Flavour Anomalies in $b \rightarrow s\ell^+\ell^-$ Processes - a round table Discussion


$^1$University of Warwick, Coventry, UK
$^2$The University of Manchester, Manchester, UK
$^3$Institut de Ciencies del Cosmos, Universitat de Barcelona, Barcelona, Spain
$^4$University of Sussex, Brighton, UK
$^5$Institute of High Energy Physics and Theoretical Physics Center for Science Facilities, Chinese Academy of Sciences, Beijing, 100049, China
$^6$Higgs Centre for Theoretical Physics, University of Edinburgh, Edinburgh, UK

Abstract. Precision measurements of flavour observables provide powerful tests of many extensions of the Standard Model. This contribution covers a range of flavour measurements of $b \rightarrow s\ell^+\ell^-$ transitions, several of which are in tension with the Standard Model of particle physics, as well as their theoretical interpretation. The basics of the theoretical background are discussed before turning to the main question of the field: whether the anomalies can be explained by QCD effects or whether they may be indicators of effects beyond the Standard Model.

---

$^a$e-mail: Marco.Gersabeck@cern.ch
1 Introduction

Flavour physics has a long track record of discoveries that paved the way for advances in particle physics. In particular the discovery of the $B^0$ meson oscillations in 1987 \cite{1} is a great example demonstrating the potential of flavour physics to infer physics of high mass scales through precision measurements at low scales: the observed rate of oscillations was the first indication of the top quark being much heavier than the other five quark flavours. Precision flavour measurements at the LHC are sensitive to indirect effects from physics beyond the Standard Model (SM) at far greater scales than those accessible in direct searches.

This paper summarises a panel discussion focussing on a range of anomalies seen in measurements of $b \to s\ell^+\ell^-$ decays, which are a sensitive class of processes to potential effects beyond the SM (BSM). The central question is whether the observed anomalies are indeed BSM effects or whether they can be explained by QCD effects. The general $b \to s\ell^+\ell^-$ framework is briefly discussed below, the experimental situation is reviewed in Sec. 2, which are followed by theory Secs. 3-5 on the heavy quark framework, form-factor determinations and the relevance of long distance charm contributions. Global fit strategies aiming to combine experimental observables to extract a more precise theoretical picture are presented in Sec. 6, and finally Sec. 7 covers potential BSM interpretations of the flavour anomalies before the summary, which is given in Sec. 8.

1.1 General $b \to s\ell^+\ell^-$ framework

In order to maintain some coherence between the different contributions of various authors we give the effective Hamiltonian with some minimal explanation referencing to the theory based sections.

$$\mathcal{H}_{\text{eff}} = \frac{G_F}{\sqrt{2}} \left( \sum_{i=1}^{2} (\lambda_u C_i O_i^u + \lambda_c C_i O_i^c) - \lambda_t \sum_{i=3}^{10} C_i O_i \right), \quad \lambda_i \equiv V_{ub}^* V_{tb},$$

where $V_{ij}$ are CKM-elements, $\mu$ is a factorisation scale, $C_i$ the Wilson coefficients encoding the ultra-violet (UV) physics and some of the most important operators are

$$O_1 = (\bar{s}_i q_j)_{V-A}(\bar{q}_j b_i)_{V-A}, \quad O_2 = (\bar{s}_i q_i)_{V-A}(\bar{q}_j b_j)_{V-A},$$

$$O_{7[8]} = -\frac{e m_b}{8\pi^2} \bar{s}\gamma \cdot F[G](1 + \gamma_5) b, \quad O_{9,10} = \frac{\alpha}{2\pi} (\bar{\ell}\gamma^\mu[\gamma_5]\ell)(\bar{s}\gamma_\mu(1 - \gamma_5)b),$$

(2)
where $i, j$ are colour indices, $(\bar{s}b)_{V,A} = \bar{s}y^b (1 \pm \gamma_5) b$, $e = \sqrt{4\pi\alpha} > 0$ and $G_F$ is the Fermi constant. The Wilson coefficients $C_i$ are computed to next-to-next-to-leading order (NNLO) in perturbation theory, and the matrix elements of the operators are of non-perturbative nature. Those are either local short distance form-factors or non-local long distance matrix elements. The form-factors are computed in LCSR and in lattice QCD in complementary regimes as discussed in Sec. 4, and obey some symmetry relations for large large $m_b$-mass, cf. Secs 3 and 4.3 respectively. At large recoil of the lepton pair the matrix elements are evaluated outside the resonance region, either within the heavy quark expansion or LCSR as discussed in Secs. 3 and 5 respectively. At low-recoil an OPE has been proposed cf. Sec. 3, subject to potentially significant corrections from resonant charm contribution as discussed in Sec. 5.

2 Summary of experimental situation (T. Blake)

There has been huge experimental progress in measurements of rare $b \to s$ processes in the past five years. This has been driven by the large $b\bar{b}$ production cross-section in $pp$ collisions at the LHC, which enabled the LHC experiments to collect unprecedented samples of decays with dimuon final-states.

2.1 Leptonic decays

The decay $B_0^s \to \mu^+\mu^-$ is considered a golden mode for testing the SM at the LHC. The SM branching fraction depends on a single hadronic parameter, the $B_0^s$ decay constant, that be computed from Lattice QCD. Consequently the SM branching fraction is known to better than a 10% precision [2]. A combined analysis of the CMS and LHCb datasets [3] results in a time-averaged measurement of the branching fraction of the decay of

$$\overline{B}(B_0^s \to \mu^+\mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9}.$$ (3)

A recent ATLAS measurement [4] yields

$$\overline{B}(B_s^0 \to \mu^+\mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9},$$ (4)

which is consistent with the combined analysis from CMS and LHCb. These measurements are in good agreement with SM predictions and set strong constraints on extensions of the SM that introduce new scalar or pseudoscalar couplings.

2.2 Semileptonic decays

The large dataset has also enabled the LHCb and CMS experiments to make the most precise measurements of the differential branching fraction of $B \to K^{(*)}\mu^+\mu^-$ and $B_0^s \to \phi\mu^+\mu^-$ to date [5–7]. Above the open charm threshold, broad resonances are seen in the LHCb dataset [8]. The most prominent of these is the $\psi(4160)$. The regions close to the narrow charmonium resonances are excluded from the analysis. With the present binning scheme, the uncertainties on differential branching fraction measurements are limited by the knowledge of the $B \to J/\psi K^{(*)}$ and $B_0^s \to J/\psi\phi$ branching fractions that are used to normalise the signal. The measured differential branching fractions of $b \to s\mu^+\mu^-$ processes, tend to prefer smaller values than their corresponding SM predictions. The largest discrepancy is seen in the $B_0^s \to \phi\mu^+\mu^-$ decay, where the data are more than 3 $\sigma$ from the SM predictions in the dimuon invariant mass squared range $1 < q^2 < 6 \text{ GeV}^2/c^4$, see Fig 1.
While the branching fractions of rare $B \to K^{(*)}\mu^+\mu^-$ decays have large theoretical uncertainties arising from the $B \to K^{(*)}$ form-factors, many sources of uncertainty will cancel when comparing the decay rates of the $B \to K^{(*)}\mu^+\mu^-$ and $B \to K^{(*)}e^+e^-$ decays. In the range $1 < q^2 < 6 \text{ GeV}^2/c^4$, the LHCb experiment measures [12]

$$R_K[1,6] = 0.745^{+0.090}_{-0.074}\text{(stat)}^{+0.035}_{-0.035}\text{(syst)}.$$  \hfill (5)

This is approximately 2.6σ from the SM expectation of almost identical decay rates for the two channels. In order to cancel systematic differences between the reconstruction of electrons and muons in the detector, the LHCb analysis is performed as a double ratio to the rate $B^+ \to J/\psi K^+$ decays (where the $J/\psi$ can decay to a dielectron or dimuon pair). The migration of events in $q^2$ due to final-state-radiation is accounted for using samples of simulated events. QED effects are simulated through PHOTOS [13]. The largest difference between the dimuon and dielectron final-states comes from Bremsstrahlung from the electrons in the detector. This is simulated using GEANT 4 [14].

The modelling of the migration of events and the line-shape of the decay are the main contributions to the systematic uncertainty on the measurement.

The distribution of the final-state particles in the $B \to K^* \ell^+\ell^-$ decay can be described by three angles and $q^2$. The angles are: the angle between the direction of the $\ell^+$ ($\ell^-$) and the $B$ ($\bar{B}$) in the rest-frame of the dilepton pair; the angle between the direction of the kaon and the direction of the $B$ in the $K^*$ rest-frame; and the angle between the decay planes of the dilepton pair and the $K^*$ in the rest-frame of the $B$, denoted $\phi$. The resulting angular distribution can be parameterised in terms of eight angular observables: the longitudinal polarisation of the $K^*$, $F_L$; the forward-backward asymmetry of the dilepton system, $A_{FB}$; and six additional observables that cancel when integrating over $\phi$. Existing measurements of the observables $A_{FB}$ and $F_L$ are shown in Fig. 2 along with SM predictions based on Refs. [9, 10]. The most precise measurements of the $F_L$ and $A_{FB}$ come from the LHCb and CMS experiments. In general, the measurements are consistent with each other and are compatible with the SM predictions. The largest tension is seen in the BaBar measurement of $F_L$ [15].

