Towards excluding a light $Z'$ explanation of $b \to s\ell^+\ell^-$

Andreas Crivellin and Claudio Andrea Manzari
Physik-Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland and Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland

Wolfgang Altmannshofer
Santa Cruz Institute for Particle Physics and Department of Physics, University of California, Santa Cruz, CA 95064, USA

Gianluca Inguglia and Paul Feichtinger
Institute of High Energy Physics, 1050, Vienna, Austria

The discrepancies between $b \to s\ell^+\ell^-$ data and the corresponding Standard Model predictions constitute the most significant hints for new physics (at the TeV scale or below) currently available. In fact, many scenarios that can account for these anomalies have been proposed in the literature. However, only a single light new physics explanation, including also muon-specific observables, like the branching ratio of $B_s \to \phi\mu^+\mu^-$ and angular observables in $B \to K^{(*)}\mu^+\mu^-$, have been proposed in the literature [6–10]. Furthermore, global analyses including also muon-specific observables, like the branching ratio of $B_s \to \phi\mu^+\mu^-$ and angular observables in $B \to K^{(*)}\mu^+\mu^-$, show preferences compared to the SM hypothesis with pulls of up to even more than 7σ, depending on theoretical assumptions and data included in the fits [9, 11–13].

Because the $b \to s\ell^+\ell^-$ anomalies constitute such tantalizing hints for NP, a plethora of SM extensions have been proposed in the literature, including leptoquarks [14–16], models with loop effects of new scalars and fermions [13–19] and in particular models with new neutral gauge bosons, i.e. $Z'$s [50–107]. Because all these solutions, except the $Z'$ one, involve charged particles they must be realized at the electroweak scale, or even significantly above, due to the constraints from direct searches. However, a $Z'$ boson can be light and, in fact, such solutions to the $b \to s\ell^+\ell^-$ anomalies have been proposed and studied in the literature [108–115]. Importantly, this is the only NP model addressing the $b \to s\ell^+\ell^-$ anomalies which predicts $R(K^{(*)}) > 0$ in the high $q^2$ bins, i.e. above the charm resonances, where precise experimental data is still missing but expected in the near future.

In recent years, multiple hints for the violation of lepton flavour universality (LFU), which is satisfied by the Standard Model (SM) gauge interactions, have been accumulated (see Refs. [11, 2] for recent reviews). Among them, the discrepancies between $b \to s\ell^+\ell^-$ data and the corresponding SM predictions are statistically most significant (see Refs. [3, 5] for an overview). Combining the current measurements of the LFU ratios $R(K^{(*)})$ one observes that several new physics (NP) scenarios are statistically preferred over the SM hypothesis with significances close to 5σ [6, 10]. Furthermore, global analyses including also muon-specific observables, like the branching ratio of $B_s \to \phi\mu^+\mu^-$ and angular observables in $B \to K^{(*)}\mu^+\mu^-$, like $P_\ell^\mu$, show preferences compared to the SM hypothesis with pulls of up to even more than 7σ, depending on theoretical assumptions and data included in the fits [9, 11–13].

Because the $b \to s\ell^+\ell^-$ anomalies constitute such tantalizing hints for NP, a plethora of SM extensions have been proposed in the literature, including leptoquarks [14–16], models with loop effects of new scalars and fermions [13–19] and in particular models with new neutral gauge bosons, i.e. $Z'$s [50–107]. Because all these solutions, except the $Z'$ one, involve charged particles they must be realized at the electroweak scale, or even significantly above, due to the constraints from direct searches. However, a $Z'$ boson can be light and, in fact, such solutions to the $b \to s\ell^+\ell^-$ anomalies have been proposed and studied in the literature [108–115]. Importantly, this is the only NP model addressing the $b \to s\ell^+\ell^-$ anomalies which predicts $R(K^{(*)}) > 0$ in the high $q^2$ bins, i.e. above the charm resonances, where precise experimental data is still missing but expected in the near future.

