Search for Lepton-Universality Violation in $B^+ \to K^+ \ell^+ \ell^-$ Decays

R. Aaij et al. (LHCb Collaboration)

(Received 25 March 2019; published 13 May 2019)

A measurement of the ratio of branching fractions of the decays $B^+ \to K^+ \mu^+ \mu^-$ and $B^+ \to K^+ e^+ e^-$ is presented. The proton-proton collision data used correspond to an integrated luminosity of 5.0 fb$^{-1}$ recorded with the LHCb experiment at center-of-mass energies of 7, 8, and 13 TeV. For the dilepton mass-squared range $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$ the ratio of branching fractions is measured to be $R_K = 0.846^{+0.090+0.016}_{-0.054-0.014}$, where the first uncertainty is statistical and the second systematic. This is the most precise measurement of $R_K$ to date and is compatible with the standard model at the level of 2.5 standard deviations.

DOI: 10.1103/PhysRevLett.122.191801

Decays involving $b \to s \ell^+ \ell^-$ transitions, where $\ell$ represents a lepton, are mediated by flavor-changing neutral currents. Such decays are suppressed in the standard model (SM), as they proceed only through amplitudes that involve electroweak loop diagrams. These processes are sensitive to virtual contributions from new particles, which could have masses that are inaccessible to direct searches for resonances, even at Large Hadron Collider experiments.

Theoretical predictions for exclusive $b \to s \ell^+ \ell^-$ decays rely on the calculation of hadronic effects, and recent measurements have therefore focused on quantities where the uncertainties from such effects are reduced to some extent, such as angular observables and ratios of branching fractions. The results of the angular analysis of the decay $B^0 \to K^0 \mu^+ \mu^-$ [1–9] and measurements of the branching fractions of several $b \to s \ell^+ \ell^-$ decays [10–13] are in some tension with SM predictions [14–19]. However, the treatment of the hadronic effects in the theoretical predictions is still the subject of considerable debate [20–30].

The electroweak couplings of all three charged leptons are identical in the SM and, consequently, the decay properties (and the hadronic effects) are expected to be the same up to corrections related to the lepton mass, regardless of the lepton flavor (referred to as lepton universality). The ratio of branching fractions for $B \to H \mu^+ \mu^-$ and $B \to H e^+ e^-$ decays, where $H$ is a hadron, can be predicted precisely in an appropriately chosen range of the dilepton mass squared $q^2_{\text{min}} < q^2 < q^2_{\text{max}}$ [31,32]. This ratio is defined by

$$R_H = \frac{\int_{q^2_{\text{min}}}^{q^2_{\text{max}}} \frac{d^3 \Gamma(B \to H \ell^+ \ell^-)}{dq^2} dq^2}{\int_{q^2_{\text{min}}}^{q^2_{\text{max}}} \frac{d^3 \Gamma(B \to H \ell^+ \ell^-)}{dq^2} dq^2},$$

where $\Gamma$ is the $q^2$-dependent partial width of the decay. In the range $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$, such ratios are predicted to be unity with O(1%) precision [33]. The inclusion of charge-conjugate processes is implied throughout this Letter.

The most precise measurements of $R_K$ in the region $1.0 < q^2 < 6.0$ GeV$^2$/c$^4$ and $R_K^\ell$ in the regions $0.045 < q^2 < 1.1$ GeV$^2$/c$^4$ and $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$ have been made by the LHCb collaboration and, depending on the theoretical prediction used, are 2.6 [34], 2.1–2.3, and 2.4–2.5 standard deviations [35] below their respective SM expectations [20,21,33,36–43]. These tensions and those observed in the angular and branching-fraction measurements can all be accommodated simultaneously in models with an additional heavy neutral gauge boson [44–47] or with leptoquarks [48–52].