The ATLAS, CMS, LHCb and Belle experiments have also measured the remaining angular observables that are usually cancelled by integrating over $\phi$. The LHCb collaboration performed a first
Figure 2. Observables $F_L$ and $A_{FB}$ measured by the BaBar [15], Belle [16], CDF [17], CMS [7] and LHCb [18] experiments for the $B \rightarrow K^* \mu^+ \mu^-$ decay as a function of the dimuon invariant mass squared, $q^2$. The shaded region indicates a theoretical prediction for the observables based on Refs. [9, 10]. No data point is shown for CMS in the range $q^2 < 1 \text{ GeV}^2/c^4$, due to the thresholds used in the CMS trigger system.

Figure 3. Observable $P'_5$ measured by LHCb [18] and Belle [19] as a function of the dimuon invariant mass squared, $q^2$, in the $B \rightarrow K^* \mu^+ \mu^-$ decay. Preliminary results from ATLAS [20] and CMS [21] are also included. The shaded regions indicate theoretical predictions from Ref. [22].

full angular analysis of the decay in Ref. [18]. The majority of these additional observables are consistent with SM predictions. However, a tension exists between measurements of the observable $P'_5$ and their corresponding SM prediction in the region $4 < q^2 < 8 \text{ GeV}^2/c^4$. This tension is illustrated in Fig. 3. In the region $4 < q^2 < 8 \text{ GeV}^2/c^4$, the data from ATLAS, Belle and LHCb are significantly above the SM predictions. The CMS result is more consistent.

The experimental measurements of the angular observables are currently statistically limited. The largest sources of systematic uncertainty arise from modelling of the experimental angular acceptance and the background angular distribution.
3 Rare $B$-decays and the heavy quark expansion (S. Jäger)

3.1 Context

In the Standard Model (SM), the amplitude for a rare semileptonic decay $\bar{B} \to M\ell^+\ell^-$, with $M$ a hadronic system such as $\bar{K}, \bar{K}\pi, \ldots$ can be written, to leading order in the electromagnetic coupling $\alpha_{EM}$ but exact in QCD, as

$$\mathcal{A}(\bar{B} \to M\ell^+\ell^-) = L^\mu a_{V\mu} + L^{5\mu} a_{A\mu},$$

where $L^\mu$ and $L^{5\mu}$ are the vector and axial lepton currents. One can trade $a_{V\mu}, a_{A\mu}$ for helicity amplitudes $H_{V,\lambda}(q^2)$ and $H_{A,\lambda}(q^2)$, where $\lambda$ is the helicity of the dilepton, which coincides with that of the hadronic state. For a kaon, $\lambda = 0$ and there are two hadronic amplitudes. For a narrow $K^*$, $\lambda = \pm 1$ and there are six amplitudes. These amplitudes determine all rate and angular observables measured at $B$-factories and LHCb (where the lepton spins are not measured).

In the conventions of [23], $H_{A,1} \propto C_{10} V_1(q^2)$, where $V_1$ are form factors and $C_{10}$ is the axial semileptonic Wilson coefficient. (We omitted a normalisation free of hadronic uncertainties.) Note that $C_{10}$ and $V_1$ are renormalisation-scale and -scheme-independent. $C_{10}$, in fact, receives no contributions from below the weak scale at all — this is what makes $B_s \to \mu^+\mu^-$ such a precision observable. The vector amplitudes can be written [23]

$$H_{V,\pm} \propto \left[ C_9(\mu) V_{\pm}(q^2) + \frac{2m_b m_B}{q^2} C_7^{\text{eff}}(\mu) T_{\pm}(q^2,\mu) - 16\pi^2 \frac{m_B^2}{q^2} h_{\pm}(q^2,\mu) \right],$$

$$H_{V,0} \propto \frac{\lambda^{1/2}}{2 m_B \sqrt{q^2}} \left[ C_9(\mu) V_{0}(q^2) + \frac{2m_b}{m_B} C_7^{\text{eff}}(\mu) T_{0}(q^2,\mu) - 16\pi^2 \frac{m_B^2}{q^2} h_{0}(q^2,\mu) \right].$$

Here $C_9(\mu)$ denotes the vector semileptonic Wilson coefficient, $T_{\lambda}$ tensor form factors, multiplied by the “effective” dipole Wilson coefficient $C_7^{\text{eff}}$, and $h_{\lambda} \sim (M(\lambda)/T \{ \frac{1}{2} + \left. \frac{1}{2} \right. )$ $|\bar{B}\rangle$ denotes the contribution from the hadronic weak Hamiltonian (see [23] for a precise definition). So far everything is exact from a QCD point of view. The theoretical difficulty resides in the evaluation of $h_{\lambda}$, which the heavy-quark expansion achieves, and of the form factors, for which the heavy-quark expansion provides relations.

One notes that $C_9(\mu)$ and $C_7^{\text{eff}}(\mu)$ are strongly scale-dependent. E.g., at NNLL order, $C_9(10 \text{GeV}) = 3.75$, $C_9(5 \text{GeV}) = 4.18$, $C_9(2.5 \text{GeV}) = 4.49$. (One may compare this variation to a putative $\Delta C_9^{\text{BSM}} \sim -1$.) The scale dependence must be precisely cancelled by the non-perturbative object $h_{\lambda}$. (This follows from RG-invariance of the Hamiltonian, which is exact.) Any quantitative theoretical description of $h_{\lambda}$ must correctly incorporate this.

3.2 The heavy-quark expansion

In the heavy-quark, large-recoil limit $E_M, m_B \gg \Lambda$, where $\Lambda$ is the QCD scale parameter, $h_{\lambda}$ factorizes into perturbative hard-scattering kernels multiplying form factors and light-cone distribution amplitudes for the light meson $M = K, K^*$ [25–27], known as QCD factorization. It applies in the $q^2$-region below the narrow charm resonances, and perhaps above the charm threshold up to $q^2 \sim 15 \text{GeV}^2$.

---

1. We neglect the lepton mass, as the anomalies occur at $q^2 \gg m_\ell^2$. This eliminates a seventh amplitude. We will also neglect the strange quark mass and CKM-suppressed terms throughout.

2. A recent LHCb paper [24] models $h_{\lambda}$, for $B \to K$, as a sum of Breit-Wigner resonances. While this more or less fits the data ($p = 0.01$), it has no scale-dependence, which is one reason why the coefficient $C_9$ in that paper cannot be identified with $C_9(\mu)$ for any scale $\mu$. 

---
QCDF can be formulated in soft-collinear effective field theory (SCET) language [28]. “Factorization” is used in the Wilsonian sense of separating physics of the scales $\sqrt{\Lambda m_b}, m_b$, and above from the physics of the scale $\Lambda$, not to be confused with “naive factorization.” There are two distinct classes of effect. Vertex corrections may be compactly written as

$$C_9 \rightarrow C_9(\mu) + Y(\mu, q^2, m_c) + \frac{\alpha_s}{4\pi} Y^{(1)}(\mu, q^2, m_c), \quad (9)$$

$$C_7^{\text{eff}} \rightarrow C_7^{\text{eff}}(\mu) + \frac{\alpha_s}{4\pi} Z^{(1)}(\mu, q^2, m_c), \quad (10)$$

where $Y, Y^{(1)},$ and $Z^{(1)}$ contain loop functions and Wilson coefficients. This form makes the helicity independence of these effects evident. The combination $C_9 + Y$ is traditionally called $C_9^{\text{eff}}(q^2)$; the heavy-quark limit justifies its use, but also predicts model-independent higher-order corrections. (Note that $C_9$ starts at $O(1/\alpha_s)$ in the logarithmic counting, so $Y$ is formally a NLL correction.) The r.h.s. of (9),(10) are $\mu$-independent up to $O(\alpha_s^2)$ corrections. The other class of effects constitutes so-called hard spectator scattering and includes an annihilation contribution at $O(\alpha_s^0)$, though the latter comes with small CKM-factors and Wilson coefficients. Spectator scattering probes the structure of the $B$ and $K^*$-mesons through their light-cone distribution amplitudes and is helicity dependent (and vanishes for $\lambda = +$). Schematically,

$$h_\lambda^{\text{spec}} \propto T_4(\alpha_s) \ast \phi_{B_s} \ast \phi_{K^*}. \quad (11)$$

This expression is separately scale-independent, resulting in formally $\mu$-independent observables. Corrections to the heavy-quark limit scale like $\Lambda/m_B$ and do not factorize, and must be estimated in other ways.