While a light $Z'$ explanation of $b \to s\ell^+\ell^-$ data is experimentally well constrained, it still remains viable if it is assumed that the $Z'$ decays dominantly to invisible final states. This avoids direct searches such as $e^+e^- \to 4\mu$ and provides the sizable width necessary for the $Z'$ to affect multiple $q^2$ bins in $b \to s\ell^+\ell^-$ observables and thus give a good fit to data. However, also processes with invisible final states are constrained experimentally, such as the di-muon invariant mass distribution in Drell-Yan production close to the $Z$ mass [113] or $e^+e^- \to \mu^+\mu^- \text{invisible}$ at Belle II [116]. In this letter, we analyze these processes together with $B \to K^{(*)}\nu\bar{\nu}$ [117] using a proper treatment of the $Z'$ contribution, including the effects of its large width and the mass of the invisible states it decays to. We assess the possibility that a light $Z'$ can in fact be responsible for the $b \to s\ell^+\ell^-$ anomalies and how this option can

*Electronic address: andreas.crivellin@cern.ch
†Electronic address: claudioandrea.manzari@physik.uzh.ch
‡Electronic address: waltmann@ucsc.edu
§Electronic address: gianluca.inguglia@oeaw.ac.at
¶Electronic address: paul.feichtinger@oeaw.ac.at
**Electronic address: jcamalich@iac.es
We solely include couplings to muons because the ones \( \chi \mathcal{Z} \) be tested in the future with the forthcoming BELLE II 2

As outlined in the introduction, we supplement the SM by a light neutral gauge boson, i.e. a \( Z' \), with a mass below the \( B \)-meson mass scale (\( m_{Z'} \lesssim 6 \) GeV). We will be agnostic about the origin of this new state and simply parametrize its couplings necessary to explain \( b \to s \ell^+\ell^- \) data by the Lagrangian

\[
\mathcal{L}_{Z'} \supset \left( \bar{b} \left( g^{V}_{\mu\mu} \gamma^\mu + g^{A}_{\mu\mu} \gamma^\mu \gamma_5 \right) \mu + g^{L,R}_{sb} \bar{s} \gamma^\mu P_{L,R} b \right) Z'_\mu \, . \tag{1}
\]

We solely include couplings to muons because the ones to electrons are not necessary to explain the \( b \to s \ell^+\ell^- \) anomalies and are experimentally well constrained, in particular when they appear simultaneously with muon couplings. Furthermore, we do not consider couplings to muon neutrinos as they are very stringently constrained by neutrino trident production. Note that this is possible even for left-handed muon couplings because we assume our model to be realized below the EW symmetry breaking scale such that \( SU(2)_L \) invariance is not necessarily obeyed by the \( Z' \) couplings.

Furthermore, in order to achieve the large width necessary to affect multiple \( q^2 = s \) bins in \( b \to s \ell^+\ell^- \) observables such that a good fit to data is possible, we will assume that the \( Z' \) has a large decay rate to invisible final states \( \chi \) with \( m_\chi < m_{Z'}/2 \). As the couplings to \( \bar{s}b \) and \( \bar{\mu}\mu \) turn out to be small, we will assume that the branching ratio to invisible final states is to a good approximation 100%. Furthermore, for specificity \( \chi \) is taken to be a fermion with vectorial couplings to the \( Z' \).

**A. \( b \to s \ell^+\ell^- \)**

Using the standard parametrization of semi-leptonic \( B \) decays, the effect of a light \( Z' \) can be described by a \( q^2 \) dependent contribution to the effective Wilson coefficients

\[
C_{9(10)} = \frac{g_{9(10)}^{V(A)}}{q^2 - m_{Z'}^2 + i m_{Z'} \Gamma_{Z'}} \, .
\]

\[
C'_{9(10)} = \frac{g_{9(10)}^{A(V)}}{q^2 - m_{Z'}^2 + i m_{Z'} \Gamma_{Z'}} \, .
\]

defined at the \( B \) meson scale. Here \( \Gamma_{Z'} \) is the width of the light vector boson which we approximate here to be \( q^2 \) independent. For the phenomenological analysis, we implemented these contributions into \texttt{flavio} \[118\] to perform the global fit of our model to data. This includes e.g. the measurement of the LFU ratios \( R_K \) \[119\], \( R_{K^{(*)}} \) \[120\], \( R_{K_S} \) \[121\] and \( R_{K^{+\to \mu^+\mu^-}} \) \[124\], as well as the branching ratio for \( B_s \to \mu^+\mu^- \) \[122\], the angular observables of \( B \to K^+\mu^+\mu^- \) \[123\] and the branching ratio and angular distribution of \( B_s \to \phi \mu^+\mu^- \) \[126\] which exhibit the most significant deviations from SM predictions.

