the relevant matrix element may be represented as follows

$$\text{Im}M_{12} = \text{Im}M_{12}^{(6)} + \text{Im}M_{12}^{\text{non-local}} + \text{Im}M_{12}^{(8)} + \dots =$$

$$\text{Im} \left( \begin{array}{c} \text{Diagram 1} \\ \uparrow \\ \text{dim-6 } O(G_F^2) \end{array} + \begin{array}{c} \text{Diagram 2} \\ \uparrow \quad \uparrow \\ \text{dim-6 } O(G_F) \quad \text{dim-6 } O(G_F) \end{array} + \begin{array}{c} \text{Diagram 3} \\ \uparrow \\ \text{dim-8 } O(G_F^2) \end{array} + \dots \right). \quad (2.55)$$

The first diagram in this expansion denotes the leading contribution, coming from the local dimension-6  $|\Delta S| = 2$  operator  $\mathcal{Q}_1$  in eq. (2.48), whose coefficient is of  $O(G_F^2)$ . This operator insertion is indicated with an empty square to remind that it is the local ‘remnant’ of the box diagram in fig. 1, at energies way below those of the heavy states, the  $W$  and the top, propagating in the box itself. We will henceforth refer to this contribution as  $\text{Im}M_{12}^{(6)}$ . The second diagram depicts one of the possible non-local contributions arising from two insertions of the leading  $|\Delta S| = 1$  four-quark operators, with Wilson coefficients of  $O(G_F)$ . The quark d.o.f. propagating in the loop are light ones, therefore these diagrams are rather calculated in Chiral Perturbation Theory (ChPT), whose d.o.f. are  $\pi$  and  $K$  mesons. This contribution will be referred to as  $\text{Im}M_{12}^{\text{non-local}}$ . The third diagram represents local contributions from dimension-8  $|\Delta S| = 2$  operators of  $O(G_F^2)$ , yielding  $\text{Im}M_{12}^{(8)}$ . Finally, ellipses denote other possible contributions from operators of dimension higher than 8 and/or with coupling order higher than  $G_F^2$ , that are negligible.

Let us dwell on the  $\text{Im}M_{12}^{\text{non-local}}$  and  $\text{Im}M_{12}^{(8)}$  contributions, starting from the latter. With respect to dimension-6 operators, contributions from dimension-8 ones will have a further suppression by two powers of UV mass scales. The largest contributions will therefore arise below the charm threshold, yielding a suppression factor of  $1/m_c^2$  relative to the leading dimension-6 contributions.<sup>27</sup> Therefore, by a purely dimensional argument, the contributions from these operators are expected to scale as  $m_K^2/m_c^2$  times the corresponding matrix elements from the leading dimension-6 operator. Note also that the CKM suppression factor of these dimension-8 operators is likewise the same as the CKM factor of the pure charm contribution in  $\mathcal{H}_{|\Delta S|=2}^{(6)}$  of eq. (2.48), namely  $(V_{cs}V_{cd}^*)^2$ , whose imaginary part is very small. On the whole, one naively expects a contribution of the order of 15% (the  $m_K^2/m_c^2$  ratio) times 15% (the size of the dimension-6 charm contribution in units of the total leading contribution), yielding about 2% [3]. According to the detailed analysis in ref. [64], the actual numerical impact is even smaller.

<sup>27</sup> Actually these operators do not appear in the quark effective Lagrangian, hence they do not come with short-distance suppression factors such as  $1/M_W^2$  [64].

The  $\text{Im}M_{12}^{\text{non-local}}$  contribution is instead non-negligible, and for consistency needs to be included whenever the  $\xi$  term is. The line of argument is as follows [3]. The key quantity to evaluate is

$$\mathcal{T}_{12} = -i \int d^4x \langle K^0 | T \left[ \mathcal{H}_{|\Delta S|=1}^{(u,c)}(x) \mathcal{H}_{|\Delta S|=1}^{(u,c)}(0) \right] | \bar{K}^0 \rangle, \quad (2.56)$$

where the  $(u, c)$  superscript denotes the two dynamical up-type quarks. The absorptive part of  $\mathcal{T}_{12}$  contributes to  $\text{Im}\Gamma_{12}$ , namely to  $\xi$ , whereas the dispersive part contributes to  $\text{Im}M_{12}$ . The crucial point is that within the  $|\epsilon_K|$  formula (2.46) these two contributions are multiplied by coefficients of the same order of magnitude. In fact, using eqs. (2.30) and (2.25), one may rewrite eq. (2.46) as

$$|\epsilon_K| = \sin \phi_\epsilon \left( \frac{\text{Im}M_{12}}{\Delta m_K} - \underbrace{\frac{\text{Im}\Gamma_{12}}{\Delta \Gamma_K}}_{+\xi} \right). \quad (2.57)$$

Since  $\Delta m_K$  and  $\Delta \Gamma_K$  are of similar size, it is clear that, if one is to include the  $\xi$  contribution, that comes from the absorptive part of  $\mathcal{T}_{12}$ , then by consistency also the dispersive part of  $\mathcal{T}_{12}$  should be included.

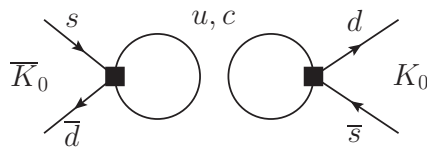
### $\text{Im}M_{12}^{\text{non-local}}$ within ChPT

For the purpose of evaluating  $\text{Im}M_{12}^{\text{non-local}}$  within ChPT, let us first rewrite  $\text{Im}M_{12}$  separating explicitly short- from long-distance contributions

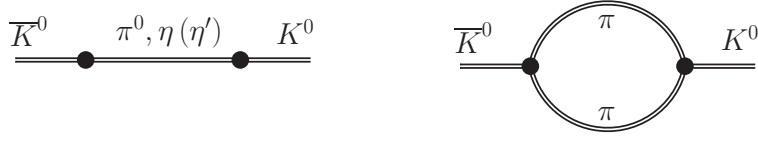
$$\text{Im}M_{12} = \text{Im}M_{12}^{\text{SD}} + \text{Im}M_{12}^{\text{LD}}, \quad \text{with} \quad \begin{cases} \text{Im}M_{12}^{\text{SD}} \equiv \text{Im}M_{12}^{(6)} \\ \text{Im}M_{12}^{\text{LD}} = \text{Im}M_{12}^{\text{non-local}} + \text{Im}M_{12}^{(8)} \end{cases}. \quad (2.58)$$

Both of the terms in  $\text{Im}M_{12}^{\text{LD}}$  are in principle captured by the matrix element in eq. (2.56). Its direct evaluation would therefore provide the best strategy, also because the result would be manifestly scale-independent, whereas  $\text{Im}M_{12}^{\text{non-local}}$  and  $\text{Im}M_{12}^{(8)}$  are separately not so. However, the evaluation of the matrix element (2.56) is far from being straightforward in first-principle approaches such as lattice QCD, particularly because of the presence of disconnected diagrams, such as the one depicted in fig. 2.

We will therefore adopt an effective-theory strategy, based on ChPT and on the phenomenological enhancement of  $\Delta I = 1/2$  amplitudes. In the ChPT framework, one trades quark and gluon fields for  $\pi$ ,  $K$  and  $\eta$  fields, identified with the would-be Goldstone bosons arising from the  $SU(3)_L \times SU(3)_R \rightarrow SU(3)_{L+R}$  symmetry breaking of the QCD action in the limit of vanishing light quark masses. One thereby writes an effective Lagrangian in terms of the pseudo-Goldstone boson fields, and evaluates low-energy amplitudes involving these mesons as expansions in powers of their masses and momenta. The lowest-order effective Lagrangian describing non-leptonic  $|\Delta S| = 1$  decays consists of only two operators, transforming as  $(8_L, 1_R)$  and respectively  $(27_L, 1_R)$  under the  $SU(3)_L \times SU(3)_R$  chiral group. Between them, only the  $(8_L, 1_R)$  operator has a phenomenologically large coefficient, having to be responsible for the enhancement of  $\Delta I = 1/2$  amplitudes. This



**Figure 2:** Example of disconnected diagram involved in the evaluation of  $\mathcal{T}_{12}$ , eq. (2.56).



**Figure 3:** Tree-level and one-loop diagrams contributing to  $\bar{K}^0$ - $K^0$  mixing in ChPT.

Lagrangian term reads

$$\mathcal{L}_{|\Delta S|=1}^{(2)} = F^4 G_8 \left( \partial^\mu U^\dagger \partial_\mu U \right)_{23} + \text{h.c.} \quad (2.59)$$

and for the above reason represents the only term in the ChPT Lagrangian relevant to our calculation. In eq. (2.59), we have defined as usual

$$U = \exp(i\sqrt{2}\Phi/F), \quad \Phi = \begin{bmatrix} \frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & \pi^+ & K^+ \\ \pi^- & -\frac{\pi^0}{\sqrt{2}} + \frac{\eta}{\sqrt{6}} & K^0 \\ K^- & \bar{K}^0 & -\frac{2\eta}{\sqrt{6}} \end{bmatrix}, \quad (2.60)$$

and  $F$  can be identified with the pion decay constant in the chiral limit ( $F \simeq 92$  MeV). The effective coupling  $G_8$  can be determined by using the  $\Delta I = 1/2$  rule. To this purpose, one can calculate with the interaction in eq. (2.59) the tree-level contribution to the amplitude already defined in eq. (2.16), obtaining

$$\mathcal{A}(K^0 \rightarrow (\pi\pi)_0) = \sqrt{2}FG_8(m_K^2 - m_\pi^2) = a_0 e^{i\delta_0} \equiv A_0. \quad (2.61)$$

This implies  $|G_8| \simeq 9 \times 10^{-6} (\text{GeV})^{-2}$ , namely that  $G_8$  has Fermi-coupling strength. Furthermore, the imaginary over real part of this relation allows to identify the weak phase of  $G_8$  with that of  $a_0$ :

$$\frac{\text{Im}G_8}{\text{Re}G_8} = \xi. \quad (2.62)$$

Having determined both the magnitude and the weak phase of  $G_8$ , we can now proceed to estimate the  $\mathcal{L}_{|\Delta S|=1}^{(2)}$  contribution to  $M_{12}$ . In principle  $\mathcal{L}_{|\Delta S|=1}^{(2)}$  could contribute to  $M_{12}$  already at  $\mathcal{O}(p^2)$ , via the tree-level diagram in fig. 3 (leftmost diagram). However, this contribution vanishes [65] when one takes into account the  $\mathcal{O}(p^2)$  relation among  $\pi^0$ ,  $\eta$  and kaon masses, namely the Gell-Mann–Okubo mass formula. Hence the first non-vanishing contribution to  $M_{12}$  generated by  $\mathcal{L}_{|\Delta S|=1}^{(2)}$  arises only at  $\mathcal{O}(p^4)$ .

At  $\mathcal{O}(p^4)$ , contributions include loop diagrams with two insertions of  $\mathcal{L}_{|\Delta S|=1}^{(2)}$ , as well as tree diagrams with insertion of appropriate  $\mathcal{O}(p^4)$  counterterms. Among all these terms, the only model-independent, and presumably dominant, contribution to  $M_{12}$  is the non-analytic one generated by the pion-loop amplitude in fig. 3 (rightmost diagram),

$$\begin{aligned} T_{12}^{(\pi\pi)} = \mathcal{A}^{(\pi\pi)}(\bar{K}^0 \rightarrow K^0) &= -\frac{3}{16\pi^2} F^2 (G_8^*)^2 (m_K^2 - m_\pi^2)^2 \times \\ &\times \left[ \sqrt{1 - 4r_\pi^2} \left( \log \frac{1 + \sqrt{1 - 4r_\pi^2}}{1 - \sqrt{1 - 4r_\pi^2}} - i\pi \right) + \log \left( \frac{m_\pi^2}{\mu^2} \right) \right], \end{aligned} \quad (2.63)$$

with  $r_\pi^2 = m_\pi^2/m_K^2$  and where we have absorbed all finite (mass-independent) terms in the definition of the renormalization scale  $\mu$ . *This is the only contribution which has an absorptive*

part. As a consequence, its weak phase can be unambiguously related to the weak phase of the  $K^0 \rightarrow (\pi\pi)_{I=0}$  amplitude to all orders in the chiral expansion. In addition, it is the only contribution that survives in the  $SU(2)_L \times SU(2)_R$  limit of ChPT, which is known to represent a good approximation of the full  $O(p^4)$  amplitude in several  $K$ -decay observables where contributions from counterterms are fully under control (see e.g. ref. [66]).

A ChPT calculation of  $M_{12}$  complete to  $O(p^4)$  would require consideration of loops involving kaons and  $\eta$ 's, as well as  $O(p^4)$  local counterterms. However, all these additional pieces are not associated with any physical cut. As such, they can effectively be treated as a local term whose overall weak phase cannot be related to the phase of the  $K^0 \rightarrow (\pi\pi)_{I=0}$  amplitude. On account of the above considerations, in ref. [3] we refrained from a full  $O(p^4)$  ChPT calculation, and we focused on the pion-loop non-analytic contribution only. Using the relation  $T_{12}^{(\pi\pi)} = 2m_K M_{12}^{(\pi\pi)}(\mu)$ , the result in eq. (2.63) implies

$$M_{12}^{(\pi\pi)}(\mu) = -\frac{3}{64\pi^2 m_K} (\mathcal{A}_0^*)^2 \left[ \log \left( \frac{m_K^2}{\mu^2} \right) + O \left( \frac{m_\pi^2}{m_K^2} \right) \right]. \quad (2.64)$$

with  $A_0$  defined in eq. (2.61). The absorptive part in eq. (2.63) is nothing but the leading  $|(\pi\pi)_{I=0}\rangle$  contribution to  $\Gamma_{12}$ , which gives rise to the relation (2.30). The dispersive part is the dominant contribution to  $M_{12}$  in the leading-log approximation. The close link of these two terms is a further confirmation that we cannot neglect the long-distance contribution to  $\text{Im}M_{12}$  if we want to keep track of all the  $O(\xi)$  terms in  $\epsilon_K$ .

The result in eq. (2.64) can now be used to estimate the corresponding contribution to  $\text{Im}M_{12}$ . Setting  $\mu = 800$  GeV and varying it in the interval  $0.6 - 1$  GeV leads to

$$\frac{\Delta m_K^{\text{LD}}|_{G_8^2}}{\Delta m_K^{\text{exp}}} = \frac{2\text{Re}M_{12}^{(\pi\pi)}}{\Delta m_K^{\text{exp}}} = 0.4 \pm 0.2. \quad (2.65)$$

We now need to discuss how our calculated quantity  $\Delta m_K^{\text{LD}}|_{G_8^2}$  is related with our main quantity of interest in this section, namely  $\text{Im}M_{12}^{\text{LD}}$ , eq. (2.58). In the framework of ChPT, the latter can be decomposed as a leading term arising from the Lagrangian (2.59), hence proportional to  $G_8^2$  and a subleading term with different effective coupling

$$\text{Im}M_{12}^{\text{LD}} = \text{Im}M_{12}^{\text{LD}}|_{G_8^2} + \text{Im}M_{12}^{\text{LD}}|_{\text{non-}G_8^2}. \quad (2.66)$$

Recalling eq. (2.63) we can write

$$\text{Im}M_{12}^{\text{LD}}|_{G_8^2} = \text{Re}M_{12}^{\text{LD}}|_{G_8^2} \times \frac{\text{Im}[(G_8^*)^2]}{\text{Re}[(G_8^*)^2]}, \quad (2.67)$$

and identify the weak phase of  $G_8$  with  $\xi$  using eq. (2.62). It follows

$$\text{Im}M_{12}^{\text{LD}}|_{G_8^2} \approx \text{Re}M_{12}^{\text{LD}}|_{G_8^2} \times (-2\xi) \approx -\xi \times (\Delta m_K^{\text{LD}}|_{G_8^2}). \quad (2.68)$$

This relation, connecting our calculated quantity (2.65) with  $\text{Im}M_{12}^{\text{LD}}$ , that appears in the  $\epsilon_K$  formula, is what we were after. Using eqs. (2.58) and (2.68), eq. (2.46) can be rewritten as follows

$$|\epsilon_K| = \sin \phi_\epsilon \left[ \frac{\text{Im}M_{12}^{(6)}}{\Delta m_K} + \xi \left( 1 - \frac{\Delta m_K^{\text{LD}}|_{G_8^2}}{\Delta m_K} \right) + \delta_{\text{Im}M_{12}} \right], \quad (2.69)$$

where  $\delta_{\text{Im}M_{12}}$  encodes the subleading terms in  $\text{Im}M_{12}^{\text{LD}}|_{\text{non-}G_8^2}$  (including also  $\text{Im}M_{12}^{(8)}$ ). Eq. (2.69) is our final formula for  $\epsilon_K$ . This equation shows that the long-distance contribution

**Im** $M_{12}^{\text{LD}}$ :  
**result**



to  $\text{Im}M_{12}$ , estimated in eq. (2.65) tends to cancel against the corresponding long-distance contribution to  $\text{Im}\Gamma_{12}$ , parameterized by  $\xi$ . Note in this respect that the result in eq. (2.65) has a well-defined sign since  $G_8$  appears squared in  $M_{12}^{(\pi\pi)}$ . According to our estimate, this cancellation is only partial, at the level of 40%. In the limit where the contribution of  $G_8$  saturates  $\Delta m_K$ , the  $\xi$  term would be absent altogether, i.e. the long-distance contributions to  $\text{Im}M_{12}$  and to  $\text{Im}\Gamma_{12}$  would cancel exactly. This is precisely what we should expect, since in this limit  $\text{Im}M_{12}$  and  $\text{Im}\Gamma_{12}$  would have the same weak phase but for the short-distance contribution to  $\text{Im}M_{12}$ .

### Consistency check

Since our estimate of  $\Delta m_K^{\text{LD}}|_{G_8^2}$  is not the result of a complete calculation at fixed order in the chiral expansion, it is worthwhile to cross-check it using a different argument. For this purpose, we note that the only relevant contribution to  $M_{12}$ , beside the one from the two-pion intermediate state estimated above, is expected to arise from tree-level  $\eta'$  exchange (fig. 3 left) [67]. We can thus decompose  $M_{12}$  as follows:

$$M_{12} \approx M_{12}^{\text{SD}} + M_{12}^{\text{LD}}|_{\pi\pi} + M_{12}^{\text{LD}}|_{\eta'} . \quad (2.70)$$

According to this decomposition it is clear that, as far as long-distance contributions are concerned, we can trade the evaluation of  $M_{12}^{\text{LD}}|_{\pi\pi}$  for that of  $M_{12}^{\text{LD}}|_{\eta'}$ . An estimate of the  $\eta'$  contribution to  $M_{12}$  goes beyond pure ChPT, where it can be considered as a free parameter – the leading contribution to the  $\mathcal{O}(p^4)$  local terms. However, its impact can be estimated in the large  $N_c$  limit, extending the underlying symmetry from  $SU(3)_L \times SU(3)_R$  to  $U(3)_L \times U(3)_R$ . Within this framework the operator basis must be extended and we cannot directly relate the phase of the  $\eta'$  exchange amplitude to the phase of  $G_8$ . According to the recent analysis in ref. [67], the  $\eta'$  amplitude gives a negative contribution to  $\Delta m_K$ :

$$2\text{Re}M_{12}^{\text{LD}}|_{\eta'} = \Delta m_K^{\text{LD}}|_{\eta'} \approx -0.3 \Delta m_K^{\text{exp}} . \quad (2.71)$$

Most importantly for our analysis, this contribution is found to be induced at the quark level by the operator  $(\bar{s}d)_{V-A} \times (\bar{u}u)_{V-A}$  only [67]. This implies that the  $\eta'$  exchange has a vanishing weak phase in the standard CKM phase convention:

$$\text{Im}M_{12}^{\text{LD}}|_{\eta'} = 0 . \quad (2.72)$$

Using this result in eq. (2.46), and using the relation (2.68) for the  $\pi\pi$  contribution, we get

$$|\epsilon_K| = \sin \phi_\epsilon \left[ \frac{\text{Im}M_{12}^{(6)}}{\Delta m_K^{\text{exp}}} + \xi \frac{\Delta m_K^{\text{SD}} + \Delta m_K^{\text{LD}}|_{\eta'}}{\Delta m_K^{\text{exp}}} \right] . \quad (2.73)$$

Denoting as  $\rho$  the fraction multiplying  $\xi$  in eq. (2.73), and combining eq. (2.71) with the NLO short-distance estimate of  $\text{Re}M_{12}$ , namely  $\Delta m_K^{\text{SD}} = (0.7 \pm 0.1)\Delta m_K^{\text{exp}}$  [61, 68], we get  $\rho = 0.4 \pm 0.1$ . This result is well consistent with the value  $\rho = 0.6 \pm 0.2$  obtained from eq. (2.69) with the direct evaluation of the  $\pi\pi$  contribution in eq. (2.65).

\*\*\*

We rate the direct evaluation of the  $\pi\pi$  loop as the most reliable estimate of  $\rho$ . As a consequence, our final phenomenological expression for  $\epsilon_K$  is [3]

**$\epsilon_K$ : final formulae**

$$\epsilon_K = \sin \phi_\epsilon e^{i\phi_\epsilon} \left[ \frac{\text{Im}M_{12}^{(6)}}{\Delta m_K} + \rho \xi \right] \quad \text{with} \quad \rho = 0.6 \pm 0.3 , \quad (2.74)$$

where we have conservatively increased by 50% the error in eq. (2.65) to take into account the sub-leading contributions of  $\text{Im}M_{12}^{\text{LD}}|_{\text{non-}G_8^2}$ . For  $\rho = 1$  the result in eq. (2.74) reduces to the one in [1]. The contribution calculated in [3], resulting in  $\rho < 1$ , completes the estimate of the terms of  $\mathcal{O}(\xi)$  in  $\epsilon_K$ .

Following the notation of ref. [1], we summarize the corrections to  $\epsilon_K$  due to LD effects and  $\phi_\epsilon \neq 45^\circ$ , via the introduction of the phenomenological factor  $\kappa_\epsilon$ , defined by

$$\epsilon_K = \kappa_\epsilon \frac{e^{i\phi_\epsilon}}{\sqrt{2}} \left[ \frac{\text{Im}M_{12}^{(6)}}{\Delta m_K} \right]. \quad (2.75)$$

According to eq. (2.65), and taking into account the estimate of  $\xi$  obtained in [1], namely  $\xi = -(6.0 \pm 1.5) \times 10^{-2} \times \sqrt{2}|\epsilon_K|$ , the new numerical value of  $\kappa_\epsilon$  is [3]

$$\kappa_\epsilon = \frac{\sin \phi_\epsilon}{1/\sqrt{2}} \times \left( 1 + \rho \frac{\xi}{\sqrt{2}|\epsilon_K|} \right) = 0.94 \pm 0.02. \quad (2.76)$$

This should be compared with  $0.92 \pm 0.02$  in [1] and  $0.92 \pm 0.01$  in [69], where only the long-distance contributions to  $\text{Im}\Gamma_{12}$  (not those to  $\text{Im}M_{12}$ ) have been included. As mentioned, there is an ongoing effort in the lattice-QCD community (see [48, 49] and talk by A. Jüttner at [50]) to evaluate eq. (2.76), or in fact eq. (2.56) that is the most challenging part, from first principles.

## 2.6 Status of the $\epsilon_K - \sin 2\beta$ tension

The discussion in this whole chapter can be summarized by eq. (2.74) [3], that supersedes the corresponding  $\epsilon_K$  formula proposed in ref. [1]. Eq. (2.74) can be used in phenomenological applications, including CKM fits [32, 70] and analyses aimed at constraining new physics within specific models (an example among many being [71]). In this section we would like to summarize the  $\epsilon_K - \sin 2\beta$  tension [1],<sup>28</sup> discuss some of its implications, and its status as of the time of writing of the present manuscript.

### Parametric error in $\epsilon_K$

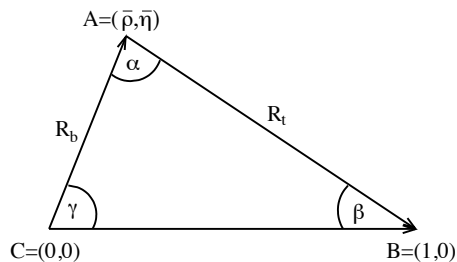
In fact, we already introduced the essential aspects of this tension at the beginning of this chapter (see end of sec. 2.1), but discussed only central values. We would like now to convey an intuitive picture of the main components to the  $\epsilon_K$  *parametric* error. In fact, we expect that, once the long-distance shifts discussed in the previous sections, and summarized by the  $\xi$  term in eq. (2.74), are properly included, the residual *systematic* error be negligible. By looking at eq. (2.74) it is clear that the bulk of the parametric uncertainty in  $\epsilon_K$  arises from  $\text{Im}M_{12}^{(6)}$  – in fact  $\phi_\epsilon$  and  $\Delta m_K$  can be taken from data, and have a negligible error, and the  $\xi$  term is itself a correction.  $\text{Im}M_{12}^{(6)}$  has been discussed in some detail in sec. 2.5. Eq. (2.48) in particular, displaying the relevant effective Hamiltonian, allows to easily identify the main components to the parametric error. A first obvious component is the bag parameter  $\hat{B}_K$ , eq. (2.51). At the time of ref. [1], the reference value used was the one from a then-new lattice-QCD evaluation, yielding  $\hat{B}_K = 0.720(13)(37)$  [73], lower by roughly 9% with respect to previous determinations. The  $\hat{B}_K$  5% error, and the fact that  $\hat{B}_K$  enters only linearly in the  $\epsilon_K$  formula, made the  $\hat{B}_K$  error play a subdominant role with respect to other error components to be discussed next. This is even more true today, considering that lattice-QCD

<sup>28</sup> For a different point of view on the same tension, see [72].

evaluations of  $\hat{B}_K$  have reached a striking  $O(1\%)$  error, as we will discuss in more detail in chapter 4.

We conclude that the dominant component to the  $\epsilon_K$  uncertainty must be in the CKM factor, that we need to analyze in more detail. To this end, it is useful to introduce a specific parameterization of the CKM matrix. As already remarked, the CKM matrix is described by four physical parameters. A possible parameterization, particularly useful here, is in terms of  $\lambda$ ,  $\sin 2\beta$ ,  $V_{cb}$  and  $R_t$  [74], where  $R_t$  denotes one of the sides of the ‘Unitarity Triangle’ depicted in fig. 4. The latter is constructed by taking the first and the complex conjugate of the third column of the CKM matrix (for an introduction see ref. [53]). In terms of the usual Wolfenstein parameters  $\lambda$ ,  $A$ ,  $\bar{\rho}$  and  $\bar{\eta}$ , one has

$$\begin{aligned} R_t &\equiv \sqrt{(1 - \bar{\rho})^2 + \bar{\eta}^2} , \\ \sin 2\beta &= \frac{2\bar{\eta}(1 - \bar{\rho})}{R_t^2} , \\ |V_{cb}| &= A\lambda^2 + O(\lambda^8) , \end{aligned} \quad (2.77)$$



**Figure 4:** The Unitarity Triangle (figure taken from [75]).

so that indeed the parameterization in terms of  $\lambda$ ,  $\sin 2\beta$ ,  $V_{cb}$  and  $R_t$  is equivalent to the Wolfenstein one – let alone issues of accuracy in  $\lambda$ . Let us now recall that the aim of this re-parameterization is to help us intuitively understand the magnitude and the origin of the CKM error in  $\epsilon_K$ . We can easily work this out by recalling that the dominant CKM dependence in  $\epsilon_K$  arises from a top-top loop, yielding

$$\text{Im}(V_{ts}V_{td}^*)^2 \simeq \text{Im}[-A\lambda^2 \cdot A\lambda^3(1 - \bar{\rho} + i\bar{\eta})]^2 \propto |V_{cb}|^4 \lambda^2 (1 - \bar{\rho}) \bar{\eta} \propto |V_{cb}|^4 \lambda^2 R_t^2 \sin 2\beta \quad (2.78)$$

where in the last two members we have used eqs. (2.77). Eq. (2.78) displays that the dominant CKM combination in  $\epsilon_K$  depends on the fourth power of  $|V_{cb}|$ . This CKM entry is incidentally among the most theoretically debated at present, because of the marginal consistency between determinations from exclusive and inclusive semileptonic  $b \rightarrow c$  decays. In ref. [1] we took

$$|V_{cb}| = (41.2 \pm 1.1) \cdot 10^{-3} , \quad (2.79)$$

as the average quoted in the 2008 edition of the PDG.<sup>29</sup> Eq. (2.79) corresponds to a 2.7% error on  $|V_{cb}|$ , implying an 11% relative error on  $\epsilon_K$ . This constitutes by far the dominant component to the  $\epsilon_K$  error.

Let us next turn to  $R_t$ . To get an idea of the contributed uncertainty, we note that  $R_t$  can be calculated from (see e.g. [74])

$$R_t \approx \frac{\xi_s}{\lambda} \sqrt{\frac{m_{B_s}}{m_{B_d}}} \sqrt{\frac{\Delta m_d}{\Delta m_s}} , \quad (2.80)$$

where  $m_{B_{d,s}}$  are the masses of the  $B_{d,s}$ -mesons,  $\Delta m_{d,s}$  the mass differences between the mass

<sup>29</sup> The  $|V_{cb}|$  average in the latest PDG review [26] reads  $|V_{cb}| = (40.9 \pm 1.5) \cdot 10^{-3}$ . To be noted is the even larger error with respect to eq. (2.79).

eigenstates in the  $B_{d,s}$ -systems, and  $\xi_s$  denotes the ratio

$$\xi_s = \frac{f_{B_s} \sqrt{\hat{B}_{B_s}}}{f_{B_d} \sqrt{\hat{B}_{B_d}}}, \quad (2.81)$$

involving the  $B_{d,s}$ -meson decay constants  $f_{B_{d,s}}$  and the bag parameters of the analogues for the  $B_{d,s}$ -systems of the operator  $\mathcal{Q}_1$  in eq. (2.49). The parameter  $\xi_s$  can be determined with accuracy of O(5%) from lattice QCD, hence one can expect the  $R_t$  error to be of the same order. In practice, CKM-fit determinations of  $R_t$  tend to give a somewhat better uncertainty. We can assume, as in ref. [1],  $R_t \simeq 0.92 \cdot (1 \pm 3.5\%)$ , implying about 7% error at the level of  $\epsilon_K$ . By inspection of recent CKM fits [32, 70] within the SM, one can realize that this  $R_t$  figure has barely changed. However, as evident from eq. (2.80),  $R_t$  is sensitive to new-physics contributions, in particular those affecting  $\Delta m_{d,s}$ . Hence its central value as well as its error do change according to the new-physics scenario assumed. The  $R_t$  value mentioned above assumes the SM, or the wider class of models characterized by MFV, whereby new physics contributions cancel in the  $\Delta m_d/\Delta m_s$  ratio, as discussed in ref. [36]. The issue of the  $R_t$  model dependence will be settled once an accurate determination of the CKM angle  $\gamma$  will be available. This can be understood by noting that  $R_t$  and  $\gamma$  are opposite to one another in the Unitarity Triangle of fig. 4.