This Letter presents the most precise measurement of the ratio $R_K$ in the range $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$. The analysis is performed using 5.0 fb$^{-1}$ of proton-proton collision data collected with the LHCb detector during three data-taking periods in which the center-of-mass energy of the collisions was 7, 8, and 13 TeV. The data were taken in the years 2011, 2012, and 2015–2016, respectively. Compared to the previous LHCb $R_K$ measurement [34], the analysis benefits from a larger data sample (an additional 2.0 fb$^{-1}$ collected in 2015–2016) and an improved reconstruction; moreover, the lower limit of the $q^2$ range is increased, in order to be compatible with other LHCb $b \to s \ell^+ \ell^-$ analyses and to suppress further the contribution from $B^+ \to \phi(\to \ell^+ \ell^-)K^+$ decays. The results supersede those of Ref. [34].

Throughout this Letter, $B^+ \to K^+ \ell^+ \ell^-$ refers only to decays with $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$, which are denoted nonresonant, whereas $B^+ \to J/\psi(\to \ell^+ \ell^-)K^+$ decays are...
The simulation is calibrated using data-derived control

ci
des required to form Eq. (2) are taken from simulation. The requirements applied to the resonant and nonresonant
different decay modes are shown in Table I. The selection

Table I. Resonant and nonresonant mode \(q^2\) and \(m_{J/\psi}(K^+\ell^-\ell^-)\) ranges. The variables \(m(K^+\ell^-\ell^-)\) and \(m_{J/\psi}(K^+\ell^-\ell^-)\) are used for nonresonant and resonant decays, respectively.

<table>
<thead>
<tr>
<th>Decay mode</th>
<th>(q^2) [GeV^2/c^2]</th>
<th>(m_{J/\psi}(K^+\ell^-\ell^-)) [GeV/c^2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonresonant (e^+e^-)</td>
<td>1.1–6.0</td>
<td>4.88–6.20</td>
</tr>
<tr>
<td>Resonant (e^+e^-)</td>
<td>6.00–12.96</td>
<td>5.08–5.70</td>
</tr>
<tr>
<td>Nonresonant (\mu^+\mu^-)</td>
<td>1.1–6.0</td>
<td>5.18–5.60</td>
</tr>
<tr>
<td>Resonant (\mu^+\mu^-)</td>
<td>8.68–10.09</td>
<td>5.18–5.60</td>
</tr>
</tbody>
</table>

The measurement requires knowledge of the observed
yield, the efficiency to trigger, reconstruct, and select each
decay mode. The use of this double ratio exploits the fact
that \(J/\psi \rightarrow \ell^+\ell^-\) decays are observed to have lepton-
universal branching fractions within 0.4% [53,54]. Using
Eq. (2) then requires the nonresonant \(B^+ \rightarrow K^+\ell^+\ell^-\) detection efficiency to be known only relative to that of
the resonant \(B^+ \rightarrow J/\psi(\rightarrow e^+e^-)K^+\) decay, rather than the \(B^+ \rightarrow K^+\mu^+\mu^-\) decay. As the detector signatures of each
resonant decay are similar to those of the corresponding
nonresonant decay, systematic effects are reduced and the
precision on \(R_K\) is dominated by the statistical uncertainty.