**B. \( B \to K^{(*)} + \text{invisible} \)**

The most important constraints on \( Z' - b - s \) couplings, in case the \( Z' \) decays dominantly to invisible final states, can be obtained from the processes \( B \to K^{(*)}\nu\bar{\nu} \) measured most precisely at BaBar \[128\] and Belle \[117\]. However, only the latest Belle II analysis \[129\] with the bound

\[
\mathcal{B}(B^+ \to K^+\nu\bar{\nu}) < 4.1 \times 10^{-5} \, ,
\]

provides the necessary \( q^2 \) dependence of the experimental efficiency necessary to easily recast it in terms of the decay \( B^+ \to K^+\chi\chi \).

In the case of a large \( Z' \) width, the branching ratio \( \mathcal{B}(B \to K^{(*)}\chi\chi) \) can be approximated by

\[
\mathcal{B}(B \to K^{(*)}\chi\chi) = \frac{\Gamma_{Z'} \times (\text{factor})}{\Gamma_{Z'}(s)} = \frac{\Gamma_{Z'}(s)}{\Gamma_{Z'}^\text{total}(s)} \approx 1 - \frac{s_{\text{min}}}{s_{\text{max}}} \times \frac{\Gamma_{Z'}^\text{total}(s)}{\Gamma_{Z'}(s)}
\]

with \( s_{\text{min}} = 4m_\chi^2 \), \( s_{\text{max}} = (m_B - m_{K^{(*)}})^2 \) and \( \Gamma_{Z'}(s) = \int_{s_{\text{min}}}^{s_{\text{max}}} ds \Gamma_{Z'}\times (s) \text{BW}(s) \mathcal{B}(B \to K^{(*)}Z')(s) \), with the \( s_{\text{min}} \) and \( s_{\text{max}} \) adjusted such that the \( \Gamma_{Z'}(m_{Z'}/2) \) gives the desired width \( \Gamma_{Z'} \). The reason for keeping the \( s \)-dependence is that it can affect significantly the limits obtained from \( B \to K^{(*)}\nu\bar{\nu} \) searches for large \( m_\chi \).

With the SM predictions for the differential decay width \( d\Gamma(B^+ \to K^+\nu\bar{\nu})/dq^2 \) \[131\], the relevant form factors \[132\] and the experimental efficiency function reported by Belle II \[133\], we can translate Eq. (3) into a limit on our \( Z' \) model, given the masses \( m_{Z'} \) and \( m_\chi \) as well as the width \( \Gamma_{Z'} \). The experimental signal efficiency \[133\] is shown in Fig. [1] together with the form factor, the Breit-Wigner distribution of the \( Z' \) and the squared matrix element of the amplitude (excluding the form factor). The resulting branching ratio is obtained by

---

1 This is relevant for calculating the width of the \( Z' \) as a function of \( q^2 \). However, we checked that this assumption has a minor impact on our final results.

2 The bounds obtained using the full Belle \[117\] and BaBar \[128\] data sets are slightly more stringent than Eq. (3) but are more difficult to use for our purposes (see, however, Ref. \[130\] for a recast of the BaBar limits in terms of a QCD axion).
integrating the product, starting at $s_{\text{min}}$. As the amplitude and the efficiency function increase at small $s = q^2$, the bounds on $g_{ab}$ are stronger in case of a large width compared to a narrow one.