Finally  $\lambda$  is very well known and  $\sin 2\beta$  enters only linearly, therefore both of them contribute negligibly to the  $\epsilon_K$  error with respect to  $R_t$  and especially  $|V_{cb}|$ . Putting together these considerations, for the total  $\epsilon_K$  uncertainty ref. [1] obtained

$$\frac{\sigma_{\epsilon_K}}{\epsilon_K} \approx \underbrace{4 \cdot \frac{\sigma_{|V_{cb}|}}{|V_{cb}|}}_{11\%} \oplus \underbrace{2 \cdot \frac{\sigma_{R_t}}{R_t}}_{7\%} \oplus \underbrace{\frac{\sigma_{\hat{B}_K}}{\hat{B}_K}}_{5\%} \approx 14\%. \quad (2.82)$$

Among the error components in eq. (2.82), the  $\hat{B}_K$  error has today shrunk to a negligible level, as already mentioned above. The  $R_t$  relative error will likewise be significantly reduced with progress on the  $\gamma$  determination. Therefore, the  $\epsilon_K$  theoretical error – and with it the final word on the  $\epsilon_K - \sin 2\beta$  tension – will be in the hands of  $|V_{cb}|$ . On both  $|V_{cb}|$  and  $\gamma$  we will pause again in chapter 4, to discuss their prospects of improvement.

### The $\epsilon_K - \sin 2\beta$ tension as of refs. [1–3]

In short, the  $\epsilon_K - \sin 2\beta$  tension [1] is the tension between [3]

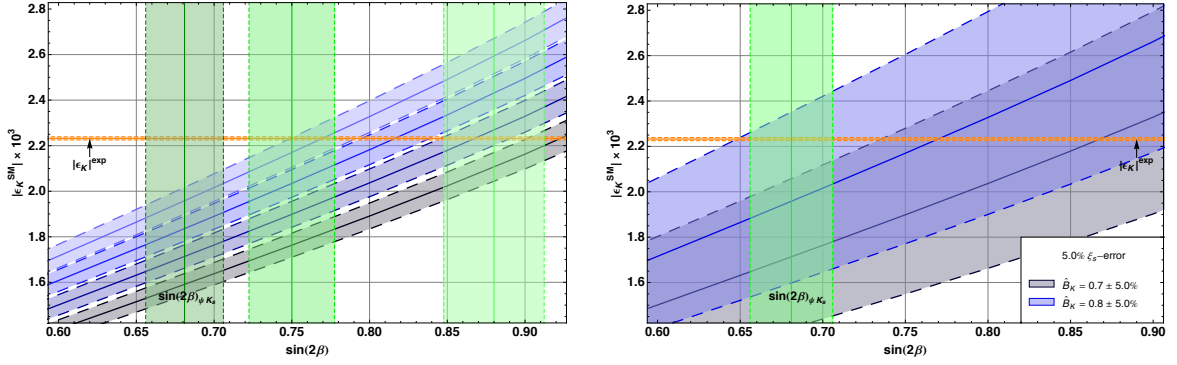
$$|\epsilon_K|_{\text{SM}} = (1.82 \pm 0.26) \times 10^{-3}, \quad (2.83)$$

whose 14% error is now clear in the light of eq. (2.82), and

$$|\epsilon_K|_{\text{exp}} = (2.229 \pm 0.012) \times 10^{-3}. \quad (2.84)$$

The ‘low’ central value in eq. (2.83) was mostly the result of two downward shifts of roughly the same magnitude: the one due to  $\kappa_\epsilon$  and the one caused by the low  $\hat{B}_K$  value of ref. [73], as already mentioned above.

Ref. [1] pointed out three scenarios to accommodate the mentioned tension. One first possibility was to advocate no new CP violation in  $K^0 - \bar{K}^0$  mixing. In this case, the size of the measured value of  $\epsilon_K$  implied  $\sin 2\beta$  by 10-20% larger [72] than  $\sin 2\beta_{\psi K_s}$ , so that a new, negative phase  $\phi_d$  was required in order to fit the experimental value of  $\sin 2\beta_{\psi K_s}$ . A second possibility was, on the contrary, that of assuming no new CP violation in the  $B_d$ -system. In



**Figure 5:** Left panel: allowed regions in the  $|\epsilon_K|_{\text{SM}}$  vs.  $\sin 2\beta$  plane. The blue areas delimited by dashed lines display the cases  $\hat{B}_K \in \{0.65, 0.70, 0.75, 0.80\} \pm 3\%$  (darker to lighter blue, or lower to higher), and no other error components are included. Vertical areas display the three advocated scenarios, with  $\sin 2\beta \in \{0.681, 0.75, 0.88\} \pm 3.7\%$  (darker to lighter green, or left to right). Right panel: same as left panel, but with inclusion of all input uncertainties. Figures taken from ref. [1], including assumed uncertainties. See text for more details.

this case, the measured size of  $S_{\psi K_s}$  implied  $\epsilon_K$  as in eq. (2.83), namely as much as 20% below the data, hinting at new CP-violating physics in  $K^0 - \bar{K}^0$  mixing. Finally, a third scenario involved two smaller new-physics phases in both the  $K^0$ - and the  $B_d$ -systems. These three scenarios can be visualized in fig. 5, taken from ref. [1].

The left panel of this figure reports  $|\epsilon_K|_{\text{SM}}$  as a function of  $\sin 2\beta$  for  $\hat{B}_K \in \{0.65, 0.70, 0.75, 0.80\} \pm 3\%$ . The vertical ranges centered at  $\sin 2\beta \in \{0.681, 0.75, 0.88\}$ , with a relative error chosen at 3.7% as in the  $\sin 2\beta_{\psi K_s}$  case, define the scenarios in question. Specifically, in the region where  $|\epsilon_K|_{\text{SM}} \approx |\epsilon_K|_{\text{exp}}$ , as in the first scenario, namely the rightmost region where the blue bands cross the horizontal  $|\epsilon_K|_{\text{exp}}$  band, no new contributions are required to fit the data on  $\epsilon_K$ , even for  $\hat{B}_K \approx 0.65$ . However, the  $\sin 2\beta$  prediction (vertical green band) tends to be in excess of the  $\sin 2\beta_{\psi K_s}$  determination, revealing the need for a new phase in the  $B_d$  system as large as  $-9^\circ$ . Conversely, in the region where  $\sin 2\beta \approx \sin 2\beta_{\psi K_s}$  (leftmost vertical green band), the discrepancy between the  $|\epsilon_K|_{\text{SM}}$  prediction and data would require a new contribution to  $\epsilon_K$ , unless  $\hat{B}_K \gtrsim 0.85$ .

In the left panel of fig. 5 only the  $\epsilon_K$  error component due to  $\hat{B}_K$  (at the time of [1] an important one) was included. An analogous plot, but with all uncertainties on the input taken into account, is shown in the right panel of fig. 5, also taken from [1].

The first mentioned scenario of a single additional  $B_d$ -mixing phase was found to be especially interesting. In this instance, the SM formula for the mixing-induced CP asymmetry  $S_{\psi K_s}$  generalizes to

$$S_{\psi K_s} = \sin(2\beta + 2\phi_d) = 0.681 \pm 0.025, \quad (2.85)$$

where  $\phi_d$  is the new phase. The information mentioned above points toward a small *negative* value of  $\phi_d$ . On the other hand, the mixing-induced CP asymmetry  $S_{\psi\phi}$  is given by [74]

$$S_{\psi\phi} = \sin(2|\beta_s| - 2\phi_s), \quad (2.86)$$

where the SM phases  $\beta, \beta_s$  are defined from the CKM matrix entries  $V_{td}, V_{ts}$  through

$$V_{td} = |V_{td}|e^{-i\beta}, \quad V_{ts} = -|V_{ts}|e^{-i\beta_s}, \quad (2.87)$$

with  $\beta_s \approx -1^\circ$ . The interesting point (again, at the time of ref. [1]) was that, using eq. (2.86), one could find that a negative  $\phi_s$  was also required to explain the claim of a large

phase in the  $B_s$ -system [76].<sup>30</sup> Using the analysis in [76], the range for the new phase  $\phi_s$  at 95% probability was found to be

$$\begin{aligned} \phi_s &\in [-30.45, -9.29]^\circ \cup [-78.45, -58.2]^\circ, \\ &\text{corresponding to } S_{\psi\phi} \in [0.35, 0.89], \end{aligned} \quad (2.88)$$

whereas, assuming generic new physics, the SM contribution to the  $B_s$ -system phase amounted to  $\beta_s = -1.17(11)^\circ$  [76]. It was then tempting, and meaningful, to investigate whether, at least to first approximation, the same new phase  $\phi_d \approx \phi_s \approx \phi_B$  could fit in both  $B_d$  and  $B_s$  systems, being a small correction in the former case – where the SM phase is large – and the bulk of the effect in the latter. As a matter of fact ref. [1] found

$$\phi_B = \phi_d \approx \phi_s \approx -9^\circ \Rightarrow \begin{cases} \beta_{\psi K_s} < \beta \approx 30^\circ \\ S_{\psi\phi} \approx 0.4 \end{cases} \quad (2.89)$$

in perfect consistency with eq. (2.88).

\*\*\*

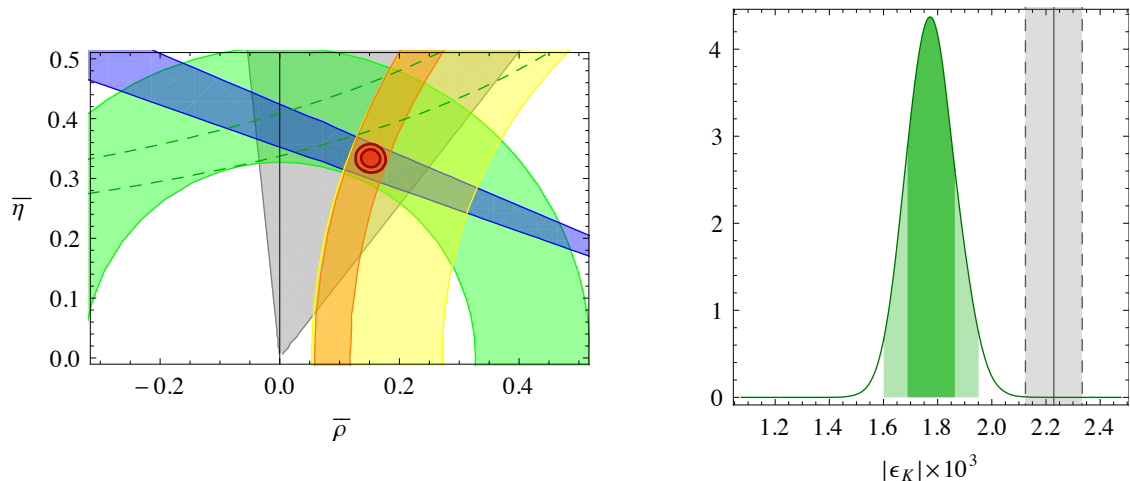
After ref. [1], numerous studies followed up, both in the context of CKM fits, and of specific model setups.<sup>31</sup> An interesting example of this second instance is ref. [71]. The authors start from a global fit of the CKM matrix using only tree-level and  $\Delta F = 2$  observables. The status, as of the time of ref. [71], of the CKM-fit compatibility with the SM can be appreciated from fig. 6, obtained excluding  $|\epsilon_K|$  from the fit itself. To be noted in particular is the fact that in the global fit (leftmost panel) the  $2\sigma$  band for  $\epsilon_K$  (enclosed by dashed green lines), is outside the region (enclosed by red solid lines) that corresponds to the  $2\sigma$  fit minimum. The fit projection onto  $\epsilon_K$  is displayed in the rightmost panel. On the basis of these results, ref. [71] proposed to move from the  $U(3)^3$  approximate symmetry normally invoked within Minimal Flavor Violation (see corresponding discussion in sec. 1.5) to a suitably broken  $U(2)^3$ , where the ‘2’ refers to the first two generations of quarks and squarks. Besides being able to perfectly account for the  $\epsilon_K - \sin 2\beta$  discrepancy, this symmetry is appealing in that it provides a rationale for lighter (heavier) first and second generations of quarks (squarks) with respect to the third generation. It is to be noted, in particular, that third-generation top partners lighter than first- and second-generation ones (the so-called inverted mass hierarchy) are particularly welcome from a theoretical point of view, in that third-generation masses are the ones that have the largest impact on naturalness constraints and on account of this are required to be not far away from the weak scale.

A persistent  $\epsilon_K - \sin 2\beta$  tension, and evidence of direct production of top-partner states at the LHC, would give strong circumstantial evidence for this scenario, and would represent a perfect example of the interplay between indirect and direct searches for new physics.

Unfortunately, as of the time of this writeup, there is no  $\epsilon_K - \sin 2\beta$  tension. This is the result of several changes on the input parameters to  $\epsilon_K$ : in particular, the  $|V_{cb}|$  uncertainty has been meanwhile somewhat increased, as noted in footnote 29; the central value of the  $\hat{B}_K$  average went slightly up; the  $\kappa_\epsilon$  correction has decreased in magnitude after the more refined analysis in ref. [3]. On a related note, it should also be remarked that the

<sup>30</sup> After the UTfit claim [76], this exciting possibility was also confirmed by early Tevatron data. Unfortunately, it was then disproved by more statistics and by LHCb measurements.

<sup>31</sup> In turn, the updated analysis in ref. [3] drifted attention towards a first-principle determination of the  $\kappa_\epsilon$  correction, spurring lattice-QCD studies, as mentioned elsewhere.

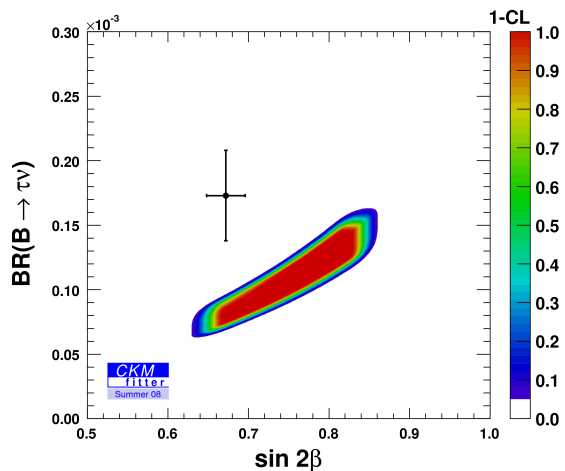


**Figure 6:** Global CKM fit using tree-level and  $\Delta F = 2$  observables only. The green dashed lines on the leftmost plot denote the  $2\sigma$  region of compatibility with the  $\epsilon_K$  constraint. The fit minimum is represented by the region delimited by red solid lines (1 and  $2\sigma$  contours). The rightmost plot represents the projection of the fit on the  $\epsilon_K$  constraint. Green shaded areas delimit the 1 and  $2\sigma$  regions for  $\epsilon_K$ , whereas the gray shaded area denotes the experimental measurement. Figure taken from ref. [71].

$B \rightarrow \tau\nu$  branching ratio average between the  $B$  factories is now consistent with the SM expectation. This experimental quantity contributed indirectly to the  $\epsilon_K - \sin 2\beta$  tension, because its high central value circa 2008 tended to prefer higher values of  $\sin 2\beta$ , as fig. 7 clearly illustrates.

As a matter of fact, by comparing the latest  $\epsilon_K$  and  $\sin 2\beta$  predictions from global CKM fits that assume the SM [32, 70], one can easily check that the latter agree with the corresponding experimental averages [26] at the level of 1 standard deviation.

It will be interesting to monitor the status of the  $\epsilon_K - \sin 2\beta$  tension after a mature measurement of the CKM angle  $\gamma$  and after further progress on the knowledge of the CKM matrix element  $|V_{cb}|$ . An accurate  $\gamma$  measurement is actually planned by LHCb, with projected accuracy of about  $1^\circ$  [77]. Albeit more difficult to predict, a substantial improvement on  $|V_{cb}|$  is likewise to be expected, taking also into account the forthcoming startup of the Super- $B$  factory Belle II. On both topics we will return in chapter 4.



**Figure 7:**  $\text{BR}(B \rightarrow \tau\nu)$  vs.  $\sin 2\beta$  as of summer 2008 (taken from the CKMfitter website [70]). The black cross displays the experimental averages, whereas the color-coded area represents the region preferred by the global CKM fit, with CL according to the color code on the right of the plot.

### 3 The rare decay $B_s \rightarrow \mu^+ \mu^-$

The decay of a  $B_s$  meson into two oppositely charged muons provides one of the best available probes of the mechanism of quark-flavor mixing discussed at length in the introductory part. As a matter of fact, this transition enjoys two crucial features that make it appealing on the theoretical and on the experimental sides alike: it is at the same time very rare (within the SM) and very clean. The extremely rare occurrence of this decay within the SM is the result of a double suppression mechanism at work: as we will detail in the next section, this transition is mediated by an FCNC and is also helicity suppressed. Conversely, the cleanliness of this decay, both theoretical and experimental, is consequence of the purely leptonic final state. The latter implies, on the theoretical side, that the hadronic matrix element is especially simple, as it involves only the initial-state pseudoscalar meson. Furthermore, on the experimental side, a final state consisting of two oppositely-charged leptons, in particular muons, with the invariant mass of the initial-state meson, is arguably among the simplest triggers one may hope for, let alone issues of combinatorics, that become important given the extremely rare nature of the signal.

By virtue of the above considerations, on which we will return in all detail in the sections to come, the  $B_s \rightarrow \mu\mu$  decay represents one of the milestones of the flavor program at the LHC.<sup>32</sup>

#### 3.1 $B_s \rightarrow \mu\mu$ within the SM: decay structure

Within the SM, the branching ratio of the non-radiative<sup>33</sup> decay  $B_s \rightarrow \mu\mu$ , to be henceforth indicated with  $\mathcal{B}_{s,\text{SM}}^{(0)}$ , obeys the following formula (see e.g. [4])

$$\mathcal{B}_{s,\text{SM}}^{(0)} = \tau_{B_s} \underbrace{\frac{G_F^2}{\pi} \left[ \frac{\alpha_{\text{em}}(M_Z)}{4\pi \sin^2 \theta_W} \right]^2 |V_{tb}^* V_{ts}|^2}_{\text{couplings: gauge and CKM}} \cdot \underbrace{f_{B_s}^2}_{\text{had. matrix elem.}} \cdot \underbrace{m_\mu^2}_{\text{helicity suppr.}} \cdot \underbrace{m_{B_s} \sqrt{1 - \frac{4m_\mu^2}{m_{B_s}^2}}}_{\text{phase space}} \cdot \underbrace{Y^2(x_{tW}, x_{ht}; \alpha_s)}_{\text{short-distance function}} \quad (3.1)$$

where the underbraces highlight the role or origin of the respective factors. Let us comment on the most important of them in turn.

First, the  $Y$  function is an appropriate loop function, consisting of  $Z$ -penguin and box-diagram contributions of the kind depicted in fig. 8. The  $Y$  function in eq. (3.1) includes QCD corrections as well as the leading EW corrections, and this explains its dependence on  $x_{ht}$  as well as on  $\alpha_s$ . The  $x$  arguments are defined as follows

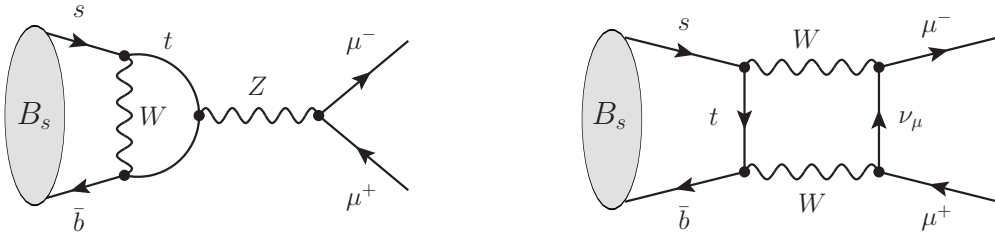
$$x_{tW} = \frac{m_t^2(\mu)}{M_W^2}, \quad x_{ht} = \frac{M_h^2}{m_t^2(\mu)}, \quad (3.2)$$

and the renormalization scheme and scale used for the top mass will be specified in due course. In the absence of QCD and EW corrections, the  $Y$  function reduces to the Inami-

<sup>32</sup> Most of the above arguments hold basically unchanged for the  $B_d \rightarrow \mu^+ \mu^-$  decay, which, albeit more challenging experimentally, is another major target of LHC searches at run II. LHC experiments are likewise planning to improve bounds on  $B_{d,s}$  decays to two taus, albeit the latter are very difficult in a hadronic environment. Finally,  $B_{d,s}$  decays to two electrons are exceedingly rare in the SM, and will not be given further consideration in this manuscript.

<sup>33</sup> The reason for this qualification will be clearer in sec. 3.4.





**Figure 8:** Example of penguin and box diagrams for the  $B_s \rightarrow \mu\mu$  decay.

Lim function [58]

$$Y_0(x) = \frac{x}{8} \left( \frac{4-x}{1-x} + \frac{3x}{(1-x)^2} \ln x \right). \quad (3.3)$$

We will come back to this function in sec. 3.3, in the discussion of the systematic error coming from the renormalization-scheme dependence.

We note incidentally that fig. 8 explains the couplings structure in eq. (3.1). The CKM factors in particular are obvious because of the two  $\bar{q}Wq'$  interactions in either diagrams. Concerning the  $G_F^2 \alpha_{\text{em}}^2$  factor, we note that the presence of  $\alpha_{\text{em}}$  is fictitious: it can be eliminated by expressing  $\alpha_{\text{em}}/\sin^2 \theta_W$  in terms of  $G_F$  and  $M_W$ . The formula in eq. (3.1) is thereby well defined in the limit  $\alpha_{\text{em}} \rightarrow 0$ .

The  $f_{B_s}^2$  factor accounts for the hadronic matrix element, whose simplicity is among the main reasons for the theoretical cleanliness of this decay. Recalling that the final state is purely leptonic, the only non-null matrix element is in fact

$$\langle 0 | \bar{b} \gamma^\alpha \gamma_5 s | B_s(p) \rangle = -i f_{B_s} p^\alpha, \quad (3.4)$$

namely the matrix element ‘defining’ the pseudo-Goldstone nature of the  $B_s$  meson. The only quantity to be determined non-perturbatively in eq. (3.4) is  $f_{B_s}$ , the  $B_s$ -meson decay constant. This quantity is among the simplest ones to estimate within lattice QCD, for various reasons, including the simplicity of the matrix element itself, and the fact that the bottom quark is heavy enough that it can be reliably described within an effective theory in powers of  $\Lambda_{\text{QCD}}/m_b$ . This makes precision LQCD computations of  $f_{B_s}$  possible and, to a large extent, already reality. We will return to this point in the discussion of the prediction for the quantity in eq. (3.1) and its associated error.

With the matrix element in eq. (3.4) at hand, it is in turn straightforward to understand the appearance of the helicity-suppression factor,  $m_\mu^2$ . In fact, the  $B_s$  four-momentum  $p^\alpha$  in eq. (3.4) needs to be contracted with the weak lepton current  $\bar{u}(p_{\mu^-}) \gamma_\alpha P_L v(p_{\mu^+})$ , with  $\bar{u}$  and  $v$  denoting the spinors of the final-state  $\mu^-$  and  $\mu^+$ , respectively, and  $p_{\mu^\mp}$  their momenta. Using  $p = p_{\mu^-} + p_{\mu^+}$  and the lepton equations of motion, one immediately finds

$$p^\alpha \bar{u}(p_{\mu^-}) \gamma_\alpha P_L v(p_{\mu^+}) = -m_\mu \bar{u}(p_{\mu^-}) \gamma_5 v(p_{\mu^+}) \quad (3.5)$$

which explains the  $m_\mu^2$  factor in the branching ratio. From a naive, dimensional argument, one would expect a mass suppression of the order of  $(m_{B_s}^2/M_W^2)^2$  at the branching-ratio level. Hence the  $m_\mu^2$  factor amounts to a further suppression of order  $m_\mu^2/m_{B_s}^2 \approx 4 \times 10^{-4}$  relative to the ‘usual’ FCNC-related suppression. This should provide an intuitive understanding of why the decay  $B_s \rightarrow \mu\mu$  occurs as rarely as a few times in a billion of  $B_s$  decays.

Besides, the helicity-suppression factor provides also a strong argument in support of the  $B_s \rightarrow \mu\mu$  new-physics sensitivity. In fact, this factor can be lifted, notably in models where

new extra fermion degrees of freedom appear in the short-distance function, so that their mass can replace the muon's.

The above discussion summarizes the typical reasons for the enormous interest of accurately measuring (and predicting) the  $B_s \rightarrow \mu\mu$  decay rate. Further arguments, less often encountered but probably even stronger, will be presented in secs. 3.8 and 3.9. Before this part, we will review in secs. 3.2 to 3.7 the status of the SM prediction, paying special attention to its associated sources of systematics.

### 3.2 Experimental status, and steps towards a precise prediction

As already mentioned, accurately measuring the decay  $B_s \rightarrow \mu\mu$  (as well as the corresponding one with a  $B_d$  as initial state) is one of the highest flavor-physics priorities at the LHC. Let us shortly review the experimental status for these observables. The first evidence for the  $B_s \rightarrow \mu\mu$  decay was announced by LHCb at the end of 2012 [78].<sup>34</sup> In 2013 new measurements were performed by both CMS [80] and LHCb [81], reporting results remarkably similar in accuracy, and both consistent with the SM expectation. The CMS analysis was based on 5 and 20 fb<sup>-1</sup> of data collected at 7 and respectively 8 TeV  $pp$  collisions and resulted in [80]

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{CMS}} = \left(3.0_{-0.9}^{+1.0}\right) \times 10^{-9}, \quad \mathcal{B}(B_d \rightarrow \mu\mu)_{\text{CMS}} = \left(3.5_{-1.8}^{+2.1}\right) \times 10^{-10}, \quad (3.6)$$

corresponding to signal significances of 4.3 and 2.0 $\sigma$  (standard deviations) respectively.

The LHCb analysis used 1 and 2 fb<sup>-1</sup> of integrated luminosity collected at 7 and respectively 8 TeV, with the following results

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{LHCb}} = \left(2.9_{-1.0}^{+1.1}\right) \times 10^{-9}, \quad \mathcal{B}(B_d \rightarrow \mu\mu)_{\text{LHCb}} = \left(3.7_{-2.1}^{+2.4}\right) \times 10^{-10}, \quad (3.7)$$

whose signal significances are 4.0 and 2.0 $\sigma$  respectively.

A preliminary combination of the above results was presented in ref. [82]. This combination is preliminary, in that the combined uncertainty is calculated within a simplified statistical approach (see paper text for details), that circumvents the explicit combination of the likelihood functions of the two measurements – approach that on the other hand would be required, especially in view of the asymmetric uncertainties of either measurement. The resulting averages read [82]

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{CMS+LHCb}} = (2.9 \pm 0.7) \times 10^{-9}, \quad \mathcal{B}(B_d \rightarrow \mu\mu)_{\text{CMS+LHCb}} = (3.6_{-1.4}^{+1.6}) \times 10^{-10}. \quad (3.8)$$

Ref. [82] does not provide (for the same reason as mentioned above) a combined significance figure, but it is clear from eq. (3.8) that the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  average is above 5 $\sigma$ , whereas  $\mathcal{B}(B_d \rightarrow \mu\mu)$  is below 3 $\sigma$ , i.e. not yet statistically significant. At the time of this writing, no further updates with respect to the results in eq. (3.8) have been announced, hence we will take eq. (3.8) as our reference experimental values. It should be stated that they are in clear consistency with the SM predictions, that at the time of the CMS-LHCb combination read [4]

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{SM}} = (3.56 \pm 0.30) \times 10^{-9}, \quad \mathcal{B}(B_d \rightarrow \mu\mu)_{\text{SM}} = (1.07 \pm 0.10) \times 10^{-10}. \quad (3.9)$$

The predictions in eq. (3.9) are the end result of a number of theoretical considerations. In particular, in order to obtain a precise prediction for the  $B_s \rightarrow \mu\mu$  rate within the SM, it

<sup>34</sup> Before ref. [78], the first one to report a signal, numerous experiments had obtained more and more stringent upper bounds, the most recent being [79].

is necessary, first, to compute the corresponding electroweak amplitude with high accuracy, and second, to assess the correspondence between the initial and final state detected by experiment, and those used in the theoretical prediction. More precisely, the following steps towards an accurate comparison between data and theory can be identified [4]

- *A reduction of the renormalization-scheme dependence of the non-radiative branching ratio.* The short-distance  $Y$  function in the branching-ratio formula of eq. (3.1) includes NLO QCD corrections [83] as well as two-loop EW corrections evaluated in the large- $m_t$  limit [84]. Hence one would conclude that the theory branching ratio can be computed with great accuracy, up to the parametric uncertainties due to the  $B_s$ -meson decay constant  $f_{B_s}$ , the CKM coupling  $|V_{tb}^* V_{ts}|$ , the  $B_s$ -meson lifetime  $\tau_{B_s}$  and the top-quark mass. (We will return to the parametric error in sec. 3.6.) In fact, as pointed out in [4], even in presence of these corrections the theory branching ratio still suffers from a sizable dependence on the renormalization scheme for EW parameters, in particular  $\sin^2 \theta_W$ . This problem will be discussed in detail in sec. 3.3.
- *The treatment of the soft-photon radiation.* Photon emission inevitably occurs in the full theory including  $\alpha_{\text{em}} \neq 0$ . The simplest infrared-safe observable is

$$\mathcal{B}(E_{\text{max}}) \equiv \mathcal{B}(B_s \rightarrow \mu^+ \mu^- + n\gamma) \Big|_{\sum E_\gamma \leq E_{\text{max}}} \quad (3.10)$$

namely the branching ratio including an arbitrary number of undetected photons with total energy in the meson rest frame less or equal to  $E_{\text{max}}$ . This topic will be discussed in detail in sec. 3.4.