The expressions so far are sufficient to express all observables in terms of form factors [29], one can then use form factor results from light-cone sum rules to compute observables. Alternatively, one can make use of the fact that the heavy-quark large-recoil limit also implies relations between different form factors [30–32]. They look extremely simple in the helicity basis,

$$T_4(q^2) = V_4(q^2)[1 + f_4(q^2, \alpha_s)] + \text{spectator scattering} + O(\Lambda/m_B) \quad (12)$$

for $\lambda = -, 0$ and $T_+^+(q^2) = V_+(q^2) = O(\Lambda/m_B)$. (The latter together with $h_+ = 0$ implies $H_{V+} = H_{A+} = 0$.) Here $f_4$ is a perturbative expression starting at $O(\alpha_s)$. The spectator-scattering contribution has a form similar to (11), and is again proportional to $\alpha_s$. All $O(\alpha_s^2)$ corrections are also known [33–36]. Note that the ratios $T_4/V_4$ are free of hadronic input in the heavy-quark limit, up to (calculable) spectator-scattering and (incalculable) power corrections.

There is an alternative expansion, applicable at $M_B^2, q^2 \gg \Lambda^2$, in particular above the $D\bar{D}$ threshold. The actual expansion is in $E_M/\sqrt{q^2}$ and expresses $h_\lambda$ in terms of matrix elements of local operators (OPE) of increasing dimension, again with perturbatively calculable coefficients [37–39]. This is not by itself a heavy-quark expansion, although the kinematics ensure that a HQE is valid and the $b$-quark field may be expanded accordingly [38], and the HQE can be used to estimate the matrix elements [39]. The leading matrix elements are again the form factors $T_4$ and $V_4$, with perturbative coefficient functions that coincide with $C_7^{\text{eff}}$ and $C_9^{\text{eff}}(q^2)$ but different $\alpha_s$-corrections. The leading higher-dimensional corrections are of order $\Lambda^3/q^2$ and negligible, in particular spectator scattering is (strongly) power-suppressed. The OPE also gives a qualitative picture how open-charm resonances arise, by means of analytic continuation of OPE remainder terms from spacelike to physical (timelike) $q$, $\exp(-ic \sqrt{q^2}/\Lambda) \rightarrow \exp(-ic \sqrt{q^2}/\Lambda)$, giving oscillatory behaviour. Note that these “duality-violating” terms are nonanalytic in $\Lambda$, hence the large-recoil $\Lambda/m_B$ expansion will not capture them either. Formally such terms are of “infinite order” in $\Lambda/\sqrt{q^2}$ or $\Lambda/m_B$. They become less important
away from the threshold and partly cancel out in binned observables (see [39, 40]). See also the discussion of charm resonances in $B \to K^{(*)} \mu^+\mu^-$ in Section 5.

### 3.3 Phenomenology

How large are the various effects and residual uncertainties? The shift to $C^\text{eff}_7$ is an $O(25\%)$ constructive correction. The $O(\alpha_s)$ correction to $C_9$ is an order 5% destructive effect which partly cancels the $O(\alpha_s^0)$ term i.e. $Y(q^2)$. The spectator-scattering corrections are smaller, though the normalisation is quite uncertain (by about a factor of two) due to the poor knowledge of the $B$-meson LCDA $\phi_B$. Overall the effects are significant. For instance, $\lim_{q^2 \to 0}[q^2 H_{V^-}(q^2)]$ gives the $B \to K^-\gamma$ amplitude, which receives a $+30\%$ correction, or a $+70\%$ correction at the rate level. Because $C^\text{eff}_7$ is tightly constrained from the measured $B \to X_s\gamma$ rate, this allowed [25] to conclude that, absent large power corrections, $T_-(0) = T_1(0) = 0.27 \pm 0.04$, updated to $T_1(0) = 0.28 \pm 0.02$ in [27] and well below LCSR predictions at the time. A recent LCSR evaluation [10] gives $T_1(0) = 0.308 \pm 0.031$. The consistency supports smallness of power corrections.

It is impossible to give a comprehensive phenomenology of angular $B \to K^*\mu^+\mu^-$ observables in this space. Two observables that are very sensitive to $C_9$ are the forward-backward asymmetry, specifically its zero-crossing, and the angular term $S_5$ (or $P_5^\gamma$). As long as Wilson coefficients are real, the FBAS zero is determined by $\Re H_{V^-}(q_0^2) = 0$, up to second-order power corrections $O(\Lambda^2/m_B^2)$. From (7) it is clear that the zero depends on $(C^\text{eff}_7/C_9) \times (T_-/V_-)$, essentially free from form-factor uncertainties in the heavy-quark limit. Given the fact that $C^\text{eff}_7$ is essentially pinned to its SM value by $B \to X_s\gamma$, a $q_0^2$-measurement may hence be viewed as a determination of $C_9$. Ref. [27] obtained $q_0^2 = 4.36^{+0.33}_{-0.31} \text{GeV}^2$ (neglecting power corrections), to be compared to the LHCb determination [18] $q_0^2 \in [3.40,4.87] \text{GeV}^2$, in good agreement. By comparison, the LCSR form factors of [10] imply a lower crossing point, around 3.5 GeV$^2$, giving a slight preference for $C_9 < C_9^\text{SM}$. This can be traced to a ratio of $T_--V_-$ that is $O(10\%)$ below the heavy-quark-limit prediction, which is still consistent with a power correction. The eventual relative accuracy on $C_9$ is limited by that on ratio $T_-/(q_0^2)/V_-/(q_0^2)$.

The observable $P_5^\gamma$ [41], which shows the most pronounced anomaly, depends on all six helicity amplitudes. It hence depends on ratios $T_-/V_-, T_0/V_0, V_+/V_-, T_+/V_-$, and $V_0/V_-$. The last of these does not satisfy a heavy-quark relation, but $P_5^\gamma$ has been constructed in such a way that dependence on it cancels in the heavy-quark limit if $\alpha_s$-corrections are neglected. It has been suggested to employ $V \propto V_++V_-$ instead of $V_-$ [22], which reduces the explicit sensitivity to $V_+$ in $P_5^\gamma$. From a pure heavy-quark perspective, the physics of $V_+$ (which involves soft physics flipping the helicity of the strange quark emitted from $b$-quark decay, or changing the helicity of the $B$-remnant absorbed into the $K^0$) and $V_-$ (which does not require such a spin-flip) seem very different, such that the conservative choice is to associate separate uncertainties to both (hence a larger one to $V$). The impact on the significance of the $P_5^\gamma$ anomaly is noticeable because of this [22, 42] and because [22] adopt central values for power corrections to match LCSR form factor central values while [42] use zero central values (the HQ limit). Further recent investigations on the role of power corrections in semileptonic $B$-decays can be found in [43–45].

### 3.4 Discussion

The heavy-quark expansion goes a long way to putting rare $B$ decays on a systematic theoretical footing, and rightly has found its place at the heart of many phenomenological works and fits to BSM effects. In particular it removes most of the ambiguities of older “$C_{7,9,10}$ + resonances” approaches. Its primary limitations are incalculable power corrections. There is no evidence that these are abnormally large. Rather, the fact that they matter in the interpretation of anomalies points to the impressive
precision that experiment has already reached (and the apparent smallness of possible BSM effects). At the moment, there is no first-principles method to compute power corrections. Light-cone sum rule calculations can provide information. As they carry their own uncontrolled systematics, it would be particularly desirable to have sum rule results directly for the power-suppressed terms, where possible. Combining these with the leading-power expressions would remove most of the systematics of either framework. One example is the sum rule [46] for \( h_m(q^2) \), which vanishes in the heavy-quark limit. One finds an extra (ie double) power suppression of this term at \( q^2 \approx 0 \), which implies an excellent sensitivity to \( C_7' \) [23, 42]. Data-driven approaches may be able to constrain some of the power corrections, especially if data in very small bin-sizes becomes available for \( B \to K^\ast \mu^+\mu^- \). This path has been followed to some extent in [47].

4 Form-factors (Z.Liu & R.Zwicky)

Form-factors (FFs) describe the short-distance part of the transition amplitudes. For hadronic transitions of the type \( B \to M \) at the quark level \( b \to q \) they consist of matrix elements of the form \( \langle M(p) | \mathcal{B} (p_B) \rangle \). For \( M \) being a light meson the FFs can be computed from light-cone sum rules (LCSR) and lattice QCD at low and high momentum transfer \( q^2 = (p_B - p)^2 \) respectively. The ones most relevant to the current discussion of flavour anomalies are the \( B \to V (V = K^\ast, \phi, \ldots) \) FFs which follow from the vector and tensor current (cf. \( O_{9,10} \) and \( O_7 \) in (1))

\[
\begin{align*}
\langle K^\ast (p, \eta) | \bar{b} \gamma^\mu (1 \mp \gamma_5) b | \bar{B}(p_B) \rangle &= \, P^\mu_1 V_1(q^2) \pm P^\mu_2 V_2(q^2) \pm P^\mu_3 V_3(q^2) \pm P^\mu_4 V_T(q^2), \\
\langle K^\ast (p, \eta) | \bar{b} i q_{\nu} \gamma^\mu (1 \mp \gamma_5) b | \bar{B}(p_B) \rangle &= \, P^\mu_1 T_1(q^2) \pm P^\mu_2 T_2(q^2) \pm P^\mu_3 T_3(q^2),
\end{align*}
\]

where we have chosen \( V = K^\ast \) as a matter of concreteness. Above \( P^\mu_1,2,3,4 \) are Lorentz structures involving the \( K^\ast \)-meson polarisation vector \( \eta \) and momenta and the structures \( V^\mu_1,2,3,4 \) are more commonly known as the \( V, A_{2,3,0} \) FFs [10]. Below we summarise the status of these computations in LCSR & lattice, discuss the issue of the finite width of the \( K^\ast \)-meson and the use of the equation of motion (EOM) on eliminating uncertainties. The FFs are fitted by a \( z \)-expansion, with flat priors on the coefficients, to the lattice [48, 49] and LCSR [10] data. The latter are pseudo-data generated from the analytic LCSR computation with a Markov chain process [10]. The thereby obtained error correlation matrix reduces the 10%-uncertainty of individual FFs by a considerable amount, which is relevant for \( B \to K^\ast \ell \ell \)-decay angular distributions.