Finally, let us remark that $B 	o K\chi\bar{\chi}$ is only sensitive to the vector current $g^L_{ab} + g^R_{ab}$ such that data from $B 	o K^*\chi\bar{\chi}$ would be required to probe the axial-vector coupling $g^A_{ab} - g^A_{ba}$. However, the former process is sufficient to constraint the NP scenarios needed to explain the $b \to s\ell^+\ell^-$ anomalies as right-handed $bs$ couplings are bounded by the fits to data.

C. $B_s - \bar{B}_s$ mixing

Tree-level exchange of the $Z'$ contributes to $B_s - \bar{B}_s$ mixing. For light $Z'$ masses one can set up an operator product expansion in $m_{Z'}/m_b$ to calculate this new physics contribution to the mixing amplitude and obtain bounds on $g_{ab}$ [130]. However, these limits turn out to be much weaker than the ones from $B \to K\chi\bar{\chi}$. While in principle for higher $Z'$ masses there could be a (close to) resonant enhancement, it is not clear how to calculate these effects reliably and we will therefore not use $B_s - \bar{B}_s$ mixing as a constraint in our analysis.

D. $(g - 2)_\mu$

The anomalous magnetic moment of the muon receives 1-loop corrections from the $Z'$. With the results given e.g. in Ref. [134] we find

$$\Delta a_\mu = \frac{m_\mu^2}{12\pi^2 M^2_{Z'}} \Re \left[ (g^V_{\mu\mu})^2 - 5(g^A_{\mu\mu})^2 \right].$$

This expression has to be compared with the experimental value $[135, 136]$ and the SM prediction $[137]$, resulting in $\Delta a_\mu = \alpha^\text{exp}_\mu - \alpha^\text{SM}_\mu = 251(59) \times 10^{-11}$.

E. $pp \to \mu^+\mu^- (+\text{anything})$

Ref. [113] pointed out that Drell-Yan (DY) searches for muon pairs at the LHC place relevant limits on the parameter space. The $Z'$ can be radiated from the final state muons and significantly modify the di-muon invariant mass distribution close the the $Z$ pole. It is found that for a $Z'$ mass between $1 - 5$ GeV the muon coupling should be smaller than $\approx 0.1$ in case of a dominant branching ratio to invisible.

F. $e^+e^- \to \mu^+\mu^- + \text{invisible}$

The Belle II experiment released a search of invisible $Z'$ decays in the process $e^+e^- \to \mu^+\mu^- + \text{invisible}$ [116] using the commissioning run data. Although limited by the size of the data sample analyzed (276 pb$^{-1}$), 90% confidence level limits on the coupling $g^V_{\mu\mu}$ of the order of $10^{-2} - 10^{-1}$ were obtained. Belle II has also provided sensitivity projections for this model for integrated luminosities up to 50 fb$^{-1}$ [163]; in addition, we obtain projections of the sensitivity up to 5 ab$^{-1}$ by accounting for a scaling factor equivalent to $L^{1/4}$. While Ref. [116] gives bounds on the vectorial coupling, the cross section for $e^+e^- \to \mu^+\mu^- + Z'$ scales as $(g^V_{\mu\mu})^2 + (g^A_{\mu\mu})^2$ and thus can be easily adjusted to the case of other chiralities. Note that the analysis of Ref. [116] was done for a $Z'$ with a narrow width. We therefore recasted the analysis such that it applies to our case with a sizable $Z'$ width by recalculating the expected signal yield in each bin of the original analysis, assuming a Breit-Wigner with $\Gamma_{Z'} = 0.1 M_{Z'}$ (Fig. 3 left) and $\Gamma_{Z'} = 0.15 M_{Z'}$ (Fig. 3 right) convoluted with a Gaussian resolution function for the signal. We then set up a binned likelihood fit and used the profile likelihood ratio method to extract the 90% C.L. intervals.