After the application of selection criteria, which are
discussed below, the four decay modes \(B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+\), \(B^+ \rightarrow J/\psi(\rightarrow e^+e^-)K^+\), \(B^+ \rightarrow K^+\mu^+\mu^-\), and \(B^+ \rightarrow K^+e^+e^-\) are separated from the
background on a statistical basis, using fits to the \(m(K^+\ell^+\ell^-)\) distributions. For the resonant decays, the mass \(m_{J/\psi}(K^+\ell^+\ell^-)\) is computed by constraining the dilepton system to the known \(J/\psi\) mass [54]. This
improves the electron-mode mass resolution (full width at half maximum) from 140 to 24.5 MeV/c^2 and the muon-
mode mass resolution from 30 to 17.5 MeV/c^2. The \(m(K^+\ell^+\ell^-)\) fit ranges and the \(q^2\) selection used for the
different decay modes are shown in Table I. The selection
requirements applied to the resonant and nonresonant
decays are otherwise identical. The two ratios of efficiencies
required to form Eq. (2) are taken from simulation. The
simulation is calibrated using data-derived control
due to the significant bremsstrahlung emission of the
electrons and the different signatures exploited in the
online trigger selection. Systematic uncertainties that
would otherwise affect the calculation of the efficiencies
of the \(B^+ \rightarrow K^+\mu^+\mu^-\) and \(B^+ \rightarrow K^+e^+e^-\) decay modes are suppressed by measuring \(R_K\) as a double ratio of
branching fractions,
Kaons and muons are identified using the output of multivariate classifiers that exploit information from the tracking system, the RICH detectors, the calorimeters, and the muon chambers [56,66–70]. Electrons are identified by matching tracks to electromagnetic calorimeter (ECAL) showers and adding information from the RICH detectors. The ratio of the energy detected in the ECAL to the original energy of the electron is measured. However, if an electron radiates a photon downstream of the dipole magnet, the photon and electron deposit their energy in the same ECAL cells as the electron. For each electron track, a determination of the energy of the photon will not be deposited in the same ECAL cells as the electron. For each electron track, a search is therefore made for ECAL showers around the extrapolated track direction (before the magnet) that are not associated with any other charged tracks. The energy of any such shower is added to the electron energy that is derived from the measurements made in the tracker.

Backgrounds from exclusive decays of $b$ hadrons and the so-called combinatorial background, formed from the reconstructed fragments of multiple heavy-flavor hadron decays, are reduced using selection criteria that are discussed below. The muon modes benefit from superior mass resolution so that a reduced mass range can be used (see Table I). Consequently, the only remaining backgrounds after the application of the selection criteria are combinatorial and, for the resonant mode, from the Cabibbo-suppressed decay $B^+ \to J/\psi \pi^+$, where the pion is misidentified as a kaon. For the electron modes, where a wider mass range is used, significant residual exclusive backgrounds also contribute. Since higher-mass $K^*$ resonances are suppressed in the mass range selected, the dominant exclusive backgrounds for the resonant and nonresonant modes are from partially reconstructed $B^{0,+} \to J/\psi(\to e^+ e^-)K^*(892)^{(0,+)}(\to K^+ \pi^{(-,0)})$ and $B^{0,+} \to K^+(892)^{(0,+)}(\to K^+ \pi^{(-,0)})e^+e^-$ decays, respectively, where the pion is not included in the candidate. At the level of $O(1\%)$ of the $K^+e^+e^-$ signal, there are also exclusive background contributions from $B^+ \to D^0(\to K^{e+e^-})e^+\nu_e$ decays and, at low $m(K^+e^+e^-)$, from the radiative tail of $B^+ \to J/\psi(\to e^+e^-)K^+$ decays. This tail is visible in the distribution of $m(K^+e^+e^-)$ versus $q_T^2$, which is given in the Supplemental Material to this Letter [71].

Cascade backgrounds of the form $H_b \to H_c(\to K^+\ell^-X)\ell^+Y$, where $H_b$ is a beauty hadron ($B^*, B^{0,0}, B_s^0, or \Lambda_b^0$), $H_c$ a charm hadron ($D^{0,+,+}, D_s^{0,+,+}, \Lambda_c^0$), and $X,Y$ are particles that are not reconstructed, are suppressed by requiring that the kaon-lepton invariant mass satisfies the constraint $m(K^+\ell^-) > m_{D^0}$, where $m_{D^0}$ is the known $D^0$ mass [54]. Cascade backgrounds with a misidentified particle are suppressed by applying a similar veto, but with the lepton-mass hypothesis changed to that of a pion (denoted $\ell^- \to \pi$). In the muon case, it is sufficient to reject $K\mu(\to \pi)$ combinations with a mass smaller than $m_{D^0}$. In the electron case, this veto is applied without the bremsstrahlung recovery, i.e., based on only the measured track momenta, and a window around the $D^0$ mass is used to reject candidates. The vetoes retain $97\%$ of $B^+ \to K^+\mu^+\mu^-$ and $95\%$ of $B^+ \to K^+e^+e^-$ decays passing the full selection. The relevant mass distributions are given in the Supplemental Material [71].