4.1 Status of LCSR and lattice computations

LCSR FFs are computed from a light-cone OPE, valid at \( q^2 \leq O(m_b \Lambda) \approx 14 \text{ GeV}^2 \) in an \( \alpha_s \)- and twist-expansion resulting in convolutions of a hard kernel and light-cone distribution amplitudes (LCDA). The LCDA are subjected to equations of motion (EOMs) and are relatively well-known to the necessary order in the conformal partial wave expansion (e.g. Gegenbauer moments). For \( B \to V (B_{(q,s)} \to (K^\ast, \phi, \rho, \omega)) \) FFs, defined in (13) the FFs are known up to twist-3 at \( O(\alpha_s^3) \) and twist-4 \( O(\alpha_s^4) \) [10, 50].\(^3\) State-of-the-art computations of \( B \to K \) LCSR FFs, for which the DAs are better known because of the absence of finite width effects, can be found in [46, 52].

Lattice QCD calculations are based on the path integral formalism in Euclidean space. The QCD theory is discretized on a finite space time lattice. Correlation functions, from which FFs can be extracted, are then obtained by solving the integrals numerically using Monte Carlo methods. Since

\[^3\] Alternatively one may use B-meson DA and an interpolating current for the FFs. E.g. [51] for a tree-level computation with therefore slightly larger uncertainties with results compatible with [10, 50].
both the noise to signal ratio of correlation functions and discretization effects increase as the momenta of hadrons increase, lattice QCD results cover the high $q^2$ region at $\sim 15 \text{ GeV}^2 \leq q^2 \leq q^2_{\text{max}}$ for $B \to V (B_{q,s} \to (K^*, \phi))$ FFs. Unquenched calculations are available for 2+1-flavor dynamical fermions [48, 49, 53] in the narrow width approximation of the vector mesons. $B \to K$ FFs can be found in [54, 55] for 2+1-flavor dynamical configurations.

### 4.2 Finite width effects

The vector meson decays via the strong force, e.g. $K^* \to K\pi$ and do therefore have a sizeable width and it is legitimate to ask how each formalism deals with this issue.

For LCSR the answer is surprisingly pragmatic in that the formalism automatically adapts to the experimental handling of the vector meson resonance. In LCSR the vector meson is described by LCDAs, which we may schematically write as

$$
\langle K^*(p, \eta) | \bar{s}(x) \gamma_\mu q(0) | 0 \rangle = m_{K^*} f_{K^*} p_\mu \int_0^1 du e^{i p \cdot x} \phi_\parallel(u) + \text{higher twist}
$$

where $p = m_{K^*} \cdot n$, valid upon neglecting the higher twist corrections $O(x^2, m_{K^*}^2)$, has been assumed. The variable $u$ is the momentum fraction of the $s$-quark in the infinite momentum frame and the DA $\phi_\parallel(u)$ parametrises the strength of the higher Gegenbauer moments (conformal spin). Since the latter contribute only about 10-15% at the numerical level the main part is effectively described by the $K^*$-meson decay constant $f_{K^*}$. Hence at the pragmatic level the issue of finite width effects is the same as to how well-defined the decay constant $f_{K^*}$ is.

Whereas the latter can be computed from QCD sum rules or lattice modulo finite width effects it is advantageous to directly extract it from experiment (e.g. $\tau \to K\pi \nu$ and cf. appendix C in Ref. [10] for a review). Consistency is ensured if in both cases, leptonic $\tau \to (K\pi)_{l=1}\nu$ and hadronic decay $B \to (K\pi)_{l=1}\ell\nu$, the experimentalist employs the same fit ansatz for the resonances and the continuous background in the p-wave ($l = 1$) $K\pi$-channel. The transversal decay constant $f_{K^*}^\perp$, is not directly accessible in experiment. It is preferably taken from the ratio $f_{K^*}^\perp / f_{K^*}^\parallel$, which can be computed from QCD sum rules or lattice QCD for which one would expect finite width effects to drop out in ratios. This is a reasonable and testable hypothesis. In conclusion a consistent treatment of the $K^*$ in experiments allows us to bypass a first principle definition of the $K^*$-meson as a pole on the second sheet and the induced error can be seen as negligible compared to the remaining uncertainty.

A fully controlled lattice QCD computation of FFs involving a vector meson needs to include scattering states. This requires much more sophisticated and expensive calculations. In existing lattice calculations of the FFs, the threshold effects are assumed to be small in the narrow width approximation. Since the $\phi$ is relatively narrow, one might expect this approximation to be better than in the case of $K^*$. For $B \to D^*$ form-factors, heavy meson chiral perturbation theory predicts 1-2% threshold effects [56, 57] which is unfortunately not indicative since it is enhanced and also outside any linear regime as compared to the broad light vector meson $\Gamma_{D^*}/\Gamma_{K^*} \sim O(10^{-2})$. An encouraging aspect is though that the continuum- and chiral-extrapolated lattice FFs obtained in [48, 49] agree with LCSR results, upon $z$-expansion extrapolation, despite the two methods having very different systematic uncertainties. Finite width effects can only be fully assessed from a systematic computation to which the Maiani-Testa no-go theorem [58] is an obstruction. The latter states that there is no simple relation between Euclidean (lattice QCD) correlators and the transition matrix elements in Minkowski space when multiple hadrons are involved in the initial or final states. The first formalism to overcome this problem is the Lellouch-Lüscher method [59] which relates matrix elements of currents in finite volume to those in the infinite volume. Further developments of this formalism can be found in, for
examples, Refs. [60–65]. Their implementations for $B \to K\pi$-type matrix elements may be hoped for in the foreseeable future.

### 4.3 The use of the equations of motion (projection on $B$-meson ground state)

Both in LCSR and lattice computations the $B$-meson is described by an interpolating current and it is therefore a legitimate question what the precision of the projection on the $B$-meson ground state is.\(^4\) Below we argue that the EOMs improve the situation.

The vector and tensor operators entering (13) are related by EOMs

\[
\langle K^*|i\partial^\nu(\bar{s}\sigma_{\mu\nu}b)|B\rangle = -(m_s \pm m_b)\langle K^*|\bar{s}\gamma_\mu b|B\rangle + i\partial_\mu\langle K^*|(\bar{s}b)|B\rangle, -2\langle K^*|\bar{s}D_\mu b|B\rangle \\
\sim T_1(q^2) \sim V_1(q^2) + \sim 0 + \sim D_1(q^2) \tag{15}
\]

where the derivative term defines a new FF $D_1(q^2)$ in analogy to $T_1(q^2)$. The EOM (15) are exact and have to be obeyed and therefore can be used as a non-trivial check for any computation. This has been done in [10] at the tree-level up to twist-4 and for the $Z$-factors describing the renormalisation of the composite operators entering the EOM (15). Furthermore at the level of the effective Hamiltonian this term is redundant, since the EOMs have been used in reducing the basis of operators. Eq.(15) therefore defines a relation between 3 FFs of which one is redundant which does not appear helpful at first. The use comes from the hierarchy $D_1 \ll T_1, V_1$ which therefore constrains the vector in terms of the tensor FF [10, 66]. At the level of the actual computation this allows us to control the correlation of the continuum threshold parameters which are a major source of uncertainty for the individual FFs. The argument is that if those parameters were to differ by a sizeable amount for the vector and tensor form factors then the exactness of (15) would impose a huge shift for the corresponding parameters of the derivative FF corresponding to an absurd violation of semi-global quark hadron duality. Since $D_1$ seems to be a FF with normal convergence in $\alpha_s$ and the twist-expansion this possibility seems absurd and therefore supports the validity of the argument. It should be mentioned that the crucial hierarchy $D_1 \ll T_1, V_1$ can be traced back to the large energy limit [30]. For more details and a plot illustrating the validity of this argument over the $q^2$-range we refer the reader to references [10, 66] and Fig.1 in [10].

On the lattice the projection on the ground state would be perfect if one could go to infinite Euclidean time and if there was no noise. In practice simulations are done at a finite $t$-interval with some noise and this sets some limitations on the projection. On the positive side these aspects are not the main sources of uncertainties in current LQCD calculations (smearing and other methods are being used to improve the projection) and are improvable with more computer power. It is conceivable that the EOMs can be used for the lattice in correlating the projection of FFs entering the same EOM.