---

3 This result is based on Refs. [138–157]. The recent lattice result of the Budapest-Marseilles-Wuppertal collaboration (BMWc) for the hadronic vacuum polarization (HVP) [158], on the other hand, is not included. This result would render the SM prediction of $a_\mu$ compatible with experiment. However, the BMWc results are in tension with the HVP determined from $e^+e^- \to$ hadrons data [122,137]. Furthermore, the HVP also enters the global EW fit [159], whose (indirect) determination is below the BMWc result [158]. Therefore, the BMWc determination of the HVP would increase tension in the EW fit [161,162] and we opted for using the community consensus of Ref. [137].
are shown in Fig. 2. In this plot we see that a large
for a 100% branching ratio to undetected final states
experimental sensitivity vanishes.  

a good fit to

$g$

as a benchmark scenario here. Note that if we choose
the limiting cases

For purely vectorial couplings, the bound from

$m_Z^2$, $\Gamma_Z$ and $g_{\mu\mu}^V × g_{\mu\mu}^A$
(with $g_{\mu\mu}^V = -\sqrt{5}g_{\mu\mu}^A$). First of all, note that a width $\approx 15\%$ gives the best fit to data with $\Delta \chi^2 = \chi^2 - \chi^2_{SM} \approx 40$
which is however still smaller than what can be achieved
with heavy NP that give a $q^2$ independent effect in the
same scenario. Furthermore, $\Delta \chi^2$ does not change signi-
ificantly for $0.1 m_Z < \Gamma_Z < 0.2 m_Z$.

In order to minimize the effect in direct searches for
the $Z'$ (i.e. DY and $e^+e^- \rightarrow \mu^+\mu^-$invisible), given
that it provides an explanation to $b \rightarrow s\ell^+\ell^-$ data,
we can assume that $g_{\mu\mu}^V$ takes its maximal value al-
lowed by $B^+ \rightarrow K^+\nu\bar{\nu}$. The resulting regions preferred
by $b \rightarrow s\ell^+\ell^-$ data in the $m_Z^2$ and $g_{\mu\mu}^V = -\sqrt{5}g_{\mu\mu}^A$
plane are shown in Fig. 3. From there, one can see
that the constraints from DY searches at the LHC and
$e^+e^- \rightarrow \mu^+\mu^-$invisible cannot exclude a light $Z'$
explanation of the $b \rightarrow s\ell^+\ell^-$ anomalies, yet. However, the
forthcoming Belle II analysis of $e^+e^- \rightarrow \mu^+\mu^-$invisible
has the potential of excluding a mass below 4 GeV de-
pending on $m_{\chi}$ and the width of the $Z'$. Alternatively,
we can use the $e^+e^- \rightarrow \mu^+\mu^-$invisible to derive an upper
limit on $g_{\mu\mu}^V = -\sqrt{5}g_{\mu\mu}^A$ and show the exclu-
sions from $B^+ \rightarrow K^+\nu\bar{\nu}$ in the $m_Z^2$-$g_{\mu\mu}^V$ as depicted in
Fig. 4 were the $50 \text{fb}^{-1}$ prospects of Belle II have
been used. Note, that for a width of 15% a $Z'$ with
$4 \text{GeV} < m_{Z'} < 4.5 \text{GeV}$ gives a good fit to data and
cannot be excluded due to the vanishing experimental
sensitivity in $B^+ \rightarrow K^+\nu\bar{\nu}$ for a DM mass close to one
half of $m_Z^2$. However, a $Z'$ with such a mass would not
lead to $R(K^{(*)}) > 1$ in the high $q^2$ bins above the $J/\Psi$
resonances.

Performing the $b \rightarrow s\ell^+\ell^-$ fit under these assumptions,
we have three free parameters, $m_Z^2$, $\Gamma_Z$ and $g_{\mu\mu}^V × g_{\mu\mu}^A$

FIG. 2: Contour lines of the bounds on $g_{\mu\mu}^V$ in the $m_Z^2$
$m_{\chi}/m_Z^2$ plane for a $Z'$ width of 10%. The region to the
top-right is not constrained as in this case the experimental
sensitivity vanishes due $s_{\min} = (2m_{\chi})^2$.