- *Connecting the  $B_s$  state in the theory calculation with the one measured by experiment.* The simplest observable accessible at hadron colliders is the flavor-averaged time-integrated distribution. To this quantity refer all the experimental averages reported above, and also the theory predictions in eq. (3.9). In fact, on the one side experimental data are summed over the decays of the  $B_s$  and of the  $\bar{B}_s$ , because there is no flavor tagging (yet). In addition, data are averaged over all the  $B_s$  ( $\bar{B}_s$ )-meson decay times, i.e. time information is not retained either for the moment being. The time-averaged branching ratio is related to the instantaneous one, calculated by theory, through a correction factor due to the large width difference in the  $B_s$  system, as pointed out in [85,86]. On the other hand this correction is irrelevant in the  $B_d$  system, because the width difference is smallish in this case. This issue will be discussed in detail in sec. 3.5.

We next proceed with the discussion of the above points in dedicated sections.

### 3.3 Renormalization-scheme dependence

Let us go back to the non-radiative branching ratio in eq. (3.1) and the mass-ratio arguments in eq. (3.2). Leaving aside parametric uncertainties, that will be discussed in sec. 3.6, two evident uncertainties are present in  $\mathcal{B}(B_s \rightarrow \mu\mu)$  if the  $Y$  function is approximated with its leading-order expression  $Y_0$ :

- The choice of the scale  $\mu$ , that can roughly span the range  $[M_W, 2m_t]$ . This unphysical scale dependence had been already reduced to less than 2% by the inclusion of NLO QCD corrections [83]. Meanwhile, even the NNLO QCD corrections to  $\mathcal{B}(B_s \rightarrow \mu\mu)$  have become available [87], bringing the residual scale dependence way below 1%, i.e. making it totally negligible.

- The choice of renormalization scheme for the EW parameters, in particular  $\sin^2 \theta_W$  and to a lesser extent  $m_t$ . In eq. (3.1) this arbitrariness amounts to a sizable error.

In order to quantify the uncertainty associated with the second item, let us evaluate eq. (3.1) with two definitions of  $\sin^2 \theta_W$ , respectively the  $\overline{\text{MS}}$  scheme and the on-shell scheme. In these two schemes  $\sin^2 \theta_W$  is very precisely known and reads [26]

$$\sin^2 \hat{\theta}_W(M_Z) = 0.23116(3) , \quad \left[ \sin^2 \theta_W \right]^{\text{OS}} \equiv 1 - M_W^2/M_Z^2 = 0.22290 . \quad (3.11)$$

We note that the second choice, with all other parameters fixed, leads to a  $\mathcal{B}(B_s \rightarrow \mu\mu)$  prediction higher by 7% than the first choice. This corresponds to a shift in the rate as large as  $0.2 \times 10^{-9}$ , equivalent to a shift in  $f_{B_s}$  (the main source of parametric error in  $\mathcal{B}(B_s \rightarrow \mu\mu)$ ) by 8 MeV. It should be noted that such a shift is larger by a factor of about two with respect to the error of the most accurate  $f_{B_s}$  determination. In short, the systematic uncertainty associated with the scheme choice of  $\sin^2 \theta_W$  is comparable with the uncertainty coming from the main source of parametric error, and needs to be removed.

### Scheme proposal in ref. [4]

The first step towards controlling this error component was proposed in ref. [4], and it will be described in detail next. This proposal prompted the authors of ref. [88] to perform a complete calculation of the EW corrections to eq. (3.1). This calculation, that supersedes the proposal of ref. [4], will be described at the end of sec. 3.6.

Ref. [4] started from the consideration of four renormalization schemes, namely two schemes for each of the parameters mostly affected by EW corrections,  $\sin^2 \theta_W$  and the top mass. These choices are:

- For  $\sin^2 \theta_W$ :

$$\sin^2 \hat{\theta}_W(M_Z) : (r_s = 0) , \quad \text{or} \quad \left[ \sin^2 \theta_W \right]^{\text{OS}} : (r_s = 1) , \quad (3.12)$$

and are labelled by the parameter  $r_s$ ;

- For the top mass:

$$m_t \equiv m_t(m_t)^{\overline{\text{MS}},\text{QCD}} : (r_t = 0) , \quad \text{or} \quad \bar{m}_t \equiv m_t(m_t)^{\overline{\text{MS}},\text{QCD}+\text{EW}} : (r_t = 1) , \quad (3.13)$$

and are labelled by the parameter  $r_t$ . These two top-mass schemes are related via [84]

$$\bar{m}_t^2 = m_t^2 (1 + \xi_t \Delta_t(\mu, x_{ht})) , \quad (3.14)$$

with

$$\xi_t = \frac{G_F m_t^2}{8\sqrt{2} \pi^2} . \quad (3.15)$$

Concerning the definitions (3.13), we note that, in the case of  $m_t$  only QCD corrections are  $\overline{\text{MS}}$ -renormalized, whereas the mass is on-shell as far as EW corrections are concerned. Conversely, in the case of  $\bar{m}_t$ , both QCD and EW corrections are  $\overline{\text{MS}}$ -renormalized. The QCD  $\overline{\text{MS}}$  top-quark mass is determined from the pole mass in table 1 using RunDec [89].<sup>35</sup> The explicit expression for  $\Delta_t(\mu, x_{ht})$  can be found in ref. [84] and has been calculated in [90].

<sup>35</sup> For the central value of  $M_t$  in table 1, ref. [4] obtains  $m_t(m_t)^{\overline{\text{MS}},\text{QCD}} = 163.2$  GeV and  $m_t(m_t)^{\overline{\text{MS}},\text{QCD}+\text{EW}} = 164.5$  GeV.

$G_F = 1.16638 \times 10^{-5} \text{ GeV}^{-2}$	$m_{B_s} = 5.36677 \text{ GeV}$
$\alpha_{\text{em}}^{-1}(M_Z) = 127.937$ [91]	$f_{B_s} = 227(8) \text{ MeV}$ [see text]
$\alpha_s(M_Z) = 0.1184(7)$ [92]	$\tau_{B_s} = 1.466(31) \text{ ps}$
$M_W = 80.385 \text{ GeV}$	$ V_{tb}^* V_{ts}  = 0.0405(8)$ [70, 93]
$M_Z = 91.1876 \text{ GeV}$	$m_{B_d} = 5.27958 \text{ GeV}$
$M_t = 173.2(0.9) \text{ GeV}$ [94, 95]	$f_{B_d} = 190(8) \text{ MeV}$ [see text]
$M_h = 125 \text{ GeV}$ [10]	$\tau_{B_d} = 1.519(7) \text{ ps}$
$m_\mu = 105.6584 \text{ MeV}$	$ V_{tb}^* V_{td}  = 0.0087(2)$ [70, 93]

**Table 1:** Input parameters used in ref. [4] for the determination of  $\mathcal{B}_{s,\text{SM}}^{(0)}$  and  $\mathcal{B}_{d,\text{SM}}^{(0)}$ . Quantities without an explicit reference are taken from ref. [26]. We do not show the errors for quantities whose uncertainty has a negligible impact on the branching-ratio determinations. The central value of  $f_{B_{s,d}}$  corresponds to the central value of the lattice averages presented in ref. [96], while the error is ref. [4]’s estimate of the present uncertainty (see text for details).

In the light of definitions (3.13), we can now define more sharply also the  $x$  parameters in eq. (3.2). They are

$$x_{tW} = \frac{m_t^2}{M_W^2}, \quad \bar{x}_{tW} = \frac{\bar{m}_t^2}{M_W^2}, \quad x_{ht} = \frac{M_h^2}{m_t^2}. \quad (3.16)$$

All the relevant parametric input to these formulae, and to eq. (3.1) itself, is collected in table 1.

Each of the four renormalization schemes of eqs. (3.12) and (3.13) is characterized by the pair  $(r_s, r_t)$ . Once this pair is fixed, we know uniquely which of the parameters listed above is to be employed in the calculation of  $Y$  in (3.1) and which value of  $\sin^2 \theta_W$  is to be used in the prefactor in this equation. Therefore in presenting a general formula for the function  $Y$  valid in all these renormalization schemes in the large  $m_t$ -limit, we can trade the mass variables for the pair  $(r_s, r_t)$ .

With this notation the loop function  $Y$ , including complete NLO QCD corrections [83] and two-loop electroweak corrections in the large- $m_t$  limit [84] is given in the  $(r_s, r_t)$  scheme as follows:

$$Y(r_s, r_t; \alpha_s) = Y_{\text{eff}}(r_s, r_t) + \frac{\alpha_s(\mu)}{4\pi} Y_1(x_{tW}), \quad (3.17)$$

where

$$Y_{\text{eff}}(r_s, r_t) = Y_0(x_0(r_t)) + \xi_t \frac{x_{tW}}{8} \left( \tau_b^{(2)}(x_{ht}) + 3 - 3r_s \frac{\cos^2 \theta_W}{\sin^2 \theta_W} - r_t \Delta_t(\mu, x_{ht}) \right) \quad (3.18)$$

with

$$x_0(r_t) = x_{tW} + r_t(\bar{x}_{tW} - x_{tW}). \quad (3.19)$$

The  $Y_{\text{eff}}$  function in eq. (3.18) is the effective Inami-Lim function for the  $(r_s, r_t)$  scheme. This expression generalizes the formulae in [84] that applied only to specific schemes. The explicit expression for  $\tau_b^{(2)}(x_{ht})$  can be found in ref. [84] and has been calculated in [97]. Finally, the function  $Y_1$ , encoding the NLO QCD corrections, can be found in the last reference in [83]. Note that  $Y_1$  is always evaluated in the  $\overline{\text{MS}}$ -QCD scheme, that is using  $x_{tW}$ , whereas  $Y_0$  is evaluated using  $\bar{x}_{tW}$  or  $x_{tW}$  depending on the presence or not of the  $-\Delta_t$  term in  $Y_{\text{eff}}$ .

In the case of complete NLO electroweak corrections, the  $r_s$ -dependence in eq. (3.18) would cancel, up to NNLO effects, the one of  $\sin^2 \theta_W$  in the prefactor in eq. (3.1). The corresponding  $r_t$  dependence in the correction term in (3.17) would in turn cancel the one present in the leading term  $Y_0$ . As evident from our formulae, where NLO electroweak

	$(r_s, r_t)$	$\mathcal{B}_{s,\text{SM}}^{(0)} [\times 10^{-9}]$
$\sin^2 \theta_W \overline{\text{MS}}, m_t \text{ OS}$	(0, 0)	3.28
$\sin^2 \theta_W \overline{\text{MS}}, m_t \overline{\text{MS}}$	(0, 1)	3.31
$\sin^2 \theta_W \text{ OS}, m_t \text{ OS}$	(1, 0)	3.42
$\sin^2 \theta_W \text{ OS}, m_t \overline{\text{MS}}$	(1, 1)	3.45

**Table 2:** Dependence of the  $\mathcal{B}_{s,\text{SM}}^{(0)}$  prediction upon the choice of the renormalization scheme  $(r_s, r_t)$  for electroweak corrections as defined in the text.

corrections are only in the large- $m_t$  limit, this cancellation is only partial, implying left-over scheme uncertainties.

Using the central input values in table 1, and taking the  $\sin^2 \theta_W$  and  $m_t$  choices in eqs. (3.11)-(3.12) and (3.13) respectively, ref. [4] obtains the central values for  $\mathcal{B}_{s,\text{SM}}^{(0)}$  in the four resulting renormalization schemes, that are collected in table 2. The central value in either of the cases has been obtained setting the QCD renormalization scale to  $\mu = m_t(m_t)^{\overline{\text{MS}},\text{QCD}}$ . This table confirms that the largest uncertainty related to the scheme choice is the one due to  $\sin^2 \theta_W$ , as already mentioned. On the other hand, the left-over uncertainty due to the scheme choice for the top-quark mass is at the level of 0.9%.

In short, even after inclusion of the NLO EW contributions in the large- $m_t$  limit [84], the left-over uncertainty is disturbing, and can only be removed by the inclusion of *full* NLO EW corrections. Here we would like to describe a more economical strategy to reduce this scheme uncertainty, as conjectured in [4]. It should be stated explicitly that whether this strategy indeed reduces the scheme uncertainty or not can only be controlled by the explicit calculation of NLO EW corrections. As alluded to, the latter have meanwhile been computed by the authors of [88], as discussed at the end of sec. 3.6.

Without a complete calculation of NLO EW effects, only a very coarse estimate of the scheme dependence can be made. We should emphasize that in all recent papers on  $B_s \rightarrow \mu^+ \mu^-$  before [4] and most earlier papers this uncertainty had been omitted. This can be justified by the fact that most authors expected non-SM effects to modify the relevant branching ratio by a large amount, rendering any shift below 10% in the SM estimate irrelevant. With the upper bounds on  $\mathcal{B}(B_s \rightarrow \mu\mu)$  becoming more and more stringent – let alone the measurement itself – the situation changed dramatically and uncertainties of this size had to be taken into account.

At the time of ref. [4] the question arose, *which value for  $\mathcal{B}_{s,\text{SM}}^{(0)}$  should be quoted in the absence of complete NLO EW corrections*. Ref. [4] proposed a ‘temporary’ solution to be described next. As already pointed out in [84] a similar scheme-dependence problem as the one described above around table 2 is present in  $K \rightarrow \pi\nu\bar{\nu}$  decays, which are theoretically even cleaner than  $B_s \rightarrow \mu^+ \mu^-$ . These decays are governed by the Inami-Lim function  $X_0(x_t)$ , which differs from  $Y_0(x_t)$  only by box contributions. Therefore, at large  $m_t$ , where only the  $Z$ -penguin is relevant (the first diagram in fig. 8), the effective EW corrections to Inami-Lim functions are identical to the ones presented above.

This observation is relevant because complete NLO EW corrections to  $K \rightarrow \pi\nu\bar{\nu}$  had been calculated [98] two years before [4]. These authors considered three renormalization schemes:

- The  $\overline{\text{MS}}$  scheme for all parameters. In our terminology – recall eqs. (3.12)-(3.13) – this is the (0,1) scheme.
- The  $\overline{\text{MS}}$  scheme for all couplings and the on-shell scheme for all masses. This is the

**Scheme dep. in short**

**Scheme proposal**

(0,0) scheme.

- The on-shell scheme for weak mixing angle and all masses and the QED coupling constant renormalized in the  $\overline{\text{MS}}$  scheme. This is the (1,0) scheme.

By calculating complete NLO electroweak corrections in these three schemes, they reduced the scheme dependence at the level of the branching ratio far below 1%. Looking at the size of different corrections they concluded that the on-shell definition of masses, together with the  $\overline{\text{MS}}$  definition of  $\sin^2 \theta_W$ , our (0,0) scheme, is the best choice of renormalization scheme – it is namely the scheme where NLO corrections are smallest in absolute value. Incidentally, we find that, in our  $B_s \rightarrow \mu\mu$  case, this scheme is also the one that exhibits the smallest dependence, below 1%, upon the choice of the renormalization scale in the range  $[M_Z, m_t]$ .

By inspection of their analysis for the mentioned scheme, in particular of equations (4.2)–(4.4) of their paper, a very simple prescription for the final result for  $K \rightarrow \pi\nu\bar{\nu}$  branching ratios (including complete NLO QCD and complete NLO EW corrections) emerges. Adapted to the  $B_s \rightarrow \mu^+\mu^-$  decay, this prescription is as follows [4]:

- Use eq. (3.1) for  $\mathcal{B}_{s,\text{SM}}^{(0)}$  with

$$\sin^2 \theta_W = \sin^2 \hat{\theta}_W(M_Z) = 0.23116(3). \quad (3.20)$$

- Set

$$Y(x_{tW}, x_{ht}; \alpha_s) = Y_0(x_{tW}) + \frac{\alpha_s(\mu)}{4\pi} Y_1(x_{tW}) \equiv \eta_Y Y_0(x_{tW}), \quad \eta_Y = 1.0113, \quad (3.21)$$

where  $x_{tW}$  is defined by eqs. (3.13) and (3.16). Our value of  $\eta_Y$  agrees well with 1.012 quoted in the last reference in [83].

The complete electroweak corrections to  $B_s \rightarrow \mu^+\mu^-$  will be different in the details, due to different box diagrams and the presence of charged leptons in the final state in place of neutrinos. Yet it is plausible to expect that the prescription given above could work in our case as well. It should be stressed that the end reason for this is the large top mass, implying a completely negligible contribution from box topologies as compared to  $Z$  penguins.

A further argument for the plausibility of the scheme proposal in ref. [4] is as follows.<sup>36</sup> It is known that, for any perturbative calculation, both in QCD and electroweak interactions, a particular definition of fundamental parameters and renormalization scale in the leading term allows to minimize NLO corrections. Specifically, one-loop corrections to  $\sin^2 \theta_W$  in general involve counterterms proportional to the two-point functions of the  $W$  and the  $Z$ , that are numerically large because of their dependence on the large top (and Higgs) mass [99]. It has been pointed out that these terms are absent if  $\sin^2 \theta_W$  is renormalized in the  $\overline{\text{MS}}$  scheme [100]. A similar argument holds for two-point Green's functions, related to the computation of renormalized masses such as  $m_t$ . In this case, the scheme that allows to absorb the numerically largest counterterms is the on-shell one. This argument supports  $r_s = 0$  and  $r_t = 0$  in eqs. (3.12) and (3.13), i.e. the scheme proposed by ref. [4].

In spite of these considerations, and as already stated, the scheme prescription described here can only be validated by a full-fledged NLO calculation of EW corrections, that has meanwhile been performed in ref. [88]. This calculation, along with the updated averages for  $\mathcal{B}_{s,\text{SM}}^{(0)}$  and  $\mathcal{B}_{d,\text{SM}}^{(0)}$ , will be described in sec. 3.6.

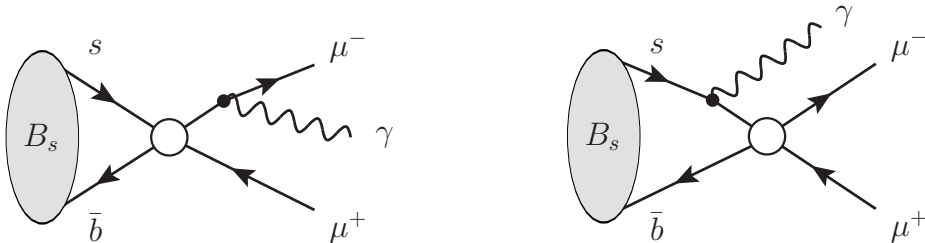
\*\*\*

<sup>36</sup> I am grateful to Paolo Gambino for an illuminating discussion in this respect.

The above discussion arguably closes the chapter of improvements needed on the short-distance function of the non-radiative branching ratio (3.1). The next step is to connect this theoretical quantity to the quantity measured by experiment, namely to relate the initial and final states detected by experiment to those used in the theory calculation. Two effects turn out to be important: on the one side, the fact that the final-state muons may actually radiate soft, undetected photons, effect to be discussed in sec. 3.4; furthermore, the fact that the initial-state  $B_s$  used in the theory branching ratio is not a mass eigenstate and oscillates over time into its antiparticle. If time information is not retrieved in the experimental observable, this effect amounts to a non-negligible correction [85,86], because of the large width difference (in units of the average width difference) between the mass eigenstates. This effect will be discussed in sec. 3.5.

### 3.4 Soft-photon corrections

As already remarked, switching on electromagnetic interactions the  $B_s \rightarrow \mu^+ \mu^-$  transition is unavoidably accompanied by real photon emission. On general grounds we can distinguish two types of radiation: bremsstrahlung and ‘direct emission’. The latter is an infelicitous shortcut for ‘emission from the  $B$ -meson valence quarks’. This component vanishes in the limit of small photon energies. It is clear in fact that this component will be relevant only for photon energies large enough to resolve the internal quark structure of the  $B$  meson, and for these energies it actually represents a background for the extraction of short-distance information on the  $B_s \rightarrow \mu^+ \mu^-$  amplitude. In the limit of small photon energies the former, bremsstrahlung component is largely dominant. Therefore, a tight cut on the  $\mu^+ \mu^-$  invariant mass ( $m_{\mu^+ \mu^-}$ ), close to  $m_{B_s}$ , allows us to treat radiative corrections in the soft-photon approximation and to suppress the background due to the direct-emission component. Two representative diagrams for the bremsstrahlung and the direct-emission amplitudes are depicted in fig. 9. (The complete set of diagrams can be found in [101].)



**Figure 9:** Example of bremsstrahlung (left) and respectively direct-emission (right) diagrams for the  $B_s \rightarrow \mu\mu + \gamma$  decay. The large empty circle denotes the ultraviolet part of the diagram, that below the weak scale is described by a suitable effective operator.

The soft-photon approximation is defined by the kinematic condition  $E_{\max} \ll m_{B_s}/2$ , with  $E_{\max}$  already introduced in eq. (3.10), and denoting the total energy of the photons emitted along with the final-state  $\mu^+ \mu^-$ . It turns out that, in the soft-photon approximation, and hence restricting to bremsstrahlung emission only, the correction can be summed to all orders, namely to an arbitrary number of *real* emitted photons, and an arbitrary number of *virtual* soft-photon insertions. The overall correction amounts to a multiplicative factor, denoted as  $\omega(E_{\max})$ , with respect to the non-radiative rate [4, 102–104]. The corrected branching ratio reads

$$\mathcal{B}^{\text{phys}}(E_{\max}) \equiv \mathcal{B}(B_s \rightarrow \mu^+ \mu^- + n\gamma) \Big|_{\sum E_\gamma \leq E_{\max}} = \omega(E_{\max}) \times \mathcal{B}^{(0)}, \quad (3.22)$$



where  $\mathcal{B}^{(0)}$  denotes the non-radiative branching ratio, whose SM expression has been given in eq. (3.1).

In the soft-photon approximation, the correction factor defined in eq. (3.22) can be expressed as [4, 102–104]

$$\omega(E_{\max}) = \omega_{\text{IB}}(E_{\max}) \times \left[ 1 + \mathcal{O}\left(\frac{\alpha_{\text{em}}}{\pi}\right) \right], \quad (3.23)$$

$$\omega_{\text{IB}}(E_{\max}) = \left( \frac{E_{\max}}{m_{B_s}/2} \right)^{\frac{2\alpha_{\text{em}} b}{\pi}}, \quad (3.24)$$

where  $\alpha_{\text{em}} = 1/137.036$  is the fine-structure constant and

$$b \equiv - \left[ 1 - \frac{1}{2\beta_{\mu\mu}} \ln \left( \frac{1 + \beta_{\mu\mu}}{1 - \beta_{\mu\mu}} \right) \right], \quad \beta_{\mu\mu} = \left[ 1 - \frac{4m_\mu^4}{(m_{\mu^+\mu^-}^2 - 2m_\mu^2)^2} \right]^{1/2}. \quad (3.25)$$

The term  $\omega_{\text{IB}}(E_{\max})$  takes into account the emission of an arbitrary number of real photons, with maximal energy in the meson rest frame less or equal to

$$E_{\max} = \frac{m_{B_s}^2 - m_{\mu^+\mu^-}^2}{2m_{B_s}}, \quad (3.26)$$

together with the corresponding virtual corrections, as mentioned. For  $E_{\max} \approx 60$  MeV, the correction amounts to

$$\omega_{\text{IB}}(60 \text{ MeV}) \approx 0.89, \quad (3.27)$$

namely to a  $\approx 11\%$  suppression of the non-radiative rate.

The correction in eq. (3.27) is numerically large. This fact can be explained intuitively by observing that the two mass scales in the ratio on the r.h.s. of eq. (3.24) are unrelated, and in general vastly different. In turn, to understand the meaning of these two mass scales, it is useful to shortly review the  $\omega_{\text{IB}}$  calculation (a detailed, lucid account can be found in [103]). The latter is composed of real-photon and of virtual-photon contributions, both of which are separately infrared-divergent, so that one has to introduce an infrared regulator  $\lambda$ . Besides, there are two physical ultraviolet cutoffs, one for each half (real- vs. virtual-photon insertions) of the calculation. In the virtual-photon calculation there is a natural ultraviolet cutoff  $\Lambda_V$ , represented by the kinematic limit of the energy of a virtual photon attached to the non-radiative amplitude. This cutoff is of order  $m_{B_s}$ . Thus, for any finite number of virtual photons attached to the non-radiative amplitude, the correction is of the form  $-A \log(\Lambda_V/\lambda)$ , with  $A$  a *positive* coefficient [103]. When summing over an arbitrary number of virtual photons, the correction exponentiates, yielding the multiplicative correction factor  $(\lambda/\Lambda_V)^A$ . Note therefore that, in the absence of real-emission corrections, and in the physical limit  $\lambda \rightarrow 0$ , the non-radiative amplitude including the virtual corrections would vanish, by virtue of  $A$  being positive. The real-emission corrections have likewise a natural ultraviolet cutoff, represented by the scale  $E_{\max}$ , the total energy of the undetected real soft photons emitted in the final state – see eq. (3.10). For any finite number of real emitted photons, corrections are again logs of the form  $\log(E_{\max}/\lambda)$ . Summing over an arbitrary number of emitted photons, the correction again exponentiates to  $(E_{\max}/\lambda)^A$ , where the exponent turns out to be the *same* as in the virtual-photon calculation.<sup>37</sup> The product between the virtual- and the real-photon calculation yields therefore  $(\lambda/\Lambda_V)^A \times (E_{\max}/\lambda)^A$ , and the  $\lambda$

<sup>37</sup> The reason for the virtual- and real-photon exponents being identical is, ultimately, the fact that the basic ingredient of both calculations – the fermion-fermion-photon interaction vertex – is one and the same.

dependence cancels out between them, as it should. The total correction therefore assumes the form in eq. (3.24), where we have assumed  $\Lambda_V = m_{B_s}/2$  (we will comment on this assumption below). It is interesting to note that the smaller  $E_{\max}$ , i.e. the tighter the requirement that the  $\mu\mu$  invariant mass be close to  $m_{B_s}$ , the more important the soft-photon correction becomes; conversely, in the limit where  $E_{\max}$  approaches its kinematic limit  $m_{B_s}$ , i.e. in the limit where the  $B_s \rightarrow \mu\mu$  rate is completely photon-inclusive, the correction factor approaches unity. It is also worth remarking explicitly that in the soft-photon limit, as assumed throughout in the above calculation, the correction factor in eq. (3.24) is universal, is namely one and the same irrespectively of the external states [103]. The reason is that the photon energy is too low to resolve, e.g., the external-state spin. Note that the soft-photon limit is the only limit in which the above calculation can easily be performed for an arbitrary number of photons. In fact, away from this limit, there is no obvious way to e.g. factorize photon loops in virtual-photon corrections.

As already alluded to, the choice of  $\Lambda_V$  in the calculation is, in principle, arbitrary: different values lead to a redefinition of the unspecified  $O(\alpha_{\text{em}}/\pi)$  finite term in eq. (3.23). Following ref. [104], in our paper [4] we took  $\Lambda_V$  to equal its kinematical limit ( $m_{B_s}/2$ ) in order to minimize the residual finite corrections in the  $O(\alpha_{\text{em}}/\pi)$  term. The latter represents the subleading model-dependent contribution due to infrared-finite virtual corrections and to the residual contribution of real emission, that vanishes in the limit of vanishing photon energy. This term is expected to be below the 1% level, as we discuss next. The  $O(\alpha_{\text{em}}/\pi)$  term can in fact be decomposed into the following three parts:

- I. The residual real and virtual corrections in the absence of direct couplings of the meson to the photon. With the choice  $\Lambda_V = m_{B_s}/2$ , these corrections amount to  $5\alpha_{\text{em}}/(4\pi) \approx 0.3\%$ , corresponding to the electromagnetic corrections for the decay of a point-like meson fully inclusive of bremsstrahlung radiation [105].
- II. The virtual structure-dependent terms (due to effective non-minimal couplings of the meson to the photon). These terms are model dependent; however, they must respect the helicity suppression of the non-radiative amplitude and do not contain large logs. As a result, they are expected to be of the same size as those in point I.
- III. The real contribution of the direct-emission amplitude. Since the direct-emission amplitude for  $B_s \rightarrow \mu^+\mu^-\gamma$  is not helicity suppressed, it may represent a significant contribution if the  $E_{\max}$  cut is not tight enough. However, according to Low's theorem [106], the interference of bremsstrahlung and direct-emission amplitudes leads to a correction to the rate that vanishes at least quadratically with the photon energy cut. From a naive dimensional analysis, the relative direct-emission contamination, for a given  $E_{\max}$  cut, is

$$\delta_{\text{DE}} \leq 2b \left( \frac{2E_{\max}}{m_{B_s}} \right)^2 \times \left[ \frac{\alpha \mathcal{B}(B_s \rightarrow \mu^+\mu^-\gamma)_{\text{DE}}}{\pi \mathcal{B}_{s,\text{SM}}^{(0)}} \right]^{1/2}, \quad (3.28)$$

where  $\mathcal{B}(B_s \rightarrow \mu^+\mu^-\gamma)_{\text{DE}}$  represents the genuine direct-emission branching fraction. According to the estimates in the literature (see ref. [101] and references therein) the latter is  $O(\text{few} \times 10^{-8})$ . Then, if we assume  $\mathcal{B}(B_s \rightarrow \mu^+\mu^-\gamma)_{\text{DE}} < 10^{-7}$  as a conservative estimate, we find that this relative correction is below 1% for  $E_{\max} < 100$  MeV. The recent explicit analysis of ref. [107] confirms this figure.

As a final point, it should be remarked that the derivation leading to eqs. (3.22) and (3.27) assumes a constant efficiency over the whole range of photon-energy integration. Experimentally, this is typically not the case. In a more rigorous treatment, one should convolute the

real- and virtual-photon correction integrals with the appropriate efficiency functions. In this case, the estimate in eq. (3.27) should be regarded as a reference figure against which to compare the rigorous result. As a matter of fact, this is the approach followed by experiments.

### 3.5 $B_s - \bar{B}_s$ oscillation effects

A last point towards connecting the theoretical with the experimental branching ratio is to take into account that the initial-state  $B_s$  is not a mass eigenstate, and oscillates into its antiparticle. In an experimental observable where the time information is not retained, this amounts to a systematic correction with respect to the corresponding ‘instantaneous’ theory quantity. This effect was initially pointed out for  $B_s$  decays in [108], and its importance in the light of the large  $B_s$ -system width difference [109] reappraised in refs. [85,86], whose line of argument will be shortly reviewed next.

At  $pp$  or  $ee$  collisions alike, one starts from the production of a  $b\bar{b}$  pair, whereby at time  $t = 0$  the  $b$  hadronizes into a  $\bar{B}_s$  or else the  $\bar{b}$  hadronizes into a  $B_s$ . The  $t = 0$  state is thus a flavor eigenstate. This state then evolves with time, according to evolution equations analogous to (2.19). The corresponding state at a generic proper time  $t$  is denoted as  $B_s(t)$  or  $\bar{B}_s(t)$ , respectively, and eventually decays into a given state of interest. As far as the  $B_s(t) \rightarrow \mu\mu$  decay is concerned, it should be noted that, experimentally

- there is no flavor tagging, namely one measures the sum of  $B_s(t) \rightarrow \mu\mu$  and  $\bar{B}_s(t) \rightarrow \mu\mu$  decays;
- the decay time information is (at least at the time of this writing) not used either.