Other exclusive $b$-hadron decays require at least two particles to be misidentified in order to form backgrounds. These include the decays $B^+ \to K^+\pi^+\pi^-$ and misreconstructed $B^+ \to J/\psi(\to e^+e^-)K^+$ and $B^+ \to \psi(2S)(\to e^+e^-)K^+$ decays, where the kaon is misidentified as a lepton and the lepton (of the same electric charge) as a kaon. The particle-identification criteria used in the selection render such backgrounds negligible. Backgrounds from decays with a photon converted into an $e^+e^-$ pair are also negligible.

Combinatorial background is reduced using boosted decision tree (BDT) algorithms [72], which employ the gradient boosting technique [73]. For the nonresonant muon mode and for each of the three different trigger categories of the nonresonant electron mode, a single BDT is trained for the 7 and 8 TeV data, and an additional BDT is trained for the 13 TeV data. The same BDTs are used to select the resonant decays. The BDT training uses nonresonant $K^+\ell^+\ell^-$ candidates selected from the data with $m(K^+\ell^+\ell^-) > 5.4 \text{ GeV}/c^2$ as a proxy for the background, and simulated nonresonant $K^+\ell^+\ell^-$ candidates as a proxy for the signal decays. The training and testing is performed using the $k$-folding technique with $k=10$ [74]. The variables used as input to these BDTs are the $p_T$ of the $B^+$, $K^+$ and dilepton candidates, and the minimum and maximum $p_T$ of the leptons, the $B^+$, dilepton and $K^+\chi_{bD}^2$ with respect to the associated PV, where $\chi_{bD}^2$ is defined as the difference in the vertex-fit $\chi^2$ of the PV reconstructed with and without the particle being considered, the minimum and maximum $\chi_{bD}^2$ of the leptons, the $B^+$ vertex-fit quality, the significance of the $B^+$ flight distance, and the angle between the $B^+$ candidate momentum vector and the direction between the associated PV and the $B^+$ decay vertex. The selection applied to the BDT output variables is chosen to maximize the predicted significance of the nonresonant signal yield. The BDT selection reduces the combinatorial background by approximately $99\%$, while retaining $85\%$ of the signal modes. The efficiency of each BDT response is independent of $m(K^+\ell^+\ell^-)$ in the regions used to determine the event yields. After the full selection is
applied, the fraction of signal candidates in each trigger category is consistent with the expectation from simulation.

An unbinned extended maximum-likelihood fit to the \(m(K^+e^+e^-)\) and \(m(K^+\mu^+\mu^-)\) distributions of nonresonant candidates is used to determine \(R_K\). In order to take into account the correlation between the selection efficiencies, the different trigger categories and data-taking periods are fitted simultaneously. The resonant decay mode yields are incorporated as constraints in this fit, such that the \(B^+\rightarrow K^+\mu^+\mu^-\) yield and \(R_K\) are fit parameters. The resonant yields are determined from separate unbinned extended maximum-likelihood fits to the \(m_{J/\psi}(K^+\ell^+\ell^-)\) distributions. For all the mass-shape models described below, the parameters are derived from simulated decays that are calibrated using data control channels.

All four signal modes are modeled by functions with multi-Gaussian cores and power-law tails on both sides of the peak [75,76]. The electron-mode signal mass shapes are described with the sum of three distributions which model whether a bremsstrahlung photon cluster was added to neither, either or both of the \(e^\pm\) candidates. The fraction of signal decays in each of the bremsstrahlung categories is constrained to the value obtained from the simulation.