III. PHENOMENOLOGY

First of all, as already noted in Ref. [108], a sizable
width for the $Z'$ is necessary such that it does not only
affect a single bin of $P_b^g$, $R(K^{(*)})$ etc. This can be achieved by assuming that the $Z'$ decays dominantly into
invisible final states $\chi$ which at the same time avoids
constraints from searches like $e^+e^- \rightarrow 4\mu$. Recasting
the $B^+ \rightarrow K^+\nu\bar{\nu}$ analysis of Belle II the limits on $g_{\mu\mu}^V$
for a 100% branching ratio to undetected final states
are shown in Fig. 2. In this plot we see that a large
$m_{\chi} \leq m_Z^2/2$ weakens the bound on $g_{\mu\mu}^V$ such that for
$2m_{\chi}^2 \geq 15 \text{GeV}^2$ no limit can be obtained because the
experimental sensitivity vanishes.

Let us now turn to the couplings of the $Z'$ to leptons.
For purely vectorial couplings, the bound from $(g - 2)_\mu$
would be so strong that it would exclude a $Z'$ explanation
of $b \rightarrow s\ell^+\ell^-$. However, the effect vanishes for $g_V = -\sqrt{5}g_A$. As this scenario (i.e. $C_9^{\text{eff}} = -\sqrt{5}C_{10}^{\text{eff}}$)
gives a good fit to $b \rightarrow s\ell^+\ell^-$ data (as any scenario between
the limiting cases $C_9$ and $C_0 = -C_{10}$) we will use it
as a benchmark scenario here. Note that if we choose
$g_V$ slightly bigger, we could account for the tension in
$(g - 2)_\mu$ while leaving the $b \rightarrow s\ell^+\ell^-$ fit unchanged to a
very good approximation.

4 Of course, the actual branching ratio cannot be 100% since de-
cays to muons must be possible where kinematically allowed.
However, as long as $Z'\rightarrow \text{invisible}$ is the dominant decay mode,
the bounds depend weakly on the branching ratio.

IV. CONCLUSIONS AND OUTLOOK

In this letter we pointed out that a light $Z'$ explanation
(with a mass below 4 GeV) of the $b \rightarrow s\ell^+\ell^-$ anomalies
can be confirmed or disproved by combining the forth-
coming Belle II searches for $e^+e^- \rightarrow \mu^+\mu^-$invisible
and $B \rightarrow K^{(*)}\nu\bar{\nu}$. Concerning the latter, it is imperative
to properly take into account the sizable $Z'$ width and
the experimental efficiencies. This endeavour is very
important to limit the number of viable models address-
ing $b \rightarrow s\ell^+\ell^-$, in particular in the absence of a sig-
nal in direct searches. Furthermore, a light $Z'$ is the
only remaining viable NP explanation of $b \rightarrow s\ell^+\ell^-$
for which the high $q^2$ bin (above the charm resonances) in
e.g. $R(K^{(*)})$ could lie above unity (assuming that the
situation in the low $q^2$ bins remains unchanged). While
a light $Z'$ with a mass between $4 \rightarrow 6 \text{GeV}$, that en-
ables the SM amplitude at high $q^2$, cannot be excluded
for $m_{\chi} \approx m_Z^2/2$ due to the limited experimental sen-
sitivity of the $B \rightarrow K^{(*)}\nu\bar{\nu}$ analysis to low energetic $K^{(*)}$,
this gap could be closed in the future, e.g. with a reliable
calculation of $B_s - \bar{B}_s$ mixing for such $Z'$ masses.
FIG. 3: Preferred regions in the $m_{Z'} - g_{\mu\mu}^V$ plane from $b \to s \ell^+ \ell^-$ (whole data-set, green) and the fit to the LFU ratios $R(K)$ and $B_s \to \mu^+ \mu^-$ (red) at the $1\sigma$, $2\sigma$ and $3\sigma$ level for $g_{\mu\mu}^V = -\sqrt{3} g_{\mu\mu}^A$ assuming that $g_{\mu\mu}^V$ takes it maximally allowed value from $B \to K \nu \nu$ for different $Z'$ widths and $\chi$ masses. The regions above the solid lines are excluded by the current DY (cyan) and $e^+ e^- \to \mu^+ \mu^-$ invisible searches (blue) while the dashed lines indicate future sensitivities. Note that a smaller width and a larger $\chi$ mass lead to weaker constraints on the model.