Therefore, the experimentally-measured branching ratio of interest can be defined as [108]:

$$\mathcal{B}_{\text{exp}} \equiv \langle \mathcal{B}(B_s \rightarrow \mu\mu) \rangle \equiv \int_0^\infty dt \frac{\Gamma(B_s(t) \rightarrow \mu\mu) + \Gamma(\bar{B}_s(t) \rightarrow \mu\mu)}{2}. \quad (3.29)$$

The integrand is a sum of two exponentials [108]

$$\Gamma(B_s(t) \rightarrow \mu\mu) + \Gamma(\bar{B}_s(t) \rightarrow \mu\mu) = R_H^{\mu\mu} e^{-\Gamma_s^H t} + R_L^{\mu\mu} e^{-\Gamma_s^L t}, \quad (3.30)$$

with  $\Gamma_s^{H,L}$  the widths of the heavier and lighter  $B_s$ -system mass eigenstates, and  $R_{H,L}$  two positive constants. The dependence on any other quantity relevant to  $B_s - \bar{B}_s$  mixing, in particular the masses  $m_{B_s^H}$  and  $m_{B_s^L}$ , cancels in the sum (3.30). This observation makes the integration of eq. (3.29) straightforward, yielding [85]

$$\mathcal{B}_{\text{exp}} = \frac{1}{\Gamma_s} \left( \frac{R_H + R_L}{2} \right) \frac{1}{1 - \Delta\Gamma_s/(2\Gamma_s)}, \quad (3.31)$$

where

$$\Gamma_s \equiv \frac{\Gamma_s^L + \Gamma_s^H}{2}, \quad \Delta\Gamma_s \equiv \Gamma_s^L - \Gamma_s^H. \quad (3.32)$$

Note that the derivation leading to eq. (3.31) assumes that the relevant experimental efficiencies, to be collectively indicated with  $\epsilon$ , be constant with respect to time. A more rigorous treatment involves measuring  $\epsilon(t)$ , or estimating it otherwise, and convoluting this function with the integrand in eq. (3.30). This procedure is currently being implemented by the LHCb collaboration, as more statistics is accumulated and time-dependent measurements become feasible.

Let us now turn to the theoretical branching ratio. The quantity defined in eq. (3.1) is equivalent to

$$\mathcal{B}_{\text{th}} \equiv \frac{\Gamma(B_s(t) \rightarrow \mu\mu) + \Gamma(\bar{B}_s(t) \rightarrow \mu\mu)}{2\Gamma_s} \Big|_{t=0} . \quad (3.33)$$

In fact, the only factor that changes in eq. (3.1) between the  $B_s(t=0)$  and  $\bar{B}_s(t=0)$  cases is the CKM coupling going into its complex conjugate. Taking also into account that eq. (3.1) is normalized to the *average*  $B_s$ -system width,  $\Gamma_s$ , it is clear that the  $B_s$ - or  $\bar{B}_s$ -decay formulae are identical, whence eq. (3.33) follows. Using eq. (3.30) one arrives at

$$\mathcal{B}_{\text{th}} = \frac{1}{\Gamma_s} \left( \frac{R_H + R_L}{2} \right) . \quad (3.34)$$

Comparing eqs. (3.31) and (3.34) one finally has [85]

$$\mathcal{B}_{\text{exp}} = \mathcal{B}_{\text{th}} \frac{1}{1 - y_s} , \quad \text{with } y_s \equiv \Delta\Gamma_s / (2\Gamma_s) . \quad (3.35)$$

Therefore, the  $B_s - \bar{B}_s$  mixing effect amounts to a multiplicative correction between the theoretical and the experimental branching ratios. This correction is of the same order as  $\Delta\Gamma_s / (2\Gamma_s)$ , namely the relative magnitude between the difference and the sum of the widths of the  $B_s$ -system mass eigenstates.  $\Delta\Gamma_s$  has been measured with good accuracy by LHCb [109], including its sign [110], yielding

$$y_s = \frac{\Delta\Gamma_s}{2\Gamma_s} = +0.088 \pm 0.014 . \quad (3.36)$$

The correction in eq. (3.35) amounts therefore to an O(10%) upward shift in the experimental branching ratio with respect to the corresponding theoretical one.

There is actually a straightforward physics explanation for the above shift. Recall that the non-radiative branching ratio in eq. (3.1) is normalized by the average lifetime  $\tau_{B_s}$  of the  $B_s$  system. One can immediately check that the correction factor  $1/(1 - y_s) = \Gamma_s / \Gamma_s^H$ . Therefore, this factor just changes the overall normalization of the branching ratio in eq. (3.1) from  $1/\Gamma_s$  to  $1/\Gamma_s^H$ , to account for the fact that it is, to first approximation, only the heavier (and longer lived) between the two mass eigenstates to actually decay into two muons. This statement is equivalent to saying that the double-exponential functional behavior in eq. (3.30) is actually well described by a single exponential.

\*\*\*

To conclude this section, we can now slightly generalize relation (3.35), and write down a final formula connecting the theoretically calculated with the experimentally measured  $B_s \rightarrow \mu\mu$  decay, including also the soft-photon correction presented in the previous section.

First of all, in eq. (3.29) we have, for the sake of presentation of the main argument, taken the upper limit of integration to be infinity. Leaving this upper limit generic, one would define the following, time-integrated branching ratio

$$\begin{aligned} \langle \mathcal{B}(B_s \rightarrow \mu\mu) \rangle_{[t]} &\equiv \int_0^t dt' \frac{\Gamma(B_s(t') \rightarrow \mu\mu) + \Gamma(\bar{B}_s(t') \rightarrow \mu\mu)}{2} \\ &= \kappa^{\mu\mu}(t, y_s) \langle \mathcal{B}(B_s \rightarrow \mu\mu) \rangle_{[t=0]} \\ &\equiv \kappa^{\mu\mu}(t, y_s) \frac{\Gamma(B_s(t=0) \rightarrow \mu\mu) + \Gamma(\bar{B}_s(t=0) \rightarrow \mu\mu)}{2\Gamma_s} . \end{aligned} \quad (3.37)$$

Here we have introduced the  $\kappa^{\mu\mu}(t, y_s)$  factor, whose SM expression reads [85]

$$\kappa_{\text{SM}}^{\mu\mu}(t, y_s) = \frac{1}{1 - y_s} \left[ 1 - e^{-t/\tau_{B_s}} \sinh\left(\frac{y_s t}{\tau_{B_s}}\right) - e^{-t/\tau_{B_s}} \cosh\left(\frac{y_s t}{\tau_{B_s}}\right) \right] \xrightarrow{t \gg \tau_{B_s}} \frac{1}{1 - y_s}, \quad (3.38)$$

and  $y_s$  has been introduced in eq. (3.35). Eq. (3.37) thus generalizes eq. (3.35), and it reduces itself to the latter in the limit  $t \gg \tau_{B_s}$ .

We can finally put together eq. (3.22) and eq. (3.37) to arrive to our final formula connecting the theoretical with the experimental branching ratio

$$\langle \mathcal{B}(B_s \rightarrow \mu^+ \mu^- (\gamma)) \rangle_{[t, E_{\text{max}}]}^{\text{SM}} = \kappa_{\text{SM}}^{\mu\mu}(t, y_s) \times \omega(E_{\text{max}}) \times \mathcal{B}_{s, \text{SM}}^{(0)}, \quad (3.39)$$

whereby the l.h.s. is the quantity accessible in experiments, whereas  $\mathcal{B}_{s, \text{SM}}^{(0)}$  is the non-radiative branching ratio of eq. (3.1).

A few concluding comments are in order:

- The quantity which is more interesting for precise SM tests, and which can easily be affected by new-physics contributions, is  $\mathcal{B}^{(0)}$ . The correction term  $\omega(E_{\text{max}})$  is insensitive to new physics, while  $\kappa^{\mu\mu}(t, y_s)$  can deviate from its SM expression only in the presence of new-physics models with new CP-violating phases and/or non-standard short-distance operators contributing to the  $B_s \rightarrow \mu^+ \mu^-$  amplitude (see [86]). Most importantly, the two correction terms  $\omega(E_{\text{max}})$  and  $\kappa^{\mu\mu}(t, y_s)$  need to be *convoluted* with the experimental efficiencies on  $E_{\text{max}}$  and  $t$ , as already mentioned. In principle, they can even be determined experimentally up to their overall normalization, although an experimental determination of both these terms will become feasible only with a significant sample of  $B_s \rightarrow \mu\mu$  events. As a result, the experimental collaborations usefully provide determinations of  $\mathcal{B}^{(0)}$ , already corrected for these two terms.
- Since  $\omega_{\text{IB}}(m_{B_s}/2) = 1$  and  $\kappa_{\text{SM}}^{\mu\mu}(t, y_s) \approx t/\tau_{B_s}$  for  $t \ll \tau_{B_s}$ , the theoretical quantity  $\mathcal{B}_{s, \text{SM}}^{(0)}$  can be identified with the SM branching ratio of a flavor-tagged  $B_s$  state at small times, fully inclusive of bremsstrahlung radiation only. We stress that the necessity to include the correction factor  $\omega(E_{\text{max}})$  does depend on the treatment of the electromagnetic radiation in the measurement. In the second reference in [79], as well as in the recent LHCb and CMS  $B_s \rightarrow \mu\mu$  measurements [80, 81], for example, the signal is simulated fully inclusive of bremsstrahlung radiation and the correction term  $\omega(E_{\text{max}})$  (properly convoluted) is taken into account in the signal efficiency.
- Finally, it is interesting to note that, for the experimental choice of  $E_{\text{max}}$  applied by LHCb ( $E_{\text{max}} \approx 60$  MeV, see second ref. in [79]) and for  $t \gg \tau_{B_s}$ , the two correction terms in eq. (3.39) tend to compensate each other to a large extent.

### 3.6 Prediction for the non-radiative decay and error budget

Barring further systematic effects not yet thought of, the discussion in the previous two sections addresses all the systematic effects that are competitive in size with the parametric errors associated to the non-radiative branching ratio in eq. (3.1). We discuss the latter error component in the present section, and also provide final predictions for the  $B_s \rightarrow \mu\mu$  and  $B_d \rightarrow \mu\mu$  decay rates.

By looking at eq. (3.1) as well as the input in table 1, four are the main sources of parametric uncertainty for the  $B_s \rightarrow \mu\mu$  decay rate. Their impact can be illustrated through

the following parametric formula [4]

$$\begin{aligned}\mathcal{B}_{s,\text{SM}}^{(0)} &= 3.2348 \times 10^{-9} \times \left(\frac{M_t}{173.2 \text{ GeV}}\right)^{3.07} \left(\frac{f_{B_s}}{227 \text{ MeV}}\right)^2 \left(\frac{\tau_{B_s}}{1.466 \text{ ps}}\right) \left|\frac{V_{tb}^* V_{ts}}{4.05 \times 10^{-2}}\right|^2 \\ &= (3.23 \pm 0.15 \pm 0.23 f_{B_s}) \times 10^{-9}.\end{aligned}\quad (3.40)$$

In order of increasing impact on the total error, the main sources of parametric uncertainty are:

- the top mass, that, while entering in eq. (3.40) with a power around three, is known with an accuracy of about 5 per mil. Its impact on the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  relative error is therefore 1.6%;
- the  $B_s$ -system average lifetime  $\tau_{B_s}$ , known with about 2% accuracy, and contributing the same relative uncertainty in  $\mathcal{B}(B_s \rightarrow \mu\mu)$ ;
- the CKM coupling, known with about 2% accuracy, but entering quadratically in eq. (3.1), and thus contributing about 4% relative error;
- the  $B_s$ -meson decay constant  $f_{B_s}$ , that, according to the 3.5% error quoted in table 1 and quadratic contribution to eq. (3.1), gives the main contribution to the parametric error, of about 7%.

In fact, the  $f_{B_s}$  error deserves a dedicated discussion. As pointed out in [111], in principle one can get rid of the quadratic  $f_{B_s}$  dependence in  $\mathcal{B}(B_s \rightarrow \mu^+\mu^-)$  by normalizing this observable to  $\Delta m_{B_s}$ , thereby taking advantage of the relatively precise lattice results on the bag parameter of the  $\bar{B}_s - B_s$  mixing amplitude, that enters only linearly the mixing amplitude itself. Moreover, this procedure removes also the dependence on the CKM parameters. Indeed, in 2003 this proposal reduced the uncertainty in  $\mathcal{B}(B_s \rightarrow \mu^+\mu^-)$  by a factor of three. However, given the recent progress in the direct determination of  $f_{B_s}$  from the lattice [96, 112, 113]<sup>38</sup> and in the determination of CKM parameters, this strategy is no longer necessary.

As far as the direct lattice determination of  $f_{B_s}$  is concerned, an impressive progress has been made in the last years [96, 112, 113]. These results are summarized in [96] and included in the world average  $f_{B_s} = (227.6 \pm 5.0) \text{ MeV}$  [69]. Using this figure at face value we would get a total error on  $\mathcal{B}_{s,\text{SM}}^{(0)}$  of  $\pm 0.2 \times 10^{-9}$ . However, given that this average is largely dominated by a single determination (the second reference in [112]), and given that all the other unquenched estimates of  $f_{B_s}$  have errors of about  $\pm 10 \text{ GeV}$ , we believe that a  $\pm 8 \text{ MeV}$  error on  $f_{B_s}$  – that we deduce from the spread of the central values – is a more conservative estimate of the present uncertainty.

With the input in table 1, and treating all errors as Gaussian, ref. [4] reported the following result for  $\mathcal{B}_{s,\text{SM}}^{(0)}$

$$\mathcal{B}_{s,\text{SM}}^{(0)} = (3.23 \pm 0.27) \times 10^{-9}, \quad (3.41)$$

with namely an 8.4% total parametric uncertainty, that can be understood as the sum in quadrature of the error components discussed below eq. (3.40). The prediction in eq. (3.41) has been quoted as the reference SM figure by the recent experimental measurements of the  $B_s \rightarrow \mu\mu$  branching ratio by LHCb [81] and CMS [80].

\*\*\*

As discussed in detail in sec. 3.3, meanwhile the complete NLO EW corrections to the short-distance function in eq. (3.1) have been made available in ref. [88]. The authors calculate the NLO EW matching conditions to the Wilson coefficient of the SM operator  $\mathcal{O}_{10}$  at the weak scale, and obtain the Wilson coefficient at the scale of the external state, of order  $m_{B_s}$ , by solving RGE equations with anomalous dimension matrices including NLO EW corrections as well. In order to get a better handle on the residual scheme dependence, the final results are presented in three schemes: an on-shell scheme<sup>39</sup>, corresponding to our scheme (1,0), as described in the itemization on page 44; an  $\overline{\text{MS}}$  scheme, where all the parameters are running  $\overline{\text{MS}}$  parameters, and corresponding to our (0,1) scheme; a hybrid scheme, corresponding to our (0,0) scheme, where the couplings  $\alpha_{\text{em}}$  and  $\sin\theta_W$  are  $\overline{\text{MS}}$ -renormalized, whereas masses are on-shell. As detailed in sec. 3.3, in ref. [4] we conjectured from the results of [98] that this scheme choice is the one that minimizes the magnitude of NLO EW corrections. This finding is confirmed by the explicit calculation in [88]. As a matter of fact, with the same parametric input, the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  prediction using full NLO EW corrections differs from the one using the prescription of [4] by just 3%. Residual uncertainties due to beyond-NLO EW corrections at the matching as well as at the low scale are estimated to be  $\pm 1\%$  at the level of the branching ratio.

Putting together the calculations of the NLO EW corrections in [88] and of the NNLO QCD ones in [87], ref. [116] presents an updated prediction for the  $B_s \rightarrow \mu\mu$  branching ratio, that supersedes the one in [4]. For the time-integrated branching ratio of eq. (3.29) they quote [116]

$$\langle \mathcal{B}(B_s \rightarrow \mu\mu) \rangle = (3.65 \pm 0.23) \times 10^{-9} , \quad (3.42)$$

which corresponds to the experimental observable. We can compare this prediction with that of ref. [4]. Using (for consistency with ref. [4]) the  $y_s$  value in eq. (3.36), one obtains from eq. (3.42) the following  $t = 0$  branching ratio

$$\mathcal{B}_{s,\text{SM}}^{(0)} = (3.33 \pm 0.21) \times 10^{-9} , \quad (3.43)$$

to be compared with eq. (3.41), corresponding to the prediction in ref. [4].

Beside a slight shift in the central value, the main difference to be noted in eq. (3.43) with respect to (3.41) is the shrinking of the total error, from 8.4% to 6.3%. Main reason is the different error assumed on  $f_{B_s}$ . Ref. [4] adopts the  $f_{B_s}$  value and error reported in table 1, and the assumed 8 MeV error figure, corresponding to a 3.5% relative error, has been justified in sec. 3.6. Conversely, ref. [116] adopts for  $f_{B_s}$  the most recent update of the  $N_f = (2 + 1)$  FLAG compilation [117], which averages (only) the  $N_f = 2 + 1$  results of refs. [112], noting that more recent calculations with  $N_f = 2 + 1 + 1$  [114] are consistent with this average. The latter yields  $f_{B_s} = 227.7 \pm 4.5$  MeV, corresponding to a relative error of about 2.0%, almost halved with respect to the  $f_{B_s}$  error assumed in [4].

A further difference is the treatment of the CKM coupling, that for ref. [116] becomes the dominant source of parametric uncertainty. The relevant CKM combination  $|V_{tb}^* V_{ts}|$  is extracted from the accurately known ratio  $|V_{tb}^* V_{ts}/V_{cb}| = 0.980(1)$  [32, 70], for which it is necessary to fix  $|V_{cb}|$ . While a clean determination of this matrix element is hampered by the long-standing  $\sim 2\sigma$  discrepancy between the determinations from inclusive and exclusive

<sup>38</sup> More results, not included in this discussion, have meanwhile appeared in [114].

<sup>39</sup> Within this scheme, the authors also consider two different normalizations for the Wilson coefficient, whereby the electroweak coupling is proportional to  $G_F^2$  or respectively to  $G_F \alpha_{\text{em}}(M_Z)$ . We have commented on this point in the paragraph following eq. (3.3). The more natural quadratic- $G_F$  normalization was put forward in ref. [115].

semileptonic decays, the authors of [116] adopt the recent  $|V_{cb}|$  fit from ref. [118],<sup>40</sup> yielding  $|V_{cb}| = 0.0424(9)$ . The resulting value for the CKM combination entering  $\mathcal{B}(B_s \rightarrow \mu\mu)$  is  $|V_{tb}^*V_{ts}| = 0.04155$ , whose 2% error is dominated by the  $|V_{cb}|$  error. Ref. [4] instead adopts the  $|V_{tb}^*V_{ts}|$  quoted in table 1, which comes with a similar error.

In short, the main difference in the treatment of parametric errors between ref. [4] and ref. [116] is the fact that the  $f_{B_s}$  error is subleading with respect to the CKM error in [116] and the other way around in [4]. This subject is still debatable.

### 3.7 The $B_d \rightarrow \mu^+\mu^-$ decay

In all of the above sections we barely have made any reference at all to the  $B_d \rightarrow \mu\mu$  decay, in order to limit clutter in the already numerous aspects of the discussion. As a matter of fact, the corresponding analysis of the  $B_d \rightarrow \mu\mu$  decay is a straightforward generalization of the one just presented for  $B_s \rightarrow \mu\mu$ . As far as the three items listed at the end of sec. 3.2 are concerned, the following comments suffice:

- Our analysis of the short-distance part (sec. 3.3), including our scheme-dependence considerations and conjecture, remains unchanged. What is trivially modified in the basic expression in eq. (3.1) are the initial-state constants  $m_{B_s}$ ,  $\tau_{B_s}$ ,  $f_{B_s}$  and the CKM coupling, as now the index  $s$  is replaced by  $d$ . The input parameters used in the  $B_d \rightarrow \mu\mu$  analysis of ref. [4] are given in table 1.
- The soft-photon corrections (sec. 3.4) remains likewise unchanged, as the  $B_s$  and  $B_d$  masses are very close to each other.
- The effect of  $\Delta\Gamma_d$  (sec. 3.5) is negligible.

Thus the final expression in eq. (3.39) is replaced by

$$\langle \mathcal{B}(B_d \rightarrow \mu^+\mu^-(\gamma)) \rangle_{[t, E_{\max}]}^{\text{SM}} = \omega(E_{\max}) \times \mathcal{B}_{d,\text{SM}}^{(0)}. \quad (3.44)$$

Furthermore, using the input in table 1 ref. [4] finds as the analogues of eqs. (3.41) and (3.40) the following results

$$\mathcal{B}_{d,\text{SM}}^{(0)} = (1.07 \pm 0.10) \times 10^{-10}, \quad (3.45)$$

$$\begin{aligned} \mathcal{B}_{d,\text{SM}}^{(0)} &= 1.0659 \times 10^{-10} \times \left( \frac{M_t}{173.2 \text{ GeV}} \right)^{3.07} \left( \frac{f_{B_d}}{190 \text{ MeV}} \right)^2 \left( \frac{\tau_{B_d}}{1.519 \text{ ps}} \right) \left| \frac{V_{tb}^*V_{td}}{8.7 \times 10^{-3}} \right|^2 \\ &= \left( 1.07 \pm 0.05 \pm 0.09_{f_{B_d}} \right) \times 10^{-10}. \end{aligned} \quad (3.46)$$

Similarly as for the  $B_s$  case, eq. (3.46) illustrates the impact of the various inputs on the quoted central value and error. In the second line of this equation, we have explicitly separated out the contribution to the error due to  $f_{B_d}$ , which is the most relevant source of uncertainty.

We have quoted the value in eq. (3.45) because it has been used as SM reference by the recent measurements of  $\mathcal{B}(B_d \rightarrow \mu\mu)$  by the LHCb and CMS collaborations. Meanwhile, this value has been superseded by the prediction of ref. [116], which reads

$$\mathcal{B}_{d,\text{SM}}^{(0)} = (1.06 \pm 0.09) \times 10^{-10}, \quad (3.47)$$

<sup>40</sup> The latter is the first global fit of semileptonic data that includes also recent determinations of heavy-quark masses by independent methods. Inclusion of the heavy-quark mass constraints is found to reduce the fit dependence on assumptions about theoretical error, thereby leading to more precise and stable fit results, a  $|V_{cb}|$  prediction among the others.



barely differing at all from the one of ref. [4].

Finally, eqs. (3.40) and (3.46) translate straightforwardly into a prediction for the ratio of the non-radiative branching ratios,  $\mathcal{B}(B_s \rightarrow \mu\mu)/\mathcal{B}(B_d \rightarrow \mu\mu)$ . Using  $f_{B_s}/f_{B_d} = 1.195$  from the separate constants in our table, and indicating its *relative* error as  $\sigma_{f_{s/d}}^r$ , one easily finds

$$\frac{\mathcal{B}_{s,\text{SM}}^{(0)}}{\mathcal{B}_{d,\text{SM}}^{(0)}} = 30.35 \left( 1 \pm 0.06 \pm 2\sigma_{f_{s/d}}^r \right). \quad (3.48)$$

As we will see in the outlook part, this ratio has the same value within the SM and the entire class of models with MFV. Therefore, a measurement not in agreement with eq. (3.48) would be a clear signal of beyond-SM physics and of new flavor spurions beyond the SM Yukawa couplings.

### 3.8 $B_s \rightarrow \mu\mu$ as a probe of the Yukawa sector

$\mathcal{B}(B_s \rightarrow \mu\mu)$  and  $\mathcal{B}(B_d \rightarrow \mu\mu)$  are among the best flavor-physics probes of beyond-SM effects, a first example being effects due to non-standard Higgs sectors. Innumerable studies exist, where plausible model setups are assumed and enhancements on these observables found, typically by orders of magnitude in case of ‘nearby’, i.e. TeV-scale new physics. Here we will not consider explicitly any of these scenarios. We will instead maintain a model-independent approach, whereby such effects can be described by contributions to operators in an effective-theory expansion, with Hamiltonian (we follow the normalization of [119])

$$\mathcal{H}_{\text{eff}} = -\frac{2G_F}{\sqrt{2}} \frac{\alpha}{2\pi \sin^2 \theta_W} V_{tb}^* V_{ts} \sum_{i=10,S,P} (c_i \mathcal{O}_i + c'_i \mathcal{O}'_i), \quad (3.49)$$

where the effective operators are defined as

$$\begin{aligned} \mathcal{O}_{10} &= (\bar{b}\gamma_L^\mu s)(\bar{\ell}\gamma_\mu\gamma_5\ell), & \mathcal{O}_S &= m_b(\bar{b}P_L s)(\bar{\ell}\ell), & \mathcal{O}_P &= m_b(\bar{b}P_L s)(\bar{\ell}\gamma_5\ell), \\ \mathcal{O}'_{10} &= (\bar{b}\gamma_R^\mu s)(\bar{\ell}\gamma_\mu\gamma_5\ell), & \mathcal{O}'_S &= m_s(\bar{b}P_R s)(\bar{\ell}\ell), & \mathcal{O}'_P &= m_s(\bar{b}P_R s)(\bar{\ell}\gamma_5\ell), \end{aligned} \quad (3.50)$$

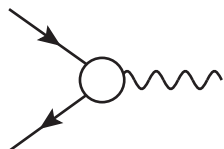
where, for the sake of simplicity, we have specialized the basis to the  $B_s \rightarrow \mu\mu$  case. The  $B_d \rightarrow \mu\mu$  ones is obtained by the obvious replacement  $s \rightarrow d$ . One may write also operators with tensor Dirac structures. However, their matrix elements necessarily vanish, because, on the quark-bilinear side, the two (antisymmetric) Lorentz indices will be contracted with the only four-vector available, the  $B$ -meson four-momentum.

The SM case in eq. (3.49) corresponds to setting  $c_{10} = -Y(x_{tW})$  (see eq. (3.1)) and all the other coefficients to zero. In fact, within the SM the occurrence of scalar and pseudoscalar structures requires a chirality flip on both the quark and the lepton bilinears, implying that the operators  $\mathcal{O}_{S,P}^{(\prime)}$  are suppressed by  $m_{b,s}m_\ell/M_W^2$ . Since the  $m_\ell/M_W$  suppression factor is present also in the  $\mathcal{O}_{10}$  case, the relative suppression of scalar operators with respect to  $\mathcal{O}_{10}$  within the SM amounts to  $m_{b,q}/M_W$  at the amplitude level. As already mentioned, this suppression can be lifted in models where extra fermionic matter can propagate in the loop, so that its mass can replace one of the quark masses (see e.g. [116, 119]).

The considerations below eq. (3.50) show already that the delicate SM structure, with the contribution by the  $\mathcal{O}_{10}$  operator being by far dominant, is ‘easily’ spoiled beyond the SM. We would like to now corroborate this statement with plausibility arguments. We will discuss two such arguments, keeping as mentioned a model-independent approach: the first

based on the sensitivity of the  $B_s \rightarrow \mu\mu$  decay to Yukawa couplings<sup>41</sup>, and the second one on  $B_s \rightarrow \mu\mu$  as a probe of anomalous  $Z$ -quark couplings, and based on the results of [5].

This section will be devoted to the first argument. The probably most elemental way to introduce it is the consideration that the short-distance function in the  $B_s \rightarrow \mu\mu$  rate depends only on one mass ratio,  $x_t = m_t^2/M_W^2$ , and that this ratio is large, equalling about 4. One is then led to entertain the case where this ratio goes to infinity, because  $m_t$  itself goes to the decoupling limit, whereas  $M_W$  is kept finite. We thus need to understand what happens in this limit of the  $Y$  function and of the  $\Delta F = 1$  effective coupling. The  $Y$  function is a gauge-invariant combination of  $Z$ -penguin and box contributions, sample diagrams of which have been depicted in fig. 8.  $Z$ -penguin and box contributions are respectively described by the  $C_0(x_{tW})$  and the  $B_0(x_{tW})$  Inami-Lim [58] functions, with gauge dependence canceling in the combination  $Y_0 = C_0 - B_0$ . (See e.g. [53] for a pedagogical introduction and explicit formulae.) For any  $m_t$  value between a few GeV and infinity the by far numerically larger contribution is the one due to the  $C_0$  function, i.e. to  $Z$ -penguin diagrams. We can therefore safely neglect box contributions for the rest of this discussion. The  $Z$ -penguin diagram can be contracted to an effective vertex, see fig. 10, whose top-loop contribution reads [53]



$$= i V_{tb}^* V_{ts} \frac{G_F}{\sqrt{2}\pi^2} \frac{g}{c_W} M_W^2 C_0(x_{tW}) \bar{b} \gamma_L^\mu s, \quad (3.51)$$

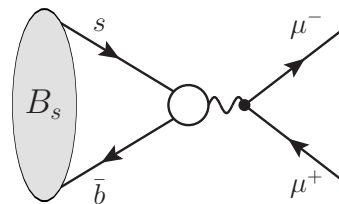
where the  $C_0$  function reads

$$C_0(x) = \frac{x}{8} \left( \frac{x-6}{x-1} + \frac{3x+2}{(x-1)^2} \log x \right). \quad (3.52)$$

Putting together eqs. (3.51) and (3.52), one sees that in the large- $m_t$  limit the  $Z$ -penguin coupling goes as  $G_F M_W^2 \times m_t^2/M_W^2$ , i.e. that it increases as  $m_t^2$ . This behavior appears to violate decoupling of the top quark.

In order to understand more closely what is going on, we should isolate Yukawa from gauge interactions. Specifically, let us switch off altogether gauge interactions, by sending  $g \rightarrow 0$ . This implies that the transverse components of the  $W, Z$  bosons do not contribute anymore to the  $B_s \rightarrow \mu\mu$  amplitude. Furthermore, since in this limit the  $W, Z$  bosons are massless, they do not eat up anymore the three Higgs-field components, to be denoted as  $\phi^\pm$  and  $\phi^0$  respectively, that provide the longitudinal d.o.f. for the  $W, Z$  themselves. It should be noted that, in the  $g = 0$  limit, and to the extent that the Higgs vev  $v$  is kept finite, all the effective-theory diagrams that result from those in fig. 8 when integrating out the  $t, W, Z$  propagator – with the diagram in fig. 10 being the dominant one – still make sense: the propagating  $W, Z$  d.o.f. are the  $\phi^\pm$  and  $\phi^0$ , with mass of  $O(v)$ . Even in the  $g = 0$  limit it is therefore sensible to consider the effective Hamiltonian in eq. (3.49). Let us isolate in its expression all the terms

**Gaugeless  
limit**



**Figure 10:** Same diagram as the leftmost one in fig. 8, but for the fact that the  $Z$ -penguin vertex has been shrunk to a local one. The very short  $Z$ -boson propagator denotes that this is also a local object, as a consistent effective-theory treatment demands.