The shape of the \(B^+\rightarrow J/\psi\pi^+\) background is taken from simulation, while its size is constrained with respect to the \(B^+\rightarrow J/\psi K^+\) mode using the known ratio of the relevant branching fractions [54,77] and efficiencies. In each trigger category, the shape and relative fraction of the background from partially reconstructed \(B^{0,+}\rightarrow K^*(892)^{(0,+)}(\rightarrow K^+\pi(-,0))e^+e^-\) or \(B^{0,+}\rightarrow J/\psi(\rightarrow e^+e^-)\) \(K^*(892)^{(0,+)}(\rightarrow K^+\pi(-,0))\) decays are also taken from simulation. The overall yield of these partially reconstructed decays is left free to vary in the fit, in order to accommodate possible lepton-universality violation in such decays. In the fits to nonresonant \(K^+e^+e^-\) candidates, the shape of the radiative tail of \(B^+\rightarrow J/\psi(\rightarrow e^+e^-)K^+\) decays is taken from simulation and its yield is constrained to the expected value within its uncertainty. In all fits, the combinatorial background is modeled with an exponential function with a freely varying yield and shape.

In order to evaluate the efficiencies accurately, weights are applied to simulated candidates to correct for the imperfect modeling of the \(B^+\) production kinematics, the particle-identification performance, and the trigger response. The weights are computed sequentially, making use of control samples of \(J/\psi\rightarrow \mu^+\mu^-\), \(D^+\rightarrow D^0(\rightarrow K^-\pi^+)\pi^+\), and \(B^+\rightarrow J/\psi(\rightarrow e^+\ell^-)K^+\) decays, and are applied to both resonant and nonresonant simulated candidates. Only subsets of the \(B^+\rightarrow J/\psi(\rightarrow e^+\ell^-)K^+\) samples are used to derive these corrections, which minimizes the number of common candidates being used for both the determination of the corrections and the measurement. The correlations between samples are taken into account in the results and cross-checks presented below. The overall effect of the corrections on the \(R_K\) measurement is at the 0.02 level, demonstrating the robustness of the double-ratio method in suppressing systematic biases that affect the resonant and nonresonant decay modes similarly.

Two classes of systematic uncertainty are considered: those that only affect the nonresonant decay yields, and those that affect the ratio of efficiencies for different trigger categories and data-taking periods in the fit for \(R_K\). The uncertainty from the choice of mass-shape models falls into the former category and is estimated by fitting pseudoexperiments with alternative models that still describe the data well. The effect on \(R_K\) is at the \(\pm0.01\) level. Systematic uncertainties in the latter category affect the ratios of efficiencies and hence the value of \(R_K\) that maximizes the likelihood. These uncertainties are accounted for through constraints on the efficiency values used in the fit to determine \(R_K\), taking into account the correlations between different trigger categories and data-taking periods. The combined statistical and systematic uncertainty is then determined from a profile-likelihood scan. In order to isolate the statistical contribution to the uncertainty, the profile-likelihood scan is repeated with the efficiencies fixed to their fitted values. For the subsamples of the electron-mode data where the trigger is based on the kaon or on other particles in the event that are not part of the signal candidate, the dominant systematic uncertainties come from the (data-derived) calibration of the trigger efficiencies. For the electron trigger, there are comparable contributions from the statistical uncertainties associated with various calibration samples and the calibration of data-simulation differences.

The migration of events in \(q^2\) is studied in the simulation. The effect of the differing \(q^2\) resolution between data and simulation, which alters the estimate of the migration, gives a negligible uncertainty in the determination of the ratio of efficiencies. The uncertainties on parameters used in the simulation decay model (Wilson coefficients, form factors, other hadronic uncertainties, etc.) affect the \(q^2\) distribution and hence the selection efficiencies determined from simulation. The variation caused by the uncertainties on these parameters is propagated to an uncertainty on \(R_K\) using predictions from the FLAVIO software package [42]. The resulting systematic effect on \(R_K\) is negligible, even when non-SM values of the Wilson coefficients are considered.