### 5 The relevance of charm contributions (R.Zwicky)

#### 5.1 The LCSR framework

The long distance contributions can be computed with LCSR as an alternative to the heavy quark framework discussed in section 3.2. The basic long distance topologies are well-known and include the chromomagnetic operator $O_8$ (1), the weak annihilation (WA), quark-spectator loop scattering

\(^4\)In both cases this is in practice achieved by an exponential suppression of the higher states. If one was able to compute the correlation function exactly in Minkowski space then one could use the LSZ-formalism. Hence the method used can be seen as approximate methods to the latter.
(QQLS) and the charm contributions. The latter three originate from the four quark operators $O_{1,6}$ in (1). The charm contributions are presumably the most relevant since they are not CKM suppressed with sizeable Wilson coefficients and are the subject of a longer discussion in Sec. 5.2. For $B \rightarrow K^{(*)}\ell\ell$ and other final state hadrons the $O_8$ and WA contribution have been computed in [67] and [68] respectively. Both come with strong phases and the first one is found to be small whereas the second one is rather sizeable albeit CKM suppressed. The quark spectator scattering diagrams have been computed in a hybrid approach of heavy quark expansion with infrared divergences removed by an LCSR computation [68].

Whereas we agree that the heavy quark expansion provides a systematic framework in the $1/m_b$-limit for regions outside the resonance regions it has to be seen that in practice these conditions are not always met for $b \rightarrow s\ell\ell$ observables. Since LCSR are not dependent on the $1/m_b$-limit they do provide an alternative to estimate these effects. In Sec. 5.2 we argue that the $1/m_b$ vertex corrections might well be sizeable, contributing to potential tension in $b \rightarrow s\ell\ell$-observables. Resonant effects can be systematically parameterised and fitted for but their inclusion into predictions requires further thought and work when a sizeable continuum contributions is present.

5.2 The charm contribution

In FCNC processes of the type $B \rightarrow K^{(*)}\ell\ell$, the subprocess $B \rightarrow K^{(*)}(\bar{c}c \rightarrow \gamma^* \rightarrow \ell\ell)$ is numerically relevant since it proceeds at tree-level. Hence the Wilson coefficients of operators of the type $O_{1,2}^{(c)}$ in (1) in the effective Hamiltonian are sizeable. The other relevant aspect is that the four momentum invariant, $q^2 \in [4m_B^2, (m_{B} - m_{K^{(*)}})^2]$, of the lepton-pair takes on values in the region of charmonium resonances. Hence the process is sensitive to resonances, with photon quantum numbers $J^{PC} = 1^{--}$, and one can therefore not entirely rely on a partonic picture. It is customary to divide the $q^2$-spectrum into three regions (cf Fig. 4). A region sufficiently well-below the first charmonium resonance at $q^2 \approx 9.6$ GeV$^2$, the region of the two narrow charmonium resonances $J/\psi$ and $\psi(2S)$ and the region of broad charmonium states above the $D\bar{D}$-threshold $q^2 \approx 14$ GeV$^2$. The crucial question is when and how partonic methods are applicable in the hadronic charmonium region. We discuss them below in the order of tractability in the partonic picture.

In the “below-charmonium region” partonic methods (i.e. $\alpha_s$-expansion) are expected to be applicable. Since this is the region where the particles are fast in the $B$-rest frame, the physics can be described within a light-cone formalism. The order $O(\alpha_s^0)$ contribution is equivalent to naive factorisation (NF) which means that the amplitude factorises as follows

$$\mathcal{A}([B \rightarrow K^{(*)}\ell\ell]_{O_{1,2}^{(c)}}[NF] \sim h(q^2)F_{B\rightarrow K^{(*)}}(q^2),$$

where $F_{B\rightarrow K^{(*)}}(q^2)$ stands for the relevant form factor (FF) combination and $h(q^2)$ is the vacuum polarisation due to the charm-part of the electromagnetic current. At this formal level the light-cone aspect is only present in that the preferred methodology for evaluating the FFs are light-cone methods (cf. Sec. 4.1). Corrections of order $O(\alpha_s)$ are expected to be sizeable since the leading order contribution is of the colour-suppressed Wilson coefficient combination. Simple factorisation formulae only hold in the $1/m_b$-limit which include hard spectator and vertex corrections [25, 26] with the latter borrowed from inclusive computations [69]. The $O(\alpha_s/m_b)$ corrections can be done within an extended LCSR approach with either a $B$- or $K^{(*)}$-meson LCDA. Only partial results exist in that the vertex and hard spectator corrections have not been evaluated in either of these methods. The third contribution, emission of the gluon into the LCDA, have been done for $K^{(*)}$- and $B$-meson [70, 71] (computations for $q^2 = 0$ only) and [46] respectively. The reason that the vertex and hard spectator corrections have not
Figure 4. Illustration of the spectrum in $q^2$, the lepton pair invariant momentum, for the $B \to K \ell \ell$ branching fraction. The three regions referred to in the text are the “below resonance region” (partonic, low-$q^2$), “narrow resonance region” and the “broad resonance region” (high-$q^2$).

been performed in either approach is that they are technically demanding. If one insists on verifying the dispersion relation involved this demands an evaluation of a two-loop graph with five scales and an integration over one LCDA-parameter. The size of these contributions is therefore unknown beyond the heavy quark limit and might well be sizeable enough to explain the current picture of deviations even within a partonic approach. A completion of this program is therefore desirable.

The broad charmonium region (see Fig. 4 on the right) is characterised by a considerable interference of the short-distance and charmonium long-distance part. The charmonium resonances $\psi(3770)$, $\psi(4040)$, $\psi(4160)$ and $\psi(4415)$ are broad because they can decay via the strong force into $\bar{D}D$-states. The resonances are necessarily accompanied by a continuum of $\bar{D}D$-states as is the case in $e^+e^-\to$ hadrons. [40]. A sketch of a realistic parametrisation is given by\footnote{More realistic parameterisation differ in two aspects. Firstly, they only parameterise the discontinuity and the remaining part is obtained by a dispersion relation e.g. $e^+e^-\to$ hadrons [72] and $B \to K \ell \ell$ [40]. Second they include energy dependent effects and interferences of the overlapping broad resonances. This has been successfully done in $e^+e^-\to$ hadrons [72] and adapted to $B \to K \mu \mu$ in [40]. The inclusion of interference effects reduces the $\chi^2$/d.o.f. from $\approx 1.4$ to $\approx 1$. Crucially the degree of model dependence is justified by the goodness of the fit.}

$$\mathcal{A}(B \to K^{(*)}\ell\ell)|_{\psi,\ell} \approx \sum_{\psi=J/\Psi, \phi(2S), \psi(3770),...} \frac{r_\psi}{q^2 - m_\psi^2 + i m_\psi \Gamma_\psi} + f_{\bar{c}c}(q^2), \quad (17)$$

where the residues $r_\psi$ and the non-resonant $\bar{c}c$-continuum function $f_{\bar{c}c}$ are the unknowns which usually have to be determined experimentally. For $e^+e^-\to$ hadrons, which correspond effectively to NF (16), $r_\psi^{NF} \sim \Gamma(\psi \to \ell\ell) F^{B \to K^{(*)}}(m_\psi^2) > 0$ and $\text{Im}[f_{\bar{c}c}(q^2)] > 0$ for which successful parametrisations are easily found and the real part follows from a dispersion relation [40, 72]. Going beyond factorisation involves determining the factors $\eta_\psi$ and the continuum function $f_{\bar{c}c}(q^2)$

$$\eta_\psi \equiv |\eta_\psi|e^{i \delta_\psi} = \frac{r_\psi}{r_\psi^{NF}} \quad f_{\bar{c}c}(q^2). \quad (18)$$
The strong-phase $\delta_\phi$ is defined relative to the FF contribution. The determination of $f_{cc}(q^2)$ is a difficult task since the $q^2$-dependence is similar to the one of the short distance contributions. On the other hand the shape of the resonances is very distinct (e.g Fig. 4) and can therefore be fitted unambiguously. This has been done in [40] using the LHCb-data [8]. A surprisingly good fit is obtained by hand the shape of the resonances is very distinct (e.g Fig. 4) and can therefore be fitted unambiguously.