FIG. 4: Preferred regions from $b \to s \ell^+ \ell^-$ ($1\sigma$, $2\sigma$ and $3\sigma$) in the $m_{Z'} - g_{\mu\mu}^A$ plane for the scenario with $g_{\mu\mu}^V = -\sqrt{3} g_{\mu\mu}^A$ assuming that $g_{\mu\mu}^V = -\sqrt{3} g_{\mu\mu}^A$ takes its maximally allowed value allowed by the $50fb^{-1}$ sensitivity of Belle II for $e^+ e^- \to \mu^+ \mu^-$ invisible. The regions above the lines, depending on the width and $m_{\chi}$, can be excluded by future $B \to K \nu \nu$ bounds.

Acknowledgments

We thank Luc Schnell for checking the heavy NP fit for the $g_{\mu\mu}^V = -\sqrt{3} g_{\mu\mu}^A$ scenario. Support by the Swiss National Science Foundation of the Project $B$ decays and lepton flavour universality violation (PP00P21.76884) is gratefully acknowledged by A. Crivellin and C. A. Manzari. G. Inguglia and P. Feichtinger acknowledge support by the European Research Council under the grant agreement No. 947006 - InterLeptons. The research of W. Altmannshofer is supported by the U.S. Department of Energy grant number DE-SC0010107. JMC acknowledges Spanish MINECO grant PGC2018-102016-A-I00 and “Ramón y Cajal” number RYC-2016-20672.

Appendix A

For the decay width of a $B$ meson into $K^{(*)}$ and a vector of invariant mass $s$, the operator product expansion analysis of Refs. [132,164] gives

$$B(B \to KZ') = \frac{(g_{\mu\mu}^V + g_{\mu\mu}^R)^2 f^2(s)}{64\pi s m_B^3} \lambda(m_B^2, m_{K^*}^2, s)^{3/2}, \quad (A1)$$

$$B(B \to K^*Z') = \frac{(g_{\mu\mu}^A + g_{\mu\mu}^R)^2 V^2(s)}{8\pi m_B^2 (m_B + m_{K^*})^2} \lambda(m_B^2, m_{K^*}^2, s)^{3/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A^2(s)(m_B^2 + m_{K^*}^2 + s^2 + 10 s m_{K^*}^2 + 2 m_B^2)}{64\pi m_B^3 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{5/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A_2(s)(m_B^2 - m_{K^*}^2 - s)}{32\pi m_B^2 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{3/2}, \quad (A2)$$

$$B(B \to K^0Z') = \frac{(g_{\mu\mu}^A + g_{\mu\mu}^R)^2 V^2(s)}{8\pi m_B^2 (m_B + m_{K^*})^2} \lambda(m_B^2, m_{K^*}^2, s)^{3/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A^2(s)(m_B^2 + m_{K^*}^2 + s^2 + 10 s m_{K^*}^2 + 2 m_B^2)}{64\pi m_B^3 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{5/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A_2(s)(m_B^2 - m_{K^*}^2 - s)}{32\pi m_B^2 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{3/2}, \quad (A3)$$

$$B(B \to K^0Z') = \frac{(g_{\mu\mu}^A + g_{\mu\mu}^R)^2 V^2(s)}{8\pi m_B^2 (m_B + m_{K^*})^2} \lambda(m_B^2, m_{K^*}^2, s)^{3/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A^2(s)(m_B^2 + m_{K^*}^2 + s^2 + 10 s m_{K^*}^2 + 2 m_B^2)}{64\pi m_B^3 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{5/2} + \frac{(g_{\mu\mu}^R - g_{\mu\mu}^A)^2 A_2(s)(m_B^2 - m_{K^*}^2 - s)}{32\pi m_B^2 m_{K^*}^2 s} \lambda(m_B^2, m_{K^*}^2, s)^{3/2}, \quad (A4)$$
where $f(s)$, $V(s)$, $A_1(s)$ and $A_2(s)$ are the form factor given in Refs. [132, 154] and $\lambda$ is the Källén function

$$\lambda(x, y, z) = x^2 + y^2 + z^2 - 2xy - 2xz - 2yz.$$