Several cross-checks are used to verify the analysis procedure. The single ratio \(r_{J/\psi} = B(B^+\rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+)/B(B^+\rightarrow J/\psi(\rightarrow e^+e^-)K^+)\) is known to be compatible with unity at the 0.4% level [53,54]. This ratio does not benefit from the cancellation of systematic effects that the double ratio used to measure \(R_K\) exploits, and is therefore a stringent test of the control of the efficiencies. The corrections applied to the simulation do not force \(r_{J/\psi}\) to be unity and some of the corrections shift \(r_{J/\psi}\) in opposing directions. The value of \(r_{J/\psi}\) is found to be
1.014 ± 0.035, where the uncertainty includes the statistical uncertainty and those systematic effects relevant to the $R_K$ measurement. It does not include additional subleading systematic effects that should be accounted for in a complete measurement of $r_{J/\psi}$. As a further cross-check, the double ratio of branching fractions, $R_K^{(2S)}$, defined by

$$R_K^{(2S)} = \frac{\mathbb{B}(B^+ \to \psi(2S)\to \mu^+\mu^-)K^+}{\mathbb{B}(B^+ \to J/\psi\to \mu^+\mu^-)K^+} / \frac{\mathbb{B}(B^+ \to \psi(2S)\to e^+e^-)K^+}{\mathbb{B}(B^+ \to J/\psi\to e^+e^-)K^+},$$

is determined to be 0.986 ± 0.013, where again the uncertainty includes the statistical uncertainty but only those systematic effects that are relevant to the $R_K$ measurement. This ratio provides an independent validation of the analysis procedure.

Leptons from $B^+ \to J/\psi K^+$ decays have a different $q^2$ value than those from the nonresonant decay modes. However, the detector efficiency depends on laboratory-frame variables rather than on $q^2$, e.g., the momenta of the final-state particles, opening angles, etc. In these laboratory variables there is a significant overlap between the nonresonant and resonant modes, even if the decays do not overlap in $q^2$ (see the Supplemental Material [71]). The $r_{J/\psi}$ ratio is examined as a function of a number of reconstructed variables. Any trend would indicate an uncontrolled systematic effect that would only partially cancel in the double ratio. For each of the variables examined, no significant trend is observed. Figure 1 shows the ratio as a function of the dilepton opening angle and other examples are provided in the Supplemental Material [71]. Assuming the deviations that are observed indicate genuine mismodeling of the efficiencies, rather than fluctuations, and taking into account the spectrum of the relevant variables in the nonresonant decay modes of interest, a total shift on $R_K$ is computed for each of the variables examined. In each case, the resulting variation is within the estimated systematic uncertainty on $R_K$. The $r_{J/\psi}$ ratio is also computed in two- and three-dimensional bins of the considered variables. Again, no trend is seen and the deviations observed are consistent with the systematic uncertainties on $R_K$. An example is shown in Fig. S7 in the Supplemental Material [71]. Independent studies of the electron reconstruction efficiency using control channels selected from the data also give consistent results.

The results of the fits to the $m(K^+\ell^+\ell^-)$ and $m_{J/\psi}(K^+\ell^+\ell^-)$ distributions are shown in Fig. 2. A total of 1943 ± 49 $B^+ \to K^+\mu^+\mu^-$ decays are observed. A study of the $B^+ \to K^+\mu^+\mu^-$ differential branching fraction gives results that are consistent with previous LHCb measurements [12] but, owing to the selection criteria optimized for the precision on $R_K$, are less precise. The $B^+ \to K^+\mu^+\mu^-$ differential branching fraction observed is consistent between the 7 and 8 TeV data and the 13 TeV data.