After this excursion let us discuss the practicalities for the low-$q^2$ (“below resonance region”) and the high-$q^2$ (“broad resonance region”) regions where phenomenologists and experimentalists compare predictions to measurements in the hope of seeing physics beyond the Standard Model. It would be desirable to obtain a coherent picture of the partonic and hadronic descriptions in both of these regions in order to validate the approaches. Both cases need more work. At low-$q^2$, as discussed above, the partonic contributions are not very complete. For example, the potentially sizeable vertex corrections of the charm loop are unknown beyond the $1/m_b$-limit. There is no universal or well-understood pattern for estimating the size of the $1/m_b$-corrections. For example the FF $1/m_b$-corrections are around 10%, for the $B$-meson decay constant $f_B$ they are 30% whereas for the $O_8$-matrix elements (chromomagnetic operator) they are 50% and more [67]. As for the hadronic data, further work is needed in order to determine the strong phases $\delta_{J/\psi,\phi(2S)}$ as well as the continuum function $f_{cc}(q^2)$. Before moving on it should be mentioned that the hadronic fits should and will be extended from $B \to K \ell\ell$ to $B \to K^* \ell\ell$ by the LHCb-collaboration. In the high-$q^2$ region a partonic picture has been advocated, known as the high-$q^2$ OPE, where one resorts to an expansion in $1/m_b$ and $1/\sqrt{q^2}$ supplemented by charm-contributions [38, 39]. The initial idea was to include charm contributions in an $\alpha_s$-expansion, relying on cancellations when averaged over large enough bins. As an estimate of these type of quark-hadron duality violations for the broad resonances, naive factorisation (i.e. $e^+e^- \to hadrons$) was taken as guidance [39], which suggests an effect of the order of 2%. One is then faced with the a posteriori fact that the actual data [8] show effects in the region of 10% [40]. Hence supplementing the high-$q^2$ OPE with a hadronic representation (17) seems attractive. The bottleneck is though the determination of the the continuum function $f_{cc}(q^2)$ (17). A promising direction could be the experimental investigation of the $B \to \bar{D}DK^{(*)}$ modes.

---

6Cf. Ref. [40] for more realistic fits, with variable individual complex residues, with slightly improved $\chi^2$/dof. These findings have recently been confirmed by the LHCb collaboration [24].

7 The gateway to quark-hadron duality are dispersion relations valid at the level of amplitudes. Averaging at the decay rate level is only valid if the rate can be written as an amplitude which is the case for inclusive modes such as $e^+e^- \to hadrons$. Averaging over the entire $B \to K^{(*)}\ell\ell$-rate, which includes the narrow resonances, fails by several orders of magnitudes as discussed and clarified in [73].

8 For angular observable effects might look though more favourable in two aspects partly at the cost of sensitivity to new physics. Firstly, in the limit of no right-handed currents and constant $\eta_{\phi}$ resonance effects drop out in a few angular observables [66]. Second at the kinematic endpoint $q^2 = (m_B - m_{cc})^2$ the observables approach exact values based on Lorentz-covariance only (i.e. valid in any model and approximation which respects Lorentz invariance) and show a certain degree of universality near the endpoint region [74].
2.3 Discussion and summary

Clarifying the role of the charm will remain an outstanding task before angular anomalies of the \( B \to K^* \ell^\pm \ell^- \)-type and branching fraction deviations, not related to violation of lepton flavour universality (LFU), can be considered to be physics beyond the Standard Model. Progress can be made by computing charm contributions consistently in one approach, complemented by more refined experimental data allowing us to extract the relevant information. Hopefully this will pursued by experimentalists and theorists.

Finally, a few brief comments on LFU-violation, the possibility of new physics in charm and right-handed currents. Clarifying the role of LFU-violating observables, e.g. \( R_K \) (5), is of importance for global fits. QED corrections, which can give rise to \( O(a_{\text{QED}} \ln^2(m_\ell/m_e)) \)-deviations from \( R_K = 1 \), can be diagnosed by higher moments [75]. The dominant real radiation effect has been estimated in [76] and crucially the effect deviates by \( O(1\%) \) as compared to the PHOTOS-implementation [13] used by the LHCb-analysis. Migration of narrow charmonium events into lower \( q^2 \)-region by misidentification of hadronic final states [75] has been checked by the LHCb-collaboration. The fact that charm contributions can mimic shifts in \( C_9 \), or any other amplitude with photon quantum numbers, has been emphasised in [40] motivated by the results on the broad charmonium resonances in \( B \to K \ell \ell \). In this work it was suggested to diagnose this effect with \( q^2 \)-dependent fits in various \( b \to s \ell \ell \)-channels, and the question as to whether the anomalies could partially be due to new physics in charm (e.g. \( \bar{b}c\bar{c}s \)-operators) was raised. The former idea was applied to \( B \to K^* \ell^\pm \ell^- \) at low \( q^2 \) for the first time in [77], and the possibility of charm in new physics was recently investigated systematically including RG-evolution and constraints (e.g. \( B_s \)-mixing) in [78]. A closely related area where charm contributions are of importance is the search for right-handed currents. The V-A structure of the weak interactions and the small ratio \( m_s/m_b \) suppress amplitudes with right-handed quantum numbers that can, for instance, be measured in time-dependent CP-asymmetries [79]. Charm contributions, of higher-twist, are a non-perturbative background to these measurements, whose understanding is important. LHCb’s first time measurement of the \( B_s \to \phi \gamma \) time-dependent CP-asymmetry \( \mathcal{A}_\Delta = -0.98(50)(20) \) [80] comes with a large uncertainty but also with a large deviation from the SM prediction \( \mathcal{A}_\Delta \approx 0.047(28) \) [71], which allows for speculations in various directions.

6 Global fits (L. Hofer)

As reported in Sec. 2, experimental results on \( B \to K^* \mu^+\mu^- \), \( B_s \to \phi \mu^+\mu^- \) and \( R_K = Br(B \to K \mu^+\mu^-)/Br(B \to K e^+e^-) \) show deviations from the SM at the 2–3\( \sigma \) level. Although none of these tensions is yet significant on its own, the situation is quite intriguing as the affected decays are all mediated by the same quark-level transition \( b \to s \ell^+\ell^- \) and thus probe the same high-scale physics. A correlated analysis of these channels can shed light on the question whether a universal new-physics contribution to \( b \to s \ell^+\ell^- \) can simultaneously alleviate the various tensions and lead to a significantly improved global description of the data.

At the energy scale of the \( B_{(s)} \) decays, any potential high-scale new physics mediating \( b \to s \ell^+\ell^- \) transitions can be encoded into the effective couplings \( C_{7,9,10}^{(r)} \) multiplying the operators

\[
\begin{align*}
O_9^{(r)} &= \frac{\alpha}{4\pi} [\bar{s}\gamma^\mu P_{L(R)} b][\bar{\mu}\gamma_\mu \mu], \\
O_7^{(r)} &= \frac{\alpha}{4\pi} m_b [\bar{s}\sigma_{\mu\nu} P_{R(L)} b] F^{\mu\nu},
\end{align*}
\]

where \( P_{L,R} = (1 \mp \gamma_5)/2 \) and \( m_b \) denotes the \( b \) quark mass. Whereas the above-mentioned semi-leptonic decays are sensitive to the full set \( C_{7,9,10}^{(r)} \) of effective couplings, the decay \( B_s \to \ell^+\ell^- \) only probes \( C_{10}^{(r)} \).
Various groups have performed fits of the couplings $C_{7,9,10}^{(i)}$ to the data [9, 81–84]. The obtained results are in mutual agreement with each other and confirm the observation, pointed out for the first time in Ref. [82] on the basis of the 2013 data, that a large negative new-physics contribution $C_9$ yields a fairly good description of the data. This is illustrated in Tab. 1, where selected results from Ref. [81] for one-parameter fits of the couplings $C_{7,9,10}^{(i)}$ are displayed. Apart from the best-fit point together with the $1\sigma$ region, the tables feature the SM-pull of the respective new-physics scenarios. This number quantifies by how many sigmas the best fit point is preferred over the SM point ($C_{NP}^{NP} = 0$) and thus measures the capacity of the respective scenario to accommodate the data. The table on the left demonstrates that a large negative new-physics contribution $C_9$ is indeed mandatory to significantly improve the quality of the fit compared to the SM. It is particularly encouraging that the individual channels tend to prefer similar values for $C_9$, as shown in the table on the right.

The results of the fit are quite robust with respect to the hadronic input and the employed methodology. This can be seen from the good agreement between the results of the analyses AS [9] and DHMV [81] which use approaches that are complementary in many respects:

- AS choose the observables $S_i$ as input for the fit to the angular distributions of $B \to K^+\ell^+\ell^-$ and $B_s \to \phi\ell^+\ell^-$ and restrict their fits in the region of large $K^*$-recoil to squared invariant dilepton masses $q^2 \in [0, 6]$ GeV$^2$. DHMV, on the other hand, choose the observables $P_i^{(i)}$, which feature a reduced sensitivity to the non-perturbative form factors, and include all bins up to $q^2 = 8$ GeV$^2$.

- AS use LCSR form factors from Ref. [10], while DHMV mainly resort to the LCSR form factors from Ref. [46].