<sup>41</sup> I am grateful to Gino Isidori for inspiring discussions on this topic. Albeit my line of argument is somewhat different, an important reference for the material to follow is [120].

with non-null mass dimensions, plus those that are functions of  $g$ . They are

$$G_F \cdot \alpha \cdot Y_0(x_{tW}) \sim \frac{1}{v^2} \cdot g^2 \cdot \frac{m_t^2}{M_W^2} \sim \frac{y_t^2}{v^2}, \quad (3.53)$$

where in the middle member we have used the fact that in the  $g \rightarrow 0$  limit one has  $Y_0(x) \rightarrow x/8$ , and on the r.h.s. we have used the relations  $M_W = gv/2$  and  $m_t = y_tv/\sqrt{2}$ , where  $y_t$  is the top Yukawa coupling.

From eq. (3.53) we see that, in fact, at the level of the effective Hamiltonian the  $g$  dependence cancels altogether in the  $g \rightarrow 0$  limit. The short-distance dynamics giving rise to this Hamiltonian is therefore dominated by Yukawa interactions, or more specifically by the large top-Yukawa entry. In short, the  $g = 0$  limit allows to see transparently that in the  $B_s \rightarrow \mu\mu$  decay it is basically the external-states peculiarity – a pseudoscalar decaying into a lepton pair, implying the absence of a photon penguin by gauge invariance – that selects the interesting configuration where only longitudinal  $W, Z$  propagate, i.e. where only Yukawa interactions are at work, as the dominant kinematic configuration.

The above discussion shows that the  $B_s \rightarrow \mu\mu$  decay is an exquisite probe of the Yukawa sector of the SM, namely of Higgs-quarks interactions. This decay provides in particular an excruciating test of the question whether the Higgs sector consists of a single doublet, as in the SM, or else of, for example, two doublets  $H_{1,2}$ <sup>42</sup> with two separate vevs  $v_1$  and  $v_2$ . One plausible case thereof is the case where the former doublet gives mass to down-type fermions and the latter to up-type ones. One could in this case assume that  $v_2 \gg v_1$ , thereby explaining the  $m_t$  vs.  $m_b$  hierarchy in terms of a hierarchy between the vevs, while allowing all the third-generation Yukawa couplings to be of  $O(1)$ <sup>43</sup>. The natural parameter to control this scenario is therefore  $\tan \beta = v_2/v_1 \gg 1$ . As a matter of fact, one finds that  $\mathcal{B}(B_s \rightarrow \mu\mu) \propto \tan^6 \beta$  in this case, namely that this observable can easily undergo enhancements by orders of magnitude for a TeV-scale Higgs sector. This example demonstrates at the same time the enormous sensitivity of this decay even to very decoupled additional degrees of freedom in the Higgs sector.

Finally, eq. (3.53) also clarifies the apparent non-decoupling behavior observed below eq. (3.52) as  $m_t \rightarrow \infty$ . Since  $m_t \sim y_tv$ , the decoupling limit is to be understood as a  $v \rightarrow \infty$ , and not as a  $y_t \rightarrow \infty$  limit. In fact we see that in the  $v \rightarrow \infty$  limit the product in eq. (3.53) – that summarizes the basic coupling vs. mass behavior of the Hamiltonian in eq. (3.49) – does go correctly to zero. Conversely, in the  $y_t \rightarrow \infty$  limit, this product increases indefinitely, confirming again that the effective interaction responsible for  $B_s \rightarrow \mu\mu$  is, to first approximation, built out of Yukawa interactions only.

### 3.9 $B_s \rightarrow \mu\mu$ as an EW precision test

In the previous section we have provided a general argument for the fact that  $B_s \rightarrow \mu\mu$  is an excellent probe of new contributions from scalar and pseudoscalar operators,  $\mathcal{O}_{S,P}^{(\prime)}$ . In this section we turn to the second general argument also mentioned there, namely the fact that  $B_s \rightarrow \mu\mu$  is likewise an outstanding probe of anomalous  $Z$ -quark-quark effective couplings, namely of new contributions from the SM operator  $\mathcal{O}_{10}$ , and especially its right-handed counterpart  $\mathcal{O}'_{10}$ , absent within the SM. Let us consider the effective  $Z$ - $\bar{d}_i$ - $d_j$  coupling depicted in fig. 11. It can be described by the following effective Lagrangian

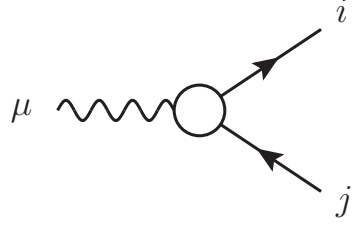
$$\mathcal{L}_{\text{eff}}^Z = \frac{g}{c_W} Z_\mu \bar{d}^i \gamma^\mu \left[ (g_L^{ij} + \delta g_L^{ij}) P_L + (g_R^{ij} + \delta g_R^{ij}) P_R \right] d^j. \quad (3.54)$$

<sup>42</sup> Provided this extended Higgs sector violates flavor.

<sup>43</sup> at a suitable unification scale.

Here  $g_{L,R}^{ij}$  denote the effective SM couplings, whereas  $\delta g_{L,R}^{ij}$  describe non-standard effects. It is useful to recall here the leading structure of the SM couplings:

$$(g_L^{ii})_{\text{tree}} = -\frac{1}{2} + \frac{1}{3}s_W^2, \quad (g_R^{ii})_{\text{tree}} = \frac{1}{3}s_W^2, \\ (g_{L,R}^{i \neq j})_{\text{tree}} = 0. \quad (3.55)$$



**Figure 11:** General  $Z\text{-}\bar{d}_i\text{-}d_j$  effective coupling.

At the one-loop level the  $g_{L,R}^{ii}$  are gauge dependent, but they assume the following simple and gauge-independent form in the already discussed limit  $m_t \gg M_W$  or  $g \rightarrow 0$ :

$$(g_L^{ij})_{1\text{-loop}}^{(g=0)} = \frac{m_t^2}{16\pi^2 v^2} V_{ti}^* V_{tj}, \quad (g_R^{ij})_{1\text{-loop}}^{(g=0)} = 0, \quad (3.56)$$

as one may easily derive from the Feynman rule (3.51) as well.

Let us now consider the flavor-conserving version of the interaction in eq. (3.54), in particular the case  $i = j = 3$ . It will contribute to  $Z$ -peak observables, accurately measured at LEP, in particular  $R_b, A^b$  and  $A_{\text{FB}}^{0b}$ , to be introduced later on. Conversely, the flavor off-diagonal part of this interaction will contribute to  $Z$ -penguin-dominated FCNC processes, most notably  $B_s \rightarrow \mu\mu$ , as already argued around fig. 10. The interesting observation [121, 122] is that there exists a wide class of models where the only relevant deviations from the SM in these two classes of observables can be described in terms of the modified  $Z$ -boson couplings of eq. (3.54).

Purpose of this section is to compare the effectiveness of  $B_s \rightarrow \mu\mu$  to that of  $Z$ -peak observables in constraining these anomalous couplings. In this comparison, we will again try to keep model dependence to a minimum. It turns out that the new-physics contributions parameterized by  $\delta g_{L,R}^{ij}$  can be related to the couplings of a manifestly gauge-invariant effective Lagrangian [5],

$$\mathcal{L}_{\text{eff}}^{\text{NP}} = -\frac{1}{2} \sum_{n,A} \sum_{i,j} \frac{c_{nA}^{ij}}{\Lambda^2} O_{nA}^{ij}, \quad (3.57)$$

with the following set of dimension-six operators:

$$\mathcal{O}_{1L}^{ij} = i \left( \bar{Q}_L^i \gamma^\mu Q_L^j \right) H^\dagger \overleftrightarrow{D}_\mu H, \quad \mathcal{O}_{1R}^{ij} = i \left( \bar{D}_R^i \gamma^\mu D_R^j \right) H^\dagger \overleftrightarrow{D}_\mu H, \\ \mathcal{O}_{2L}^{ij} = i \left( \bar{Q}_L^i \tau^a \gamma^\mu Q_L^j \right) H^\dagger \tau^a \overleftrightarrow{D}_\mu H. \quad (3.58)$$

Defining the flavor indices  $\{i, j\}$  in the mass-eigenstate basis of down-type quarks, the mentioned relation between the  $\delta g_{L,R}^{ij}$  and the  $c_{nA}^{ij}$  is easily found to be

$$\delta g_L^{ij} = \frac{v^2}{4\Lambda^2} \left( c_{1L}^{ij} + \frac{1}{4} c_{2L}^{ij} \right), \quad \delta g_R^{ij} = \frac{v^2}{4\Lambda^2} c_{1R}^{ij}. \quad (3.59)$$

In fact, the set of operators in eq. (3.58) is not the complete set of gauge-invariant dimension-six operators contributing to  $B_s \rightarrow \mu\mu$  and  $Z \rightarrow b\bar{b}$  at the tree level. In principle, we can consider also four-fermion (two-quarks/two-leptons) operators, terms of the type  $J_\nu \times D_\mu F^{\mu\nu}$ , or terms of the type  $H^\dagger J_{\mu\nu} \times F^{\mu\nu}$ , where  $J_\nu$  and  $J_{\mu\nu}$  are quark bilinears, and  $F^{\mu\nu}$  generically denotes the field-strength tensor of  $U(1)$  or  $SU(2)_L$  gauge fields. However, the effects of these operators cannot be described by means of  $\mathcal{L}_{\text{eff}}^Z$  and we lose

**Effective  
theory**

the natural correlation between these two observables.<sup>44</sup> For this reason in the following we concentrate only on the set of operators in eq. (3.58).

In order to correlate effects in  $B_s \rightarrow \mu\mu$  and in  $Z\bar{b}b$  observables, we need to specify the flavor structure of the effective theory, namely the couplings  $c_{nA}^{ij}$ . We consider two reference frameworks:

1. the hypothesis of Minimal Flavor Violation (MFV), as defined in ref. [35], already introduced in chap. 1;
2. the generic flavor structure implied by the hypothesis of Partial Compositeness (PC) [123], following the effective-theory approach described in refs. [124, 125].

To make the discussion to follow self-contained, it is useful to very shortly review here the basic aspects of the effective-theory approach to PC [124, 125] advocated above. The main advantage of this approach is that, akin to the MFV approach of [35], it allows to organize EFT operators according to their flavor suppression.

The defining property of PC is the assumption that, at a cutoff scale  $\Lambda$ , the SM fermions, to be generically denoted as  $f_i^a$ , couple *linearly* to operators  $\mathcal{O}_i^a$  of a confining sector, namely

**PC: short  
introduction**

$$\text{PC-defining interactions} = g_\rho \epsilon_i^a f_i^a \mathcal{O}_i^a, \quad (3.60)$$

where the index  $a$  labels the SM fermion fields ( $Q_L, u_R, \dots$ ) and  $i$  their flavor; the operators  $\mathcal{O}_i^a$  interpolate the resonances of the confining sector, characterized by the reference mass  $m_\rho$  and coupling  $g_\rho$ , and the  $\epsilon_i^a$  accordingly measure the degree of compositeness of the fermion  $f_i^a$ : if a given  $\epsilon$  is zero, the corresponding fermion is not coupled to the composite sector, hence it is elementary, as in the SM; conversely, for  $\epsilon \sim 1$ , the given fermion can be interpreted as a fully composite state (see [125] for a nice introduction).

From eq. (3.60) it is evident that the relevant low-energy d.o.f. are not  $f_i^a$ , but rather  $\epsilon_i^a f_i^a$ . Hence, building our effective theory out of the  $\epsilon_i^a f_i^a$ , the flavor structure will be automatically fixed – apart from O(1) factors that will be immaterial to our main points. From the form of the low-energy Lagrangian<sup>45</sup> implied by the assumption (3.60) one evinces the following structure for the familiar SM Yukawa matrices [125]

$$(Y_u)_{ij} \sim g_\rho \epsilon_i^q \epsilon_j^u, \quad (Y_d)_{ij} \sim g_\rho \epsilon_i^q \epsilon_j^d, \quad (3.61)$$

where  $\sim$  denotes that this form is fixed up to O(1) unknown matrices in flavor space, related to the hypothesis of flavor anarchy in the composite sector. Relations (3.61) can be understood intuitively by taking two instances of the interaction in eq. (3.60) and integrating out the composite operators  $\mathcal{O}$ . The result is a low-energy interaction bilinear in the SM fermion fields and in their respective amounts of compositeness.

We can now go back to our main line of discussion, concerned with studying correlated effects to  $B_s \rightarrow \mu\mu$  and to  $Z\bar{b}b$  observables within the two reference frameworks of MFV or PC. The discussion to follow will show that, in either case of MFV or PC, flavor-violating and flavor-conserving couplings will be proportional to two universal shifts:  $\delta g_L$  and  $\delta g_R$ . Once these two couplings are specified, also predictions for  $B_s \rightarrow \mu\mu$  and  $Z\bar{b}b$  observables are.

In the MFV framework there is a strict correlation between flavor-diagonal (but non-

**MFV  
case**

<sup>44</sup> The four-fermion operators do not contribute to  $\mathcal{L}_{\text{eff}}^Z$  at the tree level, hence they have a negligible impact on  $Z \rightarrow b\bar{b}$  compared to  $B_s \rightarrow \mu\mu$ . Conversely, operators with the field-strength tensor generate amplitudes suppressed by at least one power of  $p/v$ , with  $p$  the external momentum, that therefore have negligible impact on  $B_s \rightarrow \mu\mu$  compared to  $Z \rightarrow b\bar{b}$ .

<sup>45</sup> This form can be derived using naive dimensional analysis, in a fashion similar to the way ChPT is derived from QCD [125].

universal) and flavor-violating couplings of the operators listed in eq. (3.58). Restricting to the contributions relevant to this correlation, the effective couplings can be decomposed as follows:

$$(c_{nL}^{ij})^{\text{MFV}} = a_{nL} \times (Y_u Y_u^\dagger)_{ij} \approx a_{nL} \frac{2m_t^2}{v^2} V_{ti}^* V_{tj}, \quad (3.62)$$

$$(c_{1R}^{ij})^{\text{MFV}} = a_{1R} \times (Y_d^\dagger Y_u Y_u^\dagger Y_d)_{ij} \approx a_{1R} \frac{4m_{d_i} m_{d_j} m_t^2}{v^4} V_{ti}^* V_{tj}, \quad (3.63)$$

where  $a_{nL,R}$  are unknown  $O(1)$  couplings and  $Y_{u,d}$  are the SM Yukawa couplings. The last equalities in eqs. (3.62), (3.63) hold after rotating the Yukawa matrices in the mass-eigenstate basis of down-type quarks, where  $Y_u = V^\dagger \lambda_u$  and  $Y_d = \lambda_d$ , with  $\lambda_{u,d}$  diagonal matrices [35].

From eqs. (3.62) and (3.63) it follows that we can parameterize all the  $\delta g_{L,R}^{ij}$  in terms of two flavor-blind parameters,  $\delta g_{L,R}$ , as mentioned. The latter are defined by

$$(\delta g_L^{ij})^{\text{MFV}} = \frac{V_{ti}^* V_{tj}}{|V_{tb}|^2} \delta g_L, \quad (\delta g_R^{ij})^{\text{MFV}} = \frac{m_{d_i} m_{d_j}}{m_b^2} \frac{V_{ti}^* V_{tj}}{|V_{tb}|^2} \delta g_R. \quad (3.64)$$

The normalization has been chosen such that  $\delta g_{L,R}$  can be directly identified with the  $i = j = 3$  shifts relevant to  $Z\bar{b}b$  observables, namely

$$\delta g_{L(R)}^b \equiv \delta g_{L(R)}^{33} = \delta g_{L(R)}. \quad (3.65)$$

With this identification,  $\delta g_{L,R}$  corresponds to the usual definition of the modified  $Z \rightarrow b\bar{b}$  couplings [126]. As can be seen, in the left-handed sector the flavor structure is identical to the one of the leading one-loop contribution within the SM, reported in eq. (3.56). In the right-handed sector the structure is different but the effects are expected to be very small due to the strong suppression of down-type masses. Indeed, eqs. (3.59)-(3.65) imply

$$\delta g_L^{\text{MFV}} = \frac{m_t^2 |V_{tb}|^2}{2\Lambda^2} \left( a_{1L} + \frac{1}{4} a_{2L} \right), \quad \delta g_R^{\text{MFV}} = \frac{m_b^2 m_t^2 |V_{tb}|^2}{v^2 \Lambda^2} a_{1R}.$$

Since all the  $a_{nL,R}$  couplings are expected to be of  $O(1)$ , the above relation displays that, within MFV, the  $\delta g_R$  shift is naturally suppressed by  $m_b^2/v^2$  with respect to the  $\delta g_L$  one.

A completely analogous derivation can be carried out in the case of PC. In this framework the correlation between flavor-diagonal and flavor-violating couplings is determined up to unknown  $O(1)$  parameters, as already commented on after eq. (3.61). In this case, following the notation of ref. [125], we expect

$$(c_{nL}^{ij})^{\text{PC}} \sim \frac{g_\rho^2 \Lambda^2}{m_\rho^2} \epsilon_i^q \epsilon_j^q \propto |V_{ti}| |V_{tj}|, \quad (3.66)$$

$$(c_{1R}^{ij})^{\text{PC}} \sim \frac{g_\rho^2 \Lambda^2}{m_\rho^2} \epsilon_i^d \epsilon_j^d \propto \frac{m_{d_i} m_{d_j}}{v^2 |V_{ti}| |V_{tj}|}. \quad (3.67)$$

On the r.h.s. of these relations we have eliminated the  $\epsilon_i^{q,d}$  in favor of quark masses and CKM angles by means of the relations [124, 125]

$$\frac{|\epsilon_i^q|}{|\epsilon_j^q|} \sim \frac{|V_{ti}|}{|V_{tj}|}, \quad \frac{|\epsilon_i^q \epsilon_i^d|}{|\epsilon_j^q \epsilon_j^d|} \sim \frac{m_{d_i}}{m_{d_j}}. \quad (3.68)$$

Even without a detailed reading of [124, 125], the r.h.s. of eq. (3.66) is actually straightforward to obtain, keeping in mind relations (3.61) and adopting the convenient quark-field

**PC case**

basis already advocated below eqs. (3.62)-(3.63), wherein  $Y_d$  is diagonal, and  $Y_u = V^\dagger \lambda_u$ . As can be seen from eqs. (3.66)-(3.67), up to  $O(1)$  factors the flavor structure of the left-handed couplings is the same as in the MFV framework. On the other hand, the structure is significantly different in the right-handed sector, where larger effects are now possible in the flavor-violating case. Ignoring  $O(1)$  factors, we parameterize the structure of the two couplings in the PC framework as follows:

$$(\delta g_L^{ij})^{\text{PC}} = \frac{|V_{ti}||V_{tj}|}{|V_{tb}|^2} \delta g_L, \quad (\delta g_R^{ij})^{\text{PC}} = \frac{m_{d_i} m_{d_j}}{m_b^2} \frac{|V_{tb}|^2}{|V_{ti}||V_{tj}|} \delta g_R. \quad (3.69)$$

Similarly to relations (3.64), these relations are completely general within the PC framework, and parameterize the  $\delta g_{L,R}^{ij}$  in terms of two flavor-blind numbers,  $\delta g_{L,R}^{\text{PC}}$ , as advertised at the beginning of this section. Again the normalization has been chosen in order to satisfy eq. (3.65), so that  $\delta g_{L,R}^{\text{PC}}$  are directly the  $Z\bar{b}b$  shifts. With such choice, one also finds

$$\delta g_L^{\text{PC}} \sim \left( \frac{g_\rho \epsilon_3^q v}{2m_\rho} \right)^2, \quad \delta g_R^{\text{PC}} \sim \frac{1}{2} \left( \frac{m_b}{\epsilon_3^q m_\rho} \right)^2. \quad (3.70)$$

## Analysis

The previous considerations can be summarized by stating that, within the two reference frameworks of MFV or PC, possible departures from the SM predictions in the  $Z\bar{b}b$  couplings and in  $\mathcal{B}(B_s \rightarrow \mu\mu)$  can be parameterized in terms of the two couplings  $\delta g_{L,R}$  defined in eq. (3.64) for MFV or in eq. (3.69) for PC. We are now ready to discuss the observables.

Concerning  $Z$ -peak observables, the  $\delta g_{L,R}$  shifts are constrained by  $R_b$ ,  $A^b$  and  $A_{\text{FB}}^{0b}$ . A detailed discussion of all these observables can be found in [126]. To make our discussion perspicuous, we will shortly repeat their definition here.  $R_b$  is defined as the width of  $Z \rightarrow b\bar{b}$ , denoted as  $\Gamma_b$ , normalized to the total width of  $Z \rightarrow q\bar{q}$ , whereby one typically takes  $\Gamma_c \approx \Gamma_u$  and  $\Gamma_s \approx \Gamma_d$  (see e.g. [127]). For a generic fermion  $f$ , the asymmetry  $A^f$  is defined as [126]

$$A_f \equiv \frac{(g_L^f)^2 - (g_R^f)^2}{(g_L^f)^2 + (g_R^f)^2}, \quad (3.71)$$

where  $g_{L,R}^f$  denotes the left- and respectively right-handed couplings between the  $Z$  and the fermion  $f$ . Finally,  $A_{\text{FB}}^{0b}$  may be defined as follows [126]

$$A_{\text{FB}}^{0b} = \frac{3}{4} A^e A^b, \quad (3.72)$$

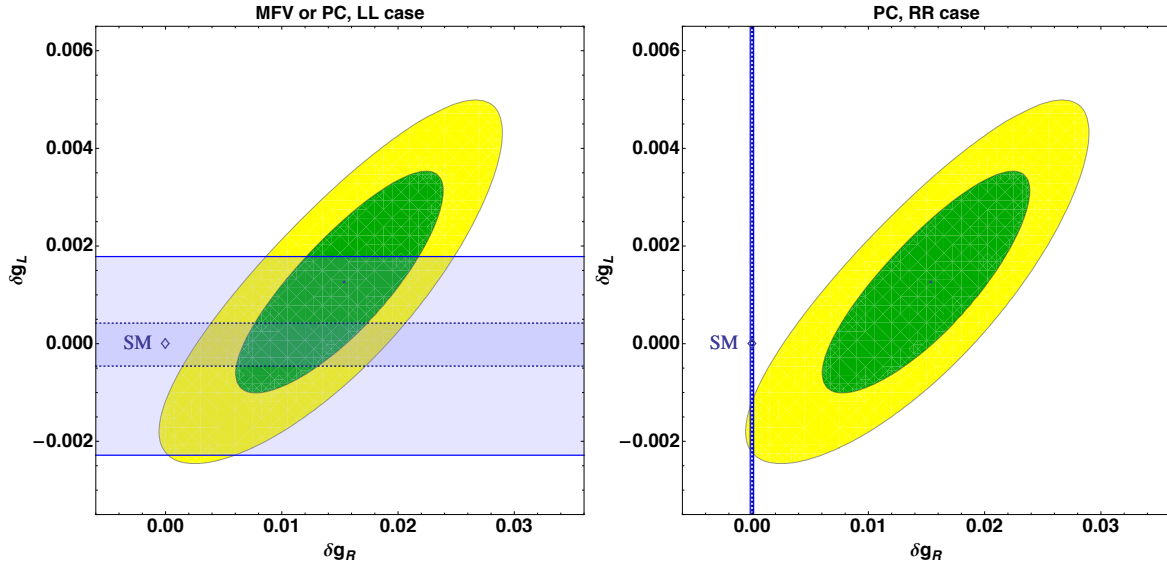
and represents the forward-backward asymmetry, averaged over the beam ( $e$ ) and final-state ( $b$ ) polarizations.

The state-of-the-art SM calculations for these quantities, to which it is straightforward to add the generic shifts in eq. (3.65), can be implemented following ref. [126], taking also into account the recent SM estimate of  $R_b$  in ref. [127]. These quantities can then be fitted to the averages of experimental results collected in table 3, where we also report the main inputs necessary for their evaluation beyond the lowest order. All the inputs collected in this table are those used in ref. [5].

The resulting allowed regions at 68% CL and 95% CL in the  $\delta g_R$ - $\delta g_L$  plane are shown in fig. 12. As can be noticed, for both  $\delta g_L$  and  $\delta g_R$  the fit prefers positive non-zero values, and

$M_h = 125 \text{ GeV}$	[18]	$\Delta\alpha_{\text{had}}^{(5)} = 0.02772$
$M_t = 173.2(0.9) \text{ GeV}$	[128]	$R_b = 0.21629(66)$
$\alpha_s(M_Z) = 0.1184(7)$	[92]	$A^b = 0.923(20)$
$\alpha^{-1}(M_Z) = 127.937$	[91]	$A_{\text{FB}}^{0b} = 0.0992(16)$

**Table 3:** Input parameters relevant for the  $Z \rightarrow b\bar{b}$  constraints as used in ref. [5]. Quantities without an explicit reference are taken from ref. [26]. We do not show the errors for quantities whose uncertainty has a negligible impact on our numerical analysis.



**Figure 12:** Constraints on the couplings  $\delta g_{L,R}$  describing the modified  $Z$ -boson couplings to down-type quarks. The inner and outer ellipses denote respectively the 68% and 95% CL regions as obtained from  $Z\bar{b}b$  observables. The region delimited by solid blue lines denotes the constraint from the 95% CL experimental range for  $\mathcal{B}(B_s \rightarrow \mu\mu)$  in eq. (3.75), while those comprised between dotted lines are obtained with the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  accuracy expected by 2018 (see text for details). *Left panel:*  $\delta g_L$  constraint from  $\mathcal{B}(B_s \rightarrow \mu\mu)$  under the hypotheses of either MFV or PC. *Right panel:*  $\delta g_R$  constraint from  $\mathcal{B}(B_s \rightarrow \mu\mu)$  under the hypothesis of PC.

the SM point ( $\delta g_R = \delta g_L = 0$ ) is outside the 95% CL region. The upper limits for the two parameters are

$$|\delta g_L|_{Z\bar{b}b} < 4.5 \times 10^{-3}, \quad |\delta g_R|_{Z\bar{b}b} < 3.0 \times 10^{-2} \quad [95\% \text{ CL}], \quad (3.73)$$

in good agreement with the results recently reported in the analysis of ref. [129].

A comment is in order concerning the  $Z$ -peak input quantities to the EW-fit ellipse in fig. 12. The  $O(2)\sigma$  tension between the EW-fit minimum and the SM expectation was, at the time of publication of ref. [5], mostly driven by  $A_{\text{FB}}^{0b}$  and, to a lesser extent, by  $R_b$ . In fact, the discrepancy between the SM point ( $\delta g_R = \delta g_L = 0$ ) and experiment was  $2.5\sigma$  and  $2.1\sigma$  for  $A_{\text{FB}}^{0b}$  and  $R_b$ , respectively. The rather large discrepancy in  $R_b$  was new, and caused by the improved two-loop calculation of ref. [127], which exhibited an unexpectedly large negative correction (see also ref. [130], in particular its figure 1). After publication of [5], the  $R_b$  result in [127] underwent a major revision (erratum), resulting in a sizable reduction of the two-loop correction. As a matter of fact, after this revision the agreement between the SM  $R_b$  prediction and experiment is at the level of  $1.2\sigma$ . As far as fig. 12 is concerned, the effect of this update is to move the EW-fit ellipse slightly upwards. The fit minimum (small dot



within the ellipse) moves from  $\{\delta g_R, \delta g_L\} = \{0.015, 0.0013\}$  to  $\{\delta g_R, \delta g_L\} = \{0.015, 0.0020\}$ , and the SM point remains outside of the outmost ellipse (95% CL region). The results in [5] are therefore hardly affected at all and we can stick with the limits in eq. (3.73).

Let us now compare these limits with those obtained from the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  measurement within the frameworks of MFV or PC. The  $\delta g_{L,R}^{32}$  couplings shift linearly the  $Z$ -penguin contribution to the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  amplitude. These shifts can easily be translated into shifts on the short-distance function appearing in the SM formula for the branching ratio (see e.g. ref. [4]). To good accuracy, the effect can simply be described by

$$\mathcal{B}(B_s \rightarrow \mu\mu) = \mathcal{B}(B_s \rightarrow \mu\mu)_{\text{SM}} \times \left| 1 + \frac{\sqrt{2}\pi^2}{G_F m_W^2 V_{tb}^* V_{ts}} \frac{(\delta g_L^{32} - \delta g_R^{32})}{Y_{\text{SM}}} \right|^2, \quad (3.74)$$

where  $Y_{\text{SM}} \approx 0.957$  denotes the SM  $Y$  function.<sup>46</sup> Using the 95% CL range on the flavor-averaged branching ratio reported by the LHCb ‘evidence-paper’ on  $B_s \rightarrow \mu\mu$  [78]

$$1.1 \times 10^{-9} < \bar{\mathcal{B}}_{\text{exp}} < 6.4 \times 10^{-9}, \quad (3.75)$$

and the central value of the SM prediction in eq. (3.1) (at this level of accuracy the theoretical error is negligible), ref. [5] obtains the following bounds on  $\delta g_L$  and  $\delta g_R$ :

$$\begin{aligned} |\delta g_L|_{B_s \rightarrow \mu\mu}^{\text{MFV,PC}} &< 2.3 \times 10^{-3}, & |\delta g_R|_{B_s \rightarrow \mu\mu}^{\text{MFV}} &< 1.0 \times 10^{-1}, \\ & & |\delta g_R|_{B_s \rightarrow \mu\mu}^{\text{PC}} &< 1.6 \times 10^{-4}. \end{aligned} \quad (3.76)$$

These bounds have been obtained considering the effects of the two couplings *separately* (i.e. barring the possibility of cancellations between  $\delta g_L$  and  $\delta g_R$ , on which we will comment at the end of this section) and ignoring the fine-tuned configuration where the non-standard amplitude is about twice, and opposite in sign, compared to the SM one (a possibility that is highly disfavored by the  $Z \rightarrow b\bar{b}$  constraints [122]).

These bounds are also depicted in fig. 12 as horizontal or vertical bands delimited by solid lines. From the figure it is evident that, even with its large error, the recent evidence for  $\mathcal{B}(B_s \rightarrow \mu\mu)$  provides a constraint on  $|\delta g_L|$  – under either of the MFV or PC hypotheses – more stringent than the one obtained from the  $Z \rightarrow b\bar{b}$  observables. Furthermore, the constraint on the  $|\delta g_R|$  coupling within PC is stronger than the one obtained from  $Z \rightarrow b\bar{b}$  by more than two orders of magnitude. This circumstance is well represented by the right panel of fig. 12, where the thickness of the  $\mathcal{B}(B_s \rightarrow \mu\mu)$ -allowed band (vertical blue ‘line’) is not resolved at the scale of the electroweak-fit ellipse. This implies that, within anarchic PC models, the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  bound forbids any significant contribution to  $Z \rightarrow b\bar{b}$  observables able to decrease the existing tension between data and theoretical predictions.