The value of $R_K$ is measured to be

$$R_K = 0.846^{+0.060+0.016}_{-0.054-0.014},$$

where the first uncertainty is statistical and the second systematic. This is the most precise measurement to date and is consistent with the SM expectation at the level of 2.5 standard deviations [21,33,36,40,42]. The likelihood profile as a function of $R_K$ is given in the Supplemental Material [71]. The value for $R_K$ obtained is consistent across the different data-taking periods and trigger categories. A fit to just the 7 and 8 TeV data gives a value for $R_K$ compatible with the previous LHCb measurement [34] within one standard deviation. This level of consistency is evaluated using pseudoexperiments that take into account the overlap between the two data samples, which are not identical due to different reconstruction and selection procedures. The result from just the 7 and 8 TeV data is
The dominant systematic uncertainty is from the limited knowledge of the $B^+ \to J/\psi K^+$ branching fraction \cite{54}. This is the most precise measurement to date and is consistent with predictions based on the SM \cite{42,78}.

In summary, in the dilepton mass-squared region $1.1 < q^2 < 6.0$ GeV$^2$/c$^4$, the ratio of the branching fractions for $B^+ \to K^+ \mu^+ \mu^-$ and $B^+ \to K^+ e^+ e^-$ decays is measured to be $R_K = 0.846^{+0.060}_{-0.054}$. This is the most precise measurement of this ratio to date and is consistent with the SM prediction at the level of 2.5 standard deviations. Further reduction in the uncertainty on $R_K$ can be anticipated when the data collected by LHCb in 2017 and 2018, which have a statistical power approximately equal to that of the full data set used here, are included in a future analysis. In the longer term, there are good prospects for high-precision measurements as much larger samples are collected with an upgraded LHCb detector \cite{79}.

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ, and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG, and MPG (Germany); INFN (Italy); NWO (Netherlands); MNI (Poland); MEN/IFA (Romania); MSHE (Russia); IndENCo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); and NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland), and OSC (USA). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany);
EPLANET, Marie Sklodowska-Curie Actions, and ERC (European Union); ANR, Labex P2IO, and OCEUV, and Région Auvergne-Rhône-Alpes (France); Key Research Program of Frontier Sciences of CAS, CAS PIFI, and the Thousand Talents Program (China); RFBR, RSF, and Yandex LLC (Russia); GVA, XuntaGal, and GENCAT (Spain); the Royal Society and the Leverhulme Trust (United Kingdom); Laboratory Directed Research and Development program of LANL (USA).


[3] B. Aubert et al. (BABAR Collaboration), Measurements of branching fractions, rate asymmetries, and angular distributions in the rare decays $B \rightarrow K\ell^+\ell^-$ and $B \rightarrow K^\ast\ell^+\ell^-$, Phys. Rev. D 73, 092001 (2006).


[8] V. Khachatryan et al. (CMS Collaboration), Angular analysis of the decay $B^0 \rightarrow K^0\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV, Phys. Lett. B 753 (2016) 424.


[26] R. R. Horgan, Z. Liu, S. Meinel, and M. Wingate, Calculation of $B^0 \rightarrow K^0\mu^+\mu^-$ and $B_s^0 \rightarrow \phi\mu^+\mu^-$ observables using form Factors from Lattice QCD, Phys. Rev. Lett. 112, 212003 (2014).


[32] Y. Wang and D. Atwood, Rate difference between $b \to s \mu^+ \mu^-$ and $b \to s e^+ e^-$ in supersymmetry with large tan $\beta$, Phys. Rev. D 68, 094016 (2003).


[34] R. Aaij et al. (LHCb Collaboration), Test of Lepton Universality using $B^+ \to K^+ \ell^+ \ell^-$ Decays, Phys. Rev. Lett. 113, 151601 (2014).

[35] R. Aaij et al. (LHCb Collaboration), Test of lepton universality with $B^0 \to K^0 \ell^+ \ell^-$ decays, J. High Energy Phys. 08 (2017) 055.


[45] A. Crivellin, G. D’Ambrosio, and J. Heeck, Explaining $h \to \mu^+ \mu^-$, $B \to K^+ \mu^+ \mu^-$ and $B \to K^+ \ell^+ \ell^-/B \to K^+ e^+ e^-$ in a two-Higgs-Doublet Model with Gauged $L_\mu - L_\tau$, Phys. Rev. Lett. 114, 151801 (2015).