- In the analysis of AS, correlations among the form factors are implemented on the basis of the LCSR calculation [10], whereas in the analysis of DHMV they are assessed from large-recoil symmetries supplemented by a sophisticated estimate of symmetry-breaking corrections of order $O(\Lambda/m_b)$. The pros and cons of these two methods complement each other: The first approach provides a more complete access of correlations at the price of a dependence on and the limitation to one particular LCSR calculation [10] and its intrinsic model-assumptions. The second approach determines the correlations in a model-independent way from first principles but needs to rely on an estimate of subleading non-perturbative $\Lambda/m_b$ corrections.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Best fit</th>
<th>1σ</th>
<th>PullSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{7,9,10}^{NP}$</td>
<td>$-0.02$</td>
<td>$[-0.04, -0.00)$</td>
<td>$1.2$</td>
</tr>
<tr>
<td>$C_{9}^{\text{SM}}$</td>
<td>$-1.09$</td>
<td>$[-1.29, -0.87]$</td>
<td>$4.5$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}}$</td>
<td>$0.56$</td>
<td>$[0.32, 0.81]$</td>
<td>$2.5$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}}$</td>
<td>$0.02$</td>
<td>$[-0.01, 0.04]$</td>
<td>$0.6$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}}$</td>
<td>$0.46$</td>
<td>$[0.18, 0.74]$</td>
<td>$1.7$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}}$</td>
<td>$-0.25$</td>
<td>$[-0.44, -0.06]$</td>
<td>$1.3$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}}$</td>
<td>$-0.22$</td>
<td>$[-0.40, -0.02]$</td>
<td>$1.1$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}} = C_{9}^{\text{SM}}$</td>
<td>$-0.68$</td>
<td>$[-0.85, -0.50]$</td>
<td>$4.2$</td>
</tr>
<tr>
<td>$C_{9}^{\text{NP}} = -C_{9}^{\text{NP}}$</td>
<td>$-1.06$</td>
<td>$[-1.25, -0.85]$</td>
<td>$4.8$</td>
</tr>
</tbody>
</table>

Table 1. Left: best-fit point, $1\sigma$ region and SM-pull for 1-parameter fits allowing new physics only in one of the couplings $C_{7,9,10}^{(i)}$. Right: fits of a new-physics contribution to the effective coupling $C_9$ using different subsets of the experimental data as input. Results are taken from Ref. [81].
The measurement of $R_K \neq 1$ hints at a possible violation of lepton-flavour universality (LFUV) and suggests a situation where the muon- and the electron-components of the operators $C_{9,10}^{\prime}$ receive independent new-physics contributions $C_{i,\mu}^{\prime \prime}$ and $C_{i,e}^{\prime \prime}$, respectively. In Fig. 5 on the left we display the result for the two-parameter fit to the coefficients $C_{9,\mu}^{NP}$ and $C_{9,e}^{NP}$. The plot is taken from Ref. [81], similar results are obtained in Refs. [9, 84]. The fit prefers an electron-phobic scenario with new physics coupling to $\mu^+\mu^-$ but not to $e^+e^-$. Under this hypothesis, that should be tested by measuring $R_{K^*\phi}$ as well as lepton-flavour sensitive angular observables [85] \(^9\), the SM-pull increases by ~ 0.5$\sigma$ compared to the value in Tab. 1 for the lepton-flavour universal scenario, except for the scenario with $C_{9}^{NP} = -C_{9}^{NP}$ where the value remains unchanged due to the absence of any contribution to $R_K$.

The fact that it is primarily the variation of the coefficient $C_9$ which is responsible for solving the anomalies unfortunately spoils an unambiguous interpretation of the fit results in terms of new physics. The reason is that precisely this effective coupling can be mimicked by non-perturbative charm loops, as discussed in Sec. 5. However, whereas these non-local effects are expected to introduce a non-trivial dependence on the squared invariant mass $q^2$ of the lepton pair, a high-scale new-physics solution would necessarily generate a $q^2$ independent $C_{9}^{NP}$. A promising strategy to resolve the nature of a potential non-standard contribution to the effective coupling $C_9$ thus consists in the investigation of its $q^2$ dependence. To this end, two different methods have been pursued so far: In Ref. [77], the authors performed a bin-by-bin fit of $C_9$ to check whether results in different bins were consistent with each other under the hypothesis of a $q^2$-independent $C_{9}^{NP}$. Their conclusion, which was later confirmed also in Ref. [81] with the plot shown in Fig. 5 on the right, was that there is no indication for a $q^2$-dependence, though the situation is not conclusive due to the large uncertainties in the single bins.

An alternative strategy to address this question has been followed recently in Refs. [44, 47] where a direct fit of the $q^2$-dependent charm contribution $C_{9}^{NP}(q^2)$ to the data on $B \rightarrow K^*\mu^+\mu^-$ (at low $q^2$) has been performed under the hypothesis of the absence of new physics. The results are in agreement with the findings from Fig. 5: in Ref. [44] it was shown that the inclusion of additional terms

\(^9\)Very recently, Belle has presented a separate measurement [86] of $P_5$ in the muon and electron channels. While the muon channel exhibits a 2.6$\sigma$ deviation with respect to the SM prediction in good agreement with the LHCb measurement, the electron channel agrees with the SM expectation at 1.3$\sigma$. 

Figure 5. Left: Fit allowing for LFUV by means of independent coefficients $C_{9,\mu}^{NP}$ and $C_{9,e}^{NP}$. Right: Bin-by-bin fit of the one-parameter scenario with a single coefficient $C_{9}^{NP}$. Reproduced from Ref. [81].
parametrising a non-trivial $q^2$-dependence does not improve the quality of the fit. On the other hand, current precision of the experimental data does not allow to exclude non-zero values for these terms.

In certain scenarios, a $q^2$-dependent contribution to $C_9$ can also have its origin at high energy scales: new physics mediating $b \to s\bar{c}\bar{c}$ transitions would induce a $q^2$-dependent contribution to $b \to s\ell^+\ell^-$ at the one-loop level in the effective theory. This possibility was proposed for the first time in Ref. [40], while a phenomenological analysis taking into account constraints from $B_s - \bar{B}_s$ mixing and $B \to X_s\gamma$ was performed recently in Ref. [78]. Note that a charm-loop contribution to $b \to s\ell^+\ell^-$, whether from high-scale new physics or from low-energy QCD dynamics, always conserves lepton flavour and thus could not account for deviations in $R_K$ or other LFUV observables.

7 BSM interpretation (L. Hofer)

As we have seen in the previous section, the observed anomalies in $b \to s\ell^+\ell^-$ decays show a coherent picture and allow for a solution at the level of the effective Hamiltonian by NP contributions to the operators $O^{(r)}_{9,10}$. At tree level, contributions to these operators can be mediated by exchange of a heavy neutral vector-boson $Z'$ (e.g. [82, 87–98]), or by scalar or vector lepto-quarks (e.g. [99–104]). At one loop, they can be generated by box diagrams involving new particles (e.g. [105–107]) or by $Z'$ penguins (e.g. [108]). The step beyond the model-independent analysis allows to attempt a common explanation of the $b \to s\ell^+\ell^-$ anomalies with other tensions in flavour data, like $R_{D^*}$, or the long-standing anomaly in the anomalous magnetic moment of the muon. It further permits to study the viability of the various model classes in the light of constraints from other flavour observables and from direct searches. In the following, we will briefly summarize typical $Z'$ and lepto-quark scenarios, and discuss bounds from $B_s - \bar{B}_s$ mixing and direct searches.

7.1 $Z'$ models

The interaction of a generic $Z'$ boson with the SM fermions is described by the Lagrangian

$$\mathcal{L}_{Z'} = \sum_{ff'} \Gamma^L_{ff'} \bar{f} \gamma^\mu P_L f' Z'_\mu + \Gamma^R_{ff'} \bar{f} \gamma^\mu P_R f' Z'_\mu + \text{h.c.},$$

(20)

where the sum is over fermions $f, f'$ with equal electric charges. The exact form of the couplings $\Gamma^L_{ij}$ depends on the $U(1)'$ charges assigned to the SM fermions and on a potential embedding of the $Z'$ in a more fundamental theory. Note, however, that SU(2)$_L$ invariance implies the model-independent relations $\Gamma^L_{uu} = V_{ud}^* \Gamma^L_{dd}$, $V_{ud}^* \Gamma_{\ell\ell}'$ and $\Gamma^L_{\ell\ell}' = \Gamma^L_{\ell\ell'}$ (with $V$ denoting the CKM matrix).

The Wilson coefficients $c^{(r)}_{9,10}$ are generated by tree-level $Z'$ exchange involving products of couplings $\Gamma^L_{qq} \Gamma^{LR}$ $\Gamma^L_{RLL'}$. Since only three out of these four products are independent, the relation $C_9 \cdot C_{10} = C_9' \cdot C_{10}$ is fulfilled in models with a single $Z'$ boson. In order to generate a non-vanishing coupling $C_{9,10}$, mandatory for a solution of the $b \to s\ell^+\ell^-$ anomalies, the couplings $\Gamma^L_{bs}$ and $\Gamma^L_{\mu\mu} + \Gamma^R_{\mu\mu}$ need to have non-vanishing values.