As far as the bounds on the effective scale of new physics are concerned, in both frameworks the constraints derived from the  $|\delta g_L|$  bound in eq. (3.76) are largely dominant. They can be summarized as follows:

$$\Lambda > 2.6 \text{ TeV} \quad [ \text{MFV} (\delta g_L) ], \quad m_\rho > (g_\rho \epsilon_3^q) \times 2.6 \text{ TeV} \quad [ \text{PC} (\delta g_L) ], \quad (3.77)$$

the equality of the numerical coefficient in the two cases being an accident due to the approximate relation  $m_t |V_{tb}| \approx v/\sqrt{2}$ . It is also worth mentioning the  $m_\rho$  bound implied by  $|\delta g_R|$  in PC,

$$m_\rho > \frac{0.23 \text{ TeV}}{\epsilon_3^q} \quad [ \text{PC} (\delta g_R) ], \quad (3.78)$$

<sup>46</sup> A similar expression holds for  $\mathcal{B}(\bar{B}_s \rightarrow \mu^+ \mu^-)$ , with the replacement  $(\delta g_L^{32} - \delta g_R^{32})/V_{tb}^* V_{ts} \rightarrow (\delta g_L^{23} - \delta g_R^{23})/V_{tb} V_{ts}^*$ . Once  $\delta g_{L,R}^{23,32}$  are expressed in terms of  $\delta g_{L,R}$ , the  $\mathcal{B}(\bar{B}_s \rightarrow \mu^+ \mu^-)$  and  $\mathcal{B}(B_s \rightarrow \mu^+ \mu^-)$  expressions are identical both in the MFV and in the PC parameterization, and can be directly compared with the flavor-averaged branching ratio reported by LHCb [81] or CMS [80].

that becomes relevant in the limit  $\epsilon_3^q \ll 1$ , in which the bound from  $|\delta g_L|$  gets weaker.

\*\*\*

While the bounds in eq. (3.76) are *per se* interesting, the present experimental error on  $\mathcal{B}(B_s \rightarrow \mu\mu)$  does not do full justice to the sensitivity of this observable to possible modified  $Z$ -boson couplings. Therefore, ref. [5] also considered the case of a  $\mathcal{B}(B_s \rightarrow \mu\mu)$  measurement with central value as in eq. (3.1) and error of  $\pm 0.3 \times 10^{-9}$ , that can be considered a realistic estimate of the experimental sensitivity to this observable around 2018. This statement takes into account the LHCb projections from ref. [77], and the fact that CMS will likely produce a measurement with similar accuracy. We also assume a still subleading theoretical error, as expected from the steady progress in the lattice determination of the  $B_s$  decay constant [96, 112–114]. With these assumptions on the projected total error on  $\mathcal{B}(B_s \rightarrow \mu\mu)$ , the 95% CL bounds on  $\delta g_{L,R}$  become

$$|\delta g_L|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{MFV,PC}} < 4.6 \times 10^{-4}, \quad |\delta g_R|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{MFV}} < 2.0 \times 10^{-2},$$

$$|\delta g_R|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{PC}} < 3.3 \times 10^{-5}, \quad (3.79)$$

and the bounds in eqs. (3.77) and (3.78) improve by a factor of about two. The comparison between eq. (3.79) and eq. (3.73) illustrates the potential of uncovering even tiny new-physics deviations in the  $Z$ -boson couplings to down-type quarks via  $\mathcal{B}(B_s \rightarrow \mu\mu)$ . Note that, in the pessimistic case where no deviations from the SM prediction are observed in  $\mathcal{B}(B_s \rightarrow \mu\mu)$ , even the bound on  $\delta g_R$  within MFV will become more stringent compared to the one obtained from the  $Z \rightarrow b\bar{b}$  observables.

Finally, as anticipated, the bounds in eq. (3.76) and eq. (3.79) do not take into account the possibility of cancellations in the case where *both*  $\delta g_L$  and  $\delta g_R$  are switched on simultaneously. In practice, admitting such possibility does not lead to any significant changes in the plots of fig. 12. As expected from the hierarchical nature of the bounds in eqs. (3.76) or (3.79), the allowed region in the case of simultaneously non-zero  $\delta g_L$  and  $\delta g_R$  is dominated by the region allowed by the strongest constraint, namely  $\delta g_L$  in the case of MFV and  $\delta g_R$  in the case of PC.

**Projected  
bounds**

## 4 Outlook

In sec. 1.4, and especially in the ‘dedicated’ sec. 1.6 we have provided numerous arguments in support of a flavor-physics experimental and theoretical program. The bottom line was the identification of rare decays and CP violation as two promising sets of observables, wherein to look for *however small, but clear-cut* deviations with respect to the SM pattern of predictions. ‘However small’ emphasizes that, in the absence of firm indications on the scale – hence on the size – of beyond-SM effects, we should be prepared to deviations as small as 1% or even less. Therefore, emphasis should be put on improving accuracies as much as possible, because this will make the scales probed by flavor-physics observables higher and higher. In turn, ‘clear-cut’ is intended as unambiguously separable from poorly calculable SM dynamics, notably non-perturbative QCD. This implies that, before any beyond-SM speculation, a substantial part of the theoretical program should be devoted to improving our knowledge of the SM ‘background’.

The previous chapters 2 and 3 provide two rather neat examples of the two directions indicated above as promising: CP violation in chapter 2 and rare decays in chapter 3. Furthermore, in the just-advocated spirit of giving theoretical priority to the SM ‘background’, chapters 2 and 3 provide examples of progress towards a better understanding of the SM prediction in the two mentioned directions. As such, these two examples are well suited for an extrapolation to the future, as is the aim of the present outlook chapter.

### 4.1 Paths towards improving the accuracy of the $\epsilon_K$ prediction

In chapter 2 we have emphasized how progress on certain inputs to  $\epsilon_K$  – most notably on the bag parameter  $\hat{B}_K$  and on long-distance corrections – have permitted to unveil a tension with another SM test of CP violation, namely  $\sin 2\beta$ . Data have meanwhile evolved so as to make the tension practically disappear, as discussed in sec. 2.6. However, the  $\epsilon_K - \sin 2\beta$  test – that can be visualized as the intercept between the  $\epsilon_K$  hyperbola and the  $\sin 2\beta$  ray within ‘Unitarity-Triangle’ plots in the  $\bar{\rho} - \bar{\eta}$  plane – remains one of the most excruciating tests of CP violation, and non-SM effects may well be behind the corner. In fact, it should be kept in mind that, as discussed in sec. 1.5, CP violation in  $K^0 - \bar{K}^0$  mixing is the most powerful among the flavor-physics constraints on beyond-SM effects: by virtue of its CKM suppression, it allows to probe, with present accuracies, scales as high as 10000 TeV for generic new-physics couplings, as eq. (1.9) shows.

In order to make progress on the possibility of an  $\epsilon_K - \sin 2\beta$  tension hidden by present accuracies, it is crucial to improve in particular the uncertainties on  $\hat{B}_K$ , on the CKM matrix element  $|V_{cb}|$ , and on the CKM angle  $\gamma$ . We will next shortly review the prospects for improvements on these three quantities.

#### The bag parameter $\hat{B}_K$

The  $\epsilon_K - \sin 2\beta$  tension spurred more accurate lattice-QCD calculations of the  $\Delta S = 2$  matrix element, or of the bag parameter  $\hat{B}_K$ , in terms of which this matrix element is usually parameterized. Here we would like to discuss the  $\hat{B}_K$  uncertainty and its prospects of improvement. It is fair to say that this second issue is rather difficult to predict, especially because the lattice-QCD approach involves some intrinsic challenges, most of which have been with us since lattice-QCD first proposal [131]. To appreciate where these challenges lie it is useful to provide here a few-sentence introduction to the subject.<sup>47</sup>

---

<sup>47</sup> A clear and complete review can be found in the PDG [26]. For a well-written and concise account, see [132].

Lattice QCD is a definition of QCD in discretized, Euclidean spacetime, that allows the numerical evaluation of non-perturbative matrix elements within Feynman’s path-integral approach. Weak interactions are included through their operator product expansion, whereby one separates out short-distance and long-distance physics via the product between Wilson coefficients and local effective operators. A typical weak process has in fact the form – for concrete examples from this manuscript see eqs. (2.46)-(2.48) for  $\epsilon_K$  or eq. (3.1) for the  $B_s \rightarrow \mu\mu$  branching ratio

$$\text{observable} = \text{known factor} \times \langle \mathcal{O} \rangle , \quad (4.1)$$

where  $\mathcal{O}$  denotes an effective operator (we take here the case of a single operator in order not to uselessly clutter notation), whereas the ‘known factor’ represents the product of the Wilson coefficient, including in particular its CKM coupling, and of the phase-space factor, which is also specified once the process is.  $\langle \mathcal{O} \rangle$  denotes the matrix element between  $\mathcal{O}$  and the external states of the process of interest, and this is the object to calculate on the lattice, using its path-integral definition

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int d\mathcal{A} d\Psi_{\text{sea}} d\bar{\Psi}_{\text{sea}} e^{-S_{\text{QCD}}(\mathcal{A}, \Psi_{\text{sea}}, \bar{\Psi}_{\text{sea}})} \mathcal{O}(\mathcal{A}, \Psi_{\text{val}}, \bar{\Psi}_{\text{val}}) . \quad (4.2)$$

Here  $S_{\text{QCD}}$  indicates the QCD action in Euclidean metric, and  $\Psi_{\text{sea}}$  and  $\Psi_{\text{val}}$  represent the sea and respectively the valence quark fields.  $Z$  is the partition function, and serves to properly normalize the matrix element: it is defined such that  $\langle 1 \rangle = 1$ . Lattice calculations reduce the continuously infinite-dimensional integral in eq. (4.2) into a multi-dimensional integral, where the dimension equals  $V \times N_c \times N_s \times N_f$ . Here  $V$  is the number of simulated spacetime-lattice points,  $N_c = 3$  the QCD colors,  $N_s = 4$  the d.o.f. for each spinor field, and  $N_f$  the number of flavors. Considering that a typical figure for  $V$  is  $32^3 \times 64$  (the last number being the number of points in the time direction), it is clear that the integral is still extremely large, and needs to be evaluated through Monte Carlo importance sampling.

It should be intuitively clear at this point that connecting the lattice evaluation of a given matrix element to the actual physical one requires a number of extrapolations, each one leading to an associated uncertainty. The most obvious such extrapolations are the one from the finite lattice spacing to the continuum limit of spacetime, and the extrapolation from finite to infinite volume. Furthermore, being computationally expensive to simulate  $u$  and  $d$  quarks at their physical masses, simulations are typically carried out with heavier pions, so that extrapolations to their physical mass values are also necessary, implying a chiral-extrapolation error.<sup>48</sup> Moreover, the regularization and renormalization schemes most natural for lattice quantities are those that use explicitly the lattice spacing as ultraviolet regulator. In this case, a scheme matching must be performed to express these quantities in a continuum scheme such as  $\overline{\text{MS}}$ . Since continuum schemes are perturbative, whereas (often) lattice ones are not, scheme conversion usually amounts to a further systematic uncertainty, associated in particular with the truncation of the perturbative series.

The above considerations are meant to give a perception of the challenges intrinsic in the lattice approach to QCD,<sup>49</sup> and to bring notice to the fact that lattice-QCD errors are not just the statistical errors associated with the Monte Carlo sampling; they are actually dominated by systematic components, and as such should be taken *cum grano salis*.

This being said, *lattice QCD is today a mature field*, thanks to a 30-year long progress

<sup>48</sup> In fact, more and more frequent are simulations performed directly at the physical light-quark masses, thus avoiding the chiral-extrapolation error.

<sup>49</sup> There are additional challenges/limitations related to the analytic properties of Euclidean Green’s functions, as opposed to the physical ones in Minkowski metric. One of them has been mentioned below eq. (2.13), and is known as the Maiani-Testa no-go theorem [46].

in algorithms, computational resources and remarkable conceptual developments in each of the issues mentioned in the above paragraphs. That the field is mature and ready to meet the challenges of precision physics can be testified by two arguments. First, while it was not mentioned at all above, in lattice-QCD simulations there used to be one more source of uncertainty, due to the so-called quenched approximation. The latter consisted in taking as constant the fermion determinant present in the integration measure in eq. (4.2) when fermionic variables are integrated over – this approximation is thus equivalent to taking quarks as ‘frozen’, non-dynamical degrees of freedom. The quenched approximation was common practice in the past, due to the then-enormous computational cost of inverting the fermion determinant. This problem has by and large been completely bypassed today, and lattice simulations are nowadays performed with fully dynamical, ‘unquenched’ quarks. It should be noted that quenching is not a systematic approximation: quenched QCD is simply another theory (not even unitary) than QCD. The corresponding error can thus be likened to the error associated with using a hadron model, rather than QCD, to compute a certain quantity. The removal of the quenching error is one major step towards making lattice QCD a mature field.

A second argument is the fact that, by now, numerous quantities have been calculated within completely different approaches – for example, different discretized actions, implying a different procedure for most of the extrapolations mentioned above. Remarkable is the fact that all these different approaches have led nonetheless to consistent results, thus confirming universality within the accuracy reached. This offers circumstantial evidence that systematic errors are indeed under control, and has spurred the creation of a lattice averaging group at the end of 2007, FLAG [117], with a spirit akin to that of the PDG or HFAG working groups, namely with the task of collecting and systematizing the existing lattice results and of providing informed averages for quantities useful in phenomenology.

\*\*\*

Coming back to the bag parameter  $\hat{B}_K$ , it should be emphasized that it represents one of the benchmark quantities among those calculated on the lattice, so that it may be used as a yardstick for lattice-QCD progress as a whole. As a matter of fact, the latest average published by the FLAG collaboration reads [117]

$$N_f = 2 + 1 : \quad \hat{B}_K = 0.7661(99) , \quad (4.3)$$

whereby  $N_f = 2 + 1$  indicates that the lattice results included in the average are only those with three dynamical light quarks, and such that  $m_u = m_d \neq m_s$ . Especially noteworthy is the impressive 1.3% error on the  $\hat{B}_K$  average in eq. (4.3), that in turn constitutes a significant improvement over the already remarkable 2.7% uncertainty quoted in the previous FLAG average of end 2010 [133]:  $\hat{B}_K = 0.738(20)$ . The substantial reduction, by a factor of two, of the  $\hat{B}_K$  uncertainty between the two averages, and the fact that the two results are entirely consistent with each other, conveys a quantitative idea of the progress made. The latter is due in particular to improvements in the chiral extrapolation and in the perturbative-truncation error, both of which have been mentioned above.

As far as the  $\epsilon_K$  prediction is concerned, recalling the discussion in sec. 2.6 we can state that the error component due to  $\hat{B}_K$  is now definitely subleading, the dominant one being that from  $|V_{cb}|$ .

### The CKM matrix element $|V_{cb}|$

As previously remarked,  $|V_{cb}|$  enters  $\epsilon_K$  to the fourth power, and this is the reason why it represents the main component to the  $\epsilon_K$  uncertainty. As a consequence, a firm  $|V_{cb}|$

prediction is arguably the most important ingredient towards a final word on the  $\epsilon_K - \sin 2\beta$  tension. Let alone  $\epsilon_K$ , the ratio  $|V_{ub}|/|V_{cb}|$  represents more generally a crucial ingredient in the overall knowledge of the unitarity triangle depicted in fig. 4. In fact,  $|V_{ub}|/|V_{cb}|$  measures the side denoted as  $R_b$ , which is adjacent to the angle  $\gamma$ . Since both  $\gamma$  and  $R_b$  can be obtained from decays expected to be free from beyond-SM contaminations, accurately measuring  $\gamma$  and  $R_b$  can be translated into a determination of the unitarity triangle which is accordingly new-physics free (see e.g. [36, 53]), thereby providing a ‘standard-candle’ for beyond-SM searches.

$|V_{ub}|$  and  $|V_{cb}|$  can be extracted from semileptonic  $b \rightarrow u\ell\bar{\nu}$  and  $b \rightarrow c\ell\bar{\nu}$  transitions, respectively, with  $\ell$  denoting an electron or a muon. These transitions are dominated by a tree-level  $W$ -exchange topology, where namely a  $W$  is responsible for the  $b \rightarrow c$  or  $u$  flavor-changing charged current, and it subsequently decays into an  $\ell\bar{\nu}$  pair. The only known models able to pollute this SM topology are those with extended Higgs sectors, whereby charged-Higgs scalars may give rise to additional tree-level topologies similar to the one with the  $W$ . However, in the cases of an  $e$  or  $\mu$  in the final state, these amplitudes are highly suppressed by small Yukawa couplings. (They are on the other hand relevant for semileptonic decays to a  $\tau$ , which present their own challenges and will not be covered here.) In short, a  $|V_{ub}|/|V_{cb}|$  determination from  $b \rightarrow \{u, c\}\ell\bar{\nu}$  can be assumed to be new-physics free.

$B$  factories are especially suited for this kind of decays, because, out of the produced  $B\bar{B}$  pair, it is possible to fully reconstruct one of the  $B$  mesons and hence study the semileptonic decay of the recoiling  $B$  with high purity. One can either consider ‘exclusive’ decays, namely decays to specific final states, or else measure ‘inclusively’ the decay rate of a  $B$  to a charmed (in the  $|V_{cb}|$  case) or uncharmed (in the  $|V_{ub}|$  one) hadron plus an  $\ell\nu$  pair. We will henceforth focus on  $|V_{cb}|$ .

Exclusive and inclusive  $|V_{cb}|$  determinations are based on rather different theoretical assumptions, and the consistency among the two classes of measurements offers an important cross-check. As we will see below, the consistency between exclusive and inclusive  $|V_{cb}|$  determinations is ‘marginal’ [26]. In order to convey an impression of the progress to be expected, I will next very shortly review – as a non-expert – either of the two classes of measurements.

In exclusive modes,  $|V_{cb}|$  is determined from  $\bar{B} \rightarrow D^{(*)}$  decays. Actually, because of its smaller rate, the decay to an unstarred  $D$  is experimentally more difficult, and comes with a larger uncertainty. As such, it plays a marginal role in the  $|V_{cb}|_{\text{excl}}$  average, and will not be considered any further in the discussion. Focusing then on the decay to a starred  $D$ , the bulk of the theoretical uncertainty in the exclusive determination of  $|V_{cb}|$  arises from the knowledge of the  $B \rightarrow D^*$  transition amplitude. The starting point is the observation that, in the heavy-quark limit for both the  $b$  and  $c$  quarks, new ‘heavy-quark symmetries’ (HQS) appear in the low-energy effective Lagrangian for QCD, implying one single form factor for  $b \rightarrow c\ell\bar{\nu}$  transitions, known as Isgur-Wise function  $\mathcal{F}(w)$  [134]. The latter depends on the product  $w \equiv v \cdot v'$  of the four-velocities  $v$  and  $v'$  of the initial- and final-state hadron, respectively. In the HQS limit, the function normalization can be calculated at the kinematic point  $w = 1$ , corresponding to the maximum momentum transfer to the leptonic current.

Therefore, the strategy to determine  $|V_{cb}|$  from exclusive decays consists in extracting from experiment the product  $\mathcal{F}(1)|V_{cb}|$ ,<sup>50</sup> and in estimating  $\mathcal{F}(1)$  from theory. Of course, the accuracy attainable in the strict HQS limit for  $\mathcal{F}(1)$  is not better than  $\mathcal{O}(\Lambda_{\text{QCD}}/m_c)$ . Besides, reaching the point  $w = 1$  typically requires an extrapolation. Therefore, the main steps towards an accurate  $|V_{cb}|_{\text{excl}}$  determination include calculating corrections to  $\mathcal{F}(w)$  due to the external-quarks’ finite masses, and finding a well-motivated (and possibly economical)

$|V_{cb}|_{\text{excl}}$

<sup>50</sup> In fact, what one extracts is the product  $\eta\mathcal{F}(1)|V_{cb}|$ , where the  $\eta$  factor takes into account short-distance, perturbatively calculable corrections from electroweak [135] and QCD [136] interactions.

functional form for the extrapolation to  $w = 1$ .

Away from the infinite-mass (to be referred to also as ‘static’) limit there are more than one form factor, some of which vanish in the static limit, and others tend to the Isgur-Wise function in this limit. Because for the latter the normalization is known for  $w = 1$ , the static-limit prediction is still valid to first order in the breaking parameter  $\Lambda_{\text{QCD}}/m_c$ , namely deviations from the static limit arise only at order  $\Lambda_{\text{QCD}}^2/m_c^2$ . This circumstance is known as Luke’s theorem [137]. On the other hand, in the case of form factors that vanish in the static limit, corrections are not protected, and arise already at linear order in the HQS-breaking parameter. Therefore, in general, the corrections to  $\mathcal{F}(w)$  due to the external-quarks’ finite masses are of order  $\Lambda_{\text{QCD}}/m_c$ .

The  $B \rightarrow D^{(*)}\ell\nu$  form factors can be calculated from first principles using lattice QCD. In this approach heavy-quark symmetries are *built in* even at finite lattice spacing. This implies that also lattice-QCD uncertainties will scale as  $\Lambda_{\text{QCD}}/m_c$  or  $\Lambda_{\text{QCD}}^2/m_c^2$ , depending on the form factor vanishing or not in the static limit. Typical lattice calculations quote overall errors in the 2% ballpark, due mostly to the chiral extrapolation and to the finite lattice spacing. In order to obtain the preferred  $|V_{cb}|_{\text{excl}}$  value, the FLAG collaboration starts from the experimentally-extracted product

$$\eta_{\text{EW}} \mathcal{F}^{B \rightarrow D^*}(1) |V_{cb}| = 35.90(45) , \quad (4.4)$$

quoted from HFAG [138]. The  $\mathcal{F}^{B \rightarrow D^*}(1)$  lattice average from  $N_f = 2 + 1$  simulations [139] is [117]

$$\mathcal{F}^{B \rightarrow D^*}(1) = 0.906(4)(12) , \quad (4.5)$$

where the first and second number in parentheses represent the statistical and systematic uncertainties, respectively. To be noted is the 1.3% error of this average. Using this result, the  $|V_{cb}|_{\text{excl}}$  preferred value is found to be [117]

$$|V_{cb}|_{\text{excl}} = 39.36(56)(50) \times 10^{-3} , \quad [N_f = 2 + 1 \text{ lattice QCD}, \quad B \rightarrow D^*\ell\nu] . \quad (4.6)$$

where the errors come from the lattice calculation and experiment plus non-lattice theory, respectively. It should be noted that the two errors are comparable. According to the latest FLAG review, the exclusive determination of  $|V_{cb}|$  will improve significantly over the next year or two because of new lattice-QCD calculations of the  $B \rightarrow D^{(*)}\ell\nu$  form factors at nonzero recoil [117]. Hence it will be important to increase experimental accuracies accordingly, which should also be feasible after the BelleII start-up.

The form factor  $\mathcal{F}(1)$  can also be estimated with methods based on QCD sum rules [140,141]. In this approach, one defines a suitable scattering amplitude, that can be related, using its analytic properties, to the form factor of interest. The relevant amplitude is the Fourier transform of the two-point function of the  $b \rightarrow c$  axial current, calculated between external  $B$  mesons. The Fourier transform is used to enforce zero momentum transfer, while leaving the mass difference between the initial and final state free. The Fourier-transformed amplitude is thereby function of the parameter  $\varepsilon \equiv M_B - M_{D^*}$ , and can be calculated with an operator-product expansion in  $1/m_Q$ , with  $m_Q$  to be identified with either of  $m_b$  and  $m_c$ . Its contour integral along a circle with fixed  $|\varepsilon|$  has an elastic component which is the square of the form factor of interest. This component can be isolated provided one can estimate the other contributions, coming from excited states. For the form factor, the PDG quotes the result  $\mathcal{F}^{B \rightarrow D^*}(1) = 0.86(1)(2)$  from refs. [141], where the first uncertainty arises from the perturbative part of the calculation, and the second, dominant one from the excited-state contributions. The  $|V_{cb}|_{\text{excl}}$  prediction resulting from this approach reads [26]

$$|V_{cb}|_{\text{excl}} = (41.4 \pm 0.5_{\text{exp}} \pm 1.0_{\text{th}}) \times 10^{-3} , \quad [\text{QCDSR}, \quad B \rightarrow D^*\ell\nu] . \quad (4.7)$$

As clearly stated in [141], the main limiting factor to improvements in the above result is the fact that the OPE involves a charm-mass expansion. While one may conceive strategies – based on direct measurements or on lattice-QCD evaluations – to extract the associated non-perturbative input, it is difficult to predict the progress in this error component. Nonetheless, one can see that the result in eq. (4.7) is compatible with the lattice one in eq. (4.6), and is, for the moment being, of comparable accuracy.

\*\*\*

The most accurate determination of  $|V_{cb}|$  comes actually from inclusive decays, to which we turn our attention next. The main strategy in this instance is based on the possibility to accurately calculate the total decay rate for  $\bar{B} \rightarrow X_c \ell \bar{\nu}$ , with  $X_c$  any charmed hadron. The calculation depends on non-perturbative parameters that can be estimated from the mass spectra associated with the total decay rate. The actual possibility of a reliable calculation rests on the validity of the OPE in powers of  $1/m_b$ , assumed for the relevant matrix element. This validity has actually not been proven in the kinematic region relevant to heavy-quark decays, and it is equivalent to the assumption of ‘quark-hadron duality’ (see [142]). Violation of this duality would result in terms not described by the  $1/m_b$  expansion. The fact that fits to data are consistent with this expansion provides circumstantial evidence for the validity of this assumption.

$|V_{cb}|_{\text{incl}}$

The total inclusive decay rate is thus the sum of  $1/m_b^n$  terms, each order  $n$  being in turn an expansion in the strong coupling  $\alpha_s(\mu)$  times calculable functions of the ratio  $m_c/m_b$ . Furthermore, each of the different terms in the  $1/m_b$  expansion involves non-perturbative parameters, that depend on forward matrix elements between external  $B$  states, akin to the one discussed above in the context of the exclusive- $|V_{cb}|$  determination. While the OPE for the inclusive decay rates has the major advantage of being an expansion in  $1/m_b$  only, effects suppressed by  $1/m_c$  are introduced by the expectation values of these matrix elements [143].

Let alone the estimate of the  $1/m_c$  contributions, the main mass dependence in the OPE relevant for inclusive semileptonic  $B$  decays is thus the  $m_b$  dependence. Crucial is then an accurate  $m_b$  choice since, for example, the  $m_b$  scheme affects the convergence of the series of QCD corrections entering at the different orders in  $1/m_b$ . As a matter of fact, taking for example  $m_b$  to be the pole mass results in a very poor convergence of the perturbative series, whereas ‘short-distance’ schemes close to the  $\overline{\text{MS}}$  one display a better convergence behavior.<sup>51</sup> Several such schemes have been proposed, like the ‘kinetic’ scheme or the ‘1S’ scheme, whose precise definition is irrelevant to the present discussion.

These qualifications being made, the most consistent way to compare data and theory usually proceeds through global fits. In fact, experiments measure the total rate for  $\bar{B} \rightarrow X_c \ell \nu$ , as well as its moments as a function of the minimum lepton momentum, the electron energy, and the squared hadron mass spectrum [26]. All these measurements are strongly correlated, and for their treatment to be consistent their correlation matrix needs to be taken into account, which in turn calls for a global-fit approach. At the theoretical level, this approach allows to extract  $|V_{cb}|$ , the quark masses and the non-perturbative parameters of the heavy-quark expansion. A recent comprehensive analysis [118] within the kinetic scheme, quoted as reference by the PDG, yields

$$|V_{cb}|_{\text{incl}} = (42.42 \pm 0.86) \times 10^{-3} . \quad [m_b^{\text{kin}} \text{ scheme: ref. [118]]} \quad (4.8)$$

To be noted is the 2% relative uncertainty, comparable to the one of the  $|V_{cb}|_{\text{excl}}$  average. An alternative fit within the 1S scheme can be found in ref. [144] and yields compatible results.

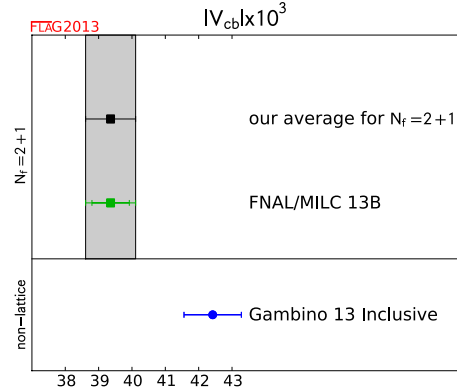
<sup>51</sup> A related issue is of course the assumption on the charm mass. In the fits to be discussed next,  $m_c$  is typically constrained from data on  $\bar{B} \rightarrow X_s \gamma$  or direct charm-mass determinations.



To conclude, while the main limitation of the inclusive approach to  $|V_{cb}|$  is the lack of knowledge of higher-order perturbative and non-perturbative contributions in the OPE expansion, its major advantage is its ‘scalability’. In fact, on the theory side, one can in particular include higher orders in perturbative QCD corrections; on the experimental side, one can pursue sensitivity to higher moments of the relevant spectra, which in turn allows global fits to gain sensitivity to higher orders in the  $1/m_b$  expansion, or to unveil duality violations. Both these routes represent clear paths towards improving accuracy and cross-checking systematics.

\*\*\*

The present status of exclusive and inclusive  $|V_{cb}|$  determinations is best summarized by fig. 13, taken from ref. [117]. This figure clearly shows the level of consistency between the two sets of determinations, already mentioned to be marginal [26]. With theory progress, and new measurements from BelleII, we can expect substantial progress to take place on this issue. However, it is difficult to make projections about the ‘final’  $|V_{cb}|$  error. In fact, one should keep in mind that this error has a substantial systematic component, on which progress is notoriously difficult to forecast. As a matter of fact, and as already remarked, over the last few years the quoted uncertainty on  $|V_{cb}|$  has increased, albeit only slightly (see footnote 29).