[53] M. Ablikim et al. (BESIII Collaboration), Precision measurements of $B[\psi(3686) \to \pi^+ \pi^- J/\psi]$ and $B[J/\psi \to \ell^+ \ell^-]$, Phys. Rev. D 88, 032007 (2013).


[77] R. Aaij et al. (LHCb Collaboration), Measurement of the ratio of branching fractions and difference in CP asymmetries of the decays $B^+ \rightarrow J/\psi \pi^+$ and $B^0 \rightarrow J/\psi K^0$, J. High Energy Phys. 03 (2017) 036.

[78] A. Khodjamirian and A. V. Rusov, $B_s \rightarrow K\ell\nu_\ell$ and $B^0(\ell) \rightarrow \pi K' e^- \bar{\nu}_e$ decays at large recoil and CKM matrix elements, J. High Energy Phys. 08 (2017) 112.

\textsuperscript{74}National Research University Higher School of Economics, Moscow, Russia (associated with Institution Yandex School of Data Analysis, Moscow, Russia) 
\textsuperscript{75}National Research Tomsk Polytechnic University, Tomsk, Russia [associated with Institution Institute of Theoretical and Experimental Physics NRC Kurchatov Institute (ITEP NRC KI), Moscow, Russia, Moscow, Russia] 
\textsuperscript{76}Instituto de Física Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain (associated with Institution ICCUB, Universitat de Barcelona, Barcelona, Spain) 
\textsuperscript{77}University of Michigan, Ann Arbor, USA (associated with Institution Syracuse University, Syracuse, New York, USA) 
\textsuperscript{78}Los Alamos National Laboratory (LANL), Los Alamos, USA (associated with Institution Syracuse University, Syracuse, New York, USA) 

\textsuperscript{a}Deceased. 
\textsuperscript{b}Also at Laboratoire Leprince-Ringuet, Palaiseau, France. 
\textsuperscript{c}Also at Università di Milano Bicocca, Milano, Italy. 
\textsuperscript{d}Also at Università di Bologna, Bologna, Italy. 
\textsuperscript{e}Also at Università di Modena e Reggio Emilia, Modena, Italy. 
\textsuperscript{f}Also at Novosibirsk State University, Novosibirsk, Russia. 
\textsuperscript{g}Also at Università di Ferrara, Ferrara, Italy. 
\textsuperscript{h}Also at LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain. 
\textsuperscript{i}Also at H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom. 
\textsuperscript{j}Also at Università di Bari, Bari, Italy. 
\textsuperscript{k}Also at Sezione INFN di Trieste, Trieste, Italy. 
\textsuperscript{l}Also at Università di Genova, Genova, Italy. 
\textsuperscript{m}Also at Università degli Studi di Milano, Milano, Italy. 
\textsuperscript{n}Also at Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil. 
\textsuperscript{o}Also at AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland. 
\textsuperscript{p}Also at Lanzhou University, Lanzhou, China. 
\textsuperscript{q}Also at Università di Padova, Padova, Italy. 
\textsuperscript{r}Also at Università di Cagliari, Cagliari, Italy. 
\textsuperscript{s}Also at MSU - Iligan Institute of Technology (MSU-IIT), Iligan, Philippines. 
\textsuperscript{t}Also at Scuola Normale Superiore, Pisa, Italy. 
\textsuperscript{u}Also at Hanoi University of Science, Hanoi, Vietnam. 
\textsuperscript{v}Also at P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia. 
\textsuperscript{w}Also at Università di Roma Tor Vergata, Roma, Italy. 
\textsuperscript{x}Also at Università di Roma La Sapienza, Roma, Italy. 
\textsuperscript{y}Also at Università della Basilicata, Potenza, Italy. 
\textsuperscript{z}Also at Università di Urbino, Urbino, Italy. 
\textsuperscript{aa}Also at Physics and Micro Electronic College, Hunan University, Changsha City, China. 
\textsuperscript{ab}Also at School of Physics and Information Technology, Shaanxi Normal University (SNNU), Xi’an, China.