The most popular class of $Z'$ models is based on gauging $L_r - L_l$ lepton number [90, 93, 94, 109, 110]. This pattern of $U(1)'$ charges avoids anomalies and is well-suited to generate the measured PMNS matrix. The vanishing coupling of the $Z'$ to electrons allows to explain LFUV in $R_K$ and helps to avoid LEP bounds on the $Z'$ mass $M_{Z'}$. The symmetry can be extended to the quark sector with a flavour non-universal assignment of $U(1)'$ charges that induces the off-diagonal couplings $\Gamma^L_{bs}$ (e.g. [91, 94]). An alternative mechanism to generate the couplings $\Gamma^L_{bs}$ consists in the introduction of additional vector-like quarks that are charged under the $U(1)'$ symmetry and that generate an effective $bsZ'$ coupling via their mixing with the SM fermions [90, 93, 109, 111].
Several \( Z' \) scenarios have been proposed that are capable of solving not only the \( b \rightarrow s\ell^+\ell^- \) anomalies but at the same time also other tensions in the data. Embedding the \( Z' \) in a SU(2)\(^\prime\) gauge model allows to address the anomalies in \( R_{D^{0}} \) with a tree-level contribution to \( b \rightarrow c\ell^+\bar{\nu} \) mediated by the \( W' \)-boson (e.g. \cite{97, 112}). It is also possible to solve the anomaly in \( (g - 2)_\mu \) in a \( Z' \) scenario, provided the \( Z' \) coupling to muons is generated at the loop-level so that both the NP contributions to \( b \rightarrow s\ell^+\ell^- \) and \( (g - 2)_\mu \) are loop-suppressed \cite{108}.

7.2 Lepto-quark models

Lepto-quarks are new particles \( \Delta_k \) beyond the SM that couple leptons to quarks via vertices \( \ell_i q_j / \Delta_k \). Different lepto-quark models can be classified according to the spin of the lepto-quarks and their quantum numbers with respect to the SM gauge groups. Since the \( \ell_i q_j \Delta_k \) couplings violate lepton-flavour, lepto-quark models are excellent candidates to explain LFUV observables like \( R_\ell \) and \( R_{D^{0}} \). Indeed, various representations of lepto-quarks have been studied with respect to their capability of accommodating the measured values of \( R_\ell \) and \( R_{D^{0}} \) by tree-level lepto-quark contributions \cite{99–104}. In Ref. \cite{106} it was further proposed that an SU(2)\(_L\) singlet scalar lepto-quark could explain \( R_{D^{0}} \) by a tree-level and \( R_K \) by a loop contribution. This possibility was later shown in Ref. \cite{113} to be challenged by other flavour data.

7.3 Constraints from \( B_s - \bar{B}_s \) mixing and direct searches

A NP model generating \( b \rightarrow s\ell^+\ell^- \) necessarily also contributes to \( B_s - \bar{B}_s \) mixing. In lepto-quark models, \( b \rightarrow s\ell^+\ell^- \) is typically mediated at tree level, while \( B_s - \bar{B}_s \) mixing contributions are loop-suppressed and thus do not pose relevant constraints. In \( Z' \) models, on the other hand, both processes are usually generated by tree-level exchange of the \( Z' \) boson. The constraint on \( |\Gamma_{bs}^{L,R}/M_{Z'}| \) from \( B_s - \bar{B}_s \) mixing then imposes a lower limit of typically \( |\Gamma_{\mu\mu}^{L,R}/M_{Z'}| \gtrsim 0.3/(1 \text{ TeV}) \) that needs to be reached for a solution of the \( b \rightarrow s\mu^+\mu^- \) anomalies. In models with box contributions to both \( b \rightarrow s\ell^+\ell^- \) and \( B_s - \bar{B}_s \) mixing, the analogous constraint is more severe due to the loop suppression: \( |\Gamma_\mu|/\sqrt{M_\Phi} \gtrsim 3/\sqrt{1} \text{ TeV} \) where \( M_\Phi \) denotes the mass scale of the new particles in the box and \( \Gamma_\mu \) their coupling strength to the muon. It was shown in Ref. \cite{107} that this bound can be relaxed in a scenario with Majorana fermions in the box where the additional crossed boxes lead to a negative interference in \( B_s - \bar{B}_s \) mixing.

Bounds from direct searches can be avoided to a large extent if the new physics couples only to the second and third fermion generation, in line with LFUV in \( R_K \) and \( R_{D^{0}} \). Collider signals are then limited to more complex final state, like e.g. \( pp \rightarrow 4\mu \) probing the muon-coupling of a possible \( Z' \) boson, or to suppressed production channels, like e.g. \( b\bar{s} \rightarrow \mu^+\mu^- \). However, it was found \cite{114} very recently that the data from Atlas/CMS already now heavily constrains \( Z' \) and lepto-quark scenarios even in the \( b\bar{b} \rightarrow \tau^+\tau^- \) channel: a solution of \( R_{D^{0}} \) by SU(2)\(^\prime\) gauge bosons \( W'/Z' \) is restricted to masses \( M_{Z'} \lesssim 500 \text{ GeV} \), and a solution via vector lepto-quarks is about to be excluded. The interplay with high-\( p_T \) searches will thus definitely play a crucial role in the quest for an explanation of the flavour anomalies.

8 Summary

The discovery of the leptonic decay \( B^0_d \rightarrow \mu^+\mu^- \) by the CMS and LHCb collaborations was a major breakthrough of precision flavour physics with data from Run 1 of the Large Hadron Collider. Since then, several anomalies have emerged in semileptonic decays. These indicate a potential violation of lepton universality in the decays \( B \rightarrow K^{(*)}\mu^+\mu^- \) and \( B \rightarrow K^{(*)}e^+e^- \) with a statistical significance of
In the angular analysis of $B \rightarrow K^+\mu^+\mu^-$ decays, tensions with the SM are seen in the longitudinal polarisation of the $K^+$ and the angular observable $P'_5$. The largest discrepancy with more than $3\sigma$ is seen in the differential branching fraction of the decay $B^0_s \rightarrow \phi \mu^+\mu^-$. The dynamics of $B \rightarrow K^+\mu^+\mu^-$ decays can be described by a set of helicity amplitudes, which, in the effective Hamiltonian formalism, split into Wilson coefficients and matrix elements of local operators. Deviations in the Wilson coefficient $C_9$, which is sensitive to new physics as well as long-distance QCD, can explain the experimental data. The most sensitive observables to $C_9$ are the $q^2$ dependence of the forward-backward asymmetry and the angular observable $P'_5$. These observables have been constructed such that they exhibit a reduced sensitivity to hadronic form factors, though a remnant dependence at order $\Lambda/m_b$ cannot be avoided and its impact cannot be predicted in the heavy quark framework. A combination of light-cone sum rule calculations with more precise measurements may be able to address this issue.

In light-cone sum rule computation finite width effects can effectively be bypassed if the vector mesons are treated consistently in all experiments, which includes those from where input is taken for form factor calculations as well as those where the form factor computations are used. Whereas current lattice QCD studies do not include finite width effects, recent developments indicate that this may change in the foreseeable future. The use of equation of motion reduces the uncertainty of the projection on the $B$-meson state for ratios of form factors in light-cone sum rules. It is conceivable that the use of equation of motion might help to further improve lattice QCD computation as well.

The main focus of recent discussions have been the so-called charm contributions, which describe the sub-process $B \rightarrow K^{(*)}(\bar{c}c\rightarrow \gamma^* \rightarrow \ell\ell)$. The data are typically studied in different regions of $q^2$: the ‘partonic’ well below the $J/\psi$ resonance, the ‘narrow’ between the $J/\psi$ and $\psi(2S)$ resonances, and the ‘broad’ in the high-$q^2$ region that is dominated by broad charmonium resonances. Ideally, a coherent description of the low and high $q^2$ regions should be obtained. Overcoming these challenges requires close collaboration of experimentalists and theorists to pursue new approaches such as a detailed study of the decays $B \rightarrow D\bar{D}K^{(*)}$.

Global fits aim to exploit a maximum of information of the range of observables in the framework of an effective theory. This allows the splitting of the Wilson coefficients in a SM part and a component to encapsulate effects beyond the SM. Most fits favour a non-SM value of the coefficient $C_9$ of about $-1$ with the pull of the SM scenario exceeding $4\sigma$ in several cases. As contributions from high scales beyond the SM create $q^2$-independent effects, it is instructive to perform fits in several regions of $q^2$ and test for $q^2$-dependent effects that would indicate low-scale SM effects. At the current level of precision these tests are consistent with a $q^2$-independent shift of $C_9$.

Possible explanations involving particles beyond the SM exist in the form of lepto-quarks or $Z'$ bosons, typically mediating the $b \rightarrow s\ell^+\ell^-$ transitions through tree-level exchange. In addition, these SM extensions have the potential to simultaneously accommodate other anomalies like $R_{D^{(*)}}$, or the anomalous magnetic moment of the muon. Direct searches pose tight constraints on some of these models and are expected to either probe or severely challenge them in the near future.

The number of flavour anomalies that appear to fit a common picture is intriguing. The analysis of data taken during the ongoing Run 2 of the LHC will yield powerful new insight both into the observables of interest and into new strategies to control uncertainties. The interpretation of these results requires close collaboration with theory, where advances are required in several areas for which promising strategies exist.

References

[18] R. Aaij et al. (LHCb collaboration), JHEP 02, 104 (2016), 1512.04442
[20] Measurement of the P1 and P′ 5 angular parameters of the decay B0 → K*0µ+µ− in proton-proton collisions at √s = 8 TeV, CMS-PAS-BPH-15-008
[41] J. Matias, F. Mescia, M. Ramon, J. Virto, JHEP 04, 104 (2012), 1202.4266