**Figure 13:**  $|V_{cb}|$  status. Figure taken from ref. [117].

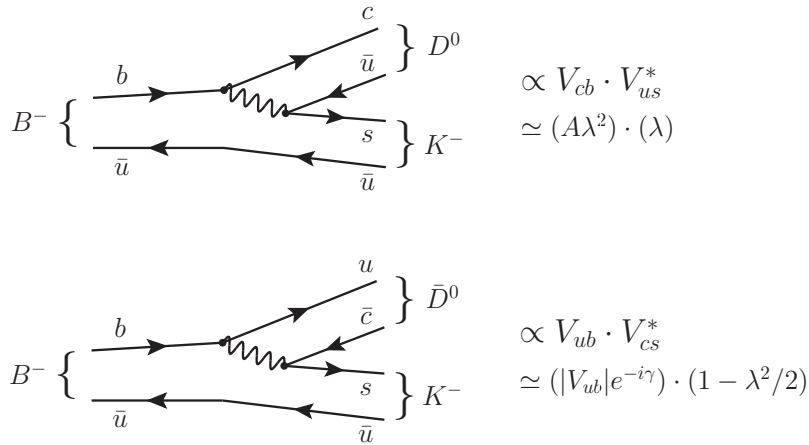
### The CKM angle $\gamma$

By inspection of unitarity-triangle fits, one can see that the angle  $\gamma$  remains one of the least well known parameters of the CKM matrix. Yet,  $\gamma$  can be determined in a self-contained way through decays of the type  $B \rightarrow DK$ , that involve only tree-level amplitudes and as such are expected to undergo little to no pollution at all from non-standard effects. This circumstance is a unique property of  $\gamma$  amongst all CP-violating parameters, and makes an accurate, model-independent  $\gamma$  measurement a ‘standard candle’ in flavor physics. As a matter of fact, the determination from tree-level  $B \rightarrow DK$  decays has essentially negligible theoretical uncertainty, at the level of  $10^{-7}$  [145], as we will detail below.

An accurate and model-independent  $\gamma$  measurement will bring more than just accuracy – it will boost sensitivity to beyond-SM physics in several key quantities. For a conspicuous example we just need to go back to sec. 2.6, in particular eq. (2.82). Here we noted that the CKM side  $R_t$  contributes, after  $|V_{cb}|$ , the most important component to the  $\epsilon_K$  uncertainty, and we remarked that the  $R_t$  input comes with a sizable model dependence, as  $R_t$  is determined from  $\Delta m_d/\Delta m_s$  (see eq. (2.80)), a quantity very sensitive to new physics. Since  $R_t$  is opposite to  $\gamma$ , one sees that the  $R_t$  error and model dependence in  $|\epsilon_K|$  can be bypassed in one stroke by an accurate  $\gamma$  determination.

The measurement of  $\gamma$  from  $B \rightarrow DK$  decays exploits direct CP violation in these decays. As originally pointed out in refs. [146],  $B \rightarrow DK$  decays consist of two interfering tree-level amplitudes,  $b \rightarrow c\bar{u}s$  and  $b \rightarrow u\bar{c}s$ , whose relative weak phase is exactly  $\gamma$ , see fig. 14. These transitions mediate respectively  $B^- \rightarrow D^0(\rightarrow f)K^-$  and  $B^- \rightarrow \bar{D}^0(\rightarrow f)K^-$  decays. Note

**$\gamma$  from  
 $B \rightarrow$   
 $DK$   
decays**



**Figure 14:**  $B^- \rightarrow DK^-$  decays and sensitivity to  $\gamma$ . The  $V_{ub}$  magnitude is  $A\lambda^3\sqrt{\bar{\rho}^2 + \bar{\eta}^2}$ , hence at the level of CKM couplings the two diagrams are competitive.

that  $D^0$  and  $\bar{D}^0$  are assumed to decay into a common final state  $f$ , which makes interference possible.

Besides  $\gamma$ , the mentioned decays depend on the ratio between the  $b \rightarrow u$  and the  $b \rightarrow c$  amplitude magnitudes, on the corresponding ratio involving the  $D \rightarrow \bar{f}$  and the  $D \rightarrow f$  amplitudes, as well as on the two associated strong-phase differences. Once  $f$  is fixed to be a certain ‘class’ of decays (see below), these parameters can be extracted directly from data. (A possible theoretical evaluation of these parameters is in general challenging, albeit proposals for lattice-QCD evaluations have been put forward lately, that may well be viable [147].)

Along the above lines, several by now established methods have been proposed to measure  $\gamma$  in tree decays of the kind  $B^- \rightarrow D^{(*)}K^{(*)-}$ . These methods differ by the  $D^{(*)}$  final state  $f$  that is being used. Restricting to the decay categories so far accessed by experiments, the considered final states include:

- CP eigenstates. This is the original GLW method [146]. CP eigenstates are advantageous as the  $D \rightarrow f$  and  $D \rightarrow \bar{f}$  decay amplitudes are the same and their relative strong phase obviously zero. Therefore, beside  $\gamma$ , observables depend on only one amplitude ratio and one strong phase.
- A combination of Cabibbo-favored (CF) and doubly Cabibbo-suppressed (DCS) decays. This is the so-called ADS method [148]. One considers in turn the color-allowed  $B^- \rightarrow D^0K^-$  decay followed by DCS  $D^0$  decays, and the color-suppressed  $B^- \rightarrow \bar{D}^0K^-$  followed by CF  $\bar{D}^0$  decays. Thus in this method the choice of DCS final states serves to compensate the color suppression of the parent- $B$  decay, i.e. to maximize the interference effect. The ADS decay chain has been observed for the first time at LHCb, see [149].
- Three-body, self-conjugate final states, in particular  $D \rightarrow K_s^0\pi^+\pi^-$  and  $D \rightarrow K_s^0K^+K^-$ . This is the so-called GGSZ or “Dalitz” method [150] (see also [151]). The choice of this class of final states – more complicated than in the previously mentioned methods, hence necessitating a Dalitz analysis – is compensated by various advantages, including: the presence of resonances, implying large strong phases; the possibility to use only Cabibbo-favored  $D$  decays; the possibility to consider final states consisting of only charged particles, that are easily reconstructible and separable from backgrounds.

These different methods share a common, rather intuitive formalism. One needs to parameterize the four amplitudes for  $B^- \rightarrow \bar{D}^0 K^-$  and  $D^0 \rightarrow \bar{f}$ , and write in terms of them the two interfering amplitudes depicted in fig. 14. Then the formulae in the different methods are, for the most part,<sup>52</sup> special cases of this basic formalism. A very useful reference in this respect is [77]. Therefore one can expect that, let alone the case of two-body decays, the various methods overlap substantially with each other. Hence the best sensitivity to  $\gamma$  arises from a combined analysis of all the applicable methods. The availability of more than one method is also beneficial for the knowledge of the involved hadronic parameters, that are thereby constrained by different theoretical arguments.

An interesting question in the present context is what is the ultimate theoretical uncertainty attainable in the  $\gamma$  determination. This question is not only important in connection with  $\gamma$  as a standard candle to other measurements, but also to answer the related question of the largest energy scale that can be probed by the  $\gamma$  determination. The irreducible theoretical uncertainty in  $\gamma$  has been recently reassessed in ref. [145]. The authors underline from the outset that most of the theoretical uncertainties associated with any of the methods mentioned above are, or will become, statistics-dominated, hence reducible. Two further sources of reducible uncertainties can be taken into account by a straightforward modification of the formulae used for the  $\gamma$  extraction. These include the effects of  $D^0 - \bar{D}^0$  and  $K^0 - \bar{K}^0$  mixing, and QED radiative corrections. Since the latter are CP conserving, they can actually be absorbed in the definition of the CP-conserving parameters fitted along with  $\gamma$ , and as such do not affect the  $\gamma$  measurement directly [145]. The authors conclude that the first source of irreducible error comes from higher-order electroweak corrections, and derive an upper bound of  $|\delta\gamma| \lesssim \mathcal{O}(10^{-7})$ . The latter is estimated via a detailed one-loop analysis, keeping only *local* contributions to electroweak corrections. The authors neglect non-local contributions, as they are more difficult to estimate. However, they argue that these contributions are unlikely to change their  $\delta\gamma$  result, quoting as supporting argument the case of  $K^0 - \bar{K}^0$  mixing, where long-distance contributions to  $K^0 - \bar{K}^0$  mixing are one order of magnitude below the short-distance ones [3].

The  $\delta\gamma$  theoretical bound from ref. [145] is far below any foreseeable experimental accuracy. However, it is interesting to derive the new-physics scale that a deviation at that level would probe. This exercise has been carried out in ref. [77]. It turns out that, assuming general flavor-violating new physics, one can probe scales as large as  $10^3$  TeV, figure that becomes  $10^2$  TeV when assuming MFV. By comparison with the  $\Delta F = 2$  bounds in eq. (1.9), one realizes that, while  $K^0 - \bar{K}^0$  mixing remains an unbeaten test of generic new physics,  $\gamma$  from  $B \rightarrow DK$  decays *would* become the most powerful probe of MFV effects beyond the SM, *if* experimental accuracies shrunk to the level of the mentioned theoretical uncertainty. By taking the new-physics scale probed as the fourth root of the experimental yield, we can easily translate the above results to the case of datasets with realistic accuracies. Assuming, with the LHCb upgrade, a projected experimental uncertainty on  $\gamma$  as low as  $0.9^\circ$ , one gets scales in the ballpark of 5 and 50 TeV in the case of MFV and respectively generic new physics [77].

The projected LHCb  $\gamma$  error of  $0.9^\circ$  is *per se* remarkable – as a matter of fact the LHCb upgrade is the only proposed experiment able to reach sub-degree precision. As concerns the Belle upgrade, a combined analysis of the GLW, ADS and Dalitz methods is expected to yield a  $1.5^\circ$  uncertainty using a data sample of  $50 \text{ ab}^{-1}$ , according to the BelleII physics-reach

<sup>52</sup> There is actually an exception, as other methods to measure  $\gamma$  exists, that do not exploit direct CP violation in  $B \rightarrow DK$  decays, like the methods mentioned so far. An example are  $\gamma$  measurements from  $B$  decays that are still based on the interference between the two amplitudes depicted in fig. 14, but are sensitive to it via  $B_d$  or  $B_s$  mixing [152].

document [153].

An interesting and direct application of such accurate  $\gamma$  measurements from  $B \rightarrow DK$  decays would be a comparison with  $\gamma$  determinations from loop-dominated processes. An example are  $B \rightarrow \pi\pi$  and  $B \rightarrow KK$  decays, that allow to extract a combination of  $\gamma$  and  $\beta$ . Consistency of the  $\gamma$  values obtained across tree- and loop-dominated modes would provide one more excruciating test of CP violation in the SM.

## 4.2 $B_s \rightarrow \mu\mu$ : outlook and possible new directions

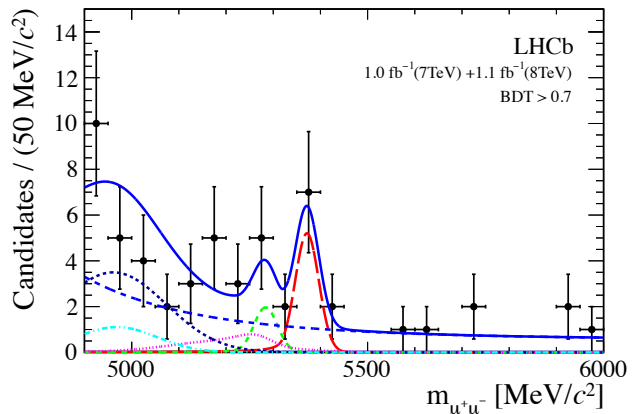
In this section we discuss the progress to be expected on the other main topic treated in this manuscript, namely the rare decay  $B_s \rightarrow \mu\mu$ .

### Outlook on $B_s \rightarrow \mu\mu$ and related observables

A first question to be addressed is that of the experimental and theoretical outlook on the total  $B_s \rightarrow \mu\mu$  decay rate. According to ref. [77], the LHCb accuracy on the time-integrated branching ratio, defined in eq. (3.29), will reach the level of  $0.5 \times 10^{-9}$  by 2018, namely by the end of run II of the LHC, and will further improve to  $0.15 \times 10^{-9}$  with  $50 \text{ fb}^{-1}$  of data collected by the LHCb upgrade [154]. ‘Extrapolating’ present performances, one can anticipate results of comparable accuracy from the CMS experiment. Therefore, it is to be expected that the experimental accuracy will be at the level of 10% by the end of LHC run II, and about 3% in the LHCb-upgrade era.

Substantial progress is also to be anticipated on the related decay  $B_d \rightarrow \mu\mu$ . Being 30 times rarer than the  $B_s \rightarrow \mu\mu$  decay, it is experimentally more challenging, especially as concerns background subtraction. To convey an idea of the challenge, it is worthwhile to note that one of the most prominent background processes is  $B_{d(s)} \rightarrow h^+ h'^-$ , with  $h^{(\prime)} = K, \pi$ , and that the latter displays a peaking structure in the very same  $m_{\mu\mu}$  region as the expected  $B_d \rightarrow \mu\mu$  signal. This fact can easily be visualized in fig. 15, taken from the LHCb ‘evidence’ paper [78]. The figure shows that this peaking background (pink dotted curve) is no less than 30% of the signal (green dashed curve) at peak. Because of this circumstance, the  $B_d \rightarrow \mu\mu$  challenge is more than just the  $\times 30$  more statistics needed with respect to the  $B_s \rightarrow \mu\mu$  counterpart. According to projections [77], by the end of run II LHCb is expected to be able to measure the ratio  $\mathcal{B}(B_d \rightarrow \mu\mu)/\mathcal{B}(B_s \rightarrow \mu\mu)$  at the level of 100% error, in case of a SM signal. This error will go down to about 30% with  $50 \text{ fb}^{-1}$  of data collected by the LHCb upgrade. In view of experimental techniques greatly outperforming projections, I am ready to bet that actual performances will substantially exceed the above forecasts.

At the theoretical level, as already remarked elsewhere in this manuscript, a measurement of  $\mathcal{B}(B_d \rightarrow \mu\mu)/\mathcal{B}(B_s \rightarrow \mu\mu)$  not in accord with the SM expectation (see end of sec. 3.7) will represent a clear-cut signal of new flavor effects beyond MFV.



**Figure 15:**  $B_s \rightarrow \mu\mu$ ,  $B_d \rightarrow \mu\mu$  signals (red dashed and green short-dashed lines, respectively) and their backgrounds. Taken from ref. [78].

Finally, theory uncertainties on  $\mathcal{B}(B_s \rightarrow \mu\mu)$  have been amply discussed in chapter 3. Specifically, secs. 3.2 to 3.5 have been concerned with all the known sources of systematic uncertainty, that is therefore expected to be under control. Besides, sec. 3.6 discussed the parametric uncertainty, stressing that it depends on the error assumed for the non-perturbative parameter  $f_{B_s}$ . In fact, the  $f_{B_s}$  determination is such that one lattice calculation by the HPQCD collaboration (second reference in [112]) comes with an error which is about half of most of the other determinations, and inevitably dominates a weighted average. Hence in ref. [4] we decided to estimate the  $f_{B_s}$  uncertainty from the spread of the lattice determinations, as is common practice in such cases [26]. With this procedure, we found for  $\mathcal{B}(B_s \rightarrow \mu\mu)$  a total 8.5% error, due mostly to  $f_{B_s}^2$  (7% at the level of the branching ratio) and to a lesser extent to the CKM combination (4%). On the other hand, in the approach of ref. [116],  $f_{B_s}$  is taken with its weighted-average error [117]. This results in a total  $\mathcal{B}(B_s \rightarrow \mu\mu)$  error of 6%, dominated by the CKM combination rather than by  $f_{B_s}$ .

While the above issue is at present still debatable to some extent – or at least it would be so in case of a discrepancy between experiment and the SM prediction – the  $f_{B_s}$  error will definitely become a negligible component of the total  $\mathcal{B}(B_s \rightarrow \mu\mu)$  uncertainty in the next few years, thanks to new lattice-QCD calculations of  $f_{B_s}$  with errors comparable to that of the present weighted average [117]. The dominant  $B_s \rightarrow \mu\mu$  uncertainty will thus come from the CKM combination, which will in turn improve from the present 4%. In short, in the foreseeable future the theory uncertainty in  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  is expected to remain at a negligible level with respect to experimental errors.

### **$B_s \rightarrow \mu\mu$ as an EW precision test: progress to be expected**

The pursuit of higher accuracy in the  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  decays is motivated by the reasonable expectation of finding, sooner or later, deviations from the SM predictions, that would signal beyond-SM effects at a scale more or less decoupled from the electroweak scale. In fact, *the more* this scale is *decoupled*, *the more* precision flavor observables such as  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  are well *motivated*, as they can probe scales far exceeding those that are reachable in direct searches of new resonances. This point has been supported by detailed arguments especially in secs. 1.4-1.6.

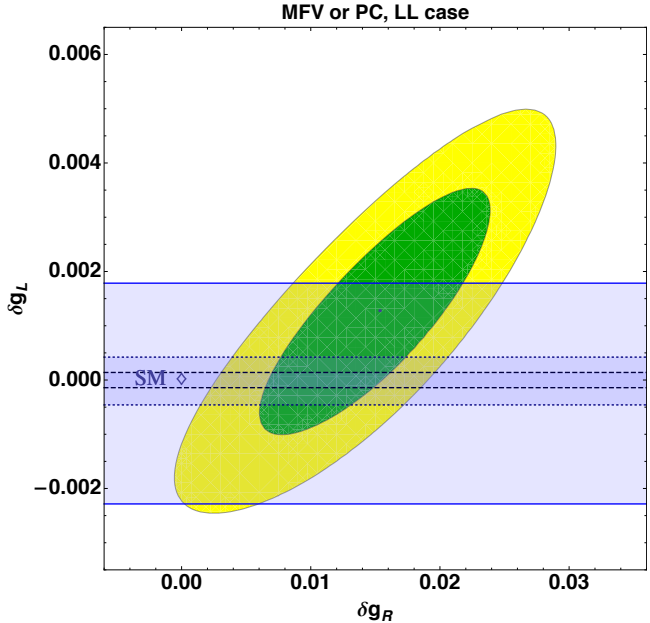
As concerns the kind of new physics that the  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  decays are suitable to probe, two prominent examples have been discussed in secs. 3.8 and 3.9. The first one is  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  as a probe of *scalar* four-fermion interactions, and thereby of extended Higgs sectors. This example is motivated by the observation that  $B_{d,s} \rightarrow \mu\mu$  is a process essentially governed by Yukawa interactions. The second example concerns the  $\mathcal{B}(B_{d,s} \rightarrow \mu\mu)$  sensitivity to anomalous  $Z$ -quark-quark couplings, feature that makes this decay competitive to, or in fact superior than, electroweak precision observables at the  $Z$  peak. This large sensitivity is motivated by the property of the  $B_s \rightarrow \mu\mu$  amplitude of being (by far) dominated by the  $Z$ -penguin diagram, the large top mass making completely negligible any other topology, notably box diagrams.

*Substantial progress* is to be expected in the years to come on both of the above directions. In this subsection and in the next one we would like to provide arguments in support of this statement.

Actually, as concerns  $\mathcal{B}(B_s \rightarrow \mu\mu)$  as an EWPT (the second of the above topics), the progress to be expected can by and large be read off from the results of ref. [5], that have been presented in sec. 3.9, and summarized by fig. 12 and eqs. (3.76) and (3.79). In particular, fig. 12 permits to visualize the relative performance of  $\mathcal{B}(B_s \rightarrow \mu\mu)$  versus  $Z$ -peak observables as constraints to anomalous right- or left-handed  $Z$ - $\bar{d}_i$ - $d_j$  couplings, reported in

the  $x$ - and respectively  $y$ -axes of the same figure. The performance of  $Z$ -peak observables is quantified by the area of the outer ellipse, whereas the constraining power of  $\mathcal{B}(B_s \rightarrow \mu\mu)$  can be appreciated by the horizontal or vertical bands. The following points can be made:

- As already detailed in sec. 3.9, a comparison between  $Z\text{-}\bar{d}_i\text{-}d_j$  deviations in  $\mathcal{B}(B_s \rightarrow \mu\mu)$  (where they are flavored) and in  $Z$ -peak observables, where they are not, can only be made once a flavor model is assumed. The results of ref. [5] hold under the fairly general and well motivated assumptions of either MFV or partial compositeness (abbreviated with ‘PC’ above the plots in fig. 12).
- Most importantly, from the width of the bands enclosed by *solid* blue lines, that correspond to the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  constraint as of the LHCb evidence paper [78], one can realize that the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  visibly outperforms  $Z$ -peak observables, already with the rather limited accuracy (of order 50% at  $1\sigma$ ) of that first LHCb measurement. This holds true for any of the scenarios displayed in fig. 12. All the corresponding numerical bounds were reported in eq. (3.76).



**Figure 16:** Same as the left panel of fig. 12, but for a  $\mathcal{B}(B_s \rightarrow \mu\mu)$  total uncertainty of 3%, corresponding to an LHCb-upgrade scenario.

- Rather striking in this respect is the case of a right-handed coupling in partial compositeness. In this instance, the width of the blue band is not even resolved at the scale of the EWPT ellipse (see vertical ‘line’ on the rightmost panel in fig. 12). Note that the latter is again obtained with the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  constraint coming from the first LHCb measurement.
- The projection to the case of  $\mathcal{B}(B_s \rightarrow \mu\mu)$  as of 2018, with namely 10% experimental error and negligible theory uncertainty, is likewise displayed in fig. 12 as bands enclosed by *dotted* blue lines. Already in the case of a left-handed anomalous coupling (leftmost panel), the  $\mathcal{B}(B_s \rightarrow \mu\mu)$  performance is visibly an order of magnitude better than that of EWPT for both of MFV or PC. The corresponding numerical bounds are reported in eq. (3.79), whence one can see that a 10%  $\mathcal{B}(B_s \rightarrow \mu\mu)$  measurement will probe anomalous  $Z\text{-}\bar{d}_i\text{-}d_j$  couplings at the level of one part in  $10^5$ .
- Finally, we can easily extrapolate the projections in the previous item to an LHCb-upgrade scenario, namely to a  $\mathcal{B}(B_s \rightarrow \mu\mu)$  expected accuracy of about 3% (again dominated by the experimental component). Results in the left-handed case, valid for either of MFV or PC, are shown in fig. 16 as a long-dashed blue band (the innermost band). The corresponding plot for the right-handed PC case is not shown, because the obtained band is, like the ‘band’ on the rightmost plot of fig. 12, resolved as a line.

Numerically, the bounds expected in the era of the LHCb upgrade read

$$|\delta g_L|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{MFV,PC}} < 1.4 \times 10^{-4}, \quad |\delta g_R|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{MFV}} < 6.3 \times 10^{-3},$$

$$|\delta g_R|_{[\sigma(B_s \rightarrow \mu\mu)=3 \times 10^{-10}]}^{\text{PC}} < 1.0 \times 10^{-5}. \quad (4.9)$$

Comparing with eqs. (3.76) and (3.79), one can see that the bounds on the couplings improve linearly with the improvement in accuracy, as expected from eq. (3.74). From eqs. (3.59) one likewise sees that the corresponding increase in the largest new-physics scale probed goes as the square root of the improvement in the coupling bounds, or in accuracy.

### $B_s \rightarrow \mu\mu$ as a probe of the Yukawa sector: new observables

In this outlook chapter, the discussion on  $\mathcal{B}(B_s \rightarrow \mu\mu)$  has so far been concerned with the progress to be expected on new *vector* interactions, namely on new contributions from the operators  $\mathcal{O}_{10}$  and  $\mathcal{O}'_{10}$  introduced in eq. (3.50). From the very same equation, it is evident that observables related to the  $B_s \rightarrow \mu\mu$  decay can be affected by a wealth of new scalar interactions as well, described by the effective operators  $\mathcal{O}_{S,P}^{(i)}$ . As a matter of fact, the rest of sec. 3.8 was devoted to showing that contributions to these operators (that are a null test within the SM) are very plausible. To this end, we have provided arguments based on general physics considerations, in particular the fact that the  $B_s \rightarrow \mu\mu$  decay is governed essentially by Yukawa interactions, fact that can easily be seen in the gaugeless limit of the SM, that in turn is a meaningful limit because of the large  $(m_t/m_W)^2 \approx 4$  ratio.

The above line of argument is aimed at reiterating that the  $B_s \rightarrow \mu\mu$  decay is an exquisite probe of Yukawa interactions beyond the SM paradigm. Correspondingly, one can expect that observables related to this decay may be able to extract valuable information on the couplings of the scalar operators  $\mathcal{O}_{S,P}^{(i)}$ , that are still poorly constrained, especially as concerns their *phases*.

This section is devoted to presenting a class of observables specifically designed for the above purpose. The reader may have noticed that in the previous paragraphs I have already used twice the phrasing ‘observables related to’ the  $B_s \rightarrow \mu\mu$  decay. This is because the observables to be invoked will be more involved than the sheer branching fraction, i.e., in order to be constructed, they will necessitate more than ‘just’ the count of  $B_s$  or  $\bar{B}_s$  mesons that have decayed into a  $\mu\mu$  pair. Specifically, they will require progress on decay-time information, and eventually on tagging – that for a decay as rare as  $B_s \rightarrow \mu\mu$  is well-known to be challenging. Nonetheless, it is to be expected that experimental progress on the  $B_s \rightarrow \mu\mu$  decay will almost necessarily lead to the experimental feasibility of these observables. In fact, experimental progress will mean in the first place larger and purer  $B_s \rightarrow \mu\mu$  samples, and this will accordingly increase sensitivity to decay-time and tagging information. In short, the ultimate aim of this section will be to show that the LHCb-upgrade era will bring substantial progress on the knowledge of fermion-fermion-scalar interactions, in particular CP-violating effects thereof.

The observables to be introduced are natural byproducts of the effect described in sec. 3.5, and likewise due to work by De Bruyn *et al.*, specifically ref. [86] as concerns the  $B_s \rightarrow \mu\mu$  decay. The large- $\Delta\Gamma_s$  effect presented in sec. 3.5 is summarized by eq. (3.35), relating the theoretical (SM) branching ratio with the experimental one via the factor  $1/(1 - y_s)$ , with  $y_s = \Delta\Gamma_s/(2\Gamma_s)$ . In fact, as shortly mentioned in the first of the items at the end of that section, this factor is model dependent, and as a matter of fact eq. (3.35) tacitly assumes the SM. Its generalization for arbitrary contributions to the Hamiltonian in eqs. (3.49) and

(3.50) reads [86]

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{th}} = \frac{1 - y_s^2}{1 + \mathcal{A}_{\Delta\Gamma} y_s} \cdot \mathcal{B}(B_s \rightarrow \mu\mu)_{\text{exp}} , \quad (4.10)$$

where we have introduced the quantity  $\mathcal{A}_{\Delta\Gamma}$ . This quantity bears sensitivity to the scalar- and pseudoscalar-operator Wilson coefficients, in particular to their phases. In fact, in the notation of eqs. (3.49) and (3.50) its explicit expression reads [86]

$$\mathcal{A}_{\Delta\Gamma} = \frac{|P|^2 \cos 2\phi_P - |S|^2 \cos 2\phi_S}{|P|^2 + |S|^2} , \quad (4.11)$$

where  $P$  and  $S$  are ‘normalized’ Wilson coefficients (in terms of the SM one), encoding contributions from operators with pseudoscalar or axial leptonic currents, and from those with scalar leptonic currents, respectively. They are defined as

$$\begin{aligned} P &\equiv \frac{c_{10} - c'_{10}}{c_{10}^{\text{SM}}} + \frac{m_{B_s}^2}{2m_\mu} \left( \frac{m_b}{m_b + m_s} \right) \left( \frac{c_P - c'_P}{c_{10}^{\text{SM}}} \right) , \\ S &\equiv \sqrt{1 - 4 \frac{m_\mu^2}{m_{B_s}^2}} \frac{m_{B_s}^2}{2m_\mu} \left( \frac{m_b}{m_b + m_s} \right) \left( \frac{c_S - c'_S}{c_{10}^{\text{SM}}} \right) , \end{aligned} \quad (4.12)$$

and the CP-violating phases  $\phi_{S,P}$  are such that  $P = |P|e^{i\phi_P}$  and  $S = |S|e^{i\phi_S}$ . By the definitions in eq. (4.12) one has  $P = 1$  and  $S = 0$  within the SM, implying  $\mathcal{A}_{\Delta\Gamma}^{\text{SM}} = +1$ .

Eq. (4.11) demonstrates in a self-commenting way the high sensitivity of  $\mathcal{A}_{\Delta\Gamma}$  to new effects, in particular to shifts from scalar and pseudoscalar Wilson coefficients, and their phases. In principle,  $\mathcal{A}_{\Delta\Gamma}$  can be constrained already from eq. (4.10), because  $y_s$  is also a measurable (and in fact measured, see eq. (3.36)) quantity. However, estimating  $\mathcal{A}_{\Delta\Gamma}$  from eq. (4.10) is not a clean procedure, because the l.h.s. of this equation is the theoretical branching ratio.

The question then arises, how to determine  $\mathcal{A}_{\Delta\Gamma}$  from measurable quantities only. It turns out that a quantity suited to this purpose is the so-called ‘‘effective lifetime’’ of the  $B_s \rightarrow \mu\mu$  decay, defined as [85]

$$\tau_{\mu\mu} \equiv \frac{\int_0^\infty \bar{\Gamma}(B_s(t) \rightarrow \mu^+\mu^-) t dt}{\int_0^\infty \bar{\Gamma}(B_s(t) \rightarrow \mu^+\mu^-) dt} , \quad (4.13)$$

where  $\bar{\Gamma}$  denotes the ‘‘untagged’’ rate<sup>53</sup>

$$\bar{\Gamma}(B_s(t) \rightarrow \mu^+\mu^-) \equiv \Gamma(B_s(t) \rightarrow \mu^+\mu^-) + \Gamma(\bar{B}_s(t) \rightarrow \mu^+\mu^-) , \quad (4.14)$$

namely the sum of the decay widths of the  $B_s(t)$  and of the  $\bar{B}_s(t)$  – implying in turn that eq. (4.13) can be measured without the need of retaining tagging information. As pointed out in ref. [86],  $\tau_{\mu\mu}$  allows a clean determination of  $\mathcal{A}_{\Delta\Gamma}$  via the relation [86]

$$\mathcal{A}_{\Delta\Gamma} = \frac{1}{y_s} \cdot \frac{(1 - y_s^2)\tau_{\mu\mu} - (1 + y_s^2)\tau_{B_s}}{2\tau_{B_s} - (1 - y_s^2)\tau_{\mu\mu}} , \quad (4.15)$$

whereby on the r.h.s. only measurables appear. We can also use this relation to eliminate  $\mathcal{A}_{\Delta\Gamma}$  in eq. (4.10), obtaining [86]

$$\mathcal{B}(B_s \rightarrow \mu\mu)_{\text{th}} = \left( 2 - (1 - y_s^2) \frac{\tau_{\mu\mu}}{\tau_{B_s}} \right) \cdot \mathcal{B}(B_s \rightarrow \mu\mu)_{\text{exp}} . \quad (4.16)$$

<sup>53</sup> Refs. [85, 86] use the notation  $\langle \Gamma \rangle$  for the untagged rate. However, we have already used the symbol  $\langle \Gamma \rangle_{[t]}$  in sec. 3.5 to denote time integration, consistently with the notation adopted in ref. [4].



The interest of this relation is that it allows to correct for the large- $\Delta\Gamma_s$  effect, thereby relating the theory with the experimental branching ratio, but, again, it does so using only measurable quantities on the r.h.s., and is furthermore valid for any model. Note also that in eq. (4.16) the dependence on  $y_s$  is only quadratic, hence it may be neglected to an approximation as good as about 1%.

In short, on the one side the measurement of the effective lifetime  $\tau_{\mu\mu}$  should be seen as an integral part of the determination of the theoretical  $B_s \rightarrow \mu\mu$  branching ratio from the experimental one; furthermore, as evident from eqs. (4.11) and (4.15), the effective lifetime offers an observable sensitive to new physics in a way that is complementary to the measurement of  $\mathcal{B}(B_s \rightarrow \mu\mu)$ .

At the sheer experimental level, effective lifetimes can be extracted from the decay-time distributions of the very same sample of *untagged* events used for the measurement of the total branching ratio. This procedure has already been applied to the  $B_s \rightarrow J/\psi f_0$  and  $B_s \rightarrow K^+ K^-$  decays by both the CDF [155] and LHCb [156] collaborations, with accuracies of about 7% with O(500) events. While in the post-run-II era of the LHC both of (at least) LHCb and CMS will have way more than 500 events for the  $B_s \rightarrow \mu\mu$  decay, it is next to impossible to make projections about the performance on  $\tau_{\mu\mu}$ , because the efficiency in the event-by-event extraction of time information does not, of course, depend on statistics alone.

A final point that should be stressed is that a determination of  $\mathcal{A}_{\Delta\Gamma} \neq +1$ , where, we recall, plus unity is the SM expectation, does not, *per se*, imply new CP-violating new physics, because the four parameters  $|S|$ ,  $|P|$ ,  $\phi_S$  and  $\phi_P$  will still be underconstrained from eq. (4.11) alone. However, a measurement of  $\mathcal{A}_{\Delta\Gamma} \neq +1$  will certainly indicate a beyond-SM effect. To ascertain whether this new physics is CP-violating or still compatible with CP conservation will call for the measurement of the time-dependent CP asymmetry, namely for a full-fledged time-dependent *and tagged* analysis. While this will be an overly challenging undertaking – even with the most optimistic assumptions on the LHCb-upgrade performances – one may argue that the result  $\mathcal{A}_{\Delta\Gamma} \neq +1$  will spur strong experimental motivation towards this new challenge.

## Acknowledgments

The work presented in this manuscript originates from several collaborations with Andrzej J. Buras and Gino Isidori. It is impossible to overstate my appreciation for the opportunity I had to collaborate with these two people. Both of them gave me rare, even unique, examples of enthusiasm for physics, of trust, and of creativity. I would like to thank the members ‘rapporteurs’ of the HDR jury, Damir Becirevic, Paolo Gambino and Marie-Noëlle Minard, for their commitment, their careful reading of the manuscript, and their feedback. A special acknowledgement to Damir goes also in connection with the topic discussed in sec. 3.4. If I turned my attention to this issue at all, I owe it to a large extent to several discussions with Damir in Orsay. I would also like to thank the other members of the HDR jury, Augusto Ceccucci, Guido Martinelli and Achille Stocchi, for kindly accepting this role. Finally, I would like to thank the director of my laboratory, Fawzi Boudjema, for his constant encouragement of my HDR undertaking, and the secretariat, especially Dominique Turc, for their professionalism and reliability.

## References

- [1] A. J. Buras and D. Guadagnoli, Phys.Rev. **D78**, 033005 (2008), 0805.3887.
- [2] A. J. Buras and D. Guadagnoli, Phys.Rev. **D79**, 053010 (2009), 0901.2056.
- [3] A. J. Buras, D. Guadagnoli, and G. Isidori, Phys.Lett. **B688**, 309 (2010), 1002.3612.
- [4] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, Eur.Phys.J. **C72**, 2172 (2012), 1208.0934.
- [5] D. Guadagnoli and G. Isidori, Phys.Lett. **B724**, 63 (2013), 1302.3909.
- [6] S. Weinberg, Eur. Phys. J. C **34** (2004) 5 [hep-ph/0401010]. See also ref. [19].
- [7] C.-N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
- [8] E. C. G. Sudarshan and R. E. Marshak, Phys. Rev. **109** (1958) 1860. R. P. Feynman and M. Gell-Mann, Phys. Rev. **109** (1958) 193.
- [9] J. Goldstone, Nuovo Cim. **19** (1961) 154. J. Goldstone, A. Salam and S. Weinberg, Phys. Rev. **127** (1962) 965.
- [10] P. W. Higgs, Phys. Lett. **12** (1964) 132; Phys. Rev. Lett. **13** (1964) 508.
- [11] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).
- [12] G. 't Hooft, Nucl. Phys. B **35** (1971) 167. G. 't Hooft and M. J. G. Veltman, Nucl. Phys. B **44** (1972) 189; B **50** (1972) 318.
- [13] G. Arnison *et al.* [UA1 Collaboration], Phys. Lett. B **122** (1983) 103; B **126** (1983) 398.
- [14] A. Djouadi, Phys.Rept. **457**, 1 (2008), hep-ph/0503172.
- [15] M. J. G. Veltman, Acta Phys. Polon. B **8** (1977) 475. T. Appelquist and C. W. Bernard, Phys. Rev. D **22** (1980) 200. A. C. Longhitano, Nucl. Phys. B **188** (1981) 118. M. B. Einhorn and J. Wudka, Phys. Rev. D **39** (1989) 2758. See also ref. [14].
- [16] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, R. Barate *et al.*, Phys.Lett. **B565**, 61 (2003), hep-ex/0306033.
- [17] See e.g. A. Djouadi, Phys. Rept. **457** (2008) 1 [hep-ph/0503172].
- [18] G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **710** (2012) 49 [arXiv:1202.1408 [hep-ex]]. S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **710** (2012) 26 [arXiv:1202.1488 [hep-ex]].
- [19] S. Weinberg, “The quantum theory of fields. Vol. 2: Modern applications”, Cambridge, UK: Univ. Pr. (1996) 489 p.
- [20] G. 't Hooft, NATO Adv.Study Inst.Ser.B Phys. **59**, 135 (1980).
- [21] S. Weinberg, “The Quantum theory of fields. Vol. 1: Foundations”, Cambridge, UK: Univ. Pr. (1995) 609 p.

- [22] P. Minkowski, Phys. Lett. B **67** (1977) 421. M. Gell-Mann, P. Ramond and R. Slansky, proceedings of the supergravity Stony Brook workshop, New York, 1979 (ed.s P. Van Nieuwenhuizen and D. Freedman, North-Holland, Amsterdam; see also P. Ramond, arXiv:hep-ph/9809459). T. Yanagida, proceedings of the workshop on unified theories and baryon number in the universe, Tsukuba, Japan 1979 (ed.s O. Sawada and A. Sugamoto, KEK Report No. 79-18, Tsukuba). S.L. Glashow in “Quarks and Leptons”, Cargèse, 1979 (ed.s M. L vy et al., North Holland 1980, Amsterdam). R.N. Mohapatra, G. Senjanovic, Phys. Rev. Lett. **44** (1980) 912. J. Schechter and J. W. F. Valle, Phys. Rev. D **22** (1980) 2227.
- [23] A. Strumia and F. Vissani, (2006), hep-ph/0606054.
- [24] N. Cabibbo, Phys. Rev. Lett. **10** (1963) 531. M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49** (1973) 652.
- [25] J. M. Cline, (2006), hep-ph/0609145.
- [26] Particle Data Group, J. Beringer *et al.*, Phys.Rev. **D86**, 010001 (2012).
- [27] L. Wolfenstein, Phys.Rev.Lett. **51**, 1945 (1983).
- [28] L. Maiani, arXiv:1303.6154 [hep-ph].
- [29] S. Glashow, J. Iliopoulos, and L. Maiani, Phys.Rev. **D2**, 1285 (1970).
- [30] B. Ioffe and E. Shabalin, Yad.Fiz. **10**, 363 (1969).
- [31] J. J. Aubert *et al.* [E598 Collaboration], Phys. Rev. Lett. **33** (1974) 1404. J. E. Augustin *et al.* [SLAC-SP-017 Collaboration], Phys. Rev. Lett. **33** (1974) 1406.
- [32] M. Bona *et al.* [UTfit Collaboration], JHEP **0803** (2008) 049 [arXiv:0707.0636 [hep-ph]]. See [utfit.org](http://utfit.org) for updates.
- [33] G. Isidori, Y. Nir, and G. Perez, Ann.Rev.Nucl.Part.Sci. **60**, 355 (2010), 1002.0900.
- [34] G. Isidori, arXiv:1302.0661 [hep-ph].
- [35] G. D’Ambrosio, G. Giudice, G. Isidori, and A. Strumia, Nucl.Phys. **B645**, 155 (2002), hep-ph/0207036.
- [36] A. J. Buras, P. Gambino, M. Gorbahn, S. Jager and L. Silvestrini, Phys. Lett. B **500** (2001) 161 [hep-ph/0007085]. See also ref. [75].
- [37] L. B. Okun, Int. J. Mod. Phys. A **17S1** (2002) 105 [hep-ph/0112031].
- [38] M. K. Anikina *et al.*, ZhETF 42 (1962) 130 [Sov. Phys. JETP **15** (1962) 93].
- [39] J. Christenson, J. Cronin, V. Fitch, and R. Turlay, Phys.Rev.Lett. **13**, 138 (1964).
- [40] L.D. Landau, ZhETF **32** (1957) 405 [Sov. Phys. JETP **5** (1957) 336]; Nucl. Phys. **3** (1957) 127.
- [41] A. D. Sakharov, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32 [JETP Lett. **5** (1967) 24] [Sov. Phys. Usp. **34** (1991) 392] [Usp. Fiz. Nauk **161** (1991) 61].
- [42] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, Eur.Phys.J. **C72**, 2172 (2012), 1208.0934.

- [43] K. Anikeev *et al.*, (2001), hep-ph/0201071.
- [44] M. Gell-Mann and A. Pais, Phys. Rev. **97** (1955) 1387. M. Gell-Mann and A. H. Rosenfeld, Ann. Rev. Nucl. Part. Sci. **7** (1957) 407.
- [45] RBC, UKQCD, P. Boyle *et al.*, Phys.Rev.Lett. **110**, 152001 (2013), 1212.1474.
- [46] L. Maiani and M. Testa, Phys.Lett. **B245**, 585 (1990).
- [47] K. M. Watson, Phys.Rev. **88**, 1163 (1952).
- [48] N. H. Christ *et al.* [RBC and UKQCD Collaboration], Phys. Rev. D **88** (2013) 1, 014508 [arXiv:1212.5931 [hep-lat]]. N. H. Christ, PoS LATTICE **2012** (2012) 008 [arXiv:1301.4239 [hep-lat]].
- [49] T. Blum *et al.*, Phys.Rev.Lett. **108**, 141601 (2012), 1111.1699.
- [50] Workshop on “Implications of LHCb measurements and future prospects”, 14-16 Oct., 2013, CERN, [indico.cern.ch/event/255380](http://indico.cern.ch/event/255380).
- [51] L.-L. Chau, Phys.Rept. **95**, 1 (1983).
- [52] V. Weisskopf and E. P. Wigner, Z. Phys. **63** (1930) 54. T. D. Lee, R. Oehme and C. -N. Yang, Phys. Rev. **106** (1957) 340.
- [53] A. J. Buras, p. 281 (1998), hep-ph/9806471.
- [54] E. A. Andriyash, G. G. Ovanesyan and M. I. Vysotsky, Phys. Lett. B **599** (2004) 253 [hep-ph/0310314]. E. A. Andriyash, G. G. Ovanesyan and M. I. Vysotsky, Phys. Atom. Nucl. **69** (2006) 286 [hep-ph/0502111].
- [55] A. J. Buras and M. Jamin, JHEP **0401**, 048 (2004), hep-ph/0306217.
- [56] A. J. Buras, M. Jamin, and M. E. Lautenbacher, Nucl.Phys. **B408**, 209 (1993), hep-ph/9303284.
- [57] A. J. Buras and M. E. Lautenbacher, Phys.Lett. **B318**, 212 (1993), hep-ph/9309203.
- [58] T. Inami and C. Lim, Prog.Theor.Phys. **65**, 297 (1981).
- [59] A. J. Buras, M. Jamin, M. E. Lautenbacher and P. H. Weisz, Nucl. Phys. B **370** (1992) 69 [Addendum-ibid. B **375** (1992) 501]. A. J. Buras, M. Jamin, M. E. Lautenbacher and P. H. Weisz, Nucl. Phys. B **400** (1993) 37 [hep-ph/9211304]. A. J. Buras, M. Jamin and M. E. Lautenbacher, Nucl. Phys. B **400** (1993) 75 [hep-ph/9211321]. M. Ciuchini, E. Franco, G. Martinelli and L. Reina, Nucl. Phys. B **415** (1994) 403 [hep-ph/9304257].
- [60] A. J. Buras, M. Jamin and P. H. Weisz, Nucl. Phys. B **347** (1990) 491. J. Urban, F. Krauss, U. Jentschura and G. Soff, Nucl. Phys. B **523** (1998) 40 [hep-ph/9710245].
- [61] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, Rev.Mod.Phys. **68**, 1125 (1996), hep-ph/9512380.
- [62] J. Brod and M. Gorbahn, Phys.Rev. **D82**, 094026 (2010), 1007.0684.
- [63] J. Brod and M. Gorbahn, Phys.Rev.Lett. **108**, 121801 (2012), 1108.2036.
- [64] O. Cata and S. Peris, JHEP **0303**, 060 (2003), hep-ph/0303162.

- [65] J. F. Donoghue, E. Golowich, and B. R. Holstein, *Phys.Lett.* **B135**, 481 (1984).
- [66] G. Colangelo and G. Isidori, p. 333 (2000), hep-ph/0101264.
- [67] J.-M. Gerard, C. Smith, and S. Trine, *Nucl.Phys.* **B730**, 1 (2005), hep-ph/0508189.
- [68] S. Herrlich and U. Nierste, *Nucl. Phys. B* **419** (1994) 292 [hep-ph/9310311]. S. Herrlich and U. Nierste, *Nucl. Phys. B* **476** (1996) 27 [hep-ph/9604330].
- [69] J. Laiho, E. Lunghi, and R. S. Van de Water, *Phys.Rev.* **D81**, 034503 (2010), 0910.2928.
- [70] CKMfitter collaboration: `ckmfitter.in2p3.fr`. Statistical method described in J. Charles *et al.*, *Eur. Phys. J. C* **41**, 1-131 (2005) [hep-ph/0406184].
- [71] R. Barbieri, G. Isidori, J. Jones-Perez, P. Lodone, and D. M. Straub, *Eur.Phys.J.* **C71**, 1725 (2011), 1105.2296.
- [72] E. Lunghi and A. Soni, *Phys. Lett. B* **666** (2008) 162 [arXiv:0803.4340 [hep-ph]]. E. Lunghi and A. Soni, *JHEP* **0908** (2009) 051 [arXiv:0903.5059 [hep-ph]].
- [73] RBC Collaboration, UKQCD Collaboration, D. Antonio *et al.*, *Phys.Rev.Lett.* **100**, 032001 (2008), hep-ph/0702042.
- [74] M. Blanke, A. J. Buras, D. Guadagnoli, and C. Tarantino, *JHEP* **0610**, 003 (2006), hep-ph/0604057.
- [75] A. J. Buras, *Acta Phys.Polon.* **B34**, 5615 (2003), hep-ph/0310208.
- [76] UTfit Collaboration, M. Bona *et al.*, *PMC Phys.* **A3**, 6 (2009), 0803.0659.
- [77] LHCb Collaboration, R. Aaij *et al.*, *Eur.Phys.J.* **C73**, 2373 (2013), 1208.3355.
- [78] LHCb Collaboration, R. Aaij *et al.*, *Phys.Rev.Lett.* **110**, 021801 (2013), 1211.2674.
- [79] S. Chatrchyan *et al.* [CMS Collaboration], *JHEP* **1204** (2012) 033 [arXiv:1203.3976 [hep-ex]]. R. Aaij *et al.* [LHCb Collaboration], *Phys. Rev. Lett.* **108** (2012) 231801 [arXiv:1203.4493 [hep-ex]]. G. Aad *et al.* [ATLAS Collaboration], *Phys. Lett. B* **713** (2012) 387 [arXiv:1204.0735 [hep-ex]]. V. M. Abazov *et al.* [D0 Collaboration], *Phys. Rev. D* **87** (2013) 7, 072006 [arXiv:1301.4507 [hep-ex]]. T. Aaltonen *et al.* [CDF Collaboration], *Phys. Rev. D* **87** (2013) 072003 [arXiv:1301.7048 [hep-ex]]. ATLAS Collaboration, ATLAS-CONF-2013-076.
- [80] CMS Collaboration, S. Chatrchyan *et al.*, *Phys.Rev.Lett.* **111**, 101804 (2013), 1307.5025.
- [81] LHCb collaboration, R. Aaij *et al.*, *Phys.Rev.Lett.* **111**, 101805 (2013), 1307.5024.
- [82] CMS and LHCb Collaborations, CMS-PAS-BPH-13-007.
- [83] G. Buchalla and A. J. Buras, *Nucl. Phys. B* **398** (1993) 285. G. Buchalla and A. J. Buras, *Nucl. Phys. B* **400** (1993) 225. M. Misiak and J. Urban, *Phys. Lett. B* **451** (1999) 161 [hep-ph/9901278]. G. Buchalla and A. J. Buras, *Nucl. Phys. B* **548** (1999) 309 [hep-ph/9901288].
- [84] G. Buchalla and A. J. Buras, *Phys.Rev.* **D57**, 216 (1998), hep-ph/9707243.

- [85] K. De Bruyn *et al.*, Phys.Rev. **D86**, 014027 (2012), 1204.1735.
- [86] K. De Bruyn *et al.*, Phys.Rev.Lett. **109**, 041801 (2012), 1204.1737.
- [87] T. Hermann, M. Misiak, and M. Steinhauser, JHEP **1312**, 097 (2013), 1311.1347.
- [88] C. Bobeth, M. Gorbahn, and E. Stamou, Phys.Rev. **D89**, 034023 (2014), 1311.1348.
- [89] K. Chetyrkin, J. H. Kuhn, and M. Steinhauser, Comput.Phys.Commun. **133**, 43 (2000), hep-ph/0004189.
- [90] B. A. Kniehl and A. Sirlin, Nucl.Phys. **B458**, 35 (1996).
- [91] Particle Data Group, ref. [26]. The LEP Electroweak Working Group, <http://lepewwg.web.cern.ch>.
- [92] S. Bethke, Eur.Phys.J. **C64**, 689 (2009), 0908.1135.
- [93] UTfit collaboration: [www.utfit.org](http://www.utfit.org). Averages prepared with the PDG 2012 input. Statistical method described in M. Ciuchini *et al.*, JHEP 0107 (2001) 013 [hep-ph/0012308].
- [94] Tevatron Electroweak Working Group, CDF and D0 Collaborations, (2011), 1107.5255.
- [95] CDF Collaboration, D0 Collaboration, T. Aaltonen *et al.*, (2012), 1207.1069.
- [96] C. Davies, PoS **LATTICE2011**, 019 (2011), 1203.3862.
- [97] R. Barbieri, M. Beccaria, P. Ciafaloni, G. Curci and A. Vicere, Phys. Lett. B **288** (1992) 95 [Erratum-ibid. B **312** (1993) 511] [hep-ph/9205238]. R. Barbieri, M. Beccaria, P. Ciafaloni, G. Curci and A. Vicere, Nucl. Phys. B **409** (1993) 105. J. Fleischer, O. V. Tarasov and F. Jegerlehner, Phys. Rev. D **51** (1995) 3820.
- [98] J. Brod, M. Gorbahn, and E. Stamou, Phys.Rev. **D83**, 034030 (2011), 1009.0947.
- [99] A. Sirlin, Nucl.Phys. **B332**, 20 (1990).
- [100] A. Sirlin, Phys. Lett. B **232** (1989) 123. See also G. Degrossi, S. Fanchiotti and A. Sirlin, Nucl. Phys. B **351** (1991) 49.
- [101] D. Melikhov and N. Nikitin, Phys.Rev. **D70**, 114028 (2004), hep-ph/0410146.
- [102] D. Yennie, S. C. Frautschi, and H. Suura, Annals Phys. **13**, 379 (1961).
- [103] S. Weinberg, Phys.Rev. **140**, B516 (1965).
- [104] G. Isidori, Eur.Phys.J. **C53**, 567 (2008), 0709.2439.
- [105] P. Janot, Phys.Lett. **B223**, 110 (1989).
- [106] F. Low, Phys.Rev. **110**, 974 (1958).
- [107] Y. Aditya, K. Healey, and A. A. Petrov, Phys.Rev. **D87**, 074028 (2013), 1212.4166.
- [108] I. Dunietz, R. Fleischer, and U. Nierste, Phys.Rev. **D63**, 114015 (2001), hep-ph/0012219.

- [109] G. Raven [LHCb Collaboration], arXiv:1212.4140. R. Aaij *et al.* [LHCb Collaboration], Phys. Lett. B **713** (2012) 378 [arXiv:1204.5675 [hep-ex]]. R. Aaij *et al.* (LHCb Collaboration), LHCb-CONF-2012-002.
- [110] LHCb Collaboration, R. Aaij *et al.*, Phys.Rev.Lett. **108**, 241801 (2012), 1202.4717.
- [111] A. J. Buras, Phys.Lett. **B566**, 115 (2003), hep-ph/0303060.
- [112] A. Bazavov *et al.* [Fermilab Lattice and MILC Collaborations], Phys. Rev. D **85** (2012) 114506 [arXiv:1112.3051 [hep-lat]]. C. McNeile, C. T. H. Davies, E. Follana, K. Hornbostel and G. P. Lepage, Phys. Rev. D **85** (2012) 031503 [arXiv:1110.4510 [hep-lat]]. H. Na, C. J. Monahan, C. T. H. Davies, R. Horgan, G. P. Lepage and J. Shigemitsu, Phys. Rev. D **86** (2012) 034506 [arXiv:1202.4914 [hep-lat]].
- [113] P. Dimopoulos *et al.* [ETM Collaboration], JHEP **1201** (2012) 046 [arXiv:1107.1441 [hep-lat]]. B. Blossier, J. Bulava, M. Della Morte, M. Donnellan, P. Fritzscht, N. Garron, J. Heitger and G. von Hippel *et al.*, PoS LATTICE **2011** (2011) 280 [arXiv:1112.6175 [hep-lat]]. E. Gamiz *et al.* [HPQCD Collaboration], Phys. Rev. D **80** (2009) 014503 [arXiv:0902.1815 [hep-lat]]. C. M. Bouchard, E. D. Freeland, C. Bernard, A. X. El-Khadra, E. Gamiz, A. S. Kronfeld, J. Laiho and R. S. Van de Water, PoS LATTICE **2011** (2011) 274 [arXiv:1112.5642 [hep-lat]].
- [114] R. J. Dowdall *et al.* [HPQCD Collaboration], Phys. Rev. Lett. **110** (2013) 22, 222003 [arXiv:1302.2644 [hep-lat]]. N. Carrasco *et al.* [ETM Collaboration], JHEP **1403** (2014) 016 [arXiv:1308.1851 [hep-lat]].
- [115] M. Misiak, p. 301 (2013), 1112.5978.
- [116] C. Bobeth *et al.*, Phys.Rev.Lett. **112**, 101801 (2014), 1311.0903.
- [117] S. Aoki *et al.* (Flavour Lattice Averaging Group), arXiv:1310.8555 [hep-lat]. See [itpwiki.unibe.ch/flag](http://itpwiki.unibe.ch/flag) for updates.
- [118] P. Gambino and C. Schwanda, Phys.Rev. **D89**, 014022 (2014), 1307.4551.
- [119] C. Bobeth, A. J. Buras, F. Kruger, and J. Urban, Nucl.Phys. **B630**, 87 (2002), hep-ph/0112305.
- [120] G. Isidori, (2013), 1302.0661.
- [121] M. S. Chanowitz, (1999), hep-ph/9905478.
- [122] U. Haisch and A. Weiler, Phys.Rev. **D76**, 074027 (2007), 0706.2054.
- [123] D. B. Kaplan, Nucl.Phys. **B365**, 259 (1991).
- [124] S. Davidson, G. Isidori, and S. Uhlig, Phys.Lett. **B663**, 73 (2008), 0711.3376.
- [125] B. Keren-Zur *et al.*, Nucl.Phys. **B867**, 429 (2013), 1205.5803.
- [126] K. Hagiwara, Ann.Rev.Nucl.Part.Sci. **48**, 463 (1998).
- [127] A. Freitas and Y.-C. Huang, JHEP **1208**, 050 (2012), 1205.0299.
- [128] CDF Collaboration, D0 Collaboration, T. Aaltonen *et al.*, Phys.Rev. **D86**, 092003 (2012), 1207.1069.

- [129] B. Batell, S. Gori, and L.-T. Wang, JHEP **1301**, 139 (2013), 1209.6382.
- [130] M. Baak and R. Kogler, p. 349 (2013), 1306.0571.
- [131] K. G. Wilson, Phys.Rev. **D10**, 2445 (1974).
- [132] J. Laiho, PoS **KAON13**, 003 (2013).
- [133] G. Colangelo *et al.*, Eur.Phys.J. **C71**, 1695 (2011), 1011.4408.
- [134] N. Isgur and M. B. Wise, Phys. Lett. B **232** (1989) 113. N. Isgur and M. B. Wise, Phys. Lett. B **237** (1990) 527.
- [135] A. Sirlin, Nucl.Phys. **B196**, 83 (1982).
- [136] A. Czarnecki and K. Melnikov, Nucl.Phys. **B505**, 65 (1997), hep-ph/9703277.
- [137] M. E. Luke, Phys. Lett. B **252** (1990) 447. This result is an heavy-quark-theory application of the Ademollo-Gatto theorem: M. Ademollo and R. Gatto, Phys. Rev. Lett. **13** (1964) 264.
- [138] Heavy Flavor Averaging Group, Y. Amhis *et al.*, (2012), 1207.1158.
- [139] S. -W. Qiu *et al.* [Fermilab Lattice and MILC Collaboration], arXiv:1312.0155 [hep-lat]. J. A. Bailey *et al.* [Fermilab Lattice and MILC Collaborations], PoS LATTICE **2010** (2010) 311 [arXiv:1011.2166 [hep-lat]]. C. Bernard, C. E. DeTar, M. Di Pierro, A. X. El-Khadra, R. T. Evans, E. D. Freeland, E. Gamiz and S. Gottlieb *et al.*, Phys. Rev. D **79** (2009) 014506 [arXiv:0808.2519 [hep-lat]].
- [140] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev and A. I. Vainshtein, Phys. Rev. D **52** (1995) 196 [hep-ph/9405410]. A. Kapustin, Z. Ligeti, M. B. Wise and B. Grinstein, Phys. Lett. B **375** (1996) 327 [hep-ph/9602262].
- [141] P. Gambino, T. Mannel and N. Uraltsev, Phys. Rev. D **81** (2010) 113002 [arXiv:1004.2859 [hep-ph]]. P. Gambino, T. Mannel and N. Uraltsev, JHEP **1210** (2012) 169 [arXiv:1206.2296 [hep-ph]].
- [142] I. I. Bigi and N. Uraltsev, Int.J.Mod.Phys. **A16**, 5201 (2001), hep-ph/0106346.
- [143] I. I. Bigi, N. Uraltsev and R. Zwicky, Eur. Phys. J. C **50** (2007) 539 [hep-ph/0511158]. C. Breidenbach, T. Feldmann, T. Mannel and S. Turczyk, Phys. Rev. D **78** (2008) 014022 [arXiv:0805.0971 [hep-ph]]. I. Bigi, T. Mannel, S. Turczyk and N. Uraltsev, JHEP **1004** (2010) 073 [arXiv:0911.3322 [hep-ph]].
- [144] C. W. Bauer, Z. Ligeti, M. Luke, A. V. Manohar and M. Trott, Phys. Rev. D **70** (2004) 094017 [hep-ph/0408002]. See HFAG update [138].
- [145] J. Brod and J. Zupan, JHEP **1401**, 051 (2014), 1308.5663.
- [146] M. Gronau and D. London, Phys. Lett. B **253** (1991) 483. M. Gronau and D. Wyler, Phys. Lett. B **265** (1991) 172.
- [147] C. Aubin, C. D. Lin, and A. Soni, Phys.Lett. **B710**, 164 (2012), 1111.4686.
- [148] D. Atwood, I. Dunietz and A. Soni, Phys. Rev. Lett. **78** (1997) 3257 [hep-ph/9612433]. D. Atwood, I. Dunietz and A. Soni, Phys. Rev. D **63** (2001) 036005 [hep-ph/0008090].



- [149] LHCb Collaboration, R. Aaij *et al.*, Phys.Lett. **B712**, 203 (2012), 1203.3662.
- [150] A. Giri, Y. Grossman, A. Soffer, and J. Zupan, Phys.Rev. **D68**, 054018 (2003), hep-ph/0303187.
- [151] Y. Grossman, Z. Ligeti and A. Soffer, Phys. Rev. D **67** (2003) 071301 [hep-ph/0210433].  
A. Bondar and A. Poluektov, Eur. Phys. J. C **47** (2006) 347 [hep-ph/0510246].
- [152] I. Dunietz and R. G. Sachs, Phys. Rev. D **37** (1988) 3186 [Erratum-ibid. D **39** (1989) 3515].  
R. Aleksan, I. Dunietz and B. Kayser, Z. Phys. C **54** (1992) 653. R. Fleischer, Nucl. Phys. B **671** (2003) 459 [hep-ph/0304027].
- [153] T. Aushev *et al.*, (2010), 1002.5012.
- [154] R. Forty, CERN Report No. LHCb-PUB-2012-010. CERN-LHCb-PUB-2012-010, 2012 (unpublished).
- [155] CDF Collaboration, T. Aaltonen *et al.*, Phys.Rev. **D84**, 052012 (2011), 1106.3682.
- [156] LHCb Collaboration, R. Aaij *et al.*, Phys.Lett. **B707**, 349 (2012), 1111.0